1-1-2015

The ‘Peter Pan Syndrome’ in Emerging Markets: The Productivity-Transparency Tradeoff in IT Adoption

K. Sudhir

Debabrata Talukdar

THE "PETER PAN SYNDROME" IN EMERGING MARKETS: THE PRODUCTIVITY-TRANSPARENCY TRADEOFF IN IT ADOPTION

By

K. Sudhir and Debabrata Talukdar

January 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 1980

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

The "Peter Pan Syndrome" in Emerging Markets:
The Productivity-Transparency Tradeoff in IT Adoption

K. Sudhir
Debabrata Talukdar*

January 2015

* K. Sudhir is James L. Frank Professor of Private Enterprise and Management, Yale School of Management, New Haven, CT 06520. Ph: 203-432-3289 (e-mail: k.sudhir@yale.edu). Debabrata Talukdar is Professor of Marketing, School of Management, State University of New York at Buffalo, Buffalo, NY 14260. Ph: 716-645-3243 (e-mail: dtalukda@buffalo.edu). We thank Mushfiq Mobarak, Juanjuan Zhang, participants at the 2012 Marketing Science Conference in Boston, the 2012 Marketing Science Emerging Markets Conference at Wharton, the 2013 China India Insights Conference in New York, the 2014 NBER Summer Workshop on the Economics of Information Technology and Digitization, the 2015 AIM-AMA Sheth Doctoral Consortium in Dubai, the marketing workshops at Universities of Connecticut, Massachussetts and Toronto for their helpful comments.

The "Peter Pan Syndrome" in Emerging Markets: The Productivity-Transparency Tradeoff in IT Adoption

Abstract

Firms make investments in technology to increase productivity. But in emerging markets, where a culture of informality is widespread, information technology (IT) investments leading to greater transparency can impose a cost through higher taxes and need for regulatory compliance. This tendency of firms to avoid productivity-enhancing technologies and remain small to avoid transparency has been dubbed the "Peter Pan Syndrome." We examine whether firms make the tradeoff between productivity and transparency by examining IT adoption in the Indian retail sector. We find that computer technology adoption is lower when firms have motivations to avoid transparency. Specifically, technology adoption is lower when there is greater corruption, but higher when there is better enforcement and auditing. So firms have a higher productivity gain threshold to adopt computers in corrupt business environments with patchy and variable enforcement of the tax laws. Not accounting for this motivation to hide from the formal sector *underestimates* productivity gains from computer adoption. Thus in addition to their direct effects on the economy, enforcement, auditing and corruption can have indirect effects through their negative impact on adoption of productivity enhancing technologies that also increase operational transparency.

1. Introduction

For many businesses in emerging markets, information technology (IT) is a double-edged sword. On the one hand, IT systems can help improve productivity and thus help firms gain a competitive advantage. But the same systems that improve productivity also increase transparency of transactions by leaving a clear audit trail. Such increased transparency makes it easier for the government to collect taxes and enforce regulatory compliance by bringing these transactions into the formal sector of the market, potentially increasing the cost of operations, relative to those who do not use IT systems. In emerging markets, where enforcement is patchy and corruption is rampant, firms who keep much of their transactions in the informal sector can therefore gain a competitive advantage.[1] In such settings, the gains in productivity from adoption of IT are moderated by the attendant costs of making the transactions subject to taxation and regulatory compliance (Bird and Zolt 2008; Johnson et al. 2000; Mishra, Subramanian and Topalova 2008; Sinha 2003). At the margin, firms may therefore limit investments in IT, to the detriment of overall productivity, especially if their beliefs about the relative magnitude of productivity gains versus transparency costs are underestimated (Gatti and Honoratti 2008; Smith 2013). This tendency of firms in emerging markets to shun growth and remain small at the expense of efficiency, technology adoption and innovativeness to avoid taxes and regulatory scrutiny has been dubbed the "Peter Pan Syndrome."[2] Sunder (2012) summarizes the dilemma in the context of domestic Indian retailers' reluctance to modernize through IT systems: "The system that serves to manage large retail organizations is also convenient for tax payment and collection…Indian retailers can and should break out of the self-defeating confines of the beliefs about the profitability of tax evasion."

There is some intuitive appeal to the conjecture that transparency concerns might impede IT adoption among emerging market retailers; however the conjecture has not received empirical scrutiny. Just as important, the productivity enhancing benefits of IT adoption in emerging

[1] The informal or grey economy is here defined as trade, services or production, that is noncompliant in any aspect(s) of company registration, tax declaration/payment, business regulation (e.g., employer's national insurance, public/employer's liability insurance), and/or licensing requirements for the specific trade (e.g., health and safety certificate).

[2] In an article titled "The Peter Pan Syndrome," the Economist (May 17, 2014) states: "Manuel Milano of the Mexican Competitiveness Institute, a think-tank, calls this a "Peter Pan System" in which firms prefer to stay small than to grow, mostly because of tax and regulation. "It is easier to fly under the radar when you are microscopic." "The article goes on to discuss the large opportunity costs of firms for remaining small—higher interest rates from banks, lack of efficiency, technology and innovation.

markets cannot be taken for granted. For example, it is possible that given the low cost of labor and the lack of complementary infrastructures, the gains through productivity enhancement from IT adoption by retailers in emerging markets may not be sufficiently high to warrant IT adoption. For IT to help improve productivity, the business ecosystem and organization should be able take advantage of the technology. For example, in the absence of supply chain and cold chain infrastructure in emerging markets, the value of computers for efficient supply chain management may be quite limited. Similarly, when a retailer's employees are older and untrained and unfamiliar with using IT systems, installing IT systems will not lead to productivity gains. This is a particularly relevant concern, given that even within an advanced high-income economy such as the United States there was much academic debate till the mid-nineties on whether IT in fact improves productivity.

For instance, much of early research on IT productivity claimed a "IT-productivity paradox" in that it was not possible to reject the hypothesis that computers add nothing to total output (e.g., Loveman 1994), or found that the marginal costs exceeded marginal benefits (Morrison and Berndt 1990).[3] It was not until Brynjolffsson and Hitt (1996) showed through detailed firm-level survey data that dollar for dollar, spending on computer capital created more value than spending on other types of capital that the tide began to turn and researchers were able to demonstrate that IT does increase productivity. The literature discusses two reasons for the divergence of the results. First, the results reporting insignificant effects were from an older period in the seventies when IT productivity might indeed have been lower. Second, as discussed in the examples earlier, complementary infrastructure and the organizational redesign necessary to exploit IT may not have been present (Commander, Harrison & Menezes-Filho 2011). As complementary infrastructures may be inadequate and firms could still be in the low productivity part of the experience curve in emerging markets, the conjecture that IT improves productivity in an emerging market deserves systematic empirical scrutiny.

Our goal in this paper is to therefore empirically answer three questions about the use of IT by businesses in emerging markets: First, do operational transparency concerns impede IT adoption by businesses? Second, does IT adoption have a positive impact on productivity and how much? Importantly, the magnitude of the impact may be underestimated at the margin, if

[3] Robert Solow, the Nobel Prize winning economist characterized the IT productivity paradox thus: "we see computers everywhere except the productivity statistics."

firms with potentially high productivity gains do not adopt computers due to transparency concerns that impede IT adoption. Third, we seek to understand how this tradeoff varies by size of firms. Do transparency concerns reduce IT adoption among smaller or larger firms? Do larger firms gain more in productivity than smaller firms through IT adoption? We answer these questions using detailed firm-level survey data on 1948 retail firms covering a broad cross-section of states and cities in India. In addition, we use several other sources to augment this firm level survey data with state level data in terms of a number of relevant variables like corruption level, minimum wages rate, and overall socio-economic development indices.

India presents an ideal setting to study these questions. First, the retail sector is at an early stage of modernization, labor is still relatively cheap, the complementary infrastructures are still not fully available; and hence the productivity gains from IT adoption is a priori ambiguous, requiring systematic empirical analysis. Specifically, the minimum wages and literacy levels vary across states, giving us state level variation on the labor-saving productivity benefits of using computers. Second, with high level of corruption in India, the transparency concerns are especially acute as India scores a poor 36 (out of 100) in the Transparency International (2012) report and ranks at 94 out of 176 countries. Further, given India's federal system of government where states have significant power, there is considerable variation in the levels of corruption, enforcement and auditing across different states in India. These variations are valuable in identifying the empirical link between IT adoption and transparency motivations. To address the concern as to whether the link between IT adoption and transparency levels across states are not merely due to another unobserved factor that varies across states and is correlated with transparency and technology adoption, we perform a falsification test. Specifically, we tested the link between generator adoption and transparency variables, as transparency concerns should not affect generator adoption. We find that consistent with our hypotheses, unlike IT adoption, generator adoption is not linked to transparency related factors.

Further, in evaluating the effect of IT adoption on productivity, there are obvious selection concerns because business computer adoption is not random. We assess selection concerns using two approaches. First, we use propensity score matching to ensure that inferences of productivity differences between adopters and non-adopters are between firms that are "comparable" in their propensity to adopt. We also test for potential "selection on unobservables" using a Rosenbaum bounds approach to assess whether unobservable factors relating to computer adoption might drive the positive estimates of productivity effects. Second, we estimate a model of self-selection using transparency variables as instruments. Variations in corruption and enforcement levels across states and across firms, serve as exclusion restrictions in that they impact computer adoption by firms, but does not directly impact their revenues.

Our key findings are as follows: (1) At the margin, higher corruption levels are related to lower computer adoption. (2) Better regulation enforcement increases computer adoption because it creates a level playing field across firms, reducing transparency concerns. (3) Generators do increase productivity, but as one would expect, their adoption is not affected by transparency concerns. (4) Computer adoption increases store productivity on average by about 50 to 70 percent. The effects of transparency on computer adoption and the impact of computer adoption on productivity are both greater for larger than for smaller firms. (5) Not taking into account the endogenous effects of transparency related variables on computer adoption underestimates the productivity gains from IT adoption, suggesting that productivity estimates in emerging markets with non-transparent environments should account for such concerns.

Our results have obvious implications for policy makers, as they show that corruption and lax enforcement of tax laws can not only lead to direct losses in tax revenues, but also indirect losses due to productivity drop from reduced adoption of productivity enhancing systems that increase transparency. From a marketer perspective, our results show that transparency concerns will reduce market sizes of productivity-enhancing products that also increase transparency (e.g., computers, cash registers which maintain records in memory, and credit card machines). Further, they suggest that marketers should use variables measuring corruption, enforcement levels and audit mechanisms as predictors for market potentials for such products.

The rest of the paper is organized as follows. Section 2 provides background on the Indian retail sector, issues of informal sector in emerging markets and the literature on IT productivity. Section 3 describes the data. Section 4 describes the empirical analysis and the results. Section 5 concludes.

2. Background

We position this paper against two streams of literature: the IT- productivity relationship and the culture of informality in emerging markets. Finally, we discuss why the Indian retail sector is a particularly appropriate setting to study the productivity-transparency tradeoff.

2.1 IT Adoption and Productivity

As discussed in the introduction, the link between IT adoption and productivity was the subject of much controversy in the eighties and early nineties. Early analysis using firm level data from 1978-82 did not find evidence of productivity increases (Loveman 1994; Barua et al. 1995). It is possible that productivity gains were not large in the early stages of IT adoption; others have argued that the inability to detect productivity gains could be due to aggregation and measurement bias (Brynjolfsson and Hitt 1996; Stiroh 2004). This “productivity paradox” was resolved through analysis of later data during the 1987-92 period by Brynjolfsson and Hitt (1996). Since then a number of studies have found strong and positive association between IT adoption and productivity (e.g., Ichniowski, Shaw & Prennushi 1997; Black & Lynch 2001; Bartel et al. 2002; Bartel, Ichniowski & Shaw 2005). At the same time, the magnitude of IT productivity gains is found to vary significantly across countries, with estimates for European economies far lower than for the US (Basu et al 2003; Jorgenson 2001; Stiroh 2002). It is well recognized that one needs complementary logistics, supportive regulatory environments for the effective use of IT within a national economy (Commander et al. 2011). Emerging markets may lack these complementary logistics and regulatory environments, potentially limiting productivity gain from IT. Often organizations need to be redesigned to support IT; as organization redesign lags IT adoption, the benefits of IT adoption may not be immediately seen.

As for the link between firm size and productivity gains from IT, it is theoretically unclear. While larger firms have more complex coordination needs that can aid greater productivity gains (Dasgupta et al. 1999), smaller firms may be more flexible to take better advantage of IT (Morgan et al. 2006). Not surprisingly, empirical results also remain mixed though most papers report a positive relationship (e.g., Delone 1981; Fabiani et al. 2005; Morgan et al. 2006; Thong 1999); while some report insignificant (e.g., Lefebvre et al. 2005; Love et al. 2005) and negative relationships (e.g., Dewett and Jones 2001; Harris and Katz 1991).

2.2 Culture of Informality

A culture of informality – where firms keep business outputs hidden or opaque from the formal system of monitoring and thus avoid being subject to government taxation and regulation – varies across economies (Dabla-Norris, Gradstein & Inchauste 2008). The share of informal business activities is estimated at between 10% and 20% of GNP for developed countries; it ranges from 33% to 50% for developing countries (Schneider and Enste 2002). It is important to recognize that the practice is not limited to firms in the informal sector only, especially in emerging markets. A report by McKinsey Global Institute (Farrell 2004) notes: "The informal economy is not just the unregistered street vendors and tiny businesses that form the backbone of marketplaces in Asia and other emerging markets. It includes many established companies, often employing hundreds of people, in industries as diverse as retail, construction, consumer electronics, software, pharmaceuticals and even steel production."

Firms prefer informality as it helps them avoid taxes and costly regulation; unilaterally avoiding taxes becomes a competitive advantage when firms are unlikely to be caught and punished. For example, when corruption is high or enforcement is patchy, tax avoidance is feasible through paying bribes. Further, by keeping tax-related operational activities informal and avoid transparency enhancing technologies, they can reduce the level of "electronic trail" government officials can have in demanding bribes[4] (Mishra et al. 2008; Russell 2010). Unfortunately, the culture of informality leads to a vicious cycle of further tax avoidance and drive to informality as governments are forced to increase tax rates from the smaller set of compliant firms which incentivizes them further to become non-compliant (e.g., Azuma and Grossman 2002; Dabla-Norris et al. 2008; Marcouiller and Young 1995). In contrast, when the enforcement environment is excellent and there are auditing mechanisms to ensure that such tax avoidance is harder, firms are less likely to be in the informal sector and less motivated to avoid transparency enhancing technologies. This is because computers provide productivity enhancing benefits, but do not put the firm at a competitive disadvantage because better enforcement ensures that all players are on a level playing field.

