Essays in development economics and public finance

Hoseini, Mohammad

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Essays in Development Economics and Public Finance

Mohammad Hoseini

4 November 2015

Essays in Development Economics and Public Finance

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof.dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op woensdag 4 november 2015 om 14.15 uur door

Mohammad Hoseini

geboren op 11 juli 1984 te Gonbad Kavous, Iran.

Promotiecommissie:

Promotor:	prof. dr. Thorsten Beck
Copromotor:	prof. dr. Jan Boone
Overige Leden:	
	prof. dr. Erwin Bulte
	dr. Benedikt Goderis
	prof. dr. Manuel Oechslin
	prof. dr. Harrie Verbon

Acknowledgements

First, I want to express my gratitude to Thorsten Beck. I started working with him in 2009 and I never forget all of his support during my research master and PhD. I am happy that we are still working together and I hope this collaboration continues in the future.

I also want to express my gratitude to Jan Boone who was always patient and helpful to me. I learned a lot from discussions with him, especially in last two chapters of my dissertation.

My thanks also go to Jenny Ligthart, my kind supervisor, who passed away after the first year of my PhD. She was the one who entered me to the world of public finance.

I would also like to thank Meghana Ayyagri for sharing research ideas and contributing to the first chapter of my dissertation.

I am also grateful to Erwin Bulte, Benedikt Goderis, Manuel Oechslin, and Harrie Verbon for serving as my committee members.

Many thanks to all of my friends during PhD, especially Ali Motavasseli for 14 years of friendship and dr. Ehsan van Mehdad who I have known since my first day in Tilburg.

Words cannot express how indebted I am to my mother, my father, and my dear sisters for all sacrifices that they have made for me and all prayers that sustained me thus far.

Finally but foremost, I would like express my greatest appreciation to my beloved wife Mahsa who was always there with her beautiful smile.

Contents

Chapter 1

Introduction

This dissertation revolves around two main questions: How does financial development affect poverty and informality? How can the government reduce tax evasion in a value-added tax system? The first two chapters contain empirical studies on India addressing the first question. Using time and state-level variation across Indian states, chapter 2 examines the effect of financial liberalization in 1991 on poverty and investigates the underlying mechanisms. Chapter 3 studies the effect of financial deepening and bank outreach on informality using micro data from the Indian manufacturing sector. For answering the second question, my approach uses both theory and empirical examination to find out the optimal enforcement strategies for minimizing value-added tax evasion at the intensive and extensive margins. Chapter 4 addresses the problem of misreporting by the registered traders. Chapter 5 models the role of inter-sectoral linkages on tax evasion in an input-output framework and confirms the results using Indian data. The remainder of this introduction sets out the structure of the dissertation by a brief review of each chapter.

Financial liberalization has been controversial among academics and policymakers as it is not clear whom the benefits of expanded credit allocation accrue to. Using state-level data from India over the period 1983 to 2005, Chapter 2 gauges the effect of financial deepening and outreach on rural poverty. Following the 1991 liberalization episode, we find a strong negative relationship between financial deepening, rather than financial inclusion, and rural poverty. Instrumental variable regressions suggest that this relationship is robust to omitted variable and endogeneity biases. We also find that financial deepening has reduced poverty rates especially among self-employed in the rural areas, while at the same time it supported an inter-state migration trend from rural areas into the tertiary sector in urban areas, consistent with financial deepening being driven by credit to the tertiary sector. This suggests that financial deepening contributed to poverty alleviation in rural areas by fostering entrepreneurship and inducing geographic-

sectoral migration.

Chapter 3 gauges the effect of financial deepening and bank outreach on informality using micro data from the Indian manufacturing sector and exploiting cross-industry variation in the need for external finance. In this chapter, we distinguish between two channels through which access to finance can reduce informality: reducing the entry barrier to the formal sector and increasing productivity of formal firms. We find that bank outreach has a stronger effect on reducing the incidence of informality by cutting barriers to entering the formal economy, especially for smaller firms, and thus diminishing opportunistic informality. In comparison, financial deepening increases the productivity of formal sector firms while it has no significant impact on informal sector firms.

The last two chapters concern tax evasion in the value-added tax (VAT) system. Essentially, the VAT has an intrinsic third-party reporting feature making inter-firm transactions less vulnerable to fraud and thus its enforcement design has additional considerations. In a general framework, VAT evasion can be classified into two forms. At the intensive margin, the registered trader under-reports the sale or over-reports the purchase (misreporting fraud). At the extensive margin, the informal firm fails to register and is hidden from the government (informality fraud). In high-income countries, the major loss in VAT revenue is due to misreporting, while in developing countries, both types of fraud seem to be extensive.

Chapter 4 looks into the misreporting fraud by linking the level of evasion to the degree of convexity of the cost function of taxpayers and the level of transactions with final consumers. In addition, it analyses the enforcement consequences of the new developments in information reporting and electronic invoicing, which enable the tax authority to randomly cross-check the invoices. The results highlight the importance of taxpayer's subjective beliefs in shaping audit policy of the tax authority. The optimal audit rate for firms with low cost convexity is an increasing function of transaction with final consumers, but this relationship may turn to be negative when the cost function becomes very convex. Moreover, the optimal level of invoice cross-checking on transactions of each commodity is positively associated with the number of trading firms.

Chapter 5, in comparison, assumes no misreporting and addresses informality fraud at a macro perspective by analyzing inter-sectoral linkages. Specifically, this chapter models and empirically tests the self-enforcing feature of the VAT which is absent in the

theoretical literature: An incentive that makes formal traders report their purchases to the government for a deduction in their VAT bill. In addition, it explores how the government can deploy this feature to reduce the size of shadow economy in the VAT system by reallocating enforcement type and spending among different sectors. The results suggest that the government should identify informal firms more strictly in the backwardly linked sectors – which buy their inputs from the others – and focus on revealing within-firm information. In contrast, in forwardly linked industries, the government should concentrate on cross-checking the input credit claims it receives by the corresponding VAT payments. Empirical evidence from Indian service sector enterprises suggests a significant increase in formality following the VAT adoption episode in 2003. This increment is positively correlated with the sector's forward linkage implying the existence of the self-enforcement effect of the VAT.

Although the last two chapters both study VAT evasion, they have differences in terms of objectives and assumptions. While chapter 4 addresses the intensive margin from a micro perspective, chapter 5 focuses on the extensive margin from a macro perspective. In the misreporting fraud, the government is aware about the existence of the firm, but in the informality fraud, informal firms are hidden. As a result, the policy recommendation of chapter 4 is based on firm's characteristics, while it depends on sectoral linkages – not firm-level factors – in chapter 5. Moreover, in chapter 5, the production decision of informal sector is affected by the government policy and there is an optimal level of enforcement based on cost-benefit analysis. Chapter 4, however, assumes that production decision of a single firm is independent of the audit rate and they take their tax obligation as given when deciding how much to misreport. This is a plausible assumption for the registered traders in the VAT, since it is a tax on final consumers and does not distort registered firm's profit maximization.

Chapter 2

Finance and Poverty: Evidence from India [1]

2.1. Introduction

Finance as a fundamental driver of economic growth has been largely accepted after several decades of research in this area.[2] The debate today has shifted to the multifaceted nature of financial development, specifically on the role of financial depth versus access. While financial deepening has accelerated in emerging markets, it has not always been accompanied by increased use of financial services. Previous empirical evidence has shown that financial deepening fosters economic growth and reduces income inequality (Beck, Levine, and Levkov, 2010; Bruhn and Love, 2014), but the effects of financial access are less understood, even as financial inclusion is being adopted as a top development priority by policymakers worldwide.

This paper contributes to a better understanding of the role of financial access versus depth by using annual household census data from India over the period 1983 to 2005. Specifically, we exploit geographic and time variation in both financial depth (commercial bank credit to SDP), and financial inclusion (bank branch penetration), to explore the relative importance of financial depth and inclusion on changes in rural and urban poverty and to explore the channels and mechanisms through which financial development alleviates poverty.

There are two novel components to our empirical design. First, India offers the perfect landscape to examine these issues because it has a long history of implementing policies targeting financial breadth and has recently become the poster child for financial inclusion with the Prime Minister making a bank account for each household a national

[1] This chapter is coauthored with Meghana Ayyagri and Thorsten Beck.
[2] See Levine (2005) for a review.

priority.[3] Furthermore there is large sub-national variation in socio-economic and institutional development, and significant policy changes over the sample period (Besley, Burgess, Esteve-Volart, and Louise, 2007). By focusing on a specific country, using data from a consistent data source and exploiting pre-determined cross-state variation in socio-economic conditions, we alleviate problems associated with cross-country studies, including measurement error, omitted variable and endogeneity biases.

Second, we incorporate the policy changes in our empirical design to address endogeneity concerns. First, we follow Burgess and Pande (2005) and exploit the policy driven nature of rural bank branch expansion across Indian states as an instrument for branch penetration and thus financial breadth. Next we exploit the important watchdog function of the relatively free and independent press in India (Besley and Burgess, 2002), which has repercussions for corporate finance and governance, and ultimately financial sector competition. We use the large cross-state variation in national English-language newspaper penetration that deepens after India's liberalization in 1991 as an instrument for financial depth.[4]

We find that financial depth has a negative and significant impact on rural poverty in India over the period 1983-2005. This is robust to using different measures of rural poverty, controlling for time-varying state characteristics, and state and year fixed effects. We find no effect of financial depth on urban poverty rates. The effect of financial depth on rural poverty reduction is also economically meaningful. One standard deviation in Credit to SDP (within-state, within-year) explains 17 percent of demeaned variation in the proportion of the population below the poverty line (Headcount ratio). We also find that over the time period 1983-2005, financial depth has a more significant impact on poverty reduction than financial inclusion. Our measure of financial inclusion, rural branches per capita, has a negative but insignificant effect on rural poverty over this period.

Our micro-data also allows us to explore different channels identified by theory through which financial development lowers rural poverty. On the one hand, better access to credit enables the poor to pull themselves out of poverty by investing in their human

[3] On August 28, 2014, the Prime Minister of India launched Jan Dhan Yojana, a national campaign for financial inclusion under which 18 million bank accounts were opened during the first week alone.

[4] English language newspaper circulation is highly correlated with economic newspaper circulation. In placebo tests, we find no association between local language newspaper penetration and financial depth. More details on the instruments are presented in sections 2.2.2.

capital and microenterprises, thus reducing aggregate poverty (Aghion and Bolton, 1997; Banerjee and Newman, 1993; Galor and Zeira, 1993). On the other hand, more efficient resource allocation by the financial sector (not necessarily to the poor, though), will benefit especially the poor if – as a result – they are included in the formal labor market. We find evidence for the entrepreneurship channel, as the poverty-reducing impact of financial deepening falls primarily on self-employed in rural areas. We also identify migration from rural to urban areas as an important channel through which financial depth reduces rural poverty. In particular, there is inter-state migration of workers for employment reasons towards financially more developed states, suggesting that poorer population segments in rural areas migrated to urban areas. The rural primary and tertiary urban sectors benefited most from this migration, consistent with evidence showing that the Indian growth experience has been led by the services sector rather than labor intensive manufacturing (Bosworth, Collins, and Virmani, 2007). We also find that it is specifically the increase in bank credit to the tertiary sector that accounts for financial deepening post-1991 and its poverty-reducing effect.

This paper contributes to the recent literature on the role of financial sector development on poverty reduction. Theory makes contradictory predictions about which income group should benefit most from financial sector deepening. Some studies argue that credit constraints are particularly binding for the poor (Aghion and Bolton, 1997; Banerjee and Newman, 1993; Galor and Zeira, 1993) and that finance helps overcome barriers of indivisible investment (McKinnon, 1973). Other studies have claimed that only the rich can pay the "entry fee" into the financial system (Greenwood and Jovanovic, 1990) and credit is channeled to incumbents, not to entrepreneurs with the best opportunities (Lamoreaux, 1986). In a cross-country setting, Beck et al. (2010) find that banking sector development reduces income inequality and poverty.[5] By contrast our paper looks at the effect of financial sector development and rural poverty in a single country setting allowing us to better address identification issues. Furthermore, we study the impact of both financial depth and inclusion on poverty and find that financial depth has a greater

[5] Other cross-country studies have studied the relationship between financial development and the level of income inequality. Li, Squire, and Zou (1998) and Li, Xu, and Zou (2000) find a negative relationship between finance and the level of income inequality as measured by the Gini coefficient, a finding confirmed by Clarke, Xu, and Zou (2006), using both cross-sectional and panel regressions and instrumental variable methods. Honohan (2004) shows that even among societies with the same average income, those with deeper financial systems have lower absolute poverty.

impact on poverty reduction than financial inclusion. Most other papers only look at the impact of either financial depth or inclusion (e.g. Beck et al., 2010; Bruhn and Love, 2014; Burgess and Pande, 2005). Our findings also contribute to the literature on the channels through which finance should affect income equality and poverty ratios. Giné and Townsend (2004) find for Thailand that financial liberalization benefited would-be entrepreneurs but also resulted in wage increases through higher labor demand. Consistent with this, Beck et al. (2010) find that the main effect of branch deregulation in the United States on income inequality was through the indirect effects of higher labor demand and higher wages for lower income groups. Our paper finds that financial sector development reduces rural poverty in India both by fostering entrepreneurship in rural areas and by facilitating migration of workers from rural secondary and tertiary sectors to the urban tertiary sector.

Our paper also links to an increasing literature in banking and corporate finance focusing on the role of media, in general, and the business press, in corporate finance and governance.[6] The media can serve as a watch dog for both banks and publicly held companies as an information intermediary and provider of additional analysis. Several papers have shown that media pressures can entice firms to change governance structures and business strategies (Bednar, 2012; Joe, Louis, and Robinson, 2009; Kuhnen and Niessen, 2012) and can have an impact on firms' financing structure and cost (Bushman, Williams, and Wittenberg-Moerman, 2013). While helping to reduce information asymmetries between investors, on the one hand, and enterprises and banks, on the other hand, it can also have an important impact on the competitive environment in the financial and real sector. Specifically, by more information about specific firms but also industries and sectors, in general, media can help reduce private information and control rents, with a positive impact on the overall investment environment and resource allocation in a country. Dyck and Zingales (2004) show that private control benefits of majority shareholders are lower in countries with higher press penetration and thus higher media pressure, while Perotti and Volpin (2014) show that newspaper penetration explains cross-country variation in investor protection. Beck and Demirgüç-Kunt (2008) show in a cross-country setting that a free media is correlated with lower barriers to

[6] A recent finance literature has established the importance of the information content in news media for trading in financial markets (Engelberg and Parsons, 2011) and stock market returns (Griffin, Hirschey, and Kelly, 2011; Tetlock, 2007; Tetlock, Saar-tsechansky, and Macskassy, 2008).

financial inclusion. Our paper adds to this micro- and macro-level literature by linking media penetration to financial deepening across Indian states.

Finally, our paper also adds to a flourishing literature on economic development in India, which has linked sub-national variation in historic experiences and policies to differences in growth, poverty levels, political outcomes and other dependent variables (see Besley et al., 2007 for an earlier survey). Specifically, researchers have focused on differences in political accountability (Besley and Burgess, 2002; Pande, 2003), labor market regulation (Besley and Burgess, 2004; Dougherty, Robles, and Krishna, 2011; Hasan, Mitra, and Ramaswamy, 2007), land reform (Besley and Burgess, 2000), trade liberalization (Edmonds, Pavcnik, and Topalova, 2010; Topalova, 2010) and gender inequality (Iyer, Mani, Mishra, and Topalova, 2012). Directly related to our paper, Burgess and Pande (2005) relate a social banking policy on branching to differences in poverty alleviation across states. Our paper adds to this literature by focusing on cross-state differences in financial deepening after the 1991 liberalization episode and by comparing the effect of two different dimensions of financial development – total credit volume and branch penetration of financial institutions.

The remainder of the paper is organized as follows. Section 2.2 presents data and methodology. Section 2.3 discusses our main results, documenting the relationship between financial development and poverty using both OLS and IV regressions. Section 2.4 explores different channels through which finance affects poverty. Section 2.5 concludes.

2.2. Data, methodology, and summary statistics

In this section, we describe the data sources from which we construct our measures of financial development and poverty, present summary statistics, and discuss the empirical research design used for examining the relationship between finance and the poverty. Table 2.1 presents the descriptive statistics for the poverty measures, the financial development indicators and the control variables. Panel A presents the summary statistics for the whole of India while Panel B presents a state-wise breakdown. In Panel A, we present mean, standard deviation as well as cross-state, cross-time and within-state-within-time standard deviations.

2.2.1. Data and descriptive statistics

We construct poverty measures across 15 Indian states[7] covering 95% of India's population, using 20 rounds of the Indian household expenditure surveys. The Indian National Sample Survey Organization (NSSO) has been conducting Consumer Household Expenditure surveys since the 1950s, eliciting detailed household level information on household characteristics such as household size, education, socio-religious characteristics, demographic characteristics of household members and detailed expenditure patterns. Our panel dataset extends from 1983 to 2005 and builds on the state-level aggregates, complemented by data provided in Özler, Datt, and Ravallion (1996). In robustness tests for our baseline regressions, we also use data for the period 1965 to 2005.[8]

We construct two measures of poverty. First, Headcount is the proportion of the population below the poverty line, as defined by the Planning Commission (1993)[9] and adjusted yearly by price increases, and measures the incidence of poverty. Second, Poverty Gap is the mean distance separating the poor population from the poverty line as a proportion of poverty line. The calculation process of the poverty measures is described in detail in the data appendix. We compute Headcount and Poverty Gap separately for rural and urban areas.[10] Figure 2.1 charts the average evolution of the Rural and Urban Headcount ratios across the 15 states in our sample. The overall pattern suggests that both measures of poverty declined over the sample period except for sharp fluctuation in the early 1990s following economic liberalization.

Table 2.1 shows that mean Rural Headcount in our sample period is 31.9 percent and larger than the corresponding Urban Headcount of 25.9 percent. While there is a large variation in both rural and urban poverty levels across states and over time, there is a smaller, although significant, variation within states over time. State level summary

[7] The states are: Andhra Pradesh, Assam, Bihar, Gujarat, Haryana, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh and West Bengal. They contained 95.5% of Indian population in the 2001 nationwide census. Where states split during the sample period, we continued to consider them as one unit, using weighted averages for variables, with population shares being the weights.

[8] Detailed household survey data are not available before 1983 and we can therefore not run the channel regressions of section 2.4 over longer time periods.

[9] We test the robustness of our results to the new poverty line measures suggested by the Tendulkar Committee of the Planning Commission of India. See data appendix for details.

[10] The poverty line and price indices differs between rural and urban areas. Consistent with Topalova (2010), we adjusted the measures for the schedule change in the survey. In addition, we controlled for the seasonality bias due to different timing of the surveys. See data appendix for details.

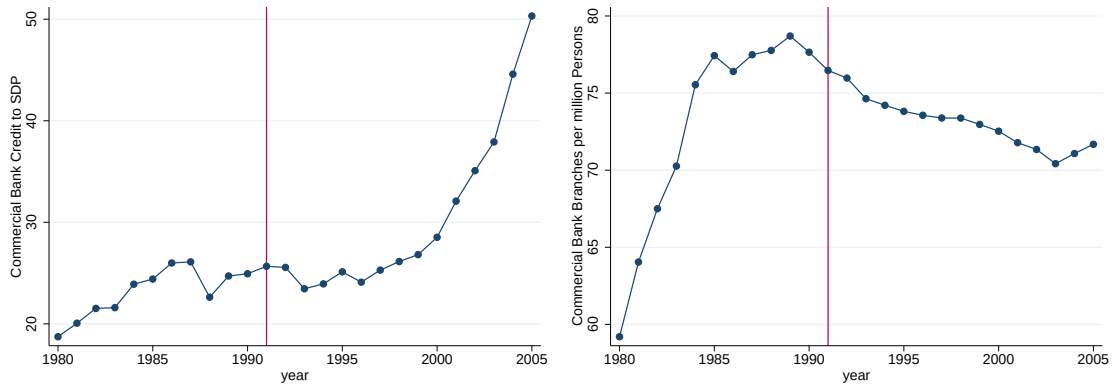
Figure 2.1: Rural and urban poverty in India over time –This figure shows the trend in Rural and Urban Headcount ratios in India. Rural and Urban Headcount ratios are the percentage of rural and urban population with monthly per capita expenditure less than the official poverty line respectively. The vertical line represents the starting year (1991) of financial liberalization. The definitions and sources of all variables are in the appendix.

statistics show that the mean Rural Headcount varies from 14.1 percent in Punjab to 49.5 percent in Bihar. We find Punjab to also have the lowest Urban Headcount of 9.8 percent[11] while the highest Urban Headcount is in the state of Orissa with 37.9 percent. In most states, we find urban poverty numbers to be lower than rural poverty except in the case of Andhra Pradesh, Uttar Pradesh and Orissa. Assam in particular looks unique given the large gap in the percentage of people below the poverty line in rural areas (37.4 percent) compared to urban areas (11.6 percent). The average Rural Poverty Gap in India is 7.5 percent and varies from 2.4 percent in Punjab to 12.6 percent in Bihar. The Urban Poverty Gap varies from 1.9 percent in Punjab to 10.6 percent in Orissa with an all-India average of 6.5 percent.

We use two different indicators of financial development at the state level, with underlying data from the Reserve Bank of India. The first indicator, Credit to SDP, is the ratio of total commercial bank credit outstanding to the Net State Domestic Product and gauges the depth of financial development. The second indicator of financial development is Branches per Capita, which is the total number of operating bank branches per million persons in each state and is a measure of the extent of financial penetration.

[11]Historically the Punjab-Haryana region has been one of the richest regions in the country.

Table 2.1 shows that the standard deviation of both measures over time is higher than that across states, reflecting the upward trend in depth and trend reversal in inclusion over the sample period. Commercial Bank Credit to SDP varies from 11.0 percent in Assam to 58.5 percent in Maharashtra with a national average of 27 percent. Figure 2.2a shows an upward trend of commercial bank credit over the sample period. On average across the 15 states, commercial bank credit increased from 18.7 percent of SDP in 1980 to 50.3 percent in 2005. In our sample, Punjab has the highest number of branches per million people (112) compared to Assam which has fewer than 50 branches per million people. Figure 2.2b illustrates the evolution of branch opening per capita in India. The data show trend breaks around 1990, which may be attributed to the suspending of the 1:4 branch license rule in 1990 according to which commercial banks were required to open 4 new branches in previously unbanked locations for every branch opening in an already banked location.



(a) Credit to SDP

(b) Bank branches per capita

Figure 2.2: Financianl development in India over time —This figure shows the trend in the ratio of total commercial bank credit outstanding to net state domestic product and the ratio of commercial bank branches over population (in million). Commercial bank credit comprises term loans, cash credit, overdrafts and bills purchased and discounted. The rural branch expansion program was in place up to 1989. The vertical line represents the starting year (1991) of financial liberalization. The definitions and sources of all variables are in the appendix.

In investigating the relationship between financial sector development and poverty, we will control for several other time-varying state characteristics. The data appendix details sources and provides extensive definitions. Specifically, we include the following variables: SDP per capita, which is net state domestic product per capita and a proxy for income levels, Rural Population Share, which is rural share of total population in each state, Literacy Rate, which is defined as proportion of persons who can both read and

write with understanding in any language among population aged 7 years and above, and State Government Expenditure to SDP defined as total state government expenses over SDP. As panel B of Table 2.1 shows, there is great variation in income levels across states with SDP per capita ranging from 3,509 Rupee in Bihar to 14,968 Rupee in Punjab, with a country-level mean of 8,781. The mean rural population share is 74 percent and ranges from 88.5 percent in Assam to 60.6 percent in Maharashtra showing that over 60 percent of the population in all states live in rural areas. The mean literacy rate in the country is 56 percent and average government expenditures are 19.3 percent of SDP.

Table 2.2 presents correlations between our main variables of interest and the control variables. The incidence and depth of poverty are highly correlated in both rural and urban areas (correlation coefficient ≥ 0.96), but we also find a significant correlation between the different rural and urban poverty measures: states with higher rural poverty also tend to have higher urban poverty. We also find that both measures of financial development are positively correlated with each other, with a correlation coefficient of 40.5%, and a negative correlation between the measures of financial development and rural and urban poverty measures. The only association that is not significant is between Urban Poverty Gap/Headcount and Credit to SDP. When we look at the control variables we find that states with higher SDP per capita, greater government expenditures to SDP, higher literacy rates and smaller rural populations have lower rural and urban poverty and greater financial development. Critically, there is a high negative correlation between the rural population share and Credit to SDP. All of our regression results are robust to dropping rural population share. Similarly, we also run all of our regressions without SDP per capita given the relatively high correlation with Credit to SDP and again all our results hold without SDP per capita.

Table 2.1: Summary statistics – The mean and standard deviation (in parentheses) of the main variables across all of India and states over the period 1983-2005. For all India one additional standard deviations are measured: within state and year which is standard deviation of $(x_{sy} - m_s - m_y)$ where m_s is the average value of x in state s over the sample period, and m_y is the average value of x in year y. The definitions and sources of all variables are in the appendix.

Across all India	**Rural Head-count**	**Rural Poverty gap**	**Urban Head-count**	**Urban Poverty gap**	**Credit to SDP**	**Branch per capita**	**SDP per capita**	**Rural popula-tion**	**Gov. exp. to SDP**	**Literacy rate**
Mean	31.935	7.521	25.89	6.515	26.962	74.1	8,781	74.487	19.274	55.981
Standard deviation	14.898	4.685	12.076	3.882	14.387	18.732	6,897	8.555	41.173	14.083
SD within state and year	4.946	1.822	4.21	1.535	4.869	2.568	2,404	1.14	2.633	2.061
Across States										
Andhra Pradesh	25.489	5.574	27.744	6.814	28.984	70.422	8600	73.772	18.932	47.389
	(10.169)	(2.574)	(8.015)	(2.512)	(4.857)	(2.995)	(6129)	(1.157)	(1.766)	(9.940)
Assam	37.410	7.845	11.631	2.117	10.990	49.558	6270	88.525	22.026	54.360
	(13.892)	(3.408)	(7.455)	(1.776)	(2.003)	(4.508)	(3465)	(1.013)	(5.219)	(7.073)
Bihar	49.461	12.567	33.704	8.670	15.934	51.310	3509	86.911	22.971	40.570
	(14.827)	(5.712)	(7.844)	(3.021)	(4.826)	(4.787)	(1657)	(0.276)	(7.269)	(7.833)
Gujarat	29.235	6.350	28.181	6.448	25.078	79.889	11316	65.096	17.702	60.749
	(11.819)	(3.383)	(11.755)	(3.285)	(3.530)	(3.036)	(7563)	(2.095)	(2.959)	(7.673)
Haryana	19.889	4.371	14.899	2.940	18.543	79.081	13096	74.253	17.584	56.388
	(10.840)	(2.739)	(7.185)	(1.551)	(2.865)	(4.183)	(8837)	(2.581)	(3.248)	(9.729)
Karnataka	35.740	8.917	26.261	7.138	38.069	95.866	9138	68.308	19.117	56.518
	(13.995)	(4.777)	(9.684)	(3.237)	(8.298)	(3.282)	(6422)	(1.831)	(0.940)	(8.630)
Kerala	24.965	5.951	24.071	6.306	35.525	106.713	9001	75.393	21.789	86.246
	(13.424)	(3.804)	(12.890)	(4.273)	(6.250)	(2.448)	(6584)	(2.295)	(2.237)	(5.986)
Madhya Pradesh	38.587	9.528	37.584	10.065	19.228	61.144	5962	76.717	20.097	48.314
	(11.124)	(3.980)	(7.603)	(3.623)	(6.179)	(4.080)	(3400)	(1.536)	(3.723)	(11.727)
Maharashtra	39.160	10.223	30.906	8.873	58.493	76.792	13533	60.626	15.228	65.783
	(13.936)	(4.960)	(6.878)	(2.496)	(20.349)	(2.133)	(8998)	(2.511)	(1.433)	(9.076)
Orissa	35.891	8.329	37.938	10.624	17.923	63.054	5665	86.335	22.119	51.401
	(11.209)	(3.789)	(7.650)	(2.752)	(5.485)	(4.138)	(3420)	(1.090)	(3.728)	(9.239)
Punjab	14.115	2.410	9.784	1.908	25.771	111.961	14968	69.109	16.442	59.165
	(6.979)	(1.465)	(5.781)	(1.556)	(4.405)	(5.121)	(9647)	(2.429)	(3.015)	(8.831)
Rajastan	38.840	9.975	23.467	5.343	19.315	66.249	6618	77.303	21.514	43.784
	(11.951)	(4.805)	(7.970)	(2.365)	(5.574)	(5.133)	(4055)	(0.724)	(3.003)	(11.909)
Tamil Nadu	33.703	8.330	29.721	7.761	45.914	81.318	9873	62.350	19.439	63.561
	(14.915)	(5.319)	(11.647)	(3.835)	(10.020)	(2.507)	(7010)	(4.635)	(1.955)	(8.228)
Uttar Pradesh	32.820	7.484	32.488	8.410	17.084	58.230	5333	80.090	19.357	47.050
	(10.118)	(3.324)	(9.052)	(3.331)	(2.712)	(4.639)	(2949)	(1.000)	(3.812)	(12.318)
West Bengal	23.719	4.956	19.964	4.310	27.575	59.914	8822	72.514	14.792	58.440
	(12.892)	(3.980)	(7.732)	(2.162)	(2.677)	(3.908)	(6101)	(0.499)	(1.444)	(8.539)

Table 2.2: Correlation table – This table presents pair-wise correlation coefficients between the main variables. The definitions and sources of all variables are in the appendix. *, **, and *** shows significance at 10 %, 5% and 1% level.

	Rural Head-count	**Rural Poverty gap**	**Urban Head-count**	**Urban Poverty gap**	**Credit to SDP**	**Branches per capita**	**SDP per capita**	**Rural popula-tion**	**Gov. exp. to SDP**	**Literacy rate**
Rural Poverty gap	0.962***									
Urban Headcount	0.714***	0.717***								
Urban Poverty gap	0.693***	0.718***	0.970***							
Credit to SDP	-0.248***	-0.198***	-0.0708	-0.0234						
Branches per capita	-0.314***	-0.239***	-0.187***	-0.142***	0.405***					
SDP per capita	-0.699***	-0.655***	-0.622***	-0.564***	0.487***	0.263***				
Rural population	0.316***	0.247***	0.135**	0.112**	-0.757***	-0.553***	-0.526***			
Government exp./SDP	-0.0823	-0.133**	-0.112**	-0.126**	-0.0921*	-0.206***	0.0395	0.361***		
Literacy rate	-0.547***	-0.543***	-0.447***	-0.400***	0.537***	0.471***	0.660***	-0.467***	0.174***	
English newspapers PC in 1991	0.318	0.288	0.305	0.335	0.842***	0.195	0.469*	-0.663***	-0.369	0.406

2.2.2. Identification strategy

We are interested in using our state-level panel data on financial indicators and poverty outcomes to examine whether financial development reduced poverty in Indian states over the period 1983 to 2005. To control for reverse causation and omitted variable bias, we utilize an instrumental variable approach using two instruments for financial development. In this section, we first discuss India's financial liberalization in the 1990s, then explain our instruments and specify the estimation methodology.

India's financial liberalization experience

Prior to financial liberalization in the 1990s, India's financial system was characterized by nationalized banks and directed credit that led to a complex structure of administered interest rates. There was detailed regulation of lending and deposit rates so as to maintain the spread between cost of funds and return on funds (Reddy, 2002). Thus India's public banks lacked proper lending incentives and had a high number of non-performing loans.[12]

Following a severe balance of payments crisis in 1991, there was a substantial liberalization of India's financial sector as part of an economy-wide liberalization process to move towards a market economy and increase the role of the private sector in development. The Government of India set up the Committee on the Financial System which released the Narasimhan Committee Report I that outlined a blueprint for financial reform in 1991. Following its recommendations, the government reduced the volume and burden of directed credit so as to increase the flow of credit to the private sector. The statutory liquidity ratio (SLR) and cash reserve ratio (CRR) that were previously maintained at high levels of 38.5 and 15 percent respectively to lock up bank resources for government use were reduced so as to allow greater flexibility for banks in determining lending terms and increase productivity (Reserve Bank of India, 2004). A second major component of the banking sector reforms was de-regulation of interest rates. Government controls on interest rates were eliminated and the concessional interest rates for priority sectors were phased out to promote financial savings and growth of the organized financial system. There was also greater competition introduced into the banking system by

[12]See Sen, Vaidya, and Sen (1997) and Hanson (2001) for further details on the state of India's banking sector in the pre-liberalization period.

granting licenses to new private banks and new foreign banks and easing of restrictions on foreign banks' operations.

The financial liberalization was also accompanied by strengthening bank regulation and supervision, such as setting minimum capital adequacy requirements for banks (the Basel Accord was adopted in April 1992) and tightening the classification of non-performing loans. Several of the public sector banks were recapitalized and also partially privatized. They were also given more autonomy to enhance competitiveness and efficiency. Given the large proportion of non-performing loans that the public sector banks were saddled with following restrictive policies prior to liberalization, special debt recovery tribunals were set-up in 1993 to streamline the legal procedures and ensure speedy adjudication and recovery of debt (Visaria, 2009). A second committee was established in 1998 that released the Narasimhan Committee Report II, reviewing the banking reform progress and outlining further reforms for strengthening the financial institutions of India.

It is important to note that – unlike the branching policy described below – these reforms were implemented over several years after 1991. In addition, we do not expect any immediate effect of individual policy measures on lending, as banks have to adjust their lending policies and risk management systems to the new regulatory framework.

Role of media

We link cross-state variation in the effects of financial liberalization on financial deepening to cross-state variation in the media environment. The finance literature has explored the role of a free and independent media both on the micro-level for corporate finance and governance, as well as on the macro-level in fostering competition in financial and real sector and ultimately in improving resource allocation. Djankov, La Porta, Lopez-de Silanes, and Shleifer (2003) find that countries with greater state ownership of media (in particular, newspapers) have less free press, fewer political rights for citizens, inferior governance, less developed markets, and do little to meet social needs of the poor. In the Indian context, Besley and Burgess (2002) show that governments are more responsive to natural calamities in states with more developed media presence such as greater newspaper circulation.

Following this literature, we argue that the media in India play an important role in

financial sector development. The information flows resulting from a free media should result in better informed citizenry that stimulates competition in the financial sector leading to greater financial sector deepening. A free media not only makes customers more financially savvy in evaluating financial products and banks but also makes banks better informed leading to better resource allocation overall.

Following Besley and Burgess (2002), we use per capita newspaper circulation as a proxy for media development. The Indian newspaper industry is one of the largest in the world with more than 74,000 newspapers in 22 languages and a readership of 325 million.[13] Newspapers in India are published in a number of languages to cater to the linguistic diversity of the country and most are concentrated in circulation to particular states and cover more localized events. By contrast, English language newspapers have greater national coverage and more business and financial news coverage and are thus more likely to influence financial sector development. There are a number of economic newspapers among which, The Economic Times, published in English, has been the dominant publication.[14] First published in 1961, it is the world's second-most widely read English-language business newspaper (after the Wall Street Journal), and is sold in all major cities in India. Its main content is based on the Indian economy, international finance, share prices, prices of commodities as well as other matters related to finance. Based on the decision makers survey,[15] on average 76% of all decision makers (including Chairmen, CEO, MD) in auto industry, consumer durable industry, telecom industry, financial sector list The Economic Times as their media habit.

Unlike circulation of English newspapers, the state-wise data on circulation of The Economic Times in 1991 is not available. However, we find a very strong (88%) correlation between circulation of The Economic Times and circulation of non-economic English language newspapers in 2005-2009. Therefore, we proxy media coverage of economic and business news in 1991 by per capita circulation of English language newspapers.[16]

[13] M&E newsreel, Ernst & Young, February 2011; "More than 74,000 newspapers are registered in India," The Pak Banker Daily, 29 July 2009, via Dow Jones Factiva 2009, Right Vision Communications Private Limited.

[14] The circulation of the four major economic and business newspapers of India in 2010: The Economic Times (642,443), Hindu Business Line (170,749), Business Standard (141,725), Financial Express (31,000).

[15] The Decision Makers' Survey conducted in 2006 by AC Nielsen ORG-MARG covered senior executives, GM's and above, across 500 private sector, 100 public sector and 100 financial companies. The study looks at the media habits and lifestyles of corporate decision makers in India.

[16] The correlation is the same, if instead of The Economic Times, we use total circulation of the 4 major economic newspapers in India. The registrar of newspapers for India publishes circulation of

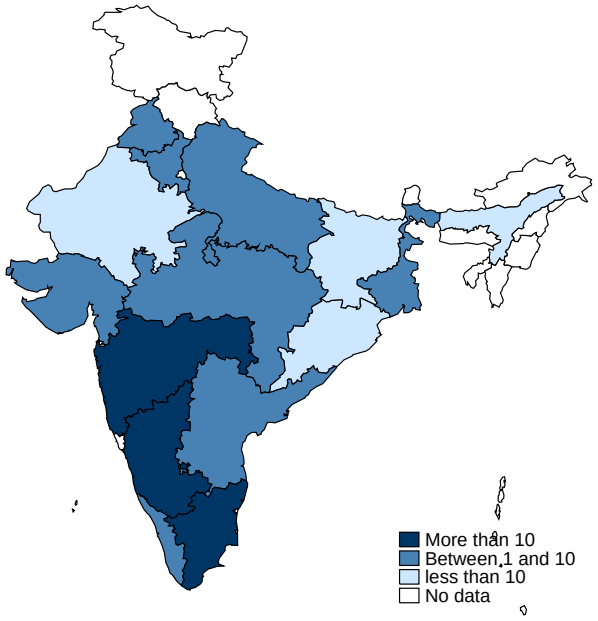


Figure 2.3: English newspaper circulation across India —This figure shows the variation in English newspaper circulation per 1000 persons in 1991 across the different states of India.

Figure 2.3 shows the variation across states in the circulation of English language newspapers per 1,000 people, with the highest levels in Tamil Nadu, Karnataka, and Maharashtra, and lowest levels in Rajasthan, Bihar, Orissa, and Assam. Figure 2.4 illustrates the variation over time – we divide the states into two groups, above (represented by circles in the figure) and below (represented by crosses in the figure) the median (=2) of English language newspaper circulation per 1,000 people and then draw the trend of Credit to SDP in them. It can be clearly seen that the growth in Credit to SDP is more or less the same before liberalization, but afterwards it appears steeper in states with higher level of newspaper circulation, a difference that is statistically significant. Moreover, the growth rate accelerates as the distance from the starting point of liberalization becomes bigger. Hence, we use the cross-state variation of per-capita circulation of English newspapers in 1991 multiplied by a time trend to capture the differential impact of the media across time after liberalization in 1991 as an instrument for financial depth. In robustness tests, we provide a placebo test using local language newspaper penetration, which should not be significantly positive in predicting cross-state variation in financial depth over time.

all newspapers in India according to their title, place and the language of publication. The data from 2005 onwards is available at `http://rni.nic.in/`

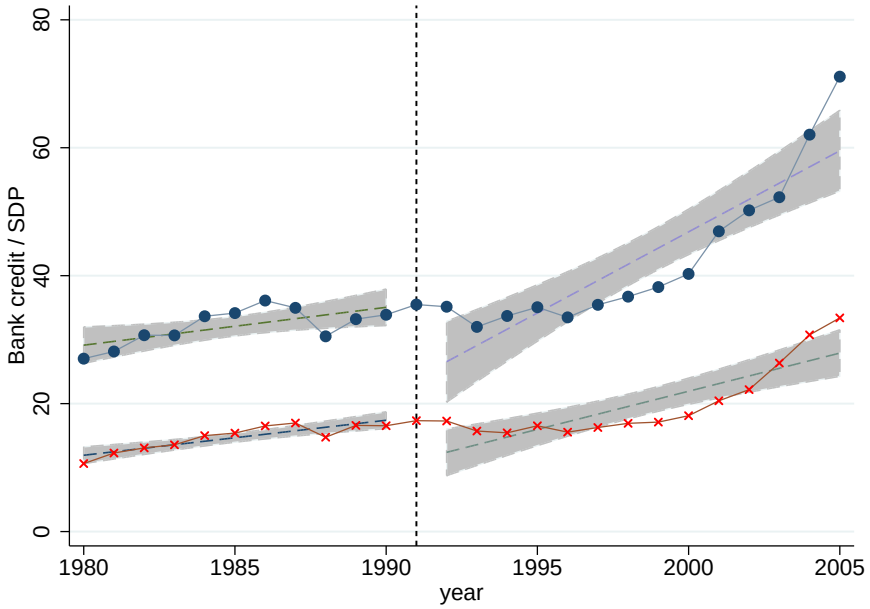


Figure 2.4: Effect of newspaper circulation on financial depth—This figure shows the effect of English newspaper circulation on Bank Credit. The circles (•) show states that had above the median (=2) English newspaper circulation per 1000 persons. The crosses (×) show the rest of states that had below the median circulation of English newspapers per 1000 persons in 1991. The vertical line represents the starting year (1991) of financial liberalization. Prior to 1991, the fitted lines have slopes 0.0055 and 0.0059 for the below and above median groups, but afterwards the slope of the below median is 0.012 and above median is 0.025. The definitions and sources of all variables are in the appendix.

India's social banking experiment

Following independence in 1949, India went through a wave of bank nationalization in 1969 which brought the fourteen largest commercial banks under the direct control of the Indian central bank. Shortly thereafter, the government launched a social banking program with the goal of opening branches in the most populous unbanked rural locations. To further facilitate rural branch expansion, the RBI announced a new licensing policy in 1977 whereby, to obtain a license for a new branch opening in an already branched location (one or more branches), commercial banks had to open branches in four unbanked locations. This rule remained in effect for thirteen years until it was revoked officially in 1990. Burgess and Pande (2005) show that between 1977 and 1990, rural branch expansion was relatively higher in financially less developed states while it was the reverse before 1977 and after 1990. Thus, following Burgess and Pande's approach, we use the resulting trend reversals between 1977 and 1990 and post-1990 in how a state's initial financial development affects rural branch expansion as instruments for branch openings

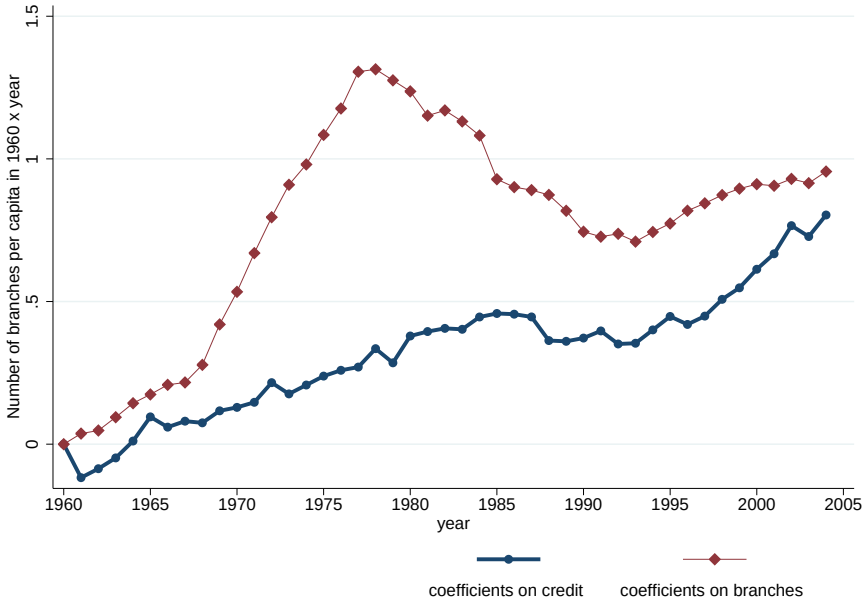


Figure 2.5: Year effects of initial financial development on branch penetration —This figure plots the η_k coefficients obtained from the regression (2.1). The definitions and sources of all variables are in the appendix.

in rural unbanked locations. Figure 2.5 illustrates this trend reversal in bank branches across states and over time, based on the following regression (Burgess and Pande, 2005). For state i in year t,

$$\begin{aligned} \text{Branches}_{it} = \eta_0 + \eta_1(B_{i60} \times D_{60}) + \eta_2(B_{i60} \times D_{61}) + \cdots + \eta_{46}(B_{i60} \times D_{05}) \\ + s_i + y_t + \epsilon_{it}, \quad i = 1, \ldots, 15; \quad t = 1960, \ldots, 2005 \end{aligned} \tag{2.1}$$

where D_t equals 1 in year t and zero otherwise, B_{i60} is the initial level (in 1960) of branch penetration in that state, and s_i and y_t are state and year dummies.

Figure 2.5 graphs the η_k coefficients for the number of branches per million persons as dependent variable. We can see two clear trend reversals in 1977 and 1990. Prior to 1977, the η_k coefficients have an upward trend suggesting that financially developed states provide a more profitable environment for the new branches. With the imposition of the 1:4 rule in 1977, the trend overturns and slopes downward until the rule was repealed in 1990. After 1990, the η_k coefficients are almost unchanging and just slightly grow over time. This reflects that more or less all states were equally likely to attract

new rural branches after the rural branch expansion ended.[17] When we examine the effect of rural branch expansion on overall banking development by estimating equation (2.1) for bank credit, we find no evidence of similar trend reversals, consistent with Joshi and Little (1996) who point out that although the number of bank branches increased over the period 1969-1991, many banks were inefficient and unsound due to poor lending strategies under government control.

In sum, the results from sections 2.2.2 imply that after financial liberalization in 1991, financial deepening increased considerably in states with higher English newspaper penetration. The rural branch expansion policy had a significant impact on the number of bank branches and increased the access of rural areas to banking but did not affect the depth of the banking sector.

Empirical strategy

Following section 2.2.2, we use the following set-up for our instrumental variable specification to address endogeneity issues in the relationship between financial sector development and poverty. The first stage regression of our instrumental variable specification is as follows:

$$\begin{aligned} FD_{it} = \lambda_0 &+ \theta(M_{i91} \times [t - 1991] \times D_{91}) + \delta_1(B_{i60} \times [t - 1960]) \\ &+ \delta_2(B_{i60} \times [t - 1977] \times D_{77}) + \delta_3(B_{i60} \times [t - 1990] \times D_{90}) + \lambda X_{it} \\ &+ s_i + y_t + \varepsilon_{it}, i = 1, \ldots, 15, t = 1983, \ldots, 2005 \end{aligned} \tag{2.2}$$

where FD_{it} is Credit to SDP or Branches per capita, D_{year} is a dummy which equals one post-year, M_{i91} is the state-wise per capita circulation of English newspapers in 1991, B_{i60} is the state-wise per capita rural branches in 1960, X_{it} is the set of control variables and includes SDP per capita, rural population share, literacy rate and state government expenditure to GDP. s_i and y_t are state and year fixed effects to control for any unobserved heterogeneity across states and years.

The main coefficients of interest are θ and δ_i, where θ measures the relationship

[17] Panagariya and Mukim (2014) and Kochar (2011) argue that India had a policy of linking urban branch expansion to rural branch expansion well before bank nationalization and 1977 is not a sharp break from the prior period in terms of the branch expansion rule. This does not concern our estimations since 1977 is not a trend break in our sample period of 1983-2005.

between media freedom interacted with a post-liberalization time trend and financial development and the δ_i's check for trend breaks due to the 1:4 licensing rule. The coefficient δ_1 measures the trend relationship between initial financial development in 1960 and FD (specifically branch expansion). The trend reversals in this relationship are given by δ_2 and δ_3. In the estimations that cover the time period 1983-2005, we skip the first trend dummy, δ_1, since it would be collinear with δ_2.

To analyze the relation between finance and poverty across Indian states, we estimate the following second stage regression:

$$\text{Poverty}_{it} = \beta_0 + \beta_1 \text{Credit}_{it-1} + \beta_2 \text{Branches}_{it-1} + \beta_3 X_{it-1} + s_i + y_t + \varepsilon_{it} \tag{2.3}$$

where Poverty_{it} is a measure of poverty in state i and time t and is one of the four poverty indicators – Rural Headcount, Rural Poverty Gap, Urban Headcount, Urban Poverty Gap. Bank Credit and Branches are the predicted values from the first stage regressions in (2.2) and the remaining variables are also the same as in (2.2). The coefficients of interest are β_1 and β_2 which measure the effect of financial deepening and broadening access on poverty, respectively. We use one-period lags of all the explanatory variables.

All the regressions have a difference-in-difference specification where by including state and time dummies we control for omitted variables that might drive the dependent variable over time or across states. We thus focus on the within-state, within-year variation in the relationship between finance and poverty alleviation, controlling for other time-variant state characteristics. We apply double clustering,[18] both within states and within years to resolve the problem of underestimated standard errors arising from serial correlation of the error terms in difference-in-difference estimations as suggested by Bertrand, Duflo, and Mullainathan (2004).[19] In further regressions and to disentangle the channels through which finance affects rural and urban poverty levels, we use different dependent variables, as we will discuss in detail below.

[18] Our results are materially similar when we cluster only at the state level.

[19] The significance levels we obtain with this method should be treated as conservative because Cameron, Gelbach, and Miller (2008) suggest that when the number of clusters is less than 50, standard errors may be biased and need small sample correction such as the wild boostrap-t procedure. However, as reported by Angrist and Pischke (2008, page 323), Hasan et al. (2007) shows that the clustered standard errors reported by the software program Stata is reasonably good at correcting for serial correlation in panels even when the number of clusters is small.

2.3. Empirical results

In this section, we examine if there is a causal relationship between financial development and poverty using two instruments for financial development, the trend reversals induced by the rural branch expansion program and the differential English newspaper circulation across states after financial liberalization. We first present and discuss the first-stage regressions, before moving to the second stage estimations.

2.3.1. Finance, media and branching policy: first stage results

Table 2.3 presents the first stage regressions following model (2.2). Specifically, we regress Credit to SDP and branch penetration on (i) the interaction between per capita English language newspaper circulation in 1991, a post-liberalization dummy that takes the value 1 for the years 1992 and beyond, and a time trend, (ii) the interaction between bank branches in 1960, a post-1977 dummy and a time trend, and (iii) the interaction between bank branches in 1960, a post-1990 dummy and a time trend. We also control for other time-variant state characteristics included in the second stage, namely SDP per capita, literacy, government expenditures to SDP and the rural population share.

The results in column (1) of Table 2.3 show that states with higher English-language newspaper circulation post-1991 have higher levels of Credit to SDP. The relationship is not only statistically significant, but also economically meaningful: one additional English newspaper per 1,000 persons in 1991 translates into an increase in Credit to SDP by 0.1 percent per year after liberalization. This compares to an average of English newspaper circulation of 5.51 per 1,000 people and a standard deviation of 8.72. On the other hand, the trend reversals in branch penetration associated with the social banking program cannot explain variation in financial depth.

Table 2.3: Finance, Newspaper Circulation and Branching Policy –The regression equation estimated is: Credit to SDP (or Branches per capita)$_{it}$ = $\lambda_0 + \theta_1$Per Capita Circulation of English newspapers in 1991$_i$ $+\theta_2$Per Capita Circulation of Non-English newspapers in 1991$_i$ $+\delta_1$(year$-$1960)$\times B_{60} + \delta_2$(year$-$1977)$\times B_{60} \times D_{77} + \delta_3$(year$-$1990)$\times B_{60} \times D_{90} + \lambda_1$Log (SDP per capita)$_{it}$ $+\lambda_2$Literacy Rate$_{it}$ $+\lambda_3$Government exp./SDP$_{it}$ $+\lambda_4$Rural population$_{it}$ $+ s_i + y_t + e_{it}$, where $D_{77(90)}$ is dummy for post 1977(1990), $B60$ is No. bank branches/Mill. capita in 1960, s_i and y_t are state and year dummies. All regressions are estimated by ordinary least squares and with time-variant independent variables all lagged by one period. Standard errors clustered at state and year level are in parentheses. AP-chi2 is Angrist and Pischke (2009) test of weak instruments. Weak ID test is Stock-Yogo weak identification test with critical values: 10% maximal LIML size=4.72 15%=3.39 20%=2.99 25%=2.79. *, **, and *** shows significance at 10%, 5% and 1% level. The definitions and sources of all variables are in the appendix.

Time period	1983-2005				1965-2005			
	Bank Credit to SDP (%)	Branches per capita	Bank Credit to SDP (%)	Branches per capita	Bank Credit to SDP (%)	Branches per capita	Bank Credit to SDP (%)	Branches per capita
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percapita circulation of english newspapers in 1991 × (year-1991) × D92	104.214***	9.516**			97.973***	-6.855		
	(8.846)	(4.826)			(9.486)	(7.777)		
Percapita circulation of Non-english newspapers in 1991 × (year-1991) × D92			0.101	0.027**			0.092	0.056**
			(0.065)	(0.013)			(0.064)	(0.027)
(year-1977) × B60 × D77	-0.020	-0.139***	-0.041	-0.142***	-0.065***	-0.389***	-0.073**	-0.389***
	(0.019)	(0.023)	(0.031)	(0.023)	(0.017)	(0.061)	(0.029)	(0.062)
(year-1990) × B60 × D90	0.064*	0.144***	0.061	0.137***	-0.012	0.132***	-0.015	0.113***
	(0.033)	(0.027)	(0.064)	(0.024)	(0.027)	(0.017)	(0.051)	(0.017)
(year-1960) × B60					0.066***	0.272***	0.065***	0.271***
					(0.013)	(0.064)	(0.017)	(0.064)
Constant	133.265***	1.400	260.657***	20.555	197.535***	11.396	298.293***	12.440
	(46.601)	(43.077)	(87.746)	(45.197)	(64.468)	(63.233)	(90.648)	(60.980)
Observations	345	345	345	345	597	597	597	597
R-squared	0.962	0.992	0.915	0.992	0.956	0.982	0.926	0.983
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AP-chi2	169.388	44.400	4.204	43.493	160.092	245.911	2.073	155.457
P-value	0.000	0.000	0.122	0.000	0.000	0.000	0.355	0.000
Weak ID test	74.336	74.336	15.955	15.955	134.912	134.912	13.067	13.067

The results in column (2) of Table 2.3 show that both English-language newspaper circulation and the social banking policy can explain cross-state, cross-year variation in branch penetration. Again, the results are not only statistically, but also economically significant. One additional English newspaper per 1,000 people in 1991 is associated with 9.5 more branch establishments per million population annually after liberalization. Moreover, one additional branch per million capita in 1960 translates to 0.139 fewer annual branches per million people during the rural branching expansion, but after the program, it is associated with 0.05 (0.144-0.139) branches more per million persons annually. The Cragg-Donald F-statistic test, with critical values complied by Stock and Yogo (2005), a weak identification test for the excluded exogenous variables, is highly significant. This test is essential when the number of endogenous variables is more than one and the standard F-test may not truly reflect the relevance of instruments (for details see Baum, Schaffer, and Stillman, 2007). We also report the Angrist-Pischke first-stage F-statistics, which are highly significant, indicating that our instruments are relevant (Angrist and Pischke, 2008).[20] In summary, we find that the differential English newspaper across states explains financial depth better than trend instruments while the reverse is true for branch penetration.

