ESSAYS ON THE ECONOMICS OF THE INFORMAL SECTOR IN INDIA

by

ABHINAV NARAYANAN
(Under the Direction of Santanu Chatterjee)

Abstract

As of 2011-12, informal workers in India accounted for 85.8 percent of the labor force, while the informal sector contributed nearly 55 percent to the GDP. This dissertation presents three essays on the economics of the informal sector in India. In Chapter 2, I use Indian labor market data to test whether workers are able to self-select into formal and informal employment. I find no evidence of workers being able to self-select into formal employment. Thus, workers may face entry barriers that restrict them from entering formal employment and they choose informal employment as a last resort. In Chapter 3, we use firm-level data on formal and informal production in the manufacturing sector in India to examine the sectoral consequences of government investment in public infrastructure. The average output elasticity of the flow of public investment for an informal sector firm is 3 times smaller than its formal counterpart. For the accumulated stock of public capital, this difference increases to a factor of 7. In Chapter 4, I set up a two-sector endogenous growth model incorporating the informal sector to examine the implications of efficiency of public capital on the optimal tax rate, and on the optimal allocation of tax revenues toward maintenance of public capital. Results show that in presence of the informal sector, the growth-maximizing tax rate in the decentralized economy is lower, and the welfare-maximizing optimal share of maintenance expenditure is shown to be positively related to the informal to formal output ratio.

INDEX WORDS: India, Informal employment, Informal sector, Segmentation, Selection bias, Public investment, Output elasticity, Quantile regressions, Efficiency, Tax rate, Maintenance

ESSAYS ON THE ECONOMICS OF THE INFORMAL SECTOR IN INDIA

by

ABHINAV NARAYANAN

B.A., Jadavpur University, India, 2007

M.A., Jawaharlal Nehru University, India, 2009

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2016

ESSAYS ON THE ECONOMICS OF THE INFORMAL SECTOR IN INDIA

by

ABHINAV NARAYANAN

Approved:

Major Professor: Santanu Chatterjee

Committee: Julio Garín
David B. Mustard
Ian Schmutte

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2016

To my *Parents* and *Emon*

Acknowledgments

This dissertation is a product of a collective effort. Below is just a partial list of the people, events and circumstances that contributed toward completing this dissertation.

First, I would like to thank Dr. Santanu Chatterjee for convincing me to join the University of Georgia- a decision I will cherish for the rest of my life. Dr. Chatterjee is also my Ph.D advisor, and any amount of words would be less to describe his contribution toward this dissertation. His critical and insightful feedback had a tremendous effect on both the quality and impact of my research. His dedication and tenacity toward research, his compassion and patience toward his students have taught me to grow as an academician. I hope I can meet his expectations during my tenure as his Ph.D student.

I would like to acknowledge my committee members Drs. Julio Garín, David Mustard, and Ian Schmutte for their constructive feedback on my research, and their constant support and encouragement without which this dissertation would not have been possible. I would also like to thank the immensely supportive faculty and staff of the entire Economics Department. Particularly, I would like to thank Drs. John Turner, William Lastrapes, Meghan Skira, Laura Zimmerman, Svetlana Pashchenko, and Ronald Warren for their comments and feedback and for helping me to prepare for the job market.

I am indebted to my friends and ex-roommates Arnab and Dipesh for taking the pain of proof-reading my papers. I owe a lot to the endless banters we shared that have helped me to relieve a lot a stress while writing this dissertation. A special thanks to Nabaneeta and Thomas for being available and patient enough to help solve any problem that I got stuck with. I would like to acknowledge Dr. Saumen Chattopadhyay for his insightful feedback on my earlier drafts. I would also like to thank a bunch of awesome friends in Athens –Anandi, Shuvankar, Debkanta, Roshini, Pradeep, Arunava and, Dipanwita for their various forms of support during all these years. I would like to acknowledge my friends- Diptayan, Suvadip, Gogol and Rishabh for always standing by my side during these years.

I cannot express in words my family's unmeasurable love and support without which this dissertation would never have been possible. I would like to thank my parents Seba Narayanan and P.R Narayanan for their sacrifice, hard work, staying power, and for making sure I never face any obstacles toward achieving my dreams. I would also like to thank my extended family members for their constant support and encouragement all through these years without which it wouldn't have been possible.

My wife, Emon, has been my strength. This dissertation wouldn't be complete without the motivation, inspiration and enthusiasm she has provided me during these five years. I owe a lot to our endless discussions and debates while writing this dissertation. I am indebted to her for her never ending love and encouragement.

Lastly, I would like to thank The University of Georgia Graduate School for providing me funding through the Dean's Award, and to the South Asia Research Network for their research grant through their Young South Asian Scholar's program that supported my research.

Contents

Acknowledgments **v**

List of Figures **viii**

List of Tables **x**

1 Introduction and Literature Review **1**

2 Understanding Informal Employment in India: Competitive Choice or a Result of Labor Market Segmentation? **10**

2.1 Introduction 10

2.2 Literature Review 14

2.3 Definitions and Conceptual Framework 16

2.4 Data and Summary Statistics 16

2.5 Empirical Methods 24

2.6 Results 35

2.7 Conclusions 44

Appendix 1 58

3 The Spillover Effects of Public Investment: Implications for Formal and Informal Sector Firms in India **60**

3.1 Introduction 60

3.2 Data 65

3.3 Empirical Specification 69

3.4 Results 73

3.5 Conclusions 80

Appendix 2 97

4 Informality, Congestion, and Public Capital Efficiency: A Case for Optimal Taxation and Maintenance Allocation 106

4.1 Introduction 106

4.2 Analytical Framework 111

4.3 A Centrally Planned Economy 115

4.4 A Decentralized Economy 117

4.5 Optimal Fiscal Policy 122

4.6 Conclusions 124

Appendix 3 129

5 Concluding Remarks 131

List of Figures

2.1 Share of Informal Employment in the Non-Agricultural Sector, by Countries 57

2.2 Predicted Formal Wage for Informal Workers (Males) 57

2.3 Predicted Formal Wage for Informal Workers (Females) 57

3.1 Sectoral Capital Intensity, 1999 and 2010 92

3.2 Sectoral Output Per Worker, 1999 and 2010 92

3.3 Share of Informal Sector in GDP, 1999-2010 93

3.4 Share of Infrastructure Spending in GDP, 2006-2010 93

3.5 Quantile Output Elasticity of Public Investment, Formal Sector 94

3.6 Quantile Output Elasticity: Economic Services, Formal Sector 94

3.7 Quantile Output Elasticity: Social Services, Formal Sector 94

3.8 Quantile Output Elasticity of Public Investment, Informal Sector 95

3.9 Quantile Output Elasticity: Economic Services, Informal Sector 95

3.10 Quantile Output Elasticity: Social Services, Informal Sector 95

3.11 Public Investment and Firm-level Capital Intensity, Formal Sector 96

3.12 Public Investment and Firm-level Capital Intensity, Informal Sector 96

A.2.1 Quantile Output Elasticities for Transport and Energy Expenditures, Formal Sector 100

A.2.2 Quantile Output Elasticities for Transport and Energy Expenditures, Informal Sector 100

A.2.3 Economic Services Expenditures and Firm-level Capital Intensity, Formal Sector . . 101

A.2.4 Social Services Expenditures and Firm-level Capital Intensity, Formal Sector 101

A.2.5 Economic Services Expenditures and Firm-level Capital Intensity, Informal Sector . 102

A.2.6 Social Services Expenditures and Firm-level Capital Intensity, Informal Sector . . . 102

A.2.7 Average State-wise Public Development Expenditures, 2006-10 (in 2004-05 Rs) . . . 103

A.2.8 Average State-wise Public Development Expenditures, 2006-10 (Share of NSDP) . . 103

4.1 Road Infrastructure Investment and Maintenance (% of Total spending), Average 2006-2010 126

4.2 Road Investment, Maintenance and Informal Sector, India, 2004-2011 126

4.3 Informal Sector Employment (% of Non-Agricultural Employment) vs No. of Power Outages in a Firm in a Typical Month . . . 127
4.4 Informal Sector Employment (% of Non-Agricultural Employment) vs Value Lost Due to Electrical Outages (% of Sales) . . . 127
4.5 Informal Sector Employment (% of Non-Agricultural Employment) vs PIMI scores, cross country . . . 128

List of Tables

2.1 Classification of Informal Employment and Informal Sector 47

2.2 Identification of Formal/Informal Sector, by Type of Enterprise 47

2.3 Identifying Broad Categories of Employment, by Principal Activity Status 47

2.4 Percentage of Workers Receiving Employment Benefits 48

2.5 Formal and Informal Employment (Treating Self Employed as a Separate Category) 48

2.6 Identification of Formal and Informal Sector and Employment Based on Enterprise Type and Activity Status 48

2.7 Formal and Informal Employment in the Formal and Informal Sectors 48

2.8 Summary Statistics (Males) 49

2.9 Summary statistics (Females) 50

2.10 Wage Gap Decomposition (Males) 51

2.11 Wage Gap Decomposition (Females) 51

2.12 Multinomial Logit Estimates of the Sectoral Allocation Equations (Males) 52

2.13 Multinomial Logit Estimates of the Sectoral Allocation Equations (Females) 53

2.14 Selection Corrected Wage Estimates for Formal and Informal Workers (Males) . . . 54

2.15 Selection Corrected Wage Estimates for Formal and Informal Workers (Females) . . 55

2.16 Oaxaca-Blinder Decomposition 56

3.1 Summary Statistics for Formal and Informal Sectors, 2010 83

3.2 Average State-wise Public Development Expenditures (2004-05 Rs) 84

3.3 State-wise Public Development Expenditures (2004-05 Rs) 85

3.4 OLS Estimation of Production Function, Formal Sector with Flow Measure of Public Investment 86

3.5 LP-S Estimation of Production Function, Formal Sector with Flow Measure of Public Investment 87

3.6 LP-S Estimation of Production Function, Formal Sector with Stock Measure of Public Investment 88

3.7 OLS Estimation of Production Function, Informal Sector with Flow Measure of Public Investment 89
3.8 LP-S Estimation of Production Function, Informal Sector with Flow Measure of Public Investment 90
3.9 LP-S Estimation of Production Function, Informal Sector with Stock Measure of Public Investment 91
A.2.1 Data Sources 97
A.2.2 Summary Statistics for Formal and Informal Sectors, 1999 97
A.2.3 OLS Estimation of Production Function, Formal Sector with Flow Measure of Public Investment 98
A.2.4 OLS Estimation of Production Function, Informal Sector with Flow Measure of Public Investment 99
A.2.5 Firm Age Statistics (Formal Sector) 105
A.2.6 Output Elasticities of Public Investment across Firm Age (LP-S method) 105

Chapter 1

Introduction and Literature Review

This dissertation presents three essays on the economics of the informal sector in developing countries with a special focus on India. Informal production is a pervasive feature of most developing countries. As such, this sector consists of small, unregistered firms that typically produce very labor intensive non-traded goods and services, with little or no access to capital markets, and very limited outward labor mobility to the formal or organized sector (LaPorta and Shleifer (2014)). However, this sector plays an important role in the structural evolution of these countries, accounting for about 42 percent of GDP, and absorbing between $48 - 54$ percent of the labor force (Schneider et al. (2010)). Despite being a high-growth emerging market, the Indian economy is largely informal, with this sector contributing to 55 percent of GDP and employing about 84 percent of the non-agricultural labor force in 2010 (ILO (2013)).

There are two broad views that explain the large share of informal employment in developing countries: labor market segmentation and the competitive choice. According to the labor market segmentation hypothesis, informal workers are disadvantaged and are waiting to be employed as formal workers (Lewis, 1954 and Harris and Todaro, 1970). Employers ration formal jobs that results in a queue for these jobs. Institutional barriers also restrict workers from entering formal employment. In the absence of such entry barriers, a worker would choose to work in the sector that pays higher wages and other non-wage benefits. In the labor economics literature, this line of argument is commonly known as labor market segmentation. A similar line of argument is put forward by efficiency wage proponents (Stiglitz, 1981 and Solow, 1980). According to the efficiency wage argument, formal wages are set higher than the market-clearing rate to induce worker productivity, which in turn creates segments within the labor market. Other factors that may lead to labor market segmentation include search frictions (Burdett and Mortensen, 1998) and monopsonistic power of firms (Ashenfelter et al., 2010 and Alan, 2011).

In Chapter 2, I use India's National Sample Survey data on Employment and Unemployment (2011-12) to examine the underlying factors that are associated with the large share of informal employment in India. Specifically, there are two objectives of this paper. First, I estimate the formal-informal wage gap across different quantiles of the wage distribution and decompose this gap into coefficient and endowment effects. The coefficient effect explains what fraction of the overall wage gap is due to the difference in returns to human capital and individual characteristics of formal and informal workers. The endowment effect explains what fraction of the wage gap is due to the difference in human capital and characteristics between formal and informal workers. Second, I analyze whether the large share of informal employment in India is driven by competitive choice or labor market segmentation.

In this paper, I use the quantile regression framework that decomposes the formal-informal wage gap into coefficient and endowment effects across the wage distribution (Machado and Mata, 2005). Then I use the polychotomous choice model developed by Lee (1983) to test for labor market segmentation. The methodology is an extension of Heckman (1979) that allows for multiple labor market choices. In the first stage, I use the the multinomial logit to estimate the participation decision that includes four labor market choices: formal employment, informal employment, self-employment and staying out of the labor force. In the second stage, I estimate the wage equations for formal and informal workers taking into account the sample selection bias resulting from self-selection of workers into formal and informal employment. Gindling (1991) argues that a non-random selection of workers into a particular sector that does not affect wages in that sector implies that workers do not have full access to that sector.

The results presented here support the segmentation hypothesis for the Indian labor market. I find a significant wage gap between formal and informal workers across the wage distribution. At the lower end of the wage distribution, differences in returns to human capital and individual characteristics between formal and informal workers explain a major part of the wage gap. The informal workers at the lower end of the wage distribution may be identified as the disadvantaged workers who earn lower returns to their skills compared to their formal counterparts. At the higher end of the distribution, the major part of the wage gap is explained by differences in human capital and individual characteristics between formal and informal workers. Second, results from the polychotomous choice model show no evidence of workers being able to self-select into formal employment. Thus, we cannot reject the labor market segmentation hypothesis. The counterfactual wages show that 85 percent of the male and 83 percent of the female informal workers would have earned higher wages if they were formal workers.

Given capital and labor market rigidities, informal firms may have to rely heavily on government-provided investment goods such as transportation, power, water, etc. for production purposes. However, very little, if anything, is known about the benefits of government investment (and the resulting stock of public capital) for informal production in developing countries. In Chapter 3, we use two large firm-level datasets on formal and informal production in the manufacturing sector in India to examine the sectoral consequences of government investment in public infrastructure.

In this paper, we attempt to bridge a gap between two strands of research that have evolved largely independently of each other. On the one hand, starting with the work of Aschauer (1989b), a voluminous empirical literature has explored the productivity benefits of public investment in infrastructure, with a rich diversity of results. However, these studies have, without exception, considered the consequences for either industrialized countries (where the share of informal production is relatively small), or only for the formal sector in developing countries. On the other hand, the literature on the informal sector has mainly focused on issues of measurement of its output share (Schneider and Enste (2000), LaPorta and Shleifer (2014, 2008), and Gomis-Porqueras et al. (2014)), or issues pertaining to tax policy and enforcement (Rauch (1991), Ihrig and Moe (2004), Turnovsky and Basher (2009), Prado (2011), and Ordonez (2014)). The importance of public investment for this type of production has generally been ignored. Consequently, by examining the benefits of government investment expenditures for private production in the formal and informal sectors, we seek to fill an important gap in this literature. This is the first contribution of our paper. Second, while most studies on public investment are conducted at a fairly aggregated level (at the level of a country, state or region), we attempt to estimate its sectoral productivity benefits at the level of the individual firm. In the case of India, for example, while Binswanger et al. (1993), Lall (1999), Mitra et al. (2002), Zhang (2004), and Hulten et al. (2006), among others, have examined the effects of public infrastructure for the formal sector at the state, district, or industry level, there is no current evidence of its sectoral importance at the level of the firm. The firm-level datasets we use for our study enable us to shed light on the role of public investment and infrastructure at a much more disaggregated level than previously studied. We view this as an additional contribution to the literature. Finally, from the perspective of designing public policy, it is important to know how the spillovers from public investment are dispersed over the size distribution of firms in each sector. In other words, do larger firms tend to benefit more or less relative to their smaller counterparts from government spending on public goods? This may help determine how public goods should be targeted to firms in each sector. To the best of our knowledge, our analysis is the first to shed light on this issue.

In India, the main source of information at the firm level for the formal sector is the Annual Survey of Industries (ASI), while for the informal sector it is the surveys conducted by the National Sample Survey Organization (NSSO). Though the ASI surveys firms on an annual basis, the NSSO survey is conducted once every 10 years. We use data from the 2010 round for each of these surveys, since that is the latest round for which firm-level information is available for *both* sectors. Restricting our coverage to only the manufacturing sector, we obtain a cross-section of $32,388$ formal-sector firms (from the ASI) and $82,748$ informal-sector firms (from the NSSO) for 2010. We proxy public investment by state-level data on government *Development Expenditures*, obtained from the Reserve Bank of India. Here, we consider two sub-categories of expenditures: (i) *Economic Services*, which include public expenditures on transport, communications, and energy, and (ii) *Social Services*, which include expenditures on health, education, water, sanitation, and other welfare programs, and construct measures of both the *flow* of public investment, using average annual expenditures over the $2006 - 2010$ period, as well as its accumulated *stock* at the per-capita level for each state, using data over the period $2000 - 2010$. The flow measure is intended to capture the short-term effects of public investment, while the stock measure captures its effects over the longer term. Henceforth, we will interchangeably refer to the broad category of Development Expenditures as *public investment*, and the corresponding stock measure as *public capital*.

Our empirical strategy can be described as follows. First, we estimate the output elasticities of the flow of public investment and the accumulated stock of public capital at the firm level in the formal and informal sector. While this gives us information on how the average firm in each sector is affected by public investment, it masks the distribution of the sectoral elasticities across firms. We therefore employ quantile regressions (QR) to examine how the sectoral output elasticities vary across the size distribution of firms. Further, we also examine the relationship between public investment and the capital intensity in production across this size distribution. The QR approach is of critical importance from the policy perspective, since public investment may serve as a potential mechanism through which the government may aim to reduce not only the size of the informal sector, but also increase the relative usage of private capital in that sector.

The empirical analysis also raises some important econometric issues. First, it is plausible that the inclusion of public investment generates a reverse causality problem with output. Using firm-level data along with state-level government expenditures helps alleviate this problem, as it is unlikely that an individual firm will have any systematic effect on public spending at the state level. Second, the usage of private inputs like capital and labor may be endogenous to the firm's decision to produce output. Here, we use a method suggested by Levinsohn and Petrin (2003) and Sivadasan (2009) that uses past values of intermediate inputs and exploits the repeated cross-sectional nature of our dataset to control for the unobserved productivity shock at the firm-level in each sector.

Our results indicate that though public investment has a positive and statistically significant effect on both formal and informal sector firms,there are important sectoral and distributional consequences. With the flow specification of public investment, we estimate an output elasticity of 0.088 for formal sector firms. The corresponding output elasticity for the average informal sector firm is about 3 times lower, at 0.027. When we consider the stock of public capital, the difference in sectoral output elasticities is much larger, with the estimate for formal firms at about 0.17, about 7 times larger relative to their informal counterparts. Since the stock measure of public investment is intended to capture its long term productivity spillovers, these results suggest that the benefits accruing to formal sector firms from the accumulated stock of public capital are much larger relative to those for informal sector firms.Within the sub-categories of public investment, we find that Economic Services is associated with higher productivity spillovers relative to Social Services in both sectors, irrespective of whether we use the stock or flow specification.

While these results only represent the effects associated with the *average* firm in each sector, it is important to consider how the estimated output elasticity of public investment varies across the size distribution of firms (measured by their gross value added). For the formal sector, we find that there is very little variation in the output elasticity of public investment across the size distribution of firms. By contrast, for the informal sector this association is strictly positive across the entire size distribution of firms. Another important issue is whether public investment influences the relative capital intensity of firms in each sector. This is especially relevant for firms in the informal sector, who tend to have extremely low capital-labor ratios. Here, we find that while public investment generally raises the capital intensity of informal sector firms, the effects are relatively stronger for the top 20 percent of firms, suggesting that the complementarities between public investment and capital intensity are the highest for the largest firms in the informal sector. This has important implications for public policy: rather than a one-size-fits-all approach, more public investment goods might be targeted for the largest informal sector firms. By contrast, public investment is negatively associated with the capital intensity of formal sector firms, suggesting that it may be a substitute for private factors in this sector.

The beneficial role of public infrastructure on economic activity has been well debated in the literature. While most studies adhere to Barro's specification of a public capital flow that enters the production function as an input, studies like Futagami et al. (1993), Turnovsky (1997), and more recently Turnovsky (2004) and Chatterjee and Turnovsky (2007) formalize the concept of public capital stock that has its own accumulation process. The conceptual notion of public capital stock envisages an important role to maintenance of the existing stock owing to depreciation. The role of maintenance of public infrastructure was highlighted by the World Bank (1994) study, which said *"timely maintenance expenditure of $12 billion would have saved road constructions costs of $45 billion in Africa"*. The study also claimed that curbing maintenance expenditure in times of budgetary austerity is a wrong policy as high costs have to be incurred later for rehabilitation and reconstruction.

The role of maintenance expenditure on economic growth, however, lacked a formal theoretical exposition until very recently. Rioja (2003) formally introduced the concept of maintenance by endogenizing the depreciation rate (as a function of maintenance expenditure) and showed that reallocating funds from new investment to maintenance has a positive effect on economic growth. However, in Rioja's model maintenance expenditure is funded by tax revenues while new investment in infrastructure is funded by foreign aid. This assumption precluded a discussion on optimal allocation of tax revenues between maintenance and new investment. Kalaitzidakis and Kalyvitis (2004) take into account government budgetary constraint and derive optimal allocations toward maintenance and new investments. But their model departs from household optimizing behavior ignoring the possibility of welfare loss through higher taxation. Dioikitopoulos and Kalyvitis (2008) expound on the implication of public capital on the trade-off between maintenance expenditure and new investments. They use an endogenous growth model to show that countries facing low congestion in public infrastructure would require a threshold level of maintenance expenditure to experience a balanced growth in output. On the other hand, countries facing high congestion would require a threshold level of new investments for balanced growth. Agenor (2009) departs from the previous studies and models maintenance expenditure through an additional efficiency parameter keeping intact the concept of endogenous depreciation rate as in Rioja (2003). He shows that the growth-maximizing tax rate in a decentralized economy is equal to the output elasticity of public capital as in the Barro model. The welfare-maximizing share of spending on maintenance is shown to be identical to the growth-maximizing share when the tax rate is set at the level implied by the Barro rule.

The studies cited above however fail to provide a perspective on developing countries. An important feature of a developing country is the existence of a large informal sector and its equally large contribution to the GDP. In India for example, the unregistered and unincorporated small production units that usually do not fall under the tax administration, contribute almost 55 percent to the GDP. Chapter 4 tries to incorporate the informal sector into a two-sector endogenous growth model and examine it's implication on optimal fiscal policies. The objective of this paper is to draw some inference regarding the optimal tax rate, and the optimal allocation of tax revenues toward maintenance and new investments for a country characterized by a large informal sector. This is the main contribution of this study.

But what role does the informal sector play in determining the optimal tax rate and the optimal allocation of spending toward maintenance of public capital? The mechanism through which the informal sector affects the provision of infrastructure and maintenance spending is through congestion and efficiency. Since the informal sector firms usually do not fall under the tax administration, the government's main source of tax revenues is the formal sector. Tax revenues that are collected from the formal sector are then spent on infrastructure (like roads, railways, and power). Spending on infrastructure has two components: new investments and maintenance. Congestion of public infrastructure stems from the non-excludable nature of public infrastructure that can be used by the formal sector and informal sector in their production processes. For example, the benefits of building a new road (or the benefits of maintaining a good condition of roads) accrue to both formal and informal productions units as none of them can be excluded from using it. But financing new investments and maintenance of existing infrastructure are however borne out of the tax revenues that are collected from the formal sector. Put differently, the informal sector poses a free-rider problem by using public infrastructure that is financed by taxes levied on the formal sector.

An implication of such a congestion effect is the loss of efficiency of public infrastructure which in turn has a negative effect on overall production. In addition to this, corrupt practices, inefficient appraisal processes, and fund disbursal systems also make public investment inefficient in developing countries. So, to contend that higher public investment has a positive impact on output and productivity would be wrong if not controlled for inefficiency of public investments. An interesting study by Hulten (1996) showed that a large portion of differential growth rate between Africa and East Asia can be explained by the difference in effective use of infrastructure capital in the two regions. Pritchett (2000) argued on methodological grounds that not all of public investments translate into capital which is quite pertinent to developing countries. Dabla-Norris et al. (2012) expounded on this idea and created the Public Investment Management Index (PIMI). Using the PIMI, Gupta et al. (2014) produced cross country estimates for efficiency adjusted public capital. The estimated factor share of efficiency adjusted public capital was in the range of 0.143 to 0.158. Chakraborty and Dabla-Norris (2009) made a serious effort in formalizing a model with the efficiency parameter. They argue that simply increasing public investment could be highly inefficient in low income countries where effort must be put in to maintain quality of public investment. Agenor (2009) also models the maintenance expenditure through an additional efficiency parameter keeping intact the concept of endogenous depreciation rate as in Rioja (2003). This paper extends Agenor (2009) framework by defining the efficiency parameter as a function of the share of maintenance expenditure in total spending, taking into account the congestion effect of public capital due to the informal sector. In my model, the efficiency of public infrastructure increases with the share of maintenance expenditure and decreases with the size of the informal sector relative to formal sector. This is the second major contribution of this paper.

Results in the paper show that the growth-maximizing tax rate is a function of the output elasticities of public infrastructure (for formal and the informal sectors), and the efficiency parameter which is exogenously given in the decentralized economy. This tax rate is however lower than the Barro (1990) tax rate which is equal to the output elasticity of public infrastructure in a one-sector growth model. The growth-maximizing share of maintenance expenditure is shown to be a function of the production elasticities, the efficiency elasticity of maintenance expenditure, and the responsiveness of depreciation rate to maintenance spending. The welfare maximizing tax rate and the share of maintenance are not separately identified in a centrally planned economy. But if the social planner imposes a tax rate that maximizes the decentralized growth rate, the welfare-maximizing share of maintenance share is shown to be positively related to the ratio of informal to formal sector output. Thus economies with large informal sectors would benefit by devoting more resources toward maintenance of existing public infrastructure. This paper, by deriving the optimal tax rate and optimal share of maintenance would help the policy makers in developing countries to optimally spend resources on public infrastructure.

Chapter 2

Understanding Informal Employment in India: Competitive Choice or a Result of Labor Market Segmentation?

2.1 Introduction

In developing countries, informal workers - those with no social security benefits, job contract and paid leave - constitute about half of the total labor force. These workers are not only employed by small, unregistered, and unincorporated firms - commonly known as the informal sector, but by large formal sector firms as well. In South Asia for example, while informal employment constitutes 75 percent of total non-agricultural employment, about 15 percent can be attributed to employment in the formal sector.[1]

There are two broad views that explain the large share of informal employment in developing countries: labor market segmentation and the competitive choice. According to the labor market segmentation hypothesis, informal workers are disadvantaged and are waiting to be employed as formal workers (Lewis, 1954 and Harris and Todaro, 1970). Employers ration formal jobs that results in a queue for these jobs. Institutional barriers also restrict workers from entering formal employment. In the absence of such entry barriers, a worker would choose to work in the sector that pays higher wages and other non-wage benefits. In the labor economics literature, this line of argument is commonly known as labor market segmentation. A similar line of argument is put forward by efficiency wage proponents (Stiglitz, 1981 and Solow, 1980). According to the

[1] Figures based on ILO (2012). South Asia includes three countries for which the data are available: India, Pakistan and Sri Lanka.

efficiency wage argument, formal wages are set higher than the market-clearing rate to induce worker productivity, which in turn creates segments within the labor market. Other factors that may lead to labor market segmentation include search frictions (Burdett and Mortensen, 1998) and monopsonistic power of firms (Ashenfelter et al., 2010 and Alan, 2011).

The competitive choice hypothesis, emphasizes the role of costs and benefits of informal employment (Fields, 1990 and Maloney, 2004). According to this view, informal employment may be associated with desirable non-wage benefits like tax evasion and flexible work hours that attract voluntary movement from formal to informal employment. Fields (1990) argues that the informal sector comprises two segments: the upper tier and lower tier. The upper tier informal segment comprises of self-employed workers who voluntarily move out of formal jobs. For these upper-tier informal workers, informal employment is a competitive choice over formal employment. The lower tier informal segment comprises of disadvantaged workers who do not find formal employment and eventually settle for low paying informal jobs. Thus there are two contrasting viewpoints: one that views informal employment as a competitive choice and the other leans on the labor market segmentation hypothesis that emphasizes the entry barriers and rationing of formal jobs.

As of 2011-12, informal workers in India accounted for 85.8 percent of the labor force. In this paper I use India's National Sample Survey data on Employment and Unemployment (2011-12) to examine the underlying factors that are associated with the large share of informal employment in India. Specifically, there are two objectives of this paper. First, I estimate the formal-informal wage gap across different quantiles of the wage distribution and decompose this gap into coefficient and endowment effects. The coefficient effect explains what fraction of the overall wage gap is due to the difference in returns to human capital and individual characteristics of formal and informal workers. The endowment effect explains what fraction of the wage gap is due to the difference in human capital and characteristics between formal and informal workers. Second, I analyze whether the large share of informal employment in India is driven by competitive choice or labor market segmentation.

Due to data constraints, there is a dearth of research that analyzes the segmentation hypothesis for South Asian countries (like India, Bangladesh, Sri Lanka, Pakistan). However, the questions are rather important for a country like India where a large fraction of the labor force (see Figure 2.1) is informally employed. Another point of concern that has challenged policy makers lately is the growing informalization of jobs. Mehrotra et al. (2013) show, the share of informal employment in the formal sector has increased from 32 percent in 1999-2000 to 54 percent in 2004-05 to 67 percent in 2011-12. This informalization of formal employment is a result of an increase in contractual jobs within the formal sector in which the firms do not pay any employment benefits to the workers. Mehrotra et al. (2012) argue that the increasing prevalence of informal employment in the formal sector poses a tough challenge to policy makers in achieving inclusive growth and sustainable development in the future. Moreover, informal workers are vulnerable to idiosyncratic economic shocks since they do not have access to social security benefits. This aspect of informal employment has important implication for poverty and human development for a developing country like India. Recent public policies in India focus on skill development as the main instrument to increase the employability of the workforce. The basic assumption behind these policies is the direct link between skill level and better pay, and hence better living and working conditions for the workers. Given the large share of informal employment in India, the pertinent question is whether these policies play a positive role in reducing the share of informal employment. Interestingly however, 28.2 percent and 29.2 percent of the workers having some technical education are distributed across formal and informal employment respectively.[2] Thus, skill is not the sole factor that allocates the workers into formal and informal employment. Present policy initiatives, however, assume that worker characteristics are the sole determinant of job choices, and do not take into account entry barriers to formal employment or rationing of formal jobs by the employers.

In this paper, I use the quantile regression framework that decomposes the formal-informal wage gap into coefficient and endowment effects across the wage distribution (Machado and Mata, 2005). Then I use the polychotomous choice model developed by Lee (1983) to test for labor market segmentation. The methodology is an extension of Heckman (1979) that allows for multiple labor market choices. In the first stage, I use the the multinomial logit to estimate the participation decision that includes four labor market choices: formal employment, informal employment, self-employment and staying out of the labor force. In the second stage, I estimate the wage equations

[2] Figures based on Employment and Unemployment Survey (2011-12), conducted by the National Sample Survey Organization.

for formal and informal workers taking into account the sample selection bias resulting from self-selection of workers into formal and informal employment. Gindling (1991) argues that a non-random selection of workers into a particular sector that does not affect wages in that sector implies that workers do not have full access to that sector.

The results presented here support the segmentation hypothesis for the Indian labor market. I find a significant wage gap between formal and informal workers across the wage distribution. At the lower end of the wage distribution, differences in returns to human capital and individual characteristics between formal and informal workers explain a major part of the wage gap. The informal workers at the lower end of the wage distribution may be identified as the disadvantaged workers who earn lower returns to their skills compared to their formal counterparts. At the higher end of the distribution, the major part of the wage gap is explained by differences in human capital and individual characteristics between formal and informal workers. Second, results from the polychotomous choice model show no evidence of workers being able to self-select into formal employment. Thus, we cannot reject the labor market segmentation hypothesis. The counterfactual wages show that 85 percent of the male and 83 percent of the female informal workers would have earned higher wages if they were formal workers.[3]

The contribution of this paper is twofold. First, this paper looks at an emerging South Asian economy that has not been studied in earlier works on labor market segmentation. This is important because the shares of informal employment in South Asian countries (see figure 2.1) are quite high compared to the other regions of the world. Moreover, informal employment is an important issue for a developing country like India because lack of benefits for informal workers makes it difficult for them to get insurance against idiosyncratic productivity shocks, health injuries etc. Second, this paper provides a useful platform for policy makers to address the issue of formal-informal wage gap and to address the entry barriers that workers face in the labor market. This study bridges these

[3]It may be argued that informal employment has other non-wage benefits, but this argument is not plausible because by definition the informal workers do not receive any employment benefits.

gaps by using the 68^{th} round of National Sample Survey Organization (India) data on Employment and Unemployment. To the best of my knowledge, no study has analyzed these questions for India using this data set as of now.

The remainder of this paper is organized as follows. Section 2.2 reviews the literature on labor market segmentation focusing on developing countries. Section 2.3 presents the definitions of informal employment and the informal sector. Section 2.4 discusses the data and summary statistics and more importantly the identification of formal and informal workers. Section 2.5 sets out the empirical models. The results are presented in Section 2.6. Section 2.7 concludes.

2.2 Literature Review

The main focus of this discussion is to review the evidence on the labor market segmentation hypothesis in the context of formal and informal employment in developing countries. Most studies on labor market segmentation focus on Latin American countries. Maloney (1999) uses a dynamic panel data for the Mexican labor market and provides evidence for the competitive choice hypothesis. He concludes that earnings differentials do not offer compelling evidence in favor of the segmentation hypothesis because of the difficulty of quantifying unobservable variables. Navarro-Lozano and Schrimpf (2004) use a discrete choice model to test for segmentation in the Mexican labor market. They find no evidence of rationing of jobs in the formal sector and thus reject the segmentation hypothesis. The evidence is consistent with a market in which comparative advantage determines the sectoral allocation of workers. Pratap and Quintin (2006) test for the segmentation in the Argentinean labor market. Specifically, they test for whether observationally similar workers earn higher wages in the formal sector versus the informal sector in developing nations. They find no evidence of a formal sector wage premium in Buenos Aires and its suburbs. They find higher wages on average in the formal sector, but this apparent premium disappears after semi parametrically controlling for individual and employer characteristics. Although they do not perform a formal test on whether informal sector workers voluntarily choose informal sector over the formal sector , the near zero wage gap between formal and informal sector workers insinuates that workers are indifferent between the formal and informal sectors. Carneiro and Henley (2001) provide further evidence in favor of the competitive choice hypothesis in the Brazilian labor market.

Tannuri-Pianto and Pianto (2002) and Günther and Launov (2012) provide evidence in favor of labor market segmentation in the context of formal and informal employment. Tannuri-Pianto and Pianto (2002) test for the segmentation hypothesis in the Brazilian labor market. They use quantile regressions to test for sample selection bias at different quantiles of income. They find that earnings gap between formal and informal workers are wider at the lower quantiles than at the high ones. Returns to attributes explain around 30 percent of the earnings at low quantiles. At high quantiles the earnings gap is completely explained by individual characteristics. Informal workers in the lower quantiles receive lower returns to their skills compared to their formal counterparts. Based on this observation they cannot reject the hypothesis of labor market segmentation. Günther and Launov (2012) use data from Côte d'Ivoire and reject the hypothesis of fully competitive labor markets. They use an augmented two step Heckman procedure to correct for sample selection bias to estimate the number of segments within the informal sector. Their results show that the informal sector comprises of two segments: the upper tier and the lower tier. The lower tier informal sector is a result of labor market segmentation. However, they conclude that comparative advantage is the cause for the existence of the upper tier informal sector. In summary, the evidence on labor market segmentation mainly comes from Latin American countries and majority of them provide evidence for the competitive hypothesis.