There is also face validity that business computerization increases operational transparency and helps better enforcement by creating easily detectable "digital traces" of taxable business activities through a transparent record keeping system (Friedman et al. 2000; International Tax Compact 2010; Russell 2010). For example, the governments of Bangladesh

[4] Miller and Tucker (2014) find that U.S. hospitals are one-third less likely to adopt "electronic medical records" (EMR) systems in those states that allow search and use of electronic records in ligation cases, even though EMR systems enhance operational productivity and cost efficiency. Their results suggest that even in developed countries like the United States, transparency concerns can reduce IT adoption.

(The Daily Star 2007), China (People's Daily 2000) and Ethiopia (Mesfin 2012) have recently mandated use of computerized systems to facilitate easy enforcement and minimize tax evasion.[5]

2.3 Choice of Setting: Indian Retail Sector

The Indian retail sector is the fifth largest in the world with a current market size of about US$ 500 billion and average growth rates of between 8-10%. Yet, the Indian retail sector lags behind those of peer emerging markets like China when it comes to adoption of modern management technologies and IT systems to help and guide retail business practices (Reardon & Gulati 2008; Sunder 2012). The retail sector in India thus offers an ideal setting for studying the productivity-transparency tradeoff. It is often argued that the sector is well positioned to gain in productivity from IT adoption through improvement in inventory management, pricing and customer relationship management (Foster, Haltiwanger and Krizan 2002; Sunder 2012). Yet, the low rate of IT adoption could be because the Indian environment is not conducive to productivity gains from adoption. For example, lack of complementary infrastructures (e.g., logistics and supply chain, road infrastructure) may reduce the productivity gain from IT adoption. Or low labor costs may reduce the gains from IT adoption.

Also, transparency concerns are quite real in India with an endemic national "culture" of corruption. Transparency International (TI) found that more than half of those surveyed had firsthand experience of paying bribe or peddling influence to get a job done in a public office, and it ranked India 94 among 176 countries for lack of transparency (TI 2012). India tops the worldwide list for "black money" with almost $1,456 billion stashed in Swiss banks (Nayar 2011; Rao 2010); an amount 13 times the country's total external debt. The popular press is replete with articles that note how "tax evasion is a national sport" for both businesses and individuals (Chopra 2011; Dhara & Thomas 2011). Understanding the relative importance of productivity and transparency in the low rate of computer adoption can be a critical aid to policy prescriptions on how to improve productivity in one of the world's largest retail markets.

3. Data

We collated the data necessary for the analysis from multiple sources. We first discuss the sources and then provide descriptive statistics of the variables used in our empirical analyses.

[5] The People's Daily of China (2000) reported: "China has stepped up its efforts to fight against tax evasion by requiring selected companies to print invoices using a computerized system connected to taxation authorities."

3.1 **Data Sources**

Our primary data source for this study is a large scale World Bank survey of Indian retailers conducted in 2006. As part of its private sector development project and research initiatives, the World Bank conducts regular surveys of individual firms in many developing countries. Such firm level surveys are used to guide internal bank policies, but have also been occasionally used to address academic research questions in economics and finance (e.g., Angelini & Generale 2008; Cull & Xu 2005). In particular, Amin (2010) uses the data from this particular survey to study the effect of labor regulations on computer adoption.

The survey consists of a stratified random sample of 1948 retail stores operating in the formal sector and located in 16 major states and federal territories across 41 Indian cities. The National Industrial Classification groups Indian retailers into those operating through registered stores and the rest who usually operate informally from home (NIC 1998, Industry Division 52). All stores in our sample belong to the former group.

The sampling was carried out with a first level stratification of three segments by retail store type: (i) traditional stores - which include general and departmental stores, grocers, chemists, food stores, etc., (ii) consumer durable stores - which are specialized stores carrying durable items like televisions, home appliances, etc., (iii) modern format stores - which are large stores and part of a shopping complex. These three store types account for 64%, 26% and 10% of the sample, respectively. Within each store type segment, a secondary stratification was based on operation size. The overall sample size was determined so as to minimize the standard error in the sample variables, given the available resources for each surveying stratum.

The survey was conducted by the Indian unit of a reputed international market research firm and involved personal interviews with store managers. The store managers were told that the goal of the survey was to gather opinions about the investment climate for the retail sector in the country. They were also told that the information obtained will be held in the strictest confidentiality; neither their names nor the names of their businesses would be used in any document based on the survey. The survey collected information on a variety of store characteristics such as annual sales, key operational costs, employment, availability of infrastructure, access to finance, etc. It also reports on the store manager's perceptions about various aspects of the business climate including competition and corruption culture.

We augment this store level survey data from the World Bank with relevant state level data from other sources. Specifically, the state level corruption index is obtained from the 'Indian Corruption Study' that was undertaken by Transparency International (TI) and released in October of 2005. It is one of the largest corruption surveys ever conducted, with a total of 14,405 respondents; spread over 151 cities and 306 rural areas within 20 Indian states. We also collected data from Indian government sources on three other state level variables to capture state level differences that can affect IT adoption. One is labor cost, as computers and electronic cash registers can replace (1) competent and experienced accounting and stock-keeping staff who use traditional manual account keeping books, and (2) experienced and trusted cashiers who are competent in mental computations to total up bills and produce change. So, higher labor costs make automation through computers to increase productivity more appealing (Amin 2010).We operationalize labor costs through the minimum wage rates in the retail services sector across states set under the Shops and Establishments Act (SEA) of India. We use data from the 2001 report of India's Labor Bureau as it is the closest year to our World Bank survey year (2006) for which the data was available for all the states in our sample. Though this differs from the actual wages in 2006, we believe that the relative values are likely to be quite comparable.

Another state level variable on which we collect data is the adult literacy rate as a proxy for relative education level differences across states. In states with a less educated workforce, it would be harder to find employees who can use computers effectively, therefore leading to lower adoption. A less educated public may also tolerate more corruption. We use the average of the states' adult literacy rates from the 2001 and 2011 census data by Indian government. The average is likely to be close to the 2006 literacy level. Finally, we use Human Development Index (HDI) to capture differences in socio-economic development across states, as lower development can inhibit technology adoption as well as foster corruption. As these two state level variables impact both corruption and IT adoption, it is important to control for these variables in isolating the direct effect of corruption and other transparency metrics on adoption. We obtain the HDI data for our sample states from the India Human Development Report 2011 (Govt. of India 2011), which computes the index values based on 2007-08 national survey data. The HDI for a state is a composite relative indicator of socio-economic development stage for the state along three key dimensions – education, health and income levels of its population.

3.2 Descriptive Statistics

Table 1 shows the summary statistics of the variables used in the empirical analyses. [6] We measure store performance through gross revenue generated at the store in the latest fiscal year, normalized for "size" of the retailer in terms of employees and store area. Specifically our performance metrics are gross annual revenue per employee (labor productivity) and gross annual revenue per sq. feet (floor area productivity).[7] For our sample stores, mean gross revenues are Rs. 1.90 million, while revenues net of operational costs are Rs. 1.62 million. The median gross revenues is Rs. 0.50 million; for firms at or below the median, the average revenues is Rs. 0.22 million; while for firms above the median, the average revenues is Rs. 3.79 million. With respect to our measure of labor productivity, the mean and median values are Rs. 0.55 million/employee and Rs.0.25 million/employee respectively. For the floor area productivity measure, the mean and median values are Rs. 7450/sq. ft. and Rs. 3330/sq. ft. respectively.

We consider the adoption of two productivity enhancing technologies by each retail store in our sample: business computer and in-store electricity generator. Only 17% of the stores have business computer systems; but 30% own a generator, while 27% own an in-store security system. This suggests that lack of computer adoption may not be entirely due to financial constraints. With 83% of stores facing power outages in the previous year, perhaps greater generator adoption might be optimal. It also reflects the tremendous loss of efficiency and wasted capital in emerging markets, where lack of infrastructure (power) necessitates what might be otherwise wasteful investment in in-store generators.

We consider both state level and firm level perceptions of corruption as it is an "experiential" phenomenon that occurs out of public glare. Hence even though perceptions of corruption by individual firms within a state will be correlated to the aggregate state level index, it will also vary across firms due to firms' differential experiences in the context of their business operations. For example, the officials in local regulatory offices who deal with a particular firm are likely to be different in their "corruption propensity." Similarly, the peer group of firms --

[6] Additional background information, in terms of the rationale for their inclusion and operationalization, for some of the variables are available in the online appendix.
[7] In addition, since the World Bank survey collected data on some key annual operational costs – viz., labor, electricity, communication services and rent or loan payment on land/building, equipment and furniture, we also tested the robustness of our results for productivity measures based on gross revenues net of those costs for the latest fiscal year. The key results are qualitatively identical and are available in the online appendix.

whose actual and/or perceived operational practices shape the perception of a particular firm about corruption prevalence -- will differ across firms, even within the same geography.

The data shows that the stores operate in business environments that vary quite a bit in terms of corruption related factors expected to discourage operational transparency, as well as in terms of regulatory enforcement related factors expected to encourage transparency. For the states included in the World Bank survey, the values of the TI corruption index (measured on a 1-10 scale) range from a low of 2.40 (Kerala) to a high of 6.95 (Bihar). Other variables also vary significantly across the sample states. For instance, the minimum wages rate varies from a low of Rs. 42.50 to a high of Rs. 99.70 with an average value of Rs. 72.38. Figure 1 and Figure 2 show the sample distribution of adoption level of computers by store type and state respectively. While the overall adoption level is low at 16.8%, there is significant variation across both store type and state. Figure 3 shows our primary explanatory variables for why computer adoption varies by (1) enforcement and (2) corruption across the 16 states in our sample. There is substantial variation across states in terms of enforcement and corruption.

In terms of store-specific characteristics, the average number of employees in a store is about 6, but there is substantial standard deviation around the mean. The median number of employees is 2; for firms at or below the median number, the average number of employees is 1.3; while for firms with above the median number of employees, the average number of employees is 12. The average store size is about 600 sq. ft., but here also there is a large standard deviation around the mean. The median size is 150 sq. ft.; for firms at or below the median, the average size is 90 sq. ft; while for firms above the median, the average size is 1167 sq. ft. The average age of the store is 12 years; here the standard deviation is less than the mean, and the mean and median are roughly the same unlike in size and number of employees.

The average experience of the store manager is 13 years. Most stores are owned by single owners with the mean value of the share of the store owned by its largest owner is about 96%. As expected, only 1% of stores are owned by the government, given that retailing is almost virtually a private sector activity, except for fair price shops meant to distribute staple groceries to the poor. Stores vary in their level of access to formal financing. About 36% of stores do not have bank accounts; while 78% do not have access to overdraft facilities—suggesting credit constraints are significant. Stores keep about 12 days of inventory for their main selling products.

[Insert Table 1 about here]

[Insert Figures 1 -3 about here]

4. Empirical Analyses

We begin with bivariate descriptive analyses to obtain a graphic sense of the nature of relationships among the key variables of interest. We then follow this with relevant statistical analyses that control for other variables that can impact the outcome variables. We also account for potential endogeneity concerns in estimating the productivity impact of computer adoption, as computer adoption is an endogenous variable.

4.1 Bivariate Relationships

We first report the relationship between computer adoption levels and transparency and enforcement. Figures 4 and 5 show the state-level scatter plots of computer adoption levels with corruption and regulatory inspection levels. The correlation between computer adoption and corruption levels is -0.53, consistent with the premise that higher the overall culture of corruption, higher will be the propensity towards business tax evasion and thus lower the incentive to adopt transparency enhancing business computer technology. Similarly, the correlation of 0.24 between adoption and inspection level is consistent with the premise that higher the regulatory enforcement, lower will be the propensity towards business tax evasion and higher the incentive to adopt transparency enhancing computer technology.[8]

[Insert Figures 4 and 5 about here]

We further test whether the overall correlations reported above hold within more specific sub-groups. First we do a median split of states by the level of corruption and test whether computer adoption is lower in states with higher corruption levels. We also check if the relationship holds within sub-groups of retailers (traditional, modern and durables). The results are reported in Figure 6. We find that the computer adoption rate is lower by 5% to 12% in higher corruption states. Interestingly, the gap is larger among modern stores; suggesting that it is modern retailers who are being most strategic about transparency concerns when adopting computers. Figure 7 shows a similar graph for computer adoption rates as a function of enforcement levels. The gap is larger here; states with better enforcement have 30% to 40%

[8] Kerala (KL) and Bihar (BR) appear to be outliers in Figure 4. We therefore assessed the robustness of our results by dropping Kerala and Bihar from the states included in the analysis. All of the reported relationships in the paper continue to hold and the key results without the outlier observations are available in the online appendix.

higher adoption rates. Clearly, one has to control for other variables to quantify the effect of transparency and enforcement on computer adoption; but there is prima facie evidence here that higher enforcement and better governance are correlated with lower transparency concerns.

We also assessed the transparency concern by using another variable that indicates the competitive disadvantage issue of tax evasion. We use store managers' perceptions of the level of dishonesty among their peers in terms of hiding revenues for tax evasion. Figure 8 reports how computer adoption varies among different store types, based on the managers' perception of perceived dishonesty among peers. For all store types, managers who have not adopted computers in their stores believe that there is a higher level of tax evasion in the industry.

[Insert Figures 6-8 about here]

Finally, Figure 9 reports the relationship between productivity and computer adoption using our two productivity metrics: the revenues per employee (labor productivity) and revenues per sq. ft. (floor area productivity). The graphs show that productivity is higher for stores that adopt computers. Obviously, other variables need to be controlled for and there are potential selection concerns, which we now address in the subsequent statistical analyses.

[Insert Figure 9 about here]

4.2 Computer Adoption

We begin by discussing the findings from the probit computer adoption regression. The first set of results is for the full sample. The first column excludes gross annual revenues, while the second column includes it. The results in both columns are consistent with our hypotheses about transparency variables: corruption, enforcement/audits and regulatory consistency. The corruption variables are negatively related to adoption, while enforcement/audits and regulatory consistency are positively related to adoption. Thus the primary hypothesis of the paper—that computer adoption is systematically correlated with transparency concerns is supported.