In columns (3) and (4), we conduct a placebo test by checking whether circulation of non-English newspapers, which are less likely to report economical and financial news, explains financial development. We find that the coefficients are mostly insignificant for credit to SDP suggesting that the circulation of non-English newspapers is not associated with financial sector development. This also suggests that the relationship between newspaper penetration and financial depth is not spurious and not driven by positive impact that more vibrant media have on government accountability and thus possibly indirectly on competition and depth in the financial system. We do however find a strong positive relationship between circulation of non-English newspapers and branch penetration. Finally, in columns (5) to (8), we show the robustness of our first-stage results to using the 1965 to 2005 sample period.[21]

[20] Unlike other F-statistics, which test the first stage regression as a whole, the Angrist-Pischke first-stage F-test gauges the relevance of each endogenous variable.

[21] Over this period we have three missing points for Assam so the number of observations is 597.

2.3.2. Finance and poverty: second-stage results

We present both OLS and IV regressions of the relationship between financial development and indicators of the incidence and extent of poverty in rural and urban areas. While the OLS regressions do not control for endogeneity and simultaneity bias, we still present them for purposes of comparison.

Table 2.4: Finance and Poverty: OLS estimations —The regression equation estimated is: $\text{Poverty}_{it} = \beta_0 + \beta_1 \text{Credit to SDP}_{it} + \beta_2 \text{Branches per capita}_{it} + \beta_3 \text{Log (SDP per capita)}_{it} + \beta_4 \text{Literacy Rate}_{it} + \beta_5 \text{Government exp./SDP}_{it} + \beta_6 \text{Rural population}_{it} + s_i + y_t + e_{it}$ where s_i and y_t are state and year dummies. Poverty is one of four measures Rural Headcount, Urban Headcount, Rural Poverty gap, and Urban Poverty gap. All explanatory variables are entered with one year lag. All regressions are estimated by ordinary least squares and with time-variant independent variables all lagged by one period. Standard errors clustered at state and year level are in parentheses. The definitions and sources of all variables are in the appendix. *, **, and *** shows significance at 10%, 5% and 1% level.

	Rural poverty	**Rural poverty gap**	**Urban poverty**	**Urban poverty gap**
	(1)	(2)	(3)	(4)
Lag of Credit to SDP	-0.082	-0.081***	0.034	-0.010
	(0.051)	(0.029)	(0.061)	(0.025)
Lag of Branches per capita	-0.220	-0.070	-0.129	-0.053
	(0.182)	(0.100)	(0.110)	(0.042)
Lag of Log(SDP/capita)	-0.664	0.082	-7.075	-2.015
	(5.395)	(3.575)	(4.782)	(1.489)
Lag of literacy rate	0.309***	-0.014	0.344	0.061
	(0.105)	(0.114)	(0.244)	(0.092)
Lag of rural population ratio	0.243	-0.080	0.781**	0.222
	(0.507)	(0.219)	(0.370)	(0.152)
Lag of Gov. exp. / SDP	-0.048	-0.007	-0.269**	-0.093**
	(0.180)	(0.075)	(0.121)	(0.042)
Constant	33.430	24.142	32.318	11.982
	(52.666)	(33.875)	(48.542)	(16.350)
Observations	345	345	345	345
R-squared	0.896	0.857	0.894	0.855

The results in Table 2.4 show a negative relationship between Credit to SDP and the incidence and extent of rural poverty, although the estimate only enters significantly in the case of the rural poverty gap. The relationship between Credit to SDP and urban poverty is not only statistically insignificant but also enters with different signs in the urban Headcount (positive) and urban Poverty Gap (negative) regressions. While branch penetration enters negatively in all four regressions, it does not enter with a significant coefficient. When excluding the rural population share, however, we find that Credit to SDP enters negatively and significantly in both the Rural Headcount (though only at the 10% level) and the Rural Poverty Gap regressions. The difference in significance

between controlling and not controlling for the rural population share can be due to multi-collinearity between credit to SDP and rural population share with correlation coefficient -0.757. Credit to SDP continues to enter insignificantly in the regressions of Urban Headcount and Urban Poverty Gap, while Branches per Capita does not enter significantly in any of the regressions. The insignificant effect of credit to SDP on urban poverty can be a first indication of a possible migration channel through which Credit to SDP impacts poverty. Columns (3) and (4) suggest that urban poverty is negatively associated with Government expenditure as a share of SDP. In all regressions except rural poverty gap, SDP per capita has a negative but insignificant association with poverty measures which can be due to the price adjustment of poverty line overtime. Finally, literacy rate appears with positive and significant coefficients in rural headcount regressions even though according to Table 2.2 they are negatively correlated. This may be due to the multi-collinearity between literacy rate and SDP per capita.

The IV regressions in Table 2.5 show a negative and significant relationship between Credit to SDP and rural poverty whereas there is no significant relationship between branch penetration and rural poverty. As in the case of the OLS regressions, neither Credit to SDP nor branch penetration enter significantly in the regressions of the urban poverty measures. The relationship between Credit to SDP and rural poverty is not only statistically but also economically significant. Specifically, the point estimates in columns (1) and (2) imply that one within-state, within-year standard deviation in Credit to SDP explains 17 percent of demeaned variation in the Headcount and 30 percent of demeaned variation in the Poverty Gap.[22] The Hansen over-identification tests reported in columns (1) to (4) are not rejected suggesting that the instruments are valid instruments.

The insignificant results on branch penetration are due to restrictions of the sample period to 1983 to 2005. As the results on branch penetration are in contrast to the finding by Burgess and Pande (2005), we try to reconcile our results with their findings in columns (5) and (6) by expanding the sample period back to 1965. We find that branch penetration enters negatively and significantly in the regressions of Rural Headcount and Rural Poverty Gap. The insignificant relationship between branch penetration and poverty, found above, is thus due to the shorter time span that does not include the

[22]The effect of credit/SDP on rural headcount and poverty gap are calculated as -0.18*4.87/4.95 = -0.17, and -0.111*4.87/1.82 = -0.30, respectively.

Table 2.5: Finance and Poverty: Instrumental Variable results –This table presents the second stage of instrumental variable regressions estimated by LIML method. The regression equation estimated is: Poverty$_{it} = \beta_0 + \beta_1$Credit to SDP$_{it} + \beta_2$Branches per capita$_{it} + \beta_3$Log (SDP per capita)$_{it} + \beta_4$Literacy Rate$_{it} + \beta_5$Government exp./SDP$_{it} + \beta_6$Rural population$_{it} + s_i + y_t + e_{it}$ where s_i and y_t are state and year dummies. Poverty is one of four measures Rural Headcount, Urban Headcount, Rural Poverty gap, and Urban Poverty gap. The instrumented values are obtained from first stage regressions in Table 2.3. All independent variables are lagged by one period. Standard errors clustered at state and year level are in parentheses. The definitions and sources of all variables are in the appendix. The OID test is the Hansen J statistic over-identification test of all instruments. *, **, and *** shows significance at 10%, 5% and 1% level.

Time period	**1983-2005**				**1965-2005**	
	Rural poverty	**Rural poverty gap**	**Urban poverty**	**Urban poverty gap**	**Rural poverty**	**Rural poverty gap**
	(1)	(2)	(3)	(4)	(5)	(6)
Lag of Credit to SDP	-0.176**	-0.111***	0.013	-0.025	-0.402**	-0.178**
	(0.084)	(0.039)	(0.100)	(0.036)	(0.174)	(0.081)
Lag of Branches per capita	-0.273	-0.107	-0.045	0.008	-0.310***	-0.118**
	(0.198)	(0.096)	(0.144)	(0.089)	(0.083)	(0.055)
Lag of Log(SDP per capita)	-1.594	-0.222	-7.249	-2.142	-5.667	-0.401
	(4.299)	(3.124)	(4.419)	(1.432)	(8.132)	(3.797)
Lag of Literacy rate	0.265***	-0.039	0.386*	0.091	0.394*	0.127
	(0.091)	(0.098)	(0.234)	(0.100)	(0.225)	(0.143)
Lag of Rural population	0.115	-0.109	0.695*	0.159	-0.403	-0.138
	(0.461)	(0.181)	(0.390)	(0.167)	(1.066)	(0.391)
Lag of Gov. exp./SDP	-0.071	-0.019	-0.252**	-0.080**	0.184	0.056
	(0.159)	(0.063)	(0.099)	(0.034)	(0.409)	(0.160)
Observations	345	345	345	345	597	597
R-squared	0.010	0.046	0.003	-0.000	-0.014	0.036
OverID test	0.932	0.326	0.610	0.075	1.227	0.981
OID P-value	0.334	0.568	0.435	0.785	0.268	0.322

starting point of rural branching program. Even over the longer time period, however, Bank Credit to SDP continues to enter negatively and significantly in the regressions of Rural Headcount and Rural Poverty Gap.

To compare the economic effect of depth with breadth, we take a look at de-trended standard errors and use the longer sample period over which both financial depth and inclusion are shown to have a significant relationship with rural poverty gauges. Between 1965 and 2005, the within state and year standard deviations of rural poverty, credit to SDP and branches per capita are 5.910, 7.715, and 5.339 respectively. Using the coefficient estimates from columns (5) and (6) we compute that one standard deviation increase in credit to SDP reduced Rural Headcount by 3.10, while a one standard deviation in branch penetration reduces Rural Headcount by 1.65. Thus, over the pe-

riod 1965 to 2005, variation in branch penetration explains 28 percent of rural poverty reduction in India which is lower than the contribution of credit to SDP (52 percent).[23] Over the longer time period, financial depth was more important than financial inclusion in reducing poverty, while in the more recent sample period, after 1983, only financial deepening can explain reductions in rural poverty.

In further sensitivity tests, available in the Appendix 2.C, we control for additional time-variant state factors, most of which, however, are not available for the whole sample period. First, we include the state government development expenditures as ratio to SDP, which might explain variation in poverty rates across states and over time. While this variable enters negatively and significantly, it does not change the economic or statistical significance of Credit to SDP. Second, we include an indicator to gauge the degree to which a state is open to trade with other countries, with annual data available for the period 1980 to 2002 (Marjit, Kar, and Maiti, 2007). While trade openness does not enter significantly, Credit to SDP continues to enter negatively and significantly.

Third, we control for an indicator of labor market regulation, based on Besley and Burgess (2004) and Gupta, Hasan, and Kumar (2009) that indicates whether labor market regulation in a given state and year can be considered flexible, neutral or inflexible. As the labor market indicator does not vary after 1991, we also interact it with a time trend to test whether states with initially more flexible labor market regulation experienced faster poverty reduction post-1991 liberalization. While the labor market index enters negatively, it does not enter significantly and our financial depth indicator continues to enter with a negative and significant coefficient. Fourth, we control for two indicators of physical infrastructure; specifically, the log of unit costs of electrical power supply, which we have available for the period up 2001 and after 2007, with data from the Planning Commission. We extrapolate for the period in between with linear extrapolation. We also control for road density, measured by the total length of roads in km per 1000 km^2, with data for 1990 to 1996 from Ghosh and De (2005) and for 1998 to 2008 from the Central Statistics Organization. While both indicators show a negative but insignificant relationship with rural poverty, our main findings are confirmed. Fifth, since poverty has convergence characteristic, we include the lag of the dependent variables in the second stage regressions. Although the lag of the dependent variables appears with positive

[23]The effect of credit is -0.402*7.715/5.910=-0.52, and for branches it is -0.310*5.339/5.910=-0.28.

and significant sign in the regressions, the coefficient of credit to SDP remains negative and significant. As an additional check to gauge whether the increasing importance of communication technology might not drive both financial deepening and reductions in poverty, we add phone penetration time trend defined as "Telephone Subscribers share in 1991 $\times$ (year-1991) $\times$ D_{92}" in the control variables of all IV regressions and do not observe any change in the qualitative results.

Finally, we control for two political variables. Specifically, we include (i) the share of votes won by the ruling coalition and (ii) the share of seats won by the ruling coalition. Lower values of the share of votes or share of seats won are likely to represent competitive districts where the ruling and opposition parties have won a similar share of votes. While both indicators show an insignificant relation with rural poverty, we find that controlling for political competition financial depth continues to have a negative and significant impact on rural headcount ratios.

Overall, this shows that even when controlling for development expenditures, trade openness, infrastructure and political structure, some of which are also significantly correlated with financial depth, Credit to SDP instrumented by newspaper circulation interacted with a post-1991 time trend, continues to be negatively and significantly associated with rural poverty. As the tests of overidentifying restrictions are notoriously weak, we also reran the Table 2.4 regressions including newspaper penetration as additional explanatory variable. It never enters significantly in the rural poverty regressions, suggesting that there is no direct impact of a thriving English-language media on rural poverty reduction other than through financial deepening or any of the other explanatory variables.

Overall, IV and OLS results suggest that higher levels of financial depth are associated with both a lower incidence and depth of rural poverty but not with incidence or depth of urban poverty. Financial inclusion, as gauged by branch penetration, is not significantly associated with lower poverty level unless we consider a longer sample period including the period before the social banking policy. These initial regressions thus show that financial deepening is more robustly related to poverty reduction than financial inclusion in recent periods. We next turn to the channels and mechanisms through which financial deepening is related to poverty reduction.

2.4. Finance and poverty: channels

So far the results show that financial deepening since the liberalization in 1991 has helped reduce rural poverty in India. However, understanding the underlying channels is as important for policy makers who try to maximize the benefits of financial development. In this section, we explore different channels through which financial development helped reduce rural poverty. Specifically, we explore whether financial depth helped reduce rural poverty by enabling more entrepreneurship, by fostering human capital accumulation, or by enhancing migration and reallocation across sectors.

2.4.1. Financial depth and entrepreneurship

Theory and empirics have shown that financial imperfections represent particularly severe impediments to poor individuals opening their own businesses for two key reasons: (i) the poor have comparatively little collateral and (ii) the fixed costs of borrowing are relatively high for the poor (Banerjee and Newman, 1993; De Mel, McKenzie, and Woodruff, 2008). The microfinance movement has been built on the premise that enabling the poor to become entrepreneurs will allow them to pull themselves out of poverty.

To assess whether higher entrepreneurship among the poor can account for the significant relationship between financial depth and rural poverty identified in section 2.3, we test whether financial depth, instrumented by English newspaper penetration interacted with a post-liberalization time trend, can explain reduction in poverty among different occupational groups. Specifically, we distinguish between (i) self-employed in agriculture, (ii) self-employed in non-agriculture, (iii) agricultural labor, (iv) other labor and (v) a residual group, which comprises economically non-active population not fitting in the above categories. While we focus our discussion on IV regressions, our findings are robust to using OLS regressions. We also focus on Credit to SDP as our main indicator of financial sector development. Robustness tests including branch penetration yield similar findings for credit depth, while the financial sector outreach measure does not enter significantly in any of the regressions. We focus on rural areas since this is where we found a negative and significant relationship between financial depth and poverty in the previous section.

The results in Table 2.6 show that Credit to SDP is negatively and significantly

associated with the Headcount and the Poverty Gap among the rural self-employed in non-agriculture and in agriculture. Financial depth does not enter significantly in any of the other regressions.[24] Notably, financial deepening cannot explain variation in Headcount or Poverty Gap among laborers or employed workers; while the coefficients enter negatively, the standard errors are far from standard levels of significance. Together, these results suggest financial deepening after the liberalization in the 1990s was associated with a reduction in both the share of the poor and the poverty gap in the population segment of self-employed in the rural areas. Overall, this provides evidence for the entrepreneurship channel, as the reduction in poverty rates fell on self-employed.

2.4.2. Financial depth and human capital accumulation

Financial imperfections in conjunction with the high cost of schooling represent particularly pronounced barriers to the poor purchasing education, perpetuating income inequality (Galor and Zeira, 1993). An extensive empirical literature has shown a relationship between access to finance and child labor, both using country-specific household data[25] and cross-country comparisons (Flug, Spilimbergo, and Wachtenheim, 1998). Theory and previous empirical evidence would thus suggest that financial reforms that ease financial market imperfections will reduce income inequality and poverty levels by allowing talented, but poor, individuals to borrow and purchase education or parents to send their children to school rather than forcing them to earn money to contribute to family income. We test these hypotheses with our data focusing on different educational segments of the rural population across Indian states and gauge whether financial deepening is associated with an increase in the educational attainment in rural India. Specifically, we distinguish between (i) illiterates, (ii) population with primary education, (iii) population with middle school education and (iv) population with high school degree or higher. Unlike in the previous regressions, we also test for longer-run trends by running regressions with five and ten-year lags.

[24] In unreported regressions, we also look into the share of each occupational group in total population and witness, credit to SDP is positively associated only with the share of self-employed in agriculture.

[25] Specifically, survey data for Peru suggest that lack of access to credit reduces the likelihood that poor households send their children to school (Jacoby and Skoufias, 1997), while studies for Guatemala, India and Tanzania point to households without access to finance as being more likely to reduce their children's school attendance and increase their labor if they suffer transitory income shocks compared to household with more assets (Beegle, Dehejia, and Gatti, 2006; Guarcello, Mealli, and Rosati, 2010; Jacoby and Skoufias, 1997).

Table 2.6: Entrepreneurship channel –This table presents the second stage of instrumental variable regressions estimated by LIML method. The regression equation estimated is the same as Table 2.5. Rural Poverty is one of two measures. Rural headcount and Rural poverty gap in each of 5 categories of rural household employment type: (i) self-employed in agriculture, (ii) self-employed in non-agriculture, (iii) agricultural labor, (iv) other labor and (v) others, a residual group that comprises economically non-active population not fitting in the above categories. All explanatory variables are entered with one year lag. Weak ID test is Stock-Yogo weak identification test. *, **, and *** shows significance at 10%, 5% and 1% level.

	rural & self-employed in non-agriculture		**rural & self-employed in agriculture**		**rural & agricultural labour**		**rural & other labour**		**rural & others**	
	HC	PG	HC	PG	HC	PG	HC	PG	HC	PG
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lag of Credit to SDP	-0.226*	-0.120***	-0.436***	-0.211***	-0.143	-0.096	-0.046	-0.033	0.156	0.023
	(0.120)	(0.035)	(0.167)	(0.065)	(0.116)	(0.075)	(0.095)	(0.035)	(0.108)	(0.058)
Observations	299	298	294	297	298	299	297	298	296	297
R-squared	-0.003	0.023	-0.010	0.013	0.005	0.001	-0.001	0.000	0.005	-0.000
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weak ID test	427.372	427.560	443.976	446.658	446.546	450.719	447.204	449.689	442.968	420.545

Table 2.7: Education channel –This table presents the second stage of instrumental variable regressions estimated by LIML method. The regression equation estimated is the same as Table 2.5. Education is the education segment of the rural population and is one of four variables proportion of illiterates, proportion of population with primary education, proportion of population with middle school education, and proportion of population with high school degree or higher.

	proportion of illiterates			**proportion of up to primary**			**proportion of middle school**			**proportion of High school diploma and above**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Lag of Credit to SDP	0.085**			-0.062*			0.019			-0.056		
	(0.039)			(0.032)			(0.025)			(0.048)		
5 lag of Credit to SDP		0.385***			-0.026			-0.025			-0.260***	
		(0.069)			(0.140)			(0.101)			(0.059)	
10 lag of Credit to SDP			-0.188			0.075			-0.199***			0.289
			(0.223)			(0.204)			(0.075)			(0.276)
Observations	285	285	285	285	285	285	285	285	285	285	285	285
R-squared	0.005	0.004	-0.027	0.007	0.002	-0.006	0.003	0.001	-0.025	0.002	0.016	-0.101
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weak ID test	456.701	93.259	18.531	456.701	93.259	18.531	456.701	93.259	18.531	456.701	93.259	18.531

Financial sector deepening that results in more human capital accumulation cannot be expected to have an effect immediately but rather after a certain time lag. Testing for the relationship across different lag structures also allows gauging whether any significant relationship is spurious or not.

The results in Table 2.7 do not show any consistent and significant impact of financial deepening on human capital allocation. The regression results do not show any increase in educational attainment, either immediately or after a five or 10 year lag from financial deepening. Rather, we find that the five-year lag of Bank Credit to SDP is positively and significantly associated with the share of illiterates, while it is negatively and significantly associated with the share of population with a high school education or higher. We also find that the 10-year lag of Bank Credit to GDP is negatively associated with the share of middle school graduates. Overall, these results suggest that financial deepening has not led to increases in educational attainment in rural India.[26]

2.4.3. Financial depth, migration and reallocation across sectors

In a world with perfect factor mobility, workers and entrepreneurs would migrate to regions or sectors with better opportunities. Market frictions, however, might prevent such reallocation. Financial deepening can thus also contribute to poverty alleviation by helping households move to areas and sectors with higher earning opportunities. Giné and Townsend (2004) show that financial liberalization in Thailand has resulted in important migration flows from rural subsistence agriculture into urban salaried employment and ultimately in lower poverty levels, while Beck et al. (2010) show that financial liberalization in the U.S. in the 1970s and 80s has helped tighten income distribution by pulling previously unemployed and less educated into the formal labor market. In both countries, financial liberalization broadened opportunities for entrepreneurs, both incumbent and new ones, who in turn hired more workers. If we apply the same argument to the Indian context, we should therefore observe an increase in migration with financial deepening and sectoral reallocation of labor.

As we want to gauge whether finance provided enough incentives for migration within India, we obtain migration data from the NSS surveys for the following years: 1983, 1987-

[26] In unreported regressions, we also limited our sample to children below the age of 18 years to gauge whether financial deepening increases schooling and thus literacy in this specific group and find no effect. Results are available on request.

88, 1993, 1999-00, and 2007-08. These surveys have comprehensive data on migration including data on household migration, characteristics of migrants, years since migration, whether they are short-term migrants or out-migrants,[27] reasons for migration, employment type and the sector from and into which they migrate. We divide households in each state in each year into six groups based on region (rural or urban) and occupational sector (primary, secondary, or tertiary).

As a first step, we present summary statistics on migration in India in panel A of Table 2.8. The migration rate is computed as the ratio of the number of households that migrated to state s in year t to the total number of households sampled in state s. Intra-state migration is computed as the fraction of people who migrated within the state, either between or within the districts and inter-state migration is computed as the fraction of people migrating from another state to this state. For each year, we used the closest survey to estimate the rates. Specifically, we used round 38 in 1983 for estimating the rates in 1980-82, round 43 in 1987 for estimating the rates in 1983-86, round 49 in 1993 for estimating the rates in 1987-92, round 55 in 1999 for estimating the rates in 1993-98, and round 64 in 2007 for estimating the rates in 1999-2005. The estimations start from 1980 because if the migration occurred further past the survey year, it is usually not reported precisely. For instance, immigrants from over 10 years ago tend to report years since migration as multiples of five or ten, creating a peak in migration rate of those years.

The data show that, while overall migration, both inter- and intra-state, is at 1.4 percent of a state's population, on average, per year, it is dominated by intra-state migration, which constitutes about 80 percent of overall migration. Assuming one migration per household, during the period 1983-2005, around 30% of population experienced a migration.[28] When we look at the migration between rural and urban sectors, we find that, as expected, urban to rural migration is the smallest and accounts for an average of 0.2% of total population through the years. Rural to urban migration is the highest though we find that there is comparable amount of migration from urban to urban areas and

[27] Short-term migrants are persons who had stayed away from the village/town for a period ≥ 1 month but ≤ 6 months during the past year for employment. Out-migrants are former members of a household who left the household any time in the past to stay outside the village/town (and are still alive on the date of survey).

[28] In the migration surveys just the earliest migration is reported. The number is computed as $1.373 \times 22 = 30.2$.

since 2000, there has also been a comparable share of rural to rural migration. When we look at occupational sectors, we find that migration into the tertiary sector has been the largest. In unreported charts of migration trends over time, we find that while the primary sector used to be smallest target sector, it overtook the secondary sector in most years after financial liberalization.

Next, we explore the finance and migration channel in more detail with regression analysis. In panel B of Table 2.8, we regress overall migration, intra-state, and inter-state migration on Credit to SDP, instrumented by English newspaper penetration interacted with a post-liberalization time trend and including our other control variables. To be consistent with the benchmark regression we estimate it for the period 1983-2005. Panel B shows that while financial deepening is not significantly associated with overall migration or intra-state migration, there is a significant impact of financial deepening on inter-state migration. The economic size of this effect is reasonable, with one demeaned standard deviation in Credit to SDP explaining around 30 percent of variation in demeaned variation of inter-state migration.[29] In the following, we therefore focus on inter-state migration. Specifically, we use household-level data for inter-state migrants to gauge the impact of financial development on (i) sectoral migration decisions and (ii) reasons for migration. We have data available for around 28,000 inter-state migrant households across the four surveys described above.

[29]The demeaned standard errors of credit and inter-state migration are 0.049 and 0.001 respectively, so the number will be 0.049*0.006/0.001= 0.294.

Table 2.8: Financial deepening and migration –Panel A presents summary statistics of the migration variables. All variables are in percentage terms. Standard errors are computed similar to Table 2.1. Panel B presents second stage of instrumental variables estimated by LIML method. The estimated regression equation is the same as Table 2.5. All explanatory variables are entered with one year lag. Weak ID test is Stock-Yogo weak identification test. *, **, and *** shows significance at 10%, 5% and 1% level.

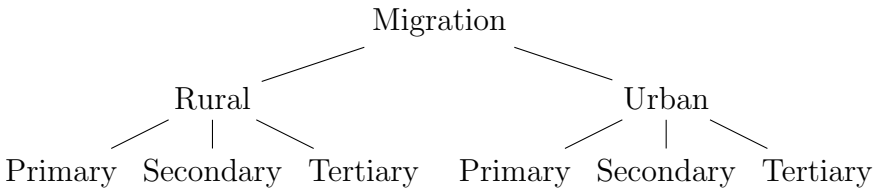
Panel A: Summary Statistics of migration variables (1983-2005)

	Migration rate	**Intra-state**	**Inter-state**	**Migration from rural to rural**	**Migration from rural to urban**	**Migration from urban to rural**	**Migration from urban to urban**	**Migration to primary**	**Migration to secondary**	**Migration to tertiary**
Mean	1.373	1.093	0.292	0.363	0.49	0.147	0.401	0.219	0.257	0.517
SD	0.669	0.592	0.241	0.224	0.271	0.103	0.229	0.135	0.198	0.317
SD within state	0.529	0.445	0.15	0.167	0.224	0.089	0.183	0.11	0.157	0.27
SD within year	0.508	0.477	0.224	0.193	0.217	0.089	0.187	0.118	0.16	0.223
SD within state and year	0.3	0.273	0.12	0.121	0.159	0.071	0.125	0.087	0.106	0.149

Panel B: IV results

	Migration rate (%)	**Intrastate (%)**	**Interstate (%)**
	(1)	(2)	(3)
lag of Credit to SDP	0.002	-0.004	0.006***
	(0.003)	(0.003)	(0.002)
Observations	345	344	330
Control variable	YES	YES	YES
Fixed Effects	YES	YES	YES
Weak ID test	442.121	439.732	420.446

In Table 2.9, we focus on inter-state migration and explore how financial development influences migration into different occupational sectors – primary, secondary, and tertiary. Migrant households can choose between six alternatives – rural primary, rural secondary, rural tertiary, urban primary, urban secondary, and urban tertiary sectors which we group by geographic area (rural or urban). Thus the tree structure of a migrant's decision would be as follows:



We estimate our model as sequential logit model, first testing to which extent the decision to move into urban or rural areas depends on differences in Credit to SDP across origin and destination states and, second, gauging whether the decision to work in the primary, secondary or tertiary sector depends on these differences and controlling for the decision to move into the rural or urban area. Unlike in the previous regressions, we thus focus on differences in financial development and other state-level variables rather than levels at the year of migration. Hence, they compare the level of variables between the destination and origin states when the households decided to migrate. We also control for two household characteristics, household size and per capita expenditure, that might influence migration decisions. We also control whether the migrant household used to live in an urban or rural area.

Table 2.9 shows that financial depth is significantly associated with inter-state migration flows into the rural primary and urban tertiary sectors. The results in columns 1 show that a higher difference in Credit to SDP between destination and origin state increases the likelihood that migrants move into urban areas though this is not statistically significant. We also find that a higher difference in SDP per capita and government expenditure and a lower difference in literacy is associated with a higher likelihood of inter-state migrants moving into urban areas. In addition, richer and smaller migrant households coming from urban areas are more likely to move into urban areas in the destination state. Considering interstate migrants into urban areas, we find that a higher difference in Credit to SDP between destination and origin states results in a higher likelihood that migrants allocate into the tertiary sector and a lower likelihood that mi-

grants allocate into the secondary sector. We also find that interstate migrants into the rural areas are more likely to allocate into the primary sector, the higher the difference in Credit to SDP between origin and destination state. Thus the primary rural sector and the urban tertiary sector were the sectors that benefitted most from the inter-state migration associated with financial deepening.

In Table 2.10, we explore the reasons for inter-state migration for a smaller sample of inter-state migrant households, for which we have such data available. Here, we use multinomial logit regressions and report marginal effects. We find that a higher difference in Credit to SDP between destination and origin states is associated with a higher share of migrants that state "search for employment", "under transfer", and "parents migration" as reason for migration and a lower share of migrants that state "search for better employment" as reason for migration. As in Table 2.9, these findings are robust to controlling for other state-level differences and characteristics of the migrant households. This suggests that higher financial development in the destination state (as compared to the origin state) is associated with migration due to search for employment, though not with the search for better employment. In a further test, available upon request, we re-estimate Table 2.10 for the sample of inter-state migrants below poverty line who emigrated from rural area to urban tertiary. The only reason of those migrants which is positively associated with the difference in credit to SDP between destination and origin is search for employment.

2.4.4. Sectoral credit and reallocation across sectors

In a final step, we relate the relationship between financial deepening and geographic-sectoral migration trends to the sectoral credit portfolio of the Indian banking system. Specifically, which sector drives the cross-state variation in financial deepening observed after the 1991 liberalization? And can we link this through to the poverty-reducing effect in rural areas documented in section 2.3?

Table 2.9: Financial deepening and inter-state migration, sequential logit estimation. –This table presents sequential logit regressions for inter-state immigrants. The regression equation is $Y_{kit} = \beta_1\Delta[\text{Credit to SDP}_{it}]+\beta_2\Delta[\text{Log (SDP per capita)}_{it}]+\beta_3\Delta[\text{Literacy rate}_{it}]+\beta_4\Delta[\text{Government exp./SDP}_{it}]+\beta_5\Delta[\text{Rural population}_{it}]+s_i+y_t+e_{it}$ where s_i and y_t are state and year dummies. Y is a vector of dummy variables taking on value one if household k migrates to an urban area, into the primary, secondary and tertiary sector. Column 1 presents a logit regressions, columns (2) to (4) and columns (5) to (7) present multinominal regressions. Δ indicates the difference between destination and origin (= destination - origin). All explanatory variables are entered with one year lag. The definitions and sources of all variables are in the appendix. The reported coefficients are marginal effects and multiplied by 100 for better illustration. The definitions and sources of all variables are in the appendix. *, **, and *** shows significance at 10%, 5% and 1% level.

	urban after migration	**rural after migration**			**urban after migration**		
		Primary	Secondary	Tertiary	Primary	Secondary	Tertiary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lag of difference in Credit to SDP	0.003	0.178***	-0.089*	-0.088*	-0.04	-0.211***	0.252***
	(0.029)	(0.061)	(0.046)	(0.049)	(0.027)	(0.045)	(0.046)
Lag of difference in SDP per capita	11.152***	-3.455	5.059**	-1.604	2.479**	18.072***	-20.551***
	(1.196)	(2.713)	(2.044)	(2.230)	(1.164)	(2.044)	(2.108)
Lag of difference in Literacy Rate	-0.112***	-0.240***	0.046	0.193***	0.017	-0.126***	0.109**
	(0.028)	(0.060)	(0.046)	(0.048)	(0.028)	(0.047)	(0.048)
Lag of difference in Rural population	-0.893***	0.286**	-0.174*	-0.112	0.099	-0.132	0.033
	(0.062)	(0.136)	(0.104)	(0.111)	(0.061)	(0.107)	(0.109)
Lag of difference in Gov. exp./SDP	0.296***	0.395*	-0.089	-0.306*	0.145	0.587***	-0.732***
	(0.091)	(0.213)	(0.159)	(0.174)	(0.090)	(0.156)	(0.161)
Monthly per capita expenditure	0.07317***	-4.406***	0.659	3.747***	-3.309***	0.055	3.254***
	(0.417)	(0.953)	(0.695)	(0.683)	(0.454)	(0.519)	(0.543)
Household size	-2.593***	2.282***	-1.288***	-0.994***	0.808***	-1.522***	0.714***
	(0.090)	(0.239)	(0.194)	(0.201)	(0.106)	(0.216)	(0.217)
Urban before migration	17.906***	-11.723***	3.846***	7.877***	-2.822***	-6.394***	9.216***
	(0.547)	(1.312)	(1.013)	(1.082)	(0.605)	(1.047)	(1.054)
Observations	28549	5419	5419	5419	11061	11061	11061

Table 2.10: Financial deepening and reasons for inter-state migration. — This table presents multinomial logit estimation for households with inter-state migration. The regression equation is $Y_{kit} = \beta_1\Delta[\text{Credit to SDP}_{it}] + \beta_2\Delta[\text{Log (SDP per capita)}_{it}] + \beta_3\Delta[\text{Literacy rate}_{it}] + \beta_4\Delta[\text{Government exp./SDP}_{it}] + \beta_5\Delta[\text{Rural population}_{it}] + s_i + y_t + e_{it}$ where s_i and y_t are state and year dummies. Y is one of eight reasons for migration. The reported coefficients are marginal effects and multiplied by 100 for better illustration. Δ indicates the difference between destination and origin (= destination - origin). All explanatory variables are entered with one year lag. The definitions and sources of all variables are in the appendix. *, **, and *** shows significance at 10%, 5%, and 1% level.

	search for employment	**search for better employment**	**under transfer**	**studies**	**marriage**	**parents migration**	**political problems**	**others**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lag of Difference in Credit to SDP	0.081***	-0.141***	0.039**	0.009	-0.023	0.054***	-0.012*	-0.007
	(0.022)	(0.027)	(0.015)	(0.008)	(0.029)	(0.010)	(0.007)	(0.018)
lag of Difference in SDP per capita	12.648***	13.404***	-5.071***	-2.134***	-10.358***	-0.319	-1.695***	-6.474***
	(0.921)	(1.103)	(0.665)	(0.359)	(1.218)	(0.466)	(0.330)	(0.795)
lag of Difference in Literacy rate	-0.130***	0.018	0.035**	-0.007	-0.138***	0.015	0.014**	0.194***
	(0.022)	(0.026)	(0.014)	(0.007)	(0.028)	(0.011)	(0.007)	(0.017)
lag of Difference in Rural population	-0.257***	-0.114*	0.014	-0.041**	-0.015	0.03	-0.021	0.404***
	(0.050)	(0.060)	(0.033)	(0.017)	(0.063)	(0.023)	(0.016)	(0.040)
lag of Difference in Gov. exp./SDP	0.833***	0.211**	-0.246***	-0.109***	-1.118***	0.152***	0.021	0.255***
	(0.071)	(0.084)	(0.050)	(0.026)	(0.092)	(0.036)	(0.025)	(0.061)
Monthly per capita expenditure	-0.551**	0.516	3.162***	0.670***	-4.720***	0.404***	-0.149	0.668***
	(0.277)	(0.324)	(0.131)	(0.057)	(0.400)	(0.125)	(0.108)	(0.226)
Household size	-2.816***	-2.605***	-0.143**	-0.544***	5.474***	0.098**	0.081***	0.455***
	(0.101)	(0.111)	(0.061)	(0.049)	(0.093)	(0.038)	(0.024)	(0.061)
Observations	28455	28455	28455	28455	28455	28455	28455	28455

Figure 2.6 graphs the trends of sector-wise credit to SDP over time. For this purpose, we construct credit to SDP measure in the primary, secondary, and tertiary sectors by dividing RBI's sector-wise credit data with the corresponding net state domestic product in that sector. The detail of the source and construction of these measures are described in Appendix 2.B. It can be clearly seen that credit to SDP in the tertiary sector started to grow sharply a few years after liberalization, but this pattern does not exist in the other sectors and there is even a downward trend in credit to the secondary sector.

Table 2.11: Sector-wise financial development. — The estimated regression equation estimated is the same as Table 2.3. All regressions are estimated by ordinary least squares and with time-variant independent variables all lagged by one period. Standard errors clustered at state and year level are in parentheses. AP-chi2 is Angrist and Pischke (2009) test of weak instruments. Weak ID test is Stock-Yogo weak identification test. *, **, and *** shows significance at 10%, 5% and 1% level.

Credit to SDP in	**Primary sector**	**Secondary sector**	**Tertiary sector**
	(1)	(2)	(3)
Percapita circulation of english newspapers in 1991 × (year-1991) × D_{92}	8.425	0.152	156.396***
	(7.657)	(27.025)	(28.891)
(year-1977) × D_{60} × D_{77}	-0.075**	0.008	-0.011
	(0.034)	(0.066)	(0.044)
(year-1990) × D_{60} × D_{90}	0.142***	0.011	0.067
	(0.046)	(0.092)	(0.060)
Constant	104.724**	236.989*	110.347
	(44.591)	(138.616)	(96.320)
Observations	270	270	270
R-squared	0.840	0.693	0.866
Control variables	Yes	Yes	Yes
AP-chi2	1.085	2.397	74.656
P-value	0.781	0.494	0.000
Weak ID test	1.061	1.701	72.687

Table 2.11 confirms in a regression framework that our findings so far are driven by credit to the tertiary sector. Using the same first-stage specification as in Table 2.3, we see that it is just Credit to SDP in the tertiary sector that is strongly associated with newspaper penetration and its interaction with a post-1991 time trend.[30] There is no significant relation between bank credit to primary or secondary sector and newspaper penetration. Not surprisingly, primary credit to SDP (and thus rural credit) is significantly associated with trend breaks of rural branching program, while neither credit to the secondary nor the tertiary sectors are. Overall, this suggests that financial liberalization after 1991 resulted in financial deepening benefitting mostly the tertiary sector.

[30] Compared to the regressions in Tables 2.3 and 2.5, we lose 5 years of data, because our sectoral credit data is not available in 1984-1986, 1988 and 1995.

Figure 2.6: Sectoral credit to SDP in India over time —This figure shows the trends in sector-wise credit to SDP. The primary sector consists of agriculture, fishing, forestry, mining and quarrying; the secondary sector is composed of manufacturing, construction, electricity, gas and water; and the tertiary sector is all services including trade, hotels and restaurants, transport, communication, storage, banking, insurance, real estate, ownership of dwelling, business services, public administration, and other services. The definitions and sources of all variables are in the appendix.

Table 2.12: Poverty and Tertiary sector credit. —This table presents the second stage of instrumental variable regressions estimated by LIML method. The estimated regression equation is the same as Table 2.5. The instrumented values are obtained from first stage regressions in Table 2.11. All independent variables are lagged by one period. Standard errors clustered at state and year level are in parentheses. The OID test is the Hansen J statistic over-identification test of all instruments.

	Rural poverty	**Rural poverty gap**	**Urban poverty**	**Urban poverty gap**
	(1)	(2)	(3)	(4)
Lag of Tertiary Credit to SDP	-0.147**	-0.083**	-0.008	-0.023
	(0.075)	(0.035)	(0.065)	(0.025)
Lag of Branches per capita	-0.159	-0.086	0.043	0.067
	(0.175)	(0.074)	(0.056)	(0.059)
Observations	270	270	270	270
R-squared	-0.051	-0.078	-0.006	-0.056
Control variables	Yes	Yes	Yes	Yes
Over ID test	0.624	0.173	0.489	0.092
OID P-value	0.430	0.678	0.484	0.761

In Table 2.12, we replicate the Table 2.5 regressions, using tertiary Credit to SDP rather than overall Credit to SDP, instrumented by English newspaper penetration in 1991 interacted with a post-1991 time trend. Our Table 2.5 results are confirmed using this sectoral credit measure. Tertiary Credit to SDP enters negatively and significantly in the regressions of Rural Headcount and Rural Poverty Gap, but not in the regressions of the Urban Headcount or Poverty Gap. As in Table 2.5, branch penetration does not enter significantly. The coefficient sizes of Tertiary Credit to SDP are only slightly smaller than those of overall Credit to SDP in Table 2.5.

While we provide statistically and economically strong evidence on the relationship between financial deepening following the 1991 liberalization, geographic-sectoral migration trends and reductions in poverty rates, we have to be careful on our interpretation. Our results do not imply that the increase in credit to the tertiary sector is purely supply-driven. Rather, we interpret our findings as suggesting that financial deepening has supported growth opportunities in the tertiary sector by providing credit to enterprises in this sector, which in turn through labor market effects resulted in the geographic-sectoral migration documented above.

2.5. Conclusion

Academics and policy makers disagree on the effect of financial liberalization and deepening on poverty levels. While some argue that the benefits of liberalization accrue to the upper income segments, others point to pro-poor effects of financial liberalization, by fostering entrepreneurship, human capital accumulation or important labor market effects. Our findings speak directly to this debate.

Using state-level indicators on financial depth, branch penetration and poverty for 1983 to 2005 across 15 Indian states, we show a negative relationship between financial deepening post-1991 and rural poverty. Exploring different channels, we find evidence that the poverty reduction effects of financial deepening fell on the self-employed in rural areas. We also find evidence that financial liberalization resulted in inter-state migration towards states with deeper financial systems, benefiting the rural primary and urban tertiary sectors. Together, these results suggest two related effects of financial deepening

in rural areas: fostering entrepreneurship and migration of the poorest towards financially more developed states. Consistent with the migration trend into the urban tertiary sector we also find that the pro-poor effects of financial deepening are associated with credit to the tertiary sector only.

Our findings suggest that financial deepening can have important structural effects, including through structural reallocation and migration, with consequences for poverty reduction. The pro-poor effects of financial development are multi-faceted and can arise through different channels. There is some evidence that financial development can reduce poverty through fostering entrepreneurship, although this does not necessarily happen through more inclusive but rather more efficient systems. We also show that financial deepening can result in important labor market and migration effects. These effects are consistent with findings by Beck et al. (2010) for the U.S. and Giné and Townsend (2004) for Thailand. On the other hand, we cannot find significant evidence for a human capital channel of financial deepening on poverty reduction.

Our paper has important policy repercussions. The pro-poor effects of financial deepening do not necessarily come just through more inclusive financial systems, but can also come through more efficient and deeper financial systems. Critically, the poorest of the poor not only benefit from financial deepening by directly accessing financial services, but also through indirect structural effects of financial deepening.

Appendix

Data Appendix

2.A. Variable Definitions and Sources

Variable	Source	Definition
Rural Headcount	Authors' calculation using NSSO surveys + Datt et al (1996)	Proportion of the population below the poverty line in rural areas
Rural Poverty gap		Mean distance of the poor from the poverty line –normalized by poverty line– in rural areas
Urban Headcount		Proportion of the population below the poverty line in urban areas
Urban Poverty gap		Mean distance of the poor from the poverty line –normalized by poverty line– in urban areas
Credit to SDP	Burgess & Pande (2005) + Besley & Burgess (2004) + updates from the Reserve Bank of India (RBI) (http://dbie.rbi.org.in)	Credit given by scheduled commercial banks over net state domestic product. Credit data is taken from Burgess & Pande (2005) till 2000 and updated to 2005 using RBI's data. Net state domestic product is provided in Besely & Burgess (2004) till 2002 and is available at EOPP website. For the remaining years it is updated using RBI's data.
Branches per capita	RBI's publications Directory of Bank Offices (http://dbie.rbi.org.in)	Number of back branches per million persons.
SDP per capita	LSE Economic Organisation and Public Policy Programme Indian States Database (EOPP) + updates from RBI (http://dbie.rbi.org.in)	Net state domestic product per person.
Rural population	EOPP + updates from Indian census	Share of rural population to total. Constructed using census data from the five censuses for 1961, 1971, 1981, 1991, 2001. Between any two successive censuses, the state-sectoral populations are assumed to grow at a constant rate, derived from the respective census population totals.
Government exp. / SDP	EOPP + updates from RBI	Total state government expenditures over net state domestic product
Literacy rate	EOPP + updates from Indian census	Proportion of persons who can both read and write in any language among population aged 7 years and above. Constructed using census data from the five censuses for 1961, 1971, 1981, 1991, 2001. Between any two successive censuses, the state-sectoral populations are assumed to grow at a constant rate, derived from the respective census population totals.
Per capita circulation of English newspaper in 1991	EOPP + updates from indiastat.com	Circulation of English newspaper over total population
Per capita circulation of non-English newspaper in 1991	EOPP + updates from indiastat.com	Circulation of non-English newspaper over total population

Variable	Source	Definition
Credit to SDP in Primary sector	RBI's publications Basic Statistical Returns of Banks and Banking Statistics 1972-2002 .	Credit given by scheduled commercial banks to the primary sector over net state domestic product of primary sector (agriculture, fishing, forestry, mining and quarrying). The data is from RBI's online publications Basic Statistical Returns of Banks and Banking Statistics 1972-2002. The data is on an annual basis under the heading Occupation-wise Classification of Credit, but not available for the full sample period and has some missing value in between. The classification of occupation is different from NSDP, so we divide them to three main groups to construct the depth measures: primary (agriculture, mining and quarrying), secondary (industry excluding mining and quarrying, electricity, gas, and water) and tertiary (the rest minus personal loans).
Credit to SDP in Secondary sector	RBI's publications Basic Statistical Returns of Banks and Banking Statistics 1972-2002 (http://dbie.rbi.org.in)	Credit given by scheduled commercial banks to the secondary sector over net state domestic product of secondary sector (manufacturing, construction, electricity, gas and water). The classification of occupation is different from NSDP, so we divide them to three main groups to construct the depth measures: primary (agriculture, mining and quarrying), secondary (industry excluding mining and quarrying, electricity, gas, and water) and tertiary (the rest minus personal loans).
Credit to SDP in Tertiary sector	RBI's publications Basic Statistical Returns of Banks and Banking Statistics 1972-2002 (http://dbie.rbi.org.in)	Credit given by scheduled commercial banks to the tertiary sector over net state domestic product of tertiary sector (trade, hotels and restaurants, transport, communication, storage, banking, insurance, real estate, ownership of dwelling, business services, public administration, and other services). The classification of occupation is different from NSDP, so we divide them to three main groups to construct the depth measures: primary (agriculture, mining and quarrying), secondary (industry excluding mining and quarrying, electricity, gas, and water) and tertiary (the rest minus personal loans).
Rural head count and poverty gap in: **(1) self-employed in non-agriculture,** **(2) self-employed in agriculture,** **(3) agricultural labor,** **(4) other labor,** **(5) other**	Authors' calculation using NSSO surveys	Proportion of the population below the poverty line among (1) self-employed in non-agriculture (2) self-employed in agriculture (3) agricultural labors (4) other labors (5) non-active population which not fitting in the above four categories
Proportion of illiterates / up to primary / middle school / high school and above	Authors' calculation using NSSO surveys	Share of illiterates / literate people who at most have a primary school degree / a middle school degree / high school degree and above in total population

Variable	Source	Definition
(1) Migration rate **(2) Intra-state migration** **(3) Inter-state migration** **(4) Migration from rural to rural** **(5) Migration from rural to urban** **(6) Migration from urban to rural** **(7) Migration from urban to urban** **(8) Migration to primary** **(9) Migration to secondary** **(10) Migration to tertiary**	Authors' calculation using NSSO migration surveys	Ratio of the number of households that migrated to (1) state s in year t to the total number of households sampled in state s. (2) state s in year t from the same states to the total number of households sampled in state s. (3) state s in year t from other states to the total number of households sampled in state s. (4) rural areas of state s in year t from rural areas (either the same state or not) to the total number of households sampled in state s. (5) urban areas of state s in year t from rural areas (either the same state or not) to the total number of households sampled in state s. (6) rural areas of state s in year t from urban areas (either the same state or not) to the total number of households sampled in state s. (7) urban areas of state s in year t from urban areas (either the same state or not) to the total number of households sampled in state s. (8) primary sector of state s in year t to the total number of households sampled in state s. (9) secondary sector of state s in year t to the total number of households sampled in state s. (10) tertiary sector of state s in year t to the total number of households sampled in state s.
Reason for migration	NSSO migration surveys	Reason for migration of immigrants is one of the following categories: search for employment, search for better employment, under transfer, studies, marriage, parents migration, political problems, others.
Household size	NSSO migration surveys	Number of person in the household
Monthly per capita expenditure	NSSO migration surveys	Monthly expenditure of household over household size
Development exp./SDP	EOPP + updates from RBI	State government development expenditures over net state domestic product
Trade openness	Marjit et al. (2007)	A time varying index to measure the openness of states to trade with other countries. It is available from 1980 to 2002.
Labor regulation	Gupta et al. (2009)	States are divided into flexible, neutral, and inflexible labor regulation. The categories are based on Besley and Burgess (2002).
% of votes received by the ruling coalition	Alok and Ayyagari (2014)	Percentage of votes received by the ruling coalition in the federal election. It is available from 1991-2009 at the state level.
% of seats won by the ruling coalition	Alok and Ayyagari (2014)	Percentage of seats received by the ruling coalition in the federal election. It is available from 1991-2009 at the state level.
Cost of power supply	Planning Commission (http:// planningcommission.nic.in)	Unit cost of electric power supply in Paise/kwh. The data is drawn from two reports on the working of state electricity boards and electricity department (2001-02 and 2011-12), prepared by Planning Commission, (Power & Energy Division), Government of India. The first report in 2001-2 includes data between 1974 and 2001 and the second one covers 2007-2010. For 2002-2006 we generate data using linear interpolation.
Road density	Ghosh & Prabir (2005) + updates from CSO	Total length of roads in km per 1000 km2. It includes both surfaced and unsurfaced roads. Data is available in Ghosh & Prabir (2005) for 1990-96. Updates for 1998-2008 are drawn from infrastructure statistics published by Central Statistical Organization (CSO) and available at: http://www.transportindia.org, and http://mospi.nic.in/

2.B. Construction of poverty and migration variables

Poverty and migration measures are calculated using socioeconomic surveys of India. The National Sample Survey Office or NSSO is the largest organization in India conducting regular socio-economic surveys. The schedule 1.0 of each round is a survey of household consumer expenditures which has been carried out in India since 1950s. However, prior to 1990s, they were not evenly spaced and sampled. The "thick" (large-sample) rounds are conducted about every five years and some "thin" rounds are in between. Datt et al. (1996) provides the time series of state-wise headcount and poverty gap measures from 1951-1992. Since 1986, NSSO has started to conduct and make available "thin" surveys on an annual basis and thick surveys every five years. We obtain the data of 20 rounds (38, 43, and 45 to 62) and among them; the thick surveys are 38th, 43th, 50th, 55th, and 61th rounds. For the missing years, we make use of Datt et al. (1996) data.[31]

round	time span	round	time span	round	time span	round	time span
38	1983	48	Jan-Dec1992	53	Jan-Dec1997	58	July-Dec2002
43	July87-June88	49	Jan-June1993	54	Jan-June1998	59	Jan-Dec2003
45	July89-June90	50	July93-June94	55	July99-June2000	60	Jan-June2004
46	July90-June91	51	July94-June95	56	july2000-june01	61	July04-June05
47	July-Dec1991	52	July95-June96	57	July2001-June02	62	July05-June06

The NSSO's household expenditure survey has a variety of data at household level. It provides information on expenditure patterns, employment (self-employed, labor, etc.), education, occupation, and some other characteristics of households and individuals which enable us to compute a variety of within group measures. It covers all Indian states and follows the Indian Census definition of urban and rural areas. To be classified urban, an area needs to meet several criteria regarding size and density of the population, and the share of male working population engaged in non-agricultural pursuits. However, the surveys are not quite the same and to make comparable indices over time, we make two adjustments: one for a methodology change and the other for seasonal effects.

There was a change in recall period of surveys in 51th to 54th rounds. Until the round 50 and after the round 55 food, tobacco and intoxicant items were asked and reported by a 30-day recall period, but in the rounds 51 to 54 two sub-samples are defined: one with 30-

[31] The data is available at: `http://go.worldbank.org/YMRH2NT5V0`. We use their data for 1960-82 and 1984-87.

day and the other with 7-day recall period for those items.[32] Deaton (2003) and Tarozzi (2007) show that there is an upward bias in total expenditures when the recall period is shorter. To achieve comparability, they suggest using the goods with unchanging recall period to find the true distribution of total expenditures. With plausible assumptions, Tarozzi (2007) shows that if τ represents survey type, and v is the bundle of goods that have the same recall period (non-food items), the distributions of income y in the two sub-samples have the following relation

$$f(y|\tau=1) = f(y|\tau=0) \times E\Big[\frac{P(\tau=1|v)P(\tau=0)}{P(\tau=0|v)P(\tau=1)}\Big|y, \tau=0\Big] \tag{2.4}$$

Where $P(\tau|v)$ is estimated by a logit regression. Using this approach, we impute the correct poverty measures of the rounds 51 to 54.

