Competition and Labor Productivity in India's Retail Stores

Mohammad Amin*
World Bank

August 2007

Abstract The paper analyzes the effect of product market competition on the average productivity of labor in India's retail sector. We use a new dataset of 1,948 retail stores located in 41 cities of India compiled by the World Bank's Enterprise surveys. According to the survey, 62% of the stores do not face any significant competition. The empirical analysis establishes a strong causal effect of competition on labor productivity of stores. Our estimates suggest an increase of 87% in labor productivity from pro-competitive reforms.

Keywords: Competition, Productivity, Retail, India
JEL: D24, L11, L81

*The World Bank, Washington DC, 20433. Email: mamin@worldbank.org. Phone: (202)-473-1915.

1. Introduction

The World Bank conducted a survey of 1,948 retail stores in 16 states and 41 cities of India in 2006. The survey shows that 62% of the retailers do not face any significant competition in the product market (henceforth, competition). There is substantial variation in the figure across cities and store-types. We exploit this variation to estimate the effect of competition on the average productivity of labor (efficiency) at the store level. We find a strong positive effect of greater competition on efficiency. The estimates suggest an increase of 87% in the efficiency level of an average store from pro-competitive reforms.[1]

The retail sector in India is the second largest employer (after agriculture) providing jobs to 9.4% of all workers and its contribution to GDP is 14%. The sector showed a remarkable turnaround during the 1990s when it grew at an annual rate of 7.3% per annum compared to 5.9% per annum during the 1980s. The overall growth rate of the economy was roughly same over the two decades at 5.8% per annum.[2] Despite this impressive performance, the retail sector in India is known to be highly inefficient which, according to many, may impede its future development. There is no formal work on the determinants of efficiency in the sector and the present paper attempts to fill this gap by focusing on the level of competition as one such determinant.

There is a general belief that competition is good for efficiency. Djankov and Murrell (2002) provide an overview of the competition-performance literature highlighting X-inefficiency and industry rationalization as possible reasons for this

[1] We are not aware of any previous work on the competition-efficiency nexus for the retail sector in a developing country.

[2] Figures reported in the paragraph relate to the distribution sector (retail plus wholesale). They are taken from Banga (2005) and Gordon and Gupta (2004) and are based on official CSO data.

belief.[3] The X-inefficiency view is that managerial effort is under-supplied in the absence of vigorous competition, lowering firm performance. More competition puts a downward pressure on price-cost margins forcing managers to exert greater effort in order to survive (Horn et al 1995, Sorenson 2000, Bruhn 2007). Industry rationalization leads to a shift of resources from inefficient to efficient firms within and between sectors, improving overall efficiency (Schmidt 1997). Empirical evidence on the competition-performance nexus, however, is limited and mixed (Nickell 1996, Djankov and Murrell 2002). For example, while Nickell (1996) and Nickell et al (1992) find a positive effect of greater competition on the performance of British manufacturing firms, Blanchflower and Machin (1996) and Januszewski et al (2001) find no such effect for British and German firms, respectively. These studies suggest that the relationship between competition and performance is an empirical question which motivates the present work.

Another strand of the literature focuses on the determinants of competition in retailing and we draw on this literature to generate exogenous variations in the level of competition for the instrumental variables estimation strategy. Studies in this literature show that competition in retailing is higher when consumers search more intensively for best prices (Marvel 1976, Calem and Mester 1995, Knittel 1997 and Sorenson 2000). Direct measures of search intensity are typically not available and the approach in the literature is to use some proxy measure (of search intensity) instead. For example, Sorenson argues that the benefit to consumers from searching and therefore the equilibrium level of search is higher for those products that require more frequent

[3] Competition may affect a firm's incentive to innovate and the cost of monitoring managers and therefore efficiency. Nickell (1996) provides an overview of the literature. These factors are of little relevance to India's retail sector because formal R&D (or large investments) in the sector is rare and most stores are run by the owner with only a few employees (4 workers in our data).

purchases. He provides confirmatory evidence for the prescription-drugs retail market in the U.S. In another study, Marvel argues that a consumer's opportunity cost of time is one of the most important determinants of search intensity. Higher opportunity cost of time implies less equilibrium search and therefore lower competition. He also suggests that a consumer's wage rate (or income level) is a good proxy for the opportunity cost of her time. Unfortunately, due to data limitations, we cannot use either product characteristics in the sense of Sorenson or wage rate in the sense of Marvel to instrument for the level of competition. Instead, we follow Goldman et al (2002) who argue that the number of non-working adults in the household is a good surrogate for the "household shopping time opportunity-cost". Goldman et al do not analyze the relationship between the level of competition and the number of adult non-workers per household and the present paper makes a first such effort.[4] Briefly, we use (lagged values of) the number of adult non-workers per household in the city to instrument for the level of competition faced by stores in the city. The motivation here is that more non-workers imply lower opportunity cost of time spent shopping and therefore more intensive search and greater competition. One novelty in our approach is that we contrast the effects of non-workers and children in the household on competition which provides additional support to our search-based identification strategy. Further, in the sections below, we argue that any possible direct effect of the instrument on efficiency is most likely to bias our main results towards zero, yielding predictions on the conservative side. This is an added advantage of the instrument.

[4] Goldman et al estimate how the number of adult non-workers in the household affects household's choice of shopping at wet markets relative to superstores and conventional supermarkets.

The rest of the paper is as follows. In section 2 we describe the data and the empirical methodology. In section 3 we present the empirical findings using the ordinary least squares (OLS) method. Instrumental variable (IV) regression results are reported in section 4. A summary of the main findings is provided in the concluding section.

2. Data and Main Variables

We use store level data collected by the World Bank in 2006 (Enterprise survey).[5] The data are a cross section of 1,948 retail stores spread over 16 states and 41 cities of India. Stores in our sample are a mix of small and large ones. Some sell grocery items while others consumer durables. The National Industrial Classification groups retailers into those operating through established stores and the rest who usually operate from home (NIC 1998, Industry Division 52). All stores in our sample belong to the former group.

Information on products carried by stores is not available. However, the survey does classify stores into the following types: (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.

2.1 *Dependent variable*

A formal definition of all the variables used in the regressions is provided in Table 1. Our dependent variable is the level of efficiency of the stores. We follow the literature in

[5] The survey and methodology for data collection are available at www.entersurveys.org.

using store's average productivity of labor to measure efficiency.[6] This equals (log of) annual sales to employment ratio of the stores for the fiscal year 2005-06 (*Efficiency*).