As expected, higher labor costs are positively related to computer adoption as firms substitute technology for labor. Computer adoption is positively related to generator adoption, suggesting positive correlation in preferences for productivity enhancing technologies. Further, as expected from a cost affordability perspective, computer adoption is positively correlated with a store's gross annual revenues.[9] In terms of control variables such as state literacy rates (or

[9] As we report in Section 4.4, gross annual revenues have a reverse causal link to computer adoption through the productivity link. Here we do not account for this endgoeneity, as our primary interest here is on the transparency

HDI),[10] other store and management characteristics, they generally have the right sign. Higher literacy rates (or HDI) are correlated with greater computer adoption. Larger stores are more likely to adopt computers. Interestingly, older stores and managers with greater experience are less likely to adopt computers; suggesting (not surprisingly) that experience is negatively related to new technology adoption. The negative relationship with experience potentially captures the relative lack of comfort of older managers with computers. We revisit the net effect of managerial experience on productivity later based on our self-selection model estimation results in Section 4.6. Finally, we find that ownership characteristics (concentration or government ownership) are not related to computer adoption. Also, interestingly the power supply related factors do not have a significant relationship, suggesting that generators and power back up equipment are perhaps being used as appropriate to address power supply problems.

We conclude this section by assessing the heterogeneity in the effect of transparency concerns by firm size. We use a median split of firms by size and estimate the probit model of computer adoption separately for large and small firms. The direction of the estimates is qualitatively identical across firm size for all variables. However, the differences in magnitudes of the effects for small and large firms show that corruption and enforcement concerns systematically have larger impact on larger firm's adoption decisions than on smaller firms.[11] Specifically, corruption suppresses computer adoption more among larger firms, while enforcement increases computer adoption among larger firms. This suggests that the losses from corruption and the gains from enforcement are greater for larger firms. External auditors have a higher positive relationship with computer adoption among small firms relative to large firms.

[Insert Table 2 about here]

4.3 Falsification Check: Generator Adoption

---

variables. In Section 4.6, we estimate the gross annual revenue and computer adoption equations as a simultaneous equations model accounting for self-selection with appropriate exclusion restrictions, where we control for the endogeneity of the revenue and computer adoption variables. Our results in Table 8 remain robust.

[10] The correlation between literacy rate and HDI in our sample of 16 Indian states is high at .85. The high correlation is not surprising as education (proxied by literacy rate) is one of the three dimensions for HDI. Literacy rate is also strongly correlated with income per capita and life expectancy—the proxies for the other two dimensions, viz., economic development and health. To avoid multicollinearity, we use literacy rate and HDI separately in the regression analysis, but not together. The results are qualitatively identical. To conserve space, we present only the results using the literacy rate in the paper.

[11] Rather than estimate the probit separately for smaller and larger firms, one could have estimated a pooled regression with interaction terms between the relevant variables with a large or small firm dummy. This would have highlighted whether the differences are significant, but would not have given us a direct estimate of the effect for small and large firms. We estimated the pooled regression with interactions and all differences that we discuss here are indeed significant.

One interesting possibility is whether there are some unobserved characteristics which are correlated with transparency concerns that might drive productivity enhancing technology adoption. For example, states that have high corruption might systematically cause non-adoption of productivity enhancing technologies for reasons that might not be associated with transparency concerns. To address this issue, we conduct a falsification check. We choose a technology like generator, whose adoption increases productivity, but is not connected to transparency and test whether its adoption is linked to transparency concerns. Further, to assess face validity, we include literacy rate in the regression which is shown to impact computer adoption, but should not directly impact generator adoption. The results of the generator adoption regression for the full sample, and the large and small firms are presented in Table 3.

The falsification check is validated. The transparency variables i.e., corruption, enforcement and regulatory consistency become insignificant for generator adoption both in the aggregate as well as for small and large firms separately. On the other hand, the electric power related factors turn out highly significant; larger stores are more likely to use generators. One possibility is that electricity infrastructure is worse in less developed areas; in that sense the electric power factors might be capturing some other elements of transparency we have not considered. However, state literacy rate is insignificant as literacy is not required for generator adoption. Overall, this rules out the possibility that literacy, corruption/enforcement and electric power factors are all proxies for an unobservable development variable that might commonly affect all types of technology (computer and generator) adoption, and also productivity. This gives us greater faith in the transparency mechanism affecting computer adoption.

Further, it should be noted that some transparency related variables included in the regression are at the firm level, obtained through the survey (e.g., perceived informality by peers, external audit, etc.), while others are at the state level (e.g., TI corruption index, regulatory consistency). Neither the state level nor firm-specific transparency measures are significant in the generator adoption equation though they were significant in computer adoption. As transparency metrics at the state level are potentially correlated with other omitted factors such as lower levels of development and less-established infrastructure in the state, it is gratifying that not only the state, but also firm specific (local) transparency factors are both not significant.[12]

[12] We thank an anonymous reviewer for highlighting the importance of finding insignificant relationships not only at the state but also at the firm (local) level. As an additional robustness check, we also ran the probit regression of

[Insert Table 3 about here]

4.4 Impact of Computer Adoption on Productivity

We next report the results of the productivity regressions. We start with OLS regressions. Table 4a and Table 4b report the results of regression for labor productivity (revenue/employee) and floor area productivity (revenue/square foot). For each productivity variable, we report the results for the full sample and for large and small firms.[13] We begin with the results for the full sample. Productivity is higher among firms adopting computers as reflected by the positive coefficient on computer. Given that the dependent variable enters the regression equation in logs, the productivity multiplier on revenue per employee for firms adopting computers is 25% (exp (0.224) = 1.25), while on revenue per square foot is 29% (exp(0.259)).

The other variables in the regression have the expected signs for both metrics of productivity. Firms that adopt generators have higher productivity. Larger stores are more productive. Interestingly, manager experience is positively related to store productivity. Thus even though experience is negatively related to computer adoption and thus may be associated with lower productivity due to lack of computer adoption, the direct relationship between managerial experience and productivity is positive. However, concentrated ownership is negatively related to productivity, perhaps reflecting the fact that much of single ownership is driven by subsistence stores. In terms of store characteristics, having access to banking and financing is positively correlated with greater productivity. Use of in-store security is also positively correlated with productivity, as theft is widely considered a serious drain in retailing.

We have included a number of controls that are potentially correlated with transparency variables to rule out spurious relationships between transparency variables and productivity. Yet it is possible that there are other potentially omitted variables that could be correlated with the transparency variables, and have their own direct effect on computer adoption, but not on generator adoption. An example of such a variable is literacy rate. Demands from an educated workforce potentially reduce corruption and increases enforcement; which can increase computer adoption, but not generator adoption. We therefore estimate the model including literacy rate in

generator adoption with state level fixed effects by including only the firm level variables but excluding all the state level variables. The results are qualitatively identical to those in Table 3 for all the firm level variables in terms of their statistical significance levels and directionality; they are available in the online appendix. See Table 8b for the simultaneous equations MLE results on computer adoption with state fixed effects.

[13] Our analysis indicated statistically significant heterogeneity in gains from computer adoption by firm size, but not across other firm characteristics. Hence we focus only on large and small firm differences in the rest of the paper.

the productivity and computer/generator adoption regression. Literacy rates are significant in the computer adoption and productivity equation, but not in the generator adoption regression as noted earlier.[14] The fact that productivity enhancing effects of computers remain significant even after controlling for the effects of moderating variables such as literacy (or HDI) that are correlated not only with computer adoption and productivity, but also correlated with transparency variables, lends confidence to the conclusion.

[Insert Table 4a and Table 4b about here]

As before, we report the results of the productivity regressions for large and small firms based on a median split. The direction of the estimates is qualitatively identical for both large and small firms as in the full sample estimates for both labor and floor area productivity metrics. Comparing the results by firm size, larger firms gain more in terms of labor productivity from computer adoption (29% i.e., exp(0.253)) relative to smaller firms (24%). The corresponding numbers for floor area productivity are 28% and 24% respectively. Thus our results are consistent with the notion that productivity gains are larger for larger stores that require greater coordination. The other variables in the regressions have the same signs as before.

4.5 Propensity Score Matching

One concern with the OLS estimates on the effect of computers on productivity is that stores that adopt computers are systematically different from stores that do not affect computer adoption and therefore differences in productivity across the two groups may not be due to computer adoption. Matching methods, pioneered by Rosenbaum and Rubin (1983) and refined by Heckman and colleagues (e.g., Heckman et al. 1998; Heckman, Ichimura and Todd 1997, 1998), have been developed such that the outcomes of the treated (computer adopters) denoted by Y_1 are contrasted only against the outcomes of "comparable" untreated (non-adopters) denoted by Y_0 so that productivity differences can be attributed to the treatment (computer adoption). The basic idea of the matching method is discussed below.

Let I_0 and I_1 denote the set of indices for non-treated and treated respectively. To estimate a treatment effect for each treated firm $i \in I_1$, outcome Y_{1i} is compared against the average of outcomes Y_{0j} for all matched firms $j \in I_0$ among the untreated firms. Matches are

[14] The results are qualitatively identical when HDI is included in the regression in place of literacy rate.

based on observed characteristics that affect treatment (in our case, the variables that impact computer adoption as reported in Table 2). When the observed characteristics of the untreated firm is closer to that of the treated firm, based on an appropriate distance metric, that untreated firm gets a greater weight when constructing the match. Thus the estimated gain for each firm *i* in the treated sample I_1 is $Y_{1i} - \sum_{j \in I_0 \cap S_p} W(i,j) Y_{0j}$, where S_p is the set of firms in the region of common support across the treated and non-treated i.e., $S_p = \mathrm{Supp}(X \mid D=1) \cap \mathrm{Supp}(X \mid D=0)$ and $W(i,j)$ is an algorithm-specific weight based on the distance between the propensity scores for *i* and *j*. Let n_1 be the number of treated cases; the focal parameter of interest called the average treatment effect on the treated (ATT), reflecting the average effect of computer adoption on productivity is defined as:

$$\frac{1}{n_1} \sum_{j \in I_0 \cap S_p} \left(Y_{1i} - \sum_{j \in I_0 \cap S_p} W(i,j) Y_{0j} \right)$$

Specifically, we use kernel matching advocated in Heckman, Ichimura and Todd (1997) for constructing the weighting function based on the difference in propensity scores between firm *i* and *j*. Specifically the weighting function is given by:

$$W(i,j) = \frac{G((P_j - P_i)/h)}{\sum_{k \in I_0} G((P_k - P_i)/h)}$$

where *G*(.) is the kernel function, P_j is the propensity score of firm *j*, and *h* is a tuning parameter that specifies a bandwidth for the kernel function. Specifically, we report results based the Epanechnikov kernel $G(u) = 0.75(1-u^2) I(|u| \le 1)$ (see Leuven and Sianesi 2003). The kernel has a parabolic shape with support in the region [-1, 1].[15] We estimate the model using the PSMATCH2 module in Stata.

To perform propensity score matching (PSM), a critical requirement is the ability to match treated observations with non-treated observations through propensity scores. For this, a rich set of variables that can reasonably discriminate the treated and non-treated observations is necessary. Specifically, we estimate the propensity scores through probit analysis of computer

[15] We also estimated the effect using other matching methods such as nearest neighbor matching based on both propensity scores (Leuven & Sianesi 2003) as well as actor norm based "distances" between treated and control units (Abadie, Drukker, Herr and Imbens 2004). Our results are similar in magnitude across the different methods.

adoption using the same set of observed covariates that we report in the probit regressions of Table 2. Given missing data on different covariates, we have a sample size of 1501 with 324 treated and 1,177 non-treated cases. The probit model provides a good fit with the data with an adjusted R^2 of 0.52. Given the requirements of common support for the estimated propensity scores between treated and non-treated firms, the propensity matching involves a sample of 1,199 with 269 treated and 930 non-treated units. The region of common support is [.004, .999] and the mean propensity score for computer adopters is 0.58. Table 5 shows the distribution of the estimated propensity scores and the mean values of selected (and representative of key dimensions) variables used in propensity score matching for the treated and non-treated units by distinct block grouping.[16] The block grouping ensures that within each block (PS interval), the mean values of estimated propensity scores are very comparable between treated and non-treated units. The number of blocks, here 13, is generated by enforcing that condition using the algorithm by Becker and Ichino (2002) based on repeated splitting of each blocks starting with 5 equal interval initial blocks till the comparable condition is achieved. As expected, there are more untreated firms in the low propensity score blocks, while there are more treated firms in the high propensity score blocks, suggesting that the variables included for propensity matching does indeed help discriminate the firms on computer adoption.

Table 6 reports the estimates of the average treatment effects on the treated (ATT) based on both labor productivity and size productivity measures. We first report the results for the full sample. We find that computer adoption enhances labor productivity by about 51% (ATT value of 0.409) and floor productivity by 70% (ATT value of 0.528). Clearly, the OLS results substantially underestimate the productivity increase from computer adoption. Similarly, the ATT for large and small firms reported also show that OLS estimates are substantially biased downwards. Overall, the qualitative insights that larger firms gain more in productivity from computer adoption continue to hold.

So, why are the OLS estimates on the increase in productivity due to computer adoption biased downwards? At the margin, factors affecting computer adoption (including transparency) raises the threshold of productivity required for computer adoption. If we do not control for these factors, then the threshold for productivity required for computer adoption would be lower.

[16] The results for all the variables used in our analyses are given in the online appendix.

Therefore the OLS estimates that do not account for the factors affecting adoption have a downward bias in productivity increases.

[Insert Tables 5-6 about here]

Selection on Unobservables: Rosenbaum Bounds

Within the propensity score matching framework, sensitivity to potential selection on unobservables (hidden selection not captured in the observable variables used in propensity score matching) is assessed using Rosenbaum bounds (Rosenbaum 2002; DiPrete and Gangl 2004). The basic idea is to assess how much of the variance in unobservables needs to drive selection to negate the treatment effect; the higher the variance needed, the more confident we are that the qualitative results about the role of computer adoption are robust.