The second adjustment is done for removing seasonal bias. The timing of NSS rounds shows that the surveys are not distributed evenly across time. Moreover, most rounds are conducted in two adjacent years. Therefore, poverty estimation using each survey separately poses two problems: First, it is not for one exact year, but the rest of variables in the paper are year-specific. Second, some surveys do not cover four seasons (like rounds 47), so the expenditures have a seasonal bias in them. To control for these problems, we estimate the indices for each season (sub-round) and then average them over each specific year. Before 1987, we just have data of 1983, but after 1987 the missing points are fewer (14 of 78), so we interpolate seasonal data after 1987 using Cubic Spline method. This method is a common way to impute high-frequency data from low-frequency (like seasonal from annual). If we have n point and n-1 space in between, this method assigns a cubic polynomial for each space to connect the two points and forces all first and second derivatives to be continuous at margins.

In order to estimate state-level headcount and poverty gap, we utilize the same poverty line as Özler et al. (1996) because our measures are updated for 1960-82 and 1984-87 using their data. The poverty line is recommended by the Planning Commission in 1993 based on calorie intake and adjusted for other years using price indices (for details, see notes of Özler et al., 1996). The Planning commission also has separate estimates of poverty line based on calorie intake in 1983, 1987, 1999, and 2004. As an

[32] In the round 55 these items were asked with both of the recall periods independently and we used the 30-day data.

Table 2.B.1: Headcount ratio using different poverty lines.

Headcount ratio	**our estimations (2004)**			**official estimations (2004-05)**	
poverty line	Datt et al.	Planning Commission	Radhakrishna et al.	Planning Commission	Radhakrishna et al.
Andhra Pradesh	9.51	8.20	32.08	7.5	32.3
Assam	14.21	14.72	34.21	17	36.4
Bihar	29.94	31.34	50.30	32.9	55.7
Gujarat	14.87	12.54	37.70	13.9	39.1
Haryana	7.20	6.63	20.91	9.2	24.8
Karnataka	13.92	11.22	31.68	12	37.5
Kerala	8.16	8.40	16.48	9.6	20.2
Madhya Pradesh	23.17	25.64	46.93	29.8	53.6
Maharashtra	19.16	16.36	40.08	22.2	47.9
Orissa	28.28	35.24	56.29	39.8	60.8
Punjab	4.60	4.60	18.70	5.9	22.1
Rajasthan	18.12	11.73	33.20	14.3	35.9
Tamil Nadu	14.19	14.33	33.47	16.9	37.5
Uttar Pradesh	18.90	21.88	35.83	25.3	42.7
West Bengal	12.97	18.42	30.82	24.2	38.2

alternative, we take these measures and interpolate the line using price indices for the years in between and re-estimate headcount and poverty gap. Our results are robust to this adjustment with a slight change in the level (not significance) of the coefficients. More recent poverty lines for 1993-04 and 2004-05 are presented in Radhakrishna, Sengupta, and Tendulkar (2011). Compared to the older lines their estimates are based on normative expenditure on food, education, and health and are higher than calorie intake lines. Nevertheless, applying these lines (price adjusted for the other years) results in a parallel increase in the poverty measure across states with no change in the qualitative results of our regressions. Table 2.B.1 shows the state-level headcount measures in 2004 using the two new poverty lines and compares it with official estimates. The small difference between the two groups is mainly due to seasonal adjustment because our estimates for 2004 includes the first half of 2004-05 and the second half of 2003-04 surveys. Table 2.B.2 presents our main IV regressions with the poverty estimates by the two new lines which show that the poverty-reducing effect of credit is robust to different poverty lines.

The migration surveys has been conducted in 5 rounds by NSSO since 1980 including 1983 (round 38, schedule 10), 1987-88 (round 43, schedule 10), 1993 (round 49, schedule 1.2), 1999-2000 (round 55, schedule 10), 2007-08 (round 64, schedule 10.2). Using these surveys the migration measures are estimated.

Table 2.B.2: IV results of Table 2.5 using different poverty lines.

Poverty line	**Planning commission**		**Radhakrishna et al.**	
	Rural HC	Rural PG	Rural HC	Rural PG
Lag of Credit to SDP	-0.188**	-0.082**	-0.614**	-0.295**
	(0.083)	(0.036)	(0.270)	(0.137)
Lag of Branches per capita	-0.050	-0.094	0.412	0.200
	(0.246)	(0.107)	(0.711)	(0.293)
Observations	345	345	345	345
R-squared	-0.004	0.009	0.220	0.172
Over ID P-value	0.332	0.427	0.509	0.416

2.C. Additional robustness tests

Table 2.C.1: Robustness check: first stage. — The regressions are estimated similar to Table 2.3 with additional control variables.

	Bank Credit to SDP (%)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Percapita circulation of English newspapers in 1991 × (year-1991) × D_{92}	105.334***	87.600***	104.518***	114.514***	121.457***	126.887***	126.507***
	(9.282)	(8.546)	(9.164)	(11.380)	(11.371)	(12.950)	(12.542)
(year-1977) × B_{60} × D_{77}	-0.018	-0.016	-0.019	-0.014	0.067*	0.057*	0.059*
	(0.019)	(0.023)	(0.019)	(0.021)	(0.036)	(0.030)	(0.031)
(year-1990) × B_{60} × D_{90}	0.060*	0.060*	0.063**	0.038	-0.011	0.012	0.009
	(0.031)	(0.034)	(0.032)	(0.035)	(0.053)	(0.049)	(0.051)
development exp. / SDP (%)	-0.434						
	(0.378)						
trade openness index (N.A. after 2003)		0.072					
		(0.162)					
unit cost of power supply (Paise/KWH)			-0.011**				
			(0.004)				
labor regulation type(flex=+1 neut=0 infl=-1) × post 91 trend dummy				0.288**			
				(0.145)			
road length per 1000 km^2					-0.005**		
					(0.002)		
% of votes received by the ruling coalition (N.A. before 1991)						0.467	
						(1.702)	
% of seats won by the ruling coalition (N.A. before 1991)							-0.315
							(1.140)
Constant	129.796***	142.568**	140.756***	144.454***	135.469***	199.802***	208.707***
	(49.774)	(58.285)	(48.194)	(29.311)	(51.931)	(44.697)	(48.844)
Observations	345	315	345	345	225	210	210
R-squared	0.962	0.964	0.962	0.965	0.970	0.970	0.970
Other Control	YES	YES	YES	YES	YES	YES	YES
Fixed Effects	YES	YES	YES	YES	YES	YES	YES
APchi2	162.206	139.461	156.640	112.166	180.637	161.357	172.721
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WeakID_test	78.195	64.889	73.823	82.174	206.913	183.360	185.200

Table 2.C.2: Robustness test: second stage. — The regressions are estimated similar to Table 2.5 with additional control variables.

Panel A:	**Rural Headcount**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lag of Credit to SDP	-0.160*	-0.212*	-0.230**	-0.168**	-0.096*	-0.215*	-0.207*	-0.250**
	(0.082)	(0.114)	(0.093)	(0.083)	(0.058)	(0.114)	(0.115)	(0.108)
Lag of Branches per capita	-0.282	-0.329*	-0.362*	-0.290	-0.049			
	(0.215)	(0.188)	(0.193)	(0.213)	(0.073)			
lag of development exp. / SDP (%)	-0.831***							
	(0.245)							
lag of trade openness index (N.A. after 2003)		0.167						
		(0.249)						
lag of labor regulation type(flex=+1 neut=0 infl=-1) × post 91 trend dummy			-0.141					
			(0.104)					
lag of unit cost of power supply (Paise/KWH)				-0.016*				
				(0.009)				
Lag of Rural headcount					0.470***			
					(0.075)			
lag of % of votes received by the ruling coalition(N.A. before 1991)						-4.201*		
						(2.520)		
lag of % of seats won by the ruling coalition(N.A. before 1991)							-2.789	
							(1.901)	
lag of road length per 1000 km2								-0.006
								(0.004)
Observations	345	315	345	345	345	210	210	225
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Over-identification test	1.039	0.326	1.488	1.155	1.124	0.849	0.860	0.435
Over-identification P-value	0.308	0.568	0.223	0.282	0.289	0.357	0.354	0.510

Panel B:	Rural Poverty gap							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lag of Credit to SDP	-0.106***	-0.136***	-0.141***	-0.109***	-0.043**	-0.127***	-0.122***	-0.130***
	(0.038)	(0.050)	(0.035)	(0.039)	(0.017)	(0.041)	(0.039)	(0.038)
Lag of Branches per capita	-0.110	-0.156*	-0.138	-0.111	0.005			
	(0.097)	(0.088)	(0.102)	(0.100)	(0.058)			
lag of development exp. / SDP (%)	-0.242**							
	(0.108)							
lag of trade openness index (N.A. after 2003)		0.153						
		(0.103)						
lag of labor regulation type (flex=+1 neut=0 infl=-1) × post 91 trend dummy			-0.097*					
			(0.052)					
lag of unit cost of power supply (Paise/KWH)				-0.004				
				(0.004)				
Lag of poverty gap					0.620***			
					(0.068)			
lag of % of votes received by the ruling coalition(N.A. before 1991)						-2.642***		
						(0.971)		
lag of % of seats won by the ruling coalition (N.A. before 1991)							-1.540**	
							(0.663)	
lag of road length per 1000 km2								-0.002
								(0.001)
Observations	345	315	345	345	345	210	210	225
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Over-identification test	0.316	0.075	1.101	0.390	0.331	0.548	0.577	0.079
Over-identification P-value	0.574	0.784	0.294	0.532	0.565	0.459	0.448	0.779

Chapter 3

Informality and Access to Finance: Evidence from India[33]

3.1. Introduction

A large share of private sector activity in developing countries takes place outside the formal economy. On the one hand, working in informality implies lower regulatory and tax burden. On the other hand, informal firms have limited access to formal services like the legal system and they are less likely to hire skilled labor (LaPorta and Shleifer, 2014). Critically, informality is often associated with lack of access to formal sources of external finance, as both theory and empirical work has shown (Beck, Lin, and Ma, 2014; Straub, 2005). It is not clear, however, whether this relationship is a causal one and, if yes, what the driving factor is. Does lack of access to formal finance discourage entrepreneurs from entering the formal economy or does informality prevent them from accessing formal finance? How different is the effect of financial deepening on formal and informal firms? This paper exploits state-year variation within Indian manufacturing to disentangle the relationship between different types of informality and different dimensions of financial sector development, notably financial depth (commercial bank credit to SDP) and financial outreach (branch penetration). Following the seminal work by Rajan and Zingales (1998), we exploit cross-industry variation in the need for external finance to control for endogeneity biases, as well as instrumental variables.

Informality has different dimensions and can be voluntary or involuntary. From one perspective, some firms or workers exit from the formal sector based on a private cost-benefit analysis of formality, while others are excluded from state benefits because of high registration costs and regulatory burden (Perry, 2007). From a different angle, informality has both inter-firm and intra-firm margins. At the inter-firm margin, some

[33]This chapter is coauthored with Thorsten Beck.

firms, working "underground", completely hide from the state. Others, at the intra-firm margin, are partly formal and partly informal which usually happens in the form of misreported sales and hidden workers. In this paper, we focus on the inter-firm margin of informality, i.e. the exclusion of enterprises from the formal economy, be it voluntarily or involuntarily.

Previous research has shown important links between access to finance and the incidence of informality. On the theoretical level, Straub (2005) presents a comprehensive model of a firm's decision between formality and informality, which includes the decision to tap formal or informal financial markets and shows how the different constraints discussed in the empirical literature affect the threshold size of a company indifferent between formality and informality. In this paper, we use a similar conceptual framework for addressing different dimensions of informality. Consider an economy in which firms (or entrepreneurs) are heterogeneous in initial capital k and can work in either formal or informal sector. The productivity is higher in the formal sector, due to access to formal services; however, firms have to pay an entry cost to overcome the barrier of formality. This barrier includes registration costs, indivisibility of investment and formal property claims, where the latter enables entrepreneurs to use her assets as collateral and thus gain access to formal finance. Figure 3.1 plots the production versus initial capital of a firm in the formal and the informal sector. The marginal production of capital is decreasing and given the real rental price, the profit maximization in the informal sector yields the optimal use of capital as k^*. The intersection of the iso-profit line of k^* and formal production curve gives the level of initial capital $\bar{k}$ above which firms decide to work in formal sector. Based on the firm's decision, three different regions can be distinguished. In the right area, firms become formal and have the highest production and profitability. In the middle, although formality is possible, the optimal choice is producing in the informal sector; entrepreneurs thus voluntarily self-exclude from formality. The left area stands for firms not possessing enough capital to work formally and therefore excluded from the formal sector.

In this setting, better access to financial services helps reduce informality through two different channels:

(A) *Increase transparency*: Access to finance makes the operation of the enterprise at least partly observable and thus reduces asymmetric information and agency

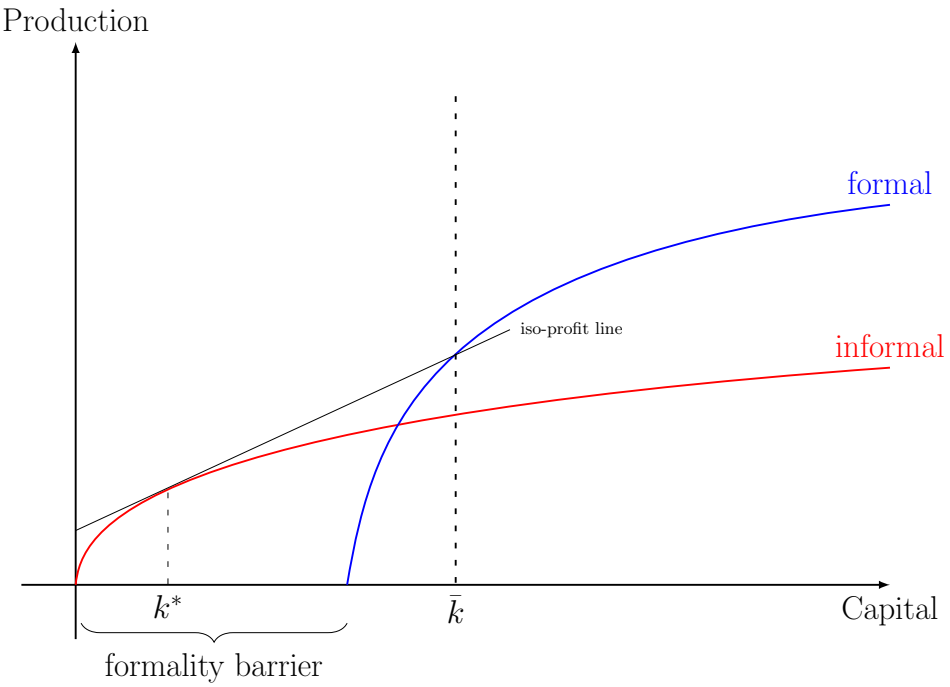


Figure 3.1: forced and voluntary informality of firms.

problems between lender and borrowers, thus facilitating the use of formal finance and other formal services. In this way, financial development helps the firm to overcome the barriers of formality shifting the formal sector production curve to the left (Figure 3.2a).

(B) *Enhance productivity*: By facilitating transactions using short-term credit and funding long-term investment, financial development shifts the productivity of formal firms upwards, while it has no significant effect on informal firms, thus increasing the benefits of producing in the formal sector (Figure 3.2b).

The transparency channel helps credit constrained firms increase their credibility to overcome barriers into the formal sector and thus reduces the incidence of informality. In contrast, the productivity channel has two effects on informality: (i) it increases the production of the formal sector for a fixed level of initial wealth; (ii) it reduces the opportunistic informality and the number of firms that voluntarily produce in the informal sector. We expect a differential importance of these channels across firms of different sizes: Channel (A) is the main mechanism through which finance affects small firms, which the literature has shown to be most likely to be credit constrained (e.g. Beck, Demirgüç-Kunt, and Levine, 2006; Beck, Demirgüç-Kunt, and Maksimovic, 2005).

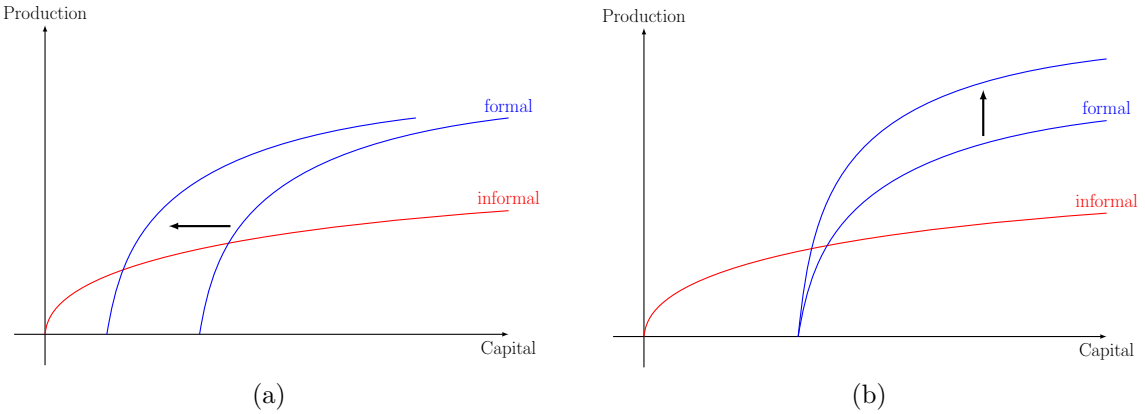


Figure 3.2: Two effects of financial development on informality: (a) reducing barriers to formality, (b) increasing productivity

In contrast, the impact of financial development on firms possessing large fixed assets is through Channel (B). Moreover, we expect both channels to be stronger in industries that are more dependent on external finance.

We examine these hypotheses using Indian manufacturing data. After examining the overall effect of financial development on the incidence of informality, we inspect whether it helps removing formality barriers, by focusing on small firms that are more likely to be excluded from the formal sector. To control for endogeneity biases related to reverse causation and omitted variables, we follow the seminal work by Rajan and Zingales (1998) and exploit cross-industry variation in the need for external finance. Using a difference-in-difference set-up, we gauge whether firms in industries more reliant on external finance are more likely to be formal in states and years with higher levels of financial development. This allows us to control for demand-side effects and for other factors co-varying on the state-year level with financial deepening.

We gauge the effect of financial development on both intensive and extensive margins of the formal sector, i.e. the number of firms and the total production share, and thus both channels discussed above, and focus on two different dimensions of financial development, namely depth, proxied by Credit to SDP, and outreach, proxied by branch penetration. Financial depth relates to the overall credit volume in the economy, independent of which enterprises have access to credit. A high credit volume could thus be mapped to different loan size distributions, including loans mainly to large firms. Financial or bank outreach relates to the ease of access to financial services, including credit. Given the importance of geographic proximity in lending relationships especially

of smaller firms (Degryse and Ongena, 2005), we conjecture that small firms stand to benefit more from financial outreach than large firms. Although these dimensions are not mutually exclusive, the emphasis of one over the other can lead to different policy recommendations.

Our results suggest that both dimensions of financial development are important for increasing the share of formal production in manufacturing. Financial outreach helps reduce formality barriers and thus increases the number of formal firms (channel A), whereas financial depth mainly affects informality through channel (B), increasing productivity of industries dependent on external finance. We also find that this effect is stronger for small firms in the case of financial outreach while financial depth is associated with the incidence of formality of larger firms.

This paper contributes to several literatures. First, we add to the literature on informality. An extensive literature has shown that informality almost always has negative consequences on the aggregate level. In addition to lack of access to formal services, hiding from the government increases distortions and reduces productivity (Gordon and Li, 2009). On the other hand, informality can indirectly hamper firm growth through lack of infrastructure caused by deficits in the government revenue (Kleven, Kreiner, and Saez, 2009). Based on the World Bank Enterprise Surveys, LaPorta and Shleifer (2014) find high levels of informality in developing countries. One of the important differences between formal and informal enterprises is that around 44 percent of informal enterprises list access to financing as the main obstacle of doing business, whereas this number is 21 and 14 percent for small and large formal enterprises, respectively. They also document a large productivity gap between formal and informal firms. In line with this, Hsieh and Olken (2014) show sharp differences in productivity and human capital of managers between formal and informal firms. Our paper investigates how variation in financial sector development across states and over time within India can explain incidence of informality and productivity differences between the formal and informal sectors.

This paper is also related to a small but growing literature on the determinants of informality, most of which focus on specific factors that can explain the incidence and extent of informality. The literature has focused on different areas to explain informality and tax evasion. First, high tax rates and other burdensome regulations increase cost and reduce benefits of formality (De Soto, 1990; Loayza, 1996; Schneider, Buehn, and

Montenegro, 2011; Schneider and Enste, 2000), although low taxation combined with deficient public services can result in similar effects (Dabla-Norris, Gradstein, and Inchauste, 2008; Friedman, Johnson, Kaufmann, and Zoido-Lobaton, 2000; Johnson, Kaufmann, McMillan, and Woodruff, 2000). The relationship between labor market rigidities and informality, on the other hand, seems to be relatively robust (Botero, Djankov, Porta, and Lopez-de Silanes, 2004; Loayza, 1996), as is the effect of entry regulations (Djankov et al., 2003; Klapper, Laeven, and Rajan, 2006). Second, weak institutions that allow rent seeking and predatory behavior by government officials drive firms into informality, an explanation often applied to post-transition economies in Eastern Europe (Shleifer and Vishny, 1993, 1994). A third explanation is that firms try to hide their profits from criminal gangs (Frye and Zhuravskaya, 2000). Fourth, deficiencies in the legal framework (Johnson, Kaufmann, and Zoido-Lobaton, 1998) reduce the benefits of formality – being able to enforce contracts through the court system and thus being able to deal with a broader set of trading partners at arms-length. In our empirical assessment, we thus have to discriminate between legal system deficiencies and financial sector development not related to the legal system. Finally, several empirical papers have shown the importance of financial constraints in explaining variation in informality. A recent cross-country study shows that firms are more likely to produce in the formal sector in countries with more effective credit registries and higher branch penetration, an effect that is stronger for smaller and geographically more remote firms and firms in industries with a higher dependence on external finance (Beck et al., 2014). Compared to this literature, we exploit within-country variation in financial development and compare the effect of two different dimensions of financial development, depth and outreach.

Second, we add to a large literature on the real effects of financial deepening. Starting with King and Levine (1993), a large literature using different aggregation levels and measures of financial depth has shown a positive relationship between financial depth and economic growth, a relationship that goes more through productivity growth than capital accumulation (e.g., Beck, Levine, and Loayza, 2000). While the recent crisis and recent studies have shown important non-linearities (e.g., Arcand, Berkes, and Panizza, 2012), there seems a wide-spread consensus in the literature on a strong effect of financial deepening on economic growth for developing countries, such as India. The literature has also related financial development to financing obstacles of small and medium-sized

enterprises, showing that obstacles are lower in countries with higher levels of financial development (Beck et al., 2006) and that these obstacles are less growth constraining in countries with deeper financial systems (Beck et al., 2005). Our paper adds to this literature by relating within-country variation in financial development to the incidence of formality, thus another important channel through which financial sector development can impact the level and structure of GDP. Unlike previous papers, we also distinguish specifically between the two dimensions of financial depth (focus of most of the finance and growth literature) and financial outreach.

Finally, our paper also adds to a flourishing literature on economic development in India, which has linked sub-national variation in historic experiences and policies to differences in growth, poverty levels, political outcomes and other dependent variables (see Besley et al., 2007 for an earlier survey). Specifically, researchers have focused on differences in political accountability (Besley and Burgess, 2002; Pande, 2003), labor market regulation (Besley and Burgess, 2004; Dougherty et al., 2011; Hasan et al., 2007), land reform (Besley and Burgess, 2000; Iyer and Banerjee, 2005), trade liberalization (Topalova, 2010) and gender inequality (Iyer et al., 2012). Directly related to our paper, Burgess and Pande (2005) relate a social banking policy on branching to differences in poverty alleviation across states. Ayyagari, Beck, and Hoseini (2013) explore the relationship between financial deepening post-1991 liberalization and poverty-levels. Our paper adds to this literature by focusing on cross-state differences in financial deepening after the 1991 liberalization episode and by comparing the effects of two different dimensions of financial development – total credit volume and branch penetration of financial institutions – on the incidence of informality.

The remainder of the paper is structured as follows. Section 3.2 describes the data we will be using and section 3.3 the methodology. Section 3.4 discusses our results and section 3.5 concludes.

3.2. Data

This section describes the different data sources and variables we use to gauge the relationship between the incidence of informality and access to formal sources of external

finance. Specifically, this section describes (i) the indicators of informality, (ii) the indicators of financial depth and outreach, and (iii) the industry characteristics that allow us to gauge the differential impact of financial sector development on the incidence of informality across different industries.

3.2.1. Gauging the incidence of informality

We use firm-level surveys for the formal and informal sectors to construct gauges of the incidence of informality on the state-industry level. Specifically, we have available data for the Indian manufacturing sector for 5 years: 1989-90, 1994-95, 2000-01, 2005-06, and 2010-11. Each year has two data sources: (i) the annual survey of industries (ASI) and (ii) the national sample survey on unorganized manufacturing sectors (NSS). The ASI covers factories employing above 10 employees using power and those with 20 employees or more without using power. In each year, all factories with more than 100 employees plus at least 12% of the rest are sampled. The sample is representative at the state and 4-digit NIC code levels.[34] The second data source is the NSS enterprise survey which covers small manufacturing units that are not covered by ASI. Its sampling strategy is based on the number of enterprises in each village/town which is computed using the most recent Economic Census.[35] Sample weights which show the number of firms the sample represents are provided for both surveys. Table 3.1 shows the number of observations for each ASI and NSS surveys across the five waves.[36]

Table 3.1: list of surveys and the number of samples

year	**1989-90**	**1994-95**	**2000-01**	**2005-06**	**2010-11**
No. sample in ASI	49,323	57,908	37,055	49,637	46,843
No. sample in NSS	123,321	192,029	222,529	80,637	99,243

To gauge the incidence of informality, we use two dummy variables at the enterprise level. The first one refers to **general registration** and indicates whether the enterprise is

[34] Up to 4-digit level, the NIC code is identical in structure to International Standard of Industrial Classification (ISIC). In the additional digits, it incorporates the national characteristics.

[35] Economic Census is a complete count of all economic units in the country synchronized with the house listing operations of the Population Census. The 6 rounds Economic censuses in India are conducted in 1977, 1980, 1990, 1998, 2005, and 2012. NSS surveys of manufacturing enterprises in the unorganized sector are conducted as a follow up of each Economic Census. The detailed information about the sample design of each survey is available at `http://mospi.nic.in/`.

[36] Given the variation in NSS coverage, we are concerned that the surveyed firm population might vary significantly over time. When comparing the share of firms in externally dependent industries across the five survey waves, however, we cannot any significant trend correlation with NSS coverage.

registered under any act or authority. The second one is **tax registration** and indicates whether the firm is registered with the tax authorities or not. All sampled firms in ASI are registered under the Factories Act and are taxpayers. The NSS sample surveys have information about registration under any act or agency. We can find out about tax registration by checking whether the firm pays any sales tax (distributive expenses) or not. Thus, a firm is registered for tax if it is in ASI or it is in NSS and has nonzero distributive expenses. We do not have information about tax registration or payment in NSS 89, and therefore, we use this year just for the regression of general registration.

One concern regarding the validity of formality measures is whether firms truthfully state their registration status. To control for such a bias, both ASI and NSS are conducted independent of tax purposes and no information about the identity of the enterprise is passed to other agencies.[37] In addition, although the honesty of firms can be questioned for tax registration, there is no meaningful incentive to overstate the registration status under other agencies. In the NSS surveys, there are at least ten listed authorities under which the firm can be registered,[38] most of which are designed to help SME growth and provide a variety of services and facilities such as training, technological and financial assistance, etc. Hence, the general registration index is less vulnerable to misreporting bias and measurement error and we will focus on it for the most part of our empirical work.

We use the information on firms' registration status to construct six different indicators of informality on the aggregate level. Panel A of Table 3.2 shows the weighted averages of the different registration indices in each year. The first two rows show general and tax registration rate among firms. Each observation is weighted with the number of firms it represents. The general registration rate increased from 8 percent to 12 percent between 1989 and 1994, declined in 2000 and 2005 to 10 percent, before it went up again to 15 percent in 2010. The tax registration rate slightly increased till 2005 but doubled from 2005 to 2010, when around 3 percent of firms were registered with tax authorities. Considering the value-added share of formal and informal firms instead of the numbers

[37] It is highlighted in the questionnaires that "according to the Collection of Statistics Act, violation of any of the confidentiality and secrecy of the information by statistics officer or field staff is prosecuted by or with the consent of appropriate agency."

[38] Some examples are District Industries Centre as Small Scale Industry (SSI), Khadi and Village Industries Commission, Council for Technological Upgradation, Small Industries Development Bank, Development Commissioners of handicraft/handloom/jute, Coir/Silk Board.

gives a somewhat different picture. In rows (3) and (4) we present the weighted sum of the value-added of registered firms divided by the weighted sum of the value-added of all firms. The numbers indicates that although the number of registered and tax-paying firms is small, they comprise a big and growing slice of the value added in the manufacturing sector, reaching 91 and 78 percent in 2010, respectively. Finally, the numbers in rows (5) and (6) are employment shares of formal firms which equal the weighted sum of the number of workers of registered firms over the weighted sum of the workers of all firms. The trends in the value-added and employment is similar to the number of firms, first dropping and then increasing again.

Table 3.2: Panel A – Summary of informality measures (weights are applied)

	name	Description	1989	1994	2000	2005	2010
(1)	**Reg**	percentage of registered under any act	7.72	11.78	10.63	10.30	15.05
(2)	**Treg**	percentage of registered under tax		1.34	1.65	1.71	3.21
(3)	**Vreg**	VA share of registered under any act (%)	80.80	83.76	79.94	85.54	91.48
(4)	**Vtreg**	VA share of registered under tax (%)		75.77	69.86	78.13	78.05
(5)	**Ereg**	employment of registered under any act (%)	24.58	33.83	32.49	32.407	45.42
(6)	**Etreg**	employment of registered under tax (%)		21.62	18.91	20.27	31.31

Panel B – Comparison of informality measures with official estimations.The base year for the first three rows is 2005. The last row is to be compared with row (4) in the Panel A.

	1989	1994	2000	2005	2010
Official Manufacturing GDP	38.9	49.3	72.8	100	160.6
Gross output	33.4	50.4	65.6	100	178.1
Gross value added	35.8	47.5	59.5	100	193.1
Official Registered Man. GDP (%)		63.2	61.6	67.5	69

To examine the robustness of our measure, we cross-check the overall numbers with comparable GDP estimations of Indian manufacturing sector published by Central Statistical Office (CSO), Government of India. Panel B of Table 3.2 compares the official estimation of net manufacturing GDP in India versus our estimations of gross output and value-added, using 2005 as the base year. The official estimations are at constant price and account for depreciation. We also normalize our estimated values by state level price indices,[39] but our measures are in gross terms. There are several reasons for differences across the different variables. First, they might be due to differences in price adjustment and depreciation. In addition, the CSO publishes net GDP data on registered and unregistered manufacturing. Compared to our methodology, the CSO's

[39]The price index is published by Labour Bureau as "consumer price index for industrial workers".

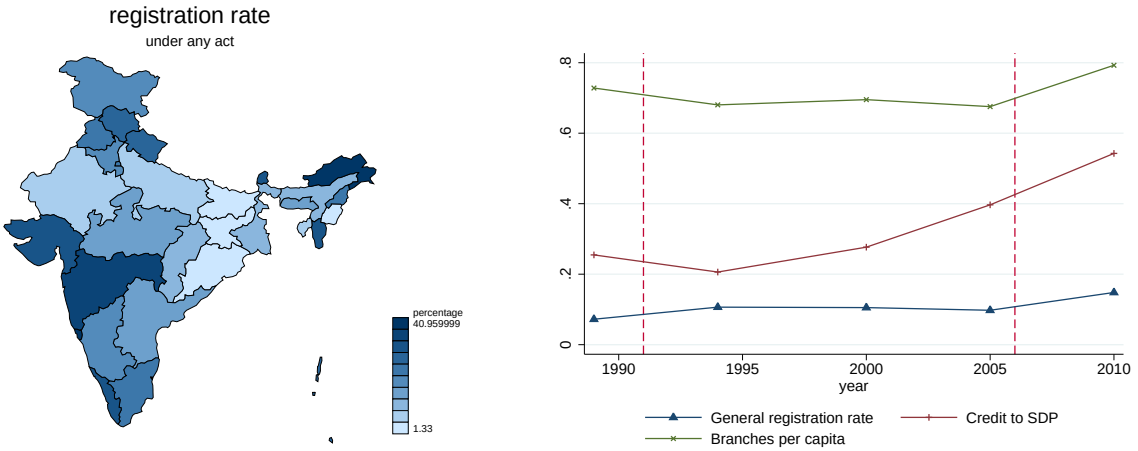
estimation is based on labor input and production per labor, counting just firms in ASI as the registered sectors.[40] Since we also take into account registered enterprises in the NSS that are not covered in ASI, our estimates of formal production tend to be higher. Nevertheless, we observe parallel trends in the value-added share of firm registered under any act and in similar estimations by CSO.

Appendix Table 3.A.1 provides the average share of registered firms across industries, both using general and tax registration and across the three dimensions of (i) share of firms, (ii) share of value added and (iii) share of employees, as well as the number of firms these averages are based on, averaged over the five survey waves. We note a substantial variation across industries in the incidence of informality. While in petroleum refineries 81% of activities are undertaken in registered companies, only 3% of companies in the tobacco industry are registered under any act and less than 0.5% are registered with tax authorities, even though their share in total employment is over 11% and their share in value added over 58%.

Appendix Table 3.A.2 provides similar information on the incidence of informality across states, again averaging over the five survey waves. While over 40 percent of firms are registered in Goa, only one percent are registered in Orissa. Figure 3.3a provides graphical illustration of cross-state variation of registration average over time and industries and Figure 3.3b shows the average trend of registration rate over time.

We use these indices of informality on the firm-level to compute gauges of the incidence of informality on the state-year-industry level. Specifically, we combine the firm-level data of ASI and NSS and then collapse them at state-year-industry level using sample weights. Since our index of financial dependence is drawn from Rajan and Zingales (1998), we aggregate our measures to the same industry categories that they suggest. Our indicators of formality are based on the share of firms, share of value added and share of employees, both for general registration and tax registration. Overall, we have 35 states, 5 years, and 34 industries, but the number of observations is only 3,335 because smaller states do not host all industries. In the regressions, we have fewer observations because of missing data on some of the independent variables in some states, industries and years. Table 3.3 Panel A provides the descriptive statistics on our indica-

[40] The methodology of CSO's is described at: `http://mospi.nic.in/Mospi_New/upload/brochure_2004-05.pdf`



(a) (b)

Figure 3.3: (a) registration rate across states. (b) registration rate, credit to SDP and Branches per capita over time. The two vertical lines show two financial reforms in India in 1991 and 2006. Registration rate is scaled between 0 and 1 for better illustration.

tors. On average, 10.6 percent of firms are registered under any act, but only 2 percent for tax authorities.

3.2.2. State level indicators of financial development and control variables

We construct several time-variant indicators of financial and economic development as well as tax enforcement on the state-level. Panel A of Table 3.3 provides descriptive statistics of the time-variant state-level variables. Appendix Table 3.A.2 provides state-level averages of the different variables.

First, we use two indicators of financial sector development, capturing the two dimensions of financial depth and financial outreach. Specifically, Credit to SDP is the outstanding amount of credit utilized in each state divided by State Domestic Product. It corresponds to a standard cross-country indicator, Private Credit to GDP, which has been extensively used in the finance-growth literature (e.g. Beck et al., 2000). Our measure of financial outreach is Branches per capita and is the number of bank branches per 10,000 people in each state and year. Average Credit to SDP varies from 0.11 in Nagaland and Manipur to 1.42 in Chandigarh, while branches per capita is 0.42 in Bihar ranging up to 3.28 in Goa. Figure 3.3b shows that Private Credit increased rapidly after 1995, while branch penetration picked up after 2005. We will come back to these time trends below when we discuss instrumental variables.

Table 3.3: Panel A: Summary statistics – Number of firms is applied as weights.

	Reg	Treg	Vreg	Vtreg	Ereg	Etreg	Credit to SDP	Branch per capita	RZ index	Forward link-ages
Mean	10.65	2.02	47.88	38.84	22.71	13.42	0.348	0.716	0.06	0.78
Standard error	15.2	5.49	29.74	29.48	23	17.32	0.225	0.22	0.3	0.54
De-trended SD across:										
state	11.32	3.91	21.66	21.4	16.84	13.15	0.183	0.21		0.17
year	7.23	2.83	13.24	12.95	9.89	6.87	0.131	0.055		0.27
industry	13.1	5.11	24.33	25.18	19.18	15.28			0.29	0.51
state-year	6.6	2.61	10.96	11.01	8.31	5.83	0.079	0.047		0.15
state-industry	9.47	3.62	16.37	17.26	13.76	11.56	0.047	0.045	0.08	0.19
year-industry	6.5	2.72	11.39	11.18	8.74	6.5	0.051	0.017	0.06	0.27
state-year-industry	6.11	2.57	10.37	10.51	7.87	5.88	0.04	0.029	0.04	0.16

Logarithm of	**Production of registered**	**Production of not registered**	**VA of registered**	**VA of not registered**	**SDP per capita(Log)**	**Gov. exp./SDP**	**Enf. exp. per firm (log)**
Mean	14.07	13.33	12.67	12.77	9.804	0.192	-0.512
Standard error	2.11	1.46	1.96	1.4	0.789	0.062	0.697
De-trended SE across:							
state	1.51	1.03	1.47	1	0.732	0.06	0.659
year	0.73	0.54	0.79	0.52	0.418	0.022	0.378
industry	1.57	1.2	1.38	1.13			
state-year	0.52	0.49	0.57	0.47	0.36	0.02	0.364
state-industry	0.92	0.71	0.91	0.7	0.176	0.012	0.165
year-industry	0.6	0.52	0.64	0.51	0.153	0.006	0.099
state-year-industry	0.52	0.48	0.55	0.46	0.173	0.01	0.134

Panel B: correlation table. Number of firms is applied as weights.

	Reg	**Treg**	**Vreg**	**Vtreg**	**Ereg**	**Etreg**	**Credit to SDP (log)**	**Branch per capita**	**Enf. exp. per firm (log)**	**SDP per capita (log)**	**Gov. exp. / SDP**	**RZ index**
Treg	0.600***											
Vreg	0.460***	0.301***										
Vtreg	0.314***	0.291***	0.850***									
Ereg	0.667***	0.399***	0.803***	0.597***								
Etreg	0.512***	0.428***	0.826***	0.850***	0.849***							
Credit to SDP (log)	0.293***	0.213***	0.370***	0.324***	0.377***	0.366***						
Branch per capita	0.314***	0.240***	0.267***	0.205***	0.333***	0.312***	0.575***					
Enf. exp. per firm (log)	0.291***	0.188***	0.219***	0.119***	0.310***	0.231***	0.619***	0.583***				
SDP per capita (log)	-0.07***	-0.08***	-0.17***	-0.17***	-0.14***	-0.156***	-0.16***	0.0383**	-0.0397**			
Gov. exp. / SDP	0.178***	0.122***	0.189***	0.110***	0.219***	0.160***	0.161***	0.161***	0.0626***	0.0404**		
RZ index	0.244***	0.231***	-0.06***	-0.09***	0.142***	0.0795***	-0.11***	-0.00353	-0.062***	0.0218	0.0163	
Forward linkages	0.0285*	0.173***	-0.11***	-0.12***	-0.04**	-0.0446**	-0.05***	-0.0103	-0.0307*	0.0052	0.0088	0.500***

In investigating the link between financial development and informality, we also control for other time-varying state characteristics. State Government Expenditure to SDP is total state government expenses over SDP and tax enforcement per firm is the component of state government expenditure on collection of taxes and duties divided by the estimated number of firms in the state. Better public service provision and higher tax enforcement might reduce the informality of enterprises. Government expenditures average 19 percent of SDP, ranging from 0.071 in Delhi to 1.119 in Sikkim. Enforcement expenditures per firm range from 0.052 in Maharashtra to 5.571 in Delhi. We also control for SDP per capita which is the net state domestic product per capita at constant price and a proxy for income levels, but since it is highly correlated with financial development measure, we drop it in the main regressions, but confirm our findings in robustness tests including it. SDP per capita ranges from 9146 in Bihar to 88845 in Chandigarh. Finally, to control for the quality of state's infrastructure for manufacturing establishment, we control for unit cost of power defined as cost of electricity in Paise/KWH. The data source for unit cost of power is Planning Commission's two reports on the working of state electricity boards and electricity department (2001-02 and 2011-12). All other state-level control variables are drawn from Reserve Bank of India's Data Warehouse.

3.2.3. Industry characteristics

To explore the differential effect of state-level policies on informality in different industries, we use an industry characteristic that captures the need for financial services and thus the potential benefit of access to formal finance or opportunity costs of informality. Specifically, we use the RZ index of financial dependence which is from Rajan and Zingales (1998) and equals the median of firm level measure "(capital expenditures cash flow)/capital expenditure averaged over 1980s" for 34 industries. This indicator, computed for a group of large listed enterprises in the U.S., for which the supply curve can be expected to be almost perfectly elastic, is supposed to indicate the need for external finance based on inherent industry characteristics and is exogenous to the actual use of external finance by firms in India. As our sample period spans the 1990s and 2000s in India and this measure is computed for the U.S. in the 1980s, concerns on different technologies in both countries might not be as critical. The level of dependence on external finance shows the potential benefits for firms from being formal and having access to

formal financial services. Appendix Table 3.A.1 shows that external dependence ranges from -0.45 in tobacco to 1.49 in drugs.

In addition, we employ another industry-level index to capture the exogenous variation in tax compliance. The Indian taxation of enterprises comprises direct and indirect taxation on both central and state level. While direct taxes are mainly levied by the central government, the main source of states' tax income is their sales tax. Union excise duties on all manufacturing products and service tax on services are also levied by the central government. Excise duties, covering all manufacturing products, turned to the VAT – named MODVAT – in 1985 and expanded to ad-valorem rates in 1993 for the majority of products. Hoseini (2014b) shows that under the value-added tax system, upstream industries that are forwardly linked to others have higher risk of detection and thus are more likely to be formal. As over the period of our study, the manufacturing sector of India has been under the value-added tax, we measure the forward linkages of each industry to capture the exogenous variation in the risk of noncompliance in the value-added tax system. The forward linkages index, based on Hoseini (2014b), is the row sum of Leontief inverse matrix of Indian economy reflecting the flow of products going to other industries rather than final consumers. This index is calculated for each industry using the input-output tables of the Indian economy. The I-O tables are available for 1993-94, 1998-99, and 2003-04 and we use the average of the index over time. The indicator ranges from 0.27 in tobacco products to 2.57 in non-ferrous metals.

Panel B of Table 3.3 presents correlations across the different state-industry level variables. We find that the share of firms registered under any act or under tax authorities is positively correlated with both financial sector indicators, with both industry characteristics, with SDP per capita and with enforcement expenditures per firm and negatively with government expenditures to SDP. Credit to SDP and branch penetration are positively correlated with each other, with a correlation coefficient of 58 percent. However, other state-level variables are also significantly correlated with financial development. Finally, external dependence and forward linkages are positively and significantly correlated with each other, with a correlation coefficient of 0.5.

3.3. Ocular econometrics and methodology

Before presenting regression results on the relationship between financial development and informality, this section provides some preliminary facts about this relationship using Indian manufacturing data and explains our methodology to identify the significance of each channel.

3.3.1. Ocular econometrics

Figure 3.4 plots the general and tax registration rates versus our two financial development variables across states. It can be seen that both financial variables have a positive relationship with general and tax registration rates, but the observations of branches per capita and tax registration are more concentrated along the fitted line, compared to the other three relationships. Both relationships are significant at the 1 percent level for branches per capita and at the 5 percent level for credit to SDP.

As shown in Figure 3.2a, theory suggests that one effect of access to finance on informality is cutting the barrier to formality and enabling firms to overcome the costs of formality. To identify this mechanism, we focus on the sample of smaller firms that are more likely to be excluded from the formal sector. Figure 3.5 plots registration rates versus our financial development indicators for the sample of smaller firms, defined as establishments with fixed assets less than the 25th percentile of the respective industry in each year. The figure suggests a positive relationship between formality and branches per capita, with a higher slope than in the overall sample (significant at the 1 percent level), while the relationship with credit to SDP is insignificant.

The second channel through which finance can alleviate informality is increasing productivity of the formal sector (Figure 3.2b). As mentioned above, this channel has two effects: reducing opportunistic informality and boosting the production of formal sector firms. In order to identify this channel, we employ the exogenous variation in the dependence on external finance among industries. In Figure 3.6, we compare the registration and financial outreach relationship between two groups of industries: above the 75th and below the 25th percentiles of the RZ index of financial dependence. It can be clearly seen that the positive relationship is stronger for industries with larger need for external finance suggesting less opportunistic informality in these industries.

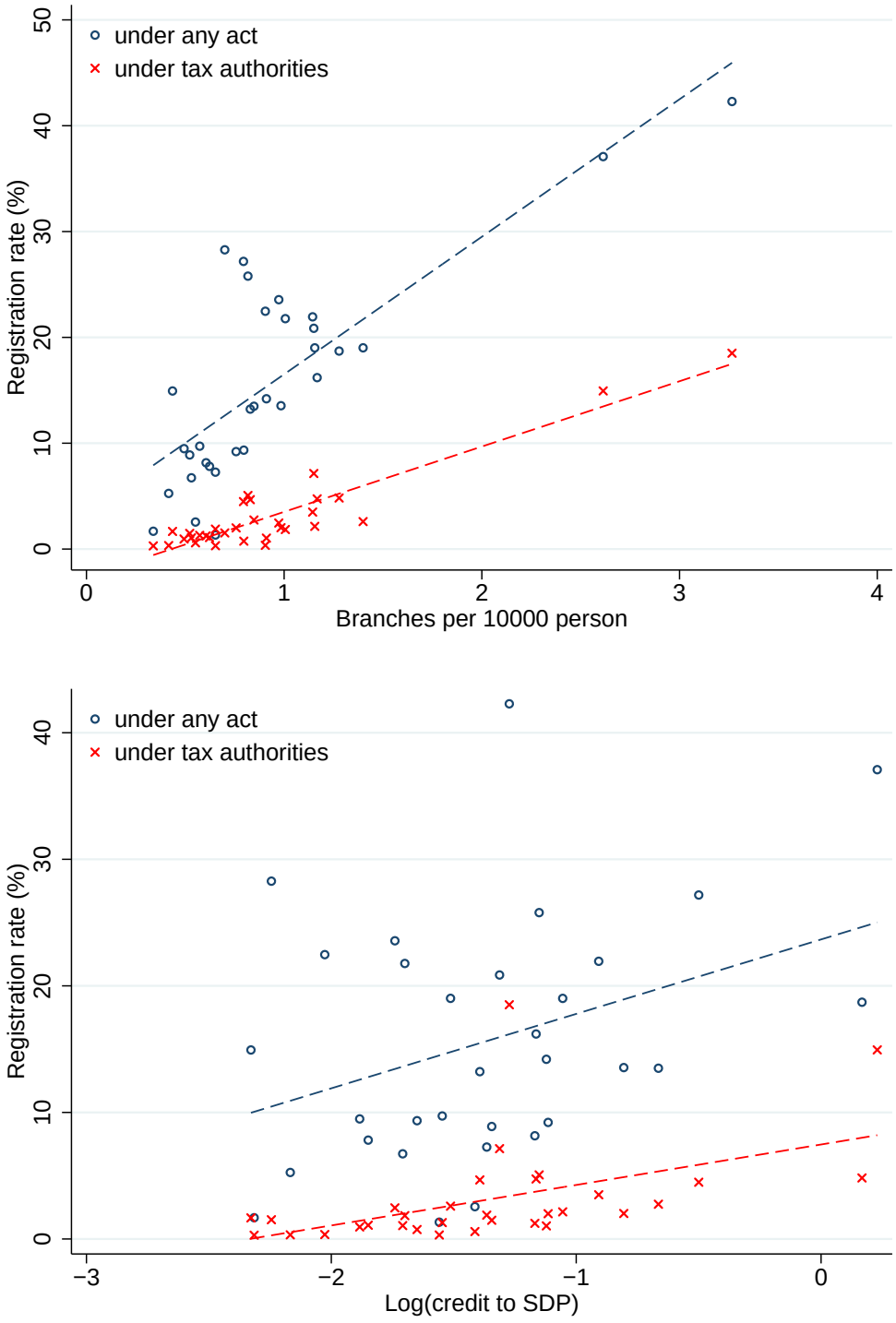


Figure 3.4: Registration rate vs. financial breadth and depth averaged over states

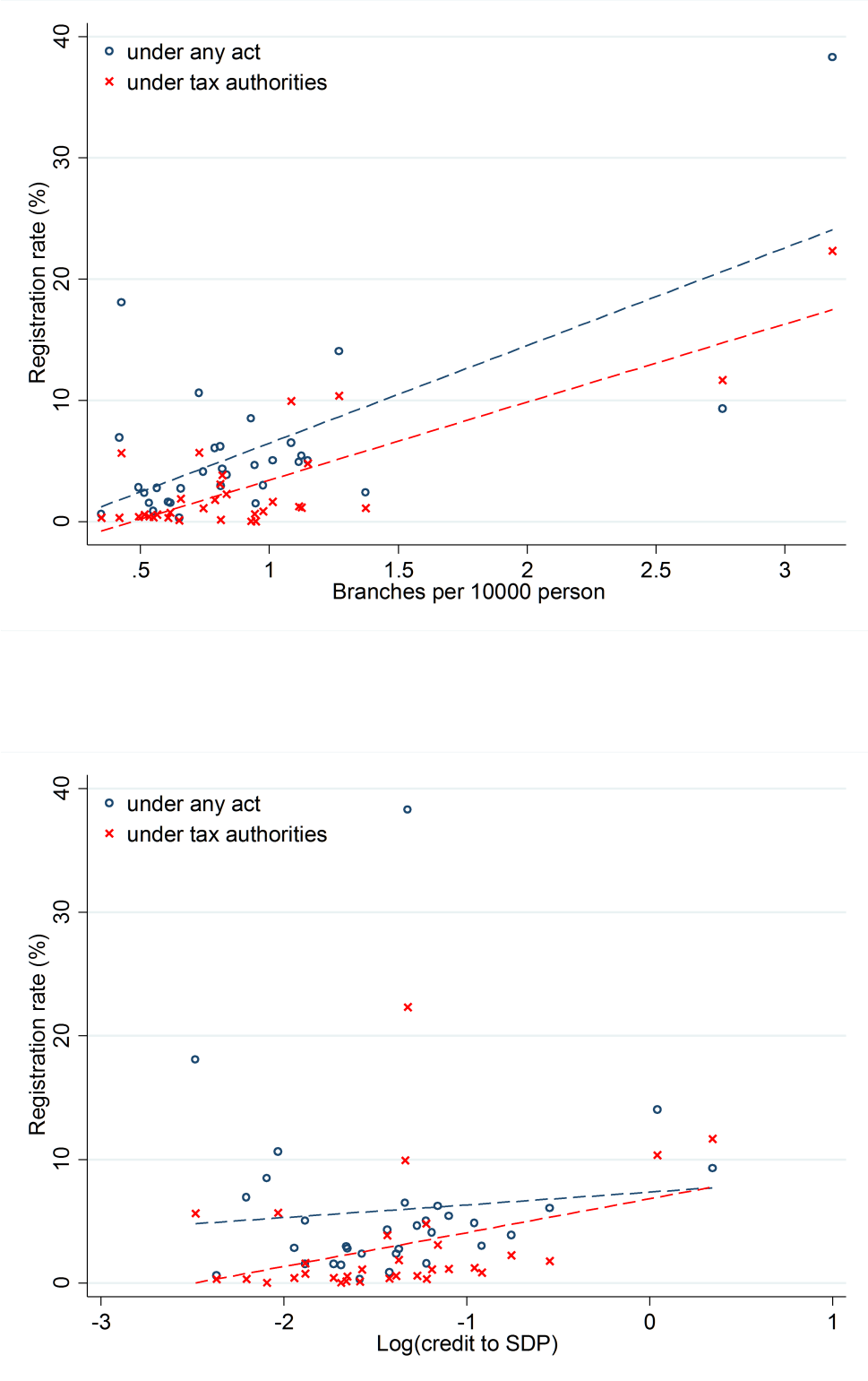


Figure 3.5: Registration rate vs. financial breadth and depth averaged over states for the Sample of firms with fixed assets below 25th percentile

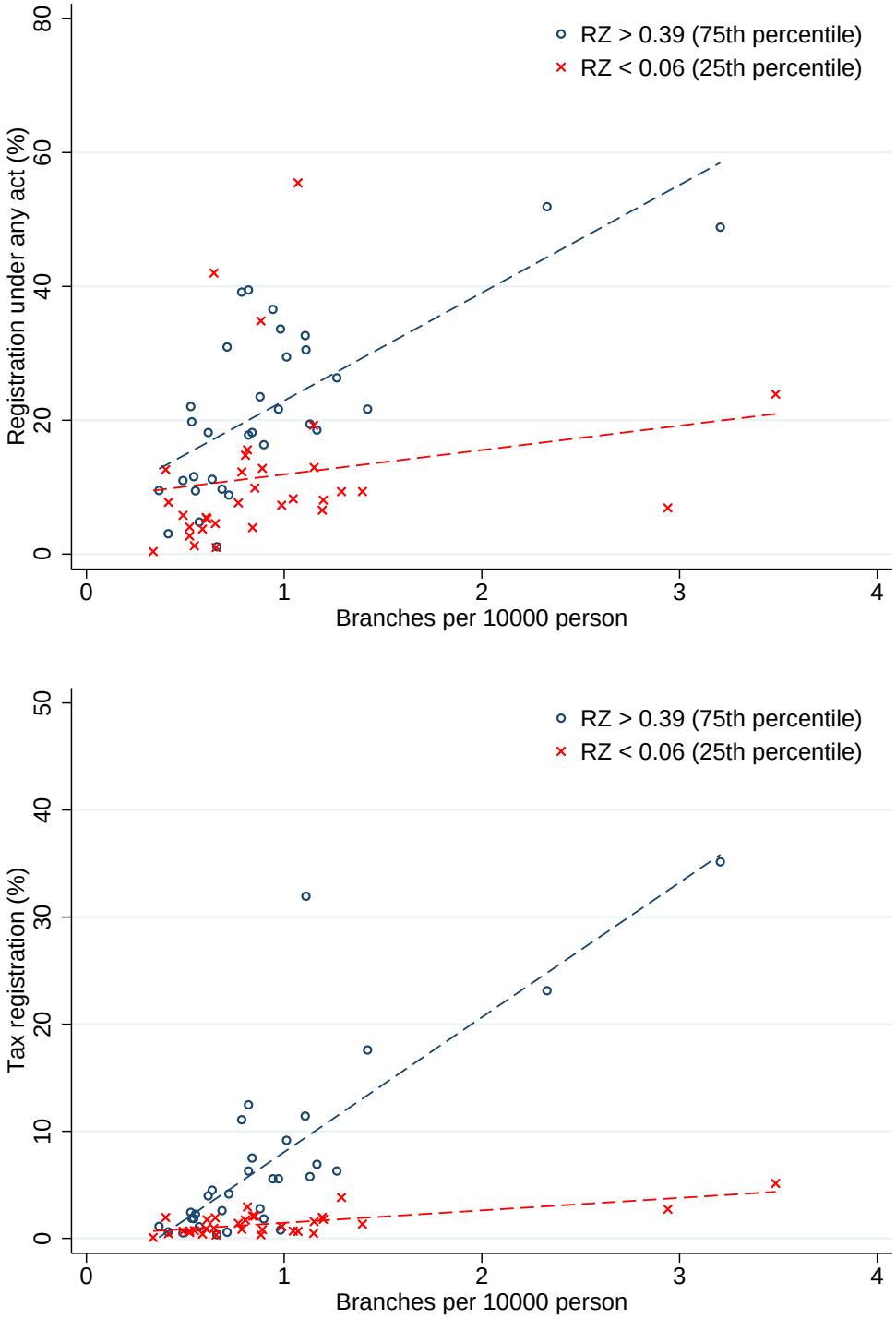


Figure 3.6: Registration rate, financial dependence and financial penetration averaged over states

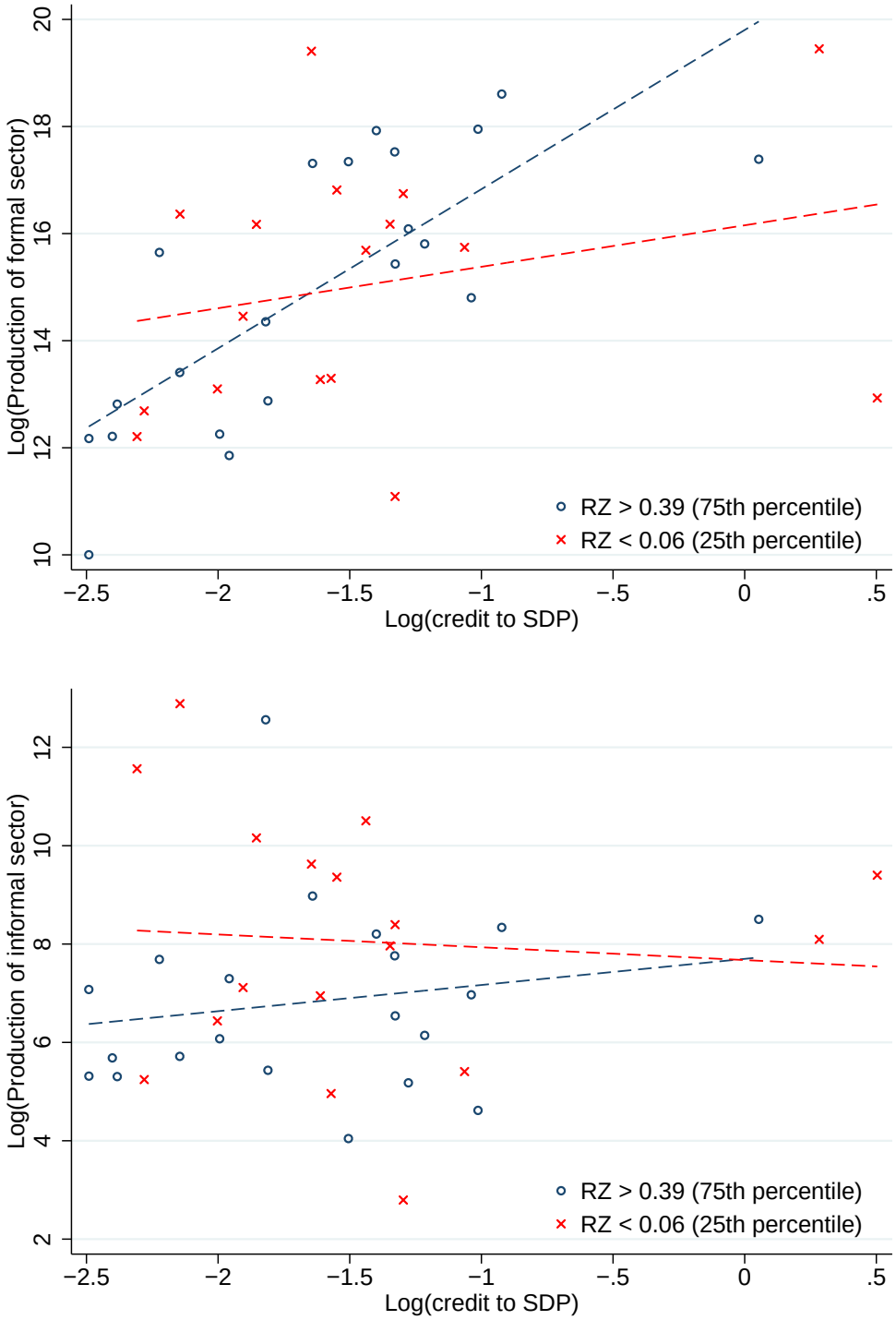


Figure 3.7: Production, financial dependence and financial deepening in formal sector (registered under any act) and informal sector (not registered for any act), averaged over states.