Efficiency varies from 8.1 to 18.4 with a mean value of 12.5. Consumer durable stores are most efficient (12.8) followed by the modern format stores (12.7) and then the traditional stores (12.2). Across cities, *Efficiency* is highest in Kochi (traditional stores), Cuttack (consumer durables) and Faridabad (modern format stores). Corresponding cities with lowest efficiency are Mysore, Nagpur and Coimbatore, respectively.

2.2 *Explanatory variables*

Our main explanatory variable is the level of competition which is constructed from responses of stores to the following question asked in the survey: "For this store, how important are each of the following influences over prices of its main products? a) Pressure/influence from domestic competitors, b) Pressure/influence from foreign competitors, c) Pressure/influence from unorganized trade (hawkers, traders sitting on pavement, people selling from home, people selling spurious good)." Stores were shown a card with the question above printed on it. Their responses to parts (a)-(c) were recorded separately on a 1-4 scale defined as: not at all important (1), slightly important (2), fairly important (3) and important (4).

We define our measure of competition as the average score on part (a) of the question above where the average is taken at the "city-store type" level (*Competition*).[7] We note a few important points here to better understand what *Competition* is picking up.

[6] Data limitations do not allow use of value added to measure efficiency. This problem is common in the literature on the services sectors. See, for example, Baily and Solow (2001) and Cainelli et al (2006).
[7] "City-store type" is the Cartesian product of the set of cities in our sample and store-types (traditional, consumer durable and modern format).

First, foreign competition is virtually non-existent in the sector and informal competition (part (c) of the question) is also small.[8] Further, policy implications of more competition via an expanded informal sector are not clear.[9] For these reasons we focus on formal competition (part (a) of the question) treating informal competition as a standard control variable in the regressions. Second, store-level responses (perceptions) about the level of competition cannot be used directly in the regressions because they could be endogenous to store characteristics such as size, age, etc. *Competition*, as defined above, is the average level of competition faced by stores within each city-store type cell. Being a group average, it suffers less from measurement errors and endogeneity problems associated with store-level responses (Krueger and Angrist, 2001) although these problems cannot be ruled out completely. We discuss this issue in detail below. Third, there is substantial variation in the reported scores (on part (a) of the question) across store-types within cities (Figure 1). Averaging these scores at the city-store type level (as opposed to averaging at the city or store-type level) allows us to exploit this variation for a better identification of our main results. Fourth, one could argue that *Competition* captures price-competition alone which is too narrow a measure of the overall competitive pressure faced by the stores. For example, pricing restrictions for certain products (by law) may blunt price-competition but stores may still compete with each other for the precious few buyers (sales competition). Also, stores may react to greater competitive pressure (for example, due to more retailers in the city) by introducing new and improved product lines rather than adjusting prices. While this problem cannot be

[8] 60% of all stores report informal competition as "not at all important" and another 21% report it as "slightly important". Corresponding figures for formal competition (part (a) of the question above) are 36% and 26%, respectively.

[9] For example, street hawkers hardly seem to be in a position to make large investments necessary for the future growth and development of the sector.

ruled out completely, we provide some evidence which suggests that it is unlikely to be serious. Specifically, in one survey question, stores were asked how important is the influence of domestic competitors for their decision to introduce new product lines. Responses were recorded on the same 1-4 scale as above. The correlation coefficient between the reported scores here (averaged at the city-store type level) and *Competition* equals .894. The high correlation is reassuring in that it suggests that *Competition* captures the broader competitive environment (density of retailers, intensity with which consumers search for best prices, availability of information on prices and product quality with the consumers, etc.) rather than the narrow specifics of price-competition. In the remainder of the paper, the word "competition" will refer to part (a) of the question above.

In the full sample, 36% of all stores report competition as "not at all important", 26% as "slightly important", 20% as "fairly important" and the rest 18% as "important". Figure 2 compares these numbers with similar ones for the retail sectors in other countries using EBRD's Business Environment and Enterprise Performance (BEEPS, 2005) survey while Figure 3 does the same using the World Bank's survey of Indian manufacturing conducted in 2005.[10] Both these figures reveal a much lower level of competition in India's retail sector and therefore ample scope for pro-competitive reforms. For example, compared to 71% of retailers in BEEPS and 82% of manufacturing firms in India, only 38% of Indian retailers find competition as significant (more than slightly important).

[10] These surveys and methodology for data collection are available at www.enterprisesurveys.org. The exact questions used from BEEPS and Indian manufacturing surveys are reported along with Figures 1 and 2.

Focusing on India's retail sector, the average score on part (a) of the competition question is highest for consumer durable stores (2.42), followed by the modern format stores (2.17) and then the traditional stores (2.11). Figure 4 shows the corresponding variation across cities. The metropolitan cities of Mumbai, Delhi, Chennai, Bangalore, Kolkatta and Hyderabad show intermediate levels of competition while some smaller cities like Madurai and Jalandhar show much higher levels of competition.

2.2.1 *Other controls*

While direct reverse causality from *Efficiency* (which varies at the store level) to *Competition* (which varies at the city-store type level) is unlikely, it cannot be ruled out completely. A relatively more serious problem relates to measurement errors with the competition variable and estimation bias due to omitted variables. We provide a few examples of these problems to motivate the identification strategy.

First, consider aspects of regulation, infrastructure, etc., that are, to some extent, commonly shared by stores in a given region. For example, more stringent business regulations are known to reduce competition by blocking new entry. Failure to control for such entry blocking laws in the regressions may bias the estimated coefficient of *Competition* upwards because these laws are likely to be correlated with efficiency and competition in the same (negative) direction. Our empirical specification controls for city and store-type fixed effects which implies that the bias here will survive only if either the business laws or their enforcement varies across store-types (otherwise city fixed effects eliminate the bias) *and* that this variation across store-types is not uniform across all cities (otherwise store-type fixed effects eliminate the bias).

Second, omitted store characteristics correlated with both, efficiency and competition may cause spurious correlation between competition and efficiency. We expect this problem to be less severe in our specification than is otherwise the case because our measure of competition is obtained by averaging across stores of different characteristics. That is, in our specification, the problem of spurious correlation could arise from variation in store attributes across but not within city-store type cells. To the extent that this source of bias does exist, it is difficult to sign its direction. For example, managers/owners with higher ability are likely to be more efficient but do these managers face more competition than others? The answer is not certain. Higher ability managers may like to compete more because they expect to win (upward bias) but their very presence may drive out the less efficient competitors or prevent new ones from entering the market (less competition and a downward bias).

Perception or understanding of what is, for example, "slightly important" level of influence (as asked in the survey question) could vary across stores. Hence, some noise or measurement error with *Competition* cannot be ruled out. While it is difficult to say how serious this problem is, a comparison of the OLS and IV regression results below provides some indication that measurement errors may be a relatively more serious problem than reverse causality or omitted variable bias for the OLS results.