Table 7 shows the level of unobserved variance necessary to make the productivity enhancing effects of computers insignificant due to unobserved selection. The treatment effect becomes insignificant when $\Gamma > 1.9$ for the labor productivity measure and at about $\Gamma > 2.2$ for the floor productivity measure. In this context, the mean propensity score for computer adopters in our study is 0.58. Since our findings of the treatment effect on the labor productivity measure remain robust to unobserved selection effects till about $\Gamma = 1.9$, it indicates that unobserved selection bias will undermine our finding of the positive productivity impact of computer adoption if the mean propensity of computer adopters increase from 0.58 to 0.58x1.9 = 1.1. Similarly, since our findings of the treatment effect on the floor productivity measure remain robust to unobserved selection effects till about $\Gamma = 2.2$, it indicates that unobserved selection bias will undermine our finding of the positive productivity impact of computer adoption if the mean propensity of computer adopters increase from 0.58 to 0.58x2.2 = 1.3. Since it is highly unlikely that the probability of computer adopters will jump from 58% (based on observables alone) to 110% or 130% (including unobservable effects), our PSM based findings on the positive productivity impact of computer adoption are reasonably robust to hidden selection bias giving us faith in the finding that computer adoption indeed enhances productivity.

[Insert Table 7 about here]

Heterogeneous Treatment Effects

We next explore heterogeneous treatment effects across firms. We use the *Matched Smoothing Method of Estimating Heterogeneous Treatment Effects* (MS-HTE) described in Brand and Xie (2010) and Xie, Brand and Jann (2011). The propensity score matching results reported the Average Treatment effect for the Treated (ATT). As reported earlier, for labor productivity, the ATT is .409; therefore computer adoption enhances labor productivity for the treated by 51%. The Average Treatment effect for the Untreated (ATU) is .605; therefore computer adoption enhances labor productivity for the untreated by 83%. The Average Treatment Effect (across Treated and Untreated) is 0.561; therefore computer adoption enhances labor productivity on average by 75%. However there is heterogeneity in these effects.

The scatter plots of the estimated treatment effects against the propensity scores for the treated and the untreated are presented in Figure 10. A regression curve fitted on these scatter plots show that overall there is a decline in treatment effects for both treated and untreated groups with increase in propensity scores. However the slope is relatively flat and insignificant for the treated, while significantly negative for the untreated. More important, we find that when propensity scores are low, the productivity increases for untreated firms is greater than the productivity increases for treated firms. But when propensity scores are high, the productivity increases for untreated firms is lower than the productivity increases for treated firms. This is consistent with the argument we made earlier that at the margin, factors affecting computer adoption (including transparency) raise the threshold of productivity required for computer adoption. Thus the productivity increases for the untreated is higher when the propensity scores are lower—despite potentially high gains in productivity, other factors reduce the propensity to adopt computers. However when such impediments to adoption are lower (as when the propensity score is higher), the productivity increases for treated firms are higher. On average, as there are far more untreated firms, and they tend to be at the low end of the propensity score spectrum, we find that ATU > ATT.

[Insert Figure 10 about here]

4.6. Modeling Self-Selection

With Rosenbaum bounds, we considered selection on unobservables, as if the selection is random. But if selection is non-random as is the case with computer adoption, one has to model self-selection. To address this concern, we next estimate the effect of computer adoption on firm productivity using Heckman's model of self-selection.[17]

[17] The propensity score matching literature argues that the Rosenbabum bounds approach does not require normal distribution assumptions and is non-parametric, unlike the bivariate normal Heckman selection model. Another advantage is that the Heckman selection model requires exclusion restrictions involving instruments—variables than impact selection, but not outcomes-which may not always be available. Further, these estimates are not the average

Let y_i denote the outcome, i.e., productivity of firm $\underline{i}$ and x_i be factors that impact productivity. Let w_i be the self-selected choice of firm *i* to adopt computers and z_i denote factors that affect the decision of firm *i* to adopt computers. Let w_i^* be a latent variable indicating the incremental value obtained by firm *i*, by adopting computers and $w_i = 1 \text{ if } w_i^* > 0 \text{ and } w_i = 0 \text{ otherwise}$. The Heckman self-selection model is described by the following two equations, the outcome and selection equations, respectively.

$$y_i = x_i\beta + \alpha w_i + \varepsilon_i$$

$$w_i^* = z_i\gamma + v_i$$

where $\begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}\right]$.

The bivariate normal model above can be estimated by Heckman's two step estimation approach or maximum likelihood (Maddala 1983). We estimate the model using maximum likelihood. We note that revenue is an endogenous variable in the computer adoption equation, and computer adoption is an endogenous variable in the revenue equation. As there are many exogenous variables, that are only present in either the revenue or the computer adoption equation and not both, the system is identified.

We report the estimation results in Table 8. While Table 8a reports the results of labor productivity equation, Table 8b reports the results of the computer adoption selection equation. Table 8c reports the results of the floor area productivity equation. We suppress the results of the computer adoption selection equation corresponding to floor area productivity as these results are virtually identical to Table 8b. We run each of the self-selection models in Tables 8a-c with and without the state level fixed effects. The models with state level fixed effects include only the firm level variables but exclude all the state level variables, and allow us to rule out any kind of potential state level unobservables. As comparison of the results across the columns in Tables 8a-c show, they remain virtually identical with or without controlling for the state level fixed

treatment effect across all firms, but a local average treatment effect (LATE) over the firms whose decision to adopt computers are affected by the instruments. When selection on unobservables is not random, and exclusion restrictions are available, the selection model is preferred.

effects.[18] Also, the signs of the variables in productivity equations in Table 8 remain essentially the same as in the OLS results reported in Table 4. However, considerable bias exists in the OLS estimates of computer adoption on productivity. Importantly, the correlation between the outcome and selection term is negative and significant with values between -0.42 to -0.53 across the different models, suggesting the importance of unobserved selection.

[Insert Table 8 about here]

Using the full sample self-selection model for labor productivity, we elaborate on the estimated average marginal effects of a few variables of interest that indirectly impact productivity through their effect on computer adoption.[19] For example, an external auditor for a retailer is related to 6.98% higher probability of computer adoption, which in turn translates to a 24.11% higher labor productivity based on simultaneous estimation of productivity and selection model. Similarly a standard deviation increase in the "regulatory inspections" and "regulatory consistency" variables is related to a 9.20% and 5.02% higher probability of computer adoption, which in turn translate to 11.75% and 4.57% higher labor productivity. Finally, we consider managerial experience that affects both computer adoption and productivity. The main effect of managerial experience on store productivity is positive; but the moderating effect of managerial experience on productivity through computer adoption is negative. The net effect of a standard deviation increase in managerial experience on productivity is higher by 17.84%, after accounting for the negative effect on computer adoption.

To facilitate comparison across the different estimation approaches, the estimated effects of computer adoption on productivity using OLS, Propensity Score Matching (PSM) and the Self-Selection Model (SSM) are reported in Table 9. The estimates using SSM are close to those from the PSM but substantially greater than those from the OLS. To be specific, in terms of labor productivity, while OLS estimates a 25% improvement, PSM estimates a 50% improvement and the SSM estimates a 60% improvement on the treated firms. The heterogeneous treatment effects model estimates the average treatment effect on the untreated as 83% and the average treatment effect across treated and untreated is 75%. Across all the results, it is clear that that not accounting for transparency leads to significant underestimation of the productivity improvement

[18] To conserve space, we have reported here the results of self-selection models with state fixed effects only for the full sample. The corresponding results by store size are qualitatively identical and are available in the online appendix.

[19] We use the margins command in Stata to compute the average marginal effects reported.

from computer adoption. The results are directionally and substantively consistent for floor area productivity and also across small and large firms.

[Insert Table 9 about here]

5. Conclusion

The tendency of firms to avoid productivity enhancing technologies and remain small due to transparency concerns has been dubbed the "Peter Pan Syndrome" in emerging markets. Though IT enhances productivity, the "culture of informality" in emerging markets causes businesses to fear IT because they remove the "veil of secrecy" around business practices that is conducive for tax evasion. This paper investigated whether emerging market firms make the tradeoff between productivity and transparency in adopting IT.

Specifically, the paper studied the productivity-transparency tradeoff in the Indian retail sector using data from a large scale national survey of 1948 Indian retailers augmented with other relevant data on corruption, enforcement and other state level control variables. We find that IT adoption is significantly affected by transparency concerns. While corruption reduces IT adoption, enforcement and auditing increases IT adoption by providing all firms a level playing field and reducing the negative impact of corruption. IT adoption increases store productivity on an average by about 50 to 70 percent. The effects of transparency on IT adoption and the impact of adoption on productivity are both greater for larger than for smaller firms. At the margin, higher corruption and lower enforcement raises the threshold of productivity required for IT adoption.

Our results are relevant to transparency enhancing IT businesses, governments and policy makers. As growth in the developed world stagnates, firms are increasingly reliant on emerging markets for their growth. To the extent that the market potential for IT among businesses is linked to the extent to which they enhance productivity, our results show that corruption and enforcement levels in a market impact not only unit sales, but also the willingness to pay (and therefore the price) in emerging markets. For governments and policy makers, our results suggest that forceful enforcement and corruption reduction can not only have a direct positive impact on tax collection, but also an indirect positive impact on the tax revenue base. The latter impact occurs through greater productivity induced by the use of modern efficiency enhancing technologies and by bringing more businesses into the transparent formal sector. Our work shows that modeling institutional characteristics of emerging markets can enhance the relevance of academic research for managers and policy makers in these markets.

We conclude with a discussion of the limitations of the paper that provide possibilities for future research. First, the India findings need to be replicated in other emerging markets. Second, the results are based on a cross-sectional data set. We used a variety of statistical tools to make appropriate inference using cross-sectional data - e.g., propensity score matching, heterogeneous treatment effects and instrumental variable methods that account for self-selection to measure the productivity effects of IT. We assessed whether the effects of transparency on IT adoption are robust through a falsification test and allowing for state level fixed effects to account for other potential omitted state level factors that may drive IT adoption. Future research should replicate our key findings with panel data that ideally have some form of experimental or quasi-experimental variation in transparency due to changes in policy or regulations.

Even though we did not find significant heterogeneity in gains from computer adoption beyond firm size, it would be useful in future research to explore firm characteristics that can drive differences in gains from computer adoption. Finally, we modeled computer adoption as a discrete variable in assessing productivity. Future research should focus on assessing the effect of IT spending rather than of merely IT adoption as a discrete variable. It would also be of interest to understand how investments in IT can impact retail prices as IT lowers marginal costs, but also increases fixed costs. We hope that our study, serves as a stimulant for further academic research on these important research questions.

In conclusion, we note that given the strong gains in productivity from IT adoption, we agree – albeit partly – with Sunder (2012) that “Indian retailers can and should break out of the self-defeating confines of the beliefs about the profitability of tax evasion” thus avoiding the “informality trap of lower productivity.” But curing the Peter Pan Syndrome among Indian retailers would require the government to improve the business environment to be free from corruption, and enhance the level and consistency of enforcement. As India opens up its markets to multinational, multi-brand retail, the need to increase productivity becomes even greater for domestic retailer survival. We hope our work encourages greater investment in productivity enhancing technologies by Indian retailers, as they prepare themselves for new levels of competition (Reardon and Gulati 2008).

References

Abadie, A., D. Drukker, J. L. Herr and G. W. Imbens (2004), "Implementing Matching Estimators for Average Treatment Effects in Stata," *The Stata Journal*: Vol. 4 No. 3: 290-311.

Amin, M. (2010), "Computer Usage and Labor Regulations in India's Retail Sector," *Journal of Development Studies*, Vol. 46, No. 9: 1572-1592.

Angelini, P. and A. Generale (2008), "On the Evolution of Firm Size Distributions," *American Economic Review*, 98(1), 426-438.

Azuma, Y. and H. Grossman (2002), "A Theory of the Informal Sector," *NBER Working Paper*, No. 8823.

Bartel, Ann, C. Ichniowski, and K. Shaw (2005), ''How Does Information Technology Really Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement and Worker Skills,'' *NBER Working Paper,* No. 11773.

Barua, A, C. H. Kriebel, and T. Mukhopadhyay (1995), "Information Technology and Business Value: An Analytic and Empirical Investigation," *Information Systems Research,* Vol. 6(1), 3-23.

Basu, S., J. G. Fernald, N. Oulton, and S. Srinivasan (2003), 'The Case of the Missing Productivity Growth: or, Does Information Technology Explain Why Productivity Accelerated in the United States But Not in the United Kingdom?'' in Mark Gertler and Kenneth Rogoff (Eds.), Macro Economics Annual, Cambridge, MA: MIT Press.

Becker, Sascha O. and Andrea Ichino. 2002. "Estimation of Average Treatment Effects Based on Propensity Scores." *The Stata Journal*: Vol. 2 No. 4: 358-377.

Black, S. and L. Lynch (2001), ''How to Compete: The Impact of Workplace Practices and Information Technology on Productivity,'' *Review of Economics & Statistics*, 83(3), 434–445.

Bird, R. M. and E. M. Zolt (2008), "Technology and Taxation in Developing Countries: From Hand to Mouse," *National Tax Journal*, Vol. LXI, No. 4, Part 2 (December), 791-821.

Brand, J. E. and Y. Xie (2010)," Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education," *American Sociological Review*, 75(2).

Brynjolfsson, E and L. Hitt (1996), "Paradox Lost? Firm-level Evidence on the Returns to Information Systems Spending," *Management Science,* 42(4), 541-558.

Chopra, A. (2011), "India's Push to Tame Tax Evasion," *The National,* June 18.

Commander, S., R. Harrison, and N. Menezes-Filho (2011), "ICT and Productivity in Developing Countries: New Firm-Level Evidence from Brazil and India," *Review of Economics & Statistics*, 93(2), 528-541.

Cull, R. and L. C. Xu (2005), "Institutions, Ownership, and Finance: The Determinants of Profit Reinvestment among Chinese Firms," *Journal of Financial Economics*, 77(1), 117-146.

Dabla-Norris, E., Gradstein, M. and G. Inchauste (2008), "What Causes Firms to Hide Output? The Determinants of Informality," *Journal of Development Economics*, 85, 1-27.

Dasgupta, S., J. Sarkis, & S. Talluri (1999) "Influence of information technology investment on firm productivity: a cross-sectional study", *Logistics Information Management*, 12(1/2), 120 – 129.

DeLone, W.H. (1981), "Firm size and characteristics of computer use," *MIS Quarterly*, 5(4), 65-77.

Dewett, T. & Jones, G. R. (2001), "The role of information technology in the organization: a review, model, and assessment," *Journal of Management*, *27*(3), 313-346.