To the best of my knowledge, Khandker (1992) is the only other study that looks into labor market segmentation for India. He uses survey data for urban slum dwellers and finds evidence of labor market segmentation resulting not from sample selection bias on the part of workers but selectivity bias by firms. The study differentiates between protected wage segments, unprotected wage segments and self employment. This specification is not consistent with the modern definitions of formal and informal employment. The study, however, does not examine the wage gap between the different segments in the labor market. Moreover, the scope of the study is limited to a small sample of urban slum dwellers that do not represent the different cohorts of the labor force. Thus no study, prior to this one has ever tested for the segmentation hypothesis for the Indian labor market in the context of formal and informal employment. As discussed in the literature review section, evidence from the Latin American countries show that informal employment is a competitive choice for workers vis-à-vis formal employment. It needs to be seen whether these predictions hold for the Indian labor market as well.

2.3 Definitions and Conceptual Framework

This paper uses the following definitions for informal sector and informal employment:[4]

Definition 1: The *informal sector* consists of small-scale, self-employed activities (with or without hired workers but less than 10 workers), typically at a low level of organization and technology, with the primary objective of generating employment and income. The activities are usually conducted without proper recognition from the authorities, and escape the attention of the administrative machinery responsible for enforcing laws and regulations.

Defintion 2: *Informal employment* is a job-based concept and encompasses those jobs that generally lack basic social or legal protections or employment benefits and may be found in the formal sector, informal sector or households.[5] Table 2.1 illustrates this categorization.[6]

2.4 Data and Summary Statistics

The National Sample Survey Organization (NSSO) in India collects data on employment and unemployment every five years. Data are collected on a number of individual characteristics, job characteristics, working conditions and social security benefits. This paper uses individual level data for the year 2011-12 (latest year available) provided by the NSSO to answer the questions outlined in the previous sections.

[4]These definitions were adopted by the International Conference of Labour Statisticians (ICLS). The International Labor Organization has implemented them based on the ICLS resolutions. The National Commission for Enterprises in the Unorganised Sector (See Sengupta et al. (2007)) restricts the informal sector to the proprietary and partnership firms that have less than 10 workers.

[5]I use workers in informal employment and informal workers interchangeably in the text.

[6]The 17th International Conference on Labor Statisticians (ICLS) provides the definition of the informal sector and informal employment based on the enterprise types and job characteristics pertaining to the non-agricultural sector. As far as the agricultural sector is concerned, no consistent definition is followed across countries. The International Labor Organization has adopted these definitions based on the ICLS resolutions. The statistical office in India does not provide a formal definition of informal sector that includes the agricultural sector, and therefore I focus on the non-agricultural sector. See http://ilo.org/public/englisg/bureau/stat/download/papers/def.pdf

The 2011-12 Employment and Unemployment Survey, studies 101,724 households that include 456,999 individuals. The sector of employment for each working individual is recorded according to the 2-digit National Industrial Classification (NIC, 2008) codes. I restrict the sample to the primary working age population (15 to 59 years) in the non agricultural sector.[7] Subject to available data on all covariates, the final sample has 109,219 observations for males and 118,620 observations for females. A detailed description of the sample selection process is provided in Appendix 1.

2.4.1 Identification of Informal Sector and Informal Employment

2.4.1.1 Informal Sector

The 17th International Conference of Labour Statisticians (ICLS) provides the definition for informal sector based on the enterprise type (firm). The type of enterprise is recorded for all workers. It is not recorded for the unemployed persons and the persons who are not in the labor force. Table 2.2 shows the distribution of the individuals across different enterprise types. Categories 1-4, 8 and 9 make up the informal sector and 5-7 comprise the formal sector.[8] In the sample, 77.5 percent of the workers are employed in the informal sector and 22.5 percent of the workers are employed in the formal sector.[9]

[7]Since unemployed persons do not have any NIC-2008 classification, I include them in the non-agricultural labor force. It is not possible to distinguish agricultural and non-agricultural unemployment.

[8]The 17th ICLS treats households as a separate category outside the formal and the informal sector. However, all workers employed by the households are classified as informal workers. For brevity, 'employer's households' are classified as informal sector enterprises in Table 2.2. This classification is innocuous since those workers are anyway treated as informal workers in the final classification.

[9]All figures in Tables 2.2-2.7 are census-adjusted

2.4.1.2 Informal Employment

The definition of informal employment is based on the job categories outlined in the 17th ICLS. The identification of informal workers is done in steps. First, the workers are separated based on their job types: self employed and employees. Self employed workers are classified as informal workers. For the employees working in the firms, those who do not receive any employment benefits are categorized as informal workers and the rest as formal workers. Tables 2.3-2.5 illustrates this identification process.

As discussed, first we need to identify the different categories of the workers based on the job types. In the data, the job categories are recorded based on the principal activity status for the last 365 days from the date of the survey. The job type of the workers may be different from the sector they work in. Table 2.3 shows the distribution of job categories based on the principal activity status. Majority of the workers are regular salaried wage employees (14.4 percent), followed by self employed (own account workers) (11.5 percent). In the sample, 62 percent of all the individuals (categories 7-12 in Table 2.3) are not in the labor force.[10]

The information provided in Table 2.3 is insufficient to categorize the workers as formal and informal workers. We need additional information on the benefits received by the workers in each category in order to identify them as formal and informal workers. Workers are classified into self-employed and employees who are employed by other persons or firms. The employees comprise of the regular/ salaried wage employees, casual wage laborers in public works and other types of work. These workers are found both in the formal and the informal sectors. As per the definition, informal workers do not receive any social security, job security and other employment benefits such as paid leave. This information on employee benefits is recorded only for the employees (categories 4-6 in Table 2.3) and not for the self employed workers (categories 1-3 in Table 2.3). Workers who receive all of the benefits above are identified as formal workers. Table 2.4 shows the percentage

[10]Note that the total number of observations are different in Table 2.2 and Table 2.3. This is because, enterprise type is recorded only for the working individuals while principal activity status is recorded for all individuals in the sample (including those who are unemployed and those who are not in the labor force).

of workers receiving each category of benefits (social security, job contract and paid leave). The last column in Table 2.4 shows that only 17.4 percent of the workers receive social security, have written job contracts and are eligible for paid leave. The rest (82.6 percent) do not receive any of these benefits.[11]

Tables 2.3 and 2.4 give us the information on the job types and the employment benefits received by the workers. Combining these two information sets, we now identify the formal and informal workers. Table 2.5 reports the measures of formal and informal employment, treating self employment as a separate category.[12] The formal and informal employment categories in Table 2.5 comprise wage earners only. The categories reported in Table 2.5 are the final categories used in this paper. Further, Table 2.5 reports the figures separately for males, females and for the entire sample. Overall, 14.2 percent of the workers in the labor force are formally employed and 44.3 percent of the workers are informally employed. If we employ the broader definition of informal workers that includes self employed workers, informal workers comprise 85.8 percent (treating the 41.5 percent of the self employed workers in Table 2.5 as informal workers) of the total labor force. However, in this study I treat self employed workers as a separate category for the reasons noted in footnote 12.

Table 2.5 shows that 29.4 percent of the working age males are out of the labor force while 88.8 percent of the working age females are out of the labor force. Labor force participation rate varies significantly for males and females across all states of India. For this reason, the empirical

[11] Again, note that the total number of observations reported in Table 2.5 is different from Table 2.2 and Table 2.3. This is because, information on benefits is recorded only for wage employees (Category 4,5 and 6 in Table 2.3). Information on benefits is not available for self employed workers.

[12] Wages are reported only for the wage earners that include regular salaried employees and casual laborers. Since it is difficult to identify the the profit and wages components from the earnings of the self employed, data on earnings of self employed persons are not collected in the survey. Although, self employed workers are categorized into formal and informal employment as per the ILO definition, in this paper I consider them as a separate category from formal and informal employment because estimation of a wage equation is not possible for the self employed workers. Furthermore, the share of self employment (14.7 percent of the working age population) is quite large in the sample representing systematically different types of jobs than the wage earners. This difference in job types compared to wage earners, and the unavailability of earnings information for the self employed workers makes the treatment of self employed workers as a separate category (from formal and informal employment) a plausible assumption.

results in this paper are reported separately for males and females. For the entire analysis in the paper, I treat unemployed persons as not in the labor force. This assumption is necessary because the empirical model I use does not allow unemployment as a separate category. Moreover, only 1.9 percent (category 7 in Table 2.3) of all individuals in the sample reports to be involuntarily unemployed that makes this assumption innocuous to the results reported in this paper.

2.4.1.3 Formal and Informal Employment in Formal and Informal Sectors

In this subsection, I allocate the formal and informal workers into the formal and informal sectors. Although this classification is not directly significant for the analysis, it gives us some useful insights on the nature of informal employment. Specifically, this classification measures the share of workers who do not receive any employment benefits in spite of working in the formal sector. Table 2.6 presents the framework for the identification of formal and informal employment in the formal and the informal sectors. The cells with 'NE' refer to non-existent. Own account workers[13] in the formal sector are identified as formal workers in the formal sector (cell A). Since all own account workers in the informal sector are identified as informal workers, there does not exist any formal own account workers in the informal sector. All own account workers working in the informal sector are identified as informal workers (cell F). Employees who run their own informal household enterprises by hiring employees are identified as informal workers (cell G). Some employers may be working in the formal sectors that are identified as formal workers. Unpaid family members who contribute to the production processes are considered as informal workers regardless of which sector they work in.

Applying the definition of informal employment on regular and casual workers based on the employment benefits, and treating self employed workers as shown in Table 2.6, the different categories (Table 2.7) of employment are identified. Each category in Table 2.7 is derived by aggregating the cells in Table 2.6. For example, the category formal employment in the formal sector comprise of own account workers in the formal sector (cell A in Table 2.6), employers who work in the formal

[13]Own-account units are owned and operated by single individuals working on their own account as self-employed persons, either alone or with the help of unpaid family members. The activities may be undertaken inside or outside the enterprise owner's home, and they may be carried out in identifiable premises, unidentifiable premises or without fixed location. See Sastry (2004).

sector (cell B in Table 2.6) and those employees who work in the formal sector but do not receive any employment benefits (cell D in Table 2.6).[14] From Table 2.7, note that 85.6 percent of the workers in the labor force are informally employed. Also, the formal sector employs 51 percent of its workers as informal workers which is a growing point of concern for policy makers.

2.4.2 Summary Statistics

Table 2.8 reports the mean and standard deviations of all the variables used in the analysis for males. Table 2.9 reports the summary statistics for females. In both the tables 2.8 and 2.9, the summary statistics are reported separately for the following categories of employment: formal employment, informal employment, self employed and not in the labor force.

2.4.2.1 Wages and Consumption Expenditure

Weekly nominal wages for formal male workers (Rs 4966) are significantly higher than the informal male workers (Rs 1771). The mean formal-informal wage gap is Rs 3195 for males and Rs 2963 for females. [15] The monthly per capita consumption expenditure (MPCE) is also significantly less for informal workers (Rs 1714 for males and Rs 2007) compared to formal workers (Rs 2902 for males and Rs 3651 for females).[16] Male self employed workers report slightly higher MPCE (Rs 1806) than the informal workers, but female self employed workers report lower MPCE (Rs 1756) than informal workers. These statistics suggest that on average informal workers maintain poor living conditions compared to the formal workers.

[14]The letters in parentheses following each category in Table 2.5 are the references to the cells in Table 2.4.

[15]The average official exchange rate during the period 2011-12 was Rs 50/$ (World Development Indicators).

[16]The MPCE is used only to highlight the difference between the different categories. This variable is not used for the analysis that follows.

2.4.2.2 Demographics

Informal workers are on average younger (34.2 years for males and 34.9 years for females) than formal workers (42.1 years for males and 39 years for females) for both males and females. The average age of females not in the labor force (32 years) is significantly higher than their male counterparts (20.1 years). Majority of the females who are not in the labor force may comprise housewives which is a common feature in rural and urban households in India. During 2009-10, 40 percent of the rural females and 48 percent of urban females were engaged in domestic duties.[17]

The number of dependents is an important factor that influences the labor market participation decision and the sectoral choice of a worker.[18] Dependents are divided into two separate categories: children under 15 years of age and elderly members greater than 60 years of age. Formal workers have on average fewer dependents than their informal counterparts and self employed workers, for both males and females. The significant variation of number of dependents across the different sectors implies that number of dependents may play a role in sector selection.

Other demographic control variables include religion, caste and marital status. Hindus comprise roughly 80 percent of the workers in all categories followed by Muslims. In India, caste is an important demographic characteristic. The category 'others' include the higher castes. Majority of the formal employment comprise of higher castes (40 percent) and Other Backward Classes (OBC). 90 percent of all males and 70 percent of all females in the sample in all categories are married.

2.4.2.3 Human Capital

General education is reflected by years of schooling.[19] Formal workers are on average more educated (12.5 years of schooling for males and 13 years of schooling for females) than informal workers (7.4 years of schooling for males and 7.3 years of schooling for female). Male self employed workers (8 years of schooling) are slightly more educated than informal male workers and but the same does not hold true for female self employed workers. Men who are not in the labor force are more

[17]Source: NSSO (2013)

[18]See Section 2.5.3 for a detailed discussion.

[19]I have used NCEUS (2007) to compute the mean years of schooling. The following classification is considered: Illiterate-0, literate below primary-1, primary-4, middle-8,) secondary-10, higher secondary-12, diploma/ certificate course - 14, graduate - 15, postgraduate and above -17.

educated than informal workers (10.1 years of schooling), but less educated than formal workers. These men may have chosen to stay out of the labor force to gain a few years of education before joining the labor force. However, this is not true for the females in the sample.

Technical education and vocational training are regarded as important attributes that contribute significantly towards the employability of workers and wages offered.[20] Roughly 14 percent of formal male workers have some technical education compared to 5 percent of informal workers. 28.2 percent and 29.2 percent of the workers having some technical education are distributed across formal and informal employment respectively. Amongst those who received formal vocational training, only 21 percent are formally employed, 30 percent are informally employed, 23 percent are self employed and the rest are not in the labor force. Females follow roughly the same pattern for technical education and vocational training as males. These facts provide preliminary evidence that technical education and vocational training may be necessary, but not sufficient, to increase the chances of being formally employed and earn more wages.

2.4.2.4 Regional Variables

Regional variables include whether a worker resides in an urban or rural area and the the region of residence – south, north, central, east, north-east and west. Workers (both male and females) in all categories are almost evenly distributed across urban and rural areas. The distribution is even across the region of residence as well.

[20] A person has some sort of technical education if he holds a degree, diploma or certificate in engineering and technology, agriculture, medicine and all other technical fields. Vocational training means some sort of expertise in the field of trade. Examples of vocational training are book binding, handicraft, medical transcriptions etc. Vocational training is more focused towards the type of job and can be formal and informal. If a worker acquires skills for a particular job from, say, family heredity then it is regarded as informal vocational training. But if a worker enrolls in a formal institution to acquire vocational training that is regarded as formal vocational training. Both technical and vocational training captures the skill level of a worker, whereas years of schooling reflect the general level of education of the workers.

2.5 Empirical Methods

The first objective of this paper is to measure the wage gap between formal and informal workers. Measuring the average formal-informal wage gap may not provide adequate information on what happens across the whole wage distribution. Pratap and Quintin (2006) showed that the wage gap between formal and informal workers decreases at the higher quantiles of the wage distribution. Thus to reveal useful information on the wage gap, I use quantile regressions that estimate the formal-informal wage gap at different quantiles of the wage distribution. I use the empirical technique proposed by Machado and Mata (2005) to decompose the wage gap into endowment and coefficient effects. The dependent variable is the weekly wage earnings. The independent variables include years of schooling, whether or not received technical education and vocational training and a number of individual characteristics like age, sex, religion, caste and region of residence. However, due to empirical complexity I am not able to control for sample selection bias in the quantile regression framework that may produce spurious results. Nonetheless, the estimates give a preliminary idea of the wage penalty faced by informal workers. I discuss the empirical strategy in Section 2.5.1.

The second objective of this paper is is to test for labor market segmentation. Several papers have empirically tested the labor market segmentation hypothesis. Heckman and Hotz (1986) test for labor market segmentation based on the earnings of Panamanian males. They find different selection corrected earnings functions for different groups (separated either geographically or by income) based on which they conclude that the labor market is segmented. However, they argue that there is little robust behavioral content to imply dual labor markets in Panama. Dickens and Lang (1985) propose a "switching model" to test for labor market segmentation. They emphasize the existence of entry barriers or an evidence of queuing of jobs in formal employment as a necessary precondition for the labor market segmentation hypothesis to be valid. Using data from the Panel Study on Income Dynamics for the year 1980, they provide evidence for the dual market hypothesis. Magnac (1991) develops a microeconomic model incorporating the cost of entry to the formal sector to test for labor market segmentation for married women in Colombia. The paper finds evidence that comparative advantages for individuals between the various economic sectors are more important than segmentation. The consensus on the methodology is that a near zero returns to human capital variables for informal workers are neither necessary nor sufficient for the segmentation hypothesis to be valid. Bosworth et al. (1996) explain that different returns to human capital variables for formal and informal workers are not necessary conditions for market segmentation because the actual wage levels may be different for the formal and the informal workers. Different returns to endowments are neither sufficient conditions because higher returns to formal workers may be compensated by lower starting salaries. Similarly, different wage determination mechanisms in the two sectors do not imply market segmentation if workers are free to choose the type of employment. Gindling (1991) argues that if workers are free to choose the type of employment, observationally identical workers would choose the jobs that pay them higher wages. Thus labor market segmentation implies a wage penalty for the observationally identical workers. In this paper, I estimate two wage equations for formal and informal workers accounting for sample selection bias. There are four labor market choices: formal employment, informal employment, self employment and not in the labor force. I use a polychotomous choice model developed by Lee (1983) which is an extension of the binary choice model developed by Heckman (1979). I lay out the model in Section 2.5.2. The identification of the model is achieved by including at least one variable in the selection equation that is excluded from the wage equation. I use two variables for this purpose: the number of dependents less than 15 years of age and the number of dependents more than 60 years of age. I discuss the rationale for this identification strategy in Section 2.5.3. The test for labor market segmentation is based on a careful interpretation of the results from the polychotomous choice model. I follow Gindling (1991) in interpreting the results. Basically, if a worker's selection into formal employment is non random and if that non randomness does not affect formal wages, then it implies that workers cannot self select into formal employment and that entry barriers exist. I discuss this interpretation and the different hypotheses in detail is Section 2.5.4.

2.5.1 The Formal-Informal Wage Gap

In this section I discuss the decomposition technique proposed by Machado and Mata (2005).[21] The objective is to estimate the wage gaps between formal and informal workers at different quantiles of the wage distribution and decompose the wage gap into coefficient effects and endowment effects. The Machado and Mata (MM) technique can be seen as generalization of the Oaxaca-Blinder decomposition method for quantile regressions. The first step of the estimation process involves estimating the conditional quantile functions for both sets of workers. Let w_i denote the log weekly wage and $\mathbf{X_i}$ denote the set of covariates for each individual i that includes age, education, caste, religion, marital status, and regional dummies. ϵ_i is a disturbance term independent of the explanatory variables. The conditional quantile function for formal workers can be specified as a linear function:

$$q_\tau^F(w_i^F|\,\mathbf{X_{F,i}}) = \mathbf{X'_{F,i}}\beta^{\mathbf{F}}{}_\tau \quad \tau \in (0,1) \tag{2.1}$$

and for informal workers:

$$q_\tau^I(w_i^I|\,\mathbf{X_{I,i}}) = \mathbf{X'_{I,i}}\beta^{\mathbf{I}}{}_\tau \quad \tau \in (0,1) \tag{2.2}$$

where $q_\tau(w_i|\,\mathbf{X_i})$ specifies the conditional quantile (τ^{th}) of the log weekly wage distribution and the set of coefficients (β_τ) are interpreted as the estimated returns to the covariates at the specified quantile. F and I denote formal and informal workers respectively. The conditional quantile regression function is then estimated using the Koenker and Bassett Jr (1978) approach that minimizes the weighted least absolute deviations. The wage gap (G) between formal and informal workers can be specified as:

$$G_\tau = q_\tau^I(w_i^I|\,\mathbf{X_{I,i}}) - q_\tau^F(w_i^F|\,\mathbf{X_{F,i}}) = X'_{I,i}\beta^{\mathbf{I}}{}_\tau - X'_{F,i}\beta^{\mathbf{F}}{}_\tau \quad \tau \in (0,1) \tag{2.3}$$

[21] See Albrecht et al. (2009) and Arulampalam et al. (2007) for applications of this technique.

The next step is to decompose the wage gap into the coefficient effect and the endowment effect. Equation 2.3 can be written as:

$$G_\tau = [(X'_{I,i}\beta^{\mathbf{I}}{}_\tau - X'_{I,i}\beta^{\mathbf{F}}{}_\tau)] + [(X'_{I,i}\beta^{\mathbf{F}}{}_\tau - X'_{F,i}\beta^{\mathbf{F}}{}_\tau)] \quad \tau \in (0,1) \tag{2.4}$$

The first term on the right hand side of expression (2.4) refers to the coefficient effect. This term shows how much of the wage gap is explained by the differences in the returns to covariates for formal and informal workers if the informal workers had retained their characteristics. The second term calculates the contribution of the differences in characteristics between formal and informal workers to the overall wage gap. The decomposition technique involves the construction of the counterfactual unconditional wage distribution $X'_{I,i}\beta^{\mathbf{F}}{}_\tau$, that is, how much the informal workers would earn if they were paid the same returns as the formal workers. However, in case of quantiles the unconditional quantile is not the same as the integral of conditional quantiles. Machado and Mata (2005) address this problem using a simulation based technique. The following steps summarizes the MM technique:

1. Sample u from a standard uniform distribution.
2. Estimate the different quantile regression coefficients, $\beta^{\mathbf{I}}{}_\tau(u)$ and $\beta^{\mathbf{F}}{}_\tau(u)$ for informal and formal workers respectively.
3. Generate a random sample with replacement from the empirical distribution of the covariates ($X_{I,i}$ and $X_{F,i}$) for each group.
4. Compute the counterfactual $X'_{I,i}\beta^{\mathbf{F}}{}_\tau(u)$.
5. Repeat steps 1 to 4 M times[22]

2.5.2 Polychotomous Choice Model with Selectivity Bias

The quantile regression technique works well in estimating the wage gap between the formal and informal workers given that the sectoral allocation of workers is exogenous. In other words, the model discussed above assumes that the allocation of workers into formal and informal employment

[22] I have used the Stata command 'mmsel' recently released by Souabni (2013). The command implements the MM technique as mentioned above. The standard errors are calculated using a bootstrapping procedure.

is completely random.[23] However, if sectoral allocation and labor force participation are non random, then the estimates from the models 2.1 and 2.2 are biased. This problem, commonly known as sample selection bias, was first proposed by Heckman (1979). In his original model, workers faced a binary decision to enter the labor force or stay out of the labor force. The decision is based on an underlying latent variable, such as the utility of the worker. If the utility from working is less than the utility from not working, the worker stays out of the labor force. Since utilities are not observed and offered wages are only observed for the workers who are in the labor force, estimating a standard Mincerian wage equation[24] would yield biased estimates, since the decision to enter the labor force is an endogenous choice. Heckman proposed a two stage method to circumvent this problem. In the first stage, a probit model is estimated with the participation decision as the dependent variable. This participation equation is also known as the selection equation. In the second step, the inverse Mill's ratio is constructed using the predicted probabilities from the first stage that is included in the standard wage equation in the second stage. This method produces consistent coefficient estimates of the wage equation. To identify the model and to avoid large standard errors in the second step, it is necessary to include at least one variable in the selection equation that is excluded from the wage equation.

The problem I am analyzing in this paper is slightly more complex. A potential worker faces the decision to enter or stay out of the labor market. If he decides to enter the labor market, he has three choices : start his own business (self employed), start working as an informal worker, or start working as a formal worker. So in the aggregate, a worker faces four potential outcomes: stay out of the labor force, accept informal employment, accept formal employment, or become a self employed worker. He chooses the outcome that gives him the maximum utility. For example, if a worker gains the maximum utility working as an informal worker compared to all other choices that he has, he chooses informal employment. Whether a worker actually faces these choices or whether the endogenous decision making on the part of the workers actually holds is an empirical question. This paper tests the hypothesis of endogenous sectoral allocation. If the evidence suggests that workers choose sectors that gives them the maximum utility, then we can infer that sectoral selection is endogenous. The Heckman model discussed above is not well suited to tackle the problem of

[23] Albrecht et al. (2009) extend theMachado and Mata (2005) technique to allow for sample selection bias. However, the model only allows a bivariate choice in the underlying selection choice is not appropriate in this case.

[24] See Mincer (1974)

polychotomous choices such as those described above. In this paper I use a polychotomous choice model developed by Lee (1983).[25] Hay (1980) proposes a similar model that deals with multiple choices with stronger assumptions than the Lee model. I use the Hay model as a robustness check for my results. Both approaches are discussed below.

Let there be M categories and one potential wage equation in each category.

$$w_{ji} = \mathbf{x_{ji}}\beta_\mathbf{j} + u_{ji} \;\; j = (1, ...M) \;\; i = (1, ...N) \tag{2.5}$$

$$I^*_{ji} = \mathbf{z_{ji}}\gamma_\mathbf{j} + \eta_{ji} \; j = (1, ...M) \; i = (1, ...N) \tag{2.6}$$

where w_{ji} is the log weekly wage in sector j for individual i. $\mathbf{x_{ji}}$ is the vector of explanatory variables that affect wages and $\beta_\mathbf{j}$ the respective coefficients. u_{ji} is the error that captures all unobserved characteristics of the workers not in $\mathbf{x_{ji}}$. I^*_{ji} is the utility derived from choosing sector j which is a function of $\mathbf{z_{ji}}$ is the vector explanatory variables and $\gamma_\mathbf{j}$ the coefficient vector. η_{ji} is the error term in the latent equation. The variables in x_{ji} and $z_{ji}$are exogenous such that, $E(u_j| \, \mathbf{x_1}, \mathbf{x_2}, ...\mathbf{x_M}, \, \mathbf{z_1}, \mathbf{z_2}, ...\mathbf{z_M}) = 0$ and $E(\eta_j| \, \mathbf{x_1}, \mathbf{x_2}, ...\mathbf{x_M}, \, \mathbf{z_1}, \mathbf{z_2}, ...\mathbf{z_M}) = 0$.

w_{ji} is observed only if the j^{th} category is chosen. Let I_j be the indicator function such that $I_j = j$ if the j^{th} category is chosen. The model can be formulated by an underlying utility maximization exercise in the following way :

$$I = j \;\; iff \; I^*_j > Max \, I^*_s(s = 1, ...4; \, j \neq s) \tag{2.7}$$

Let us define

$$\epsilon_j = Max \, I^*_s - \eta_j(s = 1, ...4; \, j \neq s). \tag{2.8}$$

So we can write the sectoral choice as :

$$I = j \;\; iff \, \epsilon_j < \mathbf{z_j}\gamma_\mathbf{j} \tag{2.9}$$

[25]See Trost and Lee (1984), Gyourko and Tracy (1988), Cohen and House (1996), Hilmer (2001), Zhang (2004) and Packard (2007) for applications of this model. Bourguignon et al. (2007) perform Monte carlo simulations to compare three methods used in the literature for selection bias correction using multinomial logit model, namely Dubin and McFadden (1984), Lee (1983) and Dahl (2002). Their results show that the semi-parametric alternative proposed by Dahl (2002) is to be preferred to Lee (1983). However, the semi-parametric approach by Dahl (2002) does not provide a robust interpretation of the coefficient on the selection term that is useful to identify the underlying selection process. One of the objective of this paper is to identify whether the workers can self select into formal and informal employment. The Lee (1983) approach allows us to interpret the coefficient on the selection term in a way to fulfill this objective. Nonetheless I have tested my results using the Dahl (2002) model but I did not find any significant differences in the estimates and the standard errors. However, in the robustness check section I have only reported the Hay (1980) model because it's interpretation is similar to the Lee (1983) model.

Assume that $\eta_j s$ are independently and identically distributed with type I extreme value distribution:

$$F(\eta_j < c) = exp[-exp(-c)] \tag{2.10}$$

Then, as shown by McFadden (1973) and Domencich and MacFadden (1975), the probability that sector j is chosen is given by:

$$Pr(I = j) = Pr(\epsilon_j < \mathbf{z_j}\gamma_\mathbf{j}) = F(\mathbf{z_j}\gamma_\mathbf{j}) = \frac{exp(\mathbf{z_j}\gamma_\mathbf{j})}{\sum_{j=1}^{M} exp(\mathbf{z_j}\gamma_\mathbf{j})} \tag{2.11}$$

The distribution of ϵ_j is given by,

$$F_j(\epsilon) = Prob(\epsilon_s < \epsilon) = \frac{exp(\epsilon)}{exp(\epsilon) + \sum_{j=1,j\neq s}^{M} exp(\mathbf{z_j}\gamma_\mathbf{j})} \tag{2.12}$$

Since wages are observed for the particular sector that the worker chooses, the conditional wage equation then becomes,

$$E(w_{ji}|I = j) = E(w_{ji}|\ \epsilon_j < \mathbf{z_j}\gamma_\mathbf{j}) = \mathbf{x_{ji}}\beta_\mathbf{j} + E(u_{ji}|\ \epsilon_j < \mathbf{z_j}\gamma_\mathbf{j}) \tag{2.13}$$

Equation (2.13) shows that if $E(u_{ji}|\ \epsilon_j < \mathbf{z_j}\gamma_\mathbf{j}) \neq 0$, the coefficient estimates from OLS will be inconsistent. If the underlying latent equation has two outcomes, we are in the standard Heckman selection world where the correction term (the inverse Mill's ratio) is included in the wage equation that yields consistent coefficient estimates. In the case where the selection equation is estimated using a multinomial logit, we run OLS on the wage equation (2.13) including an analogue of the inverse Mill's ratio (λ_j) as given below :

$$E(w_j|I = j) = \mathbf{x_j}\beta_\mathbf{j} + \delta_j\lambda_j + \vartheta_j \tag{2.14}$$

where,

$$\lambda_j = -\phi[\Phi^{-1}[F_j(\mathbf{z_j}\gamma_\mathbf{j})]]/F_j(\mathbf{z_j}\gamma_\mathbf{j}) \ \ and \ \ \delta_j = \sigma_j\rho_j$$

and σ_j is the variance of u_j and ρ_j is the correlation between u_j and $\epsilon_j^*(= \Phi^{-1}(F_j(\epsilon)\,)$. The error term ϑ_j has a zero mean and uncorrelated with u_j. Since parametric form of the variance and covariance matrix is difficult to derive, the standard errors are calculated by bootstrap methods.

Hay (1980) develops a similar approach for the polychotomous choice model where the conditional expectation of u_j conditional on the disturbances in the selection equation are assumed to linear. This linearity of the conditional expectation is a stronger assumption imposed on the model. By contrast, the Lee model does not impose such restrictions. When the linearity assumption is imposed, the analogue of the inverse Mill's ratio becomes :

$$\lambda_j = 6/\pi^2(-1)^{j+1}[\sum_{k\neq j}(1/J).(\frac{p_k}{1-p_k}).logp_k + (J-1)/J\,logp_j] \tag{2.15}$$

where $p_j = F_j(\mathbf{z_j}\gamma_\mathbf{j})$. Consistent estimates of the wage equation is derived by running OLS on (2.13) by replacing λ_j as in (2.15). [26]

2.5.3 Identification Strategy

The identification of the model is achieved by correctly specifying the selection equation and including at least one variable in the selection equation that is excluded from the wage equation. So we need to find at least one variable that affects sectoral choice but does not affect wages in that sector. I use two variables for identification purposes: number of children in the household less than 15 years of age and number of non earning elderly members in the household greater than 60 years of age. Grootaert and Mundial (1988) and Günther and Launov (2012) have used number of dependents for the identification of their model.

Theoretically, dependents do play a role in determining the sectoral choice of a worker, however the question on how it affects the sectoral decision is an empirical one. A worker with many dependents may be more likely to prefer formal employment because he values benefits like social security and job security more compared to workers with fewer dependents. So higher the number of dependents, greater is the probability of accepting formal over informal employment. On the other hand, higher the number dependents, more desperate the worker is to accept a job and earn his livelihood. In such as scenario the worker reduces his search time for formal employment and settle for informal employment. Thus the question on how the number of dependents affects sectoral choice is an empirical question that cannot be ascertained a priori. Further, children and elderly

[26] See Hill (1989) and Khandker (1992) for applications of this approach.

may have differential effects on the sectoral employment choice. To exploit this extra variation in the sample, I have included separate variables for children and elderly dependents rather than including the total number of dependents.

Pratap and Quintin (2006) have used the presence of a relative in the formal sector as the exclusion restriction. Since the share of formal workers in the entire labor force is very small, the variation obtained by including a dummy variable for a relative in the informal sector is negligible. In fact, every worker in the sample I use, has at least one relative who is an informal worker. For this reason, I chose the number of dependents over a relative who is an informal worker for identification purposes.

2.5.4 Test for Labor Market Segmentation : Interpretation of δ_j

In equation 2.14, the coefficient on the selection correction term λ_j in the wage equation for the j^{th}sector is δ_j. The coefficient δ_j has the same sign as ρ_j (since $\delta_j = \sigma_j \rho_j$ and $\sigma_j > 0$), which is the correlation coefficient between the errors in the original wage equation (u_j) and the transformed variable ϵ_j^*. Further since, ϵ_j^* is the standardized transformation of ϵ_j, they are directly proportional to each other. This proportionality implies that the correlation coefficient ρ_j (between $u_j$and ϵ_j^*), have the same sign as the correlation coefficient between u_j and ϵ_j. However, from equation (2.8) we know that ϵ_j and the error term in the selection equation (η_j) are negatively correlated. Thus a negative ρ_j implies a positive correlation between the errors in the wage equation (u_j) and the errors in the selection equation (η_j). A negative sign on δ_j (that has the same sign as ρ_j) means a positive correlation between u_j and η_j. The economic interpretation of the correlation between u_j and η_j is discussed below. [27]

Hypothesis 1: Sectoral Choice is Voluntary

The error terms u_j and η_j include unobservable characteristics of a worker. A positive (negative) correlation between between u_j and η_j (i.e $\delta_j < 0$) implies that unobserved worker characteristics that increases the worker's probability of choosing sector j, increases his wage in that sector. Worker's innate ability or productivity can be thought of as one of the unobserved factors that

[27] I follow Gindling (1991), Grootaert and Mundial (1988), Khandker (1992) and Zhang (2004) for the economic interpretation.

cannot be controlled by any observed variable. In that case, a positive correlation means that a more productive worker (in terms for unobserved innate ability) who has higher probability of selecting into sector j (controlling for observed characteristics) also earns higher wages in that sector. This implies that productive workers competitively selects sector j that offers them higher wages.

Another interpretation of δ_j is the wage difference between the workers who self select into sector j, and a randomly chosen worker in that sector. Since λ_j is negative by construction, a negative δ_j (i.e $\delta_j \lambda_j > 0$) means that a worker who self selects into sector j earns higher wages than a randomly chosen worker in that sector. Thus a negative sign on δ_j implies that workers competitively choose sector j that pays them higher rewards for their unobserved productivity reflected by higher wages in that sector.

Hypothesis 2: Adverse Selection

On the other hand, a negative correlation implies that a higher productive worker (in terms of innate ability) who has a higher probability of selecting into sector j earns lower wage in that sector. Alternatively, a lower productive worker who has a lower probability of selecting sector j earns higher wages in that sector. Thus a negative correlation implies adverse selection in sector j, where lower productive workers who select sector j earns higher wages and higher productive workers earn lower wages.

In case of a negative correlation, the term self-selection is not quite appropriate because the selection mechanism is not competitive. If a more productive worker knows that he has increased probability of entering sector j and will earn lower wages in that sector, then it may not be rational for the worker to self select into that sector. Thus a negative correlation can only happen if there is information asymmetry between the employers and the workers. Since the employers in sector j do not have perfect information on the worker's productivity, they offer lower productive workers higher wages and the higher productive workers lower wages leading to adverse selection in that sector. This argument is based on the assumption that worker's utility is linear in wages that may not be the case. A worker may receive better non-wage benefits even though receiving lower wages in a particular sector. This adverse selection argument is put forward by Grootaert and Mundial (1988).

Alternatively, a positive sign on δ_j (i.e $\delta_j \lambda_j < 0$) means that a worker who self selects into sector j earns lower wages than a randomly selected worker in that sector. Thus a positive sign on δ_j means that self-selection into sector j is not a rational choice for the workers as they do not receive higher rewards for their unobserved productivity reflected by lower earnings. This can only happen if there is information asymmetry about the worker's productivity between the employers and the workers as argued in the previous paragraph.