We address the problems discussed above by directly controlling for a large number of observables at the store, store-type, city and city-store type level and by using the instrumental variables estimation approach. In our main (OLS) specification we control for store-size proxied by the floor area of the store (*Size*), city fixed effects and store-type fixed effects. The section on robustness includes a number of additional

controls like age of the store, ownership structure, infrastructure and finance availability, regulations, aspects of business climate, etc.

Floor area of the shop is correlated with a number of observable store characteristics in our sample which suggests that it may be a good proxy for unobservable store characteristics as well.[11] One advantage of using floor area over other observables is that the floor area is largely predetermined and therefore unlikely to suffer from simultaneity problem. In India, land is acquired primarily for opening a new store rather than expanding an existing one.[12]

Efficiency could vary due to different product lines carried by the stores. For example, we mentioned above that relative to traditional stores, consumer durable stores are more competitive and also more efficient (in 36 out of 41 cities). Does this pattern reflect the competition-efficiency nexus in the sense discussed above or simply the fact that compared to traditional stores, consumer durable stores sell higher valued products (computers vs. bread) and they also happen to be more competitive for extraneous reasons?

We address this identification problem by showing that our main results hold with and without store-type fixed effects as controls.[13] The assumption here is that product lines carried by stores of a particular store-type (e.g. televisions for consumer durable stores) are roughly similar across cities and therefore absorbed by the store-type fixed effects. To see this, we regressed *Efficiency* on store-type fixed effects and took the

[11] For example, access to finance, computer usage, days of inventory maintained by the store and the availability of power supply show significant correlation with the floor area of the shop.
[12] In one survey question stores were asked if they had acquired new land in the last three years to expand operations of the current store. Less than 2% of the stores reported doing so.
[13] The use of instrumental variables estimation strategy provides another layer of defense for our main results against the identification problem here.

residuals. These residuals show that consumer durable stores are more efficient than the traditional stores in only half of the cities compared to 36 out of 41 cities earlier (Figure 5).

City fixed effects control for all determinants of efficiency which vary across cities but are common to stores within a city. Examples include the quality of roads, crime, tax rates and income and expenditure levels.

3. Estimation

In this section we estimate the relationship between competition and efficiency using the OLS method. The base specification is as follows

$$Efficiency_{ics} = \alpha_0 + \beta\, Competition_{cs} + \alpha_1 Size_{ics} + CFE_c + SFE_s + u_{ics}$$

subscript i denotes the ith store, c the city in which it is located and s the store-type (traditional, consumer durable or modern). CFE_c and SFE_s denote city and store-type fixed effects, respectively. u_{ics} is the error term. The coefficient of interest in the equation is β which we expect to be positive. In all our regressions we use Huber-White robust standard errors clustered on city-store type.

3.1 *Base regression results*

Results from the estimation of the previous equation are reported in Table 2. Without any controls, the estimated coefficient of *Competition* equals .143 significant at less than 10% level with a p-value of .063 (column 1, Table 2). The coefficient remains roughly

unchanged in magnitude but becomes significant at less than 5% level when we control for *Size* (column 2, Table 2). *Size* has a positive effect on efficiency significant at less than 1% level. Controlling for city and store-type fixed effects increases the estimated coefficient of *Competition* from .141 above to .391 significant at less than 1% level. The coefficient of *Size* decreases from .216 to .181 but remains significant at less than 1% level (column 3, Table 2).

We also check for a possible non-linear effect of *Size* on efficiency by adding $Size^2$ (square of *Size*) to the set of controls above. Estimation results reported in column 4 of Table 2 show that efficiency rises with size but at a sharply decreasing rate. The estimated coefficient of *Competition* is smaller here than above equaling .343 but still significant at less than 1% level.

The results above imply a fairly large effect of competition on efficiency. For example, increasing the level of competition faced by the traditional stores in the city of Ghaziabad (median value of *Competition*) to the level faced by the modern format stores in the city of Madurai (highest value of *Competition*) will lead to an increase of 87% in the average productivity of labor for the former set of stores.[14] Given the relatively low level of competition in the sector, our findings suggest substantial scope for improvement in efficiency through pro-competitive reforms.

3.2 *Robustness of* OLS *results*

Robustness checks for the OLS estimation method are reported in Table 3. We begin by controlling for store's current level of employment (*Employment*) and age (*Age*).

[14] The improvement in efficiency here is calculated using the estimated coefficient of *Competition* in the column 4 of Table 2.

Diminishing returns to labor implies that average productivity of labor may decline with *Employment* while learning-by-doing or selection effects imply a positive relationship between *Age* and *Efficiency*. Our results for the effect of competition on efficiency could be biased if age or employment is systematically correlated with the level of competition reported by the stores. Controlling for age and employment, we find that the estimated coefficient of competition remains significant at less than 1% level although it decreases in magnitude from .343 above to .307 (column 1, Table 3). As predicted, *Employment* has a negative effect while *Age* a positive effect on efficiency. Both these effects are significant at less than 1% level. There is no noticeable change in the estimated coefficients of the remaining variables.

In the Enterprise survey, stores reported irregular power supply and access to finance as the two biggest problems they faced in running their business. We therefore expect some effect of these variables on efficiency. The danger here is that these variables could also be correlated with competition because they are in the nature of what Bliss and Di Tella (1997) call "deeper competition parameters". If this is indeed the case for our sample, then our results for the effect of competition on efficiency could suffer from serious omitted variable bias problem. To check for this potential problem we control for the duration of power outage faced by a store per day on an average during the last fiscal year (*Outage*) and a dummy variable equal to 1 if a store has overdraft facility and 0 otherwise (*Overdraft*).[15] Regression results controlling for *Outage* and *Overdraft* are reported in column 2 of Table 3. As expected, *Outage* has a negative effect while *Overdraft* has a positive effect on efficiency although only the latter is significant at less

[15] The survey also reports on a number of additional measures of power supply and access to finance. Controlling for these measures here or elsewhere in the paper does not make any difference to our main results. We discuss this point in more detail towards the end of the section.

than 5% level. The estimated coefficient of competition here equals .282 which is slightly smaller than in the previous specification (.307) but it is still significant at 1% level.

The insignificant effect of *Outage* on efficiency above may seem a bit surprising given that irregular power supply was voted as the biggest obstacle to doing business by a majority of stores. The reason for the insignificant effect is the presence of city fixed effects which absorb much of the variation in power supply across stores. Estimation results for the previous specification but without city fixed effects show a negative effect of *Outage* on efficiency which is significant at less than 1% level.