DiPrete, Thomas and Markus Gangl. 2004. "Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments." *Sociological Methodology* 34:271-310.

Dhara, T. and C. Thomas (2011), "In India, Tax Evasion Is a National Sport," *Bloomberg Businessweek*, July 28.

Fabiani, S., Schivardi, F., & Trento, S. (2005), "ICT adoption in Italian manufacturing: firm-level evidence," *Industrial and Corporate Change*, *14*(2), 225-249.

Farrell, D. (2004), "Boost Growth by Reducing the Informal Economy," *The Asian Wall Street Journal*, October 18.

Foster, L., J. Haltiwanger, and C.J. Krizan (2002) "The Link between Aggregate and Micro Productivity Growth: Evidence from Retail Trade", *NBER Working Paper*, No.1920.

Friedman, E., S. Johnson, D. Kaufman, and P. Zoido-Lobaton (2000), "Dodging the grabbing hand: the determinants of unofficial activity in 69 countries," *Journal of Public Economics*, 76, 459–492.

Gatti, R. and M. Honoratti.(2008), "Informality Among Formal Firms: Firm-Level, Cross-Country Evidence on Tax Compliance and Access to Credit," *Policy Research Working Paper No. 4476*, World Bank, Washington, DC.

Government of India (2011), "*India Human Development Report 2011: Towards Social Inclusion*," Planning Commission, Oxford University Press.

Harris, S. E., & Katz, J. L. (1991), "Firm size and the information technology investment intensity of life insurers," *MIS Quarterly*, 15(3), 333-352.

Heckman, J., H. Ichimura, and P. Todd (1997): Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, 64, 605-654.

Heckman, James J., Hidehiko Ichimura, and Petra Todd. 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies*, 65, 261-294.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998), Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017-1098.

Ichniowski, C., K. Shaw, and G. Prennushi (1997), 'The Effects of Human Resource

Management Practices on Productivity,'' *American Economic Review*, 86, 291–313.

International Tax Compact (2010), "Avoiding tax evasion and tax avoidance in developing countries," *German Federal Ministry for Economic Cooperation and Development* (BMZ), December.

Johnson, S., D. Kaufman, J. McMillan, and C. Woodruff (2000), "What do firms hide? Bribes and unofficial activity after communism," *Journal of Public Economics*, 76, 495–520.

Jorgenson, D. (2001), ''Information Technology and the US Economy,'' *American Economic Review*, 91, 1–32.

Lefebvre, L. A., Lefebvre, É., Elia, E., & Boeck, H. (2005), "Exploring B-to-B e-commerce adoption trajectories in manufacturing SMEs," *Technovation*, *25*(12), 1443-1456.

Leuven, Edwin and Barbara Sianesi (2003), "PSMATCH2 – Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing." Software: http://ideas.repec.org/c/boc/bocode/s432001.html.

Love, P. E., Irani, Z., Standing, C., Lin, C., & Burn, J. M. (2005), "The enigma of evaluation: benefits, costs and risks of IT in Australian small–medium-sized enterprises," *Information & Management*, *42*(7), 947-964.

Loveman, G. W. (1994), "An Assessment of the Productivity Impact on Information Technologies," in T. J. Allen and M. S. Scott Morton (Eds.), *Information Technology and the Corporation of the* 1990s: *Research Studies*, MIT Press, Cambridge, MA.

Maddala, G. S. 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge, U. K.: Cambridge University Press.

Marcouiller, D. and L. Young, L. (1995), "The black hole of graft: the predatory state and the informal economy," *American Economic Review*, 85, 630–646.

Mesfin, M. (2012), "Ethiopia: Tax Authority Unhappy with Lack of Cash Register Servicing," *Addis Fortune*, March 4.

Miller, A. R., & Tucker, C. E. (2014), "Electronic discovery and the adoption of information technology," *Journal of Law, Economics, and Organization*, 30(2), 217-243.

Mishra, P., A. Subramanian, and P. Topalova (2008), "Tariffs, enforcement, and customs evasion: Evidence from India," *Journal of Public Economics*, 92, 1907-1925.

Morgan, A., Colebourne, D., & Thomas, B. (2006), "The development of ICT advisors for SME businesses: an innovative approach," *Technovation*, *26*(8), 980-987.

Morrison, C. J. and E. R. Berndt (1990), "Assessing the Productivity of Information Technology Equipment in the U.S. Manufacturing Industries," *NBER Working Paper*, No. 3582.

Nayar, K. (2011), "Laundering Black Money," *Deccan Herald*, Feb. 4.

People's Daily (2000), "China Battles Tax Evasion with Help of Computers," Dec. 4 – accessed from http://english.peopledaily.com.cn

Rao, V. V. (2010), "Black, Bold and Bountiful," *The Hindu Business Line*, Aug. 13.

Reardon, T. and A. Gulati (2008), "The Supermarket Revolution in Developing Countries," *International Food Policy Research Institute (IFPRI)*, Policy Brief 2, June.

Rosenbaum, P., and D. Rubin (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-50.

Russell, B. (2010), "Revenue Administration: Managing the Shadow Economy," Technical *Notes and Manuals No. 10/14*, International Monetary Fund, Washington, DC.

Schneider, F. and D. Enste (2002), "Shadow economies: size, causes and consequences," *Journal of Economic Literature*, 38, 77–114.

Sinha, B. (2003), "Pan masala unit raided for Rs. 6-cr sales tax evasion," *The Times of India*, Oct. 23 – accessed from http://articles.timesofindia.indiatimes.com/

Smith, M. G. (2013), "Doing It under the Table: Unreported Sales in India's Manufacturing Sector," *Working Paper, U.S. Department of Treasury*, available at SSRN: http://ssrn.com/abstract=220822.

Stiroh, K. (2002), ''Information Technology and the US Productivity Revival: What Do the Industry Data Say?'' *American Economic Review*, 92, 1559–1576.

Stiroh, K. (2004), ''Reassessing the Role of IT in the Production Function: A Meta-Analysis,'' *Federal Reserve Bank of New York*, mimeograph.

Sunder, S. (2012), "India's Retarded Retail Sector*," Live Mint & The Asian Wall Street Journal*, Jan. 8.

The Daily Star (2007), "Electronic Cash Registers to be Made Mandatory", Dec. 17, -- accessed from http://archive.thedailystar.net/newDesign/news-details.php?nid=15744

Thong, J. (1999), "An integrated model of information systems adoption in small businesses," *Journal of management information systems*, *15*(4), 187-214.

Transparency International (2012), "*Corruption Perceptions Index 2012*," www.transparency.org/poli/cpi/

Xie, Y., J. E. Brand, and B. Jann (2011), "Estimating Heterogeneous Treatment Effects with Observational Data," *Research Report II-729 (Feb.),* Population Studies Center, Institute of Social Research, University of Michigan.

Figure 1: Adoption level of business computers by store type (% of stores)

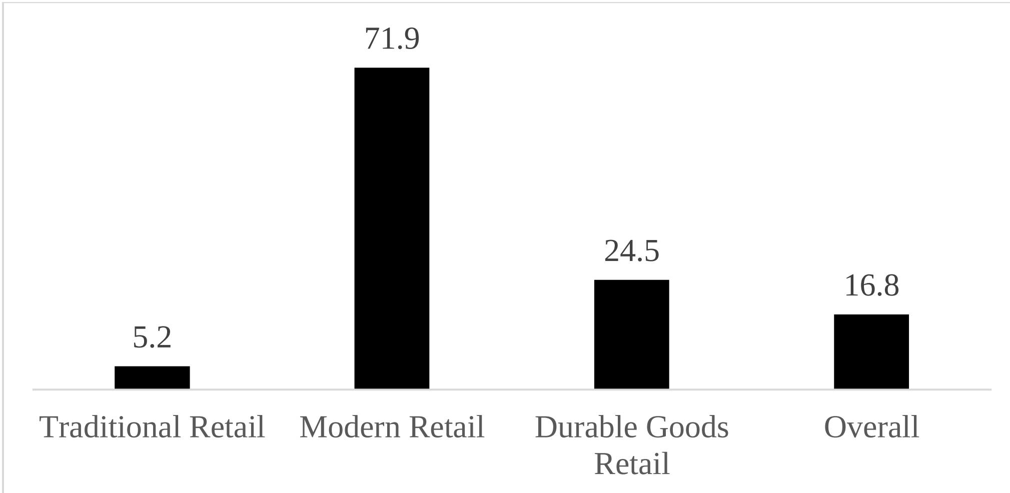
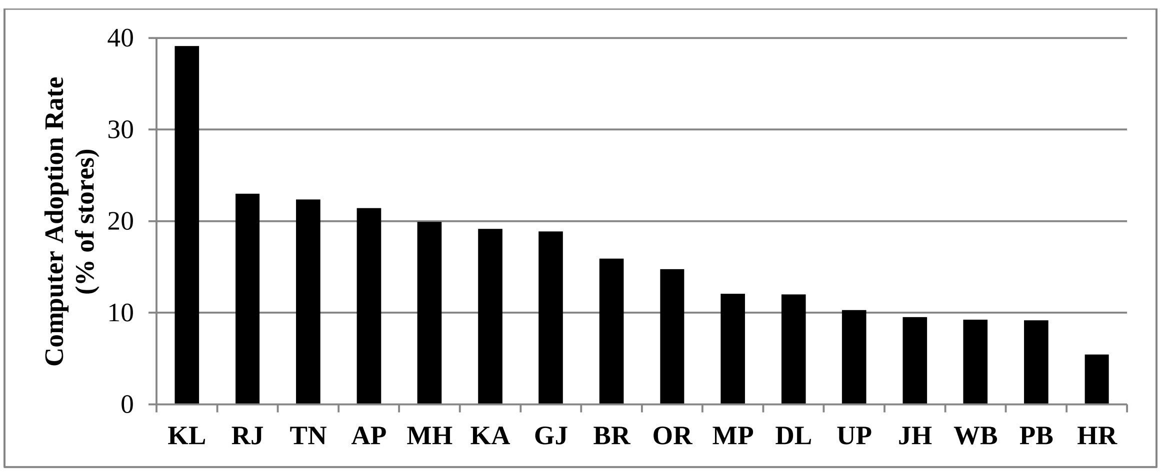


Figure 2: Adoption level of business computers by state



AP- Andhra Pradesh; BR-Bihar; DL-Delhi; GJ-Gujarat; HR-Haryana; JH-Jharkand; KA-Karnataka; KL-Kerala; MP-Madhya Pradesh; MH-Maharashtra; OR-Orissa; PB-Punjab RJ-Rajasthan; TN-Tamilnadu; UP-Uttar Pradesh WB-West Bengal