Hypothesis 3: Entry Barriers Exist and Sectoral Allocation is Involuntary

A third possibility arises if δ_j is statistically insignificant that implies three things. First, a statistically insignificant $\delta_j$means no evidence of self-selection in that sector. Second, it could be the case the the the selection process is not well identified. And lastly, as Gindling (1991) suggests, an insignificant δ_j means that worker's unobserved ability that affects his probability of choosing sector j does not affect his wages. If the coefficients of the sector allocation model are significantly different from zero as a group, then sector allocation is non random. If workers were free to choose the sectors, then they would choose the sector that pays higher rewards for their unobserved productivity. But in this case, the non random allocation of workers into sector j does not affect the wages in that sector. In other words, productivity of workers that affects the probability of a worker choosing sector j does not affect wages in that sector. Thus, there is no sample selection bias even though the sector allocation mechanism is non random. Thus, in our context if the coefficient on the selection correction term for the formal sector is statistically insignificant but the sector allocation is non random (the coefficients of the sector allocation model are significantly different from zero as a group) it means that workers do not have full access to formal employment. Evidence from counterfactual wages that show expected formal wage are higher than informal wages will further corroborate this explanation. A higher expected formal wage relative to the informal wage along with the evidence that sector allocation is non random, will provide definitive evidence of restricted entry into formal employment.

2.6 Results

2.6.1 The Formal-Informal Wage Gap

In the first model, I estimate the wage gap between formal and informal workers at different quantiles of income. I have used the MM technique to decompose the wage gap into coefficient and endowment effects. The underlying assumption of this model is that workers are randomly selected into formal and informal employment, hence there is no selectivity bias.

Tables 2.10 and 2.11, report the wage gap decomposition results for males and females respectively. For both males and females, significant wage gap exists between formal and informal workers across the wage distributions. For males, coefficient effects explain the major part of the wage gap between the 10th and the 40th quantiles. At higher quantiles, the endowment effect explains major part part of the wage gap. For females, the endowment effect explains the major part of the wage gap across the whole distribution. However, the contribution of coefficient effect cannot be ignored because even at the 90th quantile it explains around 40 percent of the formal-informal wage gap for males (47 percent for females). Two things can be inferred from these results. First, informal workers face a significant wage penalty across the wage distribution. The persistent wage gap at the higher quantiles of the wage distribution shows that informal workers earn significantly less than the formal counterparts even at the higher end of the wage distribution. Thus, workers at the upper end of the wage distribution do not receive any benefits in terms of higher wages that may induce them to choose informal employment voluntarily. This result is contrary to the argument given by Fields (1990) that informal employment may be a voluntary choice for the upper tier informal workers.[28] Second, both coefficient and endowment effects play significant roles in explaining the formal-informal wage gap. The contribution of the coefficient effects show that an informal worker with the same set of skills as a formal worker earns significantly less because he receives lower returns on his skills and individual characteristics. The contribution of the coefficient effect to the estimated wage gap insinuates to some form of discrimination against the informal workers in the labor market. The contribution of endowment effect to the estimated wage gap shows that informal

[28] A literature survey by Chen et al. (2004) provides evidence of informal wage penalties for Egypt, El Salvador and South Africa. Marcouiller et al. (1997) find significant wage premiums for formal sector workers in El Salvador and Peru, but find wage premiums for informal workers in Mexico. Using, Brazilian labor market data, Tannuri-Pianto and Pianto (2002) showed that the earnings gap between formal and informal workers are wider at the lower quantiles than at the high ones. Thus, the results presented here conform to the evidence from other developing countries.

workers are systematically different from formal workers. Basic education and skill level may be important factors that give rise to these differences. Overall, the wage gap decomposition results show that the informal workers earn less than the formal workers not only because they are less skilled, but also because they face discrimination as they receive lower returns to their endowments compared to the formal workers.. In the next section, I discuss the results of the polychotomous choice model.

2.6.2 Selection Corrected Estimates

The results presented in Section 2.6.1 are based on the assumption that sector allocation of workers is random. So, it will not be correct to base our predictions based on this specification because it does not take into account self-selection on the part of the workers. The self-selection issue may bias the estimates reported in Tables 2.10 and 2.11. To derive consistent estimates of the wage equation, we correct for sample selection bias. The estimation process is done in two stages. First, a multinomial logit equation is estimated for the selection equation. In the second stage, the wage equation is estimated separately for formal and informal workers including the selection correction term. The first stage and second stage results are reported separately in the following two subsections.

2.6.2.1 Multinomial Logit Estimates of the Sectoral Allocation Equations

Tables 2.12 and 2.13 report the multinomial logit estimation for males and females respectively. In this model a worker faces four choices : formal employment, informal employment , self employment and to stay out of the labor force.

To achieve identification and circumvent the problem of large second stage standard errors I include two variables in the selection equation that are excluded from the wage equation: children under 15 years of age and non earning-elderly persons more than 60 years of age. All other control variables that are included in the wage equation are also included in the selection equation. The coefficient estimates for each category reported in Tables 2.12 and and 2.13 are the log odds with respect to formal employment ,which is the base category. The marginal effects of each variable are also reported in the tables. A first look at the tables tells us that allocation of workers to the segments is non-random. A (χ^2) test rejects the null hypothesis that coefficients are jointly equal to zero.

For men, an additional year of schooling increases the probability of formal employment by 1.1 percent and decreases the probability of informal employment by 2.3 percent.[29] Having some technical education increases the chances of formal employment by 1.9 percent, however, it also increases the chances of informal employment by 9.2 percent. But a person having technical education is less likely to be self employed or stay out of the labor force. Formal vocational training increases the probability of formal employment by 2 percent but increases the chances of informal employment by only 0.6 percent. Like technical education, a person having formal vocational training decreases the probability of self employment or staying out of the labor force. For women (see Table 2.13), an additional year of schooling does not have a significant effect on the sector allocation. Technical education increases the chances of informal employment by 6.5 percent but increases the chances of formal employment by only 0.6 percent. Formal vocational training increases the chances of informal employment by 3.8 percent, but increases the chances of formal employment by 0.5 percent. Thus, technical education and vocational training seems to have a greater favorable effect on the likelihood of informal employment than formal employment. These results, particularly with respect to technical education and vocational training may seem counter-intuitive at first, but they provide useful insights to the central idea of the paper. First, it means that skill development programs (technical education and vocational training) are ineffective in making workers employable for formal employment. Second, it insinuates to entry barriers and rationing of formal jobs that restricts even those workers with the adequate skills from entering formal employment. It is is because of rationing of formal jobs and entry barriers that workers having technical education and vocational training are not able to enter formal employment. The formal workers with technical education and vocational training may be considered as the lucky ones, a term commonly used in the segmentation literature.

For men, an additional child in the household decreases the likelihood of formal employment by only 0.5 percent and informal employment by 1 percent. However, the probability of self employment increases by 2 percent for a person having one more child. Having one more elderly dependent

[29] Results for men refer to Table 2.12. Results for women refer to Table 2.13

in the household decreases the probability of formal employment by 0.5 percent and informal employment by 3.8 percent, but increases the probability of self employment by 2.8 percent. Thus, both categories of dependents seem to have a larger negative effect on informal employment than formal employment. So, a person with greater number of dependents is more likely to be formally employed. As was discussed earlier, the role of dependents on the sectoral choice is an empirical question. Greater number of dependents may increase the likelihood of formal employment because the workers would value the employment benefits accompanying formal employment. On the other hand, greater number of dependents may increase chances of informal employment because the worker would be more likely to reduce search length and avoid long spells of unemployment. Since the results show that workers with more dependents are more likely to be formally employed, it conforms to the hypothesis that workers with more dependents value benefits of formal employment. For women, number of dependents have lesser effects on the sectoral choice.

The control variables include a number of individual characteristics like religion, caste, marital status and the region of residence. Religion plays an important role in sector allocation for males compared to females. Muslim men are 1.1 percent less likely to be formally employed and 4.3 percent less likely to be informally employed than Hindus. However, Muslim men are 6.5 percent more likely to be self employed than Hindus. Caste is an important attribute that can be used for discrimination and hence may play an important role in sector allocation. Amongst men, scheduled tribes are 5.2 percent more likely to be formally employed, 8.7 percent more likely to be informally employed and 14.9 percent less likely to self employed than other higher castes. Scheduled castes are 1.8 percent more likely to be formally employed, 12.9 percent more likely to be informally employed and 14 percent less likely to be self employed than other higher castes. For females, however, caste is a comparatively less significant factor in sector allocation than men. Marital status also plays an important role in sectoral allocation. For men, those who are currently married compared to those who never married are 1.5 percent less likely to be formally employed, 3 percent less likely to be informally employed and 9 percent less likely to be self employed. For females, those who are currently married compared to those who are never married are 0.2 percent less likely to be formally employed, 6.5 percent less likely to informally employed and 4 percent less likely to be self employed. Thus, sector allocation of workers is not independent of the worker's background characteristics.

Sectoral allocation may depend on whether a worker resides in rural or urban areas. Males living in urban areas, are 0.9 percent less likely to be formally employed and 2.7 percent more likely to informally employed compared to those who live in rural areas. For females also, living in the urban area proves unfavorable for formal employment. Rural to urban migration may be one of the factors that may produce this kind of result. Since, the data set does not allow us to distinguish between the migrant and non-migrant workers it is hard to substantiate these results based on migration patterns.

2.6.2.2 Selection Corrected Estimates of the Wage Equations for Formal and Informal Workers

Table 2.14 presents the formal and informal wage equation estimates for males. Columns 1 and 4 report the OLS estimates for formal and informal workers respectively. Columns 2 and 5 report the selection corrected estimates for formal and informal wages respectively.[30]

Selection corrected returns to schooling are similar for formal and informal workers. An additional year of schooling increases the weekly formal and informal wage by 4.6 percent. Technical education has a higher return for informal workers than formal workers. Formal vocational training has no statistically significant effect on formal wages, but has a positive significant effect on informal wages. Informal vocational training has no statistically significant effect on formal wages but has a negative and significant effect on informal wages. Although returns to general education as measured by years of schooling are similar for formal and informal workers, returns to any specialized training like technical education and vocational training are significantly different for formal and informal workers. These results are consistent with the segmentation literature because similar returns to human capital variables in the two sectors are neither necessary nor sufficient for the segmentation hypothesis (Dickens and Lang (1985)). Here similar returns to general education do not mean that informal workers earn as much as formal workers because the quantile regression estimates show significant wage gap between formal and informal workers at all quantiles of the wage distribution. However, higher returns to technical education for informal workers than formal

[30]Estimates reported in Columns 3 and 6 correspond to Hay's methodology that will be discussed in the next section under robustness check.

workers may seem an aberration at first, but this result can be explained by the fact that the government and public sector employ majority of the formal workers. The eligibility criteria for government jobs is based on general education levels. Thus a worker qualifying for a government job based on the basic eligibility criteria, does not earn higher returns for technical education and vocational training.[31] Returns to experience (measured by age) are almost the same for formal and informal workers.

Religion does not play a significant role in determining the formal wage, but is important for the informal wage. Scheduled castes and other backward classes face wage penalties for both formal and informal workers compared to other higher castes. Marital status does not have any significant effect on the formal wage. But married informal workers earn 15 percent more compared to unmarried informal workers. Formal workers located in the urban regions earn 14 percent more than formal workers located in the rural regions. Informal workers located in the urban region earn 10.2 percent more than informal workers located in rural regions.

Table 2.15 reports the formal and informal wage estimates for female workers. As in Table 2.14, columns 1 and 4 report the OLS estimates for formal and informal wages respectively. Columns 2 and 5 report the selection corrected estimates for formal and informal wages respectively.[32]

Selection corrected returns to schooling has no significant effect on the formal wage. However, returns to schooling for informal workers are 6.7 percent. Technical education has no significant effect on formal workers but has positive and significant returns for informal workers. Formal vocational training has a negative premium attached with respect to formal wage while it has no significant effect on informal wage. The same argument follows as with men that the government jobs cover majority of formal employment, where the minimum eligibility criteria is based on general education levels. The negative premium on formal vocational training for formal workers seems to be an aberration. However, it can be argued that the quality of the training programs in the

[31] Duraisamy (2002) estimated returns to education for workers in wage employment using NSSO data for the years 1983 and 1993. He finds that men receive 6.4 percent, 15.7 percent and 8.9 percent returns on middle, secondary and higher secondary levels. Women received 10.3 percent , 33.7 percent and 11.8 percent returns for the same levels of education. Azam (2012) used NSSO data for the years 1983, 1993 and 2004 to estimate the returns to education. He uses quantile regression techniques and finds that returns to education have increased during the period 1983-2004. Both the studies do not make a distinction between formal and informal workers. The estimates reported in this paper are slightly different from the earlier studies because, I use years of schooling as a measure of education and they use dummy variables for each level of education.

[32] As in Table 2.14, estimates reported in Columns 3 and 6 correspond to Hay's methodology that will be discussed in the next section under robustness check.

formal vocational training institutes is not good enough to provide workers higher returns on their training. Experience (as measured by age) has no significant effect on formal wage, but the returns to experience is positive for informal workers.

Religion is not a significant factor that affects the formal wage. However, female Muslim informal workers earn less than Hindus. On the other hand, Christian informal workers earn more than Hindus. Schedules castes and other backward classes face a wage penalty compared to other higher castes in both formal and informal employment. Martial status has no significant effect on formal wage. But married informal workers earn less compared to unmarried informal workers. Urban formal workers earn 37.7 percent more than rural formal workers. Urban informal workers earn 22.8 percent more than rural informal workers.

The wage estimates show that significant difference exists between the formal and informal wage determination mechanisms. However, two different wage determination mechanisms do not mean that the labor market is segmented. If workers can freely move between the types of employment i.e if there are no entry barriers, they will choose the sector that pays them higher wages. We have already seen that informal workers earn significantly less than formal workers across the wage distribution. Thus, the situation where workers choose informal employment because it pays them higher wages is less likely. The next section provides evidence of entry barriers and rationing of formal jobs that substantiates the segmentation hypotheses.

2.6.2.3 Evidence for Labor Market Segmentation

Section 2.5.4 discussed the implications of the sign and the statistical significance of the estimated coefficients on the selection correction terms (λ_{Formal} and $\lambda_{Informal}$) in the wage equations. For the formal wage equation (for both males and females in Tables 2.14 and 2.15 respectively) the estimated coefficient on λ_{Formal} is positive but statistically insignificant. A statistically insignificant coefficient on λ_{Formal} implies no evidence of self-selection in the formal sector. In other words, a worker's unobserved productivity that increases his chances of formal employment does not affect his wages. Also there is no difference between a randomly selected formal worker and a worker who self selects into formal employment. No evidence of self-selection into formal employment insinuates to the fact that entry barriers may be present in the formal sector. However, it could also be the case that the model is not identified properly that would require some further testing of the model with alternative identification strategies. Decomposition of the wage gap into characteristics and coefficient effects may give further insights to labor market segmentation hypothesis. I discuss the results in the next section.

The coefficient on $\lambda_{Informal}$ is negative and statistically significant. A negative coefficient on $\lambda_{Informal}$ implies that informal employment is a competitive choice for the worker. Workers can self select into informal employment and increase the informal wage. Also, a worker self selecting into informal employment earns higher wages than a randomly selected informal worker.

Based on these two findings we cannot reject the hypothesis that the labor market is segmented. This is because there is no evidence of self-selection into formal employment. If on the contrary we had found that workers can self select into both formal and informal employment we could have safely rejected the labor market segmentation hypothesis. Since in that case, the workers would be free to choose between formal and informal employment. They would choose the sector that maximizes his utility. Decomposition of the wage gap into coefficient and characteristic effect may provide further insights to this argument. However, I have to constrain myself to the decomposition at the mean. The underlying self-selection model is a polychotomous choice model that does not allow the implementation of the MM decomposition technique across the quantiles of the wage distribution. I have used the Oaxaca-Blinder technique to decompose the formal-informal wage gap into coefficient and characteristics effect.

2.6.2.4 Wage Gap Decomposition Adjusting for Self-Selection (Oaxaca-Blinder)

Table 2.16 shows the wage gap decomposition into coefficient and and endowment effects. The first column shows the results for males and the second column for females. For both males and females, the formal-informal wage gap is higher when corrected for self-selection. For males, the coefficient effect explains 66 percent of the wage gap and the endowment effect explains the rest 34 percent. For females, the coefficient effects explain 84 percent of the formal informal wage gap. So, at the mean, the informal workers face a wage penalty because they receive lower returns on their human capital and individual characteristics. So formal workers enjoy higher returns on their human capital and individual characteristics. This finding is compatible with the efficiency wage theory in the sense that firms pay higher wages to formal workers that is set higher than than the market clearing wage to incentivize the worker. So, formal workers who receive higher returns to their skills and individual characteristics have less incentive to move elsewhere. On average they could also be more productive than the informal workers because they receive higher returns.

A significant part of the wage gap is explained by the differences in characteristics of the formal and informal workers. These differences may stem from skill differences, age and other individual characteristics. So, if workers with different characteristics have different labor supply elasticities, firms can exploit to opportunity to exercise their monopsonistic power. This line of argument is also compatible with labor market segmentation because firms having monopsonistic power can now pay different wages to the workers with different labor supply elasticities. In our case, it may be optimal for the firms to pay different wage to formal and informal workers because they have differ in skills and individual characteristics and hence have different labor supply elasticities.

Figures 2.2 and 2.3 are the kernel density plots of the predicted formal wage and the observed informal wage for the informal male workers and informal female workers respectively. The counterfactual is based on the differences in returns to human capital and individual characteristics. That is, how much the informal workers would have earned if they had received similar returns on human capital and individual characteristics as the formal workers. For males, we see that predicted mean and median formal wage is higher than the observed informal wage. Also, 85 percent of the male informal workers (83 percent of female informal workers) would have earned higher wages in formal employment than their observed informal wage. Given that informal workers do not receive any non-wage benefits as formal workers, the result shows the majority of the informal workers would choose formal employment over informal employment.The other 15 percent of the male informal workers (17 percent of the female informal workers) would have earned less as formal workers. These few people could be regarded as voluntary informal workers who would have chosen informal employment even if they had full access to formal employment. No discernible differences in observed characteristics can be found between the involuntary informal workers and the voluntary informal workers. But majority of the voluntary informal workers are employed by the proprietary and the government enterprises. So, it could be the case that these enterprises pay a higher wage to informal workers instead of providing defined employment benefits to the workers. These individuals could also posses entrepreneurial skills that are rewarded by the proprietary enterprises that compensates for the employment benefits that accompany formal employment.

2.6.2.5 Robustness Check

I conducted the robustness check for my results by employing the selection correction strategy proposed by Hay (1980). This methodology imposes stronger restrictions on the model by assuming linearity of the conditional expectation of the error terms in the wage equation conditional on the errors in the selection equation. But the results should not be different from our baseline model. The selection correction terms constructed from the first stage multinomial logit estimates are given by equation 14 in Section 2.6.2. Columns 3 and 6 in Table 2.14 and 2.15 report the wage estimates employing Hay's method. For both males and females, the results are similar to the baseline model. The coefficients on $\lambda_{Formal_{Hay}}$ are statistically insignificant that provides evidence for entry barriers in the formal employment. The coefficients on $\lambda_{Informal_{Hay}}$ are negative and significant that implies informal employment is a competitive outcome given that workers cannot self select into formal employment. Moreover, the coefficient estimates from both models are almost the same. Thus, the results presented in this paper are robust and invariant to the other polychotomous choice models used in the literature.

2.7 Conclusions

The results in this paper show that the Indian labor market is segmented between formal and informal employment. Evidence does not support the hypothesis of a fully competitive labor market and that workers choose informal employment as a last resort. This finding is in strong contrast to the empirical evidence from some Latin American countries that show informal employment is a competitive choice for workers over formal employment. The paper shows that majority of the workers would have earned more as formal workers than their current informal wage if they had full access to formal employment. Also, informal workers earn significantly less than formal workers as reflected by the wage gap between the two sets of workers at all quantiles of the wage distribution. Thus, the finding refutes the competitive hypothesis for majority of the informal workers. The current empirical and theoretical literature on informal employment is mainly based on Latin American countries. Evidence amassed in this paper show that the Indian labor market is systematically different from Latin American countries providing the platform for further research on both empirical and theoretical grounds focusing on the South Asian countries.

The wage gap decomposition results show that both coefficient and endowment effects explain significant portion of the formal-informal wage gap at all quantiles of the wage distribution. This evidence suggests that informal workers earn less than formal workers not only because they have different characteristics and human capital, but also because they receive lower returns to their endowments than their formal counterparts. Recent policies focus on skill development to address the problem of informal employment. However, policies should also address the problem of discrimination faced by informal workers as firms pay lower returns to their skills and characteristics compared to observationally similar formal workers. Moreover, results show that workers having technical education and vocational training do not have a clear advantage to enter formal employment. The Unorganized Sector Worker's Social Security Bill passed in 2007 was an important and directed step towards reducing informal employment. However, the efficacy of this initiative can be questioned. This is because, 85.8 percent of the labor force still continues to be informally employed even after four years since the bill was passed.

This paper finds no evidence of self-selection into formal employment. However, it will be interesting to see what kinds of entry barriers that workers face. Monopsonistic discrimination, efficiency wages and search frictions are some of probable causes that may give rise to discrimination and entry barriers that workers face. Ito (2009) finds that transaction costs in the Indian labor market (that includes actual expenditure and time spent traveling for finding employment) are higher for socially backward classes. On the other hand, he finds no evidence of wage discrimination in regular employment that implies caste based discrimination takes the form of job based discrimination. This job based discrimination limits the range of available jobs to some groups of people. Social norms may also play an important role on labor market outcomes. Besley and Burgess (2004) show that labor regulations in India have a significant effect on informal sector activity that hints at institutional barriers in the labor market. This paper lays the foundation for further research on the specific types of entry barriers and their impact on informal employment.

Tables

Table 2.1: Classification of Informal Employment and Informal Sector

Production Units	Informal jobs	Formal Jobs
Informal sector enterprises	A	B
Other units of production	C	D

A+C = Persons in Informal Employment

A+B = Persons Employed in the Informal Sector

C = Informal Employment outside the Informal Sector

B = Formal Employment in the Informal Sector

Table 2.2: Identification of Formal/Informal Sector, by Type of Enterprise

Enterprise type	Frequency	%
Informal sector		
1. Proprietary (male)	55702	62.3
2. Proprietary (female)	4318	4.7
3. Partnership (members from same hh)	1833	1.9
4. Partnership (with members from different hh)	1265	1.6
8. Employer's households(i.e., private households)	1290	1.6
9 . Others	4826	5.3
Formal sector		
5. Government/public sector	18591	12.8
6. Public/Private limited company	5902	8.5
7. Co-operative societies/trusts	1198	1.2
Total	94639	100.0

Table 2.3: Identifying Broad Categories of Employment, by Principal Activity Status

Principal activity status	Frequency	%
Self employed		
1. Self-employed (own account worker)	30807	11.5
2. Self-employed (Employer)	1320	0.5
3. Self-employed (unpaid family member)	7232	2.6
Employees		
4. Regular salaried/ wage employee	37329	14.4
5. Casual wage labor (public works)	902	0.4
6. Casual wage labor (other types)	17219	8.5
Not in the labor force		
7. Did not work but was seeking for work	5508	1.9
8.Attended educational institution	42415	18.5
9. Attended domestic duties only	48902	21.9
10. Attended domestic duties & other activities	33378	18.3
11. Rentiers, pensioners, remittance recipients	1263	0.6
12. Others (begging, prostitution,etc.)	1564	0.8
Total	227839	100

Table 2.4: Percentage of Workers Receiving Employment Benefits

Employment benefits	Social security	Job contract	Paid leave	All benefits
	%	%	%	%
No	69.4	77.1	32.0	82.6
Yes	30.6	22.9	68.0	17.4
Total	100	100	100	100
Observations		55,450		

Table 2.5: Formal and Informal Employment (Treating Self Employed as a Separate Category)

	Male		**Female**		**All**	
Categories	Frequency	%	Frequency	%	Frequency	%
Formal employment	11087	7.4	2377	1.3	13464	4.1
Informal employment	34921	35.9	7065	5.6	41986	19.3
Self employed	33307	27.3	6052	4.3	39359	14.7
Not in labor force	29904	29.4	103126	88.8	133030	62.0
Total	109219	100.0	118620	100.0	227839	100.0

Table 2.6: Identification of Formal and Informal Sector and Employment Based on Enterprise Type and Activity Status

<table>
<tr><td rowspan="3"></td><td colspan="5">Self employed</td><td colspan="2"></td></tr>
<tr><td colspan="2">Own account worker</td><td colspan="2">Employers</td><td>Contributing family mbrs.</td><td colspan="2">Employees</td></tr>
<tr><td colspan="2">(Activity Status:1)</td><td colspan="2">(Activity Status:2)</td><td>(Activity Status:3)</td><td colspan="2">(Activity Status:4,5,6)</td></tr>
<tr><td></td><td>Formal</td><td>Informal</td><td>Formal</td><td>Informal</td><td>Informal</td><td>Formal</td><td>Informal</td></tr>
<tr><td>Formal sector
(Enterprise types: 5,6,7)</td><td>A</td><td>NE</td><td>B</td><td>NE</td><td>C</td><td>D</td><td>E</td></tr>
<tr><td>Informal sector
(Enterprise types: 1,2,3,4,8,9)</td><td>NE</td><td>F</td><td>NE</td><td>G</td><td>H</td><td>I</td><td>J</td></tr>
</table>

Source:Sastry (2004) and own calculation

Table 2.7: Formal and Informal Employment in the Formal and Informal Sectors

Employment type, by sector	Frequency	%
Formal employment		
1. Formal employment in formal sector (A+B+D)	12770	4.4
3. Formal employment in informal sector (I)	737	0.1
Informal employment		
2. Informal employment in formal sector (C+E)	12797	8.7
4. Informal employment in informal sector (F+G+H+I+J)	68505	24.8
5. **Not in labor force**	133030	62.0
Total	227839	100.0

Table 2.8: Summary Statistics (Males)

	Formal		Informal		Self employed		Not in labor force	
	mean	sd	mean	sd	mean	sd	mean	sd
Wages and Consumption								
Weekly Wage (Rupees)	4966.4	3363.4	1771.4	1889.2	.	.	.	.
Log weekly wage (Rupees)	8.3	0.6	7.2	0.7	.	.	.	.
MPCE (Rupees)	2902.5	2621.5	1714.4	1420.8	1806.4	1585.1	1970.2	1859.1
Log MPCE	7.8	0.6	7.3	0.6	7.3	0.6	7.4	0.6
Age	42.1	9.6	34.2	10.5	37.2	10.3	20.1	7.4
Education								
Years of Schooling	12.5	4.0	7.4	5.2	8.0	4.9	10.1	3.2
Technical education	0.2	0.4	0.05	0.2	0.03	0.2	0.04	0.2
No vocational training	0.8	0.4	0.8	0.4	0.8	0.4	0.9	0.2
Vocational training: formal	0.1	0.3	0.05	0.2	0.04	0.2	0.05	0.2
Vocational training: informal	0.07	0.3	0.2	0.4	0.2	0.4	0.01	0.1
Dependents								
No of Children in the hhd (¡15 years)	1.1	1.2	1.3	1.4	1.5	1.5	0.9	1.3
No. of elders (¿60 years)	0.2	0.5	0.2	0.5	0.3	0.6	0.3	0.5
Caste								
Scheduled tribe	0.06	0.2	0.06	0.2	0.03	0.2	0.07	0.3
Scheduled caste	0.1	0.4	0.2	0.4	0.1	0.3	0.2	0.4
Other backward class	0.3	0.5	0.4	0.5	0.4	0.5	0.4	0.5
Others	0.5	0.5	0.3	0.5	0.4	0.5	0.3	0.5
Marital status								
Never married	0.1	0.3	0.3	0.4	0.2	0.4	0.9	0.3
Currently married	0.9	0.3	0.7	0.5	0.8	0.4	0.06	0.2
Widowed	0.009	0.10	0.01	0.1	0.01	0.1	0.005	0.07
Divorced/separated	0.002	0.04	0.004	0.06	0.002	0.05	0.0009	0.03
Regional variables								
Urban	0.7	0.5	0.5	0.5	0.5	0.5	0.4	0.5
North	0.2	0.4	0.2	0.4	0.3	0.4	0.3	0.4
West	0.3	0.4	0.2	0.4	0.2	0.4	0.2	0.4
East	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4
North-East	0.03	0.2	0.03	0.2	0.04	0.2	0.04	0.2
Central	0.07	0.3	0.06	0.2	0.06	0.2	0.07	0.3
South	0.2	0.4	0.3	0.4	0.2	0.4	0.2	0.4
Observations				109219				

Notes: The summary statistics for religion dummies are not reported in this table due space constraint. See section 2.4.2.2 for the discussion on the distribution of religious groups.

Table 2.9: Summary statistics (Females)

	Formal		**Informal**		**Self employed**		**Not in labor force**	
	mean	sd	mean	sd	mean	sd	mean	sd
Wages and Consumption								
Weekly Wage (Rupees)	4330.9	2956.7	1367.6	1741.4	.	.	.	.
Log weekly wage (Rupees)	8.1	0.8	6.8	0.9	.	.	.	.
MPCE (Rupees)	3651.7	3559.2	2007.2	2164.2	1756.4	1599.2	1794.4	1662.4
Log MPCE	8.0	0.7	7.4	0.6	7.3	0.6	7.3	0.6
Age	39.0	9.9	34.9	10.4	36.2	10.8	32.0	12.1
Education								
Years of Schooling	13.0	4.1	7.3	6.2	5.6	5.1	6.7	5.1
Technical education	0.2	0.4	0.06	0.2	0.03	0.2	0.01	0.1
No vocational training	0.8	0.4	0.8	0.4	0.7	0.5	1.0	0.2
Vocational training: formal	0.2	0.4	0.07	0.3	0.06	0.2	0.02	0.1
Vocational training: informal	0.04	0.2	0.09	0.3	0.2	0.4	0.03	0.2
Dependents								
No of Children in the hhd (¡15 years)	1.0	1.1	1.1	1.3	1.3	1.3	1.4	1.5
No. of elders (¿60 years)	0.4	0.6	0.2	0.5	0.2	0.5	0.3	0.6
Caste								
Scheduled tribe	0.06	0.2	0.09	0.3	0.05	0.2	0.06	0.2
Scheduled caste	0.1	0.4	0.2	0.4	0.2	0.4	0.2	0.4
Other backward class	0.3	0.5	0.4	0.5	0.5	0.5	0.4	0.5
Others	0.5	0.5	0.3	0.4	0.3	0.5	0.3	0.5
Marital status								
Never married	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4
Currently married	0.7	0.5	0.6	0.5	0.7	0.5	0.7	0.4
Widowed	0.2	0.4	0.1	0.4	0.1	0.3	0.03	0.2
Divorced/separated	0.02	0.1	0.03	0.2	0.02	0.1	0.003	0.05
Regional variables								
Urban	0.7	0.5	0.5	0.5	0.5	0.5	0.3	0.5
North	0.2	0.4	0.1	0.3	0.1	0.4	0.3	0.4
West	0.3	0.4	0.2	0.4	0.2	0.4	0.2	0.4
East	0.2	0.4	0.1	0.3	0.2	0.4	0.2	0.4
North-East	0.03	0.2	0.02	0.1	0.03	0.2	0.04	0.2
Central	0.05	0.2	0.07	0.3	0.05	0.2	0.07	0.3
South	0.3	0.5	0.4	0.5	0.4	0.5	0.2	0.4
Observations				118620				

Notes: The summary statistics for religion dummies are not reported in this table due space constraint. See section 2.4.2.2 for the discussion on the distribution of religious groups.

Table 2.10: Wage Gap Decomposition (Males)

Quantiles	Wage gap (raw)	Wage gap (predicted)	Endowment effect	Coefficient effect
10	-1.101	-1.155	-0.536	-0.619
		(0.025)	(0.024)	(0.034)
20	-1.299	-1.245	-0.592	-0.652
		(0.018)	(0.022)	(0.028)
30	-1.338	-1.278	-0.638	-0.640
		(0.015)	(0.022)	(0.027)
40	-1.325	-1.284	-0.673	-0.611
		(0.014)	(0.022)	(0.026)
50	-1.288	-1.267	-0.692	-0.574
		(0.015)	(0.021)	(0.025)
60	-1.273	-1.229	-0.693	-0.536
		(0.015)	(0.020)	(0.025)
70	-1.247	-1.169	-0.672	-0.497
		(0.016)	(0.021)	(0.027)
80	-1.170	-1.077	-0.632	-0.445
		(0.010)	(0.022)	(0.029)
90	-0.788	-0.929	-0.558	-0.372
		(0.024)	(0.025)	(0.034)
Observations		46008		

Table 2.11: Wage Gap Decomposition (Females)

Quantiles	Wage gap (raw)	Wage gap (predicted)	Endowment effect	Coefficient effect
10	-1.153	-1.278	-0.668	-0.610
		(0.032)	(0.028)	(0.043)
20	-1.440	-1.431	-0.713	-0.718
		(0.027)	(0.026)	(0.0381)
30	-1.579	-1.520	-0.749	-0.771
		(0.022)	(0.026)	(0.035)
40	-1.573	-1.549	-0.786	-0.763
		(0.019)	(0.026)	(0.033)
50	-1.661	-1.543	-0.814	-0.728
		(0.019)	(0.026)	(0.033)
60	-1.545	-1.497	-0.813	-0.684
		(0.019)	(0.026)	(0.0317)
70	-1.453	-1.413	-0.775	-0.638
		(0.021)	(0.026)	(0.033)
80	-1.342	-1.277	-0.688	-0.589
		(0.023)	(0.026)	(0.034)
90	-0.823	-1.049	-0.551	-0.497
		(0.028)	(0.028)	(0.040)
Observations		9442		

Table 2.12: Multinomial Logit Estimates of the Sectoral Allocation Equations (Males)

	Coefficient Estimates			Marginal effects			
	Informal	Self employed	Not in labor force	Formal	Informal	Self employed	Not in labor force
Education							
Years of Schooling	-0.251***	-0.214***	-0.000	0.0111	-0.0232	-0.0046	0.0166
	(0.003)	(0.003)	(0.005)				
Technical education	-0.123**	-0.567***	-0.585***	0.0197	0.0928	-0.0914	-0.0211
	(0.045)	(0.047)	(0.060)				
Voc. training: formal	-0.334***	-0.419***	-0.398***	0.0237	0.0066	-0.0277	-0.0026
	(0.047)	(0.048)	(0.060)				
Voc. training: informal	0.275***	0.631***	-2.100***	-0.0149	-0.0207	0.1336	-0.0980
	(0.044)	(0.043)	(0.077)				
Age	-0.135***	-0.099***	-0.967***	0.0104	0.0224	0.0325	-0.0653
	(0.011)	(0.011)	(0.013)				
Age^2	0.001***	0.001***	0.011***	-0.0001	-0.0004	-0.0004	0.0008
	(0.000)	(0.000)	(0.000)				
Dependents							
No. of Children (<15 years)	0.072***	0.142***	0.065***	-0.0051	-0.0107	0.0183	-0.0025
	(0.011)	(0.010)	(0.014)				
No. of elders (>60 years)	0.022	0.178***	0.283***	-0.0056	-0.0386	0.0286	0.0156
	(0.023)	(0.022)	(0.029)				
Caste							
Scheduled tribe	-0.516***	-1.136***	-0.609***	0.0526	0.0871	-0.1491	0.0094
	(0.048)	(0.049)	(0.064)				
Scheduled caste	-0.043	-0.713***	-0.380***	0.0182	0.1292	-0.1410	-0.0063
	(0.040)	(0.041)	(0.052)				
Other backward class	0.089**	0.048	-0.014	-0.0031	0.0138	-0.0046	-0.0061
	(0.031)	(0.030)	(0.039)				
Marital status							
Currently married	-0.215***	-0.025	-1.657***	-0.0150	-0.0301	-0.0946	0.1398
	(0.050)	(0.050)	(0.059)				
Widowed	-0.178	-0.183	0.424**	-0.0196	-0.157	-0.1716	0.3419
	(0.135)	(0.135)	(0.157)				
Divorced/separated	-0.050	-0.091	0.484	-0.0230	-0.1352	-0.1690	0.3272
	(0.231)	(0.234)	(0.282)				
Regional Variables							
Urban	0.138***	0.082**	-0.159***	-0.0089	0.0272	0.0077	-0.0260
	(0.026)	(0.025)	(0.032)				
Regional dummies	Yes	Yes	Yes				
Religion dummies	Yes	Yes	Yes				
Constant	8.512***	6.515***	19.582***				
	(0.207)	(0.207)	(0.224)				
χ^2	110089.51	8727.92	14897.18				
Log likelihood		-93745.166					
N				109219			

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

Notes : The reference category is formal employment.