In our next robustness check we control for a dummy variable equal to 1 if a store reports using a computer for running its business and 0 otherwise (*Computers*),[16] level of informal competition faced by stores which is equal to the average score (at the city-store type level) on part (c) of the competition question stated in section 2 (*Informal_Competition*), and the number of days of inventory maintained by the store (*Inventory*). The control for computer use is motivated by existing studies which find a strong positive effect of computer usage on labor productivity.[17] We have already discussed the issue of informal competition above (section 2). The stock of inventory is the productive capital of the store which should have a positive (complimentary) effect on labor productivity. Regression results with these additional controls are reported in column 3 of Table 3. The estimated coefficient of competition here equals .269 (p-value of .011) which is only slightly lower than the value of .282 we found in the previous specification. Computer usage and inventory show positive effects on efficiency significant at less than 1% level. Informal competition has a negative effect on efficiency

[16] Data on the number of workers using computers or hours of computer usage are not available.
[17] See, for example, Autor, Katz and Krueger (1998) and Autor, Katz and Kearney (2006).

but this is not significant at 10% or less. There is no noticeable change in the estimated coefficients of the remaining variables.

We experimented with a number of additional controls under the assumption that these controls could be correlated with both, *Competition* and *Efficiency* causing an omitted variable bias problem with our results above. We found that some of these controls had a significant effect on efficiency. These include the percentage of a store's sales in 2005-06 that were never paid for (*Non-payment*), a dummy variable equal to 1 if a store was audited in the last fiscal year and 0 otherwise (*Audited*), a dummy variable equal to 1 if a store has a female principal owner and 0 otherwise (*Female*), and a measure of overall business climate which equals the percentage of stores in each city-store type cell reporting an incidence of theft in the last fiscal year ($Theft_{cs}$). Regression results with these additional controls are reported in column 4 of Table 3. The estimated coefficient of *Competition* here equals .245 which is slightly lower than what we found above (.269) but it is still significant at less than 5% level (p-value of .018). *Non-payment* has a negative effect on efficiency significant at less than 5% level. Audited stores show higher efficiency (significant at less than 1% level) perhaps reflecting a self-selection effect where the more efficient stores prefer to get audited. Female ownership and theft show negative effects on efficiency although these effects are somewhat weak with the former significant at less than 10% level (p-value of .060) and the latter at close to 10% level (p-value of .121).

Lastly, we briefly report on some of the other controls we added to the previous specification but found that these controls did not change our main results significantly. First, we controlled for additional measures of power supply and access to finance which

include a dummy variable equal to 1 if a store owns a generator and 0 otherwise, percentage of a store's electricity derived from generator, a dummy variable equal to 1 if a store has a checking account and 0 otherwise, a dummy variable equal to 1 if a store has a line of credit and 0 otherwise, a dummy variable equal to 1 if a store reported no need to borrow from external sources during the last fiscal year and 0 otherwise. Second, we controlled for a dummy variable equal to 1 if a store is part of a larger chain and 0 otherwise, years of store manager's experience in retailing and the amount of time spent by senior management of the store in dealing with business regulations. Third, we controlled for a number of regulatory and investment climate measures. Since our specification already controls for city fixed effects, we constructed these measures at the city-store type level using store's perceptions about the quality of the investment climate. The measures are formally defined in Table 1 and they capture the extent to which the following are a problem for store's operations: court inefficiency ($Courts_{cs}$), skill shortage ($Skill\_shortage_{cs}$), tax rates ($Tax\ rates_{cs}$), corruption ($Corruption_{cs}$), cumbersome land laws ($Land\ Laws_{cs}$) and difficulty in obtaining permits and licenses ($Permits_{cs}$). With all these controls added to the previous specification, the estimated coefficient of *Competition* equaled .198 in value and was significant at less than 5% level (p-value of .033).

Summarizing, our results show a strong positive effect of greater competition on efficiency of retail stores in India. The relationship is robust to a large number of controls for store characteristics, regulatory and business environment, store-type fixed effects and city fixed effects. The robustness checks raise our confidence against possible omitted variable bias and endogeneity concerns which were discussed earlier.

4. Instrumental variables

In this section we use the number of adult non-workers per household (*Non-workers*) at the city level to instrument for *Competition*. Formally, *Non-workers* equals (1991 values of) the total number of adult non-workers in the city divided by the total number of households in the city.[18] Data source for the variable is Census of India (1991). We note that the variable is lagged to avoid possible reverse causality problem.

As discussed in the introduction, opportunity cost of time spent shopping is an important determinant of search intensity and therefore the level of competition. We treat *Non-workers* as a proxy for this opportunity cost of time. As expected, *Non-workers* and *Competition* show a positive correlation in our sample. The correlation coefficient between the two equals .176 and it rises to .263 if we drop the city of Kozhikode which is an outlier (discussed below).

We note a few important points about the instrument. First, as discussed in the introduction, direct measures of search intensity are typically not available. The approach in the literature is to use some proxy measure instead and our strategy is consistent with this approach. Second, data on non-workers is available every ten years from the Census records with the most recent year being 2001. We prefer 1991 over 2001 values of non-workers to avoid possible reverse causality related problems. However, the correlation coefficient between 1991 and 2001 values of non-workers per household equals .875 and our results are roughly similar with either of them. Third, it is possible that voluntary or involuntary unemployment may be higher in the lesser developed cities. For example,

[18] We follow the Census definition of adults and household and use log values of *Non-workers* to ensure that our results are not sensitive to extreme values of the variable. Adults are individuals above 7 years of age. Distribution of non-workers by other ages at the city level is not reported in the Census. A household is defined as a set of individuals living in one house and sharing a common kitchen.