In Figure 3.7, we use the same structure to compare the link between production and financial depth in the formal and informal sector. This figure shows that in the formal sector, production of industries more reliant on external finance is highly sensitive to credit to SDP (significant at 1 percent level), while this sensitivity is much less for other industries. One the other hand, this pattern is much weaker and insignificant in the informal sector.

3.3.2. OLS and difference-in-difference regressions

To formally estimate the overall effect of state-level financial development on registration rates, we use different methodologies. First, we use the following difference-in-difference setting as the baseline.

$$Inf_{ist} = a_i + b_s + c_t + \alpha_1 FD_{st} + \alpha_2 X_{st} + \varepsilon_{ist} \tag{3.1}$$

where Inf_{ist} is one of the informality indices in industry i, state s and year t. a_i, b_s , c_t are industry, state and year fixed effects, respectively, FD_{st} is one of our two financial development indicators in state s and year t, and X_{st} is a vector of control variables including government expenditure to SDP, unit cost of power, and enforcement expenditure per firm in state s and year t.

Because our regressions are for the whole of India, in each regression, we use the estimated number of firms in state s, year t, and industry i as weights for the observations. In addition, to control for the underestimated standard error in the difference-in-difference setting, as suggested by Bertrand et al. (2004), we cluster our estimation at the state level.

To estimate the differential effect of state-level financial development and enforcement activity on the incidence of informality across firms with different needs for external finance, we utilize the following difference-in-difference setting for estimation.

$$inf_{ist} = a_i + b_s \times c_t + \beta_1 RZ_i \times FD_{st} + \beta_2 RZ_i \times X_{st} + \beta_3 FL_i \times Y_{st} + \varepsilon_{ist} \tag{3.2}$$

where RZ_i is the Rajan-Zingales index of external dependence for industry i, and FL_i is forward linkage for industry i, Y_{st} is a vector of state-level log of enforcement per firm, and the rest of variables are the same as equation (3.1). By saturating the model with

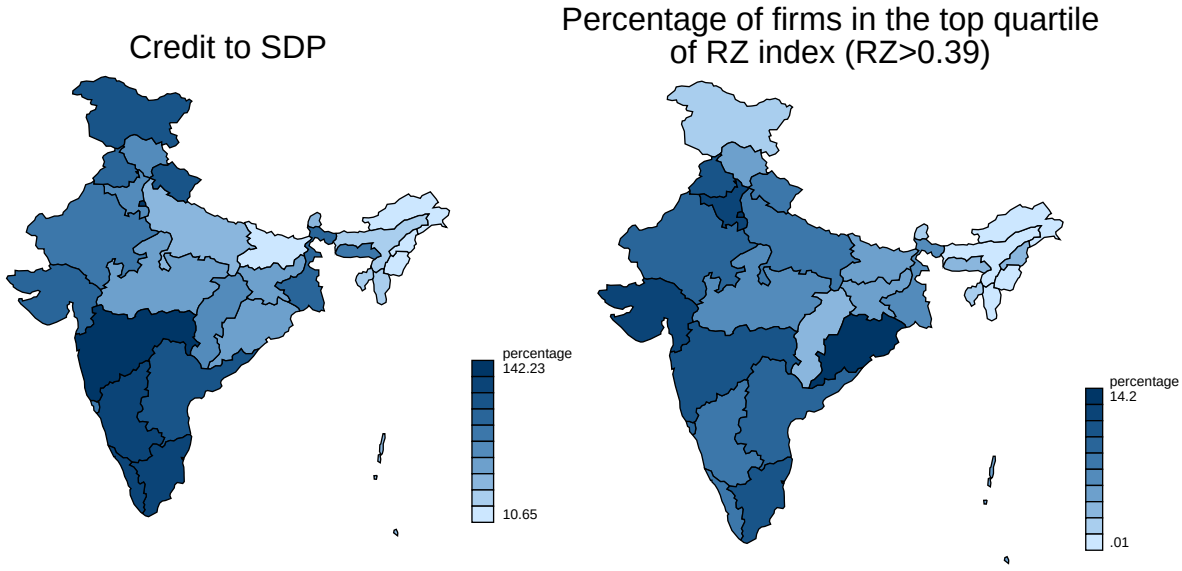


Figure 3.8: Comparison of the pattern of financial development and concentration of externally dependent industries

industry and state-year fixed effects, we focus on the relative effect that time-variant state-level variables have on the incidence of informality on the state-industry-year level. In order to examine the productivity-enhancing effect of finance on the formal sector, we use the same regression setup as (3.2), but instead of the incidence of formality, we use the levels of production and value-added in formal and informal sectors as the dependent variable.

One concern of using Rajan and Zingales methodology is the agglomeration of externally dependent industries in the states with higher level of financial development which reduces state-industry variation. To check whether industries more dependent on external finance choose states with higher levels of financial development, in Figure 3.8, we map the distribution of credit to SDP and share of externally dependent industries across different states. The comparison of two figures suggests no local concentration of externally dependent industries in financially developed states. The correlation between external dependence on the state level (weighted by number of firms per industry) and credit to SDP is 10 percent.

3.3.3. Instrumental variable strategy

In further analysis of the relationship between informality and access to finance, we employ an instrumental variable approach using two instruments for financial development

in India. This enables us to better control for reverse causation and omitted variable bias. In the period of our study, the financial sector of India experienced two major reforms, though with important differences across Indian states. The first one is the financial liberalization in 1990s which led to a rapid financial deepening in India, and the second is the post-2006 financial inclusion reform. In the following, we first explain the two policy changes, then specify the instruments and estimation methodology.

Indian Financial Liberalization

As documented in Ayyagari et al. (2013), following a severe balance of payments crisis in 1991, there was a substantial liberalization of India's financial sector as part of an economy-wide liberalization process. These reforms included de-regulation of interest rates, reduction in the volume of directed credit and entry of new privately-owned financial institutions. Reforms of the regulatory and supervisory framework and the contractual environment also supported financial deepening in the subsequent decades. Figure 3.3b shows the trends of financial depth and outreach before and after the liberalization. Although branches per capita almost remain constant for around 15 years after 1991, Credit to SDP, after an initial fall in 1994, substantially increased for the next years. As documented by Ayyagari et al. (2013), however, this financial deepening process was uneven across different Indian states. This heterogeneity over time and across states provides us a rich identification tool that we can relate to variation in the incidence of informality.

Following to Ayyagari et al. (2013), we link cross-state variation in the effects of financial liberalization on financial deepening to cross-state variation in the media environment. Free media and press help reduce information asymmetries between investors, as well as enterprises and banks, in addition to its impact on the competitive environment in the financial and real sectors. To proxy for state-level media environment at the time of financial liberalization, we use the same measure as Ayyagari et al. (2013) which is per capita English newspaper circulation in 1991. According to Besley and Burgess (2002), compared to other media, newspapers in India are relatively free and independent. Indian newspapers are published in 22 languages among which English language newspapers have greater national coverage and are more business and financial news

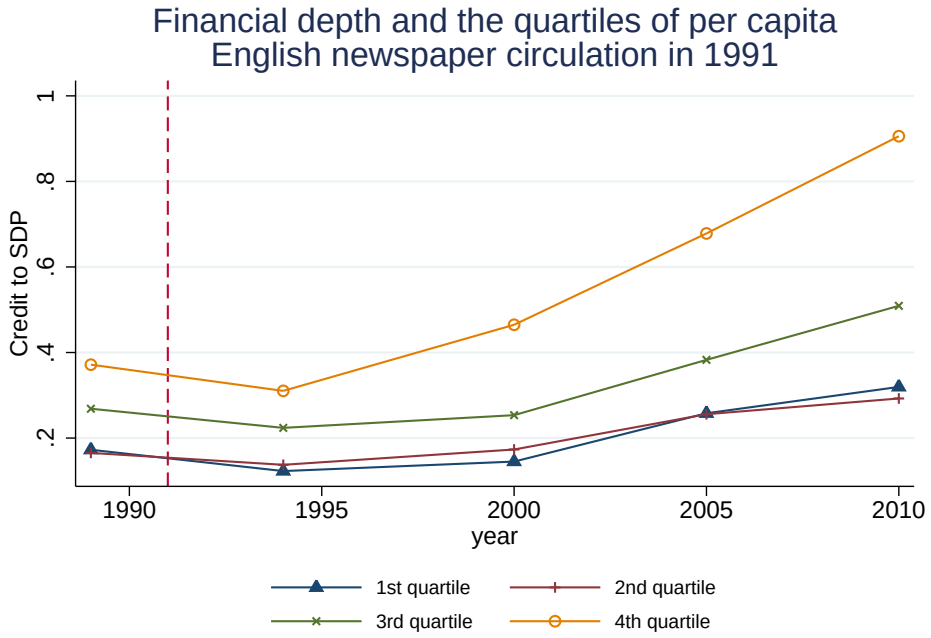


Figure 3.9: Media environment at the time of liberalization and financial depth. The lines show the weighted average of credit to SDP in the states in the respective quartile of per capita English newspaper circulation.

oriented.[41] Figure 3.9 illustrates the trend of Credit to SDP among different quartiles of English newspaper circulation in 1991. The variation in the initial media environment is strongly associated with the degree of financial deepening following to financial liberalization and in the two top quartiles the slope is increasing over time. As a result, similar to Ayyagari et al. (2013), we multiply the cross-state variation of per-capita circulation of English newspapers in 1991 with a post-1991 time trend to capture the differential impact of the media across time after liberalization in 1991 and use it as an instrument for financial depth.

Financial Inclusion Reform in 2006

Providing access to finance for the poor has been a longstanding objective of the Indian government. In 2006, the government of India introduced an ambitious reform aiming at providing formal finance for every Indian citizen based on suggestions made by the

[41] Ayyagari et al. (2013) find a high correlation (88%) between circulation of English language newspaper and business-oriented newspapers among Indian states in 2005-09. However, state-level data of circulation of business-oriented newspapers is not available for 1991 and thus they use English language newspaper circulation.

Committee on Financial Inclusion (Rangajaran Committee),[42] which recommended to introduce correspondent banking system. The goal of correspondent banking was to enable banks to use a third party agent to provide banking services on their behalf. This meant banks were no longer required to open costly branches in less populated locations and new branches could be established where the demand for service is enough. As highlighted in the Recommendations of the Committee on Financial Inclusion, "Adoption of appropriate technology would enable the branches to go where the customer is present instead of the other way round." Figure 3.3b shows that following to this reform branches per capita sharply increased in the whole country between 2005 and 2010. In order to use this trend break as an instrument, we need to find the cross-state heterogeneity in the effectiveness of the reform. Since the emphasis on demand side factors was a main pillar of reform in 2006, we expect more branch expansion in state that had higher initial demand for credit. To measure the demand for credit at the time of the reform, we use credit to SDP in 2005. Figure 3.10 graphs the growth rate of branches per capita in 2000-05 and 2005-10 across states versus initial demand for credit before 2006 reform. The figure suggests that the rise in branch penetration in 2010 is higher in states that had higher credit to SDP in 2005. This positive association, however, is much weaker for 2000-2005, where the average growth rate is negative.[43] Hence, we exploit this trend break by interacting credit to SDP in 2005 with the year dummy of 2010 – the only available year after the reform – as our instrument for branch penetration.

3.4. Empirical Results

This section reports the regression results. We will first discuss OLS regressions capturing average effects across industries, before presenting difference-in-difference regressions that gauge the differential effect of financial development on informality on industries with different needs for external finance. We then report instrumental variable regressions before turning to the effect of financial development on overall production.

[42] To see the details of recommendations visit `http://pib.nic.in/newsite/erelease.aspx?relid=35141`

[43] We have the same results, if we use credit to SDP in older years to capture the demand of financial services.

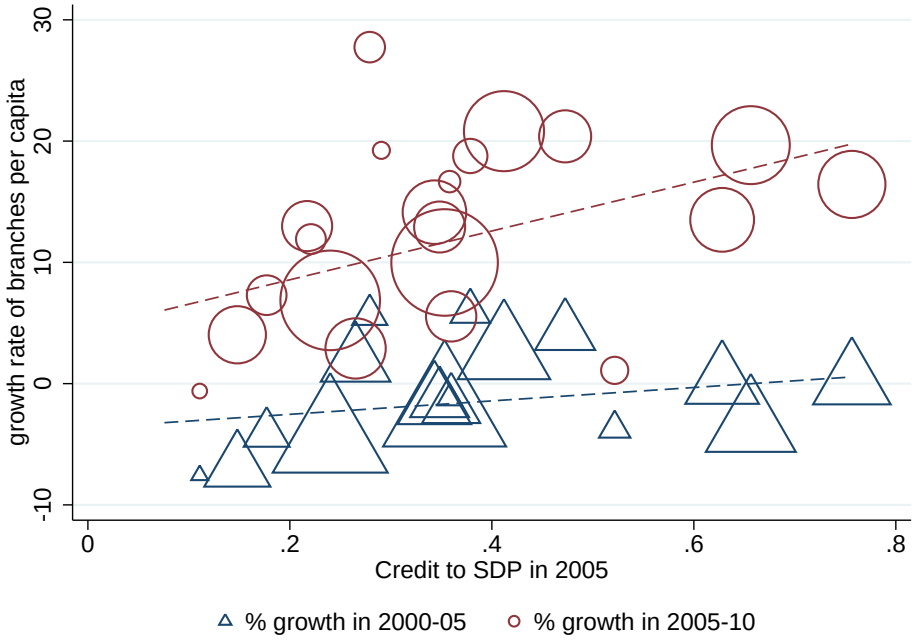


Figure 3.10: Branches expansion before and after the financial inclusion reform and demand for credit in 2005. Each observation represents one state. The growth rate is defined as $(y_{t+1} - y_t)/y_t$ in percentage term. The size of the observations is proportional to the number of enterprises in the state in 2005. For better illustration states with less than 50,000 firms (estimated using sample multiplier) are not shown in the graph.

3.4.1. OLS regressions

The results in Table 3.4 show the overall effect of financial development on the share of firms registered under any act or tax authorities. While we find a significant relationship between branches per capita and the share of formal enterprises registered under tax authorities, Credit to SDP does not enter significantly in the two regressions. The economic effect of the relationship between branch penetration and tax registration is also significant. Specifically, the standard deviation of branches per capita de-trended for state and year effects is 0.045 and this variation explains 0.045×8.411=0.38 percentage point in tax registration accounting for 15% of the standard deviation of the tax registration de-trended for state, year and industry which is 2.57.

In Appendix Table 3.A.3, we estimate the same equation, but for sub-samples of firms with smaller/bigger fixed assets to capture the effect of financial development on firms that are more/less likely to be excluded from the formal sector. Specifically, we select firms whose total fixed assets are below and above the 25th percentile of their industries in

each year, and re-compute the informality measures and the sample weights.[44] Columns (1) to (4) show that for the sample of smaller firms the effect of financial outreach is significant for both general and tax registration. Moreover, financial depth is positively associated with tax registration but not significantly with general registration for smaller firms. In terms of the economic size of the relationship, financial outreach has more explanatory power for the incidence of informality than financial deepening. For instance, for the sample of firms below 25th percentile, one de-trended standard deviation increase in branches per capita and credit to SDP increases tax registration rate by 0.045×14.58 = 0.656 and 0.047×4.21 = 0.198 percentage point, respectively. In column (5) to (8) of Table 3.A.3, however, we do not find a significant impact of financial outreach on general registration rate of bigger firms with fixed assets more that 25th percentile and instead credit to SDP becomes significant at 10% level.

Table 3.4: Financial depth vs. breadth and barriers to formality – state, year, and industry fixed effects are included in all regressions. Number of firms is applied as weight in all regressions and standard errors are clustered at state level.

	proportion of registered under any act		**proportion of registered under tax**	
	(1)	(2)	(3)	(4)
Branches per capita	25.131		8.411***	
	(15.552)		(2.750)	
Credit/SDP		13.518		1.298
		(8.263)		(2.458)
Government exp./SDP	20.588	18.775	0.456	-0.362
	(19.851)	(18.787)	(4.343)	(4.166)
Unit cost of energy	-36.525	-21.465	2.391	0.761
	(30.673)	(32.946)	(7.793)	(7.895)
Forward linkages	1.205	1.218	1.365***	1.355***
	(1.111)	(1.086)	(0.445)	(0.437)
Enforcement exp./No. firms	1.041	0.631	0.119	-0.084
	(1.419)	(1.292)	(0.417)	(0.440)
FL × Enfor. exp./No. firms	-1.024*	-1.093**	0.383	0.370
	(0.522)	(0.474)	(0.270)	(0.278)
Constant	6.040	22.688*	-6.308	-0.012
	(18.497)	(11.070)	(4.371)	(2.770)
Observations	3160	3160	2778	2778
R-squared	0.573	0.573	0.487	0.485

Overall, these estimations suggest that financial development reduces exclusion from the formal sector by reducing entry barriers to the formal sector, a relationship stronger for smaller firms; we also find that broadening access through increasing geographic

[44]The results are robust to using the percentiles just within industry or irrespective of the industry. They are stronger for smaller percentiles.

proximity between borrowers and lenders plays a more important role to increase registration rate than financial deepening, i.e. a higher credit volume.[45] These results, however, are based on average estimations across industries with different needs for external finance. The estimates are also subject to endogeneity biases, related to reverse causation (a higher share of formal firms demanding more formal finance and thus increasing both credit volume and outreach by financial institutions) and omitted variables that might drive both reduction in informality and financial deepening and broadening. In the following, we will therefore explore the differential relationship between financial development and the incidence of informality across industries with different needs for external finance.

3.4.2. Difference-in-difference regressions

The results in Table 3.5 show that the positive association of financial development and the share of firms registered under any act is stronger in industries that rely more on external finance. In columns (1) and (2), we interact the two financial development variables on the state-year level with external dependence on the industry level, including state-year and industry fixed effects. Both interaction terms enter positively and significantly at 1% level. To control for the fact that financial development and formality are correlated with public good provision and tax enforcement intensity, we also include an interaction term of external dependence with the government expenditures to SDP and unit cost of power, and an interaction term of forward linkages with the enforcement expenditure per firm. While the interaction term of public good provision enters positively and significantly in the regression including the interaction between Credit to SDP and external dependence, it enters insignificantly in the regression including the interaction of external dependence and branches per capita. The interaction term of unit cost of power is insignificant and the tax enforcement intensity appears with negative sign in all regressions.[46]

While column (1) and (2) of Table 3.5 considers only the share of firms operating in the formal sector, we now turn to alternative indicators of informality as dependent

[45] Given the high correlation between the two financial sector variables, we only include one of them at a time. If we include both at the same time, we find that just outreach is significant in Table 3.4. We also ran regressions with SDP per capita confirm our findings. Given its high correlation with financial development, we do not include it in the main regressions to avoid multi-colinearity.

[46] Tax enforcement intensity becomes positive if we use the respective shares of tax registered firms as dependent variables.

variables. In columns (3) and (4), we use value-added and employment share of formal firms in total as dependent variables. Specifically, we present results with (i) share of value added produced by firms registered under any act (columns 3 to 4) and (ii) the share of employment in firms registered under any act (columns 5 to 6), with the interaction of financial development and external dependence as the main explanatory variable of interest.

The results show that while financial depth is positively and significantly associated with the share of value added produced by formal firms and the employment share of formal firms, there is no significant impact of branch penetration on the share of value added and a less robust impact on the employment in formally registered firms. Specifically, the interaction term between Credit to SDP and external dependence enters positively and significantly at the one percent level for both of these dependent variables, while the interaction of branch penetration and external dependence does not enter significantly in column (3) and is significant at 10% level only in column (5). These findings suggest that while financial outreach is negatively associated with the incidence of informality by pushing informal firms into the formal sector, it does not significantly affect informality at the intensive margin i.e. total value-added produced by the informal sector. In comparison, financial depth reduces informality at both margins.

The results of Table 3.5 are not only statistically, but also economically significant. The difference-in-differences estimation suggest that going from a state at the 25th percentile of branches per capita (Jharkhand = 0.55) to a state at the 75th percentile (Kerala = 1.14) and an industry at the 25th percentile of the RZ external dependence index (basic metals = 0.03) to an industry at the 75th percentile of external dependence (motor vehicles = 0.39) results in an increase in registration under any act by $25.937 \times 0.59 \times 0.36 = 5.51$ percentage points. The 25th and 75th percentiles of credit to SDP are Uttar Pradesh (0.19) and Andhra Pradesh (0.36); the differential effects for Credit to SDP are therefore 2.66 percentage points. This compares to a mean registration rate of 11.3 percent. As in Table 3.4, the economic effect is thus larger for financial outreach than for financial depth on the extensive margin. Note, however, that in the case of value added share of formal enterprises, only Credit to SDP is statistically significant.

Table 3.5: The effect of financial dependence on registration, formal value-added and formal employment. – Difference-in-differences estimation: State × year, and industry fixed effects are included in all regressions. Number of firms is applied as weight in all regressions and standard errors are clustered at state level. Interquartile impact is computed as a differential effect between 75th and 25th percentiles of RZ and financial variables: $B \times (RZ_{75} - RZ_{25}) \times (FD_{75} - FD_{25})$.

	proportion of registered under any act		**VA share of registered under any act**		**employment share of registered under any act**	
	(1)	(2)	(3)	(4)	(5)	(6)
RZ × branches per capita	25.937***		23.388		20.652*	
	(6.487)		(16.320)		(11.371)	
RZ × Credit/SDP		43.410***		29.252**		49.873***
		(8.007)		(12.384)		(11.651)
RZ × Government exp./SDP	11.180	70.424**	63.533	111.533	53.804	109.366**
	(31.847)	(29.900)	(65.025)	(73.467)	(41.035)	(39.620)
RZ × unit cost of power	22.271	24.422	-80.249	-107.530	11.633	58.568
	(74.008)	(75.418)	(221.551)	(218.195)	(140.653)	(136.694)
FL × Enforcement exp./No. firms	-1.438*	-1.950**	-5.355*	-5.653**	-2.662	-3.325*
	(0.710)	(0.825)	(2.654)	(2.682)	(1.640)	(1.850)
Constant	30.132***	42.147***	48.592***	44.061***	23.359***	16.653**
	(6.430)	(6.318)	(11.832)	(12.615)	(7.565)	(6.505)
Observations	3160	3160	3153	3153	3151	3151
R-squared	0.616	0.629	0.702	0.703	0.642	0.651
Interquartile impact	5.509	2.657	4.968	1.790	4.387	3.052

The results reported Appendix Table 3.A.4 show that the effect of financial outreach is stronger for smaller firms, while we only find an effect of financial depth for larger firms. Here, we split the sample into firms below and above the 25th percentile of fixed assets for a specific industry and year. The relative economic size of the effects of financial depth and outreach is confirmed. In the case of firms above the 25th percentile, only the interaction of external finance with our measure of financial depth, Credit to SDP, enters positively and significantly in all regressions. The economic size of the impact also suggests a stronger impact of credit to SDP on firms above – compared to below – the 25th percentile. This suggests larger firms in industries relying on external finance do not benefit from higher branch penetration as much as small firms, but rather from overall financial depth, as captured by credit volume on the state level.[47] This is consistent with smaller firms depending more on geographic proximity to lenders for the cost-benefit trade-off to tip towards formality than larger firms (Beck et al., 2014).

One concern regarding the impact of financial development on informality is the reverse causation in the sense that lower informality leads to higher demand for financial services, especially in industries with higher need for external finance. To control for this effect, in Appendix Table 3.A.5, we re-estimate Table 3.5 for the sample of industries that are below the median of production level in the respective state and year. The results suggest that even if we exclude the larger industries in each state that can create such a demand effect, the interaction of RZ with both financial penetration and financial deepening are positively associated with registration rates.

3.4.3. IV regressions

To further control for the endogeneity problem, we estimate the impact of financial outreach and depth on different margins of informality using instrumental variable approach. Table 3.6 and 3.7 present both first-stage and second-stage estimation results for the regression equations (1) and (2), respectively.

In column (1) of Table 3.6, we regress the braches per capita on credit to SDP in 2005 (to capture the state-level demand for financial services), and a time dummy for 2010. Consistent with Figure 3.10, we observe a significant impact of financial inclusion reform

[47] If we include both financial outreach and depth in a single regression in Table 3.5, for the sample of small firms (fixed asset $<$ 25th percentile), branches per capita is positive and significant while credit to SDP becomes insignificant. In contrast, when the sample contains large firms (fixed asset $>$ 25th percentile), credit to SDP is significant, while branches per capita is not.

on branch establishment in state with higher financial depth. The first-stage F-statistics indicates that our instrument for branches per capita is relevant. Column (2) presents the first stage estimation for the interaction term of financial depth instrumented by per capita circulation of English newspaper in 1991 (to capture business and economic news penetration at the time of liberalization), and a time trend for the post-liberalization years. Similar to column (1), the instrument is positively and significantly associated with the endogenous variable and F-statistics confirms its relevance.

The rest of Table 3.6, reports the second stage regressions examining the impact of financial inclusion and deepening on extensive and intensive margins of informality. In comparison to Table 3.4, branches per capita appears with a positive and significant coefficient when the dependent variable is the proportion or employment share of firms registered under any act. However, similar to Table 3.4, we do not observe any significant impact of credit to SDP.

Table 3.7 presents the IV estimation results for the differential effect regressions. In column (1), we regress the interaction term of branches per capita on the interaction of RZ, credit to SDP in, and a dummy for 2010. In column (2), we instrument financial depth by the interaction between RZ, per capita circulation of English newspaper in 1991, and a trend dummy for the post-liberalization years. Similar to Table 3.6, the instruments are positively and significantly associated with the endogenous variables with high relevance. The second stage results for the impact of financial outreach and depth on extensive and intensive margins of informality are consistent with the OLS estimations. Therefore, all results of Table 3.5 are confirmed by using an IV approach.

Table 3.6: Instrumental Variable estimation: main effects. – Difference-in-differences estimation: State, year, and industry fixed effects are included in all regressions. Number of firms is applied as weight in all regressions and standard errors are clustered at state level.

	Branches per capita	**Credit/SDP**	**proportion of registered under any act**		**VA share of registered under any act**		**employment share of registered under any act**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Credit to SDP in 2005 × (year>2006)	0.216***							
	(0.061)							
Percapita circulation of english newspapers in 1991 × (year>1991) × (year-1991)		0.895***						
		(0.117)						
Branches per capita			109.211***		20.625		106.293**	
			(40.450)		(70.135)		(53.642)	
Credit/SDP				13.654		-6.610		9.187
				(9.439)		(18.392)		(14.076)
Government exp./SDP	-0.275	0.272	47.802**	19.782	-25.290	-34.940	30.495	1.319
	(0.174)	(0.343)	(23.859)	(17.239)	(39.663)	(36.309)	(26.062)	(20.294)
Enforcement exp./No. firms	-0.006	-0.033	2.840	-0.144	1.545	0.782	3.386	0.391
	(0.010)	(0.020)	(2.461)	(1.129)	(4.630)	(2.891)	(3.415)	(1.606)
Unit cost of energy	-0.015	-0.956	-18.814	-20.758	-78.043	-92.210	-39.178	-47.233
	(0.396)	(0.572)	(58.572)	(31.672)	(90.433)	(87.456)	(58.256)	(37.796)
Constant	0.858***	-0.050	-70.066*	21.990**	56.100	90.728**	-115.056	59.568***
	(0.073)	(0.359)	(36.010)	(10.049)	(101.347)	(40.946)	(75.310)	(12.878)
Observations	3160	3098	3160	3098	3153	3092	3151	3090
R-squared	0.974	0.959	0.533	0.572	0.659	0.662	0.595	0.613
F-test	12.335	58.779						

Table 3.7: Instrumental Variable estimation: differential effects. – State × year, and industry fixed effects are included in all regressions. Number of firms is applied as weight in all regressions and standard errors are clustered at state level. Interquartile impact is the computed as a differential effect between 75th and 25th percentiles of RZ and financial variables: $B \times (RZ_{75} - RZ_{25}) \times (FD_{75} - FD_{25})$.

	RZ × branches per capita	**RZ × Credit/SDP**	**proportion of registered under any act**		**VA share of registered under any act**		**employment share of registered under any act**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RZ × credit to SDP in 2005 × (year>2006)	0.248***							
	(0.048)							
RZ × Percapita circulation of english newspapers in 1991 × (year>1991) × (year-1991)		1.337***						
		(0.208)						
RZ × branches per capita			110.013***		39.841		131.210***	
			(36.635)		(62.024)		(45.482)	
RZ × Credit/SDP				47.421***		22.655***		54.866***
				(6.264)		(8.650)		(14.785)
RZ × Government exp./SDP	1.521**	0.345	-103.742	70.758**	41.042	108.519	-97.321	110.725***
	(0.616)	(0.281)	(71.950)	(29.122)	(129.683)	(72.159)	(91.820)	(38.479)
RZ × unit cost of power	-4.577***	-1.949***	430.982**	38.086	-0.264	-127.418	549.081**	74.674
	(1.043)	(0.361)	(204.047)	(72.949)	(410.727)	(207.880)	(261.303)	(140.657)
FL × Enforcement exp./No. firms	0.006	-0.003	-2.120	-2.016**	-5.489*	-5.550**	-3.558	-3.411*
	(0.006)	(0.004)	(1.319)	(0.801)	(2.933)	(2.526)	(2.422)	(1.811)
Observations	3160	3098	3160	3098	3153	3092	3151	3090
R-squared	0.323	0.728	-0.149	0.056	0.020	0.025	-0.127	0.037
Interquartile impact			23.367	2.902	8.462	1.387	27.869	3.358
F-test	26.379	41.167						

3.4.4. Production and value-added

So far, we have focused on the relative importance of formal and informal sectors within manufacturing. We now turn our attention to production and value added in the formal and informal sectors as an additional test of the second channel outlined above, i.e. the higher productivity of firms in the formal sector. We therefore use as dependent variable total production or total value added on the state-industry level for registered firms and unregistered firms. Specifically, Table 3.8 illustrates the result of estimation of equation (2) for log of production and value-added.

The results in column (1) to (4) of Table 3.8 show that production in registered firms increases with Credit to SDP in industries that depend more on external finance, while total production of unregistered firms is negatively associated with the interaction of external dependence and Credit to SDP, suggesting a positive (negative) and significant impact of financial deepening on total formal (informal) production. On the other hand, the interaction term of branches per capita does not enter significantly in any of the specifications. The results in column (5) to (8) show the same pattern for total value-added: the value-added of registered firms increases across industries with a higher need for external finance as financial systems deepen, while value-added of informal firms decreases with the interaction of Credit to SDP and external dependence. Comparing the 75th and 25th percentiles, the economic effect of credit to SDP interacted with RZ on the value-added of firms registered is $1.454 \times 0.17 \times 0.36 = 0.089$, accounting for 16% of de-trended standard deviation (0.55). As in total production, the interaction terms of branch penetration and external dependence do not enter significantly when the dependent variable is total value-added. In Appendix Table 3.A.6, we test the robustness of the finding of Table 3.8 in IV specification and observe the same results for the impact of financial depth on formal and informal production. In column (1) of Table 3.A.6, we however find a positive impact of the interaction term of branches per capita on total production of formal firms.

Table 3.8: Financial depth vs. breadth and Productivity. Difference-in-differences estimation: State × year, and industry fixed effects are included in all regressions. Number of firms is applied as weight in all regressions and standard errors are clustered at state level. Interquartile impact is the computed as a differential effect between 75th and 25th percentiles of RZ and financial variables: $B \times (RZ_{75} - RZ_{25}) \times (FD_{75} - FD_{25})$.

	production of registered		**production of not registered**		**Value-added of registered**		**Value-added of not registered**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RZ × branches per capita	1.025		-0.505		0.909		-0.604	
	(0.871)		(0.742)		(0.970)		(0.876)	
RZ × Credit/SDP		1.679***		-0.977***		1.454**		-1.198***
		(0.462)		(0.324)		(0.617)		(0.396)
RZ × Government exp./SDP	2.811	5.132	0.578	-0.647	4.713	6.753	0.864	-0.618
	(4.144)	(3.578)	(3.376)	(2.977)	(4.590)	(4.205)	(3.411)	(3.104)
RZ × unit cost of power	-0.633	-0.655	0.582	0.151	-6.742	-6.851	-0.032	-0.634
	(8.407)	(7.720)	(6.892)	(6.379)	(9.569)	(9.139)	(7.153)	(6.177)
FL × Enforcement exp./No. firms	-0.164	-0.183*	0.165	0.178	-0.075	-0.092	0.150	0.165
	(0.103)	(0.105)	(0.111)	(0.112)	(0.105)	(0.108)	(0.100)	(0.102)
Constant	11.671***	12.146***	9.852***	9.619***	6.314***	5.857***	8.036***	8.151***
	(1.054)	(0.861)	(0.839)	(0.730)	(1.163)	(1.103)	(0.367)	(0.398)
Observations	3160	3160	3160	3160	3093	3093	3153	3153
R-squared	0.824	0.825	0.756	0.757	0.784	0.785	0.729	0.731
Interquartile impact	0.218	0.103	-0.107	-0.060	0.193	0.089	-0.128	-0.073

3.5. Conclusion

This paper explores the relationship between financial sector development and the relative importance of formal and informal manufacturing in India. Previous work and theory suggest an impact of financial development on both extensive and intensive margins, i.e. pulling more firms into the formal sector and increasing total production of the formal sector. Our results provide evidence for both channels, but also distinct roles for financial depth, as proxied by Credit to SDP, and financial outreach, as proxied by branch penetration. Specifically, exploiting variation within state-years and industries with different needs for external finance, we find that financial outreach is positively associated with a higher share of formal enterprises, especially in industries with a higher demand for external finance, i.e. where firms benefit more from access to formal finance. While we also find a positive effect of financial depth on the share of formal firms, although this effect is of a smaller size. In terms of production efficiency, on the other hand, we find a positive and significant role for financial depth, especially in industries more reliant on external finance, while no significant effect for branch penetration.

Our empirical findings suggest an important impact of financial sector development on the incidence of formality. This impact works through different channels, with different dimensions of financial sector development dominating specific channels. Specifically, branch penetration is associated with a lower incidence of informality mainly through the extensive margin by helping or persuading informal firms to enter the formal sector. Overall credit volume, on the other hand, increases the productivity of formal sector and reduces informality mainly through the intensive margin, i.e. expanding the formal economy at the expense of the informal economy.

Together, these results suggest an important role for finance in reducing informality, though with important differences across industries. They also suggests that policies aimed at deepening the financial system as much as policies aimed at increasing outreach are important for increasing the share and productivity of formal enterprises in manufacturing.

Appendix

3.A. Additional Tables

Table 3.A.1: summary statistic of each industry – The numbers are estimated using sample multiplier

Description	# of firms	Reg	Treg	Vreg	Vtreg	Ereg	Etreg	RZ	FL
Tobacco	10320280	3.13	0.48	66.63	58.02	16.25	11.62	-0.45	0.274
Pottery	2890694	5.7	2.63	67.14	48.47	31.65	18.54	-0.15	0.772
Leather	10679940	7.89	0.52	51.99	38.21	22.41	11.08	-0.14	0.416
Spinning	6371518	12.17	2.53	86.78	66.44	41.79	23.58	-0.09	0.985
Footwear	738193	8.37	2.35	69.86	60.39	36.11	25.86	-0.08	0.691
Nonferrous metal	107628	33.26	11.63	98.79	88.16	77.13	56.75	0.01	2.575
Apparel	1479351	2.11	0.54	54.83	43.25	9.29	4.93	0.03	1.327
Petroleum refineries	1742	80.8	67.43	99.99	99.99	98.29	96.88	0.04	1.789
Nonmetal products	796204	20.87	10.9	96.61	87.31	62.67	41.11	0.06	0.942
Beverages	1041632	11.87	2.33	91.53	82.96	29.73	17.46	0.08	0.343
Iron and steel	42914	53.45	32.62	98.83	91.04	92.44	75.47	0.09	1.276
Food products	12094625	20.79	2	78.47	60.01	37.38	17.63	0.14	0.544
Pulp	31303	26.31	20.49	99.61	84.82	92.2	73.19	0.15	1.481
Paper	914277	10.37	4.9	94.2	85.67	50.64	36.24	0.18	1.734
Printing and Publishing	619644	50.68	8.98	90.73	68.27	69.8	29.12	0.2	1.015
Other chemicals	646187	8.17	3.37	98.48	88.9	47.22	32.77	0.22	0.622
Rubber products	115613	45.28	21.28	98.64	82.58	85.81	64.95	0.23	2.131
Furniture	2492328	10.17	1.55	46.84	27.89	20.29	5.28	0.24	0.593
Metal products	2756281	23.18	4.54	81.33	62.52	46.92	21.58	0.24	1.205
Basic chemicals	43231	78.55	62.29	99.78	95.35	95.61	80.43	0.25	1.562
Wood products	11825137	3.37	0.72	31.02	16.37	7.84	2.35	0.28	1.515
Transportation equipments	73410	42.53	11.24	95.64	78.94	78.24	42.41	0.31	1.337
Petroleum and coal products	21328	42.13	17.58	98.49	89.23	84.01	59.13	0.33	1.7
Motor vehicles	100638	61.59	26.35	99.46	91.26	95.2	77	0.39	0.507
Textile	4973730	4.14	0.84	58.4	47.66	16.67	9.18	0.4	0.52
Machinery	1099944	23.96	6.96	95.87	82.91	65.48	42.16	0.45	0.718
Ship	116768	44.86	18.46	97.01	83.74	82.95	57.93	0.46	1.038
Glass	177449	11.7	5.2	92.57	80.61	47.62	32.36	0.53	0.819
Electric Machinery	316788	39.34	13.38	97.94	84.58	82.05	58.49	0.77	0.817
Professional goods	1154791	5.48	1.6	88.02	75.24	19.61	12.35	0.96	0.786
Radio	40620	50.72	25.61	99.03	86.97	91.43	69.72	1.04	0.563
Office and computing	4540	70.1	39.44	99.85	87.8	97.86	74.72	1.06	
Plastic products	488515	31.55	11.54	93.58	82.92	66.46	44.43	1.14	1.698
Drugs	45765	60.7	42.97	99.84	95.35	93.11	83.2	1.49	1.26

Table 3.A.2: summary statistic for each state – The numbers are estimated using sample multiplier

State name	# of firm	Reg	Treg	Vreg	Vtreg	Ereg	Etreg	Credit to SDP	Branches per capita	SDP per capita	Gov. exp. / GDP	Enf. exp. per firm
Jammu & Kashmir	845350	13.56	0.93	68.8	50.92	26.63	8.47	0.37	0.91	13003	0.567	0.117
Himachal Pradesh	525488	19.13	2.39	95.59	91.28	45.34	26.64	0.25	1.4	20315	0.382	0.246
Punjab	1415414	16.83	4.82	88.65	75.02	56.2	37.91	0.35	1.17	34837	0.185	0.184
Chandigarh	19633	36.84	16.08	93.57	64.87	77.92	44.68	1.42	2.67	88845		
Uttaranchal	448464	19.59	2.18	96.65	91.72	54.59	36.07	0.35	1.16	32902	0.285	0.323
Haryana	853424	13.58	4.85	93.18	83.85	60.33	46.36	0.27	0.83	35870	0.168	0.195
Delhi	732034	17.81	4.93	61.75	44.15	32.49	16.71	1.31	1.28	58386	0.071	5.571
Rajasthan	2838819	7.13	1.9	82.07	71.59	29.63	18.72	0.29	0.65	14456	0.223	0.223
Uttar Pradesh	10916734	6.6	1.05	77.19	64.7	22.13	12.24	0.19	0.53	13446	0.185	0.15
Bihar	3518443	5.53	0.32	43.25	33.62	12.45	4.51	0.12	0.42	9146	0.178	0.133
Sikkim	8913	23.15	2.65	99.07	97.35	60.68	39.4	0.21	0.99	27620	1.084	1.649
Arunachal Pradesh	7687	28.33	1.64	65.23	13.37	65.99	12.22	0.12	0.7	17110	0.54	0.724
Nagaland	35920	14.12	1.62	65.4	38.93	38.59	17.56	0.11	0.43	17459	0.574	0.95
Manipur	238840	1.62	0.3	20.22	11.43	6.63	2.92	0.11	0.34	11653	0.509	0.116
Mizoram	22520	22.14	0.36	51.07	1.2	33.75	0.58	0.16	0.9	17891	0.787	1.328
Tripura	239238	7.75	1.05	57.36	41.08	22.07	11.96	0.16	0.62	15440	0.474	0.115
Meghalaya	119532	8.87	0.7	63.4	51.69	18.56	5.79	0.28	0.79	14808	0.392	0.297
Assam	1275659	8.8	0.93	76.99	64.44	29.05	17.59	0.16	0.49	12888	0.241	0.158
West Bengal	12751838	7.99	1.21	70.36	51.39	22.67	9.78	0.32	0.6	18553	0.168	0.055
Jharkhand	1947689	2.35	0.53	75.64	59.81	16.46	10.97	0.25	0.55	17260	0.278	0.121
Orissa	4818666	1.32	0.3	73.21	61.56	6.66	4.19	0.23	0.65	10926	0.227	0.073
Chhattisgarh	915889	8.53	1.43	86.54	71.57	26.66	13.36	0.26	0.52	23668	0.215	0.311
Madhya Pradesh	3298643	9.02	1.31	87.23	75.66	26.11	14.82	0.23	0.58	18187	0.175	0.396
Gujrat	3629548	25.73	4.95	92.49	79.86	55.32	30.09	0.33	0.82	36281	0.169	0.096
Daman & Diu	12925	62.31	47.6	99.85	99.08	97.22	92.89					
Dadra & Nagar Haveli	11053	48.49	41.72	99.77	98.14	96.03	92.2					
Maharastra	5319557	26.83	4.49	94.3	79.98	58.44	30.74	0.68	0.8	36704	0.144	0.52
Andhra Pardesh	7384581	9.23	1.97	83.11	72.63	33.04	22.45	0.36	0.76	22615	0.18	0.094
Karnataka	4384418	13.39	2	89	76.04	39.51	22.74	0.49	0.98	20723	0.196	0.11
Goa	65348	42.6	18.62	98.9	93.8	75.7	55.77	0.29	3.28	60389	0.239	0.283
Lakshadweep	1251	19.36	1.02	52.1	8.12	38.87	6.31					
Kerala	2489642	21.97	3.48	85.92	60.78	51.67	24.72	0.42	1.15	22036	0.22	0.184
Tamil Nadu	7019458	13.52	2.65	88.73	74.52	45.51	27.35	0.55	0.85	24962	0.189	0.089
Pondicheri	53972	21.33	6.66	97.74	92.78	71.23	51.64	0.28	1.16	44962	0.358	0.487
Andaman & Nicober	10348	21.31	1.75	77.67	42.74	57.1	21.88	0.23	1.01	27706		

Table 3.A.3: Financial depth vs. breadth for small and big firms. – State, year, and industry fixed effects are included in all regressions. Number of firms is applied as weight in all regressions and standard errors are clustered at state level.

	Fixed assets < 25th percentile				**Fixed assets > 25th percentile**			
	proportion of registered under any act		**proportion of registered under tax**		**proportion of registered under any act**		**proportion of registered under tax**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Branches per capita	22.432***		14.583**		22.485		8.577**	
	(5.979)		(5.395)		(17.604)		(3.688)	
Credit/SDP		9.895		4.213**		16.437*		1.652
		(6.229)		(1.819)		(8.689)		(3.077)
Government exp./SDP	-2.224	-3.863	6.361	4.171	21.864	20.366	-6.130	-7.088
	(12.823)	(14.918)	(9.945)	(9.995)	(21.625)	(19.094)	(5.507)	(4.921)
Unit cost of energy	-53.975	-44.103	3.983	5.805	-13.271	10.260	5.797	4.157
	(36.324)	(33.769)	(5.393)	(5.058)	(36.948)	(40.639)	(11.712)	(12.184)
Forward linkages	0.850	0.935	0.460	0.490*	0.832	0.823	1.250*	1.233*
	(0.636)	(0.662)	(0.277)	(0.266)	(1.902)	(1.891)	(0.718)	(0.714)
Enforc. exp./No. firms	2.033**	1.814***	-0.227	-0.336	-0.082	-0.390	0.386	0.203
	(0.739)	(0.648)	(0.240)	(0.298)	(1.655)	(1.501)	(0.622)	(0.631)
FL × Enf. exp./No. firms	-0.220	-0.201	0.450***	0.452***	-0.624	-0.716	0.325	0.299
	(0.297)	(0.302)	(0.146)	(0.151)	(0.891)	(0.851)	(0.501)	(0.504)
Constant	-21.495	3.270	-11.643	-1.197	11.010	24.843**	-1.145	8.851
	(16.613)	(16.837)	(6.929)	(5.984)	(21.363)	(11.757)	(9.313)	(5.780)
Observations	2863	2863	2497	2497	3094	3094	2714	2714
R-squared	0.333	0.331	0.229	0.225	0.633	0.635	0.501	0.499

Table 3.A.4: IV estimation: Financial depth vs. breadth and Productivity – Difference-in-differences estimation: State × year, and industry fixed effects are included in all regressions. Number of firms is applied as weight in all regressions and standard errors are clustered at state level. Interquartile impact is the computed as a differential effect between 75th and 25th percentiles of RZ and financial variables: $B \times (RZ_{75} - RZ_{25}) \times (FD_{75} - FD_{25})$.