Omitted category for Vocational training is 'No vocational training',

for Caste is 'others' (that includes higher castes) and for Marital status is 'never married'.

Coefficients on religion dummies are not reported due to space constraint.

χ^2 is the test statistic for the null hypothesis that coefficients are jointly equal to zero.

Table 2.13: Multinomial Logit Estimates of the Sectoral Allocation Equations (Females)

	Coefficient estimates			Marginal Effects			
	Informal	Self employed	Not in labor force	Formal	Informal	Self employed	Not in labor force
Education							
Years of Schooling	-0.319***	-0.381***	-0.327***	0.0013	0.0004	-0.0019	0.0001
	(0.007)	(0.007)	(0.007)				
Technical education	-0.034	-0.770***	-1.085***	0.0064	0.0653	0.0088	-0.0805
	(0.087)	(0.116)	(0.075)				
Voc. training: formal	-0.140	0.424***	-0.978***	0.0049	0.0385	0.0830	-0.1265
	(0.084)	(0.093)	(0.074)				
Voc. training: informal	0.419***	1.629***	-0.789***	0.0022	0.0499	0.2144	-0.2664
	(0.120)	(0.117)	(0.114)				
Age	-0.027	-0.128***	-0.438***	0.0016	0.0157	0.0098	-0.0270
	(0.021)	(0.021)	(0.019)				
Age^2	-0.001*	0.001**	0.005***	0.0000	-0.0002	-0.0001	0.0003
	(0.000)	(0.000)	(0.000)				
Dependents							
No. of Children (<15 years)	0.021	0.043	0.071**	-0.0002	-0.0019	-0.0009	0.0030
	(0.023)	(0.024)	(0.021)				
No. of elders (>60 years)	-0.088*	-0.003	0.005	0.0000	-0.0037	-0.0001	0.0038
	(0.043)	(0.044)	(0.037)				
Caste							
Scheduled tribe	-0.362***	-0.840***	-0.991***	0.0052	0.0290	0.0035	-0.0377
	(0.098)	(0.102)	(0.088)				
Scheduled caste	-0.079	-0.551***	-0.639***	0.0029	0.0263	0.0018	-0.0309
	(0.086)	(0.092)	(0.080)				
Other backward class	0.045	0.234***	0.022	-0.0002	0.0005	0.0073	-0.0076
	(0.066)	(0.067)	(0.059)				
Marital status							
Currently married	-0.653***	-0.389***	0.628***	-0.0020	-0.0645	-0.0393	0.1060
	(0.095)	(0.100)	(0.087)				
Widowed	-0.784***	-1.106***	-1.632***	0.0134	0.0451	0.0189	-0.0774
	(0.123)	(0.130)	(0.114)				
Divorced/separated	-0.119	-0.585**	-1.099***	0.0066	0.0583	0.0183	-0.0832
	(0.208)	(0.224)	(0.203)				
Regional Variables							
Urban	0.599***	0.407***	0.174***	-0.0008	0.0185	0.0086	-0.0263
	(0.054)	(0.055)	(0.049)				
Regional dummies	Yes	Yes	Yes				
Religion dummies	Yes	Yes	Yes				
Constant	7.166***	8.644***	15.978***				
	(0.373)	(0.379)	(0.347)				
χ^2	2904.47	3777.8	5075.33				
Log likelihood		-51015.008					
N				118620			

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Notes : The reference category is formal employment.
Omitted category for Vocational training is 'No vocational training',
for Caste is 'others' (that includes higher castes) and for Marital status is 'never married'.
Coefficients on religion dummies are not reported due to space constraint.
χ^2 is the test statistic for the null hypothesis that coefficients are jointly equal to zero.

Table 2.14: Selection Corrected Wage Estimates for Formal and Informal Workers (Males)

	(1)	(2)	(3)	(4)	(5)	(6)
		Formal			**Informal**	
	OLS	Selection (Lee)	Selection (Hay)	OLS	Selection (Lee)	Selection (Hay)
Education						
Years of Schooling	0.054***	0.046***	0.062***	0.050***	0.046***	0.044***
	(0.002)	(0.006)	(0.004)	(0.001)	(0.001)	(0.001)
Technical education	0.207***	0.191***	0.217***	0.313***	0.321***	0.325***
	(0.017)	(0.020)	(0.018)	(0.021)	(0.021)	(0.021)
Vocational training: formal	0.038*	0.022	0.050**	0.042*	0.042*	0.042**
	(0.017)	(0.019)	(0.019)	(0.017)	(0.017)	(0.016)
Vocational training: informal	0.003	0.017	-0.010	-0.031***	-0.028***	-0.025**
	(0.021)	(0.024)	(0.022)	(0.009)	(0.009)	(0.009)
Age	0.042***	0.033***	0.051***	0.025***	0.032***	0.036***
	(0.005)	(0.008)	(0.007)	(0.002)	(0.003)	(0.003)
Age^2	-0.000***	-0.000*	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Caste						
Scheduled tribe	-0.020	-0.052	0.006	-0.023	-0.009	-0.005
	(0.019)	(0.027)	(0.024)	(0.013)	(0.013)	(0.013)
Scheduled caste	-0.094***	-0.108***	-0.082***	-0.110***	-0.089***	-0.083***
	(0.017)	(0.019)	(0.018)	(0.010)	(0.011)	(0.011)
Other backward class	-0.069***	-0.066***	-0.071***	-0.095***	-0.093***	-0.092***
	(0.013)	(0.013)	(0.014)	(0.009)	(0.009)	(0.009)
Marital Status						
Currently married	0.061*	0.045	0.077**	0.140***	0.149***	0.152***
	(0.025)	(0.027)	(0.027)	(0.010)	(0.010)	(0.010)
Widowed	0.060	0.060	0.063	-0.046	-0.035	-0.032
	(0.063)	(0.064)	(0.065)	(0.035)	(0.033)	(0.035)
Divorced/separated	0.147	0.140	0.156	-0.089	-0.080	-0.078
	(0.112)	(0.117)	(0.116)	(0.049)	(0.049)	(0.049)
Regional Variables						
Urban	0.139***	0.140***	0.137***	0.096***	0.102***	0.104***
	(0.011)	(0.011)	(0.010)	(0.007)	(0.007)	(0.007)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes
Religion dummies	Yes	Yes	Yes	Yes	Yes	Yes
Selection Correction terms						
λ_{Formal}		0.097				
		(0.060)				
$\lambda_{Formal_{Hay}}$			-0.099			
			(0.056)			
$\lambda_{Informal}$					-0.107***	
					(0.019)	
$\lambda_{Informal_{Hay}}$						-0.231***
						(0.026)
Constant	6.115***	6.627***	5.699***	6.106***	5.934***	5.881***
	(0.109)	(0.328)	(0.250)	(0.041)	(0.052)	(0.048)
N	11087	11087	11087	34921	34921	34921
R^2	0.278	0.278	0.278	0.326	0.326	0.327

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bootstrap standard errors in parenthesis.
Notes : Omitted category for Vocational training is 'No vocational training',
for Caste is 'others' (that includes higher castes) and for Marital status is 'never married'.
Coefficients of religion dummies are not reported due to space constraint.

Table 2.15: Selection Corrected Wage Estimates for Formal and Informal Workers (Females)

	(1)	(2)	(3)	(4)	(5)	(6)
		Formal			**Informal**	
	OLS	Selection (Lee)	Selection (Hay)	OLS	Selection (Lee)	Selection (Hay)
Education						
Years of Schooling	0.081***	0.033	0.036	0.067***	0.066***	0.067***
	(0.004)	(0.025)	(0.021)	(0.002)	(0.002)	(0.002)
Technical education	0.205***	0.064	0.100	0.348***	0.567***	0.528***
	(0.037)	(0.076)	(0.061)	(0.040)	(0.054)	(0.052)
Vocational training: formal	-0.048	-0.159*	-0.140*	-0.070	0.083	0.061
	(0.040)	(0.073)	(0.060)	(0.038)	(0.049)	(0.046)
Vocational training: informal	-0.131	-0.159*	-0.161*	-0.041	0.136**	0.113*
	(0.071)	(0.073)	(0.071)	(0.031)	(0.043)	(0.044)
Age	0.028*	-0.026	-0.021	0.029***	0.119***	0.112***
	(0.012)	(0.029)	(0.025)	(0.007)	(0.018)	(0.017)
Age^2	-0.000	0.001	0.000	-0.000*	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Caste						
Scheduled tribe	0.059	-0.074	-0.057	0.130***	0.270***	0.252***
	(0.051)	(0.084)	(0.075)	(0.033)	(0.041)	(0.040)
Scheduled caste	-0.127*	-0.206**	-0.194**	-0.058*	0.074	0.059
	(0.052)	(0.064)	(0.059)	(0.027)	(0.038)	(0.035)
Other backward class	-0.159***	-0.152***	-0.154***	-0.103***	-0.103***	-0.103***
	(0.038)	(0.038)	(0.038)	(0.025)	(0.025)	(0.025)
Marital status						
Currently married	-0.008	0.057	0.054	0.090**	-0.213***	-0.185**
	(0.047)	(0.053)	(0.053)	(0.029)	(0.062)	(0.059)
Widowed	0.091	-0.107	-0.062	0.068	0.231***	0.179***
	(0.059)	(0.118)	(0.093)	(0.038)	(0.052)	(0.043)
Divorced/separated	0.003	-0.097	-0.076	-0.032	0.153*	0.090
	(0.105)	(0.119)	(0.114)	(0.053)	(0.065)	(0.057)
Regional Variables						
Urban	0.341***	0.377***	0.371***	0.149***	0.228***	0.238***
	(0.031)	(0.034)	(0.034)	(0.018)	(0.027)	(0.024)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes
Religion dummies	Yes	Yes	Yes	Yes	Yes	Yes
Selection correction terms						
λ_{Formal}		0.382				
		(0.196)				
$\lambda_{Formal_{Hay}}$			0.331*			
			(0.155)			
$\lambda_{Informal}$					-0.603***	
					(0.115)	
$\lambda_{Informal_{Hay}}$						-0.555***
						(0.105)
Constant	5.810***	8.397***	7.916***	5.428***	2.947***	3.490***
	(0.228)	(1.321)	(0.984)	(0.108)	(0.484)	(0.384)
N	2377	2377	2377	7065	7065	7065
R^2	0.353	0.354	0.354		0.336	0.340

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bootstrap standard errors in parenthesis.
Notes : Omitted category for Vocational training is 'No vocational training',
for Caste is 'others' (that includes higher castes) and for Marital status is 'never married'.
Coefficients of religion dummies are not reported due to space constraint.

Table 2.16: Oaxaca-Blinder Decomposition

(Selection Corrected)

	Males	Females
Overall wage gap	-1.1449	-1.337
	(0.008)	(0.022)
Predicted wage gap	-1.3699	-3.087
	(0.0901)	(0.436)
Coefficient effect	-0.906	-2.557
	(0.091)	(0.433)
Endowment effect	-0.464	-0.530
	(0.006)	(0.016)
Observations	46008	9442

Figures

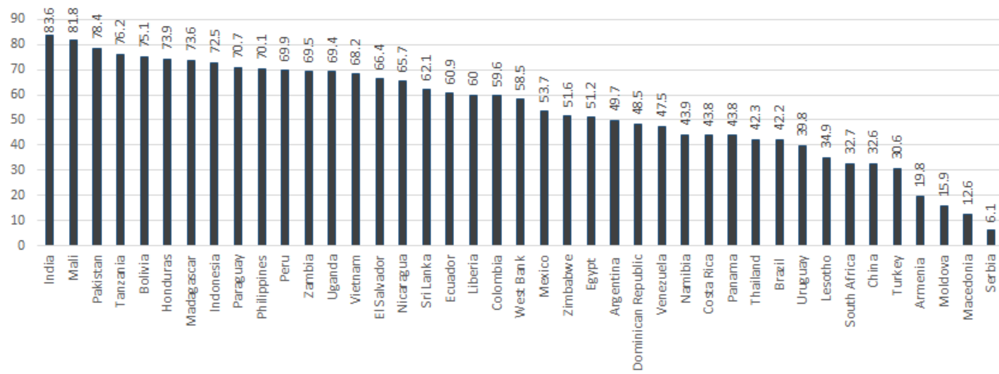


Figure 2.1: Share of Informal Employment in the Non-Agricultural Sector, by Countries
Source: ILO (2013)

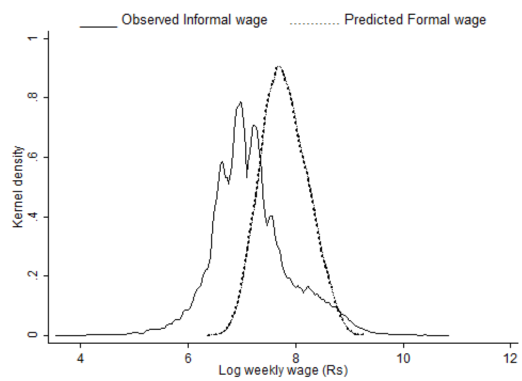


Figure 2.2: Predicted Formal Wage for Informal Workers (Males)

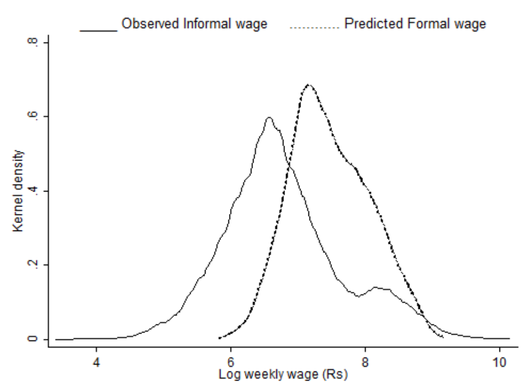


Figure 2.3: Predicted Formal Wage for Informal Workers (Females)

Appendix 1

Sample selection

The survey covered 101724 households and 456999 individuals. In this study I focus on the principal activity status recorded for each person based on their activity in the last 365 days. I deleted those individuals from the sample who were not able to work due to disability and children zero to six years of age. Persons with disabilities and children are identified by the codes 95 and 99 under ‘principal activity status code’. 41981 observations are dropped.

I restrict my sample to the working age population of 15 to 59 years. This is primarily done to avoid critical issues like child labor. Moreover, empirical literature on labor markets tends to focus on the working age population. This restriction brings down the sample to 294,044 observations.

The International Labor organization provides the definition for informal sector ad informal employment for the non-agricultural sector. The classification of individuals into agricultural and non-agricultural is done on the basis of 2-digit National Industrial Classification (NIC 2008) code that is recorded for each person. NIC codes 1, 2 and 3 are classified as agricultural sector and the rest non-agricultural sector Since the NIC code is not recorded for unemployed persons and persons who are not in the labor force, they are retained in the sample and treated as a part of the non-agricultural sector. 47,499 observations are deleted in the process.

Information on employee benefits (social security, paid leave and job contract) is recorded only for the regular workers and casual workers in public and other types of works. Information on benefits is missing for 170 workers. I decided to delete these workers from the sample because they cannot be classified as formal or informal workers. The type of enterprise which is the main identifier for formal and informal sector is missing for 363 workers. I classified these workers into formal and informal workers in the formal and informal sector respectively based on the benefits information. This classification is innocuous because I eventually sum up the formal and informal employment regardless of which sector they belong to. 206 workers out of 363 workers are assigned formal-informal worker status by this process. I delete the rest 157 workers who cannot be classified because neither benefits information nor the information on the type of enterprise is available.

In the NSSO survey, wages are collected only for the regular salaried workers and the casual workers. Wages for self employed workers are not recorded because it is difficult to separate the profit and wage components from the earnings of the self employed workers. Wages for the regular salaried and casual workers are recorded for each activity performed during the reference week. A person may engage in more than one activity in a week. The primary activity is identified by the activity serial number '1' in the data. But wages are recorded against the weekly activity status the reference period of which is the last 7 days. So weekly activity status may not match up with the principal activity status (reference period is 365 days) recorded for some workers. It turns out that 12,678 observations do not report the same weekly and principal activity status. I delete these observations and keep only those observations that report the same weekly and principal activity status. This restriction is necessary to avoid any discrepancy in the classification of formal and informal workers and the wages reported for these workers. Moreover, this restriction allows me to focus on the long term informal employment because principal activity status is based on the worker's activity in the last 365 days. I drop two observations that reported Rs 703000 and Rs 125000 as weekly wages because they were distinct outliers in the sample. I dropped the Zoroastrians (7 observations) from the sample because they were causing perfect predictability problems in the multinomial logit estimation. Further I drop 5384 observations that have no information on the relevant variables: age, sex, religion, social group, marital status, general education, technical education and vocational training. The final sample has 227,839 observations out of which 109,219 are males and 118,620 are females.

Chapter 3

The Spillover Effects of Public Investment: Implications for Formal and Informal Sector Firms in India[33]

3.1 Introduction

Informal production is a pervasive feature of most developing countries. As such, this sector consists of small, unregistered firms that typically produce very labor intensive non-traded goods and services, with little or no access to capital markets, and very limited outward labor mobility to the formal or organized sector (LaPorta and Shleifer (2014)). However, this sector plays an important role in the structural evolution of these countries, accounting for about 42 percent of GDP, and absorbing between $48 - 54$ percent of the labor force (Schneider et al. (2010)). Given capital and labor market rigidities, informal firms may have to rely heavily on government-provided investment goods such as transportation, power, water, etc. for production purposes. However, very little, if anything, is known about the benefits of government investment (and the resulting stock of public capital) for informal production in developing countries. In this paper, we use two large firm-level datasets on formal and informal production in the manufacturing sector in India to examine the sectoral consequences of government investment in public infrastructure.

Despite being a high-growth emerging market, the Indian economy is largely informal, with this sector contributing to 55 percent of GDP and employing about 84 percent of the non-agricultural labor force in 2010 (ILO (2013)).[34] Figures 3.1 and 3.2 depict the average firm-level capital intensity

[33] Co-authored with Santanu Chatterjee. Department of Economics, University of Georgia, Athens, GA 30602, USA.

[34] Mehrotra et al. (2013) document that between 2004-2012, a period of relatively high economic growth for India, the share of informal employment in the manufacturing sector was very large and persistent, at around 89 percent.

and output-labor ratio for cross-sections of manufacturing firms in the formal and informal sectors for 1999 and 2010, respectively. For example, in 2010 the capital intensity of formal sector firms exceeded that of informal firms by a factor of 5, while output per worker was higher by a factor of about 10. Interestingly, however, these gaps were smaller in 2010 than they were in 1999, suggesting that during this period, informal sector firms have indeed been able to improve both their relative usage of capital as well as labor productivity. This point is further underscored in Figure 3.3, which shows that the output share of the informal sector, though quite substantial, has been on a downward trend, declining from about 60 percent of GDP in 1999 to 55 percent in 2010.

One factor that may affect the productivity of both formal and informal sector firms is the government's provision of public goods such as infrastructure, that may serve as an input in their production processes. Essentially, public spending on roads, power, water, sanitation, communications, healthcare, and education may have complementary spillovers for private factors of production in both sectors. As such, public investment may help alleviate the credit and labor market constraints that firms typically face, especially in the informal sector. Indeed, infrastructure investment has been a centre-piece of public policy in India over the past two decades or so.[35] As shown in Figure 3.4, the share of total infrastructure spending in GDP increased from 5.8 percent in 2006 to about 8.4 percent in 2011, with more than 70 percent of this spending coming from the public sector. Further, this share is expected to rise to about 11 percent of GDP by 2017.[36] A critical consideration here is the effects of the rising share of infrastructure spending in India on the productivity of formal and informal sector firms. Given the relative magnitudes of India's public investment and the share of the informal sector, their underlying relationship (if any) is of critical importance for the design and implementation of public policy.

In this paper, we attempt to bridge a gap between two strands of research that have evolved largely independently of each other. On the one hand, starting with the work of Aschauer (1989b), a voluminous empirical literature has explored the productivity benefits of public investment in

Informal employment is a job-based concept, comprising of workers who lack access to basic legal protection, social security, and employment benefits ILO (2013).

[35]See, for example, two recent reports by the McKinsey (2013) and the Institute and Ernst&Young (2013) on trends in public infrastructure spending in emerging markets like India.

[36]Source: Planning Commission of India.

infrastructure, with a rich diversity of results.[37] However, these studies have, without exception, considered the consequences for either industrialized countries (where the share of informal production is relatively small), or only for the formal sector in developing countries. On the other hand, the literature on the informal sector has mainly focused on issues of measurement of its output share (Schneider and Enste (2000), LaPorta and Shleifer (2014, 2008), and Gomis-Porqueras et al. (2014)), or issues pertaining to tax policy and enforcement (Rauch (1991), Ihrig and Moe (2004), Turnovsky and Basher (2009), Prado (2011), and Ordonez (2014)). The importance of public investment for this type of production has generally been ignored. Consequently, by examining the benefits of government investment expenditures for private production in the formal and informal sectors, we seek to fill an important gap in this literature. This is the first contribution of our paper. Second, while most studies on public investment are conducted at a fairly aggregated level (at the level of a country, state or region), we attempt to estimate its sectoral productivity benefits at the level of the individual firm. In the case of India, for example, while Binswanger et al. (1993), Lall (1999), Mitra et al. (2002), Zhang (2004), and Hulten et al. (2006), among others, have examined the effects of public infrastructure for the formal sector at the state, district, or industry level, there is no current evidence of its sectoral importance at the level of the firm. The firm-level datasets we use for our study enable us to shed light on the role of public investment and infrastructure at a much more disaggregated level than previously studied. We view this as an additional contribution to the literature.[38] Finally, from the perspective of designing public policy, it is important to know how the spillovers from public investment are dispersed over the size distribution of firms in each sector. In other words, do larger firms tend to benefit more or less relative to their smaller counterparts from government spending on public goods? This may help determine how public goods should be targeted to firms in each sector. To the best of our knowledge, our analysis is the first to shed light on this issue.

[37]See, for example, Munnell and Cook (1990), Lynde and Richmond (1992), Glomm and Ravikumar (1997a), Gramlich (1994), and Holtz-Eakin and Schwartz (1995), and Devarajan et al. (1996) for some early contributions. Bom and Ligthart (2014) provide an excellent survey and meta-analysis of the recent empirical literature.

[38]Two recent studies, namely Datta (2012) and Ghani et al. (2015) examine the spatial role of India's recent expansion of its interstate system on plant-level production. These studies, however, do not distinguish between formal and informal production at the firm level.

In India, the main source of information at the firm level for the formal sector is the Annual Survey of Industries (ASI), while for the informal sector it is the surveys conducted by the National Sample Survey Organization (NSSO). Though the ASI surveys firms on an annual basis, the NSSO survey is conducted once every 10 years. We use data from the 2010 round for each of these surveys, since that is the latest round for which firm-level information is available for *both* sectors. Restricting our coverage to only the manufacturing sector, we obtain a cross-section of $32,388$ formal-sector firms (from the ASI) and $82,748$ informal-sector firms (from the NSSO) for 2010. We proxy public investment by state-level data on government *Development Expenditures*, obtained from the Reserve Bank of India.[39] Here, we consider two sub-categories of expenditures: (i) *Economic Services*, which include public expenditures on transport, communications, and energy, and (ii) *Social Services*, which include expenditures on health, education, water, sanitation, and other welfare programs, and construct measures of both the *flow* of public investment, using average annual expenditures over the $2006-2010$ period, as well as its accumulated *stock* at the per-capita level for each state, using data over the period $2000-2010$. The flow measure is intended to capture the short-term effects of public investment, while the stock measure captures its effects over the longer term. Henceforth, we will interchangeably refer to the broad category of Development Expenditures as *public investment*, and the corresponding stock measure as *public capital*.

Our empirical strategy can be described as follows. First, we estimate the output elasticities of the flow of public investment and the accumulated stock of public capital at the firm level in the formal and informal sector. While this gives us information on how the average firm in each sector is affected by public investment, it masks the distribution of the sectoral elasticities across firms. We therefore employ quantile regressions (QR) to examine how the sectoral output elasticities vary across the size distribution of firms. Further, we also examine the relationship between public investment and the capital intensity in production across this size distribution. The QR approach is of critical importance from the policy perspective, since public investment may serve as a potential mechanism through which the government may aim to reduce not only the size of the informal sector, but also increase the relative usage of private capital in that sector.

[39]This is the highest level of disaggregation at which public expenditure data is available, especially for infrastructure goods.

The empirical analysis also raises some important econometric issues. First, it is plausible that the inclusion of public investment generates a reverse causality problem with output. Using firm-level data along with state-level government expenditures helps alleviate this problem, as it is unlikely that an individual firm will have any systematic effect on public spending at the state level. Second, the usage of private inputs like capital and labor may be endogenous to the firm's decision to produce output. Here, we use a method suggested by Levinsohn and Petrin (2003) and Sivadasan (2009) that uses past values of intermediate inputs and exploits the repeated cross-sectional nature of our dataset to control for the unobserved productivity shock at the firm-level in each sector.

Our results indicate that though public investment has a positive and statistically significant effect on both formal and informal sector firms, there are important sectoral and distributional consequences. With the flow specification of public investment, we estimate an output elasticity of 0.088 for formal sector firms. The corresponding output elasticity for the average informal sector firm is about 3 times lower, at 0.027. When we consider the stock of public capital, the difference in sectoral output elasticities is much larger, with the estimate for formal firms at about 0.17, about 7 times larger relative to their informal counterparts. Since the stock measure of public investment is intended to capture its long term productivity spillovers, these results suggest that the benefits accruing to formal sector firms from the accumulated stock of public capital are much larger relative to those for informal sector firms. Within the sub-categories of public investment, we find that Economic Services is associated with higher productivity spillovers relative to Social Services in both sectors, irrespective of whether we use the stock or flow specification.

While these results only represent the effects associated with the *average* firm in each sector, it is important to consider how the estimated output elasticity of public investment varies across the size distribution of firms (measured by their gross value added). For the formal sector, we find that there is very little variation in the output elasticity of public investment across the size distribution of firms. By contrast, for the informal sector this association is strictly positive across the entire size distribution of firms. Another important issue is whether public investment influences the relative capital intensity of firms in each sector. This is especially relevant for firms in the informal sector, who tend to have extremely low capital-labor ratios. Here, we find that while public investment generally raises the capital intensity of informal sector firms, the effects are relatively stronger for the top 20 percent of firms, suggesting that the complementarities between public investment and capital intensity are the highest for the largest firms in the informal sector. This has important implications for public policy: rather than a one-size-fits-all approach, more public investment goods might be targeted for the largest informal sector firms. By contrast, public investment is negatively associated with the capital intensity of formal sector firms, suggesting that it may be a substitute for private factors in this sector.

The rest of the paper is organized as follows. Section 3.2 discusses the data and summary statistics, while Section 3.3 describes the empirical specification and our strategy to address the issue of endogeneity. Section 3.4 reports the results of the empirical analysis, and Section 3.5 concludes.

3.2 Data

We use firm-level data from two sources, namely the (i) Annual Survey of Industries (ASI), and (ii) National Sample Survey Organization (NSSO). The ASI covers formal sector firms registered under Sections 2(m)(i)-(ii) of India's Factories Act of 1948, and reports annual data on firm-level receipts, expenses, operational (firm-specific) characteristics. The data set is a repeated cross-section, where the sampling of firms changes in every round of the survey. The NSSO's "Survey of Unincorporated Non-Agricultural Enterprises" is the predominant source of firm-level information for the informal sector in India. The survey is conducted every ten years, and provides firm-level information on the ownership category, location, and other operational characteristics. Specifically, the NSSO survey includes household proprietary and partnership enterprises that are not registered under the Factories Act of 1948 or the Bidi and Cigar Workers (Condition of Employment) Act of 1966. Public sector enterprises and cooperatives are excluded from the survey. Since the ASI reports data on an annual frequency, while the NSSO does so on a ten-year frequency, we use the cross-sections from both surveys for 2010, which the latest available survey round for the NSSO, in order to maintain compatibility between the two sectors.

The ASI survey covers $52,243$ formal sector firms in 2010. The coverage is skewed heavily towards manufacturing firms: 93.7 percent of the firms surveyed were engaged in manufacturing. The 2010 NSSO survey of the informal sector covers $334,474$ firms. Of these, only 30 percent are in the manufacturing sector, with trading activities (36 percent) and services (34 percent) making up the rest. To ensure that the sample of formal and informal sector firms are comparable, we restrict the coverage to only manufacturing firms in both sectors. This gives us a sample of $32,388$ formal-sector firms and $82,748$ informal-sector firms in 2010.

Output for both the formal and informal sector firms is measured by the gross value added (GVA; the value of total output net of total inputs). Private capital is given by the closing balance of gross fixed capital (owned and rented) at the end of the accounting year, and labor is measured by the average number of workers employed during the accounting year. An important consideration for our empirical strategy is the value of intermediate inputs. For the formal sector, we use the value of electricity consumed at the firm level as the proxy for an intermediate input. For informal sectors firms, the value of electricity usage has many missing values, as many informal sector firms do not report electricity consumed. Therefore, we use the value of total operating expenses for the firm, which includes the combined cost of fuel, electricity, repairs, and maintenance.[40] All monetary values are expressed in terms of $2004-2005$ Indian Rupees.

Data on public investment have been collected from the State Finances Database of the Reserve Bank of India. We use state-level data on public expenditures (payments for accumulation of assets financed by borrowed funds) for two categories: (i) Economic Services, which include expenditures on transport, communications, and energy, and (ii) Social Services, which include expenditures on health, education, water and sanitation, and other welfare programs. The sum of these two categories is defined as Total Development Expenditures, and serves as a proxy for state-level public investment in our analysis. We scale each category of public expenditure by the population in each state, to obtain per-capita measures of government spending by state. To estimate the output elasticity of public investment for a firm's production function in 2010, we use average annual per-capita public expenditures at the state level for the past five years, i.e., for the period 2006-2010, to factor out any annual idiosyncratic changes to the level of public spending. This gives us an average *flow* measure for public investment.

[40] This could be due to informal sector firms using non-authorized or illegal sources of electricity, such as "borrowing" from a neighbor's or public power line. Reporting an aggregated number for operating expenses makes it difficult to distinguish different types of energy consumption. These costs are reported for the past-30 day reference period, which are then converted to an annual figure.

In addition to the flow measure, we also construct a *stock* measure for public capital using the perpetual inventory method. Specifically, we use the year 2000 to pin down the initial stock of public capital, since some Indian states before 2000 were part of bigger states. The initial level of public capital stock is measured by

$$K_{G,0} = \frac{G_{I,0}}{g + \delta_G} \tag{3.1}$$

where $G_{I,0}$ is the flow of public investment in the initial period, g is the growth rate of public investment, and δ_G is the depreciation rate for public capital. We follow Gupta et al. (2014) and set the annual depreciation rate to 2.5 percent. The stock of public capital at the end of the time period is given by the following accumulation equation

$$K_{G,t} = K_{G,0} + \sum_{t=1}^{T} (1 - \delta_G)^t \, G_{I,t} \tag{3.2}$$

We compute the stock measure of public capital in 2010 by using the public expenditure flows for each year during 2000-2010 (measured at 2004-2005 prices), using the average growth rate of public investment across the sample as a lower bound to measure the initial stock. The total stock measure is then divided by the state-level population to obtain a per-capita estimate by state.

Finally, we use several other state-level controls such as state GDP (Net State Domestic Product or NSDP), total labor force, literacy rate, dependency ratio, crime rate, and total number of enterprises. The data sources for these variables are provided in Appendix 2.

3.2.1 Summary Statistics

Table 3.1 presents the summary statistics for firm-level characteristics for the formal and informal sectors, respectively, for 2010. We see that informal firms were much smaller in size (as measured by their GVA), with average capital-labor and output-labor ratios being significantly smaller than their formal-sector counterparts. For example, capital intensity (measured by the capital-labor ratio) in production was about 5 times higher for formal firms, while output per worker was higher by a factor of about 10. About 60 percent of formal sector firms were situated in urban areas, with a large majority being privately owned. About 50 percent of informal sector firms were in urban areas, with only 20 percent being registered with some government-level authority. About 70 percent of these firms were male-owned proprietary businesses.

Table 3.2 lists the average state-wise public development expenditures, along with its two sub-categories (Social and Economic Services) (i) as a share of state GDP (Net State Domestic Product-NSDP), and (ii) in per-capita terms, for the period 2006-2010. On average, Indian states spent about 4.9 percent of state GDP on development expenditures, with about 69 percent being allocated to expenditures on Economic Services (transport, communications, and energy). There is significant variation in public expenditures on development across Indian States: while the north-eastern state of Manipur spends the most, with about 13 percent of state GDP allocated to public investment, the southern state of Kerala spends the least, at about 1.3 percent. This comparison is also consistent for the per-capita measure of government expenditure. The average per-capita level of development expenditures across states between 2006-2010 was about Rs. $1,611$ (approximately \$24 in current prices), with Economic Services again accounting for about 69 percent of per-capita development spending. Data on the average stock of pubic capital in 2010 across Indian States is presented in Table 3.3. On average, the stock of public capital represented about 37 percent of state GDP, with the stock of public capital coming from the Economic Services sub-category being about 26 percent of state GDP.[41]

[41]See Figures A.2.7-A.2.8 for heat maps generated using the state-level data.

3.3 Empirical Specification

The main objective of our empirical analysis is to estimate the output elasticity of public investment for the formal and informal sectors. To do this, we estimate a Cobb-Douglas production function without any *a priori* restriction on the returns to scale in production:

$$Y_{ist} = A_{ist} L_{ist}^{\alpha} K_{ist}^{\beta} \tag{3.3}$$

where the subscripts i refer to the firm, s to the state where the firm is located, and t denotes the time period.[42] Y_{ist} denotes the flow of output for a firm i in a given sector located in state s at time t. Similarly, L_{ist} is the labor input , private capital is given by K_{ist}, and A_{ist} represents a productivity shock. Assume that productivity at time t for a given firm i located in state s is given by

$$A_{ist} = \varepsilon_{ist} G_{st}^{\gamma} \tag{3.3a}$$

where G_{st} denotes the state-level public investment, and ε_{ist} is an unobserved productivity shock specific to the firm. The specifications in (3.3) and (3.3a) are consistent with the voluminous literature on the link between output and public investment, starting with Aschauer (1989b), and Barro (1990). Taking logs and using firm-level Gross Value-Added (GVA) as a proxy for output, we can write the empirical specification as

$$\ln GVA_{ist} = \alpha \ln L_{ist} + \beta \ln K_{ist} + \gamma \ln G_{st} + \theta X_{ist} + \rho Z_{st} + \varepsilon_{ist} \tag{3.4}$$

In (3.4), output is measured by firm-level Gross Value-Added (GVA), and α, β, and γ are the output elasticities of labor, private capital, and public investment, respectively. Since the unit of observation is the firm, X is a vector of firm-level characteristics that includes age of the firm, type of ownership, industrial category (NIC 2-digit level), and geographical location (rural or urban). We use the same set of characteristics for both formal and informal sector firms, with the addition

[42] Since we use a cross-section data for the year 2010, the time subscript t denotes the year 2010. We cannot drop the time subscript at this point because we are going to refer to some previous period's (for the year 1999) average productivity in this section.

of registration status for informal sector firms. Additionally, we control for state-level variables (Z) to factor out any state-level factors other than public investment that may have an effect on the firm's output. Z includes state GDP (Net State Domestic Product or NSDP), total labor force, literacy rate, dependency ratio, crime rate, and total number of enterprises.

One issue with the production function approach in (3.4) is that it may produce biased estimates of output elasticities if there exists reverse causality between the factors of production (including public investment) and output. Our empirical strategy addresses this concern on two fronts. First, since our unit of analysis is the firm, it is unlikely that an individual firm's output will have any influence on the level of public investment in a given state.[43] Second, there may exist an endogeneity problem with respect to the private inputs in production (labor and private capital). We use a method developed by Levinsohn and Petrin (2003) and Sivadasan (2009) to control for potential reverse causality generated by the private factors of production. We discuss this empirical strategy below.[44]

The source of endogeneity in the specification (3.4) is the unobserved productivity shock that is observed by the firm, but not by the econometrician. This may induce the firm to choose private inputs (capital and labor) endogenously. Hence, the error term that contains the unobserved productivity shock may be correlated with the choice of private inputs. To fix ideas, we start by decomposing the error term (ε) into two components

$$\varepsilon_t = \omega_t + \eta_t \tag{3.5}$$

where ω is the productivity shock observed by the firm but not by the econometrician, and η is the classical error term or the productivity shock unobserved by both the firm and the econometrician.[45] The issue here is that the firm might take the productivity shock ω into account when

[43]However, if the unit of observation had been at the industry or state level, we would not have been able to make this assumption. Another related issue is that the location of firms may reflect self-selection: states that spend more on public investment tend to attract more firms. We present some robustness checks regarding self-selection with respect to young formal sector firms in Appendix 2.