Non-workers and 1991 values of city-level adult literacy rate (*Literacy rate*), a proxy for overall development of the cities, are inversely correlated in our sample (correlation coefficient of -.12).[19] This implies that higher values of the instrument may pick up the effect of lesser development on efficiency which is likely to be negative. We note that the *negative* relationship between *Non-workers* and efficiency via lesser development implied here counters our identification strategy which is based on a *positive* relationship between *Non-workers* and efficiency via competition. In short, failure to control for differences in overall development across cities is likely to bias the estimated coefficient of *Competition* towards zero (downward bias). Empirical results reported below confirm the downward bias. Fourth, one could argue that the logic of the downward bias mentioned above may not apply to some specific aspects of retailing or to certain characteristics of stores. We looked at a number of variables but we did not find much support for this argument. For example, the correlation coefficient between our instrument and *Size* (averaged at the city level) is negative (correlation of -.19) which supports our claim of the downward bias given that *Size* and efficiency are positively correlated. Fifth, there is some work which suggests that the level of competition in retailing depends on the density of retail shops (number of retail shops per unit of city area). This can create some problem for our first and second stage IV regressions if the density of retail shops and non-workers vary systematically across cities. Unfortunately, data on the number of retail shops in India are not available. The same holds for city-area for some of the cities in our sample.[20] We looked at an alternative measure which is (1991 values of) the number of retailers in the city as a proportion of (adult) city

[19] Data on income levels of the cities are not available.
[20] City-area is reported in the Census records but this is missing for some of the cities in our sample.

population (*Retailer density*).[21] The correlation between our instrument and *Retailer density* is negative (correlation coefficient of -.21). The negative correlation suggests that failure to properly control for the density of retail shops will most probably bias the estimated coefficient of *Competition* towards zero.[22] Another implication here is that higher values of our instrument will be associated with lower values of *Retailer density* weakening the expected positive relationship between *Non-workers* and *Competition* in the first stage of the IV regressions. Lastly, looking at a number of economic and demographic variables, we found *Non-workers* and the number of children per household in the city in 1991 (*Children*) to be positively correlated (correlation coefficient of .393). The positive correlation between *Non-workers* and *Children* is both, good and bad for our identification strategy. The good part is that if search intensity is indeed important in driving competition then non-workers and children should have opposite effects on competition. More children increase the household opportunity cost of time, reducing search intensity and therefore the level of competition. The opposite holds for non-workers. This provides us with a convenient informal test for the validity of our instrument in that we expect contrasting effects of non-workers and children on competition despite the fact that children and non-workers are positively correlated and known to have common covariates (income levels, etc). The bad part is that higher values of our instrument will pick up the effect of more children on competition which we

[21] *Retailer density* captures the number of retailers chasing a consumer (adults). Our main results do not change significantly if we use the ratio of retailers to total (instead of adult) population.

[22] We expect higher values of *Retailer density* to have a positive effect on efficiency via greater competition. Further, higher values of the instrument will pick up lower values of *Retailer density* because of the negative correlation between the two. This implies a negative effect of *Non-workers* on efficiency via *Retailer density* which counters the positive association between *Non-workers* and efficiency (via competition) required for identification. The structure of correlations here suggests a downward bias in the estimated coefficient of competition from the failure to control for *Retailer density* and its covariates.

expect to be negative. This can weaken the positive relationship between non-workers and competition required for identification.

4.1 *Controls*

Our main controls for the IV regressions are based on the discussion above. The controls include store-size (*Size*, $Size^2$) and baseline city characteristics captured by *Retailer density*, *Literacy rate* and *Children*. As discussed above, the motivation for these controls is to arrest some of the downward bias we expect in our results and to properly identify the relationship between *Non-workers* and *Competition* in the first stage of the IV regressions.

4.2 IV *regression results*

IV regression results for the base specification are reported in Table 4. First and second stage results are reported in Panels B and A of the table, respectively. Without any controls, the estimated coefficient of competition equals .244 but this is not statistically significant (column 1, Table 4). The coefficient value increases sharply to .402 when we control for literacy rate and children per household (column 2, Table 4) although the coefficient is still statistically insignificant. In column 3 of Table 4 we control for *Size* and $Size^2$. The estimated coefficient of competition rises further from .402 above to .604 and it is significant at less than 10% level (p-value of .071). Adding *Retailer density* to the list of controls above, we find that the estimated coefficient of competition rises again to .620 and it is significant at 5% level.

For the remaining variables, we find that the effect of *Size* and $Size^2$ on efficiency is significant at less than 1% level while the same for literacy rate is significant at close to 5% level. However, *Children* and *Retailer density* do not show any significant direct effects on efficiency (discussed below).

The first stage IV regression results (Panel B, Table 4) reveal a positive relationship between *Competition* and *Non-workers* as expected. The relationship is slightly weak without any controls (column 1, Panel B, Table 4) but sufficiently strong in the remaining specifications. For our final specification (column 4, Table 4), the F-statistic for the significance of the (excluded) instrument equals 10.9 (p-value of .001) which is sufficiently high for proper identification.[23]

Results from the first stage of IV regressions (Panel B, Table 4) also confirm contrasting effects of *Children* and *Non-workers* on competition. The former effect is negative while the latter, positive. Both these effects are significant at less than 5% level. As discussed above, these contrasting effects raise our confidence in the validity of the instrument.

Earlier we argued that the relationship between our instrument and the level of competition is somewhat weakened by the city of Kozhikode.[24] In column 5 of Table 4 we report regression results with Kozhikode dropped from the sample. The estimated coefficient of competition remains positive and significant at less than 5% level although

[23] According to Stock and Watson (2003), an F-statistic above 10 suggests that the instrument is not weak.
[24] Using the same dataset, Amin (2007) looks at the effect of computer usage and labor regulations on employment. He reports that Kozhikode is an outlier and dropping the city from the sample has a significant effect on the estimation results.

it declines in value from .620 to .417. As expected, the F-statistic on the test of the excluded instrument (Panel B, Table 4) rises sharply from 10.9 to 24.3.[25]

We experimented with one more specification which we briefly discuss here. We found above that *Children* and *Retailer density* do not show any significant direct effects on efficiency. This implies that we could use these two variables as additional instruments. The advantage of using these additional instruments is that it makes our system overidentified, allowing us to test for the exogeneity of the instruments. Regression results for this specification (not reported) were roughly similar to our final results above as reported in column 4 of Table 4. The estimated coefficient of competition equaled .712 with a p-value of .018. Hanson overidentification *J* test statistic equaled .339 (p-value of .560) and it did not reject the null hypothesis that the instruments are exogenous.

4.3 *Robustness of* IV *results*

Robustness checks for the IV regressions are reported in Table 5. So far we controlled for store-size and some lagged city level variables. We now check for the robustness of the final results above (column 4, Table 4) with respect to a number of contemporaneous store and city characteristics.

Since our instrument varies at the city level, we cannot use city fixed effects as controls. In lieu of these fixed effects, we control for two more measures of overall development of cities for additional robustness. These are the ratio of females to males in

[25] In regression results not reported we dropped all stores which were less than 5 years old to address any possible concerns about the contemporaneous nature of the *Size* variable. This did not change our results in any significant way. For example, for the final specification (column 4, Table 4) and dropping stores less than 5 years old yielded a coefficient value of .663 (p-value of .046) for the competition variable.

the city (*Sex ratio*) and a dummy variable equal to 1 if a store is located in a metropolitan city and 0 otherwise (*Metro*). The metropolitan cities are New Delhi, Mumbai, Kolkata, Chennai, Hyderabad and Bangalore. Sex ratio is commonly used as a proxy for overall development. The metropolitan cities mentioned above are the richest, most developed and the biggest cities in the country and also the main beneficiaries of the ongoing retail boom in the country. Adding *Sex ratio* and *Metro* to the set of controls above did not change our results much. The estimated coefficient of *Competition* declined marginally from .620 above to .610 but remained significant at less than 5% level (p-value of .034). *Sex ratio* and *Metro* show positive but insignificant (at 10% or less) effects on efficiency.