	Fixed asset < 25th percentile						**Fixed asset > 25th percentile**					
	proportion of registered under any act		**VA share of registered under any act**		**employment share of registered under any act**		**proportion of registered under any act**		**VA share of registered under any act**		**employment share of registered under any act**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
RZ × branches per capita	21.187**		44.162***		24.819**		17.651**		13.204		15.933	
	(8.446)		(10.840)		(11.193)		(6.639)		(21.376)		(12.377)	
RZ × Log(Credit/SDP)		24.157**		28.168*		33.469***		40.534***		27.476**		52.055***
		(9.019)		(13.887)		(11.107)		(10.931)		(12.975)		(13.286)
RZ × Gov. exp. /SDP	2.717	47.373	-44.44	35.796	0.073	55.384	24.47	68.410**	95.838	127.334	68.149	115.291**
	(28.470)	(39.503)	(70.724)	(74.320)	(35.297)	(48.020)	(34.679)	(31.483)	(79.342)	(92.287)	(45.998)	(42.489)
RZ × unit cost of power	6.349	-34.174	62.015	-75.403	-25.626	-60.757	-29.498	15.434	-108.32	-84.071	-16.443	74.994
	(32.790)	(29.067)	(112.694)	(125.604)	(53.372)	(41.143)	(99.008)	(103.445)	(291.197)	(286.287)	(192.927)	(187.642)
FL × Enf. exp./No. firms	-0.282	-0.383	-4.921***	-4.877***	-1.563*	-1.742**	-0.94	-1.369	-1.831	-2.119	-1.405	-1.974
	(0.427)	(0.448)	(0.848)	(0.860)	(0.821)	(0.838)	(1.009)	(1.337)	(2.545)	(2.652)	(1.819)	(2.177)
Constant	-5.922	-6.105	-1.809	0.311	-6.407	-7.499	29.613***	25.928***	88.955***	86.831***	50.019***	42.451***
	(6.809)	(6.477)	(8.057)	(8.213)	(6.608)	(6.913)	(5.038)	(4.099)	(12.226)	(11.201)	(6.860)	(6.233)
Observations	2863	2863	2859	2859	2850	2850	3094	3094	3079	3079	3077	3077
R-squared	0.393	0.394	0.649	0.646	0.509	0.511	0.668	0.678	0.689	0.69	0.672	0.682
Interquartile impact	4.5	1.478	9.38	1.724	5.271	2.048	3.749	2.481	2.805	1.682	3.384	3.186

Table 3.A.5: The effect of financial dependence and forward linkages on registration in sample of small industries – The sample includes industries below median of total production in each state and year (the biggest half of industries in each state and year is excluded). State × year, and industry fixed effects are included in all regressions. Number of firms is applied as weight in all regressions and standard errors are clustered at state level.

	proportion of registered under any act		**proportion of registered under tax**	
	(1)	(2)	(3)	(4)
RZ × branches per capita	14.142**		6.765***	
	(6.039)		(2.431)	
RZ × Log(Credit/SDP)		11.544**		1.957***
		(4.220)		(0.695)
RZ × Government exp./SDP	3.362	43.763	-10.146	0.079
	(33.354)	(34.201)	(10.319)	(10.519)
FL × Enforcement exp./No. firms	2.667***	2.465***	0.417	0.274
	(0.752)	(0.704)	(0.338)	(0.336)
Constant	53.195***	50.535***	20.978*	20.730*
	(12.203)	(11.698)	(12.033)	(12.044)
Observations	955	955	843	843
R-squared	0.766	0.775	0.481	0.478

Table 3.A.6: IV estimation: Financial depth vs. breadth and Productivity. Difference-in-differences estimation: State × year, and industry fixed effects are included in all regressions. Number of firms is applied as weight in all regressions and standard errors are clustered at state level. Interquartile impact is the computed as a differential effect between 75th and 25th percentiles of RZ and financial variables: $B \times (RZ_{75} - RZ_{25}) \times (FD_{75} - FD_{25})$.

	production of registered		**production of not registered**		**Value-added of registered**		**Value-added of not registered**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RZ × branches per capita	5.337***		-1.627		4.319		-2.007	
	(1.793)		(1.497)		(2.931)		(1.439)	
RZ × Credit/SDP		1.752***		-1.034***		1.581***		-1.037***
		(0.474)		(0.374)		(0.530)		(0.355)
RZ × Government exp./SDP	-3.084	5.120	2.112	-0.584	0.061	6.818	2.783	-0.438
	(5.624)	(3.533)	(3.872)	(2.950)	(7.150)	(4.155)	(3.752)	(3.069)
RZ × unit cost of power	20.332	-0.391	-4.875	-0.111	9.811	-6.470	-6.854	-0.251
	(13.867)	(7.430)	(8.694)	(6.160)	(19.031)	(8.689)	(8.014)	(5.963)
FL × Enforcement exp./No. firms	-0.199*	-0.185*	0.174	0.179	-0.102	-0.095	0.161	0.162
	(0.119)	(0.100)	(0.117)	(0.110)	(0.115)	(0.105)	(0.107)	(0.101)
Observations	3160	3098	3160	3098	3093	3031	3153	3091
R-squared	-0.048	0.016	0.002	0.011	-0.024	0.015	-0.002	0.012
Interquartile effect	1.134	0.107	-0.346	-0.063	0.917	0.097	-0.426	-0.063
WeakID test	233.638	6168.312	233.638	6168.312	229.024	6026.338	233.122	6153.738

Chapter 4

Misreporting in the Value-added Tax and the Optimal Enforcement

4.1. Introduction

One of the most important concerns of tax administrations is choosing a simple and efficient enforcement procedure to reduce the risk of tax fraud and evasion. The value-added tax (VAT) is nowadays the preferred form of indirect tax and is believed to facilitate enforcement through its invoicing system. For tax administrators, each VAT invoice generates a piece of information on a transaction, verifiable in case of inter-firm trade with the corresponding invoice issued by the other party. This type of third-party reporting in the VAT increases the risk of hiding transactions for firms and thus reduces evasion (Pomeranz, 2013). However, this deterrence mechanism breaks down in transactions with final consumers who do not report the invoice to the government. Consequently, an approach to identify the VAT evasion potential among different taxpayers becomes critical for tax administrators. On the other hand, modern VAT systems involve extensive information reporting to achieve a high level of compliance, but the consequences of such transition are almost neglected in the tax evasion literature. This paper analyzes firm's behavior to evade the VAT and the optimal strategy of the tax administration in this respect by linking noncompliance with the level of final consumption and the subjective costs of each taxpayers. In addition, it studies the role information reporting, which enables the tax administration to cross-check the VAT invoices, on firm's VAT evasion and discusses its efficient implementation.

Broadly, VAT evasion can be classified into two forms. At the intensive margin, a registered seller gains from under-reporting the sale, while a formal buyer benefits from over-reporting its purchase. I use the generic term *misreporting* to refer to noncompliance by registered traders. In comparison, *informality* represents the extensive margin, i.e.

evasion by firms working in the shadow economy and failing to register with tax agencies. In the high-income countries, the loss in VAT revenue due to misreporting is much larger than failure to register. For instance, in 2001-02, non-compliance by registered traders in UK resulted in loss of £6.7-9.75 billion compared to £400-500 million loss from traders not registering for the VAT (Keen and Smith, 2006). However, in developing countries both frauds seem to be extensive, where according to Schneider et al. (2011) the informal sector comprises around 40% of GDP on average (ranging up to 70%). The extensive margin of VAT evasion is studied in Hoseini (2014b) in detail, and the focus of this paper is on modeling misreporting problem in the VAT system and how the government can reduce it.

The previous evidence suggests a negative relationship between final consumption and VAT compliance of a sector. Tait (1991) indicates that the mark up ratio of the VAT, defined as reported sales divided by reported purchases in a period, is less in retail and wholesale sectors. In a randomized experiment on Chilean firms, Pomeranz (2013) presents the most compelling evidence for this relationship. She observes that firms respond more to an increase in audit probability on transactions with final consumers, where there is no paper trail, suggesting more ex-ante evasion on these transactions. The basic reason for this effect is that unlike business-to-business transactions, no third-party reported invoice exists in business-to-household sales and thus they are less risky for VAT misreporting. Consequently, sectors with more inter-firm transactions (fewer sales to final consumers) have intrinsically more VAT collection efficiency. In this paper, I pinpoint how the difference in the risk of misreporting affects the optimal enforcement policy of the government.

The model of the paper is based on the standard theoretical model for analyzing tax evasion. The tax administration seeks to minimize VAT evasion given the tax base and the cost of audit. The tax base is composed of firms heterogeneous in VAT obligation, cost of evasion and the type of customers. There are two enforcement instruments available for the administration: (i) audit staff visit the enterprise and check the transactions (visiting audit); (ii) a number of reported business-to-business transactions are cross-checked for correspondence with the other party (invoice cross-checking). The second enforcement method is much easier for the administration, but it is limited to inter-firm transactions and needs infrastructure on information reporting.

The results of the model show that the optimal invoice cross-checking rate among different commodities is positively associated with the number of firms in the market of that commodity. If the cross-checking policy is not chosen optimally, it motivates firms to shift their misreporting to the commodity that generates the lowest risk. In addition, the model intuitively imply that the optimal visiting audit rate negatively depends on the taxpayer's *cost convexity* with respect to tax evasion gamble. The degree of cost convexity depends on a variety of firm's characteristics such as size, ownership, accounting system and the risk-premium of the manager. In practice, large enterprises are less likely to engage in gross evasion, such as making uninvoiced transactions, because their accounting systems would not permit this and having numerous employees makes their collusion to unreport tax fragile (Kleven et al., 2009; Tait, 1991). In addition, risk-aversion and other subjective beliefs of the manager can change the degree of convexity of the cost function. Naturally, managers with low risk premium are more involved in tax evasion and need to be audited more frequently.

The relationship between optimal visiting audit rate and the level of final consumption of a firm is however more complicated. Since transactions with final consumers are unverifiable, on the one hand, they persuade firms to evade more, on the other hand, they reduce the expected return of a visiting audit. With low cost convexity, the optimal audit rate is an increasing function of sales to final consumers. In comparison, more unverifiable transactions do not sharply change the evasion of a firm with strict cost convexity, but it makes the detection of the fraud harder. As a result, for a firm with a very convex cost of evasion, the optimal audit rate is first increasing but after some point decreasing in sales to final consumers. Therefore, given high degree of cost convexity of large enterprises, VAT collections may improve if audit is also directed to medium and small enterprises which are more risky taxpayers. This result is at odds with the conventional perception in many developing countries that the audit should primarily be devoted to larger taxpayers having higher additional tax assessment per visiting audit (Gordon and Li, 2009).

This paper contributes to several aspects of tax enforcement literature. Most studies about the optimal design of the VAT presume that it is collected costlessly, however, VAT evasion may create a critical impact on its optimality versus other taxes like tariff (Emran and Stiglitz, 2005). To the best of my knowledge, this is the first paper modeling

VAT misreporting with an important policy implications for tax administrators in both developed and developing countries.

Despite the limited research on the VAT specific aspects of tax enforcement, the broad concept of tax compliance has attracted a lot of attention in the literature. As a general classification, Slemrod and Gillitzer (2014) categorize tax evasion models based on their assumption about ethical factors, though the two groups are not mutually exclusive. The 'deterrence' models presume the actions of taxpayers are not set by morality or social norms, but are based on the possibility of audit and punishment. This branch of literature can be traced back to the seminal formulation of Allingham and Sandmo (1972), assuming that, to understate their income, risk-averse taxpayers are constrained by a possible penalty and the expected payoff determines the level of evasion. In comparison, 'non-deterrence' models (e.g. Gordon, 1989) focus on the behavioral aspects of tax evasion arguing that Allingham and Sandmo (1972) cannot fully explain the compliance rates, especially in the developed world (for a survey see Hashimzade, Myles, and Tran-Nam, 2013). This growing branch of the literature mainly studies the effect of social norms and ethical parameters such as regret, shame, or delight at cheating on the behavior of taxpayers. Andreoni, Erard, and Feinstein (1998) indicate that, by tax evasion, people may fear social stigma or damage to reputation suggesting that factors such as a moral obligation to be truthful, or the social consequences of being a known cheater, may add further compliance incentives that are not accounted for in the standard models. In the paper, using a simple deterrence framework, I explain how the subjective beliefs of a taxpayer are reflected in his cost function and analyze their role in shaping the optimal audit policy.

The paper also contributes to the literature emphasizing the importance of different forms of third party reporting in the tax enforcement. Johns and Slemrod (2010) show that in the U.S., only 1 percent of wages and salaries is misreported, but in a sharp contrast, an estimated 57 percent of self-employment business income, which is based on self-assessment, is not reported. In a field experiment on the individual income tax in Denmark, Kleven, Knudsen, Kreiner, Pedersen, and Saez (2011) show that income tax evasion is low, except for the fraction that is self-reported. Directly relevant to this paper, Pomeranz (2013) uses a randomized experiment on Chilean firms to investigate the effect of the third party reported paper trails in the VAT on the tax compliance. She

shows that in transactions with final consumers, in which the VAT paper trail is absent, the response of firms to an audit letter message is stronger. This difference in evasion opportunity of the two types of transactions builds the micro foundations of the model of this paper.

The results of the paper also add to the recent and growing literature investigating the effect of new developments in information reporting systems on tax compliance. As a general definition, information reporting are the requirements that certain transactions causing tax obligation be reported by the third party to the tax administration (Shaw, Slemrod, and Whiting, 2010). By adding information reporting into the standard tax enforcement framework, Paramonova (2014) models how tax administration can affect the accuracy of information about taxpayers and determines the optimal tax audit rule for a given information accuracy. This paper also models the consequence of information reporting system, but in a VAT setting, and characterizes the optimal way of using such a system.

Finally, this paper contributes to the literature on predictive analytics and subjective audits in tax enforcement. In the deterrence models, the audit is usually assumed to be random; however, the majority of audits conducted in practice are risk-based and not completely random. Some papers model the optimal audit strategy as a function of reported income and the dynamics of taxpayer's behavior (e.g. Andreoni et al., 1998; Reinganum and Wilde, 1985). Another important factor determining tax fraud is taxpayer's risk attitude. The effect of risk aversion on tax evasion is traced back to Allingham and Sandmo (1972) who predict, ceteris paribus, risk-averse individuals evade less than others. Considering all of these factors, some papers suggest simulation-based approaches to predict the compliance incentive of each taxpayer, and consequently, its optimal audit rate (e.g. Engel and Hines Jr, 1999; Hashimzade, Myles, and Rablen, 2014). Analyzing the evasion risk of taxpayers, helps overcome the asymmetric information problem the tax administration faces. This paper shows that such analysis is very important in increasing collection efficiency of visiting audit and must be an essential element of VAT enforcement.

After this introduction, I explain the practical VAT enforcement issues and then model the behavior of a single firm when it decides between misreporting the transactions with other firms and final consumers. The difference between the two is that the former

is more risky for misreporting and easily detectable in case of audit, but the detection of fraud in the latter is harder, especially when misreporting comprises a small fraction of total final consumption. Next, in sections 4.4 and 4.5, respectively, I study the optimal policies in the presence and absence of an information reporting system which enables the administration to perform random invoice cross-checking. Section 4.6 concludes.

4.2. VAT audit in practice

In the VAT system, firms are required to charge tax on their output and in return deduct taxes paid on inputs from their VAT bill. The input credit is given against VAT payment as a refund and thus just VAT-payers can obtain it. The implementation of the VAT in almost all countries, except Japan[48], is based on invoice-credit form, which is the focus throughout the paper. In this method, registered sellers issue an invoice corresponding to the VAT charged on sales to each customer, who if registered, can use the invoice to get refund on inputs. Like any other tax, VAT is vulnerable to evasion and the governments must choose an auditing strategy. According to Ebrill, Keen, Bodin, and Summers (2001), in general, there are three audit methods for the VAT:

Simple self-assessment without invoice reporting Each enterprise calculates its own tax liability (usually per month or quarter) and sends the aggregate tax return forms to the tax administration. It does not have to send the invoices, but has to keep them for some years and is subject to a possible visiting audit by the tax administration. During the audit, the administration checks the book of accounts and extensively cross-checks all of the invoices with the corresponding reports of the firm's suppliers and business clients. The auditors also use other possible information like bank accounts to find out the violation in other transactions. In 2001, around half of countries with the VAT apply this method of VAT administration (Ebrill et al., 2001).

Self-assessment with invoice reporting Firms calculate their own tax liability, but also send additional documents such as invoices to the tax administration. The administration then can audit the firm in person or just by cross-checking the invoices.

[48]In Japan, each trader is taxed on the difference between sales and purchases.

This method is applied in many countries and is growing because of the new developments in information technology.

Direct government audit The taxpayer files a return in the tax office, then the administrators audit and assess the tax obligation of the firm. The method is not common nowadays and broadly was conducted in 1990s in some former Soviet Union states.

There is a trade-off between enforcement method and compliance cost of taxpayers. The advocates of the first method argue that it has the least compliance costs and is more efficient. Ebrill et al. (2001) indicate that self-assessment procedures with complex requirements place excessive compliance costs on taxpayers and can be a serious impediment to detection of delinquent taxpayers. Without invoice reporting, however, cross-checking is not possible. The early attempt of heavy cross-checking was done in South-Korea in 1980s. Overall, it was unsuccessful due to complexity of the work and the lack of IT technology. At the same time, some countries like Bolivia and Chile conducted cross-checking on a random basis, but the processes were done manually by tax auditing staff and was very time-consuming (Tait, 1991). Nevertheless, the development of information technology nowadays has increased tax enforcement performance considerably and e-auditing is now growing all over the world. With the assistance of electronic invoicing and data mining – a methodology to identify specific information from rough data via computing technology – tax administrations can collect the third party information much easier to perform data matching and finding evasion cases (Wu, Ou, ying Lin, Chang, and Yen, 2012). Currently, many countries[49] are adopting integrated electronic invoicing system which enables the government to randomly cross-check the invoices. Dealing with the huge volume of invoices might be impossible and even if the list of all unmatched transactions is available by data mining methods, identifying the fraudulent party – seller or buyer – and the arrangements of proper penalty need auditing staff. Therefore, in reality, tax authorities are able to investigate a fraction of the suspicious transactions for further auditing because of limited staff resources and have to prioritize between different commodities or sectors. Nevertheless, if such a system is developed, its variable cost for cross-checking a firm is much less than the cost of a visiting audit for the tax administration.

[49] some examples are Brazil, Bolivia, Bulgaria, Chile, China, Cote d'Ivoire, Indonesia, Poland and South Korea.

In any types of assessment, if a fraud is detected, the fraudulent firm has to pay the misreported tax plus an extra penalty. Tait (1991) indicates that when a taxpayer misreports a small amount, the purpose of the penalty is to dissuade him so that he does not repeat the violation. But when the fraud goes beyond the violation and falls into the realm of crime, harsh penalties, including jail sentences, may apply. In practice, the level of penalty is different across countries. For instance, in UK it is changing from 20% to 100% of the fraud based on its magnitude but in Argentina and Bolivia the fraudulent firm has to pay from 2 to 10 times of the misreported tax. Almost in all countries large scale fraud leads to closing of the business and also years of imprisonment. Hence, in practice the penalty is an increasing and convex function of the level of evasion.

4.3. The basis of the model

In order to model misreporting and optimal enforcement in the VAT, I follow the standard model of tax evasion and extend it by differentiating between two types of transactions. Consider an economy comprising firms heterogeneous in risk-aversion and type and level of tax obligation. Based on the type of activity, each firm makes two types of transactions: with other formal businesses (b), with final consumers (c), imposing y_b and y_c as VAT obligations respectively. As a result, each firm decides about two types of evasion e_b and e_c based on the nature of the obligation. The difference between b and c comes from their transparency to the tax administration. The administrator can realize the exact amount of y_b – and thus e_b – by cross-checking the invoices, but y_c has no corresponding third-party reported information and small amount of e_c may be unrealizable. Detection of e_c is possible only by visiting audits when the auditor checks information such as bank accounts, total turnover, size, location etc. The probability that the audit detects e_c depends on the relative extent of the fraud. If the evasion comprises a very small fraction of the tax obligation, the detection would be very hard, but the probability increases when the relative extent of fraud increases. Therefore, I assume, when an auditor conducts a visiting audit, the probability that he detects the fraud on sale to final consumers is e_c/y_c.

There are two enforcement methods for the tax administration. The first one involves

random cross-checking and needs infrastructure on information reporting. The second one is self-assessment without invoice reporting. In the first method, the administration has two separate tools for each type of evasion. It randomly cross-checks a share of inter-firm transactions to detect e_b, but e_c can be detected only by visiting audit. In the simple self-assessment system, no random invoice cross-checking is possible and if a visiting audit takes place, the auditors thoroughly checks all VAT invoices for correspondence to find y_b, and also collects other information for better estimation of y_c. As a results, the fraud in transactions with formal firms (e_b) is for certain detected, but the probability that the auditor detects the fraud on sale to final consumers is e_c/y_c.

If a fraud is detected, a penalty is applied by the administration which is always greater than the amount of evasion. If the detected evasion is near zero the penalty approaches to the principal of unpaid obligation. Based on the facts explained above, I assume $\theta(e)$, an increasing and convex function of the detected evasion e with the following characteristics, determines the penalty for the firm:

$$\theta(0) = 0,\ \theta'(0) = 1, \quad \forall e > 0 :\ \theta(e) > 0,\ \theta'(e) > 1,\ \theta''(e) > 0,\ \theta'''(e) \geq 0 \tag{4.1}$$

These conditions yield $\theta(e) \geq e$.[50] Consequently, by defining e_b and e_c as above and assuming the tax administration conducts a visiting audit with probability λ, the expected cost of evasion for business and final transactions will be $\lambda\theta(e_b)$ and $\lambda e_c/y_c\theta(e_c)$ respectively. However, as a pure deterrence model, these expected amounts just reflect the monetary punishment of a risk-neutral taxpayer with no other constraint on evasion.[51] In practice, firms can have extra restrictions on tax evasion due to the structure of the company or factors like risk aversion of manager and social norms. In below, when discussing the optimal visiting audit policy, I take this cost variation into account by considering heterogeneity in the cost function $\theta(.)$. Prior to that, the results does not depend on the functional form of $\theta(e)$.

For modeling the enforcement policy, I assume that the tax administration faces an economy with heterogeneous firms and I intermediate goods. The administration is not fully informed about the heterogeneity of the firm, but using predictive analytics can

[50] The assumption about the third derivative of θ is made to have a smooth cost function and simpler calculations in Lemma 4.1, but the general results does not depend on that.

[51] As I discuss in section 4.4.4, this is a reasonable assumption for formulating cross-checking policy, where the administration audit invoices not firms.

categorize them into S different clusters based on the type of activity, the estimated amount of VAT obligation, and characteristics of the firm.[52] Each intermediate good is produced by one activity and is used as an input by a number of activities. In the presence of invoice reporting, the policy vector of the government includes μ_i which is the ratio of transactions of each commodity i to be cross-checked and λ_s which is the audit rate in each cluster s. In the self-assessment without invoice reporting, cross-checking is not available and the only policy instrument is λ_s. The aim of the tax administrator is finding the optimal allocation of λ_s and μ_i across clusters and commodities based on aggregate and firm-level factors.

In this paper, although the decision of firm for evasion (e_c and e_b) depends on the enforcement strategy of the tax administration, production and tax obligation y_b and y_c are held fixed. This is the common assumption in models for analysis of tax evasion, like the seminal formulation of Allingham and Sandmo (1972) which has been the dominating theoretical model in this literature. This assumption is especially plausible for the VAT evasion, since *the VAT is essentially a tax on final consumption and it does not distort the profit maximization of firms*. Hence, one can reasonably assume that the decisions about production and VAT evasion are orthogonal. Besides, the aim of this paper is characterizing the optimal audit strategy for the tax administration and the determinants of the tax base is not the subject of discussion.[53]

In order to find the optimal visiting audit policy, the paper seeks to find the audit rate as a function of the estimated tax obligation on sales to final consumers and risk-aversion of the firm. Moreover, as it will be discussed in section 4.4.4, the optimal policy for invoice cross-checking is independent of tax obligation and depends on the number of firms in the economy. In the following, for better mathematical tractability, I first analyze the firm's decision and the optimal policy, when an invoice cross-checking system is available for the tax administration, and then study the model in the absence of such system (simple self-assessment).

[52] The factors to help predict the heterogeneity are discussed in section 4.4.2.

[53] Minimizing evasion per se is an important objective from normative point of view, since it reduces the tax inequality between compliant and non-compliant taxpayers. Moreover, unequal tax burden intensifies tax distortion at the aggregate level.

4.4. The model with random invoice cross-checking

In this part, I model the consequences of having an integrated invoice system that enables the government for random cross-checking. As mentioned above, such a system is now more feasible with the advancements in information technology and is growing all over the world. In general, there are two types of invoices in the VAT system: (1) *sales invoice* shows the tax payment by the seller, (2) *purchase invoice* is used by the buyer to get credit on inputs. Sales can be understated in the form of not reporting a number of sales invoices by the seller, in comparison the buyer can create some fake purchase invoices to overstate the credit due on its input. Therefore, when the tax administration accumulates all reported invoices in one place, some sales invoices are missing and some extra purchase invoices are generated by the buyers. For performing cross-checking, after random selection of n purchase invoices, the administration cross-checks them with the corresponding sales invoice. The number n is a policy variable determined by the cost of cross-checking and total number of purchase invoices N in the economy.

The strategy of the tax administration after drawing and cross-checking n purchase invoices is finding the firms that made the violation. I assume, after detection of a violation, the administration cross-checks all other invoices of the suspicious firm, reimburses the principal tax, and charges an extra penalty. I also assume that, in a firm's view, invoice cross-checking is conducted independent of visiting audits and it optimizes the evasion on inter-business transactions e_b and final consumer sales e_c in two separate problems. This assumption enables me to find the analytic solutions of e_b and e_c by separating the effects of the two enforcement tools. In Appendix 4.A, I relax this assumption by assuming the two types of evasion are jointly determined in one optimizations problem and show that in practical circumstances the two problems become very similar.

Before analyzing the general random invoice cross-checking problem, I study a simple market of one intermediate good in which an upstream sector u sells the good to one downstream sector d. The optimization problem of a single firm in sector u (d) is how many sale (purchase) invoices to misreport given the risk of detection. Each invoice represents a unit of transaction with value α based on the type of the intermediate good.[54] If a firm misreports k invoices of the intermediate good, its evasion in this

[54] α shows the average value of one invoice of the good. This assumption is made to distinguish between commodities with high value per unit (e.g. Steel bar) and others (e.g. bread).

market will be $e_b = \alpha k$, and the probability that no misreported invoice is cross-checked becomes $(1 - k/N)^n$ where, N and n are the total number and the number of cross-checked invoices in the market of the intermediate good.[55] In practice, the number of reported invoices in a market is so large that a single firm cannot change it and takes it as given,[56] therefore I assume that $k << N$ and thus $k/N \approx 0$. As a result, we can use first order approximation and write $(1 - k/N)^n = 1 - kn/N$, which means that the probability of detection of at least one misreported invoice by the tax administration and getting fined is kn/N, where $k = e_b/\alpha$. Then, the optimization problem of a single firm – either in u or d – can be written as

$$\max_{e_b} \quad e_b - \mu \frac{e_b}{\alpha} \theta(e_b) \tag{4.2}$$

where $\mu = n/N$ is the share of invoices that are cross-checked by the tax administration (the policy variable) and the second term represents the expected cost of detection through invoice cross-checking. Then, the FOC becomes

$$\theta(e_b) + e_b \theta'(e_b) = \frac{\alpha}{\mu} \tag{4.3}$$

Because θ is a strictly increasing and convex function, one can easily show that e_b decreases when n/N goes up. Therefore, when the administration cross-checks a larger share of invoices, inter-firm misreporting reduces.

On the other hand, the tax administration conducts visiting audits aiming at detection of evasion on unverifiable transactions y_c. As a consequence, the firm faces another optimization problem on e_c due to the expected cost of such audits

$$\max_{e_c} \quad e_c - \lambda \frac{e_c}{y_c} \theta(e_c) \tag{4.4}$$

where λ stands for the share of firms that are randomly audited by the tax administration and e_c/y_c is the probability of detection of the unverifiable fraud. Then the choice of

[55] The probability that a random invoice is not one of the misreported by the firm is $1 - k/N$, therefore the probability that *none* of the n draws are misreported by the firm is the multiplication of n single probabilities $1 - k/N$ which becomes $(1 - k/N)^n$.

[56] For instance, around 26 billion VAT invoices are reported in Germany in 2006 (Baldwin, 2007).

firm for e_c can be found from the FOC

$$\theta(e_c) + e_c\theta'(e_c) = \frac{y_c}{\lambda} \tag{4.5}$$

Proposition 4.1. *In the presence of invoice cross-checking, total evasion $e_b + e_c$ is decreasing in visiting audit rate λ and invoice cross-checking rate μ and increasing in VAT obligation on final consumption y_c.*

Proof. Appendix 4.B. *Q.E.D.*

Proposition 4.1 indicates that if the government utilizes the cross-checking technology there is a negative relationship between final consumption and VAT compliance. Moreover, total evasion is decreasing in visiting audit and cross-checking rates. In the following, after explaining the role of subjective measures in evasion and the methods for measuring them, I determine the optimal visiting audit rate in cluster s of firms (λ_s). Then, I find the optimal policy for the cross-checking rate of the invoices of commodity i (μ_i).

4.4.1. Objective versus subjective costs of fraud

So far, I discussed the role of sectoral characteristics on firm's evasion and the assumptions about cost of detection in (4.1) just reflect the pure monetary cost of penalty for a risk-neutral taxpayer. This assumption is reasonable for formulating the optimal cross-checking policy, because in practice, the variable cost of cross-checking is small and as I will show in section 4.4.4, Proposition 4.3, this approximately yields the same cost function for all firms ($\theta(e) \approx e$). In addition, in the cross-checking process, the tax administration just deals with invoices not firms, and in his view, the invoices are just heterogeneous in terms of the activity and commodity not their issuers. Thus, it is not possible to make distinction between the cost function associated with each invoice.

For finding optimal visiting audit rate, however, the administration is in direct contact with the firm. In reality, regardless of the third-party reported business-to-business transactions, firms are heterogeneous in terms of detection probability in a visiting audit due to the other subjective costs of tax evasion. According to (4.4), the expected cost of evasion e_c is $\lambda \times e_c/y_c \times \theta(e_c)$ where λ is visiting audit probability and e_c/y_c reflects the firm's advantage in unverifiability of the sales to final consumers. In this context,

the other factors that determine the evasion cost and shape the decision of the firm are reflected in the cost function $\theta(e_c)$. Hence, the form of cost function plays an important role in determining the expected return of the auditing case for the administration. As discussed in section 4.2, in reality, the monetary penalty after detection is a convex function of the level of detected evasion. In this framework, any increase in the expected cost of evasion, which is not due to monetary penalty, can be translated into an increase in the convexity of the cost function. There are a number of firm-level factors that can cause extra evasion costs and raise the degree on convexity of $\theta(.)$. These factors can be classified as follows

1. *Enterprise structure:* Ownership, size, and the accounting system of the enterprise play important roles in shaping its tax evasion decision. In the incorporated companies, shareholders, the CEO, and the tax manager do not necessarily have similar preferences and objectives regarding tax compliance. Kleven et al. (2009) argue that the firm's employees must collude to unreport tax and such a collusion is fragile in large enterprises with numerous employees. In this context, there are fewer opportunities to evade VAT in incorporated companies rather than sole traders and small partnership firms. In addition, a stringent and transparent accounting system (e.g. electronic invoicing, external accountants, etc.) acts as another constraint for tax evasion. All of these restrictions on evasion opportunities can also be interpreted as an increase in the chance of detection and the risk of evasion for the enterprise which is not due to higher audit rate λ or more evasion on final sales e_c/y_c, rather they are reflected in the cost function $\theta(.)$ by increasing its convexity.

2. *Risk attitude:* This factor reflects the risk preference of the manager and becomes more important for unincorporated enterprises, with one decision maker about tax compliance. In general, managers do not have similar impressions about the cost of evasion and their subjective views about risk and social norms differentiates their attitudes toward tax fraud (Andreoni et al., 1998). For some people, conviction for tax evasion is unimaginable, but a habitual criminal may think about it just as a worst case scenario. Therefore, the cost convexity also depends on the risk preference of the manager, especially in smaller firms, such that the degree of convexity is higher for a firm with a risk-averse manager. In this context, Gordon

(1989) assumes that in addition to the pecuniary cost, tax evasion generate a "stigma cost" for taxpayers that can be interpreted as an increase in the convexity of the cost or the concavity of the expected payoff from evasion.

3. *Future considerations:* Fear of stigma and damage of reputation may also be translated into a monetary cost. If the detection of evasion results in loss of reputation which, in turn, leads to market losses in the future, the cost convexity increases. This extra cost might be higher for larger companies in competitive sectors.

In short, in the model, psychological and cultural aspects of taxpayers such as risk aversion, tax morale, patriotism, guilt, and shame, plus all other factors that limit evasion opportunities by raising detection probability or causing extra monetary cost for evasion, increase the convexity of cost function $\theta(.)$. In order to specify a measure for these subjective costs, first, I find evasion as a function y_c and λ.

Lemma 4.1. *In the presence of invoice cross-checking, we can write* $e_c = g(\frac{y_c}{\lambda})$*, where* $g : R^+ \to R^+$ *is a strictly increasing and concave function and* $\forall x \geq 0,\ 0 < g'(x) \leq \frac{1}{2}$.

Proof. Appendix 4.C. *Q.E.D.*

According to Section 4.2, the cost of fraud is always a convex function and this makes e_c a strictly concave function of y_c/λ. Therefore, *cost convexity* – the convexity of $\theta(.)$ – is equivalent to *evasion concavity* – the concavity of $g(.)$. In this setting, taxpayers are heterogeneous in two dimensions: One is y_c which reflects the possibility of evasion; another is $g(.)$ which reflects the subjective costs of each taxpayer. So far, I assumed the same $\theta(.)$ – and thus $g(.)$ – for all taxpayers, but hereafter, I add the possibility of different functional forms to capture the heterogeneity in cost convexity. The curvature of $\theta(.)$ – and consequently $g(.)$ – helps to characterize the cost of evasion for each enterprise. In order to measure the curvature, I use a measure similar to the conventional index of relative risk aversion (RRA).[57] Specifically, $\gamma(x) = -xg''(x)/g'(x)$ gives the degree of concavity of the function $g(.)$ and, as a result, the degree of convextiy of cost of evasion $\theta(.)$. Thus, *higher* γ *means more cost convexity.* In the next parts, after discussing the

[57] Slemrod and Yitzhaki (2002) indicate that tax evasion is akin to the choice of how much to gamble. Each unit of misreporting the VAT offers a payoff but may lead to a penalty. Consider e_c as the utility of the firm and y_c/λ as asset – which here is the possibility of evasion – then $\gamma(x) = -xg''(x)/g'(x)$ is the relative risk aversion of the taxpayer toward VAT fraud: The more curved the function $g(.)$ is, the lower will be its certainty equivalent of a risky bundle.

factors that help the government to estimate γ, I show how it affects the optimal visiting audit rate of the tax administration.

4.4.2. Predictive analytics and cost convexity estimation

To estimate the heterogeneity of taxpayers in terms of cost convexity and subjective factors, a growing literature proposes the use of predictive analytics (Hashimzade et al., 2014). This type of analysis provides a set of tools to use historical data to predict future outcomes and create a consistent risk ranking of individuals or firms. The application of predictive analytics are not limited to tax administrators and it is now spreading to other public service organizations such as police departments for crime prediction.[58] In general, there are a lot of factors that can be used for predicting the future behavior of individuals, but in the following, I only mention a number of them that can help tax administrations to estimate the subjective costs of taxpayers.

According to Tait (1991), VAT evasion is a concave function of firm's size suggesting a negative relationship between size and cost convexity. As explained above, larger firms usually have numerous employees and often hire external accountants that might mistakenly report the evasion or make a rational whistle-blowing (Kleven et al., 2009). Therefore, the collusion for hiding the evasion breaks easier in a large enterprise rather than a small one with limited workers, resulting more cost convexity for larger firms.

The ownership of the firm is another factor to predict the cost convexity. Incorporated enterprises comprise divisions with different roles and maybe conflicting incentives (shareholders, the CEO, tax agent, etc.) and this diversity reduces the chance of a compromise for tax evasion. Egger, Erhardt, and Keuschnigg (2014) indicate that the entrepreneurial enterprises, run by the owner, are more likely to evade tax rather than managerial enterprises that have an external manager.

The third factor to estimate the cost function is the age of the firm (Feinstein, 1991). Often, younger firms are more likely to evade tax, because at the early stages the enterprise faces more financial constraints. Moreover, many firms are not precisely aware about tax rules at the time of commencement of operation, and may mistakenly not pay their true tax obligation.

[58] Mastrobuoni (2014) compares the performance of two police forces in Milan and shows that the adoption of a predictive policing software in one police force has doubled the productivity of policing.

At individual level, gender and education of the taxpayer has been revealed to be significant determinants of tax evasion. Kastlunger, Dressler, Kirchler, Mittone, and Voracek (2010) show that the tax compliance of females is normally higher than males. Witte and Woodbury (1985) find a negative association between the general education level of taxpayers and tax evasion. The occupational choice and social networks are other factors revealing the risk behavior of taxpayers (Hashimzade et al., 2014).

Another important element of predictive analytics is taxpayer's past reputation and experience with respect to tax compliance. Bruttel and Friehe (2014) show that tax compliance is path dependent in the sense that past experiences with different enforcement level affects current tax compliance. When the tax administration pursues a strict enforcement policy, taxpayers who have experienced weak enforcement early on tend to be less compliant than taxpayers with strict enforcement experience in the past. Similarly, under weak enforcement, taxpayers with strict enforcement experience declare more income than taxpayers who have only experienced weak enforcement. Along the same line, Dai, Hogarth, and Villeval (2015) show that if audits are initially regular and systematic, taxpayers decrease their expectation for a new audit immediately after one is carried out. In contrast, taxpayers who have experienced less frequent and more irregular audits, maintain beliefs of new audit and less likely to show a time-dependent noncompliance behavior. Hence, the dynamics of taxpayer's behavior are also important for performing predictive analytics.

4.4.3. Optimal visiting audit rate

In order to find the optimal visiting audit λ in each cluster, the tax administration seeks to minimize evasion in sales to final consumers given visiting audit is costly. Therefore, the optimization problem is

$$\min_{\lambda_s} e_c^s + \eta\lambda_s \tag{4.6}$$

where λ_s is the rate of visiting audit in cluster s, η is the cost of auditing one firm,[59] and e_c^s is the estimation of tax administration from the evasion in cluster s. Then, FOC of

[59] An equivalent assumption about cost of auditing to make η endogenous is assuming the tax administration is able to audit a limited number of firms and its objective is minimizing total evasion. Then, the optimization problem becomes: $\min \sum_s m_s e_c^s$ subject to $\sum_s m_s \lambda_s = \bar{L}$, where m_s is the number of firms in cluster s and $\bar{L}$ is the audit capacity of the administration. In this case, the FOC becomes the same as (4.7) with η the Lagrange multiplier or the shadow cost of a single audit.

(4.6) can be written as

$$\frac{\partial e_c^s}{\partial \lambda_s} = -\eta \tag{4.7}$$

Thus, the audit rate should be adjusted at the level that the marginal reduction in the evasion of a firm is equal to the (shadow) cost of auditing. This means that at the optimum marginal reduction in the evasion is the same in all clusters.

One important issue for the policymakers is how to overcome information asymmetries to estimate e_c^s. Based on Lemma 4.1, e_c depends on y_c and $g(.)$, therefore to find e_c^s, the tax administration needs an estimation of those parameters for each firm. The size of sales to final consumer is easier to estimate and can be found by checking the information on bank accounts and location as well as the type of activity (for instance the bulk of costumers of a retailer or a barber shop are final consumers). Estimation of the cost function of each taxpayer can be done by predictive analytics as discussed above in detail.

Define y_c^s and $g_s(.)$ as the estimated value of tax obligation on sales to final consumers and the functional form of $g(.)$ in cluster s. Then, the following proposition shows how the level of audit rate depends on subjective characteristics.

Proposition 4.2. *In the presence of invoice cross-checking, if λ_s is the optimal choice of the government in cluster s, we have*

$$\lambda_s^2 = \frac{y_c^s}{\eta} g_s'\Big(\frac{y_c^s}{\lambda_s}\Big), \qquad \frac{\partial \lambda_s}{\partial y_c^s} = \frac{\lambda_s}{y_c^s}\frac{1-\gamma_s}{2-\gamma_s} \tag{4.8}$$

where $\gamma_s = -\dfrac{\frac{y_c^s}{\lambda_s} g_s''(\frac{y_c^s}{\lambda_s})}{g_s'(\frac{y_c^s}{\lambda_s})}$.

Proof. Appendix 4.D. *Q.E.D.*

Proposition 4.2 indicates that the relationship between the optimal audit rate and final consumption of a cluster depends on the cost convexity of taxpayer to evade e_c defined as γ_s. If the cost convexity in increasing in y_c/λ, it means that when the taxpayer experiences an increase in the possibility of evasion, he chooses to decrease the share of evasion in total tax obligation on final consumptions. Figure 4.1, illustrates the optimal

audit rate for two different functional forms of $g(.)$. From (4.8), we can write

$$\eta\lambda_s = \frac{y_c^s}{\lambda_s} g'(\frac{y_c^s}{\lambda_s}) \tag{4.9}$$

and therefore graphically find the optimal audit rate in each point. The slopes of the tangent lines reflect $g'(y_c^s/\lambda_s)$ and the distance from the vertical axis is y_c^s/λ_s. Thus, $\eta\lambda_s$, the multiplication of the two, is the height of the triangle they shape. The below curve with an asymptote stands for higher relative risk aversion and at the optimum has smaller audit rate. The horizontal asymptote in $g(.)$ corresponds to a vertical asymptote in the cost function $\theta(.)$. As mentioned in Section 4.2, many countries have implemented harsh penalties for big tax frauds and from one angle, the asymptote in the cost function can represent the level of evasion in which the defrauder is convicted to long-term imprisonment or life in prison. However, from different viewpoint, it reflects the subjective characteristics of a large firm that have a restricted evasion opportunities or an entrepreneur who never imagines evasion more than the asymptote due to high risk aversion or fear of stigma. In the next step, I characterize the relationship between λ_s and y_c^s based on the different amounts of cost convexity γ_s.

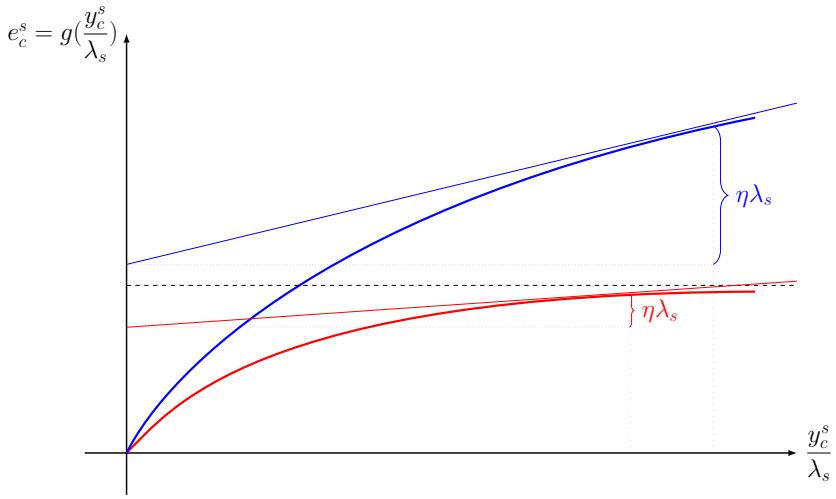


Figure 4.1: Optimal audit rate for two asymptotic and non-asymptotic cost functions.

Lemma 4.2. *For all $g : R^+ \to R^+$ that satisfy the conditions of Lemma 4.1, at the government's optimum we have*

$$\lim_{y_c^s \to 0} \lambda_s = 0, \quad \lim_{y_c^s \to 0} \gamma_s = 0, \quad \lim_{y_c^s \to 0} \frac{\partial y_c^s}{\partial \lambda_s} = +\infty,$$

Proof. Appendix 4.E. *Q.E.D.*

Lemma 4.2 shows when y_c^s is near zero, λ_s is zero but its derivative is $+\infty$. In this neighborhood the relationship between λ and y_c^s is independent of γ_s. However, for a larger level of y_c, the relationship becomes very sensitive to the cost convexity of taxpayers. Using Proposition 4.4, we can distinguish between four different cases:

1. low convexity ($\gamma_s < 1$): λ_s is increasing in y_c^s.

2. medium convexity ($\gamma_s = 1$): λ_s is constant in y_c^s.

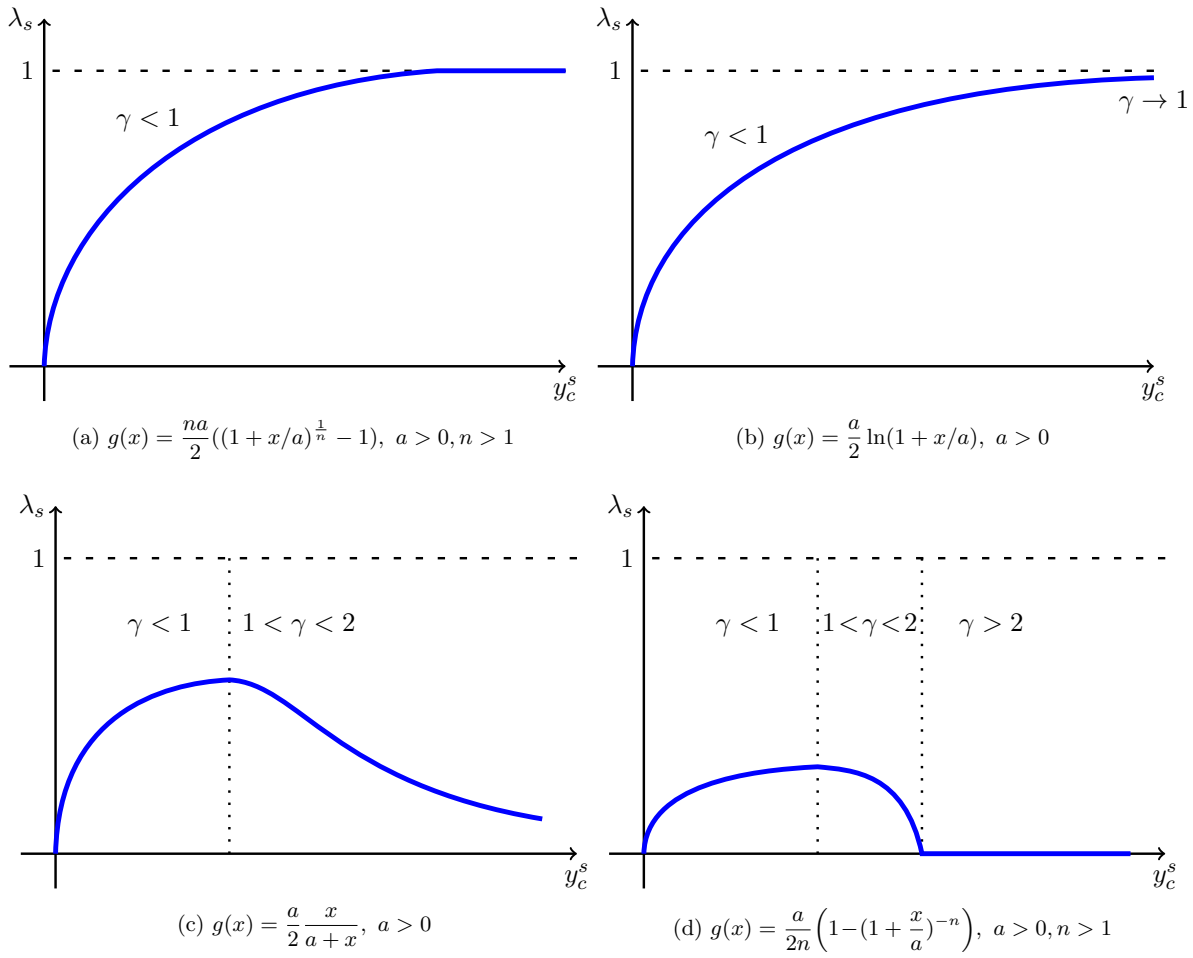
3. high convexity ($1 < \gamma_s < 2$): λ_s is decreasing in y_c^s.

4. very high convexity ($\gamma_s \geq 2$): $\lambda_s = 0$ and e_c^s is equal to its horizontal asymptote.

For any form of $g(.)$ that satisfies the conditions of Lemma 4.1, at $y_c^s = 0$ the cost convexity is zero, but it can change at higher level of y_c^s. For better interpretation, I use four different functional forms for $g(.)$ and show how they shape the link between λ_s and y_c^s. The details of calculations are available in the Appendix 4.F. Figure 4.2-a shows this relationship for $g(.)$ as a n-th root function $g(x) = \frac{1}{2}na(\sqrt[n]{1+x/a} - 1)$ where $a > 0, n > 1$. In this example, $\gamma_s = \dfrac{(n-1)x}{n(x+a)}$ and $\forall x > 0,\ \gamma_s < 1$. Therefore, λ_s is an increasing function of y_c^s and for large enough amounts of y_c^s, the optimal λ_s has a corner solution equal to one.[60] The second example is when $g(.)$ is a logarithmic function $g(x) = \frac{1}{2}a\ln(1 + x/a),\ a > 0$. Here, $\gamma_s = \dfrac{x}{x+a}$ and λ_s is increasing in y_c^s, but as y_c^s approaches infinity, $\partial\lambda_s/\partial y_c^s$ approaches zero which means, unlike 4.2-a, λ_s never reaches 1. The other two cases are for taxpayers with high cost convexity such that $g(.)$ has a horizontal asymptote. In Figure 4.2-c, $g(x) = \dfrac{ax}{2(a+x)}$ and $\gamma_s = \frac{2x}{a+x}$. Therefore, by increasing y_c^s, at first λ_s increases, then reaches a maximum when $\gamma_s = 1$ ($y_c^s = a\lambda_s$), and afterward, where $1 < \gamma_s < 2$, it becomes decreasing. The last case in Figure 4.2-d is for very high cost convexity where $g(x) = \frac{1}{2n}\left[1-(1+\frac{x}{a})^{-1/n}\right]$ and $\gamma_s = \frac{(n+1)x}{(a+x)}$. Similar to Figure 4.2-c, around $y_c^s = 0$, λ_s is increasing, and when $1 < \gamma_s < 2$, it becomes decreasing. When γ_s approaches 2, which means $\frac{y_c^s}{\lambda_s} \geq \frac{2a}{n-1}$, then $\partial\lambda_s/\partial y_c^s = -\infty$ and λ_s drops to zero and remains constant for higher levels of y_c^s. In this case, the amount of evasion stays equal to the horizontal asymptote of g(.) which is equal to $\dfrac{a}{2n}$. The

[60] In reality, because the cost of audit η is large, this hypothetical solution is reached in very large levels of y_c^s that does not happen in practice.

reason of the negative relationship between audit rate and final consumption in the last two examples is the existence of asymptote in the cost function. Basically, no evasion is possible above the asymptote and higher final consumption just makes detection of fraud difficult (lower e_c/y_c), without increasing evasion.



(a) $g(x) = \frac{na}{2}((1+x/a)^{\frac{1}{n}} - 1),\ a > 0, n > 1$

(b) $g(x) = \frac{a}{2}\ln(1+x/a),\ a > 0$

(c) $g(x) = \frac{a}{2}\frac{x}{a+x},\ a > 0$

(d) $g(x) = \frac{a}{2n}\left(1-(1+\frac{x}{a})^{-n}\right),\ a > 0, n > 1$

Figure 4.2: Government audit and firm's final consumption. For risk neutral firms the relationship is always positive, but for very risk averse taxpayer it may turn to be decreasing or become zero.

Figure 4.2 illustrates the importance of subjective restrictions and beliefs of taxpayers in shaping the optimal visiting audit rate of the tax administration. Therefore, all factors mentioned above can play a role in determining the audit rate of a firm. Importantly, as large enterprises are unlikely to engage in gross evasion due to organizational limitations or to avoid hurting their reputation, they have higher γ_s than smaller taxpayers. Therefore, when the structure of a large enterprise restricts its ability for tax evasion, the expected return of auditing it for the administration declines. This may lead to less optimal visiting audit rate for them specially, when invoice cross-checking is conducted

separately. In this case, the total revenue may decline if the audit of medium and small taxpayers is downplayed or neglected. This is in contrast with the conventional wisdom in many developing countries to over-allocate the audit staff in larger firms (Gordon and Li, 2009). In addition, the government should collect information about the history and various characteristics of taxpayer firms such as ownership, age, gender and eduction of the manager to have a better estimation of their cost convexity and risk behavior. As discussed above, a young and entrepreneurial firm with male and uneducated manager tends to have smaller cost convexity than a big incorporated firm. Overall, the results of this section highlight the importance of predictive analytics in determining visiting audit policy.

4.4.4. Optimal invoice cross-checking

The second policy instruments of the government is the cross-checking share for each commodity. With the advanced information reporting system, the government receives the invoice of each transaction and decides on the share of input credit invoices of each intermediate good to be drawn for cross-checking (μ_i). In this case, the policymaker's question is finding the optimal allocation of $\mu_1, \ldots, \mu_I$. In contrast, to optimal visiting audit policy, cross-checking policy does not depend on the subjective costs of a each single taxpayer, because the administration randomly checks invoices irrespective of the issuer firm. Therefore, for finding the optimal μ_i, the tax administration uses the cost function of a representative firm that sells or purchases the commodity i.[61] If the number of sellers and buyers of the intermediate good i are M_u^i and M_d^i and the buyers have no other input – thus they can just over-report the purchases of i –, the optimization problem for the government will be

$$\min_{\mu_i} \quad (M_i^u + M_i^d)e_b^i + \delta\mu_i N_i \tag{4.10}$$

where δ is the variable cost of cross-checking one invoice and finding its violating firm, $\bar{N}_i$ is the real number of firm-to-firm transaction invoices, and $N_i = \bar{N}_i + M_d k_i$ is total number of purchase invoices.

Proposition 4.3. *The optimal share of invoice cross-checking μ_i is positively associated*

[61] In practice the variable cost of invoice cross-checking is small, which means if such a system is implemented, the level of evasion e_b becomes small too. From (4.17), since $\theta'(0) = 1$, for small e_b we can approximate $\theta(e_b) \approx e_b$ to formulate the optimal μ_i.

with $\frac{M_i^u + M_i^d}{\bar{N}_i}$. *In addition, if the variable cost of cross-checking* δ *is small enough, then at the government's optimum*

$$\mu_i = \min\left[1\ , \sqrt{\frac{M_i^u + M_i^d}{\delta \bar{N}_i}\alpha_i}\right] \tag{4.11}$$

Proof. Appendix 4.G. *Q.E.D.*

Proposition 4.3 states the condition for optimal cross-checking rate of a commodity, when the number of traders and invoices are known. Because all firms trading the commodity i have the same chance of getting caught, and in case of detection of one misreported invoice, all verifiable transactions are cross-checked, the incentive for misreporting is the same among them. Therefore, the higher the number of firms, the more the overall misreporting. In addition, as the variable cost of cross-checking δ diminishes by new technologies, the audit rate can be estimated by (4.11). This estimation is independent of the form of cost function, because small δ results in small e_b, and near zero $\theta(e_b) \approx e_b$. If δ becomes very small, then the optimal decision of the government is cross-checking all transactions of the commodity i.

In the model, the upstream firms produce just one product and their misreporting is limited to invoices of good i, but in (4.10) and Proposition 4.3, I assume that the buyers also has just one intermediate good as input. In a general framework, the customer firms may have more than one intermediate input and, as a result, decide about the evasion of each input separately. Before studying the general case, assume that there is one downstream which buys two differentiated inputs from two upstream activities 1 and 2 and it can over-report the purchases of either 1 or 2.