Figure 3: Relative inspection versus corruption levels across the sample states

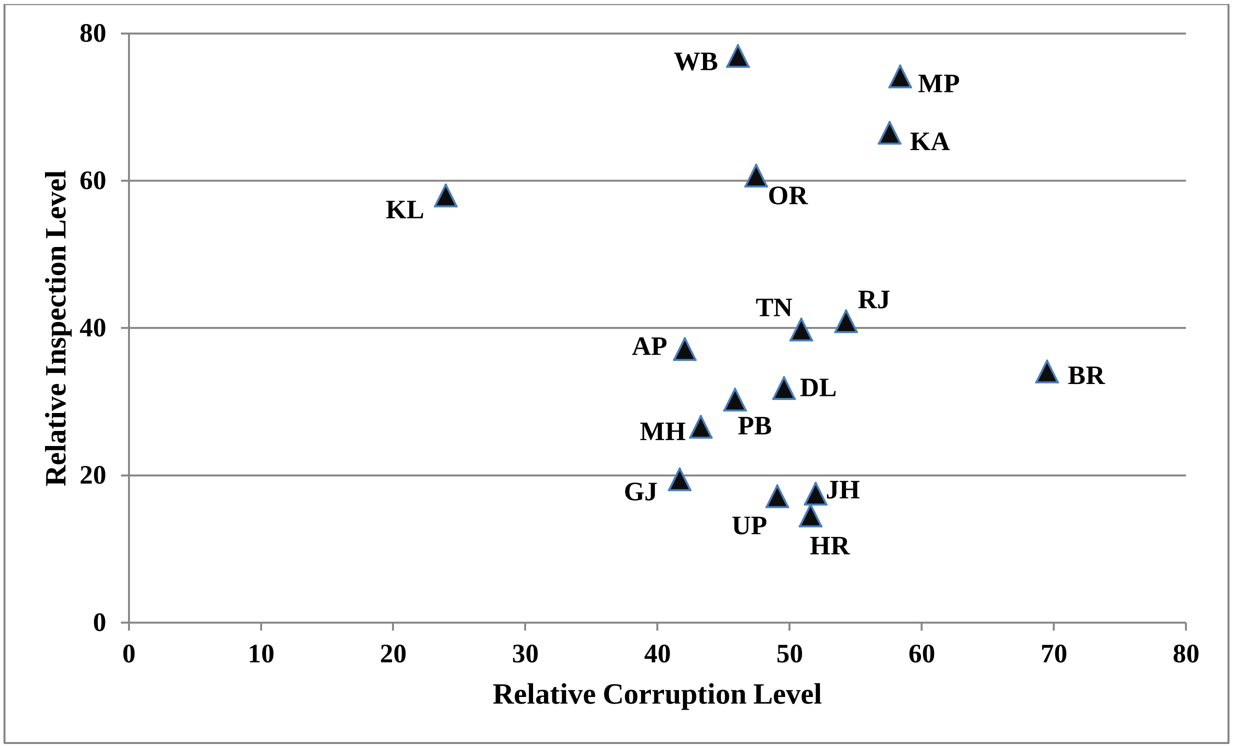


Figure 4: Corruption level and computer adoption by state

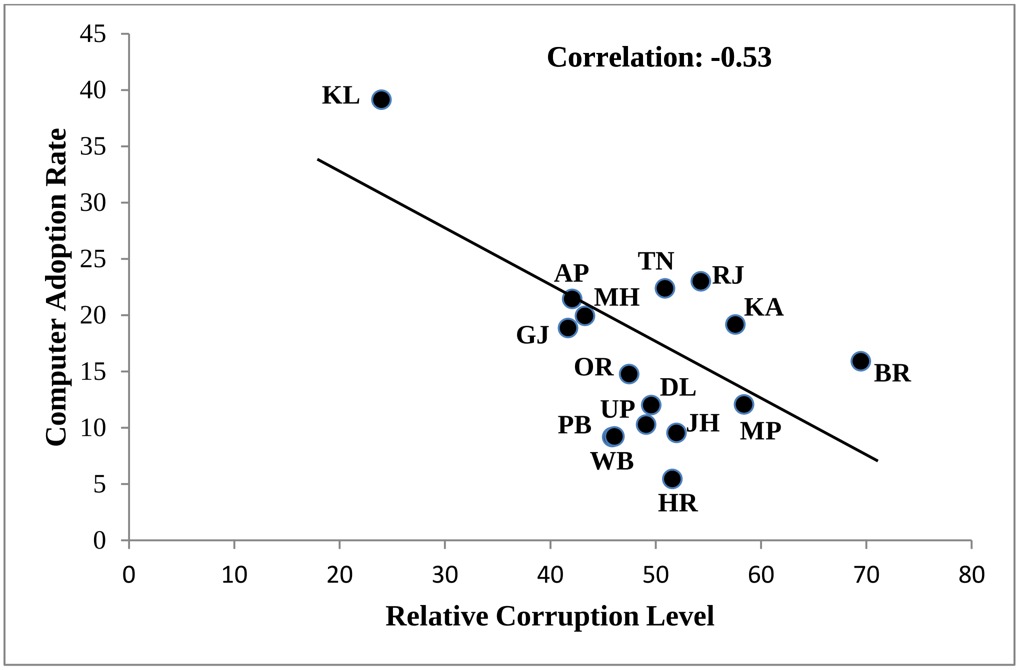


Figure 5: Inspection level and computer adoption by state

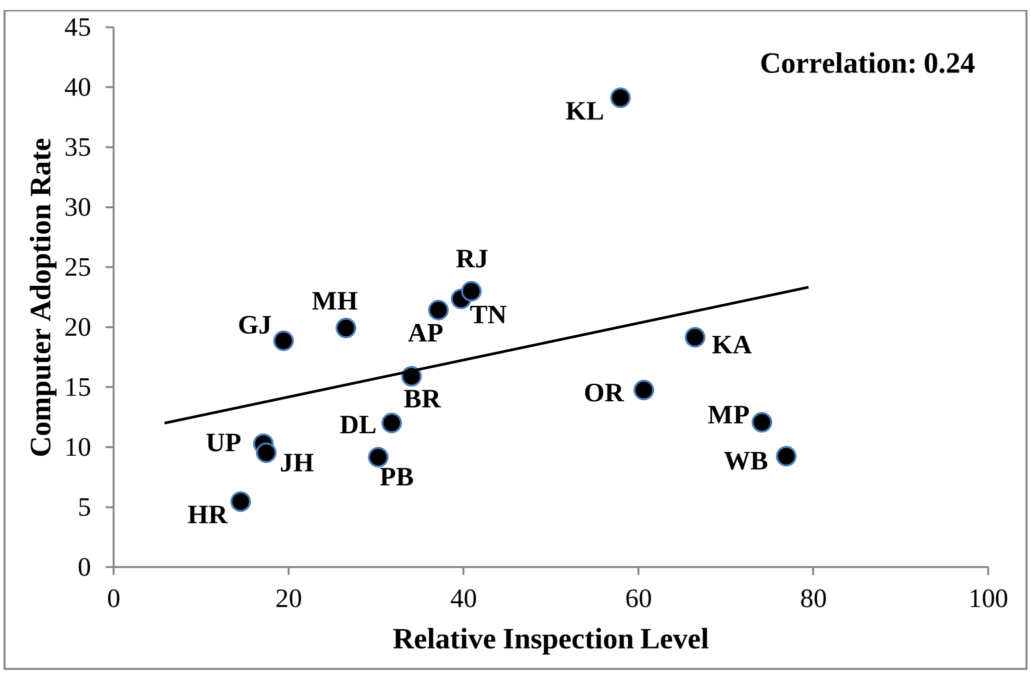


Figure 6: Computer adoption by median split of states by corruption level

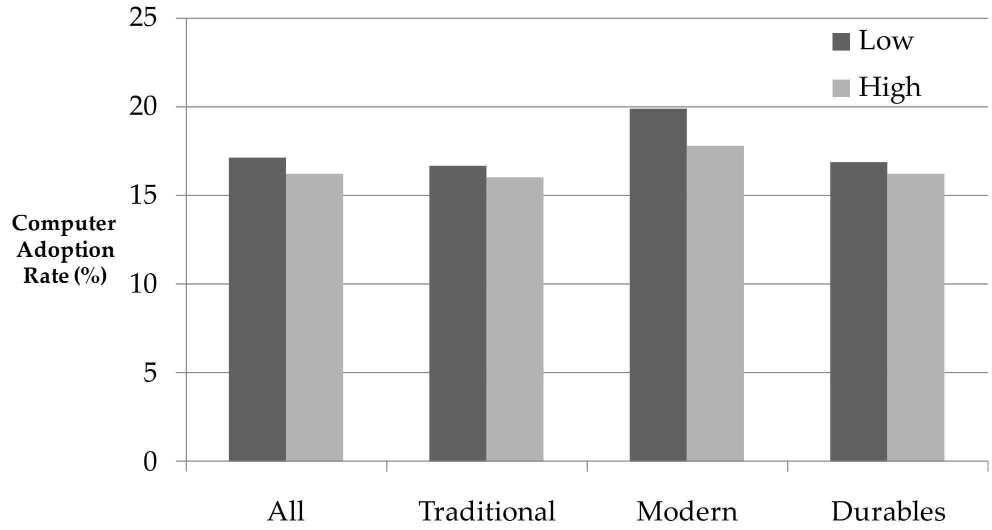


Figure 7: Computer adoption by median split of states by inspection level

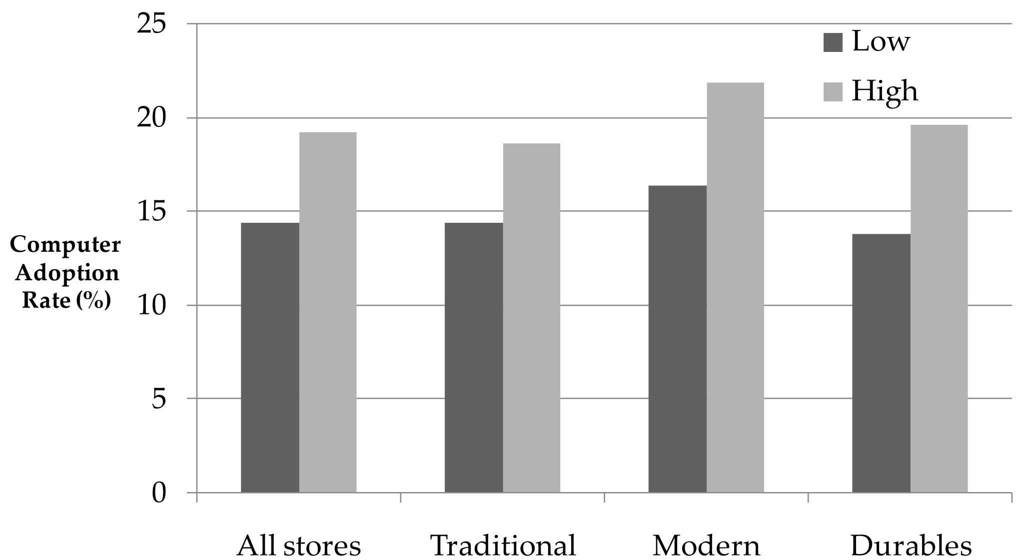


Figure 8: Perceived dishonesty among peers – by computer adoption across store types

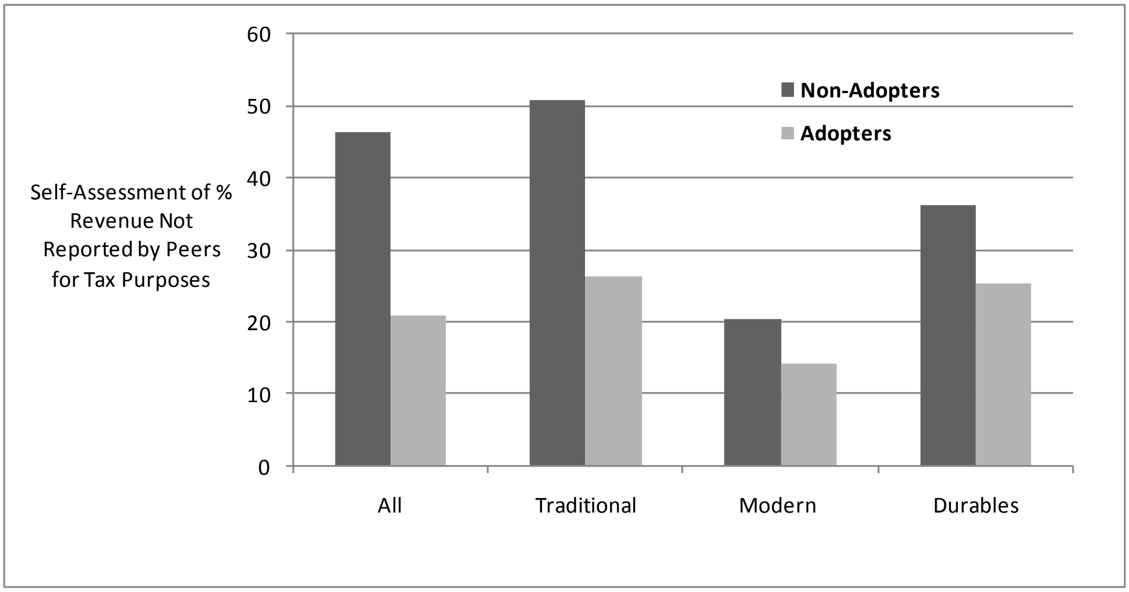
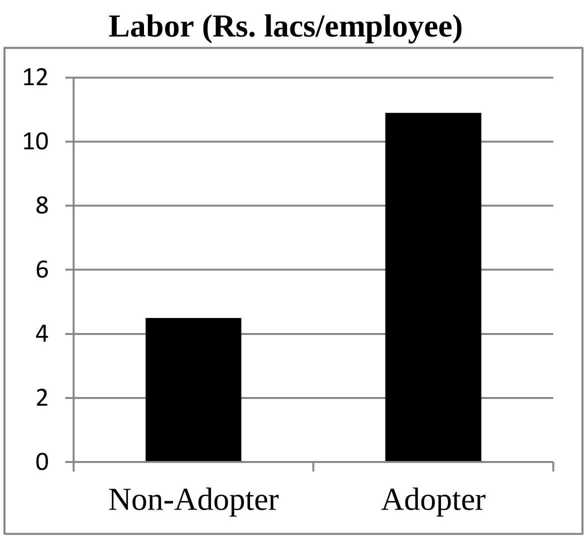


Figure 9: Productivity of stores by computer adoption status



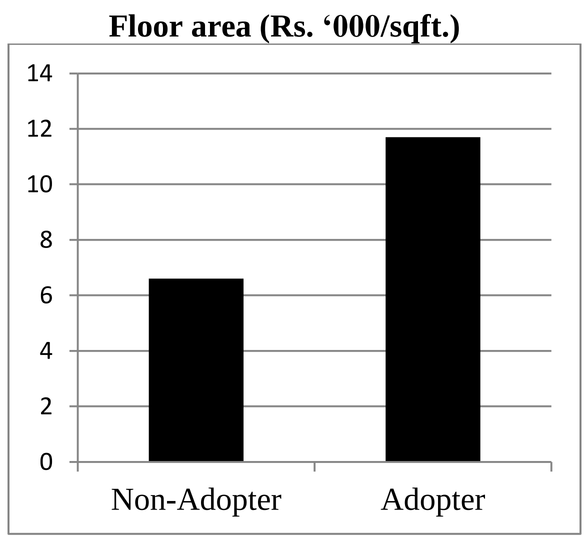


Figure 10: Heterogeneity in treatment effects

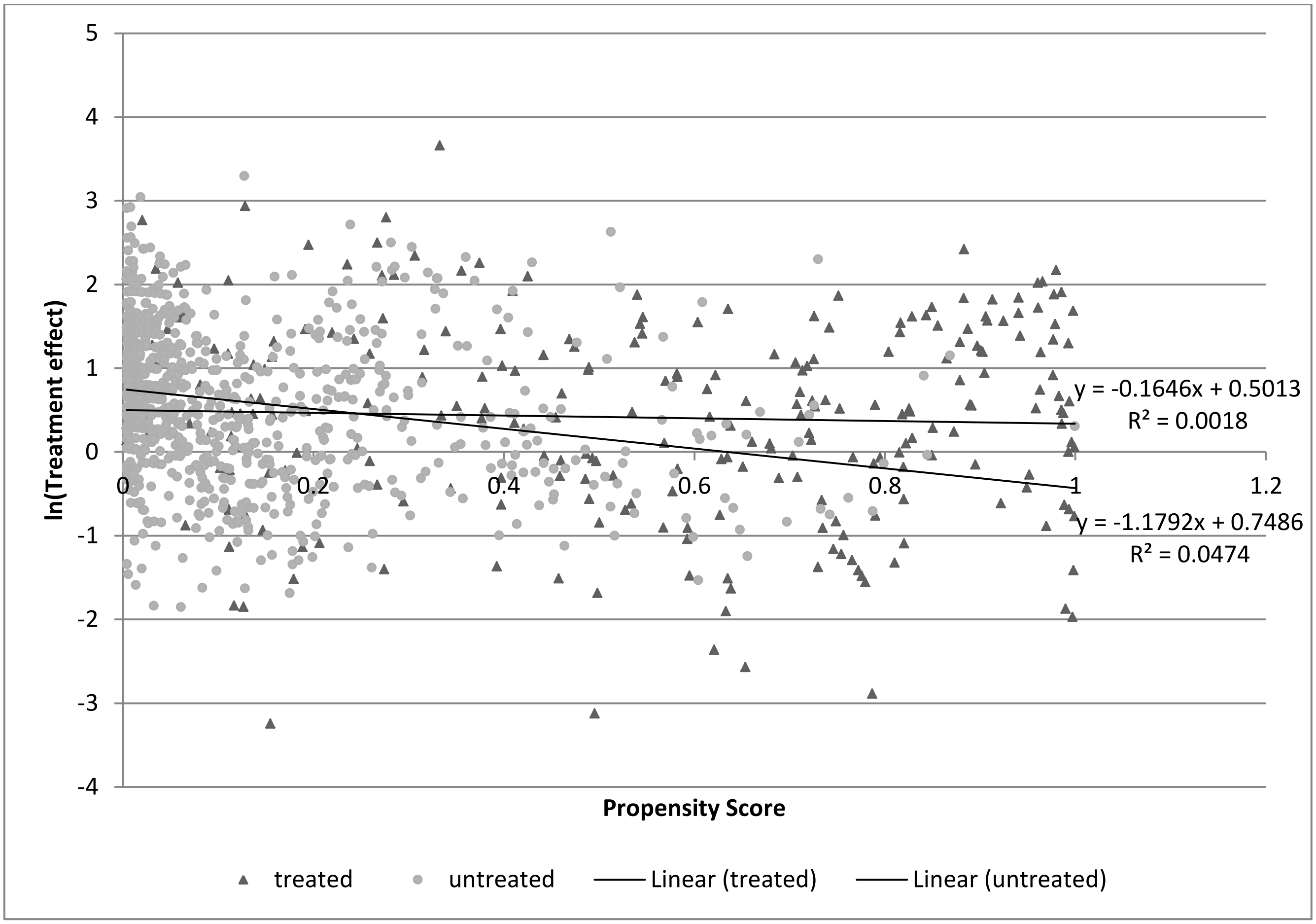


Table 1: Summary Statistics of the Analysis Variables[1]