Next, we controlled for all the variables used in the OLS robustness analysis (listed in column 4, Table 3). IV regressions results with these additional controls are reported in column 1 of Table 5. The estimated coefficient of competition here equals 1.53 significant at less than 5% level. We note that the estimated coefficient value of 1.53 here is much higher than what we found above (.610) which lends support to the downward bias view discussed above. The direction of the relationship between efficiency and the various controls here is similar to what we found in the OLS regressions but there is some difference in the significance level of these relationships. Specifically, *Overdraft* and *Audited* show only weak positive effects on efficiency here (significant at close to 10% level) while the negative effect of informal competition on efficiency is significant at less than 5% level.

In column 2 of Table 5 we report regression results controlling for various regulatory measures defined at the city level. These measures are constructed from the Enterprise survey and reflect the extent to which the following are an obstacle to the

performance of stores: tax rates ($Taxes_c$), corruption ($Corruption_c$), land laws ($Land\ Laws_c$), difficulty in obtaining permits and licenses ($Permits_c$), functioning of courts ($Courts_c$) and availability of skilled labor ($Skill\_shortage_c$).[26] With these additional controls, the estimated coefficient of competition rises from 1.53 above to 1.79 significant at less than 5% level.

We added a number of additional controls to the specification in column 2 of Table 5 but found that our main result of a positive and significant effect of competition on efficiency remained intact. Some of these additional controls are: a dummy variable equal to 1 if a store owns a generator and 0 otherwise, percentage of a store's electricity derived from generator, a dummy variable equal to 1 if a store has a checking account and 0 otherwise, a dummy variable equal to 1 if a store has a line of credit and 0 otherwise, a dummy variable equal to 1 if a store reported no need to borrow from external sources during the last fiscal year and 0 otherwise, percentage of stores' senior management's time spent in dealing with business regulations and years of managerial experience. The estimated coefficient of competition here remained almost unchanged at 1.73 (compared to 1.79 above) and was significant at less than 5% level (p-value of .049).

One concern with the robustness results above could be that the F-statistic on the test of the excluded instrument, though significant at close to 1% level, is somewhat low (Panel B, Table 5). For example, for the final specification in column 2 of Table 5, the F-statistic equals 6.43 (p-value of .011) which is below the recommended value of 10 in the literature. As for the main IV specification above, the low value of the F statistic here is largely due to the city of Kozhikode. Dropping this city from the sample we found that the F-statistic rose sharply from 6.43 to 10.6 (p-value of .002) for our final specification

[26] A formal definition of these measures is provided in Table 1.

in column 2, Table 5. The estimated coefficient of competition remained significant at less than 5% level (p-value of .017) although it declined in value from 1.79 to 1.35.[27]

Summarizing, the IV regression results reveal a significant positive relationship between competition and efficiency which is robust to a number of controls. The findings confirm our initial prediction of a downward bias in the estimated competition-efficiency relationship which explains why our results are somewhat weak without the main controls (Table 4). However, the downward bias also suggests that our main results above are on the conservative side and controlling for other relevant omitted variables will most likely strengthen these results. A comparison of the OLS and IV regression results shows that the estimated coefficient of competition is higher for the IV regressions. One reason for this could be measurement error with the *Competition* variable.[28] Additionally, we have shown that the level of competition in retailing depends on household characteristics such as the number of adult non-workers and children per household which we believe to be an important contribution of the paper to the small but growing literature on how consumer characteristics shape the level of competition in retailing. In a companion paper, we explore this issue in more detail.

5. Conclusion

The paper analyzes the impact of competition on the efficiency of retail stores in India. Our results predict a large effect of pro-competitive reforms on the efficiency of stores.

[27] All robustness specifications show a negative effect (significant at less than 5% level) of *Children* on the level of competition in the first stage of the IV regressions.

[28] The other possibility is that *Competition* may be (inversely) correlated with some other determinant of efficiency pushing the estimated coefficient of competition towards zero in the OLS specification.

Also, there is considerable scope for such reforms with over 60% of the retailers facing no significant competition. A retail boom from more competition is a likely outcome.

A number of issues need further analysis. First, it is not clear why competition is low in the sector. Previous studies suggest that this may be due to high barriers to entry (Djankov et al 2002), burdensome labor laws (Amin, 2007) or inefficient courts (Djankov et al, 2003). A formal analysis of this question is important given the significant gains from more competition predicted by the present work. Another related issue is precisely who benefits from more competition. Are these the small or the large retailers? This question is important to assess the distributional consequences of reform.

References

[1] Amin, M (2007), "Computer Usage and Labor Regulations in India's Retail Stores," mimeograph**,** http://rru.worldbank.org/PapersLinks/resources.aspx?id=7532. Forthcoming, Policy Research Working Paper Series, the World Bank, Washington DC.

[2] Autor, D.H., L.F. Katz and M.S. Kearney (2006), "The Polarization of the U.S. Labor Market," NBER Working Paper 11986.

[3] Autor, D.H., L.F. Katz and A.B. Krueger (1998), "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics*, 113, 1169-1213.

[4] Baily, Martin N. and Robert Solow (2001), "International Productivity Comparisons Built from the Firm Level," *Journal of Economic Perspectives*, 15(3), pp. 151-172.

[5] Banga, Rashmi (2005), "Critical Issues in India's Services-Led Growth," Working Paper No. 171, *Indian Council for Research on International Economic Relations*.

[6] Blanchflower, David and Stephen Machin (1996), "Product Market Competition, Wages and Productivity," *Annals of Economics and Statistics*, 17(41), pp. 219-53.

[7] Bliss, Christopher and Rafael Di Tella (1997), "Does Competition Kill Corruption," *Journal of Political Economy*, 105(5), pp. 1001=1023.

[8] Bruhn, Miriam (2007), "License to Sell: The Effect of Business Registration Reform on Entrepreneurial Activity in Mexico," Mimeograph, Massachusetts Institute of Technology. (http://econ-www.mit.edu/graduate/candidates/download_res.php?id=298).

[9] Cainelli, Guilio, Rinaldo Evangelista and Maria Savona (2006), "Innovation and Economic Performance in Services: A Firm-Level Analysis," *Cambridge Journal of Economics*, 30, pp. 435-458.