[44]It is important to note here that there are alternative approaches to estimate output elasticities of factors of production. For example, the cost function approach, based on duality theory, estimates a a translog cost function where, in our specific case, public investment would be included as an unpaid factor of production. Direct estimation of this cost function would produce an estimate of the marginal benefit (or cost reduction) from public investment. The elasticity of public investment would then be backed out with the help of duality theory; See, for example, Lynde and Richmond (1992) (1993) and Binswanger et al. (1993).

[45]Equations 3.5-3.12 describe the method proposed by Levinsohn and Petrin (2003) and Sivadasan (2009). We drop the i and s subscript for brevity in these equations and add them back in the end to fit our model.

making input choices. Since this unobserved productivity shock is included in the composite error term, ε, specification (3.4) violates the basic assumption of OLS, i.e., $E\left(\mathbf{x}^{'}\varepsilon_t\right) = 0$, where $\mathbf{x}^{'}$ is a vector for the private inputs K and L. This renders the OLS estimates inconsistent. Levinsohn and Petrin (2003) develop a strategy that uses intermediate inputs to control for the unobserved productivity shock. Specifically, they assume a demand function for intermediate inputs of the form

$$m_t = m\left(\omega_t, K_t\right) \tag{3.6}$$

where $m\left(.\right)$ is the firm's intermediate input demand function, which is assumed to be monotonically increasing in the firm's unobserved productivity (ω). Private capital, being a state variable, determines the optimal choice of intermediate inputs. Levinsohn and Petrin (2003) assume that input and output prices are constant across firms. The monotonicity assumption allows us to invert the input demand function and write the unobserved productivity shock as a function of intermediate input and private capital:

$$\omega_t = \omega\left(m_t, K_t\right) \tag{3.7}$$

Equation (3.7) can be written as a polynomial function

$$\omega_t \equiv \omega\left(m_t, K_t\right) = \sum_{i=0}^{3}\sum_{j=0}^{3-i}\delta_{ij}m_t^jK_t^i \tag{3.8}$$

Combining (3.5) with (3.8), and using in (3.4), we can write

$$\ln GVA_t = \alpha\ln L_t + \gamma\ln G_{st} + \theta X_t + \rho Z_{st} + \phi\left(m_t, K_t\right) + \eta_t \tag{3.9}$$

where, $\phi\left(m_t, K_t\right) = \beta\ln K_t + \omega\left(m_t, K_t\right)$. Note that the coefficient on private capital (β) cannot be identified from equation (3.9) because of the function $\phi(.)$.

In order to identify this coefficient, Levinsohn and Petrin (2003) use a final identification restriction by assuming that the productivity shock ω_t is governed by a first order Markov process:

$$\omega_t = E(\omega_t | \omega_{t-1}) + \xi_t \tag{3.10}$$

where ξ_t is an innovation to productivity that is uncorrelated with K_t. The underlying economic interpretation is that a firm's current productivity can be predicted by its productivity in the previous period. This assumption implies the following moment condition

$$E[\xi_t K_t] = E[K_t \{\omega_t - E(\omega_t | \omega_{t-1})\}] \tag{3.11}$$

The moment condition (3.11), used by Levinsohn and Petrin (2003) to identify the coefficient on capital relies on panel information (estimate of last period's productivity, ω_{t-1}), which is not available for our case, given the limitation of our dataset. However, given that we have repeated cross-section data comprising of different samples of firms drawn for the years 1999 and 2010 (two cross sections where *both* the ASI and NSSO data are available), we can use a modification to (3.11) proposed by Sivadasan (2009). Specifically, with repeated cross-section data we can estimate the average productivity for a particular industry in a particular state for the previous time period, which in our case is the 1999 round for both surveys (for brevity, we refer to this combination as a "cell") and use that estimate in place of ω_{t-1} in (3.11):

$$\bar{\omega}_{t-1} = \frac{1}{s_{ji}} \sum_{q=1}^{s_{ji}} \omega_{q,t-1} \tag{3.12}$$

where s_{ji} is the number of firms in a particular cell. $\omega_{q,t-1}$ is the firm-level estimate of the unobserved productivity shock located in a particular cell in the previous period. Essentially, the identification process involves three steps. First, we estimate the individual firm's unobserved productivity for the year 1999.[46] Second, we find the average productivity within a NIC 3-digit

[46] For the year 1999, the ASI survey yields 19,095 formal sector firms in the manufacturing sector. The corresponding NSSO survey covers 49,720 informal sector firms, also in the manufacturing sector.

industry code and within a particular state. The last step involves matching the average productivity of a particular cell in year 1999 to the firm located in same industry and the same state (i.e., in the same cell) in the year 2010. This process allows us to write (3.10) as

$$\omega_t = E\left(\omega_t|\bar{\omega}_{t-1}\right) + \xi_t \tag{3.13}$$

According to (3.13), firm-level productivity for the current year can be predicted by the previous period's average "cell" productivity. Now, the coefficient on private capital (β) can be identified by the following second step regression

$$v^*_{ist} = \beta K_{ist} + E\left(\omega_{ist}|\bar{\omega}_{t-1}\right) + \eta^*_{ist} \tag{3.14}$$

where, $v^*_{ist} = \ln GVA_{ist} - \alpha \ln L_{ist} - \gamma \ln G_{st} - \theta X_{ist} - \rho Z_{st}$ and $\eta^*_{ist} = \omega_{ist} - E\left(\omega_{ist}|\bar{\omega}_{t-1}\right) + \eta_{ist}$.

3.4 Results

In this section, we report the results from our empirical analysis. Specifically, we start with an OLS estimation of the output elasticity of public investment and its sub-categories (for both the flow and stock measures) for firms in the formal and informal sectors. Given the possibility of biased estimates from the OLS specification, we then re-estimate the sectoral output elasticities using the method proposed by Levinsohn and Petrin (2003) and Sivadasan (2009) (henceforth LP-S), and described in (3.5)-(3.14) above. Next, we use quantile regressions to examine the distributional consequences of public investment for the formal and informal sectors. Here, we (i) estimate the sectoral output elasticities of public investment and (ii) its effects on the capital intensity of production in each sector across the sectoral size distribution of firms, based on their gross value-added (GVA). The results are reported in Tables 3.4-3.9 and Figures 3.5-3.12.[47]

[47] Tables 3.4-3.9 report the bootstrap standard errors in parenthesis.

3.4.1 Formal Sector

We begin our empirical analysis with an OLS estimation of the output elasticity of the private factors of production (capital and labor) and public investment for firms in the formal sector. Table 3.4 reports the results of regressing firm-level GVA on the private and public inputs, along with controls at both the level of the firm and the state. Column (1) reports the results for the aggregated category of public investment, i.e., development expenditures. Columns (2) and (3) report results for its two sub-categories: social services and economic services, respectively. The OLS results suggest output elasticities of labor and private capital for formal sector firms of about 0.78 and 0.33, respectively, reflecting the presence of increasing returns to scale in the private factors of production (note that the empirical specification does not impose any *a priori* restriction on returns to scale in the production function). As for the public input, the aggregated category of development expenditures has an output elasticity of about 0.03, indicating a small, but positive effect of public investment on firm-level output in the formal sector. The sub-category of economic services expenditures has a similar elasticity measure, and social services expenditures are not statistically significant.[48]

As mentioned in the previous section, the OLS estimates reported in Table 3.4 can be biased, due to the endogeneity of private inputs in production. To address this issue, we use the strategy outlined in (5)-(14), developed by Levinsohn and Petrin (2003) and Sivadasan (2009), to obtain more robust estimates of the output elasticities of the private and public inputs. The results from this estimation (labeled LP-S) for formal-sector firms are presented in Table 3.5. As is evident from this table, correcting for the endogeneity of private inputs alters the results significantly. For example, the returns to scale for the private inputs are now much closer to 1, with the output elasticities for labor and capital being about 0.66 and 0.37, respectively. For public investment,

[48]Table A.2.3 in Appendix 2 reports the OLS estimates with standard errors clustered at the state-industry level.

the estimated elasticities are now much larger than those suggested by the OLS estimation. Development expenditures are associated with an elasticity of 0.088, while those for the sub-categories of economic and social services are also higher (and statistically significant) at 0.077 and 0.045, respectively.

Table 3.6 presents the results from a LP-S estimation of the production function, but with government investment measured as a per-capita stock variable, rather than a flow. The estimated elasticities associated with the aggregated and sub-categories of government expenditure turn out to be much larger with the stock specification. For example, the elasticity of development expenditure is now about 0.17, and that for economic services is about 0.16, indicating that the productivity benefits from the accumulated stock of public capital significantly exceed those from the flow of public investment for formal sector firms.

3.4.2 Informal Sector

Tables 3.7 and 3.8 report the estimation results for the output elasticities of private and public inputs for informal sector firms, along with firm and state-level controls. Table 3.7 reports results from the OLS estimation, while Table 3.8 reports the LP-S estimation, correcting for the endogeneity of private inputs. Comparing the OLS results from Tables 3.4 and 3.7, we see that informal sector firms have a significantly higher (lower) output elasticity for labor (private capital) relative to the formal sector. As with the OLS results for the formal sector, the informal sector also exhibits increasing returns to scale in the private inputs. The productivity effect of the public input, however, is not statistically significant, in contrast to its positive impact on formal sector firms. Within the sub-categories of public expenditures, while economic services do not have any systematic effect on private productivity, social services expenditures do have a significant, but negative effect on firm-level productivity.[49]

Given the endogeneity issue with the OLS estimation, we turn our focus to Table 3.8, which corrects for this problem, using the LP-S methodology. As with our results for the formal sector, the results change significantly. First, with respect to the private inputs, we get output elasticities of 0.63 and 0.32 for labor and capital, respectively. Interestingly, in contrast to firms in the formal

[49] Table A.2.4 in Appendix 2 reports the OLS estimates with standard errors clustered at the state-industry level.

sector, returns to scale for the private inputs in informal sector production is less than one. The output elasticity of development expenditures is about 0.027, which is about 3 times lower than the corresponding elasticity of formal sector firms. Similarly, the elasticity with respect to economic services expenditures for informal firms is lower than their formal counterparts by a factor of about 2. On the other hand, while social services expenditures had a positive impact on the productivity of formal firms, its effect on informal firms is negative and statistically significant. The intuition behind this result may have to do with the composition of social services expenditures: health, education, water and sanitation, and welfare programs. States that spend more on this category might have a more educated and healthy workforce, thereby benefiting formal sector firms at the expense of their informal counterparts (by making labor more expensive for the less productive informal sector firms).

Table 3.9 presents the LP-S estimation of the informal sector production function, but with the stock measure of public investment. Qualitatively, the results in Table 3.9 are consistent with the results for the flow specification: while the stock measure for aggregate development expenditures and the sub-category of economic services have positive and significant coefficients, social services expenditures have a negative and significant association with output. However, quantitatively, the difference in magnitudes of these sectoral effects are now much larger. For example, with the stock measure, the output elasticity of informal sector firms is lower than its formal sector counterpart by a factor of about 7, with respect to development expenditures (for the flow measure it was about 3). For the economic services category, the contribution of public investment for the informal firm is more than 4 times lower than for the formal firm (for the flow measure it was about 2). These results suggest that the benefits of the accumulated stock of public capital are much stronger for the formal sector than the informal sector, relative to the benefits from the flow measure of public investment.

3.4.3 Firm Characteristics

The firm-level control variables we use for our empirical specification include a firm's age, location (rural versus urban), and ownership status (government versus private for the formal sector, and proprietary versus partnership for the informal sector). Additionally, we use registration status for informal sector firms. Since all formal sector firms, by definition, are already registered under the Factories Act, the focus for this variable is on the informal sector. Specifically, though informal sector firms are not registered under the Factories Act, they might still be registered with other local government entities like a municipal corporation or village panchayat. Therefore, the registration status for informal sector firms includes any kind of registration *outside* of the Factories Act. In our sample, only about 20 percent of informal sector firms fall under this category.

The discussion of firm-level characteristics draws on the LP-S estimation results from Table 3.5 (formal sector) and Table 3.8 (informal sector). The age of a firm is not a systematic predictor of productivity for formal sector firms. Informal sector firms, however, are adversely affected by age, though the magnitude of the coefficient is very small, and probably not economically meaningful. Both formal and informal sector firms are affected by their location: firms in both sectors that are located in rural areas produce, on average, about 10 percent less output (in terms of GVA) than their urban counterparts. Accessibility to ports, airports, roads, and transportation infrastructure, as well as quality labor and capital could be major factors driving this result. Registration status is an important determinant of the output for informal sector firms: firms that are registered under some authority (outside of the Factories Act), produce, on average, about 16 percent more output relative to firms that are not registered. This suggests that registration might improve access for informal sector firms to credit and final goods markets, thereby improving their GVA. However, the fact that a majority (80 percent) of the informal sector firms in our sample are not registered indicates the existence of other barriers to registration, such as bureaucracy, corruption, and potential benefits from avoiding tax obligations. Finally, privately owned firms in the formal sector produce about 23 percent less output than public sector firms. For the informal sector, proprietary firms owned by the females produce about 43 percent less output compared to corresponding firms owned by males. Partnerships involving members of the same household are less productive relative to those that involve members from different households. Trusts and self-help groups also tend to have lower output within the informal sector.

3.4.4 Distributional Effects

An important consideration in the context of our empirical strategy is that the point estimates of the output elasticity of public investment reported in Tables 3.4-3.9 represent the effects for an *average* firm in each sector. However, it is plausible that the effect of public investment may vary across the size distribution of firms. In other words, do small firms in each sector benefit more or less from public investment, relative to larger firms? Another related issue is the relationship between factor usage, specifically capital intensity in production, and public investment. In other words, are there complementary effects between public and private inputs and, if so, how do they vary across the size distribution of firms in each sector? To understand this better, we employ a quantile regressions (QR) analysis for firms in each sector, by constructing a size distribution of firms based on their GVA, and then estimating the firm-level (i) sectoral output elasticities of public investment and (ii) effects on the capital-labor ratio in each quantile in this distribution. We use the same set of firm and state-level controls as in Sections 3.4.1 and 3.4.2.

3.4.4.1 Sectoral Output Elasticity

Figures 3.5-3.10 graphically illustrate how the quantile elasticities of public investment (measured by per-capita development expenditures at the state level for both the flow and stock measures) vary across the size distribution of firms in the formal and informal sectors, with the shaded areas denoting the confidence intervals.[50] As Figure 3.5 illustrates, for the flow specification, the effect of public investment is more or less even across the firm's size distribution, with elasticities in the range of 0.07-0.09, and with very little variation. For the stock specification, however, the productivity benefits of public investment do not show any trend for the bottom 60 percent of firms, but increase very gradually for firms in the top 40 percent of the size distribution. Figures 3.6 and 3.7 depict the quantile output elasticities for the two sub-categories of public investment, namely Economic and Social Services. Here, we see that while the quantile elasticity for Economic Services increases with firm size, the opposite is true for the Social Services category. When aggregated, Figures 3.6 and 3.7 help us understand why the there is so little variation in the output

[50]The quantitative results, in table format, are available on request from the authors.

elasticity for Development Expenditures across the size distribution of firms, with the quantile elasticities for the two sub-categories moving in roughly opposite directions. It also underscores the fact that formal sector firms derive larger benefits from government spending on Economic Services (transport, communications, energy), relative to Social Services (health, education, water and sanitation. etc.).

By contrast, Figures 3.8-3.10 indicate that for informal sector firms the picture is quite different. Here, the output elasticity of public investment increases persistently with firm size, irrespective of the flow or stock measure for public investment, with the largest firms again benefiting the most from government investment. Even though informal sector firms use less capital per worker than their formal counterparts, larger firms may derive strong complementary effects from public investment. Smaller informal firms, with very little capital and labor, may not be able to appropriate the productivity spillovers generated by public investment expenditures. In contrast to the case of formal sector firms, the quantile elasticities for both Economic and Social Services increase with firm size in the informal sector.[51]

3.4.4.2 Capital Intensity in Production

Since public investment generates productivity spillovers for the firm's production function, and important question is how it affects the firm's relative usage of private capital, i.e., its capital intensity, measured by the capital-labor ratio. The issue at hand is whether public investment can substitute for or complement the usage of private factors. This is especially relevant for the production structure of firms in the informal sector, which are characterized by very low levels capital intensity, as documented by LaPorta and Shleifer (2014). Therefore, low capital-labor ratios in this sector may be an impediment towards "formalization." From a policy perspective, it is natural to ask if public investment expenditures play a role in increasing this ratio for informal sector firms. Figures 3.11-3.12 present some evidence on this question, for both formal and informal sector firms across their size distribution, for both the flow and stock measures of public investment.

[51]Figures A.2.1 and A.2.2 in the Appendix plots the quantile output elasticities for two individual components of the Economic Services sub-category for each sector, namely spending on transport and energy (for the stock specification). As can be seen from these figures, the direction of the results are consistent with those for the more aggregated categories of public investment.

Figure 3.11 illustrates the quantile effects of public investment on the capital intensity of formal sector firms. Surprisingly, public investment (both flow and stock) is negatively associated with the capital intensity of formal sector firms, suggesting that it may be a substitute for private factors. The negative effect is the largest for firms in the middle of the size distribution, giving the plot for the quantile effects a non-monotonic U-shaped curvature. In sharp contrast, the relationship between public investment and capital intensity for informal sector firms, as depicted in Figure 3.12, is strictly positive across the size distribution, with larger firms benefiting significantly more than their smaller counterparts. For example, for the top 20 percent of firms in this sector, an increase in public investment is associated with about a 20 percent increase in their capital intensity, which is almost twice as large as the effect for firms in the bottom 20 percent of firms. Figure 3.12 indicates that public investment can play a complementary role in influencing the usage of private factors in the informal sector, thereby pointing to an important role for public investment goods in this sector.[52]

3.5 Conclusions

Government investment in infrastructure goods such as roads, transportation, water and sanitation, and energy is a key element of public policy is developing countries. At the same time, these countries are, on average, characterized by a significant amount of production taking place in the informal sector, populated by small, unregistered firms, which produce non-traded goods and services that are highly labor intensive . As such, these firms have very low capital intensity in production and face significant barriers to outward labor mobility, relative to the formal sector. One possible way in which productivity may be influenced in this sector is through government provision of public goods such as infrastructure. However, very little is known about the spillovers

[52]Figures A.2.3-A.2.6 in the Appendix plots the association between the two sub-categories of public investment (Economic and Social Services) and capital intensity for formal and informal sector firms. The overall direction of the effects remain similar to the aggregated level of public investment, for both the flow and stock specifications.

generated by public investment for the informal sector, both with respect to output produced, as well as factor usage. In this paper, we use two firm-level datasets from India's manufacturing sector to estimate the output elasticities of public investment for firms in the formal and informal sector. We also examine how these output elasticities and relative capital intensity vary across the size distribution of firms in each sector.

Our results indicate that while public investment is an important factor in influencing firm-level productivity in both the formal and informal sectors, there are important sectoral differences. First, the average output elasticity of the flow measure of public investment for an informal sector firm is lower than that of its formal counterpart by a factor of about 3. When we consider a stock measure for public investment, this difference increases to a factor of 7, indicating that the benefits of the accumulated stock of public capital is much larger for firms in the formal sector. The sub-category of Economic Services, containing public expenditures on goods such as transport, communications, power, etc. is associated with systematically larger output elasticities relative to Social Services, which include spending on education, healthcare, water and sanitation, etc. In estimating these sectoral elasticities, we use a method proposed by Levinsohn and Petrin (2003) and Sivadasan (2009) to control for firm-level endogeneity in the usage of private factors of production. Second, results from our quantile regressions suggest that the size distribution of firms in each sector matter for the effects associated with public investment. For example, for formal sector firms, there is very little variation in the output elasticity of public investment across their size distribution. On the other hand, the corresponding output elasticity for informal sector firms is strictly increasing in firm size. Further, the relationship between public investment and the capital intensity in production for formal sector firms is negative, with the effect being the most negative for firms in the middle of the size distribution. This suggests that public investment may be a substitute for private factors in formal production. By contrast, the relationship between public investment and capital intensity is strictly increasing with firm size for the informal sector, indicating strong complementarities. Again, the largest firms in the informal sector benefit the most from public investment.

From a policy perspective, our results suggest that firms in the informal sector do indeed benefit significantly from public investment, even though these benefits are relatively smaller on average than those for their counterparts in the formal sector. The largest firms in the informal sector benefit the most from public investment, both with respect to the overall output elasticity as well as their capital intensity. Consequently, an effective way to increase the productivity and capital usage of informal sector firms might be to send more public investment goods to the largest firms in that sector. This may have the added advantage of lowering the relative size of the informal sector, by helping to formalize the largest and most productive firms, rather than a one-size-fits-all approach.

Tables

Table 3.1: Summary Statistics for Formal and Informal Sectors, 2010

	Formal		Informal	
	mean	sd	mean	sd
Gross value added (GVA) (in thousand Rs)	97603.0	677048.7	86.7	158.0
Net Fixed Assets (K) (in thousand Rs)	169607.2	2021480.7	231.8	840.7
Total workers (L)	192.2	697.1	2.2	1.7
K/L (in thousand Rs)	476.8	2771.8	91.9	221.1
Y/L (in thousand Rs)	346.5	3029.7	34.0	33.9
Rural	0.4	0.5	0.5	0.5
Age of firm	17.1	13.0	12.3	9.4
Registered under any act/ authority?	.	.	0.2	0.4
Ownership				
Wholly Central Government	0.002	0.05		
Wholly State and/or Local Govt	0.007	0.09		
Central Government and State jointly	0.002	0.04		
Joint Sector Public	0.007	0.08		
Joint Sector Private	0.009	0.09		
Wholly Private Ownership	1.0	0.2		
Proprietary (male)	.	.	0.7	0.4
Proprietary (female)	.		0.3	0.4
Partnership (w/ members of the same hh)			0.02	0.1
Partnership (w/ members from different hh)			0.005	0.07
Not known			.	.
Self-help Group			0.0008	0.03
Trusts			0.00007	0.009
Others			0.0001	0.01
Observations	32388		82748	

Table 3.2: Average State-wise Public Development Expenditures (2004-05 Rs)

Flow measure, 2006-2010	Avg. Share (% of NSDP): 2006-2010			Avg. per capita: 2006-2010		
States	Development	Social	Economic	Development	Social	Economic
JAMMU AND KASHMIR	13.0	4.8	8.2	3259.2	1191.5	2067.7
HIMACHAL PRADESH	4.9	1.9	2.9	2008.9	784.0	1224.9
PUNJAB	1.8	0.5	1.2	800.4	252.7	547.7
HARYANA	2.4	0.7	1.7	1324.9	406.5	918.4
DELHI	3.6	1.1	2.5	3031.7	899.3	2132.4
RAJASTHAN	2.9	1.3	1.6	744.3	336.6	407.7
UTTAR PRADESH	5.1	1.0	4.1	849.6	165.6	684.0
BIHAR	5.6	0.7	4.9	618.1	78.9	539.1
NAGALAND	9.0	3.5	5.5	2828.0	1109.4	1718.7
MANIPUR	19.0	7.2	11.8	4300.6	1630.7	2669.9
TRIPURA	7.6	2.9	4.6	2188.0	845.4	1342.6
MEGHALAYA	4.5	1.8	2.8	1435.0	553.1	881.8
ASSAM	3.1	0.5	2.6	622.2	91.9	530.2
WEST BENGAL	1.2	0.3	0.9	344.3	81.8	262.5
ORISSA	2.6	0.6	2.0	645.8	149.2	496.5
MADHYA PRADESH	5.3	0.9	4.4	1075.1	174.7	900.4
GUJARAT	2.7	0.6	2.1	1247.1	296.9	950.2
MAHARASTRA	2.3	0.3	2.0	1176.1	139.7	1036.4
ANDHRA PRADESH	3.7	0.6	3.1	1340.1	229.0	1111.1
KARNATAKA	3.9	1.0	2.9	1468.8	378.6	1090.2
GOA	3.7	1.0	2.7	3885.6	1028.2	2857.4
KERALA	1.4	0.4	1.0	590.6	173.7	417.0
TAMIL NADU	2.8	0.7	2.1	1267.6	314.1	953.5
Mean	**4.9**	**1.5**	**3.4**	**1611.0**	**491.8**	**1119.1**
S.D	**4.1**	**1.7**	**2.5**	**1119.6**	**432.9**	**716.7**
Observations	23	23	23	23	23	23

Table 3.3: State-wise Public Development Expenditures (2004-05 Rs)

Stock measure, 2010	Share (% of NSDP): 2010			Per capita: 2010		
States	Development	Social	Economic	Development	Social	Economic
JAMMU AND KASHMIR	86.6	27.7	58.9	24831.3	7940.3	16891.0
HIMACHAL PRADESH	41.4	17.6	23.8	19594.7	8311.7	11283.1
PUNJAB	16.5	3.4	13.1	8486.8	1747.2	6739.6
HARYANA	16.2	4.1	12.1	10843.5	2767.3	8076.2
DELHI	29.4	8.9	20.5	28272.2	8548.5	19723.7
RAJASTHAN	23.5	9.7	13.7	7390.6	3059.6	4331.0
UTTAR PRADESH	33.4	5.5	27.8	6284.0	1042.8	5241.2
BIHAR	35.9	4.5	31.4	4856.5	612.1	4244.4
NAGALAND	83.3	34.5	48.8	27375.2	11334.4	16040.8
MANIPUR	118.8	46.1	72.7	27782.9	10780.4	17002.4
TRIPURA	75.0	28.6	46.3	23749.5	9075.0	14674.5
MEGHALAYA	47.5	18.3	29.2	16247.4	6260.1	9987.3
ASSAM	26.6	3.6	22.9	6225.4	846.4	5379.0
WEST BENGAL	14.2	2.0	12.2	4711.1	651.9	4059.2
ORISSA	21.7	5.1	16.6	6075.6	1432.2	4643.5
MADHYA PRADESH	36.5	5.8	30.7	8280.3	1320.8	6959.5
GUJARAT	21.5	5.8	15.7	11935.2	3239.0	8696.2
MAHARASTRA	20.4	2.0	18.4	11968.2	1193.8	10774.3
ANDHRA PRADESH	29.8	5.3	24.5	12681.2	2267.5	10413.7
KARNATAKA	26.5	6.3	20.2	11592.6	2758.6	8834.0
GOA	24.8	7.1	17.7	29143.2	8300.0	20843.2
KERALA	10.8	2.6	8.2	5246.2	1249.1	3997.1
TAMIL NADU	16.5	5.3	11.2	9046.4	2913.9	6132.5
Mean	**37.3**	**11.3**	**25.9**	**14027.0**	**4245.8**	**9781.2**
S.D	**27.6**	**11.9**	**16.4**	**8626.3**	**3611.3**	**5305.1**
Observations	23	23	23	23	23	23

Table 3.4: OLS Estimation of Production Function, Formal Sector with Flow Measure of Public Investment

Sector: Formal			
Dependent variable: $ln\,GVA$	(1)	(2)	(3)
ln L	0.778***	0.778***	0.778***
	(0.005)	(0.005)	(0.005)
ln K	0.325***	0.325***	0.325***
	(0.003)	(0.003)	(0.003)
ln Development expenditure per capita	0.034*		
	(0.015)		
ln Social Services expenditure per capita		0.003	
		(0.014)	
ln Economic Services expenditure per capita			0.033*
			(0.014)
Firm-level controls			
Age of the firm	0.002***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)
Rural	-0.110***	-0.110***	-0.110***
	(0.011)	(0.011)	(0.011)
Ownership			
Wholly State and/or Local Govt	-0.302**	-0.302**	-0.301**
	(0.114)	(0.114)	(0.114)
Central Govt and State jointly	-0.107	-0.106	-0.106
	(0.150)	(0.150)	(0.150)
Joint Sector Public	-0.317**	-0.317**	-0.317**
	(0.116)	(0.116)	(0.116)
Joint Sector Private	-0.305**	-0.304**	-0.305**
	(0.112)	(0.112)	(0.112)
Wholly Private Ownership	-0.303**	-0.302**	-0.302**
	(0.101)	(0.101)	(0.101)
State-level controls			
Log of NSDP per capita (2010)	0.074***	0.089***	0.075***
	(0.020)	(0.020)	(0.020)
Log of Total Labor Force (2010)	0.542***	0.549***	0.539***
	(0.025)	(0.026)	(0.026)
Literacy rate (2011)	0.012***	0.012***	0.012***
	(0.001)	(0.001)	(0.001)
Old-age Dependency ratio (2001)	-0.033***	-0.036***	-0.033***
	(0.004)	(0.004)	(0.004)
Crime rate per hundred (2010)	-0.744***	-0.741***	-0.730***
	(0.081)	(0.085)	(0.081)
Share of Unregistered Manufacturing (in total manufacturing)	-0.001*	-0.001**	-0.001*
	(0.001)	(0.000)	(0.001)
Log of total number of enterprises	-0.511***	-0.521***	-0.512***
	(0.024)	(0.023)	(0.023)
Constant	5.364***	5.514***	5.433***
	(0.295)	(0.322)	(0.287)
Industry dummies (NIC 2-digit)	Yes	Yes	Yes
N	32388	32388	32388

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

Table 3.5: LP-S Estimation of Production Function, Formal Sector with Flow Measure of Public Investment

Sector: Formal			
Dependent variable: $ln\,GVA$	(1)	(2)	(3)
ln L	0.664***	0.665***	0.665***
	(0.005)	(0.005)	(0.005)
ln K	0.369***	0.369***	0.369***
	(0.002)	(0.002)	(0.002)
ln Development expenditure per capita	0.088***		
	(0.015)		
ln Social Services expenditure per capita		0.045***	
		(0.014)	
ln Economic Services expenditure per capita			0.077***
			(0.013)
Firm-level controls			
Age of the firm	0.001*	0.001	0.001*
	(0.000)	(0.000)	(0.000)
Rural	-0.105***	-0.104***	-0.106***
	(0.010)	(0.010)	(0.010)
Ownership			
Wholly State and/or Local Govt	-0.186	-0.190	-0.184
	(0.108)	(0.108)	(0.108)
Central Govt and State jointly	-0.060	-0.062	-0.057
	(0.142)	(0.142)	(0.142)
Joint Sector Public	-0.254*	-0.254*	-0.253*
	(0.110)	(0.110)	(0.110)
Joint Sector Private	-0.248*	-0.247*	-0.247*
	(0.106)	(0.106)	(0.106)
Wholly Private Ownership	-0.225*	-0.225*	-0.223*
	(0.095)	(0.095)	(0.095)
State-level controls			
Log of NSDP per capita (2010)	-0.057**	-0.035	-0.050**
	(0.020)	(0.019)	(0.019)
Log of Total Labor Force (2010)	0.402***	0.438***	0.397***
	(0.024)	(0.025)	(0.024)
Literacy rate (2011)	0.015***	0.016***	0.015***
	(0.001)	(0.001)	(0.001)
Old-age Dependency ratio (2001)	-0.024***	-0.028***	-0.024***
	(0.004)	(0.004)	(0.004)
Crime rate per hundred (2010)	-0.621***	-0.685***	-0.586***
	(0.077)	(0.081)	(0.077)
Share of Unregistered Manufacturing (in total manufacturing)	-0.002**	-0.002***	-0.001**
	(0.000)	(0.000)	(0.000)
Log of total number of enterprises	-0.396***	-0.423***	-0.401***
	(0.022)	(0.022)	(0.022)
Constant	7.851***	7.763***	8.058***
	(0.495)	(0.511)	(0.492)
Industry dummies (NIC 2-digit)	Yes	Yes	Yes
N	32388	32388	32388

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

Table 3.6: LP-S Estimation of Production Function, Formal Sector with Stock Measure of Public Investment

Sector: Formal			
Dependent variable: $ln\,GVA$	(1)	(2)	(3)
ln L	0.664***	0.665***	0.665***
	(0.005)	(0.005)	(0.005)
ln K	0.369***	0.369***	0.369***
	(0.002)	(0.002)	(0.002)
ln Development expenditure per capita	0.173***		
	(0.019)		
ln Social Services expenditure per capita		0.036**	
		(0.013)	
ln Economic Services expenditure per capita			0.159***
			(0.017)
Firm-level controls			
Age of the firm	0.001*	0.001	0.001*
	(0.000)	(0.000)	(0.000)
Rural	-0.107***	-0.106***	-0.107***
	(0.010)	(0.010)	(0.010)
Ownership			
Wholly State and/or Local Govt	-0.183	-0.190	-0.178
	(0.108)	(0.108)	(0.108)
Central Govt and State jointly	-0.060	-0.061	-0.054
	(0.142)	(0.142)	(0.142)
Joint Sector Public	-0.256*	-0.253*	-0.256*
	(0.110)	(0.110)	(0.110)
Joint Sector Private	-0.246*	-0.246*	-0.244*
	(0.106)	(0.106)	(0.106)
Wholly Private Ownership	-0.224*	-0.224*	-0.220*
	(0.095)	(0.095)	(0.095)
State-level controls			
Log of NSDP per capita (2010)	-0.121***	-0.034	-0.115***
	(0.022)	(0.020)	(0.021)
Log of Total Labor Force (2010)	0.381***	0.437***	0.365***
	(0.024)	(0.025)	(0.025)
Literacy rate (2011)	0.015***	0.016***	0.014***
	(0.001)	(0.001)	(0.001)
Old-age Dependency ratio (2001)	-0.016***	-0.027***	-0.018***
	(0.004)	(0.004)	(0.004)
Crime rate per hundred (2010)	-0.431***	-0.684***	-0.318***
	(0.079)	(0.082)	(0.083)
Share of Unregistered Manufacturing (in total manufacturing)	-0.002**	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)
Log of total number of enterprises	-0.367***	-0.427***	-0.365***
	(0.023)	(0.022)	(0.023)
Constant	7.400***	7.794***	7.818***
	(0.499)	(0.515)	(0.493)
Industry dummies (NIC 2-digit)	Yes	Yes	Yes
N	32388	32388	32388

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

Table 3.7: OLS Estimation of Production Function, Informal Sector with Flow Measure of Public Investment

Sector: Informal **Dependent variable:** $ln\, GVA$	(1)	(2)	(3)
ln L	0.820***	0.820***	0.820***
	(0.005)	(0.005)	(0.005)
ln K	0.252***	0.254***	0.252***
	(0.002)	(0.002)	(0.002)
ln Development expenditure per capita	-0.002		
	(0.006)		
ln Social Services expenditure per capita		-0.048***	
		(0.006)	
ln Economic Services expenditure per capita			0.009
			(0.006)
Firm-level controls			
Age of the firm	-0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)
Registration	0.267***	0.263***	0.268***
	(0.007)	(0.007)	(0.007)
Rural	-0.092***	-0.094***	-0.092***
	(0.005)	(0.005)	(0.005)
Ownership			
Proprietary(female)	-0.656***	-0.656***	-0.657***
	(0.007)	(0.007)	(0.007)
Partnership with members of the same hhd	-0.189***	-0.195***	-0.188***
	(0.019)	(0.019)	(0.019)
Partnership between members from different hhd	0.125***	0.125***	0.125***
	(0.035)	(0.035)	(0.035)
Self-help Group	-1.341***	-1.345***	-1.341***
	(0.087)	(0.087)	(0.087)
Trusts	-1.045***	-1.042***	-1.046***
	(0.287)	(0.287)	(0.287)
Others	0.015	-0.002	0.019
	(0.212)	(0.212)	(0.212)
State-level controls			
Log of NSDP per capita (2010)	0.215***	0.252***	0.209***
	(0.010)	(0.010)	(0.010)
Log of Total Labor Force (2010)	-0.056***	-0.078***	-0.058***
	(0.011)	(0.012)	(0.011)
Literacy rate (2011)	-0.005***	-0.007***	-0.005***
	(0.001)	(0.001)	(0.001)
Old-age Dependency ratio (2001)	-0.008***	-0.007***	-0.008***
	(0.002)	(0.002)	(0.002)
Crime rate per hundred (2010)	0.488***	0.512***	0.489***
	(0.038)	(0.038)	(0.038)
Share of Registered Manufacturing (in total manufacturing)	-0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)
Log of total number of enterprises	-0.020*	-0.024*	-0.016
	(0.010)	(0.010)	(0.010)
Constant	5.998***	6.116***	6.085***
	(0.147)	(0.159)	(0.142)
Industry dummies (NIC 2-digit)	Yes	Yes	Yes
N	82748	82748	82748