The government draws the shares μ_1 and μ_2 input credit invoices of the two commodities and all firms take the total number of invoices N_1 and N_2 as given. In this case, the optimization problem for the upstream firms are the same as (4.2), but the probability of not being detected for the downstream is $(1-k_1/N_1)^{n_1}(1-k_2/N_2)^{n_2}$. Since N_1, N_2 are large it can be approximated by $1 - \mu_1 k_1 - \mu_2 k_2$. Because $k_i = e_{bi}/\mu_i$, the optimization problem of downstream becomes

$$\max_{e_{b1}^d, e_{b2}^d \geq 0} \quad e_{b1}^d + e_{b2}^d - (\frac{\mu_1}{\alpha_1}e_{b1}^d + \frac{\mu_2}{\alpha_2}e_{b2}^d)\theta(e_{b1}^d + e_{b2}^d) \tag{4.12}$$

which gives the FOCs as

$$1 - (\frac{\mu_1}{\alpha_1} e^d_{b1} + \frac{\mu_2}{\alpha_2} e^d_{b2})\theta'(e^d_{b1} + e^d_{b2}) - \frac{\mu_1}{\alpha_1}\theta(e^d_{b1} + e^d_{b2}) \leq 0 \tag{4.13}$$

$$1 - (\frac{\mu_1}{\alpha_1} e^d_{b1} + \frac{\mu_2}{\alpha_2} e^d_{b2})\theta'(e^d_{b1} + e^d_{b2}) - \frac{\mu_2}{\alpha_2}\theta(e^d_{b1} + e^d_{b2}) \leq 0 \tag{4.14}$$

To have nonnegative e^d_{b1} and e^d_{b2} both FOC cannot hold when $\frac{\mu_1}{\alpha_1} \neq \frac{\mu_2}{\alpha_2}$. In this case, at the optimum, at least one of the variables is zero and based on the amount of μ_1/α_1 and μ_2/α_2 we have three possibilities:

1. if $\frac{\mu_1}{\alpha_1} < \frac{\mu_2}{\alpha_2}$, we have $e^d_{b2} = 0$ and $e^d_{b1} = e_{b1}$
2. if $\frac{\mu_1}{\alpha_1} > \frac{\mu_2}{\alpha_2}$, we have $e^d_{b1} = 0$ and $e^d_{b2} = e_{b2}$
3. if $\frac{\mu_1}{\alpha_1} = \frac{\mu_2}{\alpha_2}$, the two FOC yield the same result $e_{b1} = e_{b2}$ which means d is indifferent between the two commodities and we have $e^d_{b1} + e^d_{b1} = e_{b1}$

Next, the government's optimization for μ_1 and μ_2 is obtained by solving

$$\min_{\mu_1,\mu_2} \quad M^u_1 e_{b1} + M^u_2 e_{b2} + M^d \max[e_{b1}, e_{b2}] + \delta(\mu_1 N_1 + \mu_2 N_2) \tag{4.15}$$

According to Proposition 4.3, at the government's optimum, μ_i is an increasing function of the number of evading firms. Besides, Proposition 4.1 indicates that for each commodity i, e_b is a strictly decreasing function of μ_i. Therefore, when the tax administration sets its policy from Proposition 4.3, we can write the equilibrium misreporting of intermediate good i as

$$e_{bi} = h_i(M_i) \tag{4.16}$$

where $h_i(.)$ is a decreasing function of M_i, the total number of firms that misreport the commodity i. If $\exists\ i \in \{1, 2\}$ such that $h_i(M^u_i + M^d) > h_j(M^u_j)$, then μ_1 and μ_2 are determined independently and all e^d_b is on commodity i. However, if $h_i(M^u_i + M^d) > h_j(M^u_j)$ never holds, the optimal strategy is setting $\frac{\mu_1}{\alpha_1} = \frac{\mu_2}{\alpha_2}$ such that $e_{b1} = e_{b2}$. In the general setting, the formulation of optimal policy is obtained in the same way.

Proposition 4.4. *Consider a downstream activity that uses J different intermediate goods as input. Derive the minimum number $j \in \{1, \ldots, J\}$ that can separate the goods into two mutually exclusive sets P_j and Q_j with j and $J - j$ elements respectively, in the*

sense that $\forall p \in P_j, q \in Q_j$*, we have* $h_p(M_p^u + \frac{1}{j}M^d) > h_q(M_q^u)$ *(if* $Q = \emptyset$ *put* $j = J$*). Then, the optimal policy* $\forall p_1, p_2 \in P_j$ *is*

$$\frac{\mu_{p_1}}{\alpha_{p_1}} = \frac{\mu_{p_2}}{\alpha_{p_2}}$$

In addition, $\forall q \in Q_j$*, the downstream misreporting for commodity* q *is zero and the optimal policy is determined from Proposition 4.3, given the number of other traders.*

Proof. Appendix 4.H. *Q.E.D.*

Proposition 4.4 provides the optimal policy of the government when a downstream activity has multiple inputs. At the optimal equilibrium, these conditions hold in all downstream activities that have multiple intermediate commodities as inputs. The optimal decision of a firm in such activities is over-reporting the input(s) that has the lowest cross-checking probability. Therefore, the government should allocate cross-checking rates across commodities such that a single commodity does not attract a large share of firms for misreporting. This suggests that commodities that are the input of a lot of other activities and have a lot of traders should be under more cross-checking than others. Here, for simplicity, I assumed that each activity produces just one commodity, but may use a number of inputs. In a general setting, if suppliers also have a number of products to sell, the administration should take into account that they can also switch between different products for evasion in the form of under-reporting the sales. Then, Proposition 4.4 should be extended to the case that sellers can also choose which output to under-report.

4.5. Self-assessment method without invoice cross-checking

In this section, I assume that the VAT administration uses self-assessment without invoice reporting and at each time randomly audits a share of firms in each cluster. When a visiting audit takes place, all VAT invoices are checked for correspondence and e_b is revealed, but the chance of detecting e_c depends on the relative extent of the fraud and

is equal to e_c/y_c. In this setting the optimization problem of the firms can be written as

$$\max_{e_b,e_c\geq 0} \quad e_b + e_c - \lambda\Big((1-\frac{e_c}{y_c})\theta(e_b) + \frac{e_c}{y_c}\theta(e_b+e_c)\Big) \tag{4.17}$$

The FOCs are [62]

$$\lambda\theta'(e_b) + \lambda\frac{e_c}{y_c}\Big(\theta'(e_b+e_c) - \theta'(e_b)\Big) = 1 \tag{4.18}$$

$$\lambda\frac{e_c}{y_c}\theta'(e_b+e_c) + \frac{\lambda}{y_c}\Big(\theta(e_b+e_c) - \theta(e_b)\Big) = 1 \tag{4.19}$$

Condition (4.18) indicates that, at the optimum, the marginal gain of one additional unit of e_b is equal to the sum of the expected deterministic and stochastic marginal costs of it, which are $\lambda\theta'(e_b)$ and $\lambda e_c/y_c\big(\theta'(e_b+e_c)-\theta'(e_b)\big)$ respectively. Moreover, according to (4.19), e_c results in two marginal costs: one is marginal cost of the rise in the punishment which is $\lambda e_c/y_c\theta'(e_b+e_c)$ and the other is the marginal expected cost due to the increase in detection probability of e_c. At the optimum these two costs are equal to the marginal benefit of one additional unit of evasion.

If (4.17) is a concave function, then the solutions given by FOC are the optimal choice and they maximize the objective function. However, the concavity of (4.17) near the critical point of FOC depends on y_c, λ and the cost function.

Proposition 4.5. *In the self-assessment enforcement method, if*

$$\Big(\theta'(e_b+e_c)-\theta'(e_b)\Big)^2 < \Big(2\theta'(e_b+e_c) + e_c\theta''(e_b+e_c)\Big)\frac{\theta(e_b+e_c)-\theta(e_b)}{\theta'(e_b)}\theta''(e_b) + 2e_c\theta''(e_b+e_c)\theta'(e_b)$$

then the optimal choices of firm for e_c and e_b are found using (4.18) and (4.19). Otherwise, at the optimum, $e_b = 0$ and e_c is found from $e_c\theta'(e_c) + \theta(e_c) = y_c/\lambda$.

Proof. Appendix 4.I *Q.E.D.*

Proposition 4.6 states the optimal firm's decision for evasion. When the inequality holds, the objective function is concave, otherwise the FOC results in a saddle point where

[62] Here, I implicitly assumed that the amount of tax obligation on business-to-business transactions y_b is large and imposes no restriction on e_b. If the solution of e_b in (4.17) is greater than y_b, it has the corner solution $e_b = y_b$ and the optimization is just w.r.t. e_c.

the objective function has maximum in one direction and minimum in the other. In this case, the best decision of the firm is reallocating the evasion from business transactions to sales to final consumers. The saddle point solution happens when $\theta''(e_b)$ is small and y_c – and as a results e_c – is large. For instance, if $\theta''(0) = 0$, then for small enough e_b the inequality becomes $(\theta'(e_c) - 1)^2 < 2e_c\theta''(e_c)$ which does not hold for large enough e_c. (a simple example is $\theta(e) = e^n + e$). In such situations, the firm is better off by being fully compliant for e_b. Nevertheless, if $\theta''(0)$ is large enough, for some specific polynomial or exponential cost functions, one can show that the saddle point never exists.

The optimality conditions of the firm, either if there is a maximum or a saddle point, intuitively suggest that higher audit rate increases compliance, but more final consumption reallocates evasion in favor of e_c, increasing the overall amount. Formally, we can show that the results of Proposition 4.1 are valid in this context.

Proposition 4.6. *In the self-assessment enforcement method, total evasion is increasing in transactions with final consumers y_c and decreasing in visiting audit rate λ.*

Proof. Appendix 4.J. *Q.E.D.*

Proposition 4.6 is consistent with the evidence of the positive relationship between final consumption and total evasion of a firm suggested by Pomeranz (2013). Therefore, this relationship is independent of the audit method of the government and is valid in the absence of invoice cross-checking.

4.5.1. Optimal audit policy

In the simple self-assessment method without cross-checking, the optimal audit rate is obtained similar to Section 4.4.3. Here, e_b^s also depends on λ_s and consequently the FOC w.r.t. λ_s can be written as

$$\frac{\partial(e_c^s + e_b^s)}{\partial \lambda_s} = -\eta \tag{4.20}$$

In this case, define $e^s = e_b^s + e_c^s$, and from (4.18) and (4.19) assume

$$e^s = G(\frac{y_c^s}{\lambda_s}, y_c^s), \quad G_i = \frac{\partial G(x_1, x_2)}{\partial x_i}, \quad G_{ij} = \frac{\partial^2 G(x_1, x_2)}{\partial x_i \partial x_j} \tag{4.21}$$

then in the Appendix 4.K, I show

$$\lambda_s^2 = \frac{y_c^s}{\eta} G_1, \qquad \frac{\partial \lambda_s}{\partial y_c^s} = \frac{\lambda_s}{y_c^s} \frac{1-\gamma+\zeta}{2-\gamma} \tag{4.22}$$

where $\gamma = \dfrac{G_{11} y_c^s}{G_1 \lambda_s}$ and $\zeta = y_c^s G_{12}/G_1$. In general, finding the boundaries of ζ is cumbersome, however if $0 < \zeta < 1$, the qualitative results in above about the effect of cost convexity on the link between λ_s^* and y_c^s are still valid. In an unreported robustness check, I simulate the optimal condition for λ_s using two simple quadratic and asymptotic functional forms. If $\theta(x) = \frac{1}{2}x^2 + x$ and $\theta(x) = -1 + \dfrac{1}{1-x}$ the qualitative relationship between λ_s and y_c^s are similar to Figure 4.2-a and Figure 4.2-c respectively. For the quadratic cost function the audit rate is monotonically increasing in final consumption, but for the asymptotic cost function the relationship is first increasing and then overturns at a specific point.

By comparing self-assessment method with invoice cross-checking, we can simply show that at the same rate of visiting audit, total evasion is less or equal when invoice cross-checking is available (see Appendix 4.L). However, a more general question for policymakers is the choice between different enforcement methods. Answering this question requires information about the fixed cost of implementing an integrated invoice system as well as variable costs of cross-checking and visiting audit and is not in the scope of this paper. But intuitively, we can say as new technologies reduce the fixed costs of such system the governments are better off by utilizing it.

4.6. Conclusion

The recent developments in tax enforcement literature highlight the importance of third party reporting and information technology on reducing tax fraud and evasion. Research in this field, however, has tended to focus on direct taxes. This paper models how these two instruments can be optimally utilized in the value-added tax which nowadays emerges as the indirect tax of choice and is adopted by more than 150 countries.

The previous literature has shown that final consumption of a commodity increases the intra-firm margin of VAT fraud (misreporting). This effect can be explained by

the invoice system of the VAT which raises the risk of evasion on business-to-business transaction and makes sales to final consumers more attractive for VAT fraud. Using this variation in the detection risk, the model of this paper implies that evasion is an increasing function of sales to final consumers and audit rate. It also shows how invoice reporting limits the evasion on inter-firm transactions by enabling the tax administration to do random cross-checking.

In addition to the new framework for analyzing VAT evasion, this study provides important implications for tax administrations. First, it pinpoints how invoice cross-checking can control VAT evasion and presents how to find its optimal rate, based on sector and commodity characteristics such as number of traders and the level of production. Second, the findings stress the importance of final consumption which is a sectoral characteristic, easy to measure for the policymakers. For instance, the visiting audits should be more focused on downstream sectors and services such as retailer shops, hotels, and restaurants. Third, the paper underlines the significance of risk-based visiting audits and the importance of taxpayers' subjective measures in determining their audit rate. Since large and incorporated enterprises have more transparent accounting system and restricted opportunities for tax evasion, total VAT revenue can diminish if the audit of medium and small taxpayers is neglected.

Although this paper focuses on the VAT, the results can be extended to other types of taxes in which the tax administration faces information with different levels of accuracy. For instance, withholding income taxes are third-party reported and can be checked through information reporting systems. In comparison, self-employed income tax should be investigated by visiting audits based on the characteristics of the taxpayer. Performing predictive analytics to gauge the risk premium of the self-employed person provide information for more efficient audit strategy

Despite the paper adds substantially to our understanding on VAT misreporting, a number of potential limitations need to be mentioned. The present study has only investigated one dimension of VAT fraud and tax administrations must think about other potential evasions while implementing their policies. Keen and Smith (2006) outline a list of VAT frauds and some practical solution. As mentioned above, a notable fraud in developing countries is failure to register which is studied in Hoseini (2014b) in detail. The VAT also has a big potential for cross-border frauds like missing trader (carousel)

fraud in which the importer disappears after the import (for details see Keen and Smith, 2006).

Appendix

Mathematical appendix

4.A. Relaxing the assumptions of independence of e_b and e_c in section 4.4

If e_b and e_c are not independent and jointly determined, we can write the optimization problem of a firm as the combination of (4.2) and (4.4).

$$\max_{e_b,e_c} e_b + e_c - \lambda\Big((1 - \frac{e_c}{y_c})\theta(e_b) + \frac{e_c}{y_c}\theta(e_b + e_c)\Big) - \frac{\mu}{\alpha}e_b\theta(e_b) \tag{4.23}$$

Then the FOC w.r.t e_b becomes

$$\lambda\Big((1 - \frac{e_c}{y_c})\theta'(e_b) + \frac{e_b}{y_c}\theta'(e_b + e_c)\Big) + \frac{\mu}{\alpha}(e_b\theta'(e_b) + \theta(e_b)) = 1 \tag{4.24}$$

which means that the optimal e_b in this case is smaller than in (4.3). The FOC w.r.t. e_c gives

$$e_c\theta'(e_b + e_c) + \theta(e_b + e_c) - \theta(e_b) = \frac{y}{\lambda} \tag{4.25}$$

by implicit differentiation, we can simply find the changes in the optimal level of e_c when the optimal e_b decreases

$$\frac{\partial e_c}{\partial e_b} = -\frac{e_c\theta''(e_b + e_c) + \theta'(e_b + e_c) - \theta'(e_b)}{2\theta'(e_b + e_c) + e_c\theta''(e_b + e_c)} < 0 \tag{4.26}$$

Also because $(1 + \partial e_c/\partial e_b > 0)$ total evasion $(e_b + e_c)$ becomes smaller. Therefore, by relaxing the assumption of independence, we have a reduction in e_b and an increase in e_c in the sense that total evasion declines.

The assumption about the separability of e_b and e_c in optimization is reasonable because in practical circumstances the two problems lead to close outcomes. As mentioned

above, in reality, the variable cost of cross-checking is much less than visiting audit, therefore at the government's optimal μ is much larger than λ when the firm has a high share of business customers. In this case, the second term in the left hand side of (4.24) dominates the first term and the answer of (4.24) for e_b becomes very close to (4.3). On the other hand, when y_b is small and y_c is large, as a similar case to the Proposition 4.5, the objective function in (4.23) becomes non-concave, and as a result, $e_b = 0$ while e_c is determined by (4.5) (see section 4.5 for more details). Therefore, in either case, the two problems lead to very similar results.

4.B. Proof of Proposition 4.1

Define $f(e) = e\theta'(e)+\theta(e)$. Then, because $\theta(e)$ is strictly increasing and convex, we have $f'(e) = e\theta''(e) + 2\theta'(e) > 0$. On the other hand

$$f(e_b) = \frac{\alpha}{\mu}, \qquad f(e_c) = \frac{y_c}{\lambda} \tag{4.27}$$

which results

$$\frac{\partial(e_b + e_c)}{\partial y} = \frac{1}{f'(e_c)\lambda} > 0, \frac{\partial(e_b + e_c)}{\partial \lambda} = \frac{-y}{f'(e_c)\lambda^2} < 0, \frac{\partial(e_b + e_c)}{\partial \mu} = \frac{-\alpha}{f'(e_b)\mu^2} < 0 \tag{4.28}$$

4.C. Proof of Lemma 4.1

Similar to Appendix 4.B, define $f(e) = e\theta'(e) + \theta(e)$, then according to (4.3) and (4.5) $g(x) = f^{-1}(x)$. The first and second derivative of f are equal to

$$f'(e) = 2\theta'(e) + e\theta''(e) \tag{4.29}$$

$$f''(e) = 3\theta''(e) + e\theta'''(e) \tag{4.30}$$

because $\theta' > 1$ (this comes from the assumption that the penalty is always greater than evasion) and $\theta'' > 0$, we have $f'(e) > 2$. Moreover, according to (4.1), $\theta'''(e) \geq 0$, and thus we have $f''(e) > 0$. Now, according to inverse function theorem, because $g(f(x)) = x$,

we have

$$g'(f(x)) = \frac{1}{f'(x)}, \quad g''(f(x)) = -\frac{f''(x)}{(f'(x))^3} \tag{4.31}$$

Therefore $0 < g'(x) \leq \frac{1}{2}$ and $\theta'''(e) \geq 0$ is the sufficient condition for $g''(x) < 0$.

4.D. Proof of Proposition 4.2

When cross-checking is possible, from (4.7), the optimal condition for the audit rate require

$$\frac{\partial e_c^s\big(\lambda_s(y_c^s), y_c^s\big)}{\partial \lambda_s} = -\eta \tag{4.32}$$

Here, firm-level evasion e_c^s is a function of government policy $\lambda_s(y_c^s)$ and exogenous characteristic of industry y_c^s. By implicit differentiation with respect to y_c we can write

$$\frac{\partial^2 e_c^s}{\partial \lambda_s \partial y_c^s} + \frac{\partial^2 e_c^s}{\partial \lambda_s^2}\frac{\partial \lambda_s}{\partial y_c} = 0 \quad \rightarrow \quad \frac{\partial \lambda_s}{\partial y_c^s} = -\frac{\dfrac{\partial^2 e_c^s}{\partial \lambda_s \partial y_c^s}}{\dfrac{\partial^2 e_c^s}{\partial \lambda_s^2}} \tag{4.33}$$

Based on Lemma 4.1, we can write $e_c^s = g(\frac{y_c^s}{\lambda_s})$ where $0 < g'(x) < 1/2$. In this case, we have

$$\frac{\partial e_c^s}{\partial \lambda_s} = -\frac{y_c^s}{\lambda_s^2} g'(\frac{y_c^s}{\lambda_s}) \tag{4.34}$$

By differentiating w.r.t. λ_s and y_c^s, we can write

$$\frac{\partial^2 e_c^s}{\partial \lambda_s \partial y_c^s} = -\frac{1}{\lambda_s^2} g'(\frac{y_c^s}{\lambda_s}) - \frac{y_c^s}{\lambda_s^3} g''(\frac{y_c^s}{\lambda_s}) \tag{4.35}$$

$$\frac{\partial^2 e_c^s}{\partial \lambda_s^2} = \frac{y_c^s}{\lambda_s}\Big(\frac{2}{\lambda_s^2} g'(\frac{y_c^s}{\lambda_s}) + \frac{y_c^s}{\lambda_s^3} g''(\frac{y_c^s}{\lambda_s})\Big) \tag{4.36}$$

Thus, from (4.33), we obtain

$$\frac{\partial \lambda_s}{\partial y_c^s} = \frac{\lambda_s}{y_c^s}\Big(1 - \frac{1}{2 + \dfrac{y_c^s}{\lambda_s} g''(\dfrac{y_c^s}{\lambda_s}) / g'(\dfrac{y_c^s}{\lambda_s})}\Big) \tag{4.37}$$

As a result, by defining $\gamma = -\frac{y_c^s}{\lambda_s} g''(\frac{y_c^s}{\lambda_s})/g'(\frac{y_c^s}{\lambda_s})$, we can obtain

$$\frac{\partial \lambda_s}{\partial y_c^s} = \frac{\lambda_s}{y_c^s} \frac{1-\gamma}{2-\gamma} \tag{4.38}$$

4.E. Proof of Lemma 4.2

From Lemma 4.1, we know g' is always positive and bounded. Therefore, from Proposition 4.4,

$$\lim_{y_c^s \to 0} \lambda^2 = \lim_{y_c^s \to 0} \frac{y_c^s}{\eta} g'(\frac{y_c^s}{\lambda}) = 0 \tag{4.39}$$

Similarly, because $g'(0) = 1/2$ we can write

$$\lim_{y_c^s \to 0} \frac{y_c^s}{\lambda_s} = \lim_{y_c^s \to 0} \frac{\eta \lambda_s}{g'(y_c^s/\lambda_s)} = 0 \tag{4.40}$$

On the other hand, from (4.29), $f''(0)$ and consequently $g''(0)$ are both bounded. Therefore, using (4.40)

$$\lim_{y_c^s \to 0} \gamma = \lim_{x \to 0} \frac{x g''(x)}{g'(x)} = 0 \quad \rightarrow \quad \lim_{y_c^s \to 0^+} \frac{\partial \lambda_s}{\partial y_c^s} = \lim_{y_c^s \to 0^+} \frac{\lambda_s}{y_c^s} \frac{1}{2} = +\infty \tag{4.41}$$

4.F. Finding the derivatives of Figure 4.2

	(a)	(b)	(c)	(d)
$g(x)$	$\frac{na}{2}\left((1+\frac{x}{a})^{\frac{1}{n}} - 1\right)$	$\frac{a}{2}\ln(1+\frac{x}{a})$	$\frac{a}{2}\frac{x}{a+x}$	$\frac{a}{2n}\left(1-(1+\frac{x}{a})^{-n}\right)$
$g'(x)$	$\frac{1}{2}(1+\frac{x}{a})^{\frac{1}{n}-1}$	$\frac{1}{2}(1+\frac{x}{a})^{-1}$	$\frac{a^2}{2(a+x)^2}$	$\frac{1}{2}(1+\frac{x}{a})^{-n-1}$
$g''(x)$	$\frac{1-n}{2an}(1+\frac{x}{a})^{\frac{1}{n}-2}$	$\frac{-1}{2a}(1+\frac{x}{a})^{-2}$	$\frac{-a^2}{(a+x)^3}$	$-\frac{n+1}{2}(1+\frac{x}{a})^{-n-2}$
$\gamma = \frac{-xg''}{g'}$	$\frac{1-n}{n}\frac{x}{a+x}$	$\frac{x}{a+x}$	$\frac{2x}{a+x}$	$\frac{(n+1)x}{a+x}$

4.G. Proof of Proposition 4.3

From (4.10), we can write the cost function of the government as

$$C_i = (M_i^u + M_i^d)e_b^i + \delta \mu_i (\bar{N}_i + M_i^d \frac{e_b^i}{\alpha_i}) \tag{4.42}$$

At the government's optimum, the first and second order conditions are $\frac{\partial C_i}{\partial \mu_i} = 0$ and $\frac{\partial^2 C_i}{\partial \mu_i^2} > 0$. By defining $A_i = \frac{M_i^u + M_i^d}{\delta \bar{N}_i}$ and $B_i = \frac{M_i^d}{\alpha_i \bar{N}_i}$ we can write

$$\frac{\partial C_i}{\partial \mu_i} = \frac{1}{\bar{N}_i}\Big((A_i + B_i\mu_i)\frac{\partial e_b^i}{\partial \mu_i} + B_i e_b^i + 1\Big) \tag{4.43}$$

$$\frac{\partial^2 C_i}{\partial \mu_i^2} = \frac{1}{\bar{N}_i}\Big((A_i + B_i\mu_i)\frac{\partial^2 e_b^i}{\partial \mu_i^2} + 2B_i\frac{\partial e_b^i}{\partial \mu_i}\Big) \tag{4.44}$$

Therefore, at the optimum we have

$$(A_i + B_i\mu_i^*)\frac{\partial e_b^i(\mu_i^*)}{\partial \mu_i^*} + B_i e_b^i(\mu_i^*) + 1 = 0 \tag{4.45}$$

where μ_i^* is the optimal decision of the government. Now, we can find the derivative of μ_i^* with respect to A_i

$$\frac{\partial e_b^i}{\partial \mu_i^*} + \frac{\partial \mu_i^*}{\partial A_i}\Big((A_i + B_i\mu_i)\frac{\partial^2 e_b^i}{\partial \mu_i^2} + 2B_i\frac{\partial e_b^i}{\partial \mu_i^*}\Big) = 0 \tag{4.46}$$

which means

$$\frac{\partial \mu_i^*}{\partial A_i} = \frac{-\partial e_b^i/\partial \mu_i^*}{(A_i + B_i\mu_i)\frac{\partial^2 e_b^i}{\partial \mu_i^2} + 2B_i\frac{\partial e_b^i}{\partial \mu_i^*}} \tag{4.47}$$

From the second order condition, the denominator of (4.47) is positive, thus because $\partial e_b^i/\partial \mu_i < 0$, μ_i is increasing in A_i. Now, if $\delta << (1 + \frac{M_i^u + M_i^d}{\bar{N}_i})\alpha_i$, then B_i is negligible relative to A_i and we can approximate the first order condition (4.45) as $\dfrac{\partial e_b^i}{\partial \mu_i} = -\dfrac{1}{A_i}$. On the other hand, when δ – cost of an additional cross-checking – is small, the level of evasion is small too and because $\theta' = 1$ near zero (see (4.17)) we can approximate $\theta(e) \approx e$. Then, (4.3) turns to

$$\frac{\alpha_i}{\mu_i} \approx 2e_b^i \quad \rightarrow \quad \frac{\partial e_b^i}{\partial \mu_i} \approx \frac{-\alpha_i}{2\mu_i^2} \tag{4.48}$$

As a result, $\dfrac{\alpha_i A_i}{2}$ is a good approximation of μ_i^2 when cost of cross-checking is low. But at most μ can be equal to one and if δ is very small maybe cross-checking of all invoices are optimal.

4.H. Proof of Proposition 4.4

If a downstream activity uses J intermediate goods as input and over-reports $k_1, \ldots, k_J$ invoices of each input, the probability of not being detected for a firm in that activity is $(1 - k_1/N_1)^{n_1} \ldots (1 - k_J/N_J)^{n_J}$ and its approximation will be $1 - \sum_{i=1}^{J} \mu_i k_i$. As a consequence, the optimization problem of a single firm is

$$\max_{e_{bi}^d \geq 0} \quad \sum_{i=1}^{J} e_{bi}^d - \sum_{i=1}^{J} e_{bi}^d \frac{\mu_i}{\alpha_i} \theta\big(\sum_{i=1}^{J} e_{bi}^d\big) \tag{4.49}$$

The FOC for intermediate good j becomes

$$1 - \frac{\mu_j}{\alpha_j} \theta\big(\sum_{i=1}^{J} e_{bi}^d\big) - \sum_{i=1}^{J} e_{bi}^d \frac{\mu_i}{\alpha_i} \theta'\big(\sum_{i=1}^{J} e_{bi}^d\big) \leq 0 \tag{4.50}$$

To have positive e_{b1}^d to e_{bJ}^d all FOC must hold and $\forall i, j : \dfrac{\mu_i}{\alpha_i} = \dfrac{\mu_j}{\alpha_j}$. Otherwise, at least one e_{bi}^d is zero. Because each upstream activity produces just one intermediate good, its only misreporting possibility is on one commodity, but, d has different options for misreporting. Similar to the case with two inputs, if $e_{bi} = \max\{e_{b1}, \ldots, e_{bJ}\}$, the downstream firm d is better off by over-reporting just the invoices of commodity i. However, this makes the tax administration to increase the cross-checking of i, because he knows higher number of firms are misreporting on that commodity. Then, increasing the cross-checking rate in i may make risk of other commodities smaller for d and the downstream firms switch their misreporting to other options.

To find the equilibrium, consider the administration makes the downstream firms indifferent between j commodities, that comprise the elements of a set P_j, for over-reporting their invoices and the rest of $J - j$ commodities build up the set Q_j. If downstream firms have no incentive to evade on $q \in Q_j$, on average, the share $1/j$ of them over-report the transactions of $p \in P_j$. In this case, to find the total number of misreporting firms, the tax administration adds this number to the number of the upstream firms of each intermediate good – that have just one option for misreporting. Therefore, at the government's optimum, if $\forall q \in Q_j$ downstream evasion is zero, then $\forall p \in P_j$, we have $e_{bp} = h_p(M_p^u + \frac{1}{j} M^d)$. On the other hand, P_j and Q_j characterize a Nash equilibrium, if downstream firms have no incentive to switch their over-reporting

from a commodity $p \in P_j$ to another $q \in Q_j$. This means that $\forall q \in Q_j$ and $\forall p \in P_j$, we must have

$$h_p(M_p^u + \frac{1}{j}M^d) > h_q(M_q^u) \tag{4.51}$$

For $Q_j = \emptyset$ we set $j = J$ and the Nash equilibrium always exists, but (4.51) can hold for $j < J$ too. The optimal equilibrium for the administration is the one that imposes the lowest number of constraints to its optimization problem

$$\min_{\mu_i} \quad \sum_{i=1}^{J} M_i^u e_{bi} + M^d \max[e_{b1}, \ldots, e_{bJ}] + \delta \sum_{i=1}^{J} \mu_i N_i \tag{4.52}$$

Because each additional unit of j adds a new binding constraint in form of $\frac{\mu_i}{\alpha_i} = \frac{\mu_j}{\alpha_j}$ to the optimization problem, the equilibrium with the minimum j is the optimal choice of the administration. After finding the minimum j that holds (4.51), the optimal policy for all $p \in P_j$ is making the downstream firms indifferent for misreporting which means

$$\forall p_1, p_2 \in P_j : \quad \frac{\mu_{p_1}}{\alpha_{p_1}} = \frac{\mu_{p_2}}{\alpha_{p_2}}$$

On the other hand, for all $q \in Q_j$, the downstream firms never decide to misreport commodity q and the optimal policy is determined independently from Proposition 4.3 considering no downstream firm over-reports invoices of commodity q.

4.I. Proof of Proposition 4.5

By defining $e = e_b + e_c$, $\theta_e = \theta(e_b + e_c)$ and $\theta_b = \theta(e_b)$ we can obtain an alternative formulation for firm's optimization problem

$$\max_{e \geq e_b \geq 0} E(e, e_b) = e - \lambda\Big(\theta_b + \frac{e - e_b}{y_c}(\theta_e - \theta_b)\Big) \tag{4.53}$$

Therefore, we have

$$E_e = \frac{\partial E}{\partial e} = 1 - \frac{\lambda}{y_c}\Big(\theta_e - \theta_b + (e - e_b)\theta'_e\Big) \tag{4.54}$$

$$E_b = \frac{\partial E}{\partial e_b} = -\frac{\lambda}{y_c}\Big((y_c - e + e_b)\theta'_b + \theta_b - \theta_e\Big) \tag{4.55}$$

$$E_{ee} = \frac{\partial^2 E}{\partial e^2} = -\frac{\lambda}{y_c}\Big(2\theta'_e + (e - e_b)\theta''_e\Big) \tag{4.56}$$

$$E_{bb} = \frac{\partial^2 E}{\partial e_b^2} = -\frac{\lambda}{y_c}\Big(2\theta'_b + (y_c - e + e_b)\theta''_b\Big) \tag{4.57}$$

$$E_{eb} = \frac{\partial^2 E}{\partial e \partial e_b} = \frac{\lambda}{y_c}\Big(\theta'_e + \theta'_b\Big) \tag{4.58}$$

Then, the FOC are $E_b = E_e = 0$, equivalent to (4.18) and (4.19). The second order condition of partial derivatives to have a concave objective function requires

$$\Delta = E_{ee}E_{bb} - E_{eb}^2 > 0 \tag{4.59}$$

Given the partial derivatives, the second order condition $\Delta > 0$ is equivalent to

$$(2\theta'_e + e_c\theta''_e)\big(2\theta'_b + (y_c - e_c)\theta''_b\big) - (\theta'_e + \theta'_b)^2 > 0 \tag{4.60}$$

If we expand (4.60), we must show

$$2\theta'_e\theta'_b + 2\theta'_e(y_c - e_c)\theta''_b + 2e_c\theta''_e\theta'_b + e_c(y_c - e_c)\theta''_e\theta''_b - \theta'^2_e - \theta'^2_b > 0 \tag{4.61}$$

θ is a convex and increasing function, thus $\theta' > 0$, $\theta'' > 0$ and $\theta'_e > \theta'_b$. In addition, from the FOC (4.55), $y_c - e_c = (\theta_e - \theta_b)/\theta'_b$. Therefore (4.60) is equivalent to

$$(\theta'_e - \theta'_b)^2 < (2\theta'_e + e_c\theta''_e)\theta''_b\frac{\theta_e - \theta_b}{\theta'_b} + 2e_c\theta''_e\theta'_b \tag{4.62}$$

If $\Delta < 0$, then the critical point is a saddle point and we have a corner solution. Since both e_b and e are positive and $e \geq e_b$, the corner solution that maximizes the objective

function is either $e_b = 0$ or $e_c = 0$ ($e_b = e$). In each case, we have

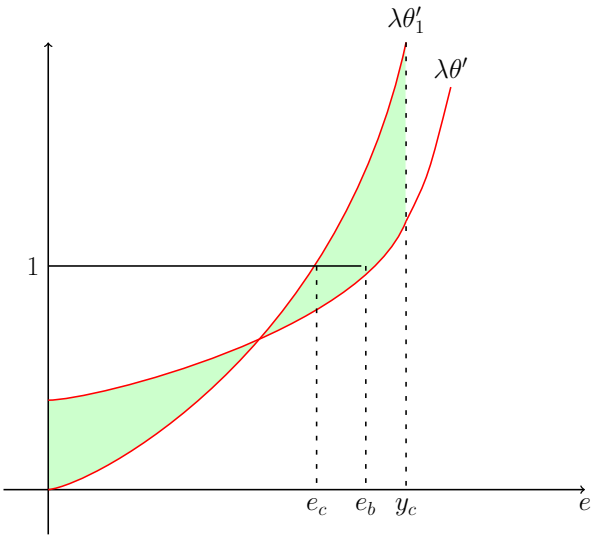
$$e_b = 0: \quad E(e,0) = e - \lambda\theta(e) \tag{4.63}$$

$$e_c = 0: \quad E(e,e) = e - \lambda\frac{e}{y_c}\theta(e), \qquad 0 \le e \le y_c \tag{4.64}$$

By defining $\theta_1(e) = e\theta(e)/y_c$, we can find the optimal decision in each case using FOC

$$e_b = 0: \quad \theta'(e) = \frac{1}{\lambda}, \qquad e_c = 0: \quad \theta_1'(e) = \frac{\theta(e) + e\theta'(e)}{y_c} = \frac{1}{\lambda} \tag{4.65}$$

To graphically see $E(e,e) < E(e,0)$, the below figure illustrates the optimal choice of e_b and e_c when the other is zero. Because $\theta(y_c) = \theta_1(y_c)$ the two shaded areas are equal. The payoff of firm from each corner solution is the area below the horizontal line at 1 and the corresponding curve. It is clearly seen that the area is bigger in case of θ_1', and therefore e_c leads to higher payoff



Hence, when $\Delta < 0$, the optimal choice of firm is $e_b = 0$ and e_c is found similar to (4.5).

4.J. Proof of Proposition 4.6

If $\Delta < 0$, then the critical point given by FOC is a saddle point which means at the optimum $e_b = 0$ and e_c is found from (4.5). In this case, the proof is similar to Proposition 4.1.

If the critical point is not a saddle point and (4.62) holds, I use the notations of

Appendix 4.I and in addition I define

$$\dot{e}_b = \frac{\partial e_b}{\partial \lambda}, \quad \dot{e}_c = \frac{\partial e_c}{\partial \lambda}, \quad \dot{e} = \frac{\partial (e_c + e_b)}{\partial \lambda}$$

At the optimum, we have

$$E_b = 0, \quad E_e = 0 \tag{4.66}$$

Now, if we differentiate the FOCs with respect to λ

$$E_{bb}\dot{e}_b + E_{be}\dot{e} + \frac{\partial E_b}{\partial \lambda} = 0 \tag{4.67}$$

$$E_{eb}\dot{e}_b + E_{ee}\dot{e} + \frac{\partial E_e}{\partial \lambda} = 0 \tag{4.68}$$

from (4.55) and (4.66), we obtain that $\partial E_b/\partial\lambda = 0$ and $\partial E_e/\partial\lambda = -1/\lambda$, therefore from (4.67), we can write

$$\dot{e}_b = -\frac{E_{be}}{E_{bb}}\dot{e} \tag{4.69}$$

By substituting (4.69) in (4.68), we obtain

$$\dot{e} = \frac{E_{bb}}{\lambda \Delta} \tag{4.70}$$

From (4.57), it turns out that $E_{bb} < 0$. Therefore, because $\Delta > 0$, we have $\dot{e} < 0$. By expanding (4.70), we can write the derivative of total evasion w.r.t. λ as

$$\dot{e}_b + \dot{e}_c = \frac{y_c}{\lambda^2} \frac{2\theta_b' + (y_c - e_c)\theta_b''}{(\theta_e' + \theta_b')^2 - (2\theta_e' + e_c\theta_e'')(2\theta_b' + (y_c - e_c)\theta_b'')} \tag{4.71}$$

To find the effect of y_c on evasion, define

$$\tilde{e}_b = \frac{\partial e_b}{\partial y_c}, \quad \tilde{e}_c = \frac{\partial e_c}{\partial y_c}, \quad \tilde{e} = \frac{\partial (e_c + e_b)}{\partial y_c}$$

From (4.54) and (4.55) we obtain $\partial E_b/\partial y_c = -\lambda\theta_b'/y_c$ and $\partial E_e/\partial y_c = 1/y_c$. Then, using similar calculation to (4.69) and (4.70), we can write

$$\tilde{e}_b = -\frac{E_{be}}{E_{bb}}\tilde{e} + \frac{\lambda\theta_b'}{y_c E_{bb}} \tag{4.72}$$

$$\tilde{e} = \frac{-1}{y_c \Delta}(E_{bb} + \lambda\theta_b') \tag{4.73}$$

By expanding (4.73), we can write

$$(\tilde{e}_b + \tilde{e}_c)\Big(2\theta'_{b+c} + e_c\theta''_{b+c} - \frac{(\theta'_{b+c} + \theta'_b)^2}{2\theta'_b + (y_c - e_c)\theta''_b}\Big) = \frac{1}{\lambda} - \frac{\theta'_b}{2\theta'_b + (y_c - e_c)\theta''_b} \tag{4.74}$$

The left hand side multiplier is equal to Δ/E_{bb} and it is positive. The right hand side is positive if

$$\lambda < 2 + (y_c - e_c)\theta''_b/\theta'_b \tag{4.75}$$

and it always holds because $0 < \lambda < 1$. Therefore, when $\Delta > 0$ we have $\tilde{e} > 0$.

4.K. The optimal audit rate in self-assessment system

Similar to above proof in section Appendix 4.D, in the absence of cross-checking, when e^s_b and e^s_c are jointly determined, define $e^s = e^s_b + e^s_c$, and from (4.18) and (4.19) assume

$$e^s = G(\frac{y^s_c}{\lambda_s}, y^s_c), \quad G_i = \frac{\partial G(x_1, x_2)}{\partial x_i}, \quad G_{ij} = \frac{\partial^2 G(x_1, x_2)}{\partial x_i \partial x_j} \tag{4.76}$$

then the condition (4.20) states that the optimal λ_s satisfies

$$\frac{\partial G}{\partial \lambda_s} = -\frac{y^s_c}{\lambda^2_s} G_1 = -\eta \tag{4.77}$$

Moreover, we can write

$$\frac{\partial^2 G}{\partial \lambda_s \partial y^s_c} = -\frac{1}{\lambda^2} G_1 - \frac{y^s_c}{\lambda^3} G_{11} - \frac{y^s_c}{\lambda^2} G_{12} \tag{4.78}$$

$$\frac{\partial^2 G}{\partial \lambda^2_s} = \frac{y^s_c}{\lambda_s}\Big(\frac{2}{\lambda^2} G_1 + \frac{y^s_c}{\lambda^3} G_{11}\Big) \tag{4.79}$$

Therefore, from (4.33) the derivative of λ_s w.r.t. y^s_c is obtained as

$$\frac{\partial \lambda_s}{\partial y^s_c} = \frac{\lambda_s}{y^s_c} \frac{1 - \gamma + \zeta}{2 - \gamma} \tag{4.80}$$

where $\gamma = \dfrac{y^s_c}{\lambda_s} \dfrac{G_{11}}{G_1}$ and $\zeta = y^s_c \dfrac{G_{12}}{G_1}$.

4.L. Comparing self-assessment and invoice cross-checking methods

The equivalent problem for self-assessment when cross-checking is possible is studied in Appendix 4.A. In both cases the FOC w.r.t. e_c yields to (4.25). Therefore, from (4.26) we can write

$$\frac{\partial(e_b + e_c)}{\partial e_b} = 1 + \frac{\partial e_b}{\partial e_c} = \frac{\theta'(e_b + e_c) + \theta'(e_b)}{2\theta'(e_b + e_c) + e_c\theta''(e_b + e_c)} > 0 \tag{4.81}$$

One the other hand, the FOC w.r.t. e_b in the two problems yield to (4.24) and (4.18). Because of additional cost due to cross-checking, (4.24) gives smaller e_b than (4.18). Hence, using (4.81) total evasion is also smaller when cross-checking is available.

Chapter 5

Value-Added Tax and Shadow economy: the Role of Inter-sectoral Linkages

5.1. Introduction

According to the public economics literature, the efficiency of the value-added tax (VAT) is tightly linked to the size of informal sector in the economy. Many developing countries, suffering from large shadow economy,[63] has adopted the VAT with a big challenge on its optimal design and enforcement. This paper looks into the VAT evasion at the extensive margin (non-registration of traders) and analyses the optimal enforcement strategy to reduce the size of underground activities. Intrinsically, the VAT has a self-enforcing feature in the sense that each formal enterprise wants to get VAT credit on its inputs and, as a result, informs the tax administration about its purchases. This incentive provides verifiable records on transactions between firms – but not with final consumers – and helps the administration to find nonregistered firms (Keen and Smith, 2006; Pomeranz, 2013). This feature provides an additional instrument for tax authorities to cross-check the two pieces of information, in addition to typical investigations by auditors. By modeling this feature in a multi-sector economy with interlinked industries, this paper analyses the optimal enforcement across activities to alleviate shadow economy problem in a VAT system. The great advantage of this approach is providing a straightforward rule for tax administration to distinguish between different activities for allocating enforcement type and spending.

As a first shot, I present a stylized fact from the Indian economy. The government of India adopted VAT on services in 2003. Figure 5.1 illustrates tax registration rate

[63] According to Schneider et al. (2011), on average the size of shadow economy in developing countries is around 40% and ranges up to 70%.

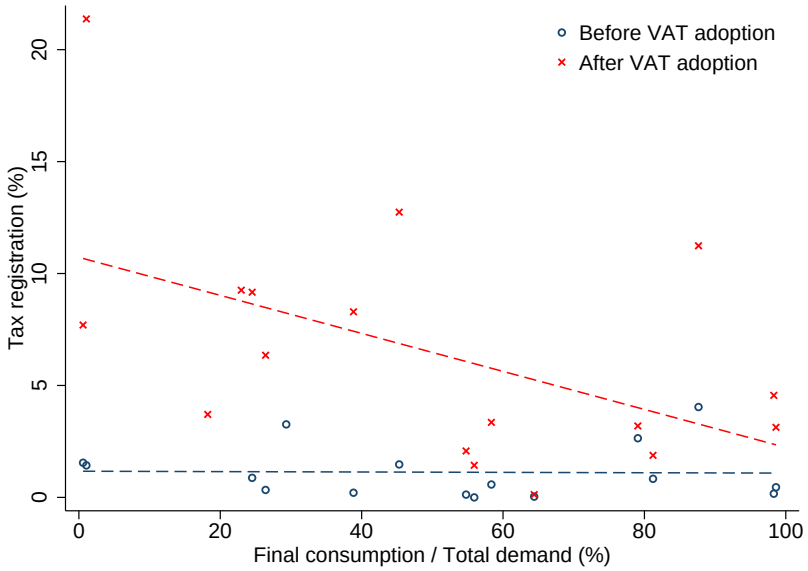


Figure 5.1: Tax registration rate and final consumption in different service sectors of India before and after VAT adoption in 2003. The ∘ and × symbols represents the registration rates in 2001 and 2006 respectively. The registration rate is for a representative sample of small enterprises with less than 10 employees. See section 5.5 for data details.

among firms in different services versus the share of final consumption in each service.[64] The ∘ and × symbols represents the percentage of enterprises registered for tax in 2001 and 2006 respectively. It can be seen that there is negative relationship between tax registration rate and the share of final consumption in an activity after – but not before – VAT adoption. In other words, this figure suggests the activities with more business-to-business (less business-to-household) transactions have more tax compliance rate when the VAT is in place. The aim of this paper is analyzing why this pattern exist in the VAT by modeling an interlinked economy, and how it changes the optimal enforcement policy across different activities. In addition, it empirically tests the existence of VAT self-enforcement using the Indian service sector data.

In order to analyze the optimal VAT enforcement policy, I start with an input-output (I-O) model which lays the groundwork for defining *forward* and *backward* linkages. These two sectoral indices respectively reflect the flow of intermediate products *to* and *from* the other industries. Next, I explain cross-checking invoices – matching each input credit claim with a corresponding output tax payment – as a distinct enforcement tool in a VAT system and elaborate how tax administration can utilize it to reduce informality. The results suggest that in each industry, the optimal enforcement is positively

[64]The sample data and the sources are described in the section 5.5.

associated with the backward linkages. Furthermore, in backwardly linked industries, tax administrators should focus on revealing internal information of firms by visiting audits and policies like rewards to whistle blowers. In comparison, in the forwardly linked industries the administration should concentrate on arm-length transactions and cross-checking the invoices with the corresponding input credit claims. In the next step, I examine the existence of the self-enforcing feature of the VAT using tow Indian service sector surveys. I take the VAT adoption on services in 2003 as a natural experiment to compare formality before and after it. The results confirm a positive relationship between activity's forward linkages and formal production share after VAT adoption. Such a relationship is not significant for backward linkages.

The findings of the paper contribute to the tax enforcement literature in several ways and have important policy implications specially for developing countries. This is the first paper that uses input-output framework and the concept of forward and backward linkages to look into the tax enforcement problem. Despite the long-established literature on linkage analysis (which dates back to Rasmussen, 1957), its applications in the economic theory is, for the most part, neglected by theorists. This paper utilizes two well-known concepts of linkages analysis in economic modeling and provides measurable industry-level indices for tax administrators to improve VAT enforcement. It shows how with the same level of effort, tax administration can raise VAT revenue by reallocating the level of enforcement among different activities. Backwardly linked industries that are usually at the bottom of production chains should be the first objective for revealing within-firms information by policies like higher audit rate and reward to whistle-blowers. On the other hand, forwardly linked industries, normally located up the production chain, are the best objectives for invoice cross-checking. A great advantage of the linkages indices is their measurability using I-O tables which eases policy implementation in developing countries. In addition, this paper test the existence of self-enforcing feature of the VAT using a natural experiment in the second most-populated country of the world and finds that the VAT adoption is the underlying reason of the positive effect of forward linkages on tax compliance.

5.1.1. Literature review

Recent studies in public economics model the VAT from different angles. Some, on a broader level, compare the efficiency of the VAT to other indirect taxes like tariff and RST (Boadway and Sato, 2009; Emran and Stiglitz, 2005; Keen, 2008), while others concentrate on its optimal design of rates and exemption levels (Keen and Mintz, 2004; Piggott and Whalley, 2001). To date, however, little research has focused on VAT evasion and avoidance by theorists. There are a number of ways of fraud in the VAT which are discussed in details in Keen and Smith (2006) and the focus of this paper is on the shadow economy fraud which refers to non-registration of traders. In high-income countries, underground activities comprise relatively smaller share in total VAT evasion rather than the frauds by registered traders such as misreporting the transactions, carousel fraud, etc.[65] In comparison, failure to register for the VAT is a common fraud in low-income countries. For instance, the below estimations show, during 2001-2006 in India, only 18 percent of production in "unorganized service sectors" is produced by tax registered firms. Therefore, the results of this paper is more applicable for low-income countries, hurting from the shadow economy fraud.

The nexus of indirect tax policy and the size of the informal sector has been the subject of a number of contributions in the literature. Emran and Stiglitz (2005) study the inefficiencies of the VAT compared to tariff in an economy with a large informal sector. They argue that, given the size of informal market in developing countries, the standard policies to replace trade taxes with the VAT reduce welfare under plausible assumptions. Their argument is based on the assumption that the informal enterprises may escape from paying the VAT, but cannot avoid border taxes. The follow-up by Keen (2008) argues that Emran and Stiglitz (2005) undervalue the power of the VAT to absorb revenue under large informal sector by neglecting the intermediate inputs in the production side of the economy. He explores the function of the VAT as a tax on the inputs of the informal sector because non-compliant firms are unable to recover the VAT paid on the inputs bought from compliant firms. However, he mentions that the existence of a large informal sector can invalidate the conventional policies on replacing trade taxes

[65] As and example, in 2001-02, VAT revenue loss in UK is estimated to be %10.2–%14.2 of total VAT revenue, among which %0.6–%0.7 is due to non-registration for the VAT (Keen and Smith, 2006). The misreporting fraud –understating sales and overstating purchases– by registered firms is studied in Hoseini (2014a) in details.

with the VAT and the ability of the tax administration to reduce informality can shift the optimal policy toward the VAT. For modeling informality, this paper uses a same framework as Emran and Stiglitz (2005) and Keen (2008), in the sense that the number of formal and informal firms are given, while informal producers set their production level according to the tax rate and enforcement policies. The empirical evidence in this field also emphasizes the importance of an efficient design and administration of the VAT in developing economies. Baunsgaard and Keen (2010) show in a panel study of 117 countries that recovering the loss in trade taxes with the VAT has failed in low income countries in spite of its success in the developed world. Although this paper does not directly discuss trade taxes, it gives straightforward policy recommendations to solve the enforcement inefficiencies of the VAT for overcoming the problem of revenue loss in replacing border tax with domestic sources.

On the tax enforcement side, the literature points to the close connection between tax administration efficiency and information gaps (for a survey see Slemrod and Yitzhaki, 2002) and a growing branch studies the role of firms in facilitating enforcement. Kopczuk and Slemrod (2006) argue that verifiable paper trails from arms-length transactions and aggregation of information within firms considerably alleviate tax enforcement problems. They show, in a simple model, that the VAT has a great administrative advantage to the retail sales tax (RST) that possesses no business-to-business transactions records. In this paper, I also emphasize the importance of arms-length transaction records as a type of third party reporting for the VAT administration and argue that as long as the share of arms-length transactions is not the same across activities (like Figure 5.1), the optimal enforcement policy also differs among them. Then, I link this variation with inter-sectoral linkages and find the optimal policies by taking such differences into consideration.

Another role of firms in facilitating tax enforcement is providing information through third party reporting of employees. Kleven et al. (2009) use this function of the firm to answer why governments can tax so much nowadays. They indicate that the standard tax evasion models like Allingham and Sandmo (1972) fail to explain the high level of tax compliance in the developed world, given the low enforcement by governments. In order to address this issue, they argue that generally a firm can collude with its employees to un- or under-report its income to the tax administration, but in practice such collusion

is fragile when the firm is big with a large number of employees. One reason for breaking the collusion is random shocks due to unintentional mistakes or conflicts within firm, another can be rational whistle blowing due to the rewards by the government. In this paper, I add this type of third-party reporting to the model by linking the probability of detection of informal firms to the number of their employees. In comparison to Kopczuk and Slemrod (2006) who emphasize the importance of inter-firm information networks, Kleven et al. (2009) focus on intra-firm information sources. In a general context both of them reflect the importance of third party reporting in facilitating tax enforcement, one by trade partners, the other by workers. The key advantage of the VAT is enabling the tax administration to use both of these information sources for detecting fraud and evasion. Therefore, this paper presents the optimal policy for the VAT enforcement by putting together two instruments for revealing firms' true information: checking arms-length transactions and third party reporting by workers.

In a closer look into VAT evasion, De Paula and Scheinkman (2010) address the decision of firms for being formal in a VAT setting. Their model assumes that firms have different managerial abilities and above a cut-off production informal firms are detected by the tax administration. In this setting, they show that low ability firms remain small and informal, but high ability firms become formal to produce with no restriction. Then, they indicate that under the VAT, (in)formal firms trade with their (in)formal peers in the production chain not with the other type. Their results are based on variation in managerial ability and are silent about the industry in which the firm works. In comparison, this paper looks from a different point of view by assuming the same ability for all managers, but different industries in which a firm can work. This approach has a great practical advantage since instead of relying on unobservable variables like ability and entrepreneurship, it addresses the optimal tax administration's policy for each industry based on measurable characteristics like linkages and connections with the other industries.

In the empirical literature, most of the papers on VAT evasion are limited to cross country or panel estimations, testing whether VAT adoption leads to more revenue (Ebrill et al., 2001; Keen and Lockwood, 2010). One notable exception is Pomeranz (2013) who provides micro evidence for the self-enforcing feature of the VAT using two randomized experiments in Chile. The first experiment implies that in the absence of the VAT

paper trail like sales to final consumers, firms respond more strongly to a rise in audit letter message, suggesting the presence of the VAT paper trail has a preventive effect on evasion. This is consistent with my finding that forwardly linked industries, having more paper trail, are more compliant. In the "spillover experiment", the Chilean tax authority randomly sends an announcement of audit for half of the suspected firms and collects information about their trading partners. The results show that the announcement reduces the evasion among the suppliers of the treated firms, verifying the presence of self-enforcement from clients to suppliers. In this paper, in addition to modeling the findings of Pomeranz (2013) theoretically, I find the optimal enforcement policy for each industry based on its linkages with the other industries.