[10] Calem, P.S. and L.J. Mester (1995), "Consumer Behavior and the Stickiness of Credit-Card Interest Rates," *American Economic Review*, 85(5), pp. 1327-1336.

[11] Diamond, P. (1971), "A Model of Price Adjustment," *Journal of Economic Theory*, 3, pp. 156-168.

[12] Djankov, Simeon and Peter Murrell (2002), Enterprise Restructuring in Transition: A Quantitative Survey," *Journal of Economic Literature*, 40(3), pp. 739-792.

[13] Djankov, Simeon, Rafael La Porta, Florencio Lopez-De Silanes and Andrei Shleifer (2002), "The Regulation of Entry," *Quarterly Journal of Economics*, 117(1), pp. 1-37.

[14] Djankov, Simeon, Rafael La Porta, Florencio Lopez-De Silanes and Andrei Shleifer (2003), "Courts," *Quarterly Journal of Economics*, 118(2), pp. 453-517.

[15] Goldman, Arieh and J.K. Johansson (1978), "Determinants of Search for Lower Prices: An Empirical Assessment of the Economics of Information Theory," *Journal of Consumer Research*, 5, pp. 176-186.

[16] Goldman, Arieh, S. Ramaswami and Robert E. Krider (2002), "Barriers to the Advancement of Modern Food Retail Formats: Theory and Measurement," *Journal of Retailing*, 78, pp. 281-295.

[17] Gordon, J. and P. Gupta (2004), "Understanding India's Services Revolution," Working Paper No. 04/171, *International Monetary Fund*.

[18] Horn, Henrik, Harold Land and Stefan Lundgren (1995), "Managerial Effort Incentives, X-Inefficiency and International Trade," *European Economic Review*, 39(1), pp. 117-38.

[19] Januszewski, Silke, Jens Kole and Joakim Winter (2001), "Product Market Competition, Corporate Governance and Firm Performance: An Empirical Analysis of Germany," Working paper, Center of European Economic Research, Mannheim.

[20] Knittel, C. (1997), "Interstate Long Distance Rates: Search Costs, Switching Costs and Market Power," *Review of Industrial Organization*, 12(4), pp. 519-536.

[21] Krueger, A.B. and J. Angrist (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments?" *Journal of Economic Perspectives*, 15, pp. 69-85.

[22] Marvel, Howard P. (1976), "The Economics of Information and Retail Gasoline Price Behavior: An Empirical Analysis," *Journal of Political Economy*, 84, pp. 1003-59.

[23] Nickell, Stephen J., Sushil Wadhwani and Martin Wall (1992), "Productivity Growth in U.K. Companies, 1974-86," *European Economic Review*, 36(5), pp. 1066-85.

[24] Nickell, Stephen J. (1996), "Competition and Corporate Performance," *Journal of Political Economy*, 104(4), pp. 724-746.

[25] Schmidt, Klaus (1997), "Managerial Incentives and Product Market Competition," *Review of Economic Studies*, 64(2), pp. 191-213.

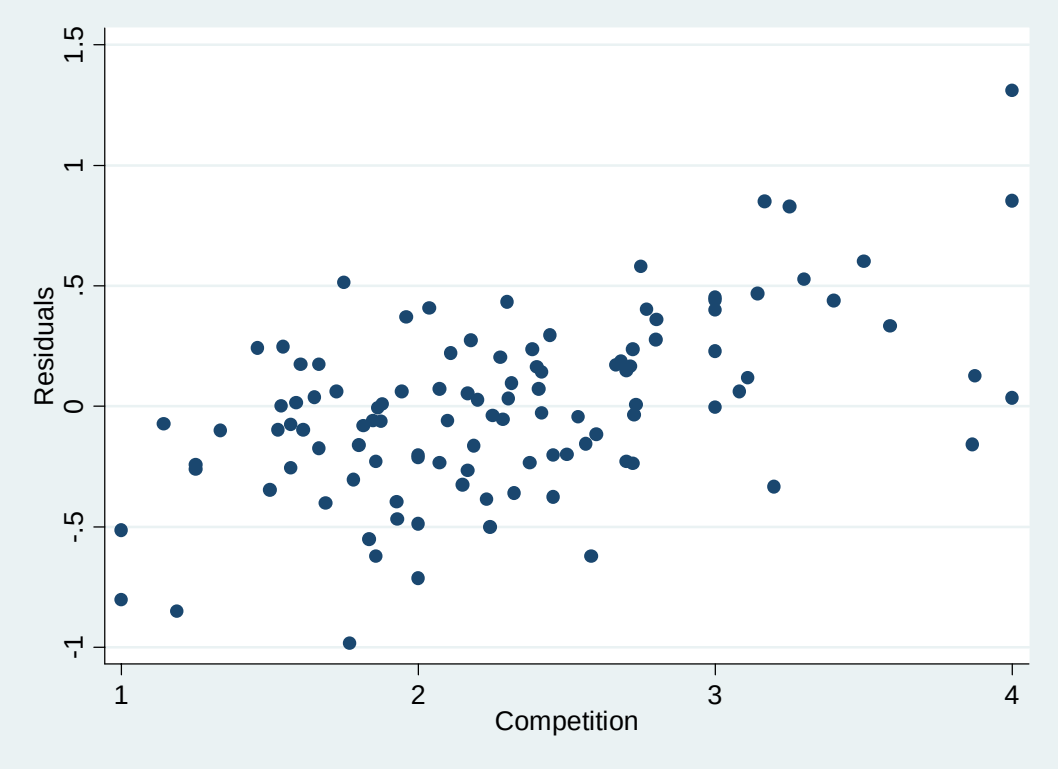
[26] Sorenson, Alan T. (2000), "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs," *Journal of Political Economy*, 108(4), pp. 833-850.

[27] Stahl II, Dale O. (1989), "Oligopolistic Pricing with Sequential Consumer Search," *American Economic Review*, 79(4), pp. 700-712.

[28] Stock, J. and M. Watson (2003), Introduction to Econometrics, Addison-Wesley, Boston.