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

Table 3.8: LP-S Estimation of Production Function, Informal Sector with Flow Measure of Public Investment

Sector: Informal			
Dependent variable: $ln\,GVA$	(1)	(2)	(3)
ln L	0.628***	0.628***	0.628***
	(0.005)	(0.005)	(0.005)
ln K	0.317***	0.319***	0.317***
	(0.001)	(0.002)	(0.002)
ln Development expenditure per capita	0.027***		
	(0.006)		
ln Social Services expenditure per capita		-0.026***	
		(0.005)	
ln Economic Services expenditure per capita			0.039***
			(0.005)
Firm-level controls			
Age of the firm	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)
Registration	0.164***	0.159***	0.164***
	(0.006)	(0.006)	(0.006)
Rural	-0.102***	-0.104***	-0.102***
	(0.005)	(0.005)	(0.005)
Ownership			
Proprietary(female)	-0.430***	-0.430***	-0.430***
	(0.007)	(0.007)	(0.007)
Partnership with members of the same hhd	-0.140***	-0.146***	-0.138***
	(0.017)	(0.017)	(0.017)
Partnership between members from different hhd	0.070*	0.069*	0.070*
	(0.031)	(0.031)	(0.031)
Self-help Group	-1.099***	-1.101***	-1.099***
	(0.078)	(0.078)	(0.078)
Trusts	-0.849***	-0.847***	-0.849***
	(0.257)	(0.257)	(0.257)
Others	0.060	0.039	0.065
	(0.190)	(0.190)	(0.190)
State-level controls			
Log of NSDP per capita (2010)	0.114***	0.151***	0.110***
	(0.009)	(0.009)	(0.009)
Log of Total Labor Force (2010)	-0.098***	-0.108***	-0.103***
	(0.010)	(0.010)	(0.010)
Literacy rate (2011)	-0.002***	-0.004***	-0.002***
	(0.001)	(0.001)	(0.001)
Old-age Dependency ratio (2001)	0.003	0.003	0.003
	(0.002)	(0.002)	(0.002)
Crime rate per hundred (2010)	0.586***	0.598***	0.592***
	(0.034)	(0.034)	(0.034)
Share of Registered Manufacturing (in total manufacturing)	-0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)
Log of total number of enterprises	0.039***	0.027**	0.043***
	(0.009)	(0.009)	(0.009)
Constant	8.031***	8.410***	8.027***
	(0.314)	(0.315)	(0.313)
Industry dummies (NIC 2-digit)	Yes	Yes	Yes
N	82748	82748	82748

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

Table 3.9: LP-S Estimation of Production Function, Informal Sector with Stock Measure of Public Investment

Sector: Informal			
Dependent variable: $ln\, GVA$	(1)	(2)	(3)
ln L	0.628***	0.629***	0.628***
	(0.005)	(0.005)	(0.005)
ln K	0.317***	0.319***	0.317***
	(0.001)	(0.002)	(0.002)
ln Development expenditure per capita	0.024**		
	(0.008)		
ln Social Services expenditure per capita		-0.014**	
		(0.005)	
ln Economic Services expenditure per capita			0.036***
			(0.007)
Firm-level controls			
Age of the firm	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)
Registration	0.162***	0.160***	0.162***
	(0.006)	(0.006)	(0.006)
Rural	-0.103***	-0.104***	-0.103***
	(0.005)	(0.005)	(0.005)
Ownership			
Proprietary(female)	-0.430***	-0.430***	-0.430***
	(0.007)	(0.007)	(0.007)
Partnership with members of the same hhd	-0.141***	-0.145***	-0.140***
	(0.017)	(0.017)	(0.017)
Partnership between members from different hhd	0.070*	0.069*	0.070*
	(0.031)	(0.031)	(0.031)
Self-help Group	-1.100***	-1.101***	-1.102***
	(0.078)	(0.078)	(0.078)
Trusts	-0.848***	-0.847***	-0.848***
	(0.257)	(0.257)	(0.257)
Others	0.058	0.042	0.060
	(0.190)	(0.190)	(0.190)
State-level controls			
Log of NSDP per capita (2010)	0.114***	0.144***	0.107***
	(0.010)	(0.010)	(0.010)
Log of Total Labor Force (2010)	-0.096***	-0.105***	-0.098***
	(0.010)	(0.010)	(0.010)
Literacy rate (2011)	-0.003***	-0.004***	-0.003***
	(0.001)	(0.001)	(0.001)
Old-age Dependency ratio (2001)	0.003	0.002	0.004*
	(0.002)	(0.002)	(0.002)
Crime rate per hundred (2010)	0.602***	0.596***	0.622***
	(0.035)	(0.035)	(0.035)
Share of Registered Manufacturing (in total manufacturing)	-0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)
Log of total number of enterprises	0.035***	0.029***	0.037***
	(0.009)	(0.009)	(0.009)
Constant	8.016***	8.338***	7.980***
	(0.317)	(0.317)	(0.315)
Industry dummies (NIC 2-digit)	Yes	Yes	Yes
N	82748	82748	82748

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

Figures

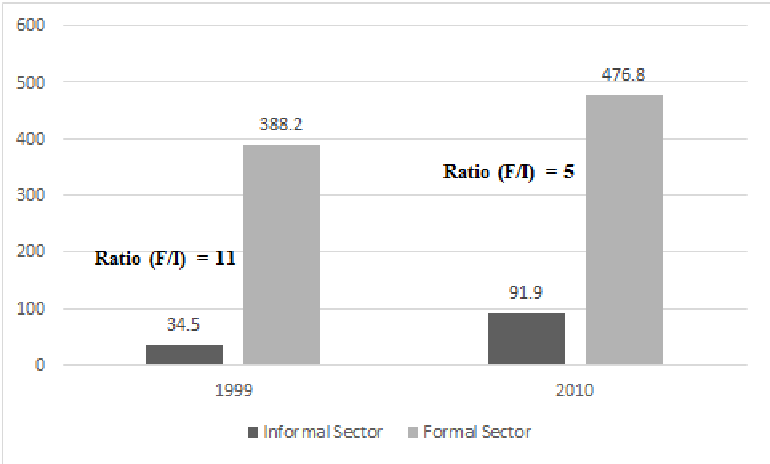


Figure 3.1: Sectoral Capital Intensity, 1999 and 2010

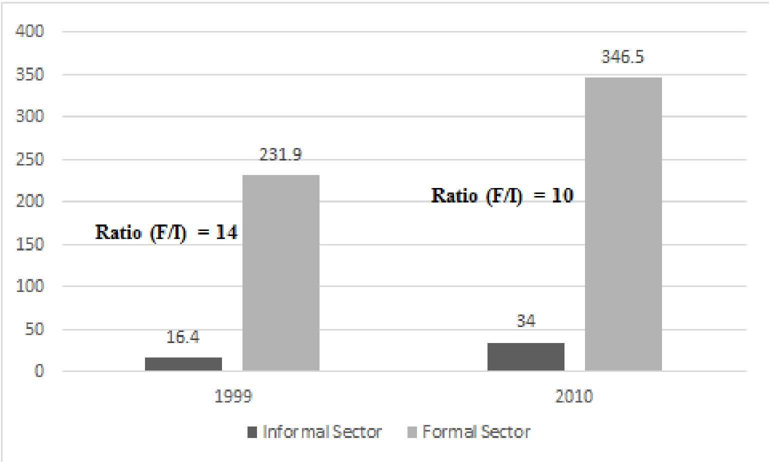


Figure 3.2: Sectoral Output Per Worker, 1999 and 2010

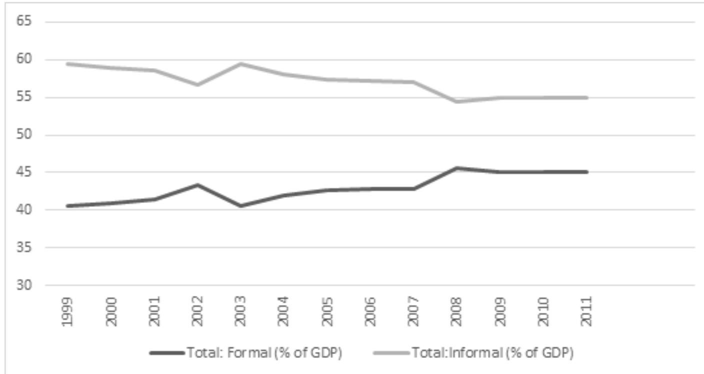


Figure 3.3: Share of Informal Sector in GDP, 1999-2010
Source: National Accounts Statistics Reports, India

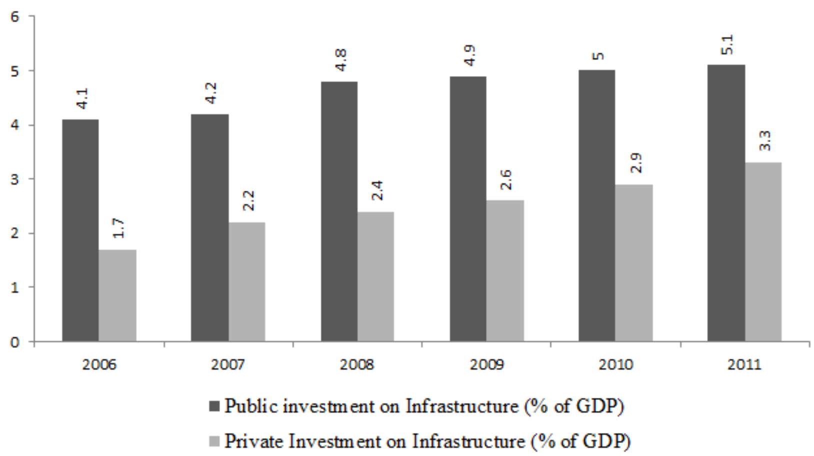
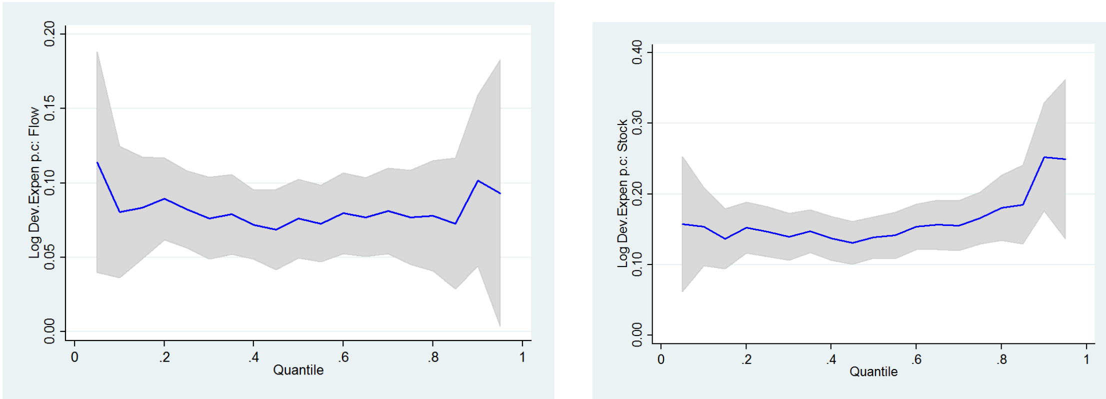
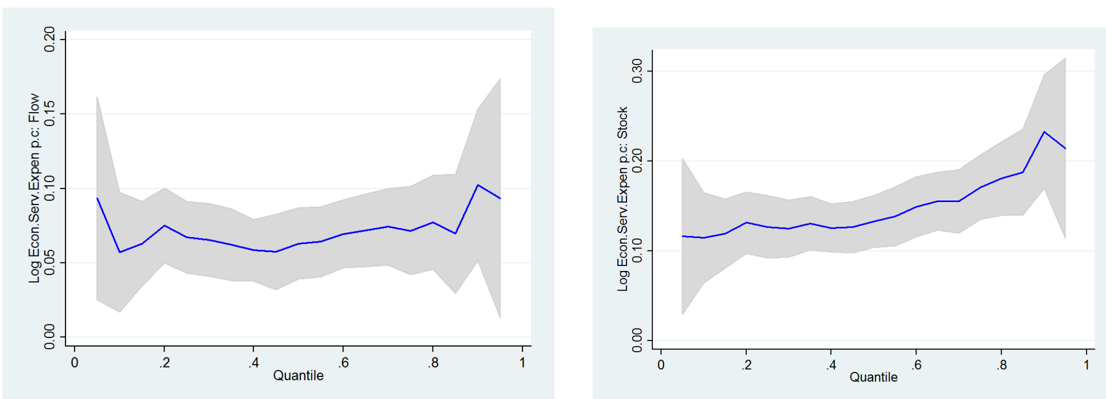


Figure 3.4: Share of Infrastructure Spending in GDP, 2006-2010
Source: Planning Commission, India



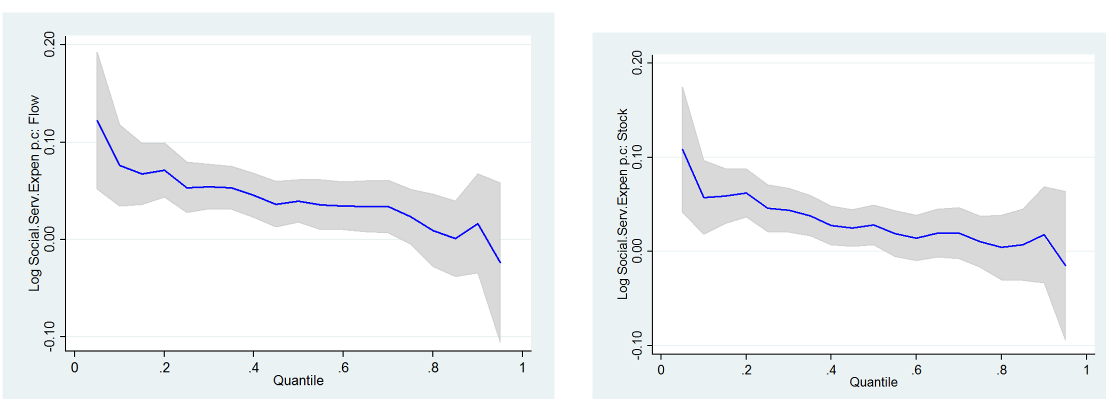
(a) Flow (b) Stock

Figure 3.5: Quantile Output Elasticity of Public Investment, Formal Sector



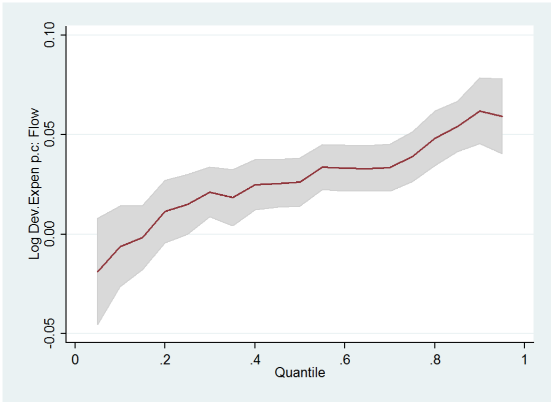
(a) Flow (b) Stock

Figure 3.6: Quantile Output Elasticity: Economic Services, Formal Sector

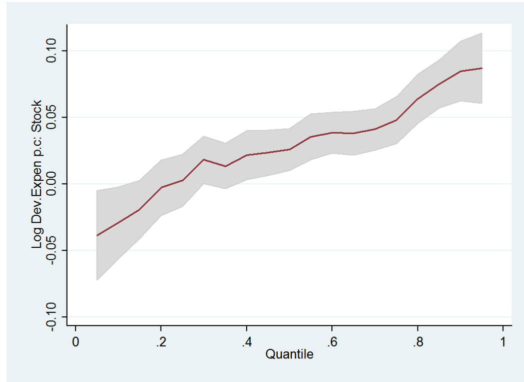


(a) Flow (b) Stock

Figure 3.7: Quantile Output Elasticity: Social Services, Formal Sector

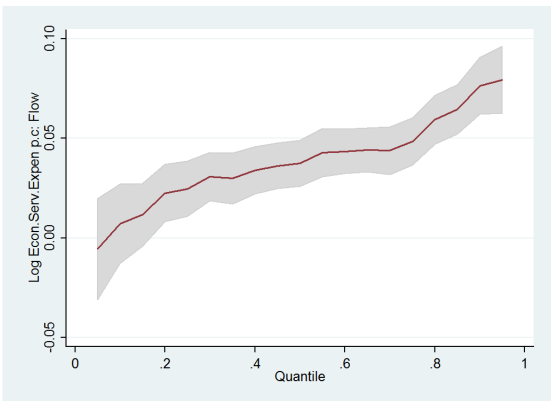


(a) Flow

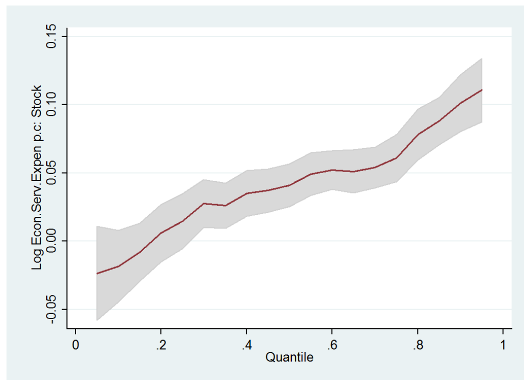


(b) Stock

Figure 3.8: Quantile Output Elasticity of Public Investment, Informal Sector

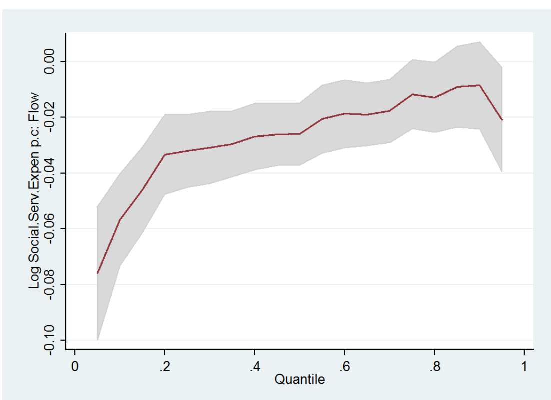


(a) Flow

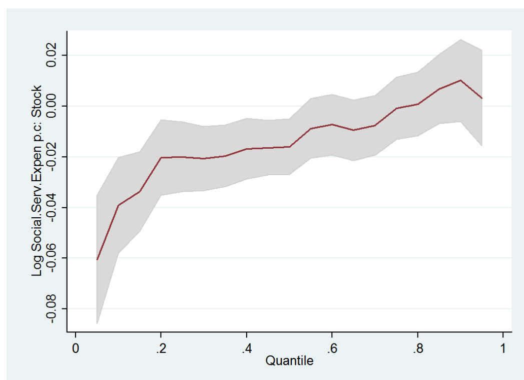


(b) Stock

Figure 3.9: Quantile Output Elasticity: Economic Services, Informal Sector

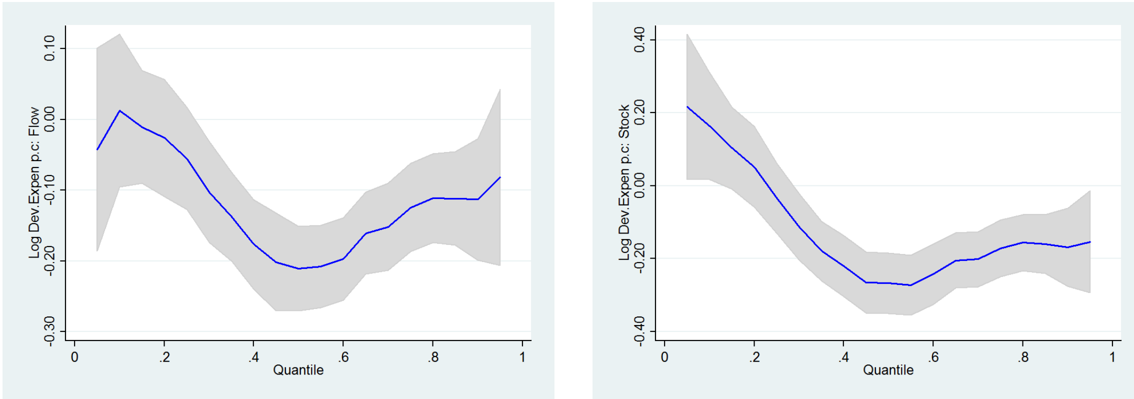


(a) Flow



(b) Stock

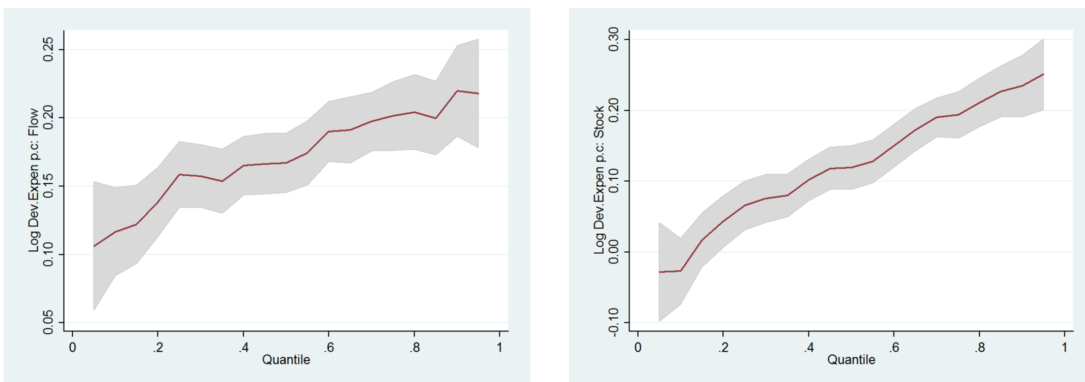
Figure 3.10: Quantile Output Elasticity: Social Services, Informal Sector



(a) Flow

(b) Stock

Figure 3.11: Public Investment and Firm-level Capital Intensity, Formal Sector



(a) Flow

(b) Stock

Figure 3.12: Public Investment and Firm-level Capital Intensity, Informal Sector

Appendix 2

Table A.2.1: Data Sources

State Controls	**Source**
NSDP per capita (2010)	Reserve Bank of India
Total Labor Force (2010)	National Sample Survey Reports, Census
Literacy rate (2011)	Planning Commission
Old-age Dependency ratio (2001)	IndiaStat, Census
Crime rate per hundred (2010)	Crime Records Bureau
Share of Registered Manufacturing (in total manufacturing)	Reserve Bank of India
Total number of enterprises	National Sample Survey Reports

Table A.2.2: Summary Statistics for Formal and Informal Sectors, 1999

	Formal		Informal	
	mean	sd	mean	sd
Gross value added (GVA) (in thousand Rs)	72453.5	600452.5	36.1	116.7
Net Fixed Assets (K) (in thousand Rs)	156303.4	1729834.5	79.6	429.1
Total workers (L)	181.8	951.2	2.0	1.4
K/L (in thousand Rs)	388.2	2282.2	34.5	96.8
Y/L (in thousand Rs)	231.9	718.8	16.4	20.8
Rural	0.3	0.5	0.6	0.5
Age of firm	16.4	11.9	.	.
Registered under any act/ authority?	.	.	0.1	0.3
Ownership				
Wholly Central Government	0.007	0.08		
Wholly State and/or Local Govt	0.01	0.1		
Central Government and State jointly	0.005	0.07		
Joint Sector Public	0.02	0.1		
Joint Sector Private	0.009	0.10		
Wholly Private Ownership	0.9	0.2		
Proprietary (male)	.	.	0.8	0.4
Proprietary (female)	.		0.2	0.4
Partnership (w/ members of the same hh)			0.01	0.1
Partnership (w/ members from different hh)			0.006	0.07
Not known			0	0
Self-help Group			.	.
Trusts			.	.
Others			.	.
Observations	19095		49720	

Table A.2.3: OLS Estimation of Production Function, Formal Sector with Flow Measure of Public Investment

(with standard errors clustered at the State-NIC2 level)

Sector: Formal			
Dependent variable: $ln\, GVA$	(1)	(2)	(3)
ln L	0.778***	0.778***	0.778***
	(0.024)	(0.024)	(0.024)
ln K	0.325***	0.325***	0.325***
	(0.019)	(0.019)	(0.019)
ln Development expenditure per capita	0.034		
	(0.042)		
ln Social Services expenditure per capita		0.003	
		(0.036)	
ln Economic Services expenditure per capita			0.033
			(0.037)
Firm-level controls			
Age of the firm	0.002*	0.001	0.002*
	(0.001)	(0.001)	(0.001)
Rural	-0.110***	-0.110***	-0.110***
	(0.017)	(0.018)	(0.017)
Ownership			
Wholly State and/or Local Govt	-0.302	-0.302	-0.301
	(0.183)	(0.183)	(0.183)
Central Govt and State jointly	-0.107	-0.106	-0.106
	(0.202)	(0.202)	(0.202)
Joint Sector Public	-0.317	-0.317	-0.317
	(0.197)	(0.197)	(0.197)
Joint Sector Private	-0.305	-0.304	-0.305
	(0.171)	(0.170)	(0.171)
Wholly Private Ownership	-0.303*	-0.302*	-0.302*
	(0.151)	(0.151)	(0.151)
State-level controls			
Log of NSDP per capita (2010)	0.074	0.089	0.075
	(0.059)	(0.057)	(0.058)
Log of Total Labor Force (2010)	0.542***	0.549***	0.539***
	(0.069)	(0.063)	(0.069)
Literacy rate (2011)	0.012***	0.012***	0.012***
	(0.003)	(0.003)	(0.003)
Old-age Dependency ratio (2001)	-0.033**	-0.036**	-0.033**
	(0.011)	(0.011)	(0.011)
Crime rate per hundred (2010)	-0.744**	-0.741**	-0.730**
	(0.261)	(0.267)	(0.262)
Share of Unregistered Manufacturing (in total manufacturing)	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Log of total number of enterprises	-0.511***	-0.521***	-0.512***
	(0.062)	(0.058)	(0.061)
Constant	5.364***	5.514***	5.433***
	(0.694)	(0.747)	(0.689)
Industry dummies (NIC 2-digit)	Yes	Yes	Yes
N	32388	32388	32388

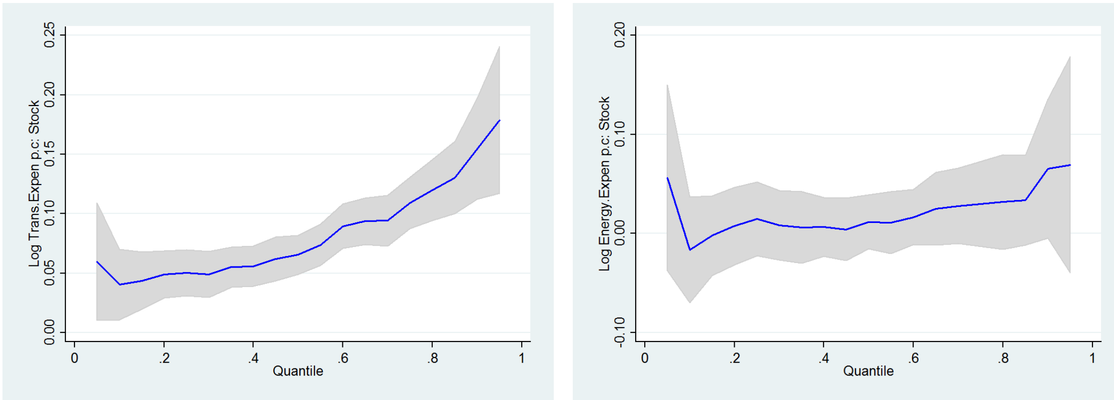
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.2.4: OLS Estimation of Production Function, Informal Sector with Flow Measure of Public Investment

(with standard errors clustered at the State-NIC2 level)

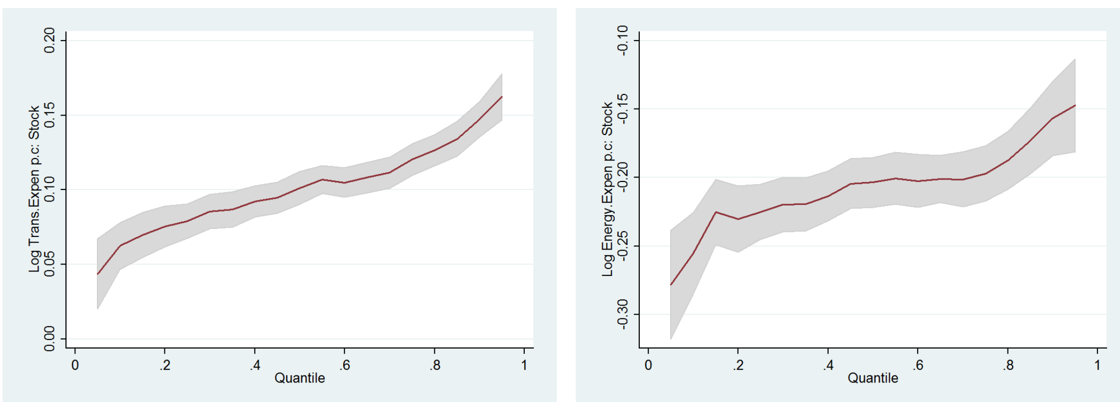
Sector: Informal **Dependent variable:** $ln\, GVA$	(1)	(2)	(3)
ln L	0.820***	0.820***	0.820***
	(0.020)	(0.018)	(0.021)
ln K	0.252***	0.254***	0.252***
	(0.008)	(0.008)	(0.007)
ln Development expenditure per capita	-0.002		
	(0.030)		
ln Social Services expenditure per capita		-0.048	
		(0.033)	
ln Economic Services expenditure per capita			0.009
			(0.025)
Firm-level controls			
Age of the firm	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)
Registration	0.267***	0.263***	0.268***
	(0.020)	(0.020)	(0.019)
Rural	-0.092***	-0.094***	-0.092***
	(0.015)	(0.013)	(0.015)
Ownership			
Proprietary(female)	-0.656***	-0.656***	-0.657***
	(0.023)	(0.024)	(0.023)
Partnership with members of the same hhd	-0.189***	-0.195***	-0.188***
	(0.033)	(0.031)	(0.033)
Partnership between members from different hhd	0.125**	0.125**	0.125**
	(0.040)	(0.040)	(0.039)
Self-help Group	-1.341***	-1.345***	-1.341***
	(0.181)	(0.173)	(0.231)
Trusts	-1.045***	-1.042***	-1.046***
	(0.256)	(0.223)	(0.253)
Others	0.015	-0.002	0.019
	(0.258)	(0.271)	(0.265)
State-level controls			
Log of NSDP per capita (2010)	0.215***	0.252***	0.209***
	(0.056)	(0.061)	(0.048)
Log of Total Labor Force (2010)	-0.056	-0.078	-0.058
	(0.059)	(0.064)	(0.055)
Literacy rate (2011)	-0.005	-0.007*	-0.005
	(0.004)	(0.004)	(0.003)
Old-age Dependency ratio (2001)	-0.008	-0.007	-0.008
	(0.012)	(0.011)	(0.011)
Crime rate per hundred (2010)	0.488	0.512	0.489*
	(0.263)	(0.314)	(0.244)
Share of Registered Manufacturing (in total manufacturing)	-0.004**	-0.004**	-0.004***
	(0.001)	(0.001)	(0.001)
Log of total number of enterprises	-0.020	-0.024	-0.016
	(0.053)	(0.057)	(0.045)
Constant	7.408***	7.789***	7.346***
	(0.696)	(0.760)	(0.683)
Industry dummies (NIC 2-digit)	Yes	Yes	Yes
N	82748	82748	82748

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



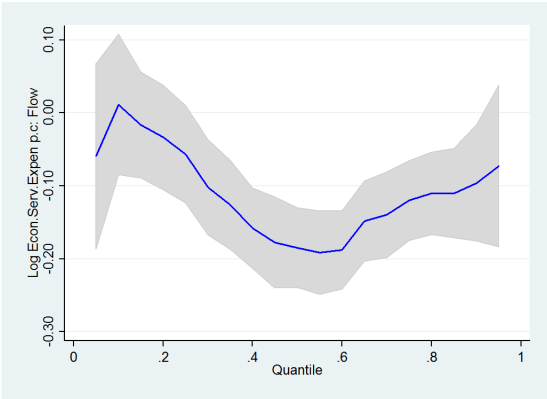
(a) Transport Stock (b) Energy Stock

Figure A.2.1: Quantile Output Elasticities for Transport and Energy Expenditures, Formal Sector

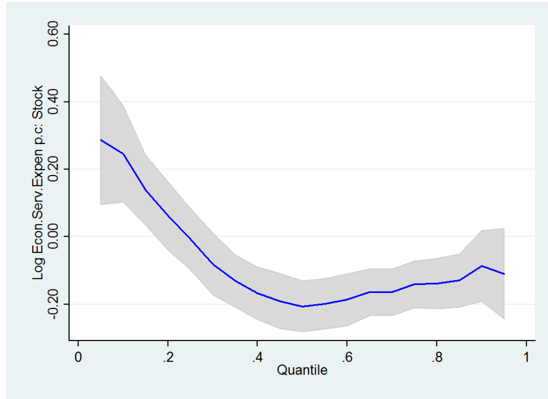


(a) Transport Stock (b) Energy Stock

Figure A.2.2: Quantile Output Elasticities for Transport and Energy Expenditures, Informal Sector

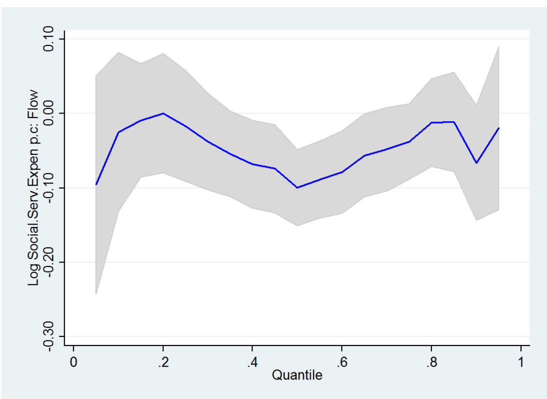


(a) Flow

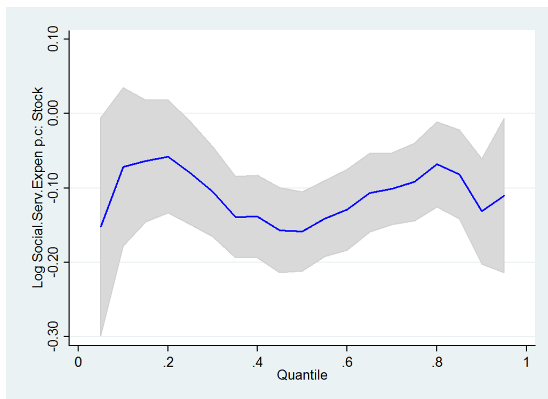


(b) Stock

Figure A.2.3: Economic Services Expenditures and Firm-level Capital Intensity, Formal Sector

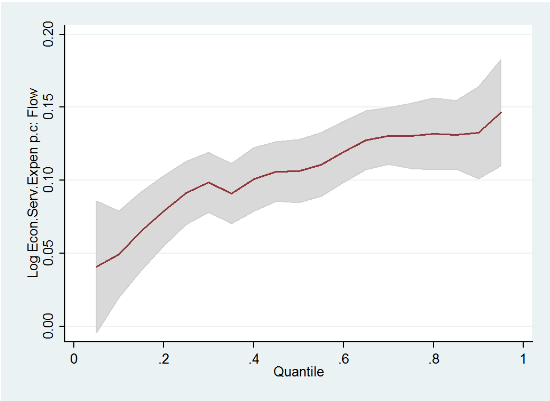


(a) Flow

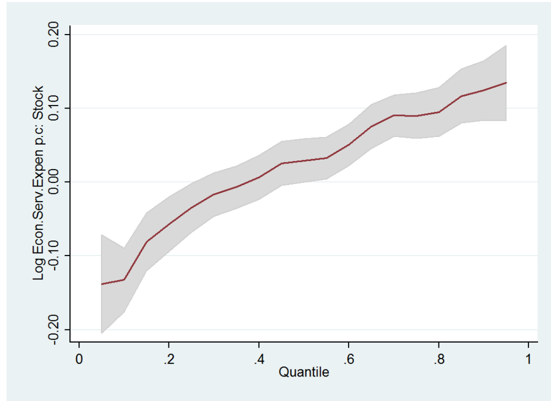


(b) Stock

Figure A.2.4: Social Services Expenditures and Firm-level Capital Intensity, Formal Sector