Description	N	Mean	SD
Store level performance measures			
Gross revenue generated – latest financial year (Rs. in million)	1918	1.900	4.081
Revenue net of operational costs – latest financial year (Rs. in million)	1849	1.619	3.105
Productivity enhancing technology adoptions			
Business computer (0= No; 1 = Yes)	1948	0.167	0.373
In-store electricity generator (0= No; 1 =Yes)	1948	0.296	0.456
Corruption factors discouraging transparency			
Self-assessment of % of revenue typically reported by peers for tax purposes	1669	58.148	39.040
Self-assessment of % of revenue typically used to bribe regulatory agencies	1808	0.835	2.366
Transparency International (TI) Corruption Index at the state level (1-10)[2]	1948	4.811	0.769
Enforcement factors encouraging transparency			
Number of times the store was inspected last year by state regulatory agencies	1948	1.512	3.580
Store has an external auditor (0= No; 1 = Yes)	1914	0.302	0.459
Perceived consistency in state's regulatory implementations (1=Low; 6=High)	1948	3.096	0.608
Other state level variables			
Labor cost in terms of minimum wages rate (Rs.)[3]	1948	73.383	13.421
Literacy rate (percentage)[4]	1948	72.922	8.335
Human Development Index (0-1)[5]	1948	0.509	0.095
Electricity power supply related factors			
Faced power outage over the last year (0=No; 1 = Yes)	1944	0.829	0.377
State's power supply as a perceived obstacle to business (0= No; 4=Severe)	1948	1.635	0.468

Table 1 (Contd.): Summary Statistics of the Analysis Variables

Description	N	Mean	SD
Store size and age characteristics			
Floor area of the store (sq. ft.)	1938	599.811	3553.710
Number of full time employees at the store	1948	5.722	24.557
Age of the store (years)	1948	14.478	12.796
Store management and ownership characteristics			
Experience of the store manager (years)	1948	12.948	9.803
Ownership concentration (% of store owned by the largest owner)	1948	96.073	16.056
Government owned store (0= No; 1 = Yes)	1948	0.011	0.103
Store finance, in-store security and competitive factors			
Business bank account (0= No; 1 = Yes)	1940	0.639	0.481
Overdraft facility (0= No; 1 = Yes)	1921	0.223	0.416
In-store security system (0= No; 1 = Yes)	1947	0.266	0.442
Perceived level of price competition (0= Low; 1 = High)	1901	0.376	0.484
Inventory level maintained for the main product (days)	1948	11.582	16.167

[1] Unless specifically noted, the data source for a variable is the World Bank 2006 survey of Indian retailers.

[2] Data source: *"India Corruption Study 2005,"* Transparency International, Centre for Media Studies, India.

[3] Data source: "*Report on the Working on the Minimum Wages Act of 1948 for the Year 2001*," Labor Bureau, Government of India. Accessed at http://www.labourbureau.nic.in/MW2K1%20Main%20Page.htm

[4] Data source: Average of the states' adult literacy rates from the 2001 and 2011 census data, Government of India.

[5] Data source: "*Human Development Report 2011*," Government of India.

Table 2: Business Computer Adoption - Probit Regression

Variables	All Stores-I Coeff.	All Stores-I SE	All Stores-II Coeff.	All Stores-II SE	Larger Stores Coeff.	Larger Stores SE	Smaller Stores Coeff.	Smaller Stores SE
Corruption factors discouraging transparency								
State TI Corruption Index	-0.169**	0.076	-0.172**	0.083	-0.179**	0.085	-0.127**	0.061
Percent revenue spent on bribe	-0.020**	0.009	-0.026**	0.011	-0.026**	0.012	-0.017**	0.008
Perceived informality by peers	-0.003*	0.002	-0.003*	0.002	-0.002*	0.001	-0.003*	0.002
Enforcement factors encouraging transparency								
Regulatory inspections	0.033**	0.014	0.034**	0.015	0.038**	0.017	0.024**	0.010
External auditor	0.321**	0.139	0.293**	0.119	0.246**	0.106	0.448**	0.202
State's regulatory consistency	0.026**	0.012	0.028**	0.013	0.027**	0.012	0.026**	0.012
Labor cost and education level								
State minimum wages	0.041*	0.021	0.032*	0.017	0.038*	0.020	0.009*	0.005
State literacy rate	0.076**	0.036	0.071**	0.033	0.074**	0.035	0.062**	0.031
Electric power supply related factors								
Power outage	-0.058	0.091	-0.051	0.094	-0.055	0.101	-0.052	0.094
State power supply problem	-0.067	0.072	-0.076	0.078	-0.077	0.073	-0.070	0.072
Productivity enhancing technology adoptions								
Generator	0.529**	0.254	0.508***	0.152	0.609***	0.152	0.541*	0.281
Store level performance measure								
Gross annual revenue	--	--	0.064***	0.022	0.062***	0.021	0.119**	0.056
Store characteristics								
Store size	0.312***	0.103	0.313***	0.098	0.225**	0.105	0.411***	0.142
Employee size	0.387***	0.088	0.401***	0.108	0.406***	0.106	0.293***	0.087
Store age	-0.185**	0.091	-0.186**	0.093	-0.222**	0.109	-0.182**	0.091
Store management & ownership characteristics								
Manager experience	-0.162**	0.078	-0.156**	0.073	-0.153**	0.071	-0.125**	0.058
Ownership concentration	-0.008	0.006	-0.005	0.005	-0.006	0.005	-0.007	0.008
Government owned	-0.274	0.469	-0.268	0.379	-0.269	0.385	-0.177	1.959
Fixed effects: Store Type and City	Yes		Yes		Yes		Yes	
Observations	1501		1501		734		767	
Model statistics	LL= -355.59		LL= -342.23		LL= -362.78		LL= -260.69	
	Chi2 = 441.36 (p =0.00)		Chi2 = 446.78 (p =0.00)		Chi2 = 467.43 (p =0.00)		Chi2 = 300.26 (p =0.00)	

*** $p < .01$; ** $p < .05$; * $p < .1$. Standard errors are based on state level clustering.

Table 3: Electricity Generator Adoption - Probit Regression

Variables	All Stores Coeff.	All Stores SE	Larger Stores Coeff.	Larger Stores SE	Smaller Stores Coeff.	Smaller Stores SE
Corruption factors discouraging transparency						
State TI Corruption Index	-0.080	0.094	-0.124	0.113	-0.044	0.218
Percentage revenue spent on bribe	-0.013	0.018	-0.031	0.028	-0.027	0.029
Perceived informality by peers	0.004	0.004	0.005	0.005	0.002	0.002
Enforcement factors encouraging transparency						
Regulatory inspections	0.026	0.031	0.009	0.022	0.075	0.083
External auditor	0.245	0.172	0.139	0.153	0.472	0.516
State’s egulatory consistency	0.011	0.015	0.009	0.014	0.013	0.018
Labor cost and education level						
State minimum wages	-0.021	0.094	-0.026	0.113	-0.019	0.087
State literacy rate	0.043	0.034	0.045	0.039	0.040	0.035
Electric power supply related factors						
Power outage	1.568***	0.289	1.801***	0.344	1.417***	0.476
State power supply problem	0.728**	0.286	1.042**	0.433	0.586**	0.251
Productivity enhancing technology adoptions						
Computer	0.518***	0.141	0.662***	0.162	0.287***	0.110
Store level performance measure						
Gross annual revenue	0.068***	0.018	0.069***	0.022	0.127***	0.046
Store characteristics						
Store size	0.306***	0.070	0.275**	0.119	0.350**	0.164
Employee size	0.368***	0.075	0.441***	0.105	0.357***	0.133
Store age	-0.081	0.080	-0.004	0.108	-0.277	0.132
Store management & ownership characteristics						
Manager experience	0.266***	0.077	0.261***	0.103	0.339***	0.126
Ownership concentration	0.005*	0.003	0.004	0.003	0.006	0.007
Government owned	-0.254	0.465	-0.765	0.523	-0.378	0.313
Fixed effects						
Store type and City	Yes		Yes		Yes	
Observations	1501		734		767	
Model statistics	LL= - 512.63		LL= - 294.97		LL= - 230.14	
	Chi2 = 481.5 (p =0.00)		Chi2 = 368.2 (p =0.00)		Chi2 = 274.2 (p =0.00)	

*** $p < .01$; ** $p < .05$; * $p < .1$. Standard errors are based on state level clustering.

Table 4a: Labor Productivity - OLS Regression

Variables	All Stores Coeff.	All Stores SE	Larger Stores Coeff.	Larger Stores SE	Smaller Stores Coeff.	Smaller Stores SE
Productivity enhancing technology adoptions						
Computer	0.224**	0.097	0.253***	0.094	0.216**	0.093
Generator	0.244***	0.080	0.233**	0.097	0.267***	0.079
Store characteristics						
Store size	0.011**	0.005	0.012***	0.002	0.011***	0.003
Store age	-0.049	0.048	-0.085	0.070	-0.064	0.050
Store management & ownership characteristics						
Manager experience	0.113**	0.046	0.220***	0.065	0.098**	0.040
Ownership concentration	-0.003*	0.002	-0.002*	0.001	-0.006**	0.003
Government owned	0.600	0.461	0.735	0.561	0.236	0.738
Finance, in-store security and competitive factors						
Bank account	0.316***	0.072	0.305**	0.124	0.319**	0.139
Overdraft facility	0.179**	0.080	0.189**	0.085	0.174**	0.077
In-store security	0.189**	0.076	0.209**	0.105	0.185**	0.085
Price competition level	0.095	0.065	0.125	0.097	0.036	0.083
Inventory level for main product	0.006***	0.002	0.007***	0.003	0.004***	0.002
State level educational factor						
Literacy rate	0.022**	0.011	0.019*	0.010	0.021**	0.009
Fixed effects						
Store type and City	Yes		Yes		Yes	
Observations	1501		734		767	
Model statistics	LL= - 2682.1		LL= - 1291.4		LL= - 1280.1	
	F-stat = 10.9 (p =0.00)		F-stat = 5.3 (p =0.00)		F-stat = 5.6 (p =0.00)	

*** $p < .01$; ** $p < .05$; * $p < .1$. Standard errors are based on state level clustering.

Table 4b: Floor Area Productivity - OLS Regression

	All Stores		Larger Stores		Smaller Stores	
Variables	Coeff.	SE	Coeff.	SE	Coeff.	SE
Productivity enhancing technology adoptions						
Computer	0.259**	0.106	0.266***	0.101	0.247**	0.102
Generator	0.213**	0.086	0.209**	0.086	0.220***	0.081
Store characteristics						
Store size	-0.057	0.047	-0.059**	0.027	-0.056**	0.026
Store age	0.082**	0.052	0.108**	0.055	0.078**	0.039
Store management & ownership characteristics						
Manager experience	0.093*	0.050	0.123**	0.057	0.087**	0.039
Ownership concentration	-0.003	0.002	-0.005*	0.003	-0.003	0.002
Government owned	0.492	0.316	0.532	0.348	0.382	0.285
Finance, in-store security and competitive factors						
Bank account	0.298***	0.078	0.348***	0.090	0.288***	0.077
Overdraft facility	0.223***	0.086	0.242***	0.092	0.221***	0.082
In-store security	0.228***	0.082	0.245***	0.090	0.213***	0.076
Price competition level	0.104	0.070	0.094	0.069	0.108	0.078
Inventory level for main product	0.002	0.002	0.003*	0.001	0.003	0.003
State level educational factor						
Literacy rate	0.026**	0.012	0.027**	0.013	0.025**	0.012
Fixed effects						
Store type and City	Yes		Yes		Yes	
Observations	1503		734		769	
Model statistics	LL= - 2607.8		LL= - 1680.9		LL= - 1342.4	
	F-stat = 10.4 (p =0.00)		F-stat = 6.06 (p =0.00)		F-stat = 6.36 (p =0.00)	

*** $p < .01$; ** $p < .05$; * $p < .1$. Standard errors are based on state level clustering.

Table 5. Distribution of Propensity Scores and Means for Selected Propensity Scoring Variables over the Common Support

Block or PS interval#	Lower bound of propensity score within block [a]	Number of sample observations		Means for selected variables in the treated and control groups [b]					
		Treated	Control	% Revenue spent on bribe	Regulatory consistency (1-6)	Power supply problem (0-4)	Gross annual revenue (million Rs.)	Store size (sq. ft.)	Managerial experience (yrs.)
1	0.000	4	328	0.442; 0.461	3.237; 3.097	1.603; 1.568	0.625; 0.597	128.50; 142.00	11.00; 11.37
2	0.025	6	132	0.985; 1.004	3.133; 3.076	1.516; 1.619	0.925; 0.792	199.67; 182.08	11.83; 11.52
3	0.050	8	134	0.939; 0.980	3.129; 3.102	1.572; 1.641	1.229; 1.090	208.75; 217.23	11.75; 13.05
4	0.100	33	135	1.012; 1.019	3.124; 3.116	1.545; 1.610	1.741; 1.592	269.73; 247.90	12.61; 12.81
5	0.200	5	59	0.874; 0.815	2.900; 3.091	1.802; 1.601	1.840; 1.589	280.40; 318.10	13.80; 13.22
6	0.250	11	33	1.708; 1.663	3.108; 3.186	1.524; 1.615	2.564; 2.235	321.82; 332.88	13.27; 13.73
7	0.300	18	33	1.585; 1.630	2.991; 3.208	1.546; 1.644	4.706; 4.962	424.72; 411.39	14.22; 13.21
8	0.400	7	23	0.947; 0.996	3.236; 3.100	1.616; 1.732	6.357; 5.966	379.71; 331.74	10.14; 11.70
9	0.450	17	10	2.276; 2.295	2.823; 2.911	1.446; 1.536	5.956; 5.827	636.59; 611.00	14.94; 14.20
10	0.500	20	17	1.034; 0.932	2.968; 3.203	1.481; 1.659	5.150; 4.986	611.80; 583.53	14.45; 14.00
11	0.600	21	13	1.133; 1.112	3.115; 3.166	1.517; 1.364	5.426; 5.512	707.62; 705.38	12.24; 12.85
12	0.700	35	9	0.955; 0.867	3.144; 3.239	1.459; 1.604	5.533; 5.609	840.86; 912.11	11.51; 12.44
13	0.800	84	4	1.028; 0.946	3.141; 3.069	1.465; 1.358	7.489; 7.750	5596.01; 5250.00	13.50; 13.25

[a] The region of common support is [0.004, 0.999].
[b] For each selected variable, mean values are shown by treated and control groups (in that sequence) within each block.