Before proceeding, some caveats are in place. The model of this paper ignores the VAT frauds by registered traders such as misreported sales, multiple rates and misclassification of commodities and missing trader (carousel) fraud which are mentioned in Keen and Smith (2006). It also dismisses benefits of formality which are stressed in the literature like access to finance and legal institutions. Considering these issues should not change the qualitative implications of the model, which just focus on reducing informality in the VAT from the view point of inter-industry connections.

This paper is organized as follows. Section 5.2, constructs the basic model and introduces forward and backward linkages. Next, in section 5.3, I explain tax administration's enforcement instruments in the VAT, and in section 5.4, I find the market equilibrium and the optimal enforcement policy in each industry. In the section 5.5, I empirically examine the existence of the self-enforcing feature and finally, I conclude.

5.2. The Basic Model

Consider an economy with $n = 1, \ldots, N$ competitive industries each producing a homogeneous and unique product x_n. Each product can either be used as an input of other industries or be consumed by final consumers. The production technology of all

industries is a Cobb-Douglas function as follows[66]

$$x_n = v_n^{\alpha_n} \prod_{k=1}^{N} x_{kn}^{\alpha_{kn}}, \quad \sum_{k=1}^{N} \alpha_{kn} + \alpha_n = 1 \tag{5.1}$$

where x_n is the production of industry n, v_n is the value-adding input,[67] and x_{kn} is the amount of product k as an intermediate input of industry n. The market of value-adding input v is also competitive and I take its price as the numeraire. In equilibrium, p_n is the price by which the product n is traded. In this setting, the profit of a representative firm in industry n becomes

$$\pi_n = p_n x_n - \sum_{k=1}^{N} p_k x_{kn} - v_n \tag{5.2}$$

Under perfect competition, the first order conditions result

$$x_{kn} = \frac{\alpha_{kn} p_n}{p_k} x_n, \qquad v_n = \alpha_n p_n x_n \tag{5.3}$$

we can find the prices by substituting (5.3) in the production function (5.1)

$$p_n = (\frac{1}{\alpha_n})^{\alpha_n} \prod_{m=1}^{N} (\frac{p_m}{\alpha_{mn}})^{\alpha_{mn}} \tag{5.4}$$

At the demand side, each product is used either as an intermediate input or for final consumption. By denoting the final consumption of the product of industry n as x_{0n}, one can write total representative demand of industry n as

$$x_n = x_{0n} + \sum_{k=1}^{N} x_{nk} \tag{5.5}$$

where the first term shows the final demand and the second term is the sum of all intermediate demands of product n. Assume that final consumers are homogeneous with a net income w and a Cobb-Douglas utility in the sense that the utility maximization

[66] The results of the model is robust to assuming a general CES function and are available upon request.

[67] For simplicity, v_n represents all inputs other than intermediate goods. More general assumption would be adding labor and other non-intermediate factors of value-added in the production function and using $l_n^{a_1} k_n^{a_2} \dots$ instead of v_n, but it does not change the qualitative results of the model.

problem for a representative final consumer is

$$\max u(x_{01},\ldots,x_{0N}) = \prod_{i=1}^{N} x_{0i}^{\beta_i}, \quad \sum_{i=1}^{N}\beta_i = 1, \qquad \text{s.t. } \sum_{i=1}^{N} p_i x_{0i} = w \tag{5.6}$$

where β_i determines per-unit fraction of the consumers' incomes used in purchasing good i. The solution of (5.6) gives the final demand of each product n as $x_{0n} = w\beta_n/p_n$. Therefore, in equilibrium, the representative production and demand are equal in all industries and by substituting x_{0n} and x_{nk} from (5.3) in (5.5), we can write

$$x_n = w\beta_n/p_n + \sum_{k=1}^{N}\frac{\alpha_{nk}p_k}{p_n}x_k, \quad \Rightarrow \quad p_n x_n = w\beta_n + \sum_{k=1}^{N}\alpha_{nk}p_k x_k \tag{5.7}$$

Define $y_n = p_n x_n$ as representative value of production in industry n. Then,

$$y_n = w\beta_n + \sum_{k=1}^{N}\alpha_{nk}y_k \tag{5.8}$$

(5.8) expresses that the total value of production of commodity n is equal to the value that is bought by final consumers who spend a fraction β_n of their total wealth on it plus the intermediate demand of other industries. The share of the value of product n consumed as the input of k in total value of output of industry k is α_{nk}. To find the equilibrium value of production in each industry, I employ vectors Y and B with dimensionality $N \times 1$ containing y_n and β_n respectively, together with an $N \times N$ matrix $\boldsymbol{A}$ built by α_{kn} as elements.[68] Then, (5.8) is written as

$$Y = wB + \boldsymbol{A}Y, \quad \Rightarrow \quad Y = (\boldsymbol{I} - \boldsymbol{A})^{-1}wB. \tag{5.9}$$

Since $wB > 0$, (5.9) imposes that $\boldsymbol{A}Y < Y$. In this case, $(\boldsymbol{I} - \boldsymbol{A})^{-1}$ in (5.9) always exists and there is unique equilibrium Y:

Theorem 1. *Matrix $\boldsymbol{A}$ is square and nonnegative, then*

$$\boldsymbol{A}V < V \Leftrightarrow (\boldsymbol{I} - \boldsymbol{A})^{-1} = \sum_{m=0}^{\infty}\boldsymbol{A}^m$$

[68] Throughout the paper, lower-case letters like x denotes a scalar, upper-case like X a vector, and bold upper-case like $\boldsymbol{X}$ a matrix.

for some column vector $V \geq 0$.

Proof. See Ten Raa (2006), theorem 2.2. *Q.E.D.*

Theorem 1 states that if the technology matrix $\boldsymbol{A}$ is capable of producing a positive output Y such that $\boldsymbol{A}Y < Y$, the inverse exists. Then, $Y = (\boldsymbol{I}+\boldsymbol{A}+\boldsymbol{A}^2+\dots)wB$, where $\boldsymbol{A}^m wB$ is the indirect effect of final consumption wB in rising the value of production of Y through m intermediate goods in between.

5.2.1. Forward and backward linkages

The above model provides a simple framework to study the effect of inter-sectoral linkages. I introduce two types of linkages indices in this regard: *forward linkage* which quantifies the value of products that is *sold to* the other industries (not final consumers), and *backward linkage*, which indexes the value of products *purchased from* other industries. In order to gauge these two properties, Rasmussen (1957) suggests the row and column sum of the inverse of the industry×industry table of the economy which is computed using the corresponding input-output transaction tables. I use the method proposed by Rasmussen (1957) and define the following formulas for forward and backward linkages vectors:

$$\text{Forward linkage:} \quad FL = (\boldsymbol{I} - \boldsymbol{Y}^{-1}\boldsymbol{A}\boldsymbol{Y})^{-1}J \tag{5.10}$$

$$\text{Backward linkage:} \quad BL = (\boldsymbol{I} - \boldsymbol{A}')^{-1}J \tag{5.11}$$

where $\boldsymbol{Y}$ is a diagonal matrix of vector Y in the sense that $diag(\boldsymbol{Y}) = Y$, $\boldsymbol{A}'$ is the transpose of $\boldsymbol{A}$, and J is a vector of ones (summation vector). If we expand (5.10) and (5.11) for each element, it gives the coefficient of each industry n:

$$fl_n = 1 + \underbrace{\sum_{k=1}^{N} \alpha_{nk}\frac{y_k}{y_n}}_{\text{direct forward linkages}} + \overbrace{\sum_{k=1}^{N}\sum_{j=1}^{N} \alpha_{nk}\alpha_{kj}\frac{y_j}{y_n}}^{\text{indirect forward linkages through one intermediate good}} + \dots \tag{5.12}$$

$$bl_n = 1 + \underbrace{\sum_{k=1}^{N} \alpha_{kn}}_{\text{direct backward linkages}} + \overbrace{\sum_{k=1}^{N}\sum_{j=1}^{N} \alpha_{jk}\alpha_{kn}}^{\text{indirect backward linkages through one intermediate good}} + \dots \tag{5.13}$$

fl_n shows how much is the value of production that goes to the other industries (not final consumers), normalized by y_n. It includes both direct linkages (customers) and indirect linkages through intermediate goods (customers of customers). Similarly bl_n measures the amount of input that comes from other industries, directly or indirectly, normalized by y_n. Note that these two indices are independent of each other. There may be an industry that has strong linkages in either directions or none of them. For instance, some agricultural products are very primitive in nature and are directly sold to the final consumers. Hence, they do not possess any forward or backward linkages. In comparison, steel is a good example of an industry that has both types of linkages: coal and iron ore are backward and many items like canned goods are its forward industries.

Furthermore, the two above linkage concepts cover both vertical and horizontal inter-industry connections. The vertical linkage is a chain of industries from up to downstream producing a final good (e.g. petroleum industries from crude oil extraction to plastics), but horizontal linkage is between two industries exchanging their differentiated products and none of them is necessarily categorized above or below the other (e.g. Fuels and Iron industries). Figure 5.1 shows a simple example of a network of industries, where there are two vertical chains 1 and 2, each including a raw material supplier (S), a manufacturer (M) and a retailer (R) such that S is the upstream and R is the downstream industry. In other words, in each chain S is just forwardly linked, R just backwardly and M has both types of linkages. These two production chains are also horizontally linked from the middle. Therefore, the upstream industry S_1, in addition to M_1 and R_1, is forwardly

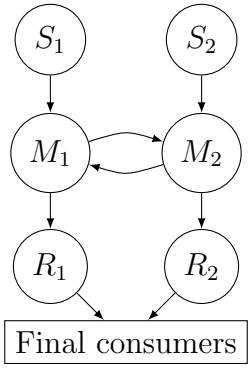


Figure 5.1: Vertical and horizontal linkages. Industry S_1 is forwardly linked to M_1 (direct linkage) and M_2, R_1, and R_2 (indirect linkages). In comparison, Industry R_1 is backwardly linked to M_1 (direct linkage) and M_2, S_1, and S_2 (indirect linkages).

linked with M_2 and R_2 through M_1. Similarly, the backward linkages index of R_1 takes into account the indirect effect of S_2 and M_2 through M_1.

5.3. Adding the VAT

Regarding taxation, consider the government imposes a uniform VAT rate t and fines all non-registered firms who have to be registered according to the law. The ad valorem tax in industry n is added to its market price p_n, in the sense that registered firms charge tax on their sales and issue the corresponding invoices to the buyers, who if registered can use this invoice to refund against their own tax liability. Unregistered firms, on the other hand, buy and sell at tax inclusive price, but they do not pay any tax to the government. For coexistence of formal and informal[69] markets, consistent with Keen (2008), I assume that the number of formal and informal firms are given but they decide about their level of production. The production function is constant returns to scale for formal firms, but due to the expected fine for VAT evasion, the profit of informal firms is subject to decreasing returns.[70]

As an example, consider a simple production chain for a final good R with three stages: raw material supplier (S), manufacturer (M), and retailer (R) such that their pre-tax prices are p_S, p_M, and p_R respectively. We can illustrate the value chain as $0 \xrightarrow{S} p_S \xrightarrow{M} p_M \xrightarrow{R} p_R$. Therefore, at the first stage, one unit of S transforms to one unit of M and then, at the second stage, the new product changes to one unit of final consumer good R. Table 5.1 shows the balance sheet of each industry if they register under the VAT or not.

Table 5.1: Prices and revenues under the VAT.

industry	output price	input price	tax payment	profit
S^f	$p_S + tp_S$	0	$tp_S x_S^f$	$p_S x_S^f$
M^f	$p_M + tp_M$	$p_S + tp_S$	$t(p_M x_M^f - p_S x_S^f)$	$p_M x_M^f - p_S x_S^f$
R^f	$p_R + tp_R$	$p_M + tp_M$	$t(p_R x_R^f - p_M x_M^f)$	$p_R x_R^f - p_M x_M^f$
S^i	$p_S + tp_S$	0	0	$(1+t)p_S x_S^i - c_S$
M^i	$p_M + tp_M$	$p_S + tp_S$	0	$(1+t)(p_M x_M^i - p_S x_S^i) - c_M$
R^i	$p_R + tp_R$	$p_M + tp_M$	0	$(1+t)(p_R x_R^i - p_M x_M^i) - c_R$

[69] From now on, by formal (informal), I mean registered and fully tax compliant (not registered) for the VAT.

[70] I ignore the possibility of misreporting in the formal sector. However, one can assume each representative firm sells one part of its production in the formal market (x_n^f) and one part in the informal (x_n^i). In this case, as shown below, if the expected fine of producing x_n^i is the same as an informal firm with that level of production, the results do not change. For detailed analysis of misreporting in the VAT see Hoseini (2014a).

The superscript f represents formal sector and i informal. The main difference between the two groups is that informal firms do not pay the tax to the government, but instead have a risk of detection and punishment by the tax administration equal to c_n. In the general model with N industries the same pattern holds. According to Table 5.1, tax earnings cancel out the payments of a formal firm to the government, therefore its profit is tax-exclusive. Then,

$$\pi_n^f = p_n x_n^f - \sum_{k=1}^{N} p_k x_{kn}^f - v_n^f, \quad x_n^f \geq 0 \tag{5.14}$$

which is similar to (5.2) and gives (5.4) as equilibrium prices.

Table 5.1 shows that informal firms benefit from evading the VAT, but bear a cost c_n. This cost reflects the risk of being detected and punished by the tax administration due to not registering for the VAT. I assume the evasion cost is a convex function of the level of production x_n^i (discussed below), and thus the production of informal firms is subject to decreasing returns reflecting the expected loss from VAT evasion. This assumption allows for the coexistence of formal and informal markets.[71] Since the production of formal firms is constant returns to scale, they may produce at any level, but informal firms choose a specific amount based on their optimization. If the optimal production of informal firms is higher than total demand in an industry, informal firms supply all demand. Otherwise, they produce at the optimal level and formal firms produce the rest. Consequently, one can easily show that if final demand is high enough in all industries, in the presence of government enforcement the formal production is always nonzero.

Following Allingham and Sandmo (1972), the literature on tax evasion normally assumes that the cost of evasion c_n depends on two factors. One is the value of the fine, reasonably proportional to the evasion level, and the other is the probability of detecting the fraud by tax administrators – see Slemrod and Yitzhaki (2002) for a survey. In this model, I enter these properties by the following simple form:

$$c_n(e_n, \eta_n) = \theta e_n \eta_n, \quad 0 \leq \eta_n \leq 1 \tag{5.15}$$

where e_n and η_n are evasion level and detection probability, respectively. $\theta > 1$ is the

[71] This structure of formal and informal production is broadly the same as Emran and Stiglitz (2005) and Keen (2008).

punishment per unit of evasion set by the government based on legal and political considerations. It should be greater than one, otherwise the marginal benefit of informality always beats its costs. Therefore, the profit function of a representative firm in the informal sector of industry n is written as

$$\pi_n^i = (1+t)p_n x_n^i - \sum_{k=1}^{N}(1+t)p_k x_{kn}^i - v_n^i - \theta e_n \eta_n, \quad x_n^i \geq 0. \tag{5.16}$$

According to the Table 5.1, the amount of evasion e_n of a representative informal firm in industry n is equal to

$$e_n = t(p_n x_n^i - \sum_{k=1}^{N} p_k x_{kn}^i), \tag{5.17}$$

Hence, we can rewrite (5.16) as

$$\pi_n^i = (1+t-t\theta\eta_n)\Big(p_n x_n^i - \sum_{k=1}^{N} p_k x_{kn}^i\Big) - v_n^i, \quad x_n^i \geq 0. \tag{5.18}$$

Proposition 5.1. *Under perfect competition, the production of a representative firm in the informal sector of industry n is obtained from*

$$x_n^i \frac{\partial \eta_n^i}{\partial x_n^i} + \eta_n^i = \frac{1}{\theta} \tag{5.19}$$

Proof. Appendix 5.A. *Q.E.D.*

According to Proposition 5.1, to compute the informal production of each industry, we need to know the functional form of detection probability η_n. In next section, I discuss in detail how we can mathematically define this function and link it with the self-enforcing feature of the VAT. Computing the level of informal production in each industry enables us to measure the share of formal production in each industry as the desired formality index of the tax administration. For simpler mathematics and notation, I redefine x_n^f and x_n as the average formal and total quantity of production per representative informal firm in the economy such that $x_n = x_n^f + x_n^i$.[72] Similarly, we can redefine $y_n^f = p_n x_n^f$ and

[72]Here, the administrations' objective is the share of formal production not the number of firms and this notational change removes the number of firms in each industry from calculations. If the number of informal and formal firms are m_n^i and m_n^f and the aggregate quantities of production are z_n^i and z_n^f, then by this definition $x_n^f = z_n^f/m_n^i$ and $x_n = (z_n^f + z_n^i)/m_n^i$. In other words, the production quantity of an average formal firm which is z_n^f/m_n^f is normalized by m_n^f/m_n^i. This change in the definition does

$y_n = p_n x_n$ as the value of formal and total production in industry n respectively. Then, the formality index in industry n can be measured as $f_n = x_n^f / x_n = y_n^f / y_n$.

5.3.1. Probability of detection

Before any further analysis, we need a reasonable assumption about the probability of detection η_n. Broadly, there are two information sources for the tax administration to check the business records. One is within-firm information such as profit and wages, the other is external information, such as transactions with other businesses and banks. Detection of within-firm information is usually done by visiting audit, but the external information is attainable by checking the bank accounts, invoices etc. In the following, I first explain these factors and how the administration can exploit them and then I define a functional form for this probability.

Firm size

On the intra-firm side, a crucial factor that determines tax evasion is the firm size. Kleven et al. (2009) show that although theoretically a firm can collude with its employees to unreport its tax liability, in practice, if the number of employees is large, maintaining such collusion becomes very difficult. They indicate this collusion may break for a number of reasons. The first set is a random shock due to unplanned events such as a conflict between employer and employee, moral concerns of a newly hired employee, or a mistake of an employee or employer to reveal the records to the tax administration. The other reason can be rational whistle blowing because of a rewards by the tax administration which is common in several countries.[73] If the number of employees is large, collusion is fragile and easily breaks down by each of these reasons. Therefore, this type of third-party reporting dramatically improves tax enforcement of bigger firms. De Paula and Scheinkman (2010) also develop a model with a trade-off between paying taxes versus the firm size. They show that tax registration is positively associated with the revenue and the number of employees among Brazilian firms. To incorporate this effect in the probability function of an informal firm, I use the size of the firm x_n^i. To ensure that

not alter any result of the model because the production of formal firms is CRS and they are fully compliant. Alternatively, one can enter the number of firms in the calculations and find formality by computing aggregate numbers and $z_n^f/(z_n^f + z_n^i)$, which leads to the exact same results.

[73] Kleven et al. (2009) mention some example from OECD countries like the US and Japan.

the probability always lies between zero and one, I define the parameter ρ_n as the share of firms that are detected by the tax administration if the whole industry n is informal and assume the chance that an informal firm in industry n is detected by visiting audit is $\rho_n x_n^i / x_n$. The variation in ρ_n across industries captures the industry-level differences in use of financial sector, labor intensity, and all other factors that can ease the audit by the tax administration for that specific industry.

Firm-to-firm transactions

The second source of information for the tax administration is arm's-length transactions of the firm (Kopczuk and Slemrod, 2006). These transaction records enable tax administrators to verify them with the accounts of their business partners. This source is especially important in the VAT because formal firms are required to keep the records of their sales and purchases. In practice, there are different methods of VAT administration based on the type of assessment, technology and invoice reporting, but in all types, firms either have to keep the sales and purchase invoices for some years or send them to the tax administration with their tax report.[74] Importantly, formal traders have a strong incentive to report their purchase invoices because they can get input credit and deduct their VAT bill. Each purchase invoice as a third-party reported piece of information enables the authority to investigate formality and registration status of the seller. In other words, the incentive of formal clients to report the transaction for getting input credit makes an additional risk of detection for the informal supplier. This risk also depends on how much the tax administration puts effort on cross-checking the input credit claim with the corresponding tax payment.

This feature of the VAT, which is labeled as self-enforcing property (Keen and Smith, 2006), is a one-way effect from formal firms to their supplier not costumers. When a formal firm purchases an input, it requests an invoice to send to the tax administration for input credit. Then, the administration can use the invoice to check whether the supplier is formal or not. However, when a formal firm sells a product, it issues the invoice irrespective of the buyer's type. If the new buyer is a formal firm, again it uses the invoice for input credit and the two records – formal seller's sale and formal buyer's

[74] With the new developments in electronic invoicing, in many countries firms report each single invoice to the tax administration electronically with substantially less compliance cost. For more information on different types of VAT administration see Ebrill et al. (2001) and Hoseini (2014a).

purchase – enable the tax administration to cross-check them together. Otherwise, the invoice just informs that the product is sold, regardless of whether the buyer is an informal firm or a final consumer. Thus, in this case, the tax administration cannot distinguish between the ones that should be punished (informal firms) and not punished (final consumers).

To incorporate the self-enforcing feature in the detection probability, I assume that risk of being detected by invoices is larger when the share of formal customers who inform the tax administration about the transaction is bigger. Hence, the detection probability is an increasing function of the share of sales to formal firms in total sales of industry n which is $\sum_{k=1}^{N} x_{nk}^{f}/x_n$. Notice that because both markets are perfectly competitive, a single informal firm cannot change x_n by changing its own production level. Thus, each firm takes x_n as given when optimizing profit.

I now discuss how tax administrations can utilize the two above information sources and then translate it to a mathematical probability function for η_n. Some of the enforcement strategies are not specific to the VAT such as identification of non-registered firms and the rate of visiting audits reflecting the efforts to reveal within-firm information. In comparison, as described above, a distinctive enforcement strategy for the VAT is the cross-checking of the invoices. Both of these strategies directly depend on tax administration expenditures and efforts for the VAT enforcement. Thus, I assume that the tax administration's policy in each industry is reflected in a function $\phi_n = \frac{g_n}{\bar{g}}$, where g_n is the expenditure to enforce tax in industry n and $\bar{g}$ is a fixed term to ensure $\phi_n \leq 1$. Altering across industries, $0 \leq \phi_n \leq 1$ is the first policy variable that the tax administration seeks to find its optimum allocation among different industries for maximizing net VAT revenue. The tax administration also determines the share λ_n of the enforcement to dedicate for invoice cross-checking versus revealing insider information in each industry. This parameter – multiplied by ϕ_n – determines the level of cross-checking which means how much the administration tries to find and check the seller of a purchase invoice that a formal customer, requesting input credit, claims to buy from.

Hence, I can combine all of the above factors in the following functional form:

$$\eta_n = \phi_n\Big(\underbrace{(1-\lambda_n)\frac{\rho_n x_n^i}{x_n}}_{\text{risk of detection thourgh within-firm information}} + \overbrace{\lambda_n \sum_{k=1}^{N}\frac{x_{nk}^f}{x_n}}^{\text{risk of detection by invoice cross-checking}}\Big) \quad 0 \leq \phi_n \leq 1, \quad 0 \leq \lambda_n \leq 1. \tag{5.20}$$

When the enforcement expenditure g_n in an industry n goes up, ϕ_n approaches one and the tax administration finds the informal firms in that industry with a higher probability. In addition, since $\sum_{k=1}^{N} x_{nk}^f + \rho_n x_n^i \leq x_n$ the probability is always less than 1.

5.4. Market equilibrium and optimal enforcement

Given the two instruments to enforce the VAT, in this part, I find the equilibrium informal production by evaluating the response of informal firms to the enforcement policy. To reduce the risk of unintentional mistakes and whistle-blowing threat, informal firms can control their size. Therefore, they cannot grow as formal firms and their size in this regard depends on the detection possibilities of the tax administration which is reflected in ρ_n. The second risk of detection is through invoice cross-checking of business transactions. To reduce this risk, informal firms should pay attention to their costumers' type. An informal firm has two choices regarding the cross-checking risk: (I) selling the product to all business costumers irrespective of their type, (II) not selling to formal costumers and match with informal firms. In action (I) the informal firm does not care about the type of their clients and issues a fake invoice if necessary which results in a risk of detection by cross-checking. The advantage of this is selling the product at the same price as formal firms $(1+t)p_n$ and earning a money equal to tp_n on each transaction. This action can be a desirable strategy when the bulk of costumers of the firm are informal and the invoice cross-checking rate λ_n is small in industry n. In comparison, action (II) is matching with an informal client, who cannot apply for input credit, to be insured against the cross-checking risk. However, in this case, the seller cannot sell its product as high as tax inclusive price of formal firms and in order to persuade the informal client to match, the firm has to suggest a lower price and share part of the VAT evasion with the buyer. Therefore, if an informal firm in industry n chooses action (II), its sales price

is between p_n and $(1+t)p_n$.

The extent of reduction in the price of informal firms under action (II) depends on the bargaining powers of the two parties. When the bulk of costumers of a product are informal firms or final consumers, the informal seller has higher bargaining power and the reduction in its sales price relative to $(1+t)p_n$ is smaller (and vice versa). In other words, the supplier has less bargaining power for matching when the bulk of clients of its product are formal firms. This means that if the supplier decides to match, the price would be a decreasing function of the share of formal clients. I therefore assume that the average reduction in price, which can be interpreted as the matching cost, for each unit of evasion is $\mu \sum_{k=1}^{N} x_{nk}^f / x_n$. Here, $0 \leq \mu \leq 1$ reflects the average share of evasion that the firm has to pay for satisfying the informal buyer to cooperate and $\sum_{k=1}^{N} x_{nk}^f / x_n$ reflects the bargaining power of the informal seller in the market. Thus, in industry n, while the sales revenue of an informal seller choosing action (I) is $(1+t)p_n x_n$, in case it chooses action (II), its sales revenue decreases to $(1+t)p_n x_n - e_n \mu \sum_k x_{nk}^f / x_n$.

We can find the turning point that informal firms decide between the two actions by comparing the profit in each case. Under action (I), the detection probability η_n is equal to (5.20) and overall profit is computed by substituting (5.20) into (5.18). Under action (II), the cross-checking risk is removed from the detection probability and we would have $\eta_n = \phi_n(1-\lambda_n)\rho_n x_n^i / x_n$, instead, the matching cost $e_n \mu \sum_k x_{nk}^f / x_n$ is subtracted from the overall profit (5.18). With straightforward calculations, we can show that the difference in profits of action (I) and (II) is

$$\pi_n^i(\mathrm{I}) - \pi_n^i(\mathrm{II}) = (\mu - \theta\phi_n\lambda_n) e_n \sum_{k=1}^{N} \frac{x_{nk}^f}{x_n} \tag{5.21}$$

Therefore, if $\mu > \theta\phi_n\lambda_n$ which means the average cost of matching is higher than the risk of cross-checking, the informal firms in industry n prefer action (I), otherwise they choose (II). In the next step, I study the pattern of formality across different industries based on their linkages. The following proposition shows the equilibrium relationship between formality defined as $f_n = x_n^f / x_n$, government enforcement, and forward linkages.

Proposition 5.2. *Define* $\gamma_n = \dfrac{\min[\lambda_n, \frac{\mu}{\theta\phi_n}]}{2(1-\lambda_n)\rho_n}$. *Then, formality index in n is given by*

$$f_n = 1 - \frac{1}{2\theta\phi_n(1-\lambda_n)\rho_n} + \gamma_n \sum_{k=1}^{N} \alpha_{nk} \frac{y_k}{y_n} f_k \tag{5.22}$$

In addition, $F_{(N\times 1)}$ *as a vector containing* $f_1 \dots f_N$*, is positively associated with the forward linkages matrix defined as* $\boldsymbol{FL} = (\boldsymbol{I} - \boldsymbol{Y}^{-1}\boldsymbol{\Gamma}\boldsymbol{A}\boldsymbol{Y})^{-1}$*, where* $\boldsymbol{\Gamma}_{(N\times N)}$ *is a diagonal matrix consisting* $\gamma_1 \dots \gamma_N$ *as diagonal elements.*

Proof. Appendix 5.B *Q.E.D.*

Since ϕ_n is directly associated with g_n, Proposition 5.2 intuitively indicates that the tax administration's attempt to find informal firms reduces the size of informal market.[75] Moreover, Proposition 5.2 indicates how formality in one is linked to formality in the other industries.[76] To provide more insight, define $\tilde{F}_{(N\times 1)}$ a vector containing $1 - \dfrac{1}{2\theta\rho_n(1-\lambda_n)\phi_n}$ which is the formality vector in the absence of cross-checking ($\forall n,\ \lambda_n = 0$) and its elements are always between zero and one. Then, $F = (\boldsymbol{I} - \boldsymbol{X}^{-1}\boldsymbol{\Gamma}\boldsymbol{A}\boldsymbol{X})^{-1}\tilde{F}$ – see (5.55) in Appendix 5.B for proof – and from Theorem 1, we can write

$$F = (\boldsymbol{I} + \boldsymbol{X}^{-1}\boldsymbol{\Gamma}\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}^{-1}(\boldsymbol{\Gamma}\boldsymbol{A})^2\boldsymbol{X} + \dots)\tilde{F} \tag{5.23}$$

where $\boldsymbol{X}^{-1}(\boldsymbol{\Gamma}\boldsymbol{A})^t\boldsymbol{X}\tilde{F}$ is the indirect effect of formality in other industries in rising the formality of an industry through t intermediate goods in between. If we expand (5.23) for each element of F

$$f_n = \underbrace{\tilde{f}_n}_{\substack{\text{no linkage}\\ \text{effect}}} + \overbrace{\sum_{k=1}^{N} \gamma_n\alpha_{nk}\frac{y_k}{y_n}\tilde{f}_k}^{\text{the firm's own formal costumers}} + \underbrace{\sum_{k=1}^{N}\sum_{j=1}^{N} \gamma_n\alpha_{nk}\gamma_k\alpha_{kj}\frac{y_j}{y_n}\tilde{f}_j}_{\substack{\text{formal customers through one intermediate good}\\ \text{(formal customers of customers)}}} + \dots \tag{5.24}$$

where $\tilde{f}_n = 1 - \dfrac{1}{2\theta\phi_n(1-\lambda_n)\rho_n}$. Therefore, by utilizing the self-enforcing feature of the

[75] Note that because formality is always between zero and one, there can be a corner solution if $\phi_n < \frac{1}{2\theta}$.

[76] The forward linkage index is not completely the same as (5.11), because of the existence of $\boldsymbol{\Gamma}$ consists of γ_n as diagonal elements. γ_n is a policy variable set by the administration and I show later that $0 \le \gamma_n < 1/\rho_n$. If it is set equal to 1 in all industries, the linkages indices become the same as (5.11) and (5.12), otherwise $\gamma_1, \dots, \gamma_N$ act as multipliers specifying a fixed weight to the linkages of each industry to compute FL and BL.

VAT, tax administration can increase the formal sector size from $\tilde{f}_n$ to f_n. The particular interest of Proposition 5.2 is in explaining Figure 5.1. From (5.8) we have

$$\sum_{k=1}^{n} \alpha_{nk} \frac{y_k}{y_n} = 1 - \frac{w\beta_n}{y_n} \tag{5.25}$$

The left hand side of (5.25) is the direct effect of forward linkages in (5.24). Thus, we can conclude that the forward linkages of an industry have a negative relationship with its final demand share.[77] Consequently, an activity with more final demand in total production is more likely to be informal and has a lower tax registration rate.

5.4.1. Optimal Enforcement

After characterizing the formality of each industry, in this section, I derive the optimal enforcement policies to control informality. To avoid mathematical complexity, in this part, I assume the informal firms are not constrained by market demand and their production is not a corner solution. In section 5.4.2, I analyze the problem when the demand side determines their production. The first step for finding the optimal policy is defining the objective function of the tax administration and its policy variables. The VAT revenue for each representative x_n is equal to $t(p_n x_n^f - \sum_{k=1}^{N} p_k x_{kn}^f)$. By substituting $x_{kn}^f = \alpha_{kn} x_n^f p_n / p_k$ from (5.3), we obtain

$$t(p_n x_n^f - \sum_{k=1}^{N} p_k x_{kn}^f) = t(1 - \sum_{k=1}^{N} \alpha_{kn}) p_n x_n^f = t\alpha_n y_n f_n \tag{5.26}$$

On the other hand, the administration faces an enforcement expense g_n in each industry which is an increasing function of ϕ_n. If g_n approaches zero, detection becomes infeasible and if it goes up ϕ_n approaches one. Knowing revenues and costs, we can write the tax administration's optimization problem

$$\max_{\substack{g1,\ldots,g_N \\ \lambda_1 \ldots \lambda_N}} \quad r = \sum_{n=1}^{N} t\alpha_n y_n f_n - g_n \tag{5.27}$$

The next two propositions indicates the optimal share of cross-checking and enforcement expenditure in each industry.

[77] w is the net-of-tax income, and in the presence of a homogeneous VAT, w represents gross household income divided by $1 + t$.

Proposition 5.3. *The optimal share of cross-checking in total enforcement spending of each industry is given by*

$$\lambda_n = \begin{cases} \dfrac{\mu}{\theta\phi_n} & \text{if} \quad 2\rho_n \dfrac{x_n^i}{x_n} < \displaystyle\sum_k \alpha_{nk} \frac{y_k}{y_n} f_k \\ 0 & \text{if} \quad 2\rho_n \dfrac{x_n^i}{x_n} > \displaystyle\sum_k \alpha_{nk} \frac{y_k}{y_n} f_k \end{cases} \tag{5.28}$$

Proof. Appendix 5.C. *Q.E.D.*

Expression (5.28) states that cross-checking is optimal in industry n when it increases detection probability at least twice higher than the firm size threat.[78] In this case, the tax administration should put effort into cross-checking as much as it does not persuade informal firms to contract with their peers. Moreover, the right-hand side of the inequality condition is an indicator of forward linkages – see the second term of (5.22). Therefore, λ_n is positively associated with forward linkages of an industry.

Proposition 5.4. *The optimal expenditure for tax enforcement in industry n is given by*

$$G^2 = \frac{t\bar{g}}{2\theta} \mathbf{\Lambda} (\boldsymbol{I} - \boldsymbol{A'\Gamma})^{-1} A \tag{5.29}$$

where G^2 and A are two $N \times 1$ vector with elements g_n^2 and α_n respectively, $\boldsymbol{A}$ is defined as (5.9), and $\mathbf{\Lambda}$ is a diagonal matrix comprising $\dfrac{y_n}{(1-\lambda_n)\rho_n}$.

Proof. Appendix 5.D. *Q.E.D.*

For better illustration, I rewrite (5.29) in the extensive form for each element

$$g_n = \sqrt{\frac{t\bar{g}y_n}{2\theta(1-\lambda_n)\rho_n} (\alpha_n + \sum_{k=1}^{N} \alpha_k \gamma_k \alpha_{kn} + \sum_{k=1}^{N} \sum_{j=1}^{N} \alpha_j \gamma_j \alpha_{jk} \gamma_k \alpha_{kn} + \dots)} \tag{5.30}$$

The immediate result of the Proposition 5.4 is that the optimal expenditure to enforce tax in each industry depends on the square root of the tax rate divided by the surcharge rate of punishment. Also, higher $\bar{g}$ and lower ρ_n mean that tax administration has to spend more money to reach the same probability of detection.

[78] Factor 2 is due to the quadratic functional form of the cost of evasion in x_n^i. If the cost function has higher degree of convexity this number increases and vice-versa.

Apart from these effects, (5.30) shows that the optimal expenditure is positively associated with backward linkages of the industry defined as $(\boldsymbol{I} - \boldsymbol{A'\Gamma})^{-1}$. This reflects the fact that the self-enforcement property of the VAT is from costumers to suppliers not vice-versa. In other words, backwardly linked industries are directly or indirectly customers of forwardly linked ones, so their compliance leads to higher compliance of their suppliers too. Therefore, they are better cases for regular tax audit than not backwardly linked industries. This result is of particular interest since it shows how inter-industry linkages can affect the optimal policy.

The backward linkage index used in Proposition 5.4 is to some extent different from (5.11), introduced by Rasmussen (1957). The first difference is weighting by the intensity of value-adding input vector A – compare the extensive forms in (5.13) and (5.30). The second is the appearance of $0 \leq \gamma_n \leq 1/\rho_n$ in (5.30). As shown in Proposition 5.3, the optimal γ_n for the administration is positively associated with the forward linkages of the industry n. Therefore, γ_n act as a weight for measuring backward linkages that overstates the linkages with the suppliers that have strong forward linkages, but rules out the linkages with industries that have low forward linkages with $\lambda_n = 0$.

Putting Propositions 5.4 and 5.3 together yields the implication of interest: *The tax administration should spend more to reveal within-firm information of backwardly linked industries and make more cross-checking in forwardly linked ones.* For instance, in the simple example of Table 5.1, the administration should put more effort on the downstream R and try to reveal the within-firm information in this industry (e.g. visiting audit, tempting reward for whistle-blowers). But in S, it should dedicate the spending to check the invoices and transactions.

Although in (5.30), just the positive and direct effect of backward linkages on g_n is visible, forward linkages can also rise g_n through λ_n. In comparison to the linearity of backward linkages, the impact of forward linkages on the optimal expenditures is non-linear. According to Proposition 5.3, if forward linkages are big enough to hold the upper inequality in (5.28), g_n increases by factor $\sqrt{\frac{1}{1-\lambda_n}}$, otherwise it does not change by the level of forward linkages. Thus, most of the enforcement spending should be dedicated to industries that are both forwardly and backwardly involved in the production chain. In contrast, the least spending should be given to the industries that do not possess any linkages. Moreover, according to (5.26) and (5.30), enforcement is not beneficial in

small industries in which $\alpha_n y_n < \frac{2\bar{g}}{t\theta\rho_n}$ and a convenient policy would be exempting these industries from the VAT.[79]

5.4.2. Market demand constraint and assortative matching equilibrium

So far in the model, I have assumed that the informal firms are not constrained by market demand and their optimal decisions are always found by the internal optimum of their profit maximization problem. In this section, I want to analyze their behavior when market demand is binding and imposes a corner solution to their decision. As mentioned above, an informal firm can choose between two actions to deal with cross-checking: (I) selling to all types of clients, (II) commit and match with an informal client. In each case, I denote the internal solution of the optimal production of an informal firm (as the case of Proposition 5.1) by x_n^{i*} and the maximum purchase of available customers for a representative informal firm by $\bar{x}_n^i$. In a similar way, I define $y_n^{i*} = p_n x_n^{i*}$ and $\bar{y}_n = p_n \bar{x}_n$. If $y_n^{i*} > \bar{y}_n^i$, the market demand is binding and producing at the level of y_n^{i*} is not possible for informal firms. In order to take market demand into consideration, I assume that final consumers are indifferent between buying from formal and informal markets such that the share of informal sales in final consumption of industry n is $\frac{y_n^i}{y_n} w\beta_n$, where $w\beta_n$ is the value of final consumption in industry n (see (5.8)).[80] Then, we can find the production level and payoff of an informal firm from choosing each action. Given the concavity of the profit, the actual production value of the informal firm after each action will be $y_n^i = \min[y_n^{i*}, \bar{y}_n^i]$. Table 5.1 shows y_n^{i*} and $\bar{y}_n^i$ after choosing action (I) and (II).

Table 5.1: Outcomes of an informal firm from action (I) (selling to all types of firms) and action (II) (not selling to formal firms) – y_n^{i*} represents the internal solution of optimization problem, while $\bar{y}_n^i$ is the market available for a representative informal firm.

	y_n^{i*}	$\bar{y}_n^i$
(I)	$\frac{1}{2(1-\lambda_n)\rho_n}\left(\frac{y_n}{\theta\phi_n} - \lambda_n \sum_k \alpha_{nk} y_k^f\right)$	$\frac{y_n^i}{y_n} w\beta_n + \sum_k \alpha_{nk} y_k^i + \sum_k \alpha_{nk} y_k^f$
(II)	$\frac{1}{2(1-\lambda_n)\rho_n\theta\phi_n}\left(y_n - \mu \sum_k \alpha_{nk} y_k^f\right)$	$\frac{y_n^i}{y_n} w\beta_n + \sum_k \alpha_{nk} y_k^i$

Moving from (I) to (II), the value of market demand $\bar{y}_n^i$ drops, but when cross-checking

[79] To show this, assume there is no cross-checking and $\forall n\ \lambda_n = 0$. Then, the revenue in each industry becomes $r_n = t\alpha_n y_n (1 - \frac{\bar{g}}{2\theta g_n \rho_n}) - g_n$ and its optimal amount is $r_n^* = t\alpha_n y_n - \sqrt{\frac{2t\alpha_n y_n \bar{g}}{\rho_n \theta}}$. If $r_n^* < 0$ the administration is better off by leaving industry n out from the VAT.

[80] alternatively, one can assume a fixed share of final consumers for each type of firms, but as long as the inter-firm factors does not affect the final demand of each group the qualitative results are the same.

is optimal and $\lambda_n > 0$ the endogenous optimum y_n^{i*} increases. The market available after action (I) is the total demand in the industry and it is the optimal choice of informal firms when $\lambda_n = 0$. On the other hand, when cross-checking is optimal and action (I) has lower y_n^{i*}, informal suppliers, missing the formal market demand, match with their informal clients.

According to Table 5.1, the equilibrium point may change based on whether $\bar{y}_n^i$ is binding or not. When there is no restriction from market demand, the results are presented in Proposition 5.4 and 5.3. In comparison, when the market demand is binding, FOC of profit maximization in proposition 5.1 and 5.2 are not valid anymore and informal production in each industry is determined by the production of the clients. There are two equilibriums in which the market demand $\bar{y}_n$ is binding. First, if $y_n^{i*}(\text{I}) > \bar{y}_n^i(\text{I})$, informal firms choose (I) and supply all market demand in industry n. In reality, this can happen when market demand for product n is so small that the tax administration leaves it out from taxation (e.g. exemption policies). The second case is when $y_n^{i*}(\text{II}) > \bar{y}_n^i(\text{II})$ and the optimal λ_n is positive. This equilibrium is an *assortative matching* (AM) equilibrium, in the sense that informal firms match with the same type customers and supply all of the informal demand while, formal firms just sell in the formal market to formal businesses or final consumers. This equilibrium is in close relation to the results of De Paula and Scheinkman (2010) which show under VAT each firm trades with the same type. The below Proposition presents the optimal policy when the equilibrium in some industries are AM.

Proposition 5.5. *If AM holds in industries indexed by $M+1, \ldots, N$, then the optimal policies are*

1. *If $M < n \leq N$ (n is under AM): $\lambda_n = 1$ and $g_n = \dfrac{\bar{g} x_n}{\theta \sum_{k=1}^{N} x_{nk}^f}$.*
2. *If $1 \leq n \leq M$ (n is not under AM): λ_n is found from Proposition 5.3 and*

$$G_1^2 = \frac{t\bar{g}}{2\theta} \mathbf{\Lambda_1} \Big(\boldsymbol{I} - \mathbf{\Gamma_1 A_{11}'} - \boldsymbol{D'} \mathbf{\Gamma_2 A_{12}'} \Big)^{-1} \big(A_1 + \boldsymbol{D} A_2 \big) \tag{5.31}$$

where G_1^2 is the vector containing $g_1^2, \ldots, g_M^2$ of the M industries that are not under AM, subscripts 1 and 2 represent the decomposition of the corresponding vector or matrix into $n \leq M$ and $n > M$ respectively, such as $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$, $\boldsymbol{A} =$

$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$, *etc., and* $\boldsymbol{D} = (\boldsymbol{I} - \boldsymbol{Y}_2^{-1}\boldsymbol{B} - \boldsymbol{A}_{22})^{-1}\boldsymbol{A}_{21}$.

Proof. Appendix 5.E. *Q.E.D.*

Here, we have two groups of firms: $1 \leq m \leq M$ in which AM is not the equilibrium and $M < k \leq N$ which are under AM. The results 1 indicates that in the industries that are under AM, the administration should focus only on invoice cross-checking. In this case, the level of expenditures are such that $\pi_k^{i*}(\text{I}) \leq \bar{\pi}_k^i(\text{II})$.

On the other hand, according to the result 2, the optimal expenditures for industries $1 \leq m \leq M$ are similar to Proposition 5.4, but with some changes due to AM. Basically, AM imposes a linear relationship between informality in industries $M < k \leq N$ and $1 \leq m \leq M$ which is reflected in a transformation matrix $\boldsymbol{D}$ in the sense that $X_2^i = \boldsymbol{D}X_1^i$. This relationship transforms X_2^i into X_1^i and reduces the dimension of optimization matrix from N to M. Therefore, in comparison to Proposition 5.4, here $\boldsymbol{\Gamma A'}$ and A (both with N rows) are replaced by $\boldsymbol{\Gamma_1 A'_{11}} + \boldsymbol{D'\Gamma_2 A'_{12}}$ and $A_1 + \boldsymbol{D}A_2$ (both with M rows) respectively. Hence, if binding constraints from demand side is added to the model, still the results of Proposition 5.4 and 5.3 are valid with the difference that here X_2^i is indirectly affected by G_1 through X_1^i but not by G_2. The role of G_2 is just to prevent firms from choosing action (I).

When AM is not the initial equilibrium, tax administration may be able to shift the equilibrium to the point that AM holds. According to Proposition 5.5, AM can imposed to be the equilibrium in industry n by setting $\lambda_n = 1$ and $g_n^{AM} = \bar{g}x_n / \sum_k \theta x_{nk}^f$. Two factors can make AM the desirable equilibrium for the administration. First, under AM, informal firms constrained by the market demand, produce below their optimal level which leads to less informality. Second, if formal firms comprise a large share of clients of industry n (large $\sum_k x_{nk}^f$), the administration can spend much less in industry n to hold AM (small g_n^{AM}). To increase the size of formal customers $\sum_k x_{nk}^f$, the administration can increase enforcement in the industries that buy product n as input. As an example, if industry m is the only customer of industry n, the administration can impose AM in n by increasing g_m and choosing $\lambda_n = 1$. Then informality in both m and n decline with the possibility that g_n under AM is less than before. Thus, imposing AM in industry n can be optimal if the benefits (higher f_n, f_m and possibly less g_n) beats the cost (higher g_m). For finding the industries that AM is the desirable equilibrium, the administration

needs to compare the overall benefits and costs with and without imposing AM in each industry which can be very complicated to model analytically. However, looking at the denominator of g_n^{AM}, we can say that the required expenditure to impose AM in industries that have low forward linkages is large and if all client of an industry are final consumers, g_n^{AM} approaches infinity and AM cannot be implemented. Therefore, AM is more likely to be the optimal equilibrium in an industry with strong forward linkages and large number of formal customers.

5.5. Empirical evidence

In this section, I provide some evidence about the existence of the VAT self-enforcement from the Indian service sector. India is a good developing country to test this hypothesis given its recent policy reform in taxation. In this paper, I concentrate on service sector given its big policy change for VAT adoption in 2003-04. Service tax in India is levied just by the central (not state) government. It was introduced in 1994 on insurance and stock brokerages, but the major change happened in 2003 by introducing the VAT in service tax rules and expanding the coverage of the rule to more than 80 activities. Before 2003, the credit for inputs was not available. The rate of service tax also has experienced a number of changes. Being 5% initially, the rate increased to 8% in 2003 and then 10.2% in 2004. In the mid-2006, it again rose to 12.24%, and then reduced to 10.3% in 2009.[81]

Given the nation-wide service tax policy change in India, I examine the existence of self-enforcing feature by testing whether the VAT has been more effective in reducing the size of shadow economy of forwardly linked activities or not. As shown in Proposition 5.2, the idea behind this hypothesis is that naturally, firms with strong forward linkages have more business customers and among them formal traders who prefer to buy from VAT compliant suppliers. Thus, after VAT adoption, we expect more tax compliance in forwardly linked industries. For this purpose, I use different data sources; two firm-level surveys on services, and the I-O table of Indian economy.

The principal data I use is two surveys on "unorganized service sector" by NSSO[82]

[81] These facts are obtained from Rao and Rao (2006) and `http://indiataxes.com`.

[82] National Sample Survey Organization, which is under Ministry of statistic of India and conducts a variety of surveys including enterprise level on unorganized manufacturing and services.

which are carried out in 2001-02 (round 57) and 2006-07 (round 63). They have nationwide representative samples (330,929 in 2001-02 and 156,352 in 2006-07) of service enterprises and include a variety of information including the status of tax registration.[83] The services they cover consist of transport, storage, communication, hotels and restaurants, real estate, financial intermediation (just round 63), renting and business activities, education, health, social work, and other community, social and personal service activities. The sample excludes government and public sector undertakings and the service sector units registered under the Factories Act. To have a balanced sample, I drop the financial intermediation services that are just surveyed in 2006. In addition, because I am interested in the effect of VAT adoption on formality, I drop all services that are not under service VAT rule in 2006. In this way, I eliminate any potential selection bias for activities that are chosen by the tax administration for service tax adoption. Finally, hair dressing and other beauty treatment services (NIC code 93020) is dropped, because service tax is assigned just for beauty treatment not hair dressing and decomposition is not possible. The list of industries used in the estimation are presented in appendix Table 5.F.2.

Having the firms' tax registration status, enables me to estimate the size of shadow economy at state-year-activity level using the sample weights. Tax registration status is not specific to the service tax and it refers to all types of registration for central or state government tax agencies. In Figure 5.1, percentage of tax registered firms in each activity – using sample weights and separately for 2001 and 2006 – is used in the y-axis. In order to measure the formality index, I define formal and informal production as the level of production that is produced by firms registered and not registered under tax respectively. Then, the **formality index** refers to the percentage of total production by firms that are registered for central or state tax agencies within each state, year, and activity. Table 5.1 shows the summary statistics of formality index for 2001-02 and 2006-07. While on average 18.2 percent of production is produced by formal sector,

[83]The 2006-07 survey has a specific question about tax registration and whether the firm has tax account number (TAN). In the 2001-02 survey, among other types of registration, the enterprise is asked about sales tax and companies act registration (which includes tax registration). To cover other types of tax, I check whether the tax payments of the enterprise are nonzero or not. It is good to mention that all NSSO surveys are conducted independent of tax purposes and no information about the identity of the enterprise is supposed to go to the other agencies. As it mentioned in the manual, according to the Collection of Statistics Act, violation of any of the confidentiality and secrecy of the information by statistics officer or field staff is prosecuted by or with the consent of appropriate agency.

formality increased from 12.79 percent in 2001 to 23.09 percent in 2006.[84] This may reflect that VAT adoption in 2003 had a significant effect on tax compliance of service sector in India. Appendix Tables 5.F.1 and 5.F.2 provide the average formality across state and industries respectively. We note a large variation in formality across states. While over 59 percent of production is created by registered firms in Maharashtra, only 2.83 percent of production is formal in Jammu and Kashmir. Looking into activities, we observe no firm is registered for tax in real estate of self-owned residential buildings, whereas around 80 percent of production in cargo handling in water transport is formal.

The second data source, which is used to measure **forward and backward linkages**, is input-output transaction tables of the national accounts statistics of Indian economy published by Central Statistical Organization in 2006-07 (No report is published in 2001-02). The I-O tables consist of 130 activities including 22 service categories[85] and include use (commodity × industry) and make (industry × commodity) matrices. Using these two, I construct the (commodity × commodity) coefficient matrix $\boldsymbol{A}$.[86] Then, similar to (5.10) and (5.11), I compute forward and backward linkages in each service sector as

$$FL = \boldsymbol{Y}^{-1}(\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{Y}J, \quad BL = (\boldsymbol{I} - \boldsymbol{A}')^{-1}J \tag{5.32}$$

Where $\boldsymbol{Y}$ is the diagonal production matrix and J the summation vector. The final consumption and total demand of each service, used in Figure 5.1, are drawn from I-O tables too. According to Table 5.1, forward and backward linkages indicators are on average 2.133 and 1.749 respectively, and they have small variation across the two survey years. Appendix Table 5.F.2, presents the variation in the linkages across industries. Forward linkages index is changing from 1.02 in education (NIC 2-digit code 80) to 3.11 in business services (NIC 2-digit code 74). In comparison, backward linkages index varies

[84] These numbers are estimated for the sample of "unorganized service sector" that described above and does not cover all service sectors and public enterprises. Therefore they do not reflect the aggregate formality in the whole service sector in India.

[85] Railway transport, Land transport, Water transport, Air transport, auxiliary transport activities, Storage and warehousing, Communication, Trade, Hotels and restaurants, Banking, Insurance, Ownership of dwellings, Education and research, Medical and health, Business services, Computer & related activities, Legal services, Real estate activities, Renting of machinery and equipment, Community, social and personal services, Other services, Public administration.

[86] The methodology for calculating matrix $\boldsymbol{A}$ from use and make tables are available in the Appendix 2 of the manual of Input-Output Transaction Tables: "Mathematical expression on the methodology of construction of associated matrices", which is available at: `http://mospi.nic.in/Mospi_New/upload/report&publication/ftest10/appendix202.pdf`

from 1.21 in education (NIC 2-digit code 80) to 2.33 in freight transport.

To investigate the effect of linkages on shadow economy after service VAT adoption in India, I also utilize the state level variation in **tax buoyancy**. Although service tax adoption in India has been nationwide and the revenue directly goes to the central government, there is a large heterogeneity between states in terms of tax capacity and effort. Rangarajan, Prasad, and Srivastava (2004, Ch.6 p.383) provide tax buoyancy estimates for 28 Indian states, measuring the percentage response of tax revenue to a one percent change in the tax base (proxied by GDP), holding constant all other parameters of tax policy. Their estimates are for a period ending to 2003 which is year of service VAT adoption. Appendix Table 5.F.1 shows the variation is in tax buoyancy is between 1.1 to 1.35 across states. We expect service VAT adoption in 2003 is more effective in reducing shadow economy in state with higher level of tax buoyancy.

Table 5.1: Summary statistics—The numbers show the summary statistics average across state and activities.