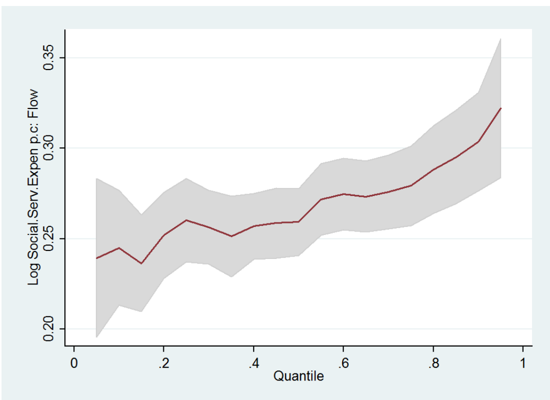


(a) Flow

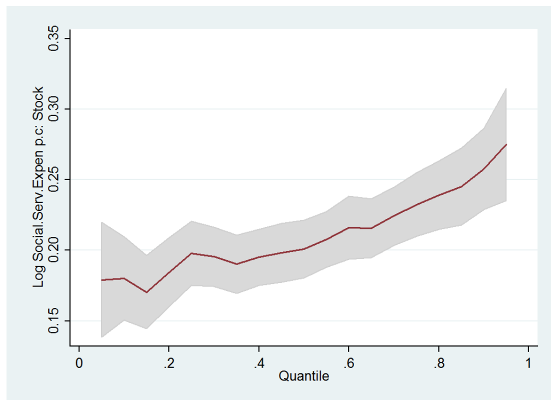


(b) Stock

Figure A.2.5: Economic Services Expenditures and Firm-level Capital Intensity, Informal Sector

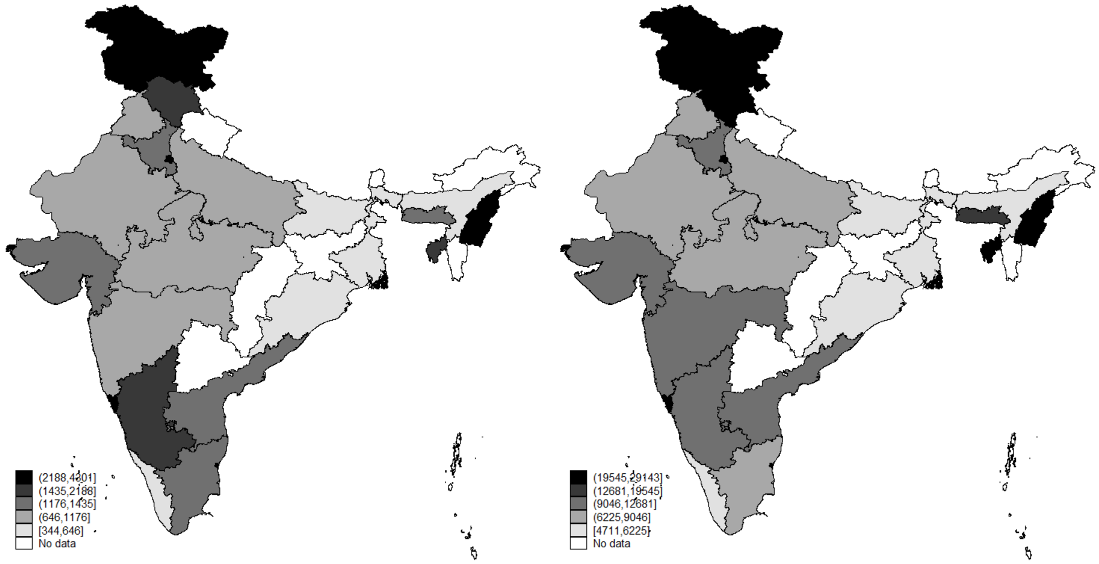


(a) Flow



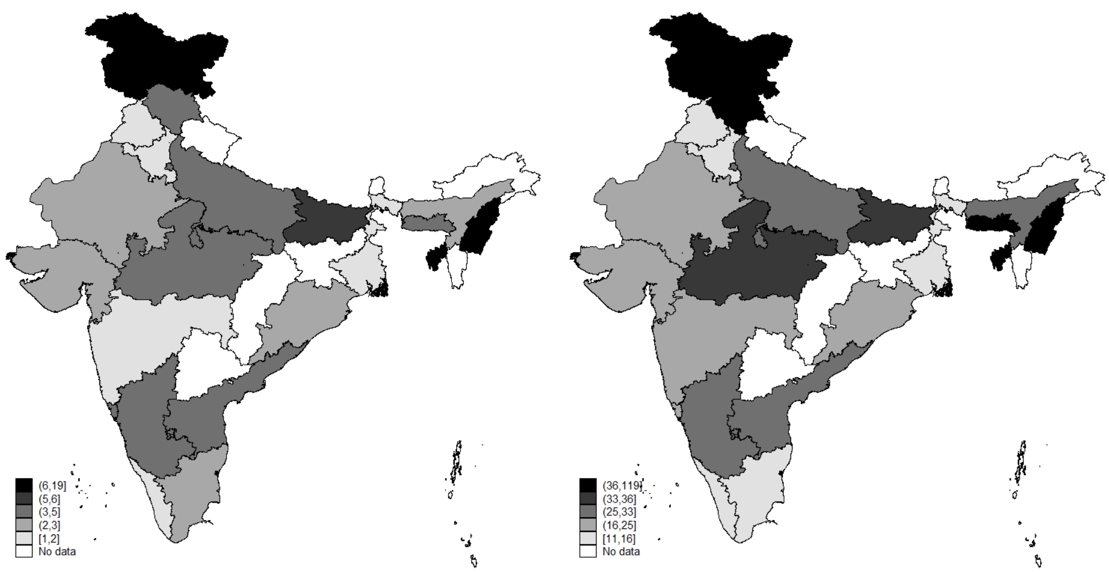
(b) Stock

Figure A.2.6: Social Services Expenditures and Firm-level Capital Intensity, Informal Sector



(a) Flow (b) Stock

Figure A.2.7: Average State-wise Public Development Expenditures, 2006-10 (in 2004-05 Rs)



(a) Flow (b) Stock

Figure A.2.8: Average State-wise Public Development Expenditures, 2006-10 (Share of NSDP)

Robustness Check for Self-Selection

In this study we use firm-level data to estimate the output elasticity of public investment. This disaggregated level of analysis allows us to assume that a firm's output is exogenous to the state-level public investment. We circumvent the problem of endogeneity of aggregate output (e.g. GDP, State Domestic Product) and public investment as highlighted in earlier studies by using firm-level data. However, the firm-level analysis poses another challenge regarding self-selection of firms. Specifically, firms may choose to be located in states where public investment is higher, and hence choosing a state with better infrastructure. We think self-selection could be an issue with the formal sector firms rather than the informal sector firms because the latter are relatively less mobile. We do some robustness checks to see if self-selection is driving our results. Specifically, we estimate the output elasticity of public investment for the young formal sector firms (age less than 1 year, 3 years and 6 years) with respect to the stock measure of public investment. I have also reproduced the estimates with respect to the median age (14 years). Table A.2.5 presents the distribution of firms across firm age. If there is selection, we would expect to see a higher output elasticity of public investment for these younger firms than the average elasticity. This is because these new firms may choose to be located in the states with higher public investment and hence would reflect higher output elasticity with respect to public investment. We would expect to see this result for the stock measure and not so much for the flow measure, because we expect the firms to look at the overall stock of public investment rather than the flow when deciding on its location. Also, the flow measure may pick up contemporaneous fluctuations. We focus our attention to Table A.2.6 that reports estimates for the stock measure. Here we see that the output elasticity of public investment for formal sector firms aged less than 1 year are a bit higher than the average elasticity but they are not statistically significant. For the firms aged less than 3 years, the elasticities are very close to the average elasticities. The elasticities are also quite close to the average elasticity for the firms less than 6 years and for firms less than the median age (14 years). Thus we cannot claim that these young formal sector firms are self-selecting into states with a large stock of public infrastructure. These robustness checks further corroborate our finding that public infrastructure do play a significant role on a firm's output. We can thus claim that the positive effect of public infrastructure on a firm's output is not picking up any self-selection of firms into the high spending states.

Table A.2.5: Firm Age Statistics (Formal Sector)

Quantile	**Firm Age (yrs)**
2	1
	[570]
11	3
	[3564]
25	6
	[8407]
50	14
	[17055]
Mean	17.1
	[20247]

Figures in brackets
represent No. of observations

Table A.2.6: Output Elasticities of Public Investment across Firm Age (LP-S method)

(with Public Capital Stock per capita)

Sector: Formal
Dependent variable: $ln\, GVA$

Age	Dev. p.c	Social p.c	Econ. p.c
At the mean	0.173***	0.036**	0.159***
Less than= 1	0.352	0.072	0.328
Greater than 1	0.178***	0.034 **	0.164***
Less than= 3	0.174**	0.062	0.143*
Greater than 3	0.180***	0.029*	0.168***
Less than= 6	0.210***	0.093***	0.170***
Greater than 6	0.138***	-0.008	0.144***
Less than= 14	0.209***	0.099***	0.166***
Greater than 14	0.075*	-0.075***	0.115***

Chapter 4

Informality, Congestion, and Public Capital Efficiency: A Case for Optimal Taxation and Maintenance Allocation

4.1 Introduction

The beneficial role of public infrastructure on economic activity has been well debated in the literature. Aschauer (1989a) provided an initial estimate of 0.39 for the output elasticity of public capital. Later studies however, criticized this apparently high elasticity on several grounds. Gramlich (1994) argued that a 0.39 elasticity estimate implies a more than 100 percent marginal product of public capital. Since in theory, it is hard to refute the hypothesis that public investment fosters economic growth, studies have tried to incorporate different factors in order to bring down the large effect to a more plausible figure. Bom and Ligthart (2014) provides an excellent literature review and performing a meta analysis, they find a conditional short run elasticity estimate of 0.085 when controlled for heterogeneity across studies. The theoretical literature, on the other hand, was mainly propounded by Barro (1990) that formalized the role of public investment in an endogenous growth framework. Evidently, Aschauer's empirical study, coupled with Barro's theoretical approach, opened a new avenue of research in the growth literature. Other notable mentions are: Glomm and Ravikumar (1997b), Devarajan et al. (1996) and, Futagami et al. (1993). The treatment of public capital in these models has also evolved over the years. While most studies adhere to Barro's specification of a public capital flow that enters the production function as an input, studies like Futagami et al. (1993), Turnovsky (1997), and more recently Turnovsky (2004) and Chatterjee and Turnovsky (2007) formalize the concept of public capital stock that has its own accumulation process. The conceptual notion of public capital stock envisages an important role to maintenance of the existing stock owing to depreciation. The role of maintenance of public infrastructure was highlighted by the World Bank (1994) study, which said *"timely maintenance expenditure of \$12 billion would have saved road constructions costs of \$45 billion in Africa"*. The study also claimed that curbing maintenance expenditure in times of budgetary austerity is a wrong policy as high costs have to be incurred later for rehabilitation and reconstruction. Data from different countries also provide evidence that maintenance expenditures comprise a major part of the total spending on infrastructure. Figure 4.1 shows the shares of maintenance expenditures and new investment in total spending on road infrastructure for select countries. It is evident that countries that spend nearly 1 percent of their GDP on an average on road infrastructure, allocate nearly 35 percent of the total spending on average to road maintenance and the rest 65 percent to building new road infrastructure. Interestingly, countries like Denmark and India devote more than 50 percent of the total spending on roads towards maintenance of existing roads.

The role of maintenance expenditure on economic growth, however, lacked a formal theoretical exposition until very recently. Rioja (2003) formally introduced the concept of maintenance by endogenizing the depreciation rate (as a function of maintenance expenditure) and showed that reallocating funds from new investment to maintenance has a positive effect on economic growth. However, in Rioja's model maintenance expenditure is funded by tax revenues while new investment in infrastructure is funded by foreign aid. This assumption precluded a discussion on optimal allocation of tax revenues between maintenance and new investment. Kalaitzidakis and Kalyvitis (2004) take into account government budgetary constraint and derive optimal allocations toward maintenance and new investments. But their model departs from household optimizing behavior ignoring the possibility of welfare loss through higher taxation. Dioikitopoulos and Kalyvitis (2008) expound on the implication of public capital on the trade-off between maintenance expenditure and new investments. They use an endogenous growth model to show that countries facing low congestion in public infrastructure would require a threshold level of maintenance expenditure to experience a balanced growth in output. On the other hand, countries facing high congestion would require a threshold level of new investments for balanced growth. Agenor (2009) departs from the previous studies and models maintenance expenditure through an additional efficiency parameter keeping intact the concept of endogenous depreciation rate as in Rioja (2003). He shows that the growth-maximizing tax rate in a decentralized economy is equal to the output elasticity of public capital as in the Barro model. The welfare-maximizing share of spending on maintenance is shown to be identical to the growth-maximizing share when the tax rate is set at the level implied by the Barro rule. There are only a handful of empirical studies that test the role of maintenance expenditures on economic activity. Owing to unavailability of cross country data on maintenance expenditures, Kalyvitis and Vella (2015) asses the productivity effects of infrastructure's operations and maintenance spending by 48 U.S states over the period 1978-2000. They find a positive and significant output elasticity of maintenance expenditures when controlling for cross-state spillovers. Kalaitzidakis and Kalyvitis (2005) use Canadian data and show that the Canadian economy would benefit by altering the allocation between maintenance expenditures and new investments in public infrastructure.

The studies cited above however fail to provide a perspective on developing countries. An important feature of a developing country is the existence of a large informal sector and its equally large contribution to the GDP. In India for example, the unregistered and unincorporated small production units that usually do not fall under the tax administration, contribute almost 55 percent to the GDP. Figure 4.2 shows how the informal sector in India has evolved during the period 2004-2011, along with the spending patterns on road infrastructure during the same period. As can be seen from the figure, there has been a slight decrease in the contribution of the informal sector to the GDP, while the share of maintenance spending in total spending on road infrastructure has increased during the period 2008-2011. This paper tries to incorporate the informal sector into a two-sector endogenous growth model and examine it's implication on optimal fiscal policies. The objective of this paper is to draw some inference regarding the optimal tax rate, and the optimal allocation of tax revenues toward maintenance and new investments for a country characterized by a large informal sector. This is the main contribution of this study.

But what role does the informal sector play in determining the optimal tax rate and the optimal allocation of spending toward maintenance of public capital? The mechanism through which the informal sector affects the provision of infrastructure and maintenance spending is through congestion and efficiency. Since the informal sector firms usually do not fall under the tax administration, the government's main source of tax revenues is the formal sector. Tax revenues that are collected from the formal sector are then spent on infrastructure (like roads, railways, and power). Spending on infrastructure has two components: new investments and maintenance. Congestion of public infrastructure stems from the non-excludable nature of public infrastructure that can be used by the formal sector and informal sector in their production processes. For example, the benefits of building a new road (or the benefits of maintaining a good condition of roads) accrue to both formal and informal productions units as none of them can be excluded from using it. But financing new investments and maintenance of existing infrastructure are however borne out of the tax revenues that are collected from the formal sector. Put differently, the informal sector poses a free-rider problem by using public infrastructure that is financed by taxes levied on the formal sector. Figure 4.3 provides cross-country evidence of a positive correlation between the size of the informal sector (using the share of informal sector employment in the total labor force as a proxy) and power outages faced by a firm in a typical month. Figure 4.4 shows a similar correlation between the size of the informal sector and the value lost due to electrical outages. Both these figures insinuate to some degree of congestion faced by economies with large informal sectors. This paper contributes to the literature on public capital maintenance and economic growth by incorporating congestion effects posed by the informal sector in developing countries.

An implication of such a congestion effect is the loss of efficiency of public infrastructure which in turn has a negative effect on overall production. In addition to this, corrupt practices, inefficient appraisal processes, and fund disbursal systems also make public investment inefficient in developing countries. So, to contend that higher public investment has a positive impact on output and productivity would be wrong if not controlled for inefficiency of public investments. An interesting study by Hulten (1996) showed that a large portion of differential growth rate between Africa and East Asia can be explained by the difference in effective use of infrastructure capital in the two regions. Pritchett (2000) argued on methodological grounds that not all of public investments translate into capital which is quite pertinent to developing countries. Dabla-Norris et al. (2012) expounded on this idea and created the Public Investment Management Index (PIMI). Using the PIMI, Gupta et al. (2014) produced cross country estimates for efficiency adjusted public capital. The estimated factor share of efficiency adjusted public capital was in the range of 0.143 to 0.158. Figure 4.5 shows a negative correlation between the size of the informal sector and the PIMI scores that underscores the fact that large informal sectors reduce the efficiency of public infrastructure. Although these studies prove in a way that inefficiency in public capital is a major hindrance to output and productivity, few studies have looked into the theoretical aspects of efficiency of public capital. Chakraborty and Dabla-Norris (2009) made a serious effort in formalizing a model with the efficiency parameter. They argue that simply increasing public investment could be highly inefficient in low income countries where effort must be put in to maintain quality of public investment. Agenor (2009) also models the maintenance expenditure through an additional efficiency parameter keeping intact the concept of endogenous depreciation rate as in Rioja (2003).[53] This paper extends Agenor (2009) framework by defining the efficiency parameter as a function of the share of maintenance expenditure in total spending, taking into account the congestion effect of public capital due to the informal sector. In my model, the efficiency of public infrastructure increases with the share of maintenance expenditure and decreases with the size of the informal sector relative to formal sector. This is the second major contribution of this paper.

Results in the paper show that the growth-maximizing tax rate is a function of the output elasticities of public infrastructure (for formal and the informal sectors), and the efficiency parameter which is exogenously given in the decentralized economy. This tax rate is however lower than the Barro (1990) tax rate which is equal to the output elasticity of public infrastructure in a one-sector growth model. The growth-maximizing share of maintenance expenditure is shown to be a function of the production elasticities, the efficiency elasticity of maintenance expenditure, and the responsiveness of depreciation rate to maintenance spending. The welfare maximizing tax rate and the share of maintenance are not separately identified in a centrally planned economy. But if the social planner imposes a tax rate that maximizes the decentralized growth rate, the welfare-maximizing share of maintenance share is shown to be positively related to the ratio of informal to formal sector output. Thus economies with large informal sectors would benefit by devoting more resources toward maintenance of existing public infrastructure. This paper, by deriving the optimal tax rate and optimal share of maintenance would help the policy makers in developing countries to optimally spend resources on public infrastructure.

[53] Loayza (1996)incorporated the informal sector into the Barro type model and showed that the optimal taxation should be less than that in the Barro model in the presence of the informal sector, but fails to recognize the efficiency and maintenance effects of the informal sector. Penalosa and Turnovsky (2005) analyzed the burden of taxation on labor and capital when one of the sectors in the economy cannot be taxed but does not take into account the effects of over utilization of public capital on efficiency and maintenance expenditure.

The rest of the paper is organized as follows. Section 4.2 lays out the analytical framework. Section 4.3 derives the welfare-maximizing equilibrium for a centrally planned economy. Section 4.4 describes the results for a decentralized economy. Section 4.5 discusses an optimal fiscal policy possibility for the social planner. Section 4.6 concludes.

4.2 Analytical Framework

We consider a closed economy with two sectors: formal and informal. The formal sector produces relatively more capital intensive goods as compared to the informal sector. We would assume the two sectors are populated by different set of individuals who consume either the formal sector good or the informal sector good but not both. In other words, we are assuming that the two markets function independently of each other. Although it seems a restrictive assumption, but it is useful to think the informal sector being populated by low-income individuals who cannot afford to buy high-quality capital intensive goods that are relatively expensive than the informal sector goods. The high income individuals, however, can afford to buy informal sector goods. But owing to the low-quality of these goods, the representative agent in the formal sector has a strict preference for the formal sector good over the informal sector good. The formal sector representative agent consumes C_F of the formal sector good. The informal sector representative agent consumes C_I of the informal sector good. Henceforth, formal and informal sectors are indexed by F and I respectively. The utility functions for the formal and informal representative agents are given by:

$$U_F = \int_0^{\infty} [\frac{1}{\gamma} C_F^{\gamma}]\, e^{-\rho t} dt, \ -\infty < \gamma \leq 1 \tag{4.1}$$

$$U_I = \int_0^{\infty} [\frac{1}{\gamma} C_I^{\gamma}]\, e^{-\rho t} dt, \ -\infty < \gamma \leq 1 \tag{4.2}$$

Each agent in the formal sector produces a private good whose output is given by Y_F, using a Cobb-Douglas production function. Private capital (K_F) and the economy-wide stock of effective public capital (eK_G) serve as the factors of production. The actual stock of public capital is adjusted with an efficiency parameter e (discussed below) to arrive at an effective stock of public capital. The informal sector is characterized by a similar technology. The production functions of the formal and informal sectors are given by:

$$Y_F = (eK_G)^{\alpha} K_F^{1-\alpha} \tag{4.3}$$

$$Y_I = (eK_G)^{\beta} K_I^{1-\beta} \tag{4.4}$$

Note that the input elasticities (α and β) are different in the formal and informal sectors. If we assume that formal sector firms are relatively more capital intensive than the informal sector firms then $1-\alpha > 1-\beta$. But since this assumption does not have any consequence on the analytical framework, we abstain from making such *a priori* assumption. It is important however, to contend different output elasticities of public capital for the formal and informal sectors. Since public capital is indivisible and non-excludable in nature, it is available to both the sectors. But the rival nature of public capital makes the degree of accessibility of the public capital different for the two sectors as captured by α and β, where $\alpha \neq \beta$. This is an important distinction of this study from the specification used by Loayza (1996) where it is assumed that the informal sector uses a fraction of public capital. It is this fact that drives congestion in public goods. Private capital depreciates at the rate $\delta_i \in (0, 1)$. We would assume the depreciation rate as exogenously given for the two sectors. If I_i is investment, then the evolution of private capital is given by:

$$\dot{K}_i = I_i - \delta_i K_i,\ i \in (F,\, I) \tag{4.5}$$

Government

The government provides public capital, the evolution of which is given by :

$$\dot{K_G} = I_G - \delta_G K_G \tag{4.6a}$$

where I_G is investment in public capital and δ_G is the depreciation rate. As in Rioja (2003), we assume the depreciation to be negatively related to maintenance expenditure. Additionally, we also assume that depreciation is directly proportional to the stock of public capital (which is used as a scaling factor as in Agenor (2009). For simplicity we assume the functional form to be linear given by:

$$\delta_G = \bar{\delta_G} - \theta_G.(\frac{M}{K_G})\,,\, \theta_G \in (0,1) \tag{4.6b}$$

Equation (4.6b) implies that when $M = 0$, $\delta_G = \bar{\delta_G}$, where $\bar{\delta_G}$ denotes the maximum depreciation rate. Each period the government invests I_G in public capital and allocates some amount M on maintenance which is funded from tax revenues collected from the formal sector. The informal sector does not pay taxes. Let τ be the output tax rate. Let v_g and v_m be the proportion of taxes that are allocated to new investment in public capital and maintenance respectively. Thus we have,

$$I_G = v_g(\tau Y_F) \tag{4.6c}$$

$$M = v_m(\tau Y_F) \tag{4.6d}$$

The government budget constraint is given by :

$$I_G + M = \tau Y_F \Rightarrow v_g + v_m = 1 \tag{4.6e}$$

Combining (4.6a)-(4.6e) we have,

$$\frac{\dot{K_G}}{K_G} = v_g(\tau Y_F/K_g) + \theta_G(\frac{v_m \tau Y_F}{K_G}) - \bar{\delta_G} \tag{4.6f}$$

Effective Public Capital Stock

The public capital stock may be non-excludable but its efficiency is a function of maintenance expenditure and the size of the informal sector. Specifically, the efficiency of public capital increases with higher maintenance expenditure and decreases with the size of the informal sector. The negative relationship between public capital efficiency and the size of the informal sector stems from the fact that both maintenance and new investment on public capital are financed by taxes levied on the formal sector. The informal sector thus poses a free rider problem that negatively affects the efficiency of public infrastructure. Because of this over utilization, efficiency of public capital goes down. We consider a concave function for the efficiency parameter given by:

$$e = (\frac{M}{Y_I})^{\chi},\ \chi \in (0,1) \tag{4.7a}$$

where χ measures the elasticity of maintenance with respect to efficiency. Substituting for $M = \upsilon_m(\tau Y_F)$ in (4.7a), the efficiency parameter can be characterized by a function:

$$e = f(\chi, \tau, \upsilon_m, s_F)\,;\ s_F = Y_F/Y_I. \tag{4.7b}$$

where $e_\chi < 0$; $e_\tau > 0$; $e_{\upsilon_m} > 0$; 0; $e_{s_F} > 0$. Thus an increase in tax rate increases the efficiency of the existing public capital as more resources become available for maintenance. A reallocation of resources toward maintenance as captured by υ_m, increases efficiency of public capital by increasing overall maintenance expenditure. As the size of the informal sector increases relative to the formal sector, efficiency of public capital decreases. Lastly, efficiency decreases with an increase in the efficiency elasticity of maintenance (as $0 < \frac{M}{Y_I} < 1$). A closed form solution for the efficiency parameter (as a function of the inputs to the production functions) is derived in Appendix 3. Substituting for the efficiency parameter, the production functions can be represented as:

$$Y_F = (\upsilon_m \tau)^{\chi\alpha/\eta}(K_G)^{\epsilon_1}(K_F)^{\epsilon_2}(K_I)^{\epsilon_3} \tag{4.8}$$

$$Y_I = (\upsilon_m \tau)^{\chi\beta/\eta}(K_G)^{\epsilon_4}(K_F)^{\epsilon_5}(K_I)^{\epsilon_6} \tag{4.9}$$

4.3 A Centrally Planned Economy

The social planner in the centrally planned economy internalizes the congestion *ex ante* by optimally choosing the tax rate and the share of maintenance expenditure in total spending. In other words, the planner optimizes the efficiency of public capital by choosing the policy instruments at his disposal. Additionally, the planner faces the market clearing condition given by equation (4.10a). He also faces the capital accumulation constraint for the economy as a whole given by (4.10b).

Market clearing condition:

$$Y_F + Y_I = C_F + C_I + I_F + I_I + I_G + M \tag{4.10a}$$

$$\Rightarrow (1 - \tau v_m)Y_F + Y_I - C_F - C_I = I_F + I_I + I_G$$

Capital accumulation constraint:

$$\dot{K_F} + \dot{K_I} + \dot{K_G} = (1 - \tau v_m)Y_F + Y_I - C_F - C_I - \delta_F K_F - \delta_I K_I - \bar{\delta}_G K_G + \tau v_m \theta_G Y_F \tag{4.10b}$$

The planner's utility function takes the form:

$$\underset{C_I, C_F}{max}\, U = \int_0^\infty \frac{1}{\gamma}[C_F^\gamma + C_I^\gamma]\, e^{-\rho t} dt \tag{4.10c}$$

The planner makes the resource allocation decision for the representative agents by choosing consumption (C_F, C_I), tax rate (τ), share of maintenance expenditure in total spending (υ_m), and the accumulation of private capital in the two sectors (K_F, K_I) and public capital (K_G) by maximizing (4.10c) subject to the resource constraint given by (4.10b). The optimality conditions are given by the following equations.

$$C_F^{\gamma-1} - \lambda = 0 \tag{4.11a}$$

$$C_I^{\gamma-1} - \lambda = 0 \tag{4.11b}$$

$$[-\upsilon_m + \theta_G \upsilon_m] Y_F \tau + \frac{\alpha\chi}{\eta}(1 - \tau\upsilon_m + \tau\upsilon_m\theta_G)Y_F + \frac{\beta\chi}{\eta}Y_I = 0 \tag{4.11c}$$

$$[-\tau + \theta_G \tau] Y_F \upsilon_m + \frac{\alpha\chi}{\eta}(1 - \tau\upsilon_m + \tau\upsilon_m\theta_G)Y_F + \frac{\beta\chi}{\eta}Y_I = 0 \tag{4.11d}$$

$$(1 - \tau\upsilon_m + \tau\upsilon_m\theta_G)\epsilon_2\frac{Y_F}{K_F} + \epsilon_5\frac{Y_I}{K_F} - \delta_F = \rho - \frac{\dot{\lambda}}{\lambda} \tag{4.11e}$$

$$(1 - \tau\upsilon_m + \tau\upsilon_m\theta_G)\epsilon_3\frac{Y_F}{K_I} + \epsilon_6\frac{Y_I}{K_I} - \delta_I = \rho - \frac{\dot{\lambda}}{\lambda} \tag{4.11f}$$

$$(1 - \tau\upsilon_m + \tau\upsilon_m\theta_G)\epsilon_1\frac{Y_F}{K_G} + \epsilon_4\frac{Y_I}{K_G} - \bar{\delta_G} = \rho - \frac{\dot{\lambda}}{\lambda} \tag{4.11g}$$

where ϵ_1, ϵ_2, ϵ_3, ϵ_4, ϵ_5, ϵ_6 are the parameters from the production functions.[54] The optimality conditions (4.11a)-(4.11g) can be interpreted as follows. Equations 4.11a and 4.11b equate the marginal utility of consumption for the formal and informal sectors to the shadow price of private capital. Equations (4.11c) and (4.11d) are the optimality conditions with respect to tax rate and the share of maintenance expenditure respectively. Equations (4.11e) and (4.11f) equate the rate of return of private capital in the formal and informal sectors to the corresponding return on

[54]See Appendix 3

consumption in the two sectors. Equation (4.11g) equates the rate of return on public capital to the corresponding return on consumption. The equilibrium growth rate ($\frac{\dot{C_F}}{C_F} = \frac{\dot{C_I}}{C_I} = \frac{\dot{K_F}}{K_F} = \frac{\dot{K_I}}{K_I} = \frac{\dot{K_G}}{K_G}$), which has many equivalent forms, can be written as:

$$\psi^* = \frac{(1 - \tau v_m + \tau v_m \theta_G)\epsilon_2 \frac{Y_F}{K_F} + \epsilon_5 \frac{Y_I}{K_F} - \delta_F - \rho}{1 - \gamma} \tag{4.12}$$

We solve equations (4.11c) and (4.11d) to find the optimal tax rate and the share of maintenance. The social planner however cannot identify the two instruments separately. So, the product of optimal maintenance and tax rate is given by:

$$(\tau v_m)^{**} = \frac{\beta\chi(Y_I/Y_F)+}{(1-\theta_G)(1+\beta\chi)} + \frac{\alpha\chi}{(1-\theta_G)(1+\beta\chi)} \tag{4.13}$$

Proposition 1. *Under a centrally planned economy, the optimal tax rate and the share of maintenance expenditure in total spending are not separately identified. The government can arbitrary set either the tax rate or the share of maintenance spending such that (4.13) always satisfies. The product τv_m is however identified which is a function of: (i) the ratio of informal to formal sector output (Y_I/Y_F), (ii) the output elasticities of public capital in the two sectors (α, β),(iii) the elasticity of maintenance expenditure with respect to public capital efficiency to the formal sector output (χ) and, (iv) the effectiveness of maintenance expenditure to depreciation rate (θ_G).*[55]

4.4 A Decentralized Economy

In this section we consider a decentralized economy where the government assumes a passive role whereas the representative agents in the two sectors make their own resource allocation decisions. In this set up, the government provides the entire stock of public capital using the policy instruments at its disposal: the tax rate and the share of maintenance expenditure in total spending. The representative agents take the stock of public capital as exogenously given when making their own

[55] The comparative statics discussion will be part of the optimal fiscal policy discussion in Section 4.5 where we derive a specific solution for the welfare-maximizing v_m.

resource allocation decisions with respect to consumption and accumulation of private capital stock. Additionally, the representative agents do not internalize the sources of congestion. They assume efficiency of public capital as exogenously given and they have no influence over it. Since there are two production sectors we have two different optimization problems: one for the formal and one for the informal sector. The utility functions are given as in (4.1) and (4.2).

Formal Sector

Household problem is characterized by :

$$\max_{C_F} U = \int_0^\infty \frac{1}{\gamma} C_F^\gamma e^{-\rho t} dt$$

subject to the budget constraint,

$$C_F + I_F = (1-\tau)Y_F$$

and

$$\dot{K_F} = I_F - \delta_F K_F$$

The household solves the optimization problem taking as given the depreciation rate of the private capital, the tax rate, the discount rate and the effective stock of public capital. Equations (4.14a) and (4.14b) describe the optimality conditions for the formal sector:

$$C_F^{\gamma-1} - \lambda = 0 \tag{4.14a}$$

$$\lambda[(1-\tau)(1-\alpha)(eK_{GF})^\alpha - \delta_F] = \rho\lambda - \dot{\lambda} \tag{4.14b}$$

Equation (4.14a) equates the marginal utility of consumption to the shadow price of private capital. Equation (4.14b) equates the rate of return on private capital in the formal sector to the corresponding return on consumption.

Combining the first order conditions and using (4.8), we find the balanced growth path ($\frac{\dot{C_F}}{C_F} = \frac{\dot{K_F}}{K_F}$) as:

$$\psi_F = \frac{(1-\tau)(1-\alpha)(v_m\tau)^{\chi\alpha/\eta}(K_G)^{\epsilon_1}(K_F)^{\epsilon_2-1}(K_I)^{\epsilon_3} - \delta_F - \rho}{1-\gamma} \tag{4.15}$$

The growth path ($\frac{\dot{C_F}}{C_F} = \frac{\dot{K_F}}{K_F}$) can also be represented by the growth path of the stationary variable $c_F = \frac{C_F}{K_F}$, where

$$\frac{\dot{c_F}}{c_F} = \frac{(\gamma-\alpha)(1-\tau)(v_m\tau)^{\chi\alpha/\eta}(K_G)^{\epsilon_1}(K_F)^{\epsilon_2-1}(K_I)^{\epsilon_3} - \gamma\delta_F + (1-\gamma)c_F - \rho}{1-\gamma} \tag{4.16}$$

where $c_F = \frac{C_F}{K_F}$.

It is interesting to note that output growth rate in the formal sector is a function of public capital (K_G), the informal sector private capital (K_I) and its own private capital (K_F).

Informal Sector

We proceed with the informal sector optimization problem in the same way as the formal sector. The only difference here is that the informal sector does not pay taxes, so the budget constraint looks different. Assuming the same utility function, the optimization problem for the informal sector is given by:

$$\underset{C_I}{max} U = \int_0^\infty \frac{1}{\gamma} C_I^\gamma e^{-\rho t} dt$$

subject to the budget constraint,

$$C_I + I_I = Y_I$$

and

$$\dot{K_I} = I_I - \delta_I K_I$$

The informal sector household solves the optimization problem taking as given the depreciation rate of the private capital, the discount rate and the effective stock of public capital. Equations (4.17a) and (4.17b) describe the optimality conditions for the formal sector:

$$C_I^{\gamma-1} - \lambda = 0 \tag{4.17a}$$

$$\lambda[(1-\beta)(eK_{GI})^{\beta} - \delta_I] = \rho\lambda - \dot{\lambda} \tag{4.17b}$$

The optimality conditions (4.17a) and (4.17b) have the same interpretations as above. Combining the first order conditions and using (4.9), we find the balanced growth path ($\frac{\dot{C}_I}{C_I} = \frac{\dot{K}_I}{K_I}$) as:

$$\psi_I = \frac{(1-\beta)(v_m\tau)^{\chi\beta/\eta}(K_G)^{\epsilon_4}(K_F)^{\epsilon_5}(K_I)^{\epsilon_6-1} - \delta_I - \rho}{1-\gamma} \tag{4.18}$$

It is interesting to note the endogeneity of the growth process for both the formal and the informal sectors. Both sector's growth rates are functions of the stock of formal and informal private capital, and the stock of effective public capital. The informal sector growth is also a function of the tax rate and the fraction of tax revenue allocated to maintenance. This may seem paradoxical at first but it actually makes intuitive sense. Since, maintenance expenditure increases efficiency of public capital, the informal sector output increases for any given level of public capital stock. Basically, the informal sector also reaps the benefits of higher maintenance expenditure on public capital. The growth path ($\frac{\dot{C}_I}{C_I} = \frac{\dot{K}_I}{K_I}$) can also be represented by the growth path of the stationary variable $c_I = \frac{C_I}{K_I}$. Thus combining equations (4.17b) and (4.18) we have,

$$\frac{\dot{c_I}}{c_I} = \frac{(-\beta)(v_m\tau)^{\chi\beta/\eta}(K_G)^{\epsilon_4}(K_F)^{\epsilon_5}(K_I)^{\epsilon_6-1} - \gamma\delta_I + (1-\gamma)c_I - \rho}{1-\gamma} \tag{4.19}$$

Using (4.6f), (4.8) and (4.9) we also have,

$$\frac{\dot{K_G}}{K_G} = \tau(v_G + \theta_G v_m)(v_m\tau)^{\chi\alpha/\eta}(K_G)^{\epsilon_1-1}(K_F)^{\epsilon_2}(K_I)^{\epsilon_3} \tag{4.20}$$

4.4.1 Growth Maximizing Tax Rate: Decentralized Economy

The objective is to find a tax rate that maximizes the growth rate in the decentralized economy given by:

$$\psi_F = \frac{(1-\tau)(1-\alpha)(v_m\tau)^{\chi\alpha/\eta}(K_G)^{\epsilon_1}(K_F)^{\epsilon_2-1}(K_I)^{\epsilon_3} - \delta_F - \rho}{1-\gamma} \tag{4.15}$$

which can also be characterized by,

$$\psi_F = \tau(v_G + \theta_G v_m)(v_m\tau)^{\chi\alpha/\eta}(K_G)^{\epsilon_1-1}(K_F)^{\epsilon_2}(K_I)^{\epsilon_3} \tag{4.20}$$

Differentiating (4.15) w.r.t τ and setting it to 0 we have,

$$\tau^* = \frac{\alpha\chi}{1+\beta\chi} \tag{4.21}$$

The growth maximizing tax rate in the decentralized economy is thus a function of the public capital elasticities for the formal and the informal sectors and the exogenously given efficiency elasticity of maintenance expenditure (χ). It can be readily seen that when there is no informal sector (or to be less restrictive: if the informal sector does not use public capital in its production process), then the parameter $\beta = 0$. Imposing this condition the optimal tax rate reduces to $\alpha\chi$, which is similar to the Barro model prediction ($\tau = \alpha$). On the other hand, as public capital becomes more productive in the informal sector it is optimal for the government to lower the tax rate. Intuitively, public capital is financed by taxing the formal sector which is freely used by the informal sector. The formal sector would get discouraged by higher taxes levied on them, which would have a negative effect on the output and growth.