Table 6: Propensity Score Analysis with Kernel Matching

Outcome	Sample	ATT[1]	S.E.
Log of labor productivity	All Stores	0.409***	0.142
Log of Floor area productivity	All Stores	0.528***	0.187
Log of labor productivity	Large Stores	0.418***	0.160
	Small Stores	0.367**	0.158
Log of Floor area productivity	Large Stores	0.557***	0.128
	Small Stores	0.527**	0.241

*** $p < .01$; ** $p < .05$

[1] All estimates are based on bias-corrected matching estimators using kernel (Epanechnikov) matching approach (Leuven and Sianesi 2003).

Table 7: Robustness to Unobserved Selection Effects: Rosenbaum Bound Analyses

Gamma (Γ)	p-Value [2]		H-L Point Estimate [3]		Conf. Interval [2]	
	(U-Bound)	(L-Bound)	(U-Bound)	(L-Bound)	(U-Bound)	(L-Bound)
For Log of Labor Productivity[1]						
1.0	0.0000	0.0000	0.4502	0.4502	0.3163	0.5713
1.1	0.0000	0.0000	0.4068	0.4886	0.2751	0.6141
1.2	0.0000	0.0000	0.3688	0.5240	0.2371	0.6519
1.3	0.0001	0.0000	0.3329	0.5570	0.2036	0.6844
1.4	0.0001	0.0000	0.3017	0.5902	0.1720	0.7150
1.5	0.0001	0.0000	0.2718	0.6195	0.1426	0.7429
1.6	0.0001	0.0000	0.2458	0.6468	0.1133	0.7688
1.7	0.0006	0.0000	0.2211	0.6700	0.0839	0.7948
1.8	0.0099	0.0000	0.1975	0.6922	0.0582	0.8182
1.9	0.0263	0.0000	0.1742	0.7133	0.0327	0.8391
2.0	0.0581	0.0000	0.1527	0.7333	0.0100	0.8601
2.1	0.1103	0.0000	0.1307	0.7526	-0.0102	0.8797
2.2	0.1847	0.0000	0.1097	0.7717	-0.0303	0.8980
2.3	0.2787	0.0000	0.0902	0.7886	-0.0503	0.9163
2.4	0.3856	0.0000	0.0709	0.8044	-0.0681	0.9329
2.5	0.4968	0.0000	0.0541	0.8203	-0.0862	0.9479
For Log of Floor Area Productivity[1]						
1.0	0.0000	0.0000	0.5919	0.5919	0.4645	0.7203
1.1	0.0000	0.0000	0.5410	0.6414	0.4120	0.7721
1.2	0.0000	0.0000	0.4950	0.6895	0.3626	0.8212
1.3	0.0000	0.0000	0.4523	0.7318	0.3166	0.8621
1.4	0.0000	0.0000	0.4116	0.7725	0.2748	0.9009
1.5	0.0000	0.0000	0.3732	0.8105	0.2324	0.9358
1.6	0.0001	0.0000	0.3382	0.8447	0.1959	0.9702
1.7	0.0004	0.0000	0.3041	0.8745	0.1617	1.0022
1.8	0.0013	0.0000	0.2729	0.9031	0.1289	1.0350
1.9	0.0037	0.0000	0.2399	0.9300	0.0991	1.0642
2.0	0.0091	0.0000	0.2112	0.9539	0.0709	1.0933
2.1	0.0195	0.0000	0.1857	0.9810	0.0419	1.1190
2.2	0.0375	0.0000	0.1601	1.0037	0.0139	1.1430
2.3	0.0654	0.0000	0.1355	1.0270	-0.0125	1.1662
2.4	0.1051	0.0000	0.1134	1.0497	-0.0397	1.1878
2.5	0.1572	0.0000	0.0919	1.0706	-0.0624	1.2116

[1] Results are based on differences between computer adopters and matched non-adopters, using kernel (Epanechnikov) matching on propensity scores through Leuven and Sianesi (2003).
[2] p-values and confidence intervals are one-sided and at the 90% level.
[3] H-L indicates Hodges-Lehmann.

Table 8a: Labor Productivity Outcome Equation - MLE for Self-Selection Model

Variables	All Stores-I Coeff.	All Stores-I SE	All Stores-II Coeff.	All Stores-II SE	Larger Stores Coeff.	Larger Stores SE	Smaller Stores Coeff.	Smaller Stores SE
Productivity enhancing technology adoptions								
Computer	0.464***	0.066	0.471***	0.074	0.517***	0.142	0.454***	0.206
Generator	0.524**	0.237	0.537**	0.224	0.473**	0.216	0.544**	0.267
Store characteristics								
Store size	0.223***	0.075	0.224***	0.076	0.230**	0.095	0.213***	0.074
Store age	-0.065	0.141	-0.082	0.124	-0.094	-0.066	-0.075	0.061
Store management & ownership characteristics								
Manager experience	0.126***	0.046	0.135***	0.046	0.138***	0.060	0.108***	0.041
Ownership concentration	-0.006***	0.002	-0.006***	0.002	-0.006**	-0.003	-0.007**	0.003
Government owned	0.475	0.319	0.465	0.310	0.686	0.501	0.385	0.335
Finance, in-store security and competitive factors								
Bank account	0.306***	0.079	0.317***	0.066	0.368***	0.097	0.283**	0.118
Overdraft facility	0.407**	0.196	0.389**	0.183	0.415**	0.187	0.370**	0.176
In-store security	0.328**	0.151	0.334**	0.151	0.343***	0.097	0.326***	0.122
Price competition level	0.077	0.066	0.077	0.059	0.122	0.088	0.067	0.114
Inventory level for main product	0.005***	0.002	0.005***	0.002	0.008***	0.003	0.005***	0.002
State level educational factor								
Literacy rate	--	--	0.019*	0.010	0.020*	0.010	0.020**	0.009
Fixed effects	Store type and State		Store type and City		Store type and City		Store type and City	
Observations	1501		1501		734		767	
Model statistics[1]	LL= - 2566.28		LL= - 2494.92		LL= - 1272.86		LL = - 1350.32	
	Chi2 = 524.7 (p =0.00)		Chi2 = 660.6 (p =0.00)		Chi2 = 550.2 (p =0.00)		Chi2 = 400.6 (p =0.00)	
	Rho = - 0.44		Rho = - 0.46		Rho = - 0.51		Rho = - 0.42	

*** $p < .01$; ** $p < .05$; * $p < .1$. Standard errors are based on clustering – at city level for models with state fixed effects and at state level for models with no state fixed effects.

[1] Rho denotes the correlation in error terms between the outcome (productivity) and the selection (computer adoption) equations.

Table 8b. Business Computer Adoption Selection Equation - MLE for Self-Selection Model

Variables	All Stores-I Coeff.	All Stores-I SE	All Stores-II Coeff.	All Stores-II SE	Larger Stores Coeff.	Larger Stores SE	Smaller Stores Coeff.	Smaller Stores SE
Corruption factors discouraging transparency								
State TI Corruption Index	--	--	-0.110***	0.036	-0.116***	0.037	-0.093***	0.028
Percent revenue spent on bribe	-0.028*	0.014	-0.024*	0.012	-0.026**	0.013	-0.019*	0.011
Perceived informality by peers	-0.004***	0.001	-0.004***	0.001	-0.004***	0.002	-0.002***	0.001
Enforcement factors encouraging transparency								
Regulatory inspections	0.040***	0.016	0.041***	0.013	0.048***	0.014	0.025***	0.008
External auditor	0.274***	0.095	0.278***	0.099	0.236***	0.082	0.304***	0.104
Regulatory consistency	--	--	0.025**	0.012	0.024**	0.012	0.026**	0.013
Labor cost and education level								
State minimum wages	--	--	0.021**	0.011	0.022**	0.011	0.015**	0.007
State literacy rate	--	--	0.065*	0.036	0.067*	0.036	0.062**	0.029
Electric power supply related factors								
Power outage	-0.063	0.102	-0.067	0.117	-0.077	0.108	-0.063	0.098
State power supply problem	--	--	-0.081	0.056	-0.084	0.059	-0.074	0.061
Productivity enhancing technology adoptions								
Generator	0.406***	0.125	0.398***	0.117	0.491***	0.139	0.519***	0.148
Store level performance measure								
Gross annual revenue	0.093***	0.008	0.095***	0.007	0.087***	0.006	0.133***	0.014
Store characteristics								
Store size	0.291***	0.063	0.299***	0.064	0.274***	0.055	0.353**	0.168
Employee size	0.239***	0.038	0.235***	0.038	0.268***	0.041	0.196***	0.019
Store age	-0.086**	0.042	-0.091**	0.045	-0.108**	0.051	-0.088**	0.042
Store management & ownership characteristics								
Manager experience	-0.146*	0.070	-0.140*	0.075	-0.177**	0.089	-0.127*	0.068
Ownership concentration	-0.009	0.006	-0.011*	0.006	-0.010*	0.006	-0.008*	0.005
Government owned	-0.322	0.223	-0.336	0.237	-0.532	0.409	-0.197	0.768
Fixed effects	Store type and State		Store type and City		Store type and City		Store type and City	
Observations	1501		1501		734		767	
Model statistics	LL= - 2566.09		LL=- 2494.94		LL=- 1272.88		LL = - 1350.33	
	Chi2 = 524.8 (p =0.00)		Chi2 = 660.6 (p =0.00)		Chi2 = 550.2 (p =0.00)		Chi2 = 400.6 (p =0.00)	

*** $p < .01$; ** $p < .05$; * $p < .1$. Standard errors are based on clustering – at city level for models with state fixed effects and at state level for models with no state fixed effects.

Table 8c. Floor Area Productivity Outcome Equation - MLE for Self-Selection Model

	All Stores-I		**All Stores-II**		**Larger Stores**		**Smaller Stores**	
Variables	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Productivity enhancing technology adoptions								
Computer	0.538***	0.105	0.543***	0.103	0.566***	0.056	0.523***	0.082
Generator	0.410**	0.178	0.418**	0.188	0.394**	0.163	0.440**	0.195
Store characteristics								
Store size	-0.065***	0.015	-0.061***	0.012	-0.066***	0.022	-0.059***	0.010
Store age	0.142**	0.063	0.147**	0.063	0.204***	0.077	0.143**	0.070
Store management & ownership characteristics								
Manager experience	0.175**	0.076	0.176**	0.076	0.214***	0.070	0.165***	0.052
Ownership concentration	-0.005**	0.002	-0.005**	0.002	-0.007***	0.003	-0.005*	0.003
Government owned	0.361	0.356	0.337	0.343	0.307	0.394	0.363	0.272
Finance, in-store security and competitive factors								
Bank account	0.342***	0.083	0.335***	0.082	0.448***	0.116	0.314**	0.138
Overdraft facility	0.143*	0.075	0.137*	0.073	0.161*	0.091	0.059*	0.034
In-store security	0.206***	0.078	0.208***	0.076	0.251***	0.095	0.204**	0.092
Price competition level	0.074	0.067	0.076	0.065	0.076	0.087	0.000	0.089
Inventory level for main product	0.003**	0.002	0.003**	0.002	0.004**	0.002	0.003**	0.001
State level educational factor								
Literacy rate	--	--	0.022**	0.011	0.025*	0.013	0.023**	0.010
Fixed effects	Store type and State		Store type and City		Store type and City		Store type and City	
Observations	1501		1501		734		767	
Model statistics[1]	LL = -2561.36 Chi2= 710.83 (p =0.00) Rho = - 0.46		LL = -2595.52 Chi2= 739.04 (p =0.00) Rho = - 0.49		LL = -1362.07 Chi2= 670.02 (p =0.00) Rho = - 0.53		LL = -1452.27 Chi2= 450.69 (p =0.00) Rho = - 0.46	

*** $p < .01$; ** $p < .05$; * $p < .1$. Standard errors are based on clustering – at city level for models with state fixed effects and at state level for models with no state fixed effects.

[1] Rho denotes the correlation in error terms between the respective productivity and the selection regression equations.
Results for the computer adoption equation with the floor area productivity outcome equation are not shown to conserve space; the estimates are virtually identical to Table 8b.

Table 9: Effect of Computer Adoption on Productivity: Summary[1]

Firm Size	OLS		Propensity Score Matching		Self-Selection Model	
	Labor	Floor	Labor	Floor	Labor	Floor
All	1.251	1.296	1.505	1.696	1.602	1.721
Large	1.288	1.305	1.519	1.745	1.677	1.761
Small	1.241	1.280	1.443	1.694	1.575	1.687

[1]The productivity effects are obtained by taking the exponential of the parameter estimates in the respective tables since productivity enters in logs as dependent variables in the regression. For example, consider the effects on labor productivity from all store analysis results. For OLS, from Table 4a, exp(0.224)=1.251; for propensity score matching from Table 6, exp(.409)=1.505; and for self-selection model from Table 8a column 2, exp(.471)=1.602.