Sample	2001-02		2006-07		The whole sample	
	Mean	SE	Mean	SE	Mean	SE
Formal share of production	12.79	(26.09)	23.09	(35.39)	18.20	(31.73)
Forward linkages index	2.208	(0.697)	2.064	(0.720)	2.133	(0.713)
Backward linkages index	1.768	(0.334)	1.733	(0.333)	1.749	(0.334)

To estimate the effect of service VAT adoption on the relative size of formal sector, I use the following regression setup

$$f_{ist} = a_0 + a_1 VAT_t \times TB_s + a_2 X_{ist} + I_i + S_s + T_t + \varepsilon_{ist} \tag{5.33}$$

where f_{ist} is formality index in Industry i, state s and time t, VAT_t is equal to one in 2006-07 and zero otherwise, TB_s is tax buoyancy estimate of state s as given in Table 5.F.1, and X_{ist} is a vector of control variables including the average number of workers – to capture the effect of firm size and third party reporting by labor – and total number of firms – to control for the fixed firm-level enforcement costs – at activity, state, and year level. In addition, I_i, S_s and T_t are vectors of activity, state, and time fixed effects. In the sample, there are 69 industries and 28 states, but some industries do not exist in a number of states and total observation is 1901 in the regressions. To control for underestimation of standard errors, the error terms are clustered at state level.

Column (1) of Table 5.2 illustrates the estimated coefficients of (5.33). The results suggest a positive and significant impact of service VAT interacted with tax buoyancy on formality. The economic effect of the impact is also significant. Formality in Punjab (tax buoyancy=1.35) is $42.27 \times 0.25 = 10.57$ percent more than Arunachal Pradesh (tax buoyancy=1.1) after VAT adoption. The sign of the average number of workers is positive and significant, suggesting higher formal production in bigger firms with more employees. This is consistent with the third party reporting hypothesis of Kleven et al. (2009) which I used in section 5.3.1. The effect of total number of firms is negative and significant which can reflect more difficult enforcement in the presence of large number of firms and fixed firm-level administrative costs (Dharmapala, Slemrod, and Wilson, 2011).

In the next step, I utilize the following regression setup to estimate the effect of forward and backward linkages after VAT adoption on formality

$$f_{ist} = b_0 + b_1 VAT_t \times TB_s \times L_i + b_2 X_{ist} + I_i + S_s \times T_t + \varepsilon_{ist} \tag{5.34}$$

where L_i is the linkages index of activity i and the rest of variables are defined in the same way as (5.33). By including state×year fixed effects in (5.34), I control for all state-level time-variant variables and focus on the effect of forward and backward linkages interacted with VAT dummy and tax buoyancy on formality. Column (2) to (4) of Table 5.2 present the results of the estimation of (5.34) for specifications with only forward linkages, only backward linkages, and both types of linkages. Column (2) shows that industries with higher forward linkages in state with higher tax buoyancy experiences more formality after VAT adoption. For example, after VAT adoption in 2003, formal production of accounting activities (forward linkages=3.11) in Punjab (tax buoyancy=1.35) increased by 11.72 percentage point more than formal production of education (forward linkages=1.02) in Arunachal Pradesh (tax buoyancy=1.1).[87] However, column (3) implies no significant relationship between formality and the interaction term of backward linkages. If we include both linkages indices in one regression, forward linkages remains significant and backward linkages is still insignificant with a sharp reduction in its coefficient (column 4). Moreover, similar to column (1), the average number of workers

[87] The economic effect is computed as $(3.11 \times 1.35 - 1.02 \times 1.1) \times 3.811 = 11.72$

and the number of firms appear with positive and negative signs respectively and their coefficient is significant at 5% level in all specifications.

In sum, the empirical evidence confirms a positive and strong impact of forward linkages of an activity on formal share of production when the VAT is conducted in the economy. This self-enforcing feature of the VAT provides a novel tool for tax administrations to move to a more efficient tax enforcement policy.

Table 5.2: Difference-in-difference regressions for the effect of VAT adoption and linkages on formal production. The estimated regressions are (5.33) for column (1) and (5.34) for column (2)-(4). The dependent variable is share of production by firms registered for tax at each activity, state, and year. VAT adoption is a dummy equal to one after service VAT adoption in 2003. State-level tax buoyancy for 2003 is drawn from Rangarajan et al. (2004). Forward and backward linkages are estimated using I-O tables of Indian economy in 2006. The standard errors are in parentheses and clustered at state level. The * and ** show significance at 10% and 5% respectively.

	Formal share of production (%)			
	(1)	(2)	(3)	(4)
VAT adoption$_t$× Tax buoyancy$_s$	42.267*			
	(21.755)			
VAT adoption$_t$× Tax buoyancy$_s$× Forward linkages$_i$		3.811**		3.769**
		(1.804)		(1.785)
VAT adoption$_t$× Tax buoyancy$_s$× Backward linkages$_i$			4.303	0.162
			(3.069)	(2.764)
Average No. workers	0.314**	0.288**	0.287**	0.288**
	(0.123)	(0.127)	(0.128)	(0.127)
Number of firms (in thousands)	-0.047**	-0.047**	-0.049**	-0.047**
	(0.019)	(0.020)	(0.020)	(0.020)
Observations	1901	1901	1901	1901
R-squared	0.261	0.284	0.282	0.284
Activity fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
State×Year fixed effects	No	Yes	Yes	Yes

5.6. Conclusion

The main contribution of this paper is providing the insight that for optimal VAT enforcement, tax administration should take account of the linkages between different activities. Many developing countries have moved toward replacement of border taxes with the VAT and the optimality of this reform crucially depends on VAT collection efficiency. The administrative advantage of the VAT is that in addition to common tools to reveal intra-firm information, it provides a unique inter-firm information source via its invoice system. This intrinsic feature makes evasion of forwardly linked activities dependent on their customers' informality. Consequently, tax administration can utilize this unique tool to improve enforcement by spending more on backwardly linked firms to reveal their internal information. In comparison, tax administrations should spend extra expenses in forwardly linked activities to cross-check invoices with the corresponding credit claims. Empirical evidence from Indian economy shows that the self-enforcing feature among forwardly linked activities increases formal share of production.

However, the underlying model has some limitations. Generally, this paper neglects some other VAT enforcement issues, as stressed in Keen and Smith (2006), like under-reported sales and over-reported purchases, multiple rates and misclassification of commodities, self-consumption and carousel fraud. In an accompanying paper (Hoseini, 2014a), the misreporting in the VAT is studied in detail, but international aspect of VAT fraud like missing trader (carousel) fraud, need more attention of theorists. In spite of these limitations, the findings of this paper have an important implication for the VAT enforcement design, specially for countries suffering from big size of shadow economy.

Appendix

Mathematical Appendix

5.A. Proof of Proposition 5.1

By defining $\tau_n = t(1-\theta\eta_n)$ and $e_n = t(p_n x_n^i - \sum_{k=1}^N p_k x_{kn}^i)$, we can write (5.18) as

$$\pi_n^i = (1+\tau_n)\frac{e_n}{t} - v_n^i \tag{5.35}$$

which gives the F.O.C as

$$\frac{\partial \pi_n^i}{\partial v_n^i} = \Big(\frac{\partial \tau_n}{\partial x_n^i}\frac{e_n}{t} + (1+\tau_n)p_n\Big)\frac{\partial x_n^i}{\partial v_n^i} - 1 = 0 \tag{5.36}$$

$$\frac{\partial \pi_n^i}{\partial x_{kn}^i} = \Big(\frac{\partial \tau_n}{\partial x_n^i}\frac{e_n}{t} n + (1+\tau_n)p_n\Big)\frac{\partial x_n^i}{\partial x_{kn}^i} - (1+\tau_n)p_k = 0 \tag{5.37}$$

On the other hand, from (5.1), we have

$$\frac{\partial x_n^i}{\partial v_n^i} = \alpha_n \frac{x_n^i}{v_n^i}, \qquad \frac{\partial x_n^i}{\partial x_{kn}^i} = \alpha_{kn}\frac{x_n^i}{x_{kn}^i} \tag{5.38}$$

Therefore, we obtain

$$v_n^i = \alpha_n x_n^i\Big(\frac{\partial \tau_n}{\partial x_n^i}\frac{e_n}{t} + (1+\tau_n)p_n\Big), \qquad x_{kn}^i = \frac{\alpha_{kn}x_n^i}{(1+\tau_n)p_k}\Big(\frac{\partial \tau_n}{\partial x_n^i}\frac{e_n}{t} + (1+\tau_n)p_n\Big) \tag{5.39}$$

Next, by substituting v_n^i and x_{kn}^i from (5.39) into (5.1) and taking (5.4) into account, it turns out

$$\frac{\partial \tau_n}{\partial x_n^i}\frac{e_n}{t} + (1+\tau_n)p_n = p_n(1+\tau_n)^{1-\alpha_n} \tag{5.40}$$

Thus, at the optimum, firm choices are

$$v_n^i = \alpha_n p_n x_n^i (1+\tau_n)^{1-\alpha_n}, \qquad x_{kn}^i = \frac{\alpha_{kn}p_n x_n^i}{p_k(1+\tau_n)^{\alpha_n}} \tag{5.41}$$

By substituting x_{kn}^i from (5.41) into $e_n = t(p_n x_n^i - \sum_{k=1}^N p_k x_{kn}^i)$, we have

$$\frac{e_n}{t} = p_n x_n^i \Big(1 - \frac{\sum_k \alpha_{kn}}{(1+\tau_n)^{\alpha_n}}\Big) = p_n x_n^i \Big(1 - \frac{1-\alpha_n}{(1+\tau_n)^{\alpha_n}}\Big) \tag{5.42}$$

and by substituting (5.42) into (5.40) and rearranging the expression, it turns out

$$\frac{\partial \tau_n}{\partial x_n^i} x_n^i = -(1+\tau_n)\frac{(1+\tau_n)^{\alpha_n} - 1}{(1+\tau_n)^{\alpha_n} - 1 + \alpha_n} \tag{5.43}$$

Since, $0 \leq \tau_n = t(1-\theta\eta_n) < 1$, and τ_n is small, I am able to simplify (5.43) by using the linear approximation of $(1+\tau_n)^{\alpha_n}$ as $1 + \alpha_n\tau_n$.[88] By doing this, we get

$$\frac{\partial \tau_n}{\partial x_n^i} x_n^i = -\tau_n \tag{5.44}$$

Finally, by substituting $\tau_n = t(1-\theta\eta_n)$ in (5.44), we obtain

$$\theta \frac{\partial \eta_n}{\partial x_n^i} x_n^i = 1 - \theta\eta_n \tag{5.45}$$

5.B. Proof of Proposition 5.2

From Proposition 5.1, at the informal firms' optimum, we have

$$\frac{1}{\theta} = \eta_n + x_n^i \frac{\partial \eta_n}{\partial x_n^i} \tag{5.46}$$

If $\mu > \theta\phi_n\lambda_n$ and action (I) is chosen, substituting (5.20) into (5.46) yields

$$\frac{1}{\theta\phi_n} = 2(1-\lambda_n)\rho_n \frac{x_n^i}{x_n} + \lambda_n \sum_{k=1}^N \frac{x_{nk}^f}{x_n} \tag{5.47}$$

Because $x_n^f = x_n f_k$ and we have $x_{nk}^f = \alpha_{nk}\frac{p_k}{p_n} x_k^f$ (see (5.2)–(5.3) for derivation), we can write

$$\frac{x_{nk}^f}{x_n} = \alpha_{nk}\frac{y_k}{y_n} f_k \tag{5.48}$$

[88] Notice that the tax rate is between 0 and 1 (around 0.2 for many countries) and when it multiplies by $0 < 1 - \theta\eta_n < 1$, it becomes even smaller. Therefore, the higher powers of τ_n are negligible for the approximation. For instance, if $\tau_n = 0.1$ and $\alpha_n = 0.5$, the true value and the first order approximation become 1.0488 and 1.05, respectively, with only 0.1% error.

Moreover, at the equilibrium, we have $x_n^i/x_n = 1 - f_n$. Therefore, by rearranging (5.47) we get

$$f_n = 1 - \frac{1}{2\theta\phi_n(1-\lambda_n)\rho_n} + \frac{\lambda_n}{2(1-\lambda_n)\rho_n}\sum_{k=1}^{N}\alpha_{nk}\frac{y_k}{y_n}f_k \tag{5.49}$$

If $\mu < \theta\phi_n\lambda_n$ and action (II) is chosen, then cross-checking risk is removed from the detection probability and we have $\eta_n = \phi_n(1-\lambda_n)\rho_n x_n^i/x_n$, but $e_n\mu\sum_k x_{nk}^f/x_n$ is subtracted from the profit of informal firms. By applying both of these changes the profit of a representative informal firm can be written in the same form as (5.35) with a change in the definition of τ_n such that

$$\tau_n = t(1 - \theta\eta_n - \mu\sum_{k=1}^{N}\frac{x_{kn}^f}{x_n}) \tag{5.50}$$

Then, we can use (5.44) to find the optimal decision of the firm.

$$\theta\frac{\partial\eta_n}{\partial x_n^i}x_n^i = 1 - \theta\eta_n - \mu\sum_{k=1}^{N}\frac{x_{kn}^f}{x_n} \tag{5.51}$$

The equilibrium formality index in each industry can be obtained similar to (5.46) – (5.49)

$$f_n = 1 - \frac{1}{2\theta\phi_n(1-\lambda_n)\rho_n} + \frac{\mu}{2(1-\lambda_n)\rho_n\theta\phi_n}\sum_{k=1}^{N}\alpha_{nk}\frac{y_k}{y_n}f_k \tag{5.52}$$

Therefore, by defining $\gamma_n = \dfrac{\min[\lambda_n, \frac{\mu}{\theta\phi_n}]}{2(1-\lambda_n)\rho_n}$, the formality index becomes

$$f_n = 1 - \frac{1}{2\theta\phi_n(1-\lambda_n)\rho_n} + \gamma_n\sum_{k=1}^{N}\alpha_{nk}\frac{y_k}{y_n}f_k \tag{5.53}$$

Next, I employ vectors F, Φ^{-1} and J with dimensionality $N \times 1$ containing f_n, $\dfrac{1}{\phi_n(1-\lambda_n)\rho_n}$, and ones (summation vector) respectively; and $\mathbf{\Gamma}$ as a diagonal $N \times N$ matrices with γ_i as diagonal elements. Then, (5.53) is written as

$$F = (J - \frac{1}{2\theta}\Phi^{-1}) + \boldsymbol{Y}^{-1}\mathbf{\Gamma}\boldsymbol{A}\boldsymbol{Y}F \tag{5.54}$$

and we immediately get

$$F = (\boldsymbol{I} - \boldsymbol{Y}^{-1}\boldsymbol{\Gamma}\boldsymbol{A}\boldsymbol{Y})^{-1}(J - \frac{1}{2\theta}\Phi^{-1}) \tag{5.55}$$

Since $\gamma_n \geq 0$, according to Theorem 1 (replace V by $\boldsymbol{\Gamma}V$), F always has a unique solution.

5.C. Proof of Proposition 5.3

From Proposition 5.2, when $\mu > \theta\phi_n\lambda_n$, at equilibrium we have

$$x_n^f = x_n - \frac{x_n}{2\theta\phi_n(1-\lambda_n)\rho_n} + \frac{\lambda_n}{2(1-\lambda_n)\rho_n}\sum_{k=1}^{N} x_{nk}^f \tag{5.56}$$

Since $x_n^i = x_n - x_n^f$, by rearranging (5.2), we can write

$$2(1-\lambda_n)\rho_n x_n^i + \lambda_n \sum_{k=1}^{N} x_{nk}^f = \frac{x_n}{\theta\phi_n} \tag{5.57}$$

Then, the derivative of x_n^i with respect to λ_n can be derived from (5.57)

$$2(1-\lambda_n)\rho_n \frac{\partial x_n^i}{\partial \lambda_n} - 2\rho_n x_n^i + \sum_{k=1}^{N} x_{nk}^f = 0, \tag{5.58}$$

so

$$\frac{\partial x_n^i}{\partial \lambda_n} = \frac{1}{1-\lambda_n}\big(x_n^i - \frac{1}{2\rho_n}\sum_{k=1}^{N} x_{nk}^f\big) \tag{5.59}$$

Because $1-\lambda_n > 0$, x_n^i is monotone in λ_n and the sign of $\dfrac{\partial x_n^i}{\partial \lambda_n}$ depends on $2\rho_n x_n^i - \sum_k x_{nk}^f$. If $2\rho_n x_n^i > \sum_k x_{nk}^f$, then x_n^i is increasing in λ_n and the best choice to reduce informality is $\lambda_n = 0$. On the other hand, if $2\rho_n x_n^i < \sum_k x_{nk}^f$, the derivative is negative and the tax administration should focus on cross checking as much as possible ($\lambda_n = 1$). However, according to Proposition 5.2, if λ_n increases such that $\mu < \theta\phi_n\lambda_n$, the informal firm start matching with their peers to eliminate the cross-checking risk from their detection probability. The turning point that makes an informal firm indifferent between matching and not matching is $\lambda_n = \dfrac{\mu}{\theta\phi_n}$. Hence, if $\lambda_n\theta\phi_n > \mu$ the administration gains nothing from increasing λ_n – even if $2\rho_n x_n^i < \sum_k x_{nk}^f$ – and its best choice is $\lambda_n = \dfrac{\mu}{\theta\phi_n}$. Finally,

I show that $\frac{\mu}{\theta\phi_n}$ is always smaller than 1 and thus λ_n is never equal to 1. When cross-checking is optimal, from (5.68) we can write

$$\phi_n^2 = \frac{z_n}{1-\frac{\mu}{\theta\phi_n}} \quad \rightarrow \quad \phi_n^2 - \frac{\mu}{\theta}\phi_n - z_n = 0 \tag{5.60}$$

where $z_n > 0$ is a positive constant. The acceptable solution of (5.60) is

$$\phi_n = \frac{1}{2}\left(\frac{\mu}{\theta} + \sqrt{(\frac{\mu}{\theta})^2 + 4z_n}\right) \tag{5.61}$$

which means that $\phi_n > \mu/\theta$. In sum, the optimal cross-checking in each industry can be written as

$$\lambda_n = \begin{cases} \frac{\mu}{\theta\phi_n} & \text{if} \quad 2\rho_n x_n^i < \sum_k x_{nk}^f \\ 0 & \text{if} \quad 2\rho_n x_n^i > \sum_k x_{nk}^f \end{cases} \tag{5.62}$$

Finally, using (5.49), we obtain (5.28).

5.D. Proof of Proposition 5.4

Using (5.24), I can rewrite (5.27) as

$$r = \sum_{n=1}^{N} t\alpha_n y_n \Big(\tilde{f}_n + \sum_{k=1}^{N} \gamma_n \alpha_{nk} \frac{y_k}{y_n}\tilde{f}_k + \sum_{k=1}^{N}\sum_{j=1}^{N} \gamma_n \alpha_{nk}\gamma_k\alpha_{kj}\frac{y_j}{y_n}\tilde{f}_j + \dots) - g_n \tag{5.63}$$

which yields

$$r = \sum_{n=1}^{N} t\alpha_n y_n \tilde{f}_n + \sum_{n=1}^{N}\sum_{k=1}^{N} t\alpha_n\gamma_n\alpha_{nk}y_k\tilde{f}_k + \sum_{n=1}^{N}\sum_{k=1}^{N}\sum_{j=1}^{N} t\alpha_n\gamma_n\alpha_{nk}\gamma_k\alpha_{kj}y_j\tilde{f}_j + \dots) - \sum_{n=1}^{N} g_n \tag{5.64}$$

By re-indexing we obtain

$$r = \sum_{m=1}^{N} tyx_m\tilde{f}_m(\alpha_m + \sum_{k=1}^{N}\alpha_k\gamma_k\alpha_{km} + \sum_{k=1}^{N}\sum_{j=1}^{N}\alpha_k\gamma_k\alpha_{kj}\gamma_j\alpha_{jm} + \dots) - g_m \tag{5.65}$$

and since $\tilde{f}_m = 1 - \dfrac{1}{2\theta\phi_m(1-\lambda_m)\rho_m}$ and $\dfrac{1}{\phi_m} = \dfrac{\bar{g}}{g_m}$

$$\frac{\partial \tilde{f}_m}{\partial g_m} = \frac{\bar{g}}{2\theta(1-\lambda_m)\rho_m g_m^2}, \tag{5.66}$$

the FOC of (5.66) becomes

$$\frac{\partial r}{\partial g_m} = \frac{t\bar{g}y_m}{2\theta(1-\lambda_m)\rho_m g_m^2}\big(\alpha_m + \sum_{k=1}^{N}\alpha_k\gamma_k\alpha_{km} + \sum_{k=1}^{N}\sum_{j=1}^{N}\alpha_k\gamma_k\alpha_{kj}\gamma_j\alpha_{jm} + \dots\big) - 1 = 0 \tag{5.67}$$

which result in the optimal expenditure in each industry as

$$g_m^2 = \frac{t\bar{g}y_m}{2\theta(1-\lambda_m)\rho_m}\big(\alpha_m + \sum_{k}\alpha_k\gamma_k\alpha_{km} + \sum_{k}\sum_{j}\alpha_k\gamma_k\alpha_{kj}\gamma_j\alpha_{jm} + \dots\big) \tag{5.68}$$

and in the matrix form

$$G^2 = \frac{t\bar{g}}{2\theta}\mathbf{\Lambda}(\boldsymbol{I} - \boldsymbol{A'}\mathbf{\Gamma})^{-1}A \tag{5.69}$$

where G^2 and A are vectors with elements g_n^2 and α_m, respectively; $\mathbf{\Lambda}$ is a diagonal matrix with elements $\dfrac{y_n}{(1-\lambda_n)\rho_n}$ and $\mathbf{\Gamma}$ is defined as Proposition 5.2.

5.E. Proof of Proposition 5.5

If $\lambda_n = 0$, informal firms are always better off by choosing (I) and AM is not the equilibrium. Thus if AM is the equilibrium $\lambda_n > 0$. Moreover, as shown in Proposition 5.3, because the profits are monotone in λ_n, its optimal level is either zero or one. Under AM, the administration does not need to make the firm indifferent between $y_n^{i*}(\mathrm{I})$ and $y_n^{i*}(\mathrm{II})$ (since $y_n^{i*}(\mathrm{II}) \geq \bar{y}_n^i(\mathrm{II})$), thus λ_n is not constrained by the corresponding search cost $\mu/\theta\phi_n$ and at optimum $\lambda_n = 1$.

On the other hand, holding AM requires $\pi_n^{i*}(\mathrm{I}) \leq \bar{\pi}_n^i(\mathrm{II}) \leq \pi_n^{i*}(\mathrm{II})$. This indicates that for imposing AM in industry n, the administration has to set the efforts such that the informal firms are indifferent between actions (I) and (II) while the market demand is binding. When $\lambda_n = 1$, we always have $\pi_n^{i*}(\mathrm{I}) \leq \pi_n^{i*}(\mathrm{II})$, and under AM the demand constraint yields $\bar{\pi}_n^i(\mathrm{II}) \leq \pi_n^{i*}(\mathrm{II})$, thus AM will be the equilibrium if g_n is chosen such that $\pi_n^{i*}(\mathrm{I}) \leq \bar{\pi}_n^i(\mathrm{II})$. By substituting (5.41) and (5.42) in (5.35), we obtain the profit

after action (I) as

$$\pi_n^{i*}(I) = (1+\tau_n)\Big(1 - \frac{1-2\alpha_n}{(1+\tau_n)^{\alpha_n}}\Big)y_n^i \tag{5.70}$$

where $\tau_n = t(1-\theta\eta_n)$. With a first order approximation similar to Proposition 5.1, (5.70) becomes

$$\pi_n^{i*}(I) = (1+\tau_n)^{1-\alpha_n}(2+\tau_n)y_n^i \tag{5.71}$$

When $\lambda_n = 1$, we have $\eta_n = \phi_n \sum_k x_{nk}^f / x_n$ –see (5.20). This means that τ_n is independent of y_n^i and if $1+\tau_n > 0$, the optimal production tends to infinity. Therefore, the level of enforcement expenditure g_n that holds $\pi_n^{i*}(\text{I}) \leq \bar{\pi}_n^i(\text{II})$ leads to $1+\tau_n \leq 0$. The level is equal to

$$g_n = \frac{\bar{g}x_n}{\theta \sum_{k=1}^N x_{nk}^f} \tag{5.72}$$

In the next step, I want to derive the optimal policy when AM is the equilibrium in some industries. Without loss of generality, I divide the industries into two groups: Industries $1,\ldots,M$ that the AM is not the equilibrium, and $M+1,\ldots,N$ in which AM is held. If AM equilibrium holds in industry n, we have

$$y_n^i = \sum_{k=1}^N \alpha_{nk} y_k^i + y_n^i \frac{w\beta_n}{y_n} \tag{5.73}$$

(5.73) gives no answer for the industries that possess no forward linkages ($\sum_k \alpha_{nk} y_k = 0$ and $y_n = w\beta_n$) since they do not have any business customers. Therefore, the industries with no forward linkage are indexed among $1,\ldots,M$. The informal firms that work in industries $1,\ldots,M$ get the same payoff as before, and the optimal cross-checking for them can be obtained from Proposition 5.3. Now, if I decompose vector Y into Y_1 and Y_2 consisting of $y_1,\ldots,y_M$ and $y_{M+1},\ldots,y_N$ respectively, and $\boldsymbol{A}$ into four corresponding sub-matrices such that $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$, we can write (5.73) as

$$Y_2^i = \boldsymbol{A}_{21}Y_1^i + \boldsymbol{A}_{22}Y_2^i + w\boldsymbol{Y}_2^{-1}\boldsymbol{B}_2 Y_2^i \tag{5.74}$$

where $\boldsymbol{Y}_2$ is a diagonal metrics consisting of $y_{M+1},\ldots,y_N$ respectively. From (5.74) we can find

$$Y_2^i = (\boldsymbol{I} - \boldsymbol{Y}_2^{-1}\boldsymbol{B}_2 - \boldsymbol{A}_{22})^{-1}\boldsymbol{A}_{21}Y_1^i, \tag{5.75}$$

so by knowing Y_1^i, Y_2^i is obtained from (5.75). Now, we can use this results to find the optimal policy. Similar to (5.27) tax administration's objective becomes

$$r = t\sum_{n=1}^{N} \alpha_n y_n^f - g_n \tag{5.76}$$

which can be rewritten in matrix form as

$$r = tA'Y^f - I'G \tag{5.77}$$

where G is a vector of g_ns. From (5.75), define $\boldsymbol{D} = (\boldsymbol{I} - \boldsymbol{Y}_2^{-1}\boldsymbol{B} - \boldsymbol{A}_{22})^{-1}\boldsymbol{A}_{21}$, then

$$Y_2^f = Y_2 - Y_2^i = Y_2 - \boldsymbol{D}Y_1 + \boldsymbol{D}Y_1^f \tag{5.78}$$

Then, we can decompose r for the two groups of industries as

$$r = t(A_1' + A_2'\boldsymbol{D})Y_1^f + A_2'(Y_2 - \boldsymbol{D}Y_1) - I'G \tag{5.79}$$

Define $\boldsymbol{\Lambda}$ and $\boldsymbol{Y}$ as diagonal matrices containing $\frac{y_n}{\rho_n(1-\lambda_n)}$ and y_n as elements respectively, then we have $\boldsymbol{Y}^f = \boldsymbol{Y}F$ and similar to (5.54) we can find

$$Y_1^f = (Y_1 - \frac{\bar{g}}{2\theta}\boldsymbol{\Lambda}_1 G_1^{-1}) + \boldsymbol{A}_{11}\boldsymbol{\Gamma}_1 Y_1^f + \boldsymbol{A}_{12}\boldsymbol{\Gamma}_2 Y_2^f \tag{5.80}$$

where G_1^{-1} is a $M \times 1$ vector comprising $1/g_1, \ldots, 1/g_M$. Using (5.78), it turns out

$$Y_1^f = \Big(\boldsymbol{I} - \boldsymbol{A_{11}\Gamma_1} - \boldsymbol{A_{12}\Gamma_2 D}\Big)^{-1}\Big(\boldsymbol{Y}_1 - \frac{1}{2\theta}\boldsymbol{Y}_1\Phi_1^{-1} + \boldsymbol{A}_{12}\boldsymbol{\Gamma}_2(Y_2 - \boldsymbol{D}Y_1)\Big) \tag{5.81}$$

Substituting (5.81) in (5.79) yields

$$r = r_0 - \frac{t\bar{g}}{2\theta}HG^{-1} - I'G \tag{5.82}$$

where r_0 is a constant independent of G, and H is a row vector equal to

$$H = \big(A_1' + A_2'\boldsymbol{D}\big)\Big(\boldsymbol{I} - \boldsymbol{A_{11}\Gamma_1} - \boldsymbol{A_{12}\Gamma_2 D}\Big)^{-1}\boldsymbol{\Lambda}_1 \tag{5.83}$$

Thus, if h_m is the m^{th} element of H

$$r = r_0 - \sum_{m \leq M} \frac{t\bar{g}}{2\theta g_m} h_m - g_m - \sum_{k>M} g_k \tag{5.84}$$

and the FOC for g_m results

$$g_m^2 = \frac{\bar{g}}{2\theta} h_m \tag{5.85}$$

which can be rewritten in vector form as

$$G_1^2 = \frac{\bar{g}}{2\theta} H' \tag{5.86}$$

Which yields to (5.31).

5.F. Data Appendix

Table 5.F.1: Tax Buoyancy (TB) and formality estimates for Indian states – TB measures the percentage response of tax revenue to a one per cent change in the tax base, holding constant all parameters of tax policy. Formality index measures the percentage of production by tax registered firms averaged across services activities as listed in Table 5.F.2. TB data is drawn from Rangarajan et al. (2004, Ch.6, Page.383). Formality is estimated by the author using representative NSSO surveys.

state	TB	formality	state	TB	formality
Andhra Pradesh	1.2	33.52	Maharashtra	1.25	59.52
Arunachal Pradesh	1.1	25.14	Manipur	1.1	23.06
Assam	1.2	20.32	Meghalaya	1.2	19.29
Bihar	1.2	8.45	Mizoram	1.1	27.17
Chattisgarh	1.2	11.10	Nagaland	1.1	27.19
Goa	1.35	19.75	Orissa	1.2	8.54
Gujarat	1.3	43.54	Punjab	1.35	15.11
Haryana	1.25	16.55	Rajasthan	1.2	18.81
Himachal Pradesh	1.3	14.73	Sikkim	1.2	21.75
Jammu and Kashmir	1.2	2.83	Tamil Nadu	1.2	19.68
Jharkhand	1.2	14.46	Tripura	1.1	9.38
Karnataka	1.3	29.92	Uttar Pradesh	1.2	15.45
Kerala	1.3	23.18	Uttarkhand	1.2	5.35
Madhya Pradesh	1.2	24.43	West Bengal	1.35	41.25

Table 5.F.2: Summary statistics at industry level—The sample includes the services under tax rule before 2006, and is chosen based on various notifications of service tax rule, Government of India (for details see `http://www.servicetax.gov.in`). Services under financial intermediation and trade categories are not sampled in the round 57 of NSS survey and are not listed here.

NIC	Description	formality (%)	FL	BL
55204	Activities of caterers	4.27	1.34	2.23
60231	Freight transport by motor vehicles	23.39	1.90	2.33
60232	Freight transport other than by motor vehicles	4.77	1.90	2.33
61100	Sea and coastal water transport	37.50	1.96	1.89
61200	Inland water transport	1.03	1.96	1.89
63011	Cargo handling, incidental to land transport	34.03	2.07	1.98
63012	Cargo handling, incidental to water transport	79.99	2.07	1.98
63013	Cargo handling, incidental to air transport	51.02	2.07	1.98
63021	Warehousing of agricultural products without refrigeration	17.16	3.00	1.94
63022	Warehousing of agricultural products with refrigeration (i.e. cold storage)	34.62	3.00	1.94
63023	Storage and warehousing n.e.c.	22.76	3.00	1.94
63031	Supporting services to land transport	7.34	2.07	1.98
63032	Supporting services to water transport	21.91	2.07	1.98
63033	Supporting services to air transport	33.33	2.07	1.98
63040	Activities of travel agencies and tour operators n.e.c.	30.94	2.07	1.98
63090	Activities of other transport agencies	35.80	2.07	1.98
64110	National post activities	23.90	2.66	1.58
64120	Courier activities other than national post activities	31.51	2.66	1.58
64202	paging, e-mail, cellular phone, videoconferencing, internet	16.15	2.66	1.58
64203	Maintenance of telecom network	27.10	2.66	1.58
64204	Activities of the cable operators	31.04	2.66	1.58
70101	Purchase, sale letting of leased residential buildings	12.18	1.40	1.48
70102	Purchase, sale letting of leased non-residential buildings	19.69	1.40	1.48
70103	Operating of real estate of self owned residential buildings	0.00	1.40	1.48
70104	Operating of real estate of self owned non-residential buildings	20.17	1.40	1.48
70105	Developing and subdividing real estate into lots	38.10	1.40	1.48
70106	Lessors of real property	11.07	1.40	1.48
70109	Real estate activities with own or leased property n.e.c.	9.75	1.40	1.48
70200	Real estate activities on a fee or contract basis.	10.27	1.40	1.48
71110	Renting of land transport equipment	10.35	2.57	1.29
71301	Renting of tents, electrical appliances, furniture, table ware, crockery and utensils.	6.11	2.57	1.29
72294	Web-page designing	25.20	1.22	1.54
72295	Software maintenance	40.22	1.22	1.54
72400	Database activities and distribution of electronic content	15.93	1.22	1.54
72501	Repair & maintenance of computers and computer based systems	27.33	1.22	1.54
72502	Repair of office, computing and accounting machinery	20.00	1.22	1.54
72901	Activities of cyber cafe	18.32	1.22	1.54
72909	Other computer related activities, n.e.c.	21.59	1.22	1.54
73100	Research and experimental development on natural sciences and engineering	51.09	1.02	1.21
74120	Accounting, book-keeping and auditing activities;	28.90	3.11	2.12
74130	Market research and public opinion polling.	37.34	3.11	2.12
74140	Business and management consultancy activities.	23.78	3.11	2.12
74210	Architectural and engineering activities	25.71	3.11	2.12
74220	Technical testing and analysis.	29.64	3.11	2.12
74300	Advertising.	30.25	3.11	2.12
74910	Labour recruitment and provision of personnel.	17.40	3.11	2.12
74920	Investigation and security activities	27.51	3.11	2.12
74930	Building-cleaning and industrial cleaning activities	27.48	3.11	2.12
74940	Photographic activities.	10.51	3.11	2.12
74950	Packaging activities.	4.86	3.11	2.12
74991	Auctioning activities of self employed auctioneers.	28.57	3.11	2.12
74994	Fashion design	1.49	3.11	2.12
74996	Interior decoration	22.98	3.11	2.12
74997	Secretarial activities	19.40	3.11	2.12
80902	Coaching centres	14.42	1.02	1.21
80903	Activities of the individuals providing tuition	8.59	1.02	1.21
80904	Activities relating to training/ education/ conduct of specialised course	14.03	1.02	1.21
91990	Activities of other membership organizations n.e.c.	31.16	2.12	1.72
92111	Motion picture production	44.85	1.80	1.51
92113	Video production and distribution	9.54	1.80	1.51
92115	Sound recording studios	12.89	1.80	1.51
92131	Radio activities	0.00	1.80	1.51
92132	Television activities	19.26	1.80	1.51
92411	Operation and maintenance of sports facilities	15.98	1.80	1.51
92412	Activities of sports and game schools,	8.12	1.80	1.51
92413	Activities relating to organisation and operation of indoor / outdoor sports	1.16	1.80	1.51
92490	Other recreational activities	2.81	1.80	1.51
93010	Washing and (dry-) cleaning of textile and fur products	3.88	2.12	1.72
93030	Funeral and related activities	2.71	2.12	1.72

Bibliography

Aghion, P. and P. Bolton (1997). A theory of trickle-down growth and development. *The Review of Economic Studies 64*(2), 151–172.

Allingham, M. G. and A. Sandmo (1972, November). Income tax evasion: a theoretical analysis. *Journal of Public Economics 1*(3-4), 323–338.

Alok, S. and M. Ayyagari (2014). Politics, State Ownership, and Corporate Investment. *SSRN Working Paper (March 27, 2014)*.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax Compliance. *Journal of Economic Literature 36*(2), pp. 818–860.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arcand, J.-L., E. Berkes, and U. Panizza (2012). *Too much finance?* International Monetary Fund.

Ayyagari, M., T. Beck, and M. Hoseini (2013). Finance and Poverty: Evidence from India. *CEPR Discussion Paper 9497*.

Baldwin, R. (2007). EU VAT fraud. `http://www.voxeu.org/article/vat-fraud-part-1`.

Banerjee, A. V. and A. F. Newman (1993). Occupational Choice and the Process of Development. *Journal of Political Economy 101*(2), 274.

Baum, C. F., M. E. Schaffer, and S. Stillman (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal 7*(4), 465–506.

Baunsgaard, T. and M. Keen (2010). Tax revenue and (or?) trade liberalization. *Journal of Public Economics 94*(9), 563–577.

Beck, T., A. Demirgüç-Kunt, and R. Levine (2006). Bank supervision and corruption in lending. *Journal of Monetary Economics 53*(8), 2131–2163.

Beck, T., A. Demirgüç-Kunt, and V. Maksimovic (2005). Financial and legal constraints to growth: does firm size matter? *The Journal of Finance 60*(1), 137–177.

Beck, T. and A. Demirgüç-Kunt (2008). Access to finance: An unfinished agenda. *The World Bank Economic Review 22*(3), 383–396.

Beck, T., R. Levine, and A. Levkov (2010). Big bad banks? The winners and losers from bank deregulation in the United States. *The Journal of Finance 65*(5), 1637–1667.

Beck, T., R. Levine, and N. Loayza (2000). Finance and the sources of growth. *Journal of financial economics 58*(1), 261–300.

Beck, T., C. Lin, and Y. Ma (2014). Why do firms evade taxes? the role of information sharing and financial sector outreach. *The Journal of Finance 69*(2), 763–817.

Bednar, M. K. (2012). Watchdog or lapdog? A behavioral view of the media as a corporate governance mechanism. *Academy of Management Journal 55*(1), 131–150.

Beegle, K., R. H. Dehejia, and R. Gatti (2006). Child labor and agricultural shocks. *Journal of Development Economics 81*(1), 80–96.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics 119*(1).

Besley, T. and R. Burgess (2000). Land reform, poverty reduction, and growth: evidence from India. *Quarterly Journal of Economics*, 389–430.

Besley, T. and R. Burgess (2002). The political economy of government responsiveness: Theory and evidence from India. *Quarterly Journal of Economics*, 1415–1451.

Besley, T. and R. Burgess (2004). Can labor regulation hinder economic performance? Evidence from India. *The Quarterly Journal of Economics*, 91–134.

Besley, T., R. Burgess, B. Esteve-Volart, and J. Louise (2007). The policy origins of poverty and growth in India. *Delivering on the Promise of Pro-Poor Growth: Insights and Lessons from Country Experience*.

Boadway, R. and M. Sato (2009). Optimal tax design and enforcement with an informal sector. *American Economic Journal: Economic Policy 1*(1), 27.

Bosworth, B., S. M. Collins, and A. Virmani (2007). Sources of growth in the indian economy. Technical report, National Bureau of Economic Research.

Botero, J., S. Djankov, R. Porta, and F. Lopez-de Silanes (2004). The regulation of labor. *The Quarterly Journal of Economics 119*(4), 1339–1382.

Bruhn, M. and I. Love (2014). The real impact of improved access to finance: Evidence from mexico. *The Journal of Finance 69*(3), 1347–1376.

Bruttel, L. and T. Friehe (2014). On the path dependence of tax compliance. *European Economic Review 65*, 90107.

Burgess, R. and R. Pande (2005). Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment. *American Economic Review*, 780–795.

Bushman, R. M., C. D. Williams, and R. Wittenberg-Moerman (2013). The Informational Role of the Media in Private Lending. *Ann Arbor 1001*, 48109.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Clarke, G. R. G., L. C. Xu, and H.-f. Zou (2006). Finance and income inequality: what do the data tell us? *Southern economic journal*, 578–596.

Dabla-Norris, E., M. Gradstein, and G. Inchauste (2008). What causes firms to hide output? the determinants of informality. *Journal of Development Economics 85*(1), 1–27.

Dai, Z., R. M. Hogarth, and M. Villeval (2015). Ambiguity on audits and cooperation in a public goods game. *European Economic Review 74*, 146162.

De Mel, S., D. McKenzie, and C. Woodruff (2008). Returns to capital in microenterprises: evidence from a field experiment. *The Quarterly Journal of Economics 123*(4), 1329–1372.

De Paula, A. and J. A. Scheinkman (2010, October). Value-Added Taxes, Chain Effects, and Informality. *American Economic Journal: Macroeconomics 2*(4), 195–221.

De Soto, H. (1990). *The other path: The invisible revolution in the third world*. Perennial Library.

Deaton, A. (2003). Adjusted Indian poverty estimates for 1999-2000. *Economic and Political Weekly*, 322–326.

Degryse, H. and S. Ongena (2005). Distance, lending relationships, and competition. *The Journal of Finance 60*(1), 231–266.

Dharmapala, D., J. Slemrod, and J. D. Wilson (2011). Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs. *Journal of Public Economics 95*(9-10), 1036–1047.

Djankov, S., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2003). Courts. *The Quarterly Journal of Economics*, 453–517.

Dougherty, S., V. C. F. Robles, and K. Krishna (2011). Employment protection legislation and plant-level productivity in India. Technical report, National Bureau of Economic Research.

Dyck, A. and L. Zingales (2004). Private benefits of control: An international comparison. *The Journal of Finance 59*(2), 537–600.

Ebrill, L. P., M. Keen, J.-P. Bodin, and V. P. Summers (2001). *The Modern VAT*, Volume 1. International Monetary Fund.

Edmonds, E., N. Pavcnik, and P. Topalova (2010). Child Labor and Schooling in a Globalized World: Some Evidence from Urban India. *Journal of European Economic Association* (7), 498–507.

Egger, P., K. Erhardt, and C. Keuschnigg (2014). Personal Income Taxes, Corporate Prot Taxes and the Heterogeneous Tax Sensitivity of Firm-level Investments. Working paper.

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

Engel, E. M. R. A. and J. R. Hines Jr (1999). Understanding tax evasion dynamics. Working paper, National Bureau of Economic Research.

Engelberg, J. E. and C. A. Parsons (2011). The causal impact of media in financial markets. *The Journal of Finance 66*(1), 67–97.

Feinstein, J. S. (1991). An econometric analysis of income tax evasion and its detection. *The RAND Journal of Economics*, 14–35.

Flug, K., A. Spilimbergo, and E. Wachtenheim (1998). Investment in education: do economic volatility and credit constraints matter? *Journal of Development Economics 55*(2), 465–481.

Friedman, E., S. Johnson, D. Kaufmann, and P. Zoido-Lobaton (2000). Dodging the grabbing hand: the determinants of unofficial activity in 69 countries. *Journal of public economics 76*(3), 459–493.

Frye, T. and E. Zhuravskaya (2000). Rackets, regulation, and the rule of law. *Journal of Law, Economics, and Organization 16*(2), 478–502.

Galor, O. and J. Zeira (1993). Income distribution and macroeconomics. *The review of economic studies 60*(1), 35–52.

Ghosh, B. and P. De (2005). *India infrastructure database 2005*, Volume 1. Bookwell.

Giné, X. and R. M. Townsend (2004). Evaluation of financial liberalization: a general equilibrium model with constrained occupation choice. *Journal of Development Economics 74*(2), 269–307.

Gordon, J. (1989). Individual morality and reputation costs as deterrents to tax evasion. *European economic review 33*, 797–805.

Gordon, R. and W. Li (2009, August). Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics 93*(7-8), 855–866.

Greenwood, J. and B. Jovanovic (1990). Financial Development, Growth, and the Distribution of Income. *The Journal of Political Economy 98*(5 Part 1), 1076–1107.

Griffin, J. M., N. H. Hirschey, and P. J. Kelly (2011). How important is the financial media in global markets? *Review of Financial Studies*, hhr099.

Guarcello, L., F. Mealli, and F. C. Rosati (2010). Household vulnerability and child labor: the effect of shocks, credit rationing, and insurance. *Journal of population economics 23*(1), 169–198.

Gupta, P., R. Hasan, and U. Kumar (2009). Big reforms but small payoffs: explaining the weak record of growth in Indian manufacturing. In *India Policy Forum*, Volume 5, pp. 59–123. The World Bank.

Hanson, J. A. (2001). Indian Banking: Market liberalization and the pressures for institutional and market framework reform. *Center for Research on Economic Development and Policy Reform, Working Paper* (104).

Hasan, R., D. Mitra, and K. V. Ramaswamy (2007). Trade reforms, labor regulations, and labor-demand elasticities: Empirical evidence from India. *The Review of Economics and Statistics 89*(3), 466–481.

Hashimzade, N., G. D. Myles, and M. D. Rablen (2014). The Use of Predictive Analytics to Increase Tax Compliance. mimeo, presented in CEBID taxation, social norm and compliance conference.

Hashimzade, N., G. D. Myles, and B. Tran-Nam (2013). Applications of behavioural economics to tax evasion. *Journal of Economic Surveys 27*(5), 941–977.

Honohan, P. (2004). *Financial sector policy and the poor: Selected findings and issues*. Number 43. World Bank Publications.

Hoseini, M. (2014a). Misreporting in the Value-Added Tax and the Optimal Enforcement Policy. *Tilburg University, mimeo*.

Hoseini, M. (2014b). Value-Added Tax and Shadow Economy: the Role of Inter-sectoral Linkages. *CentER Discussion Paper Series No. 2013*(36).

Hsieh, C.-T. and B. A. Olken (2014). The missing "missing middle". Technical report, National Bureau of Economic Research.

Iyer, L. and A. Banerjee (2005). History, institutions, and economic performance: The legacy of colonial land tenure systems in india. *American economic review 95*(3), 1190–1213.

Iyer, L., A. Mani, P. Mishra, and P. Topalova (2012). The power of political voice: women's political representation and crime in India. *American Economic Journal: Applied Economics 4*(4), 165–193.

Jacoby, H. G. and E. Skoufias (1997). Risk, financial markets, and human capital in a developing country. *The Review of Economic Studies 64*(3), 311–335.

Joe, J. R., H. Louis, and D. Robinson (2009). Managers and investors responses to media exposure of board ineffectiveness. *Journal of Financial and Quantitative Analysis 44*(03), 579–605.

Johns, A. and J. Slemrod (2010). The distribution of income tax noncompliance. *National Tax Journal 63*(3), 397–418.

Johnson, S., D. Kaufmann, J. McMillan, and C. Woodruff (2000). Why do firms hide? bribes and unofficial activity after communism. *Journal of Public Economics 76*(3), 495–520.

Johnson, S., D. Kaufmann, and P. Zoido-Lobaton (1998). Regulatory discretion and the unofficial economy. *American Economic Review*, 387–392.

Joshi, V. and I. Little (1996). *India's Economic Reforms: 1991-2001*, Volume 1. Oxford, Clarendon Press.

Kastlunger, B., S. G. Dressler, E. Kirchler, L. Mittone, and M. Voracek (2010). Sex differences in tax compliance: Differentiating between demographic sex, gender-role orientation, and prenatal masculinization (2D: 4D). *Journal of Economic Psychology 31*(4), 542–552.

Keen, M. (2008, October). VAT, tariffs, and withholding: Border taxes and informality in developing countries. *Journal of Public Economics 92*(10-11), 1892–1906.

Keen, M. and B. Lockwood (2010). The value added tax: Its causes and consequences. *Journal of Development Economics 92*(2), 138–151.

Keen, M. and J. Mintz (2004). The optimal threshold for a value-added tax. *Journal of Public Economics 88*(3), 559–576.

Keen, M. and S. Smith (2006). VAT Fraud and Evasion: What do we know, and what can be done? *National Tax Journal LIX*(4).

King, R. G. and R. Levine (1993). Finance and Growth: Schumpeter Might be Right. *The Quarterly Journal of Economics 108*(3), 717–737.

Klapper, L., L. Laeven, and R. Rajan (2006). Entry regulation as a barrier to entrepreneurship. *Journal of Financial Economics 82*(3), 591–629.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. r. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica 79*(3), 651–692.

Kleven, H. J., C. T. Kreiner, and E. Saez (2009, August). Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries. Working Paper 15218, National Bureau of Economic Research.

Kochar, A. (2011). The Distributive Consequences of Social Banking: A Micro-empirical Analysis of the Indian Experience. *Economic Development and Cultural Change 59*, 251–280.

Kopczuk, W. and J. Slemrod (2006, May). Putting Firms into Optimal Tax Theory. *American Economic Review: Papers and Proceedings 96*(2), 130–134.

Kuhnen, C. M. and A. Niessen (2012). Public opinion and executive compensation. *Management Science 58*(7), 1249–1272.

Lamoreaux, N. R. (1986). Banks, kinship, and economic development: the New England case. *The Journal of Economic History 46*(03), 647–667.

LaPorta, R. and A. Shleifer (2014). Informality and development. *Journal of Economic Perspectives*.

Levine, R. (2005). Finance and growth: Theory and evidence. *Handbook of economic growth 1*(September), 0–117.

Li, H., L. Squire, and H.-f. Zou (1998). Explaining international and intertemporal variations in income inequality. *The Economic Journal 108*(446), 26–43.

Li, H., L. C. Xu, and H.-f. Zou (2000). Corruption, income distribution, and growth. *Economics & Politics 12*(2), 155–182.

Loayza, N. V. (1996). The economics of the informal sector: a simple model and some empirical evidence from latin america. In *Carnegie-Rochester Conference Series on Public Policy*, Volume 45, pp. 129–162. Elsevier.

Marjit, S., S. Kar, and D. S. Maiti (2007). Regional trade openness index and income disparity: a new methodology and the Indian experiment. *Economic and Political Weekly*, 757–769.

Mastrobuoni, G. (2014). Crime is terribly revealing: Information technology and police productivity. *Working Paper*.

McKinnon, R. I. (1973). *Money and capital in economic development*. Brookings Institution Press.

Özler, B., G. Datt, and M. Ravallion (1996). A database on poverty and growth in India. *Poverty and Human Resources Division Policy Research Department, World Bank. Washington, DC*.

Panagariya, A. and M. Mukim (2014). A comprehensive analysis of poverty in India. *Asian Development Review 31*(1), 1–52.

Pande, R. (2003). Can mandated political representation increase policy influence for disadvantaged minorities? Theory and evidence from India. *The American Economic Review 93*(4), 1132–1151.

Paramonova, Y. (2014). The Optimal Information Reporting and Tax Audit Policy.

Perotti, E. and P. Volpin (2014). Investor protection and entry. *Econometrica 82*(2), 705–30.

Perry, G. (2007). *Informality: Exit and exclusion*. World Bank Publications.

Piggott, J. and J. Whalley (2001). VAT Base Broadening, Self Supply, and the Informal Sector. *American Economic Review 91*(4), 1084–1094.

Planning Commission (1993). Report of the expert group on estimation of proportion and number of poor. *New Delhi: Government of India*.

Pomeranz, D. (2013, July). No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. Working Paper 19199, National Bureau of Economic Research.

Radhakrishna, R., S. Sengupta, and S. Tendulkar (2011). Report of the expert group to review the methodology for estimation of poverty. Technical report, Planning Commission, Government of India.

Rajan, R. G. and L. Zingales (1998). Financial Dependence and Growth. *American Economic Review*, 559–586.

Rangarajan, C., T. R. Prasad, and D. K. Srivastava (2004). *Report of the Twelfth Finance Commission (2005-10)*. Finance Commission India.

Rao, M. G. and R. K. Rao (2006). Trends and issues in tax policy and reform in India. *India Policy Forum 2005-06*, 55–122.

Rasmussen, N. (1957). Studies in Inter-Sectoral Relations. *Revue Économique 8*(6), 1103–1104.

Reddy, Y. V. (2002). Monetary and financial sector reforms in India: a practitioner's perspective. *in India's Emerging Economy: Performance and Prospects in the 1990s and Beyond ed. Kaushik Basu, MIT Press*.

Reinganum, J. F. and L. L. Wilde (1985). Income tax compliance in a principal-agent framework. *Journal of Public Economics 26*(1), 1–18.

Reserve Bank of India (2004). Report on Trend and Progress of Banking in India 2003-04. Technical report, Reserve Bank of India Publication.

Schneider, F., A. Buehn, and C. E. Montenegro (2011). Shadow Economies all over the World: New Estimates for 162 Countries from 1999 to 2007. *Handbook on the shadow economy*, 9–77.

Schneider, F. and D. H. Enste (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature 38*, 77–114.

Sen, K., R. R. Vaidya, and V. Sen (1997). *The process of financial liberalization in India*. Oxford University Press New Delhi.

Shaw, J., J. Slemrod, and J. Whiting (2010). Administration and compliance. *Dimensions of Tax Design: the Mirrlees Review, J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba (eds), Oxford University Press*.

Shleifer, A. and R. W. Vishny (1993). Corruption. *The Quarterly Journal of Economics 108*(3), 599–617.

Shleifer, A. and R. W. Vishny (1994). Politicians and firms. *The Quarterly Journal of Economics*, 995–1025.

Slemrod, J. and C. Gillitzer (2014). *Tax Systems*, Volume 1. The MIT Press.

Slemrod, J. and S. Yitzhaki (2002). Tax avoidance, evasion, and administration. *Handbook of public economics 3*, 1423–1470.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear IV regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg 1*.

Straub, S. (2005). Informal sector: the credit market channel. *Journal of Development Economics 78*(2), 299–321.

Tait, A. A. (1991). *Value-added Tax: Administrative and Policy Issues*. International Monetary Fund.

Tarozzi, A. (2007). Calculating comparable statistics from incomparable surveys, with an application to poverty in India. *Journal of business & economic statistics 25*(3).

Ten Raa, T. (2006). *The economics of input-output analysis*. Cambridge University Press.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance 62*(3), 1139–1168.

Tetlock, P. C., M. Saar-tsechansky, and S. Macskassy (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance 63*(3), 1437–1467.

Topalova, P. (2010). Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India. *American Economic Journal: Applied Economics 2*(4), 1–41.

Visaria, S. (2009). Legal reform and loan repayment: The microeconomic impact of debt recovery tribunals in India. *American Economic Journal: Applied Economics*, 59–81.

Witte, A. D. and D. F. Woodbury (1985). The effect of tax laws and tax administration on tax compliance: The case of the US individual income tax. *National Tax Journal*, 1–13.

Wu, R.-S., C.-S. Ou, H. ying Lin, S.-I. Chang, and D. C. Yen (2012). Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications 39*(10), 8769–8777.