4.4.2 Growth Maximizing Maintenance: Decentralized Economy

In this section we look into the optimal allocation of tax revenues to maintenance in the presence of an informal sector. The objective is to find that allocation of maintenance out of tax revenues that maximizes the decentralized growth rate. Differentiating (4.20) (which is the optimal growth path for the economy) with respect to v_m and imposing the condition that $dv_m = -dv_g$ and $d\tau = 0$,which means the government balances its budget, we have,

$$v_m^* = (\frac{\alpha\chi}{1-\theta_G})(\frac{1}{1+\beta\chi}) \tag{4.22}$$

Agenor (2009) has only one production sector in the model, so the optimal maintenance is just the first expression in the parentheses which can be readily derived by setting $\beta = 0$. But in the presence of the informal sector, the allocation towards maintenance is smaller because $1 + \beta\chi > 1$. This makes intuitive sense because the presence of the informal sector means greater utilization of public capital which is actually financed by the taxes levied on the formal sector. Because of the illegal utilization of public capital by the informal sector, the government is tempted to spend less on maintenance and spend higher in new investments instead. The growth-maximizing share of maintenance is also positively related to the responsiveness of depreciation rate to maintenance (θ_G). Thus greater is the response of maintenance spending on the depreciation rate, more resources must be allocated to maintenance. The maintenance share is also negatively related to the exogenously given efficiency elasticity of maintenance (χ).

4.5 Optimal Fiscal Policy

In Section 4.3, we showed that the social planner cannot separately implement welfare maximizing tax rate and the share of maintenance expenditure on total spending. This means that the planner can arbitrarily choose any one of the policy instruments at his disposal such that equation (4.13) holds. It is instructive to ask however, which arbitrary level of τ or v_m the planner should choose. Since, taxes are imposed on the formal sector output, let us say that the one objective of the social planner is to maximize the formal sector output in the decentralized economy. Under such an objective, the planner can implement the decentralized growth maximizing tax rate and choose v_m such that equation (4.13) holds. Under such a condition, the planner achieves the welfare-maximizing share of maintenance expenditure in total spending.

Proposition 2. *If the social planner arbitrary sets the tax rate that maximizes the formal sector growth, the welfare-maximizing share of maintenance expenditure (v_m^{**}) in total spending is given by:*

$$v_m^{**} = \frac{1}{(1-\theta_G)}[Y_I/Y_F\,(\beta/\alpha)+1] \tag{4.23}$$

Some important predictions that emerge from this result are as follows. First, it is interesting to note that the welfare-maximizing share of maintenance is positively related to the ratio of informal to formal sector output. Thus as the size of the informal sector increases relative to the formal sector, the social planner should increase the share of maintenance in order to maximize overall welfare. The mechanism through which the informal-formal output ratio affects the optimal share of maintenance is through the efficiency parameter. An increase in the Y_I/Y_F decreases the efficiency of public capital. Since the social planner internalizes the efficiency parameter, he tries to maintain the same level of efficiency by devoting more resources toward maintenance. This is a major departure from the decentralized solution since in that case the grow-maximizing share of maintenance is independent of the relative size of the informal sector. Second, higher the response of depreciation rate to maintenance expenditure (higher θ_G), more resources should to be devoted to maintenance in raising the stock of public capital. The optimal share of maintenance is independent of the efficiency elasticity of maintenance (χ) because the social planner internalizes the efficiency parameter in the decision making process.

4.6 Conclusions

Efficient public infrastructure is an important part of the growth process for all advanced and developing economies. However, a fundamental difference between an advanced country and a developing country is the existence of a large informal sector in the latter. In developing countries, the informal sector contributes to almost half of the country's GDP. This sector is characterized by low-productivity unincorporated firms that operate at a low level of capital intensity relative to the formal sector firms. A large fraction of the population is dependent on the informal sector as it produces cheap low quality goods. The existence of such a sector has important implications on fiscal policies. First, the informal sector does not pay taxes which imply less amount of resources that can be devoted to the provision of public goods. Second, public infrastructure which is a non-excludable good, is used by the formal and informal sectors as an input in their production processes. Since public infrastructure is funded by taxes imposed on the formal sector, informal sector's use of public infrastructure introduces some degree of congestion with regards to public infrastructure. We set up a two-sector endogenous growth model to examine the implication of the informal sector on optimal fiscal policies.

Since congestion affects the efficiency of public infrastructure, in order to make public infrastructure efficient, the government must devote some resources to maintain the existing stock of public infrastructure. This paper throws some light on how tax revenues should be spent on investments on new infrastructure and maintenance of existing public infrastructure. We find that under a centrally planned economy, the social planner may choose an arbitrary level of tax rate and implement a welfare maximizing share of maintenance expenditure in total spending. The tax rate however may be chosen to maximize the decentralized growth rate. The results shown in this paper have important policy implications. First, it is shown that the welfare maximizing share of maintenance spending in total spending on infrastructure is positively related to the ratio of informal to formal sector output. Thus economies with large informal sectors must devote more resources toward maintaining public infrastructure. Second, we find that both welfare maximizing and decentralized fiscal policies are functions of the parameters of the formal and informal production processes. Overall, these results show a non-trivial effect of the informal sector on the dynamics of fiscal policies. Thus, policy makers in developing countries must take these into consideration when implementing fiscal polices in order to achieve their desired outcomes.

Figures

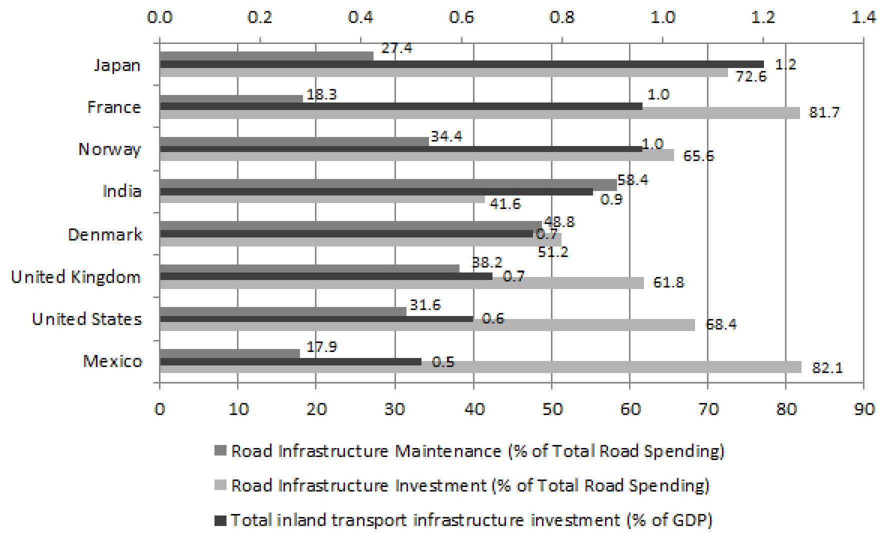
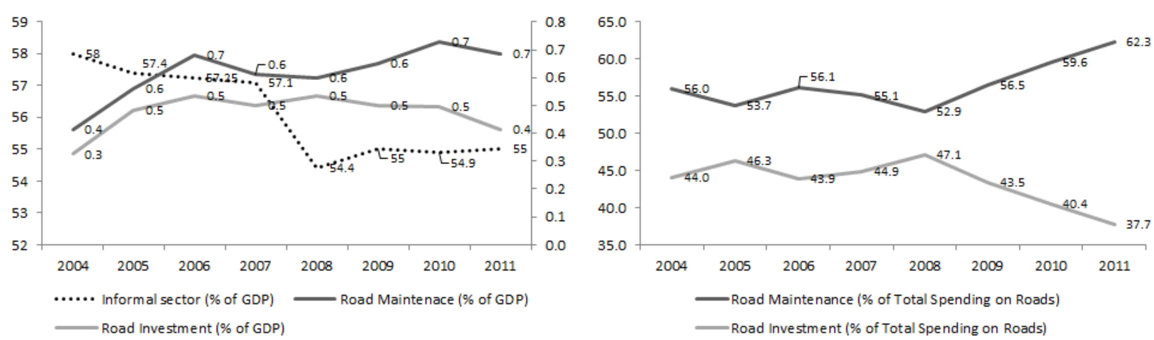


Figure 4.1: Road Infrastructure Investment and Maintenance (% of Total spending), Average 2006-2010

Source: International Transport Forum (ITF) Outlook, 2013



(a) Informal sector, Road Investment and, Maintenance (% of GDP)

(b) Road Investment and Maintenance (% of Total Spending)

Figure 4.2: Road Investment, Maintenance and Informal Sector, India, 2004-2011

Source: ITF Outlook, 2013 and National Accounts Statistics Reports (India)

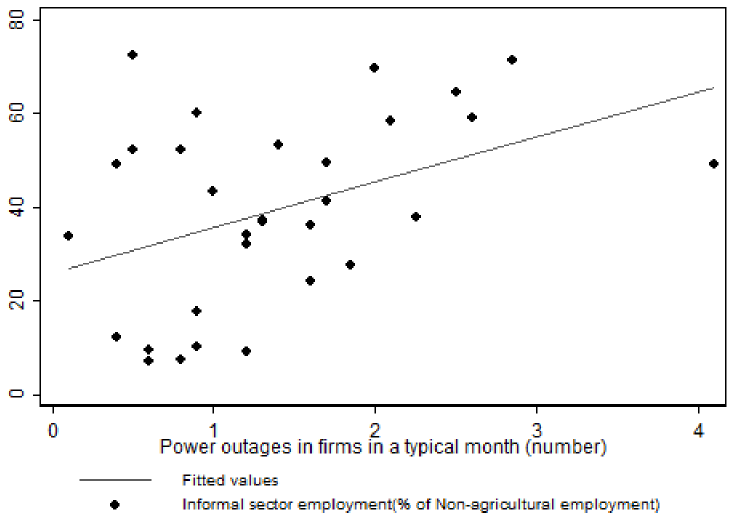


Figure 4.3: Informal Sector Employment (% of Non-Agricultural Employment) vs No. of Power Outages in a Firm in a Typical Month

Source: ILO (2012) and World Development Indicators

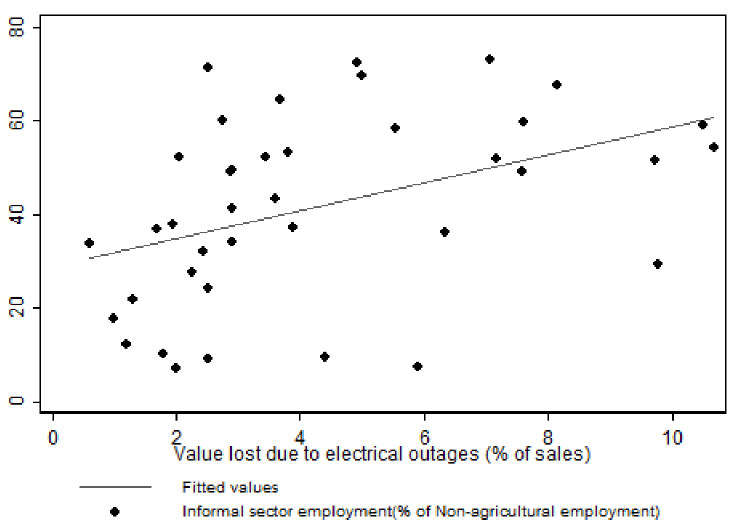


Figure 4.4: Informal Sector Employment (% of Non-Agricultural Employment) vs Value Lost Due to Electrical Outages (% of Sales)

Source: ILO (2012) and World Development Indicators

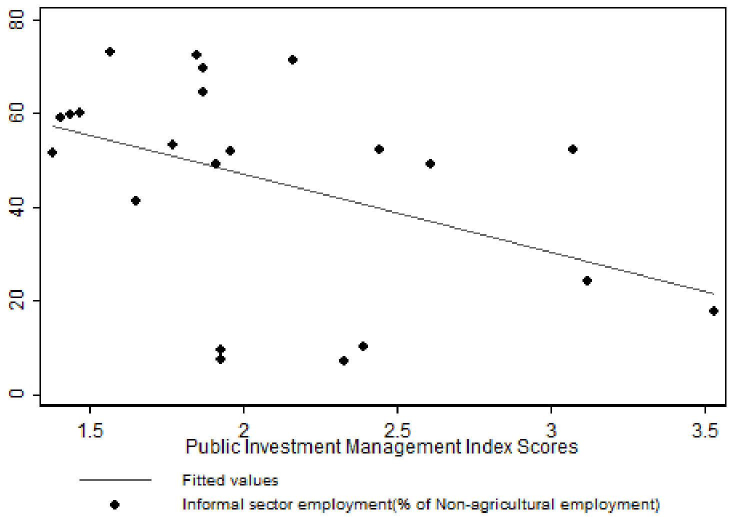


Figure 4.5: Informal Sector Employment (% of Non-Agricultural Employment) vs PIMI scores, cross country

Source: ILO (2012) and World Development Indicators

Appendix 3

Characterizing the Efficiency Parameter

From (4.7a) and (4.6d) we note that,

$$e = (\frac{v_m \tau Y_F}{Y_I})^{\chi}$$

$$\rightarrow e = (\frac{v_m \tau Y_F / K_G}{Y_I / K_G})^{\chi} \tag{A.3.1}$$

From the production functions (4.3) and (4.4) we have,

$$\frac{Y_F}{K_G} = e^{\alpha} k_{GF}^{\alpha-1}$$

and

$$\frac{Y_I}{K_G} = e^{\beta} k_{GI}^{\beta-1}$$

where $k_{GF} = \frac{K_G}{K_F}$ and $k_{GI} = \frac{K_G}{K_I}$. Substituting these into (A.3.1) we have,

$$e = (v_m \tau)^{\chi} (e^{\alpha} k_{GF}^{\alpha-1})^{\chi} (e^{-\beta} k_{GI}^{1-\beta})^{\chi}$$

Solving for e we have,

$$e = (v_m \tau)^{\chi/\eta} (k_{GF})^{\chi(\alpha-1)/\eta} (k_{GI})^{\chi(1-\beta)/\eta} \tag{A.3.2}$$

$$e^{\alpha} = (v_m \tau)^{\chi\alpha/\eta} (K_G)^{\chi\alpha(\alpha-\beta)/\eta} (K_F)^{\chi\alpha(1-\alpha)/\eta} (K_I)^{\chi\alpha(\beta-1)/\eta} \tag{A.3.3}$$

where $\eta \equiv 1 - \chi(\alpha - \beta)$.

Substituting for e, the production functions can be rewritten as:

$$Y_F = (v_m \tau)^{\chi\alpha/\eta}(K_G)^{\epsilon_1}(K_F)^{\epsilon_2}(K_I)^{\epsilon_3} \tag{A.3.4}$$

$$Y_I = (v_m \tau)^{\chi\beta/\eta}(K_G)^{\epsilon_4}(K_F)^{\epsilon_5}(K_I)^{\epsilon_6} \tag{A.3.5}$$

where $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6$are given by the following expressions:

$$\epsilon_1 = \alpha[1 + (\alpha - \beta)\chi/\eta]$$

$$\epsilon_2 = \alpha\chi/\eta(1 - \alpha) + (1 - \alpha)$$

$$\epsilon_3 = \alpha\chi(\beta - 1)/\eta$$

$$\epsilon_4 = \beta[1 + (\alpha - \beta)\chi/\eta]$$

$$\epsilon_5 = \beta\chi(1 - \alpha)/\eta$$

$$\epsilon_6 = \beta\chi/\eta(\beta - 1) + (1 - \beta)$$

Comparative Statics for Welfare-Maximizing Share of Maintenance (v_m^{**}) when $\tau = \frac{\alpha\chi}{(1+\beta\chi)}$

$$\frac{\partial v_m^{**}}{\partial(Y_I/Y_F)} = \frac{1}{(1-\theta_G)}[(\beta/\alpha)] > 0 \tag{A.3.6}$$

$$\frac{\partial v_m^{**}}{\partial(\theta_G)} = \frac{1}{(1-\theta_G)^2}[Y_I/Y_F\,(\beta/\alpha)] > 0 \tag{A.3.7}$$

Comparative Statics for Growth-Maximizing Share of Maintenance (v_m)

$$\frac{\partial v_m}{\partial(\chi)} = \frac{1}{(1-\theta_G)}[\frac{\alpha}{(1+\beta\chi)^2}] > 0 \tag{A.3.8}$$

$$\frac{\partial v_m}{\partial(\theta_G)} = \frac{1}{(1-\theta_G)^2}[\frac{\alpha\chi}{(1+\beta\chi)}] > 0 \tag{A.3.9}$$

Chapter 5

Concluding Remarks

The results in Chapter 2 show that the Indian labor market is segmented between formal and informal employment. Evidence does not support the hypothesis of a fully competitive labor market and that workers choose informal employment as a last resort. This finding is in strong contrast to the empirical evidence from some Latin American countries that show informal employment is a competitive choice for workers over formal employment. The paper shows that majority of the workers would have earned more as formal workers than their current informal wage if they had full access to formal employment. Also, informal workers earn significantly less than formal workers as reflected by the wage gap between the two sets of workers at all quantiles of the wage distribution. Thus, the finding refutes the competitive hypothesis for majority of the informal workers. The current empirical and theoretical literature on informal employment is mainly based on Latin American countries. Evidence amassed in this paper show that the Indian labor market is systematically different from Latin American countries providing the platform for further research on both empirical and theoretical grounds focusing on the South Asian countries.

In Chapter 3, we use two firm-level datasets from India's manufacturing sector to estimate the output elasticities of public investment for firms in the formal and informal sector. We also examine how these output elasticities and relative capital intensity vary across the size distribution of firms in each sector.

Our results indicate that while public investment is an important factor in influencing firm-level productivity in both the formal and informal sectors, there are important sectoral differences. First, the average output elasticity of the flow measure of public investment for an informal sector firm is lower than that of its formal counterpart by a factor of about 3. When we consider a stock measure for public investment, this difference increases to a factor of 7, indicating that the benefits of the accumulated stock of public capital is much larger for firms in the formal sector. The sub-category of Economic Services, containing public expenditures on goods such as transport, communications, power, etc. is associated with systematically larger output elasticities relative to Social Services, which include spending on education, healthcare, water and sanitation, etc. In estimating these sectoral elasticities, we use a method proposed by Levinsohn and Petrin (2003) and Sivadasan (2009) to control for firm-level endogeneity in the usage of private factors of production. Second, results from our quantile regressions suggest that the size distribution of firms in each sector matter for the effects associated with public investment. For example, for formal sector firms, there is very little variation in the output elasticity of public investment across their size distribution. On the other hand, the corresponding output elasticity for informal sector firms is strictly increasing in firm size. Further, the relationship between public investment and the capital intensity in production for formal sector firms is negative, with the effect being the most negative for firms in the middle of the size distribution. This suggests that public investment may be a substitute for private factors in formal production. By contrast, the relationship between public investment and capital intensity is strictly increasing with firm size for the informal sector, indicating strong complementarities. Again, the largest firms in the informal sector benefit the most from public investment. From a policy perspective, our results suggest that firms in the informal sector do indeed benefit significantly from public investment, even though these benefits are relatively smaller on average than those for their counterparts in the formal sector.

In Chapter 4, I set up a two-sector endogenous growth model to examine the implication of the informal sector on optimal fiscal policies. Since congestion affects the efficiency of public infrastructure, in order to make public infrastructure efficient, the government must devote some resources to maintain the existing stock of public infrastructure. This paper throws some light on how tax revenues should be spent on investments on new infrastructure and maintenance of existing public infrastructure. I find that under a centrally planned economy, the social planner may choose an arbitrary level of tax rate and implement a welfare maximizing share of maintenance expenditure in total spending. The tax rate however may be chosen to maximize the decentralized growth rate. The results shown in this paper have important policy implications. First, it is shown that the welfare maximizing share of maintenance spending in total spending on infrastructure is positively related to the ratio of informal to formal sector output. Thus economies with large informal sectors must devote more resources toward maintaining public infrastructure. Second, we find that both welfare maximizing and decentralized fiscal policies are functions of the parameters of the formal and informal production processes.

References

Agenor, P.-R. (2009). Infrastructure investment and maintenance expenditure: Optimal allocation rules in a growing economy. *Journal of Public Economic Theory 11*(2), 233–250.

Alan, M. (2011). Imperfect competition in the labor market. *Handbook of labor economics 4*, 973–1041.

Albrecht, J., A. Van Vuuren, and S. Vroman (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the netherlands. *Labour Economics 16*(4), 383–396.

Arulampalam, W., A. L. Booth, and M. L. Bryan (2007). Is there a glass ceiling over europe? exploring the gender pay gap across the wage distribution. *Industrial & Labor Relations Review 60*(2), 163–186.

Aschauer, D. A. (1989a). Does public capital crowd out private capital? *Journal of monetary economics 24*(2), 171–188.

Aschauer, D. A. (1989b). Is public expenditure productive? *Journal of monetary economics 23*(2), 177–200.

Ashenfelter, O. C., H. Farber, and M. R. Ransom (2010). Labor market monopsony. *Journal of Labor Economics 28*(2), 203–210.

Azam, M. (2012). Changes in wage structure in urban india, 1983–2004: A quantile regression decomposition. *World Development 40*(6), 1135–1150.

Bank, W. (1994). *Investing in Infrastructure.* Oxford University Press: New York.

Barro, R. J. (1990). Government spending in a simple model of endogenous growth. *Journal of Political Economy 98*(5 pt 2).

Besley, T. and R. Burgess (2004). Can labor regulation hinder economic performance? evidence from india. *The Quarterly Journal of Economics*, 91–134.

Binswanger, H. P., S. R. Khandker, and M. R. Rosenzweig (1993). How infrastructure and financial institutions affect agricultural output and investment in india. *Journal of Development Economics 41*(2), 337–366.

Bom, P. R. and J. E. Ligthart (2014). What have we learned from three decades of research on the productivity of public capital? *Journal of Economic Surveys 28*(5), 889–916.

Bosworth, D. L., P. J. Dawkins, and T. Stromback (1996). The economics of the labour market.

Bourguignon, F., M. Fournier, and M. Gurgand (2007). Selection bias corrections based on the multinomial logit model: Monte carlo comparisons. *Journal of Economic Surveys 21*(1), 174–205.

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 257–273.

Carneiro, F. G. and A. Henley (2001). Modelling formal vs. informal employment and earnings: micro-econometric evidence for brazil. *U of Wales at Aberystwyth Management & Business Working Paper* (2001-15).

Chakraborty, S. and E. Dabla-Norris (2009). *The quality of public investment*. International Monetary Fund.

Chatterjee, S. and S. J. Turnovsky (2007). Foreign aid and economic growth: The role of flexible labor supply. *Journal of development economics 84*(1), 507–533.

Chen, M. A., J. Vanek, and M. Carr (2004). *Mainstreaming informal employment and gender in poverty reduction: a handbook for policy-makers and other stakeholders*. Commonwealth secretariat.

Cohen, B. and W. J. House (1996). Labor market choices, earnings, and informal networks in khartoum, sudan. *Economic Development and Cultural Change*, 589–618.

Dabla-Norris, E., J. Brumby, A. Kyobe, Z. Mills, and C. Papageorgiou (2012). Investing in public investment: an index of public investment efficiency. *Journal of Economic Growth 17*(3), 235–266.

Dahl, G. B. (2002). Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica 70*(6), 2367–2420.

Datta, S. (2012). The impact of improved highways on indian firms. *Journal of Development Economics 99*(1), 46–57.

Devarajan, S., V. Swaroop, and H.-f. Zou (1996). The composition of public expenditure and economic growth. *Journal of monetary economics 37*(2), 313–344.

Dickens, W. T. and K. Lang (1985). A test of dual labor market theory. *The American Economic Review*, 792–805.

Dioikitopoulos, E. V. and S. Kalyvitis (2008). Public capital maintenance and congestion: Long-run growth and fiscal policies. *Journal of Economic Dynamics and Control 32*(12), 3760–3779.

Domencich, T. and D. MacFadden (1975). Urban travel demand: a behavioral analysis: a charles river associates research study. *Contributions to economic analysis (* (93).

Dubin, J. A. and D. L. McFadden (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica: Journal of the Econometric Society*, 345–362.

Duraisamy, P. (2002). Changes in returns to education in india, 1983–94: by gender, age-cohort and location. *Economics of Education Review 21*(6), 609–622.

Fields, G. S. (1990). Labour market modelling and the urban informal sector: theory and evidence.

Futagami, K., Y. Morita, and A. Shibata (1993). Dynamic analysis of an endogenous growth model with public capital. *The Scandinavian Journal of Economics*, 607–625.

Ghani, E., A. G. Goswami, and W. R. Kerr (2015). Highway to success: The impact of the golden quadrilateral project for the location and performance of indian manufacturing. *The Economic Journal 126*(591), 317–357.

Gindling, T. H. (1991). Labor market segmentation and the determination of wages in the public, private-formal, and informal sectors in san jose, costa rica. *Economic Development and Cultural Change*, 585–605.

Glomm, G. and B. Ravikumar (1997a). Productive government expenditures and long-run growth. *Journal of Economic Dynamics and Control 21*(1), 183–204.

Glomm, G. and B. Ravikumar (1997b). Productive government expenditures and long-run growth. *Journal of Economic Dynamics and Control 21*(1), 183–204.

Gomis-Porqueras, P., A. Peralta-Alva, and C. Waller (2014). The shadow economy as an equilibrium outcome. *Journal of Economic Dynamics and Control 41*, 1–19.

Gramlich, E. M. (1994). Infrastructure investment: a review essay. *Journal of economic literature*, 1176–1196.

Grootaert, C. and B. Mundial (1988). *Cote d'Ivoire's vocational and technical education.* Number 19. International Bank for Reconstruction and Development.

Günther, I. and A. Launov (2012). Informal employment in developing countries: opportunity or last resort? *Journal of development economics 97*(1), 88–98.

Gupta, S., A. Kangur, C. Papageorgiou, and A. Wane (2014). Efficiency-adjusted public capital and growth. *World Development 57*, 164–178.

Gyourko, J. and J. Tracy (1988). An analysis of public-and private-sector wages allowing for endogenous choices of both government and union status. *Journal of Labor Economics 6*(2), 229–53.

Harris, J. R. and M. P. Todaro (1970). Migration, unemployment and development: a two-sector analysis. *The American economic review*, 126–142.

Hay, J. W. (1980). *Occupational choice and occupational earnings: Selectivity bias in a simultaneous logit-OLS model.* Ph. D. thesis, Yale University.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161.

Heckman, J. J. and V. J. Hotz (1986). An investigation of the labor market earnings of panamanian males evaluating the sources of inequality. *Journal of Human Resources*, 507–542.

Hill, M. A. (1989). Female labor supply in japan: Implications of the informal sector for labor force participation and hours of work. *Journal of Human Resources*, 143–161.

Hilmer, M. J. (2001). A comparison of alternative specifications of the college attendance equation with an extension to two-stage selectivity-correction models. *Economics of Education Review 20*(3), 263–278.

Holtz-Eakin, D. and A. E. Schwartz (1995). Infrastructure in a structural model of economic growth. *Regional Science and Urban Economics 25*(2), 131–151.

Hulten, C. R. (1996). Infrastructure capital and economic growth: How well you use it may be more important than how much you have. Technical report, National Bureau of Economic Research.

Hulten, C. R., E. Bennathan, and S. Srinivasan (2006). Infrastructure, externalities, and economic development: a study of the indian manufacturing industry. *The World Bank Economic Review 20*(2), 291–308.

Ihrig, J. and K. S. Moe (2004). Lurking in the shadows: the informal sector and government policy. *Journal of Development Economics 73*(2), 541–557.

ILO (2012). Statistical update on employment in the informal economy.

ILO (2013). Women and men in the informal economy: A statistical picture. *Second Edition, International Labour Office-Geneva*.

Institute, U. L. and Ernst&Young (2013). Infrastructure 2013: Global priorities, global insights. *Washington, D.C.: Urban Land Institute*.

Ito, T. (2009). Caste discrimination and transaction costs in the labor market: Evidence from rural north india. *Journal of development Economics 88*(2), 292–300.

Kalaitzidakis, P. and S. Kalyvitis (2004). On the macroeconomic implications of maintenance in public capital. *Journal of Public Economics 88*(3), 695–712.

Kalaitzidakis, P. and S. Kalyvitis (2005). New public investment and/or public capital maintenance for growth? the canadian experience. *Economic Inquiry 43*(3), 586–600.

Kalyvitis, S. and E. Vella (2015). Productivity effects of public capital maintenance: evidence from us states. *Economic Inquiry 53*(1), 72–90.

Khandker, S. R. (1992). Earnings, occupational choice, and mobility in segmented labor markets of india. Technical report, World Bank.

Koenker, R. and G. Bassett Jr (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50.

Lall, S. V. (1999). The role of public infrastructure investments in regional development: Experience of indian states. *Economic and Political Weekly*, 717–725.

LaPorta, R. and A. Shleifer (2008). The unofficial economy and economic development. *Brookings Papers on Economic Activity 2008*(2), 275–363.

LaPorta, R. and A. Shleifer (2014). Informality and development. *The Journal of Economic Perspectives 28*(3), 109–126.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica: Journal of the Econometric Society*, 507–512.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*(2), 317–341.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The manchester school 22*(2), 139–191.

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.

Lynde, C. and J. Richmond (1992). The role of public capital in production. *The Review of Economics and Statistics*, 37–44.

Machado, J. A. and J. Mata (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics 20*(4), 445–465.

Magnac, T. (1991). Segmented or competitive labor markets. *Econometrica: journal of the Econometric Society*, 165–187.

Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? evidence from sectoral transitions in mexico. *The World Bank Economic Review 13*(2), 275–302.

Maloney, W. F. (2004). Informality revisited. *World development 32*(7), 1159–1178.

Marcouiller, D., V. R. de Castilla, and C. Woodruff (1997). Formal measures of the informal-sector wage gap in mexico, el salvador, and peru. *Economic development and cultural change*, 367–392.

McFadden, D. (1973). *Conditional logit analysis of quantitative choice behavior*, Chapter Frontiers in Econometrics. New York: Academic Press.

McKinsey (2013). Infrastructure productivity: How to save dollar 1 trillion a year. *New York: McKinsey Global Institute*.

Mehrotra, S., A. Gandhi, P. Saha, and B. Sahoo (2013). Turnaround in india's employment story-silver lining amidst joblessness and informalization? *Economic and Political Weekly 48*(35).

Mehrotra, S., A. Gandhi, B. Sahoo, and P. Saha (2012). Creating employment in the twelfth five-year plan. *Economic and Political Weekly 47*(19), 63–73.

Mincer, J. (1974). *Schooling, earnings and experience.* New York: Columbia University Press.

Mitra, A., A. Varoudakis, and M.-A. Veganzones-Varoudakis (2002). Productivity and technical efficiency in indian states'manufacturing: The role of infrastructure. *Economic Development & Cultural Change 50*(2), 395.

Munnell, A. H. and L. M. Cook (1990). How does public infrastructure affect regional economic performance? *Proceedings of Conference Series no. 34, Federal Reserve Bank of Boston.*.

Navarro-Lozano, S. and P. Schrimpf (2004). The importance of being formal: testing for segmentation in the mexican labor market. *Chicago, United States: University of Chicago. Manuscript*.

NSSO (2013). Participation of women in specified activities along with domestic duties. Technical Report Report No. 550 (66/10/5), National Sample Survey.

Ordonez, J. C. L. (2014). Tax collection, the informal sector, and productivity. *Review of Economic Dynamics 17*(2), 262–286.

Packard, T. G. (2007). Do workers in chile choose informal employment? a dynamic analysis of sector choice. *A Dynamic Analysis of Sector Choice (May 1, 2007). World Bank Policy Research Working Paper* (4232).

Penalosa, C. G. and S. J. Turnovsky (2005). Second-best optimal taxation of capital and labor in a developing economy. *Journal of Public Economics 89*(5), 1045–1074.

Prado, M. (2011). Government policy in the formal and informal sectors. *European Economic Review 55*(8), 1120–1136.

Pratap, S. and E. Quintin (2006). Are labor markets segmented in developing countries? a semi-parametric approach. *European Economic Review 50*(7), 1817–1841.

Pritchett, L. (2000). The tyranny of concepts: Cudie (cumulated, depreciated, investment effort) is not capital. *Journal of Economic Growth 5*(4), 361–384.

Rauch, J. E. (1991). Modelling the informal sector formally. *Journal of Development Economics 35*(1), 33–47.

Rioja, F. K. (2003). Filling potholes: macroeconomic effects of maintenance versus new investments in public infrastructure. *Journal of Public Economics 87*(9), 2281–2304.

Sastry, N. (2004). *Estimating informal employment & poverty in India.* Human Development Resource Centre.

Schneider, F., A. Buehn, and C. E. Montenegro (2010). New estimates for the shadow economies all over the world. *International Economic Journal 24*(4), 443–461.

Schneider, F. and D. H. Enste (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature 38*, 77–114.

Sengupta, A. et al. (2007). Report on conditions of work and promotion of livelihoods in the unorganised sector. *National Commission for Enterprises in the Unorganised Sector, Government of India, New Delhi*.

Sivadasan, J. (2009). Barriers to competition and productivity: Evidence from india. *The BE Journal of Economic Analysis & Policy 9*(1).

Solow, R. M. (1980). Another possible source of wage stickiness. *Journal of macroeconomics 1*(1), 79–82.

Souabni, S. (2013). Mmsel: Stata module to simulate (counterfactual) distributions from quantile regressions (w/optional sample selection correction). *Statistical Software Components*.

Stiglitz, J. E. (1981). *Alternative theories of wage determination and unemployment: The efficiency wage model*. Princeton University, Woodrow Wilson School.

Tannuri-Pianto, M. and D. Pianto (2002). Informal employment in brazil-a choice at the top and segmentation at the bottom: a quantile regression approach. *Anais do XXIV Encontro Brasileiro de Econometria 2*.

Trost, R. P. and L.-F. Lee (1984). Technical training and earnings: a polychotomous choice model with selectivity. *The Review of Economics and Statistics*, 151–156.

Turnovsky, S. J. (1997). Fiscal policy in a growing economy with public capital. *Macroeconomic Dynamics 1*(03), 615–639.

Turnovsky, S. J. (2004). The transitional dynamics of fiscal policy: long-run capital accumulation and growth. *Journal of Money, Credit and Banking*, 883–910.

Turnovsky, S. J. and M. A. Basher (2009). Fiscal policy and the structure of production in a two-sector developing economy. *Journal of Development Economics 88*(2), 205–216.

Zhang, H. (2004). Self-selection and wage differentials in urban china: a polychotomous model with selectivity. *Massachusetts Institute of Technology, Boston, mimeo*.