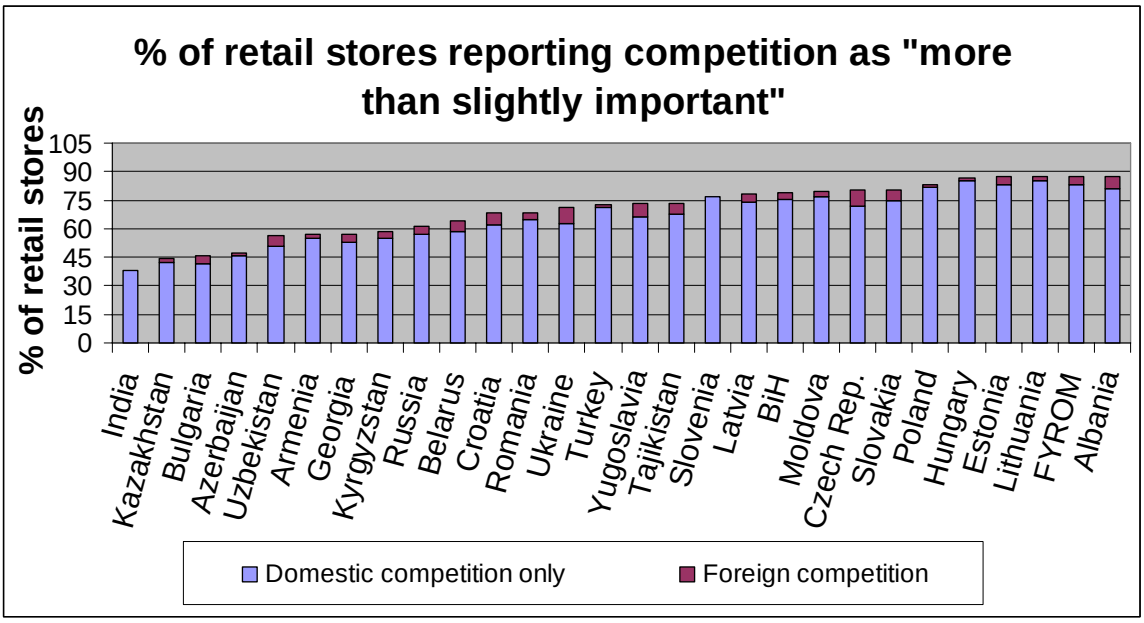
[29] Stigler, George J. (1961), "The Economics of Information," *Journal of Political Economy*, 69(3), pp. 213-225.

Figure 1



The Y axis in the graph above plots residuals obtained by regressing *Competition* on city and store-type fixed effects. The X axis plots *Competition* as defined in the sections above.

Figure 2

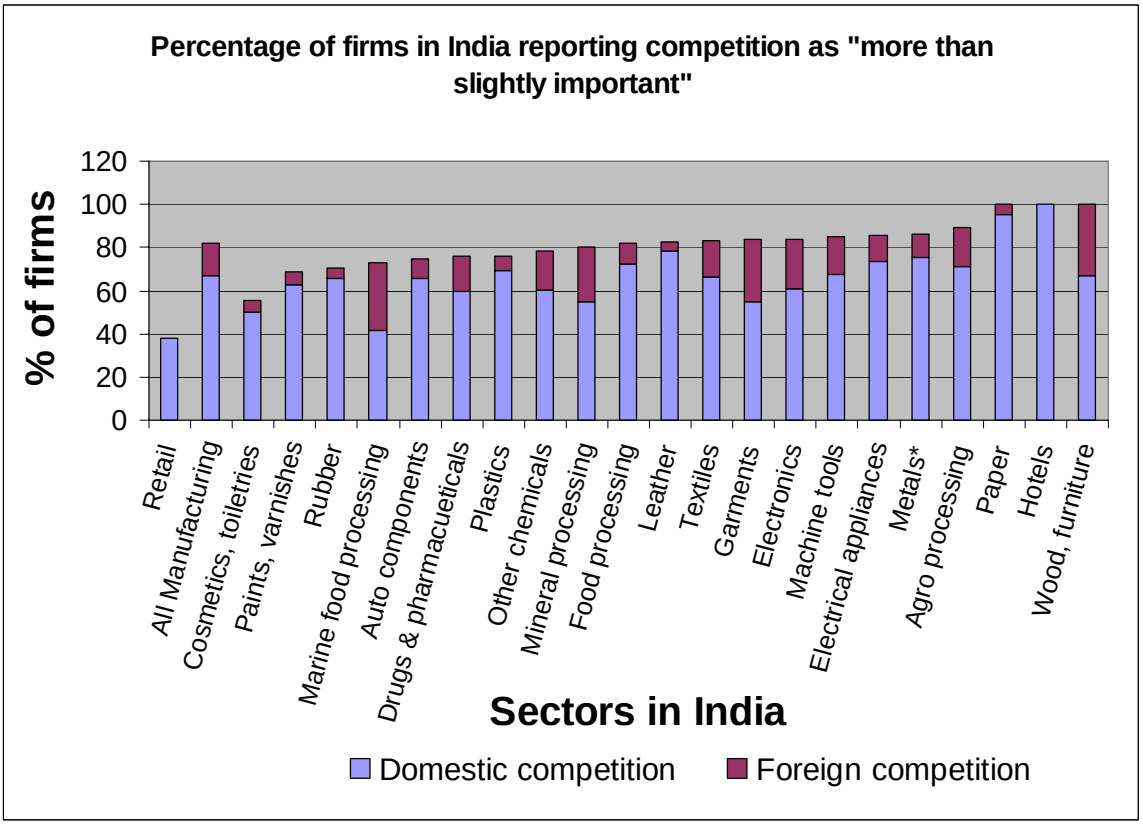


For the BEEPS countries: "Domestic competition only" in the graph above is defined as the percentage of all stores who report "Fairly important" or "Very important" on part (a) of the question stated below. "Foreign competition" in the graph above is defined as percentage of all stores who do not face domestic competition (score of 1 or 2 on pat (a) below) but report "Fairly important" or "Very important" on part (b) of the question below.

Competition question in BEEPS:
How would you rate the importance of each of the following factors on key decisions about your business with respect to reducing the production costs of existing products or services:

	Not at all important	Slightly important	Fairly important	Very important
a) Pressure from domestic competitors	1	2	3	4
b) Pressure from foreign competitors	1	2	3	4
c) Pressure from customers	1	2	3	4

Figure 3



*Metals: Structural metals and metal products

For the non-retail sectors: "Domestic competition" in the graph above is defined as the percentage of all stores who report "Fairly important" or "Very important" on part (a) of the question stated below. "Foreign competition" in the graph above is defined as percentage of all stores who do not face domestic competition (score of 1 or 2 on pat (a) below) but report "Fairly important" or "Very important" on part (b) of the question below.

Competition question in Manufacturing-India survey:

How important are each of the following influences on your establishment to reduce the production costs of existing products and/or to develop new products.

	Not at all important	Slightly important	Fairly important	Very important
a) Pressure from domestic competitors	1	2	3	4
b) Pressure from foreign competitors	1	2	3	4
c) Others	1	2	3	4

Figure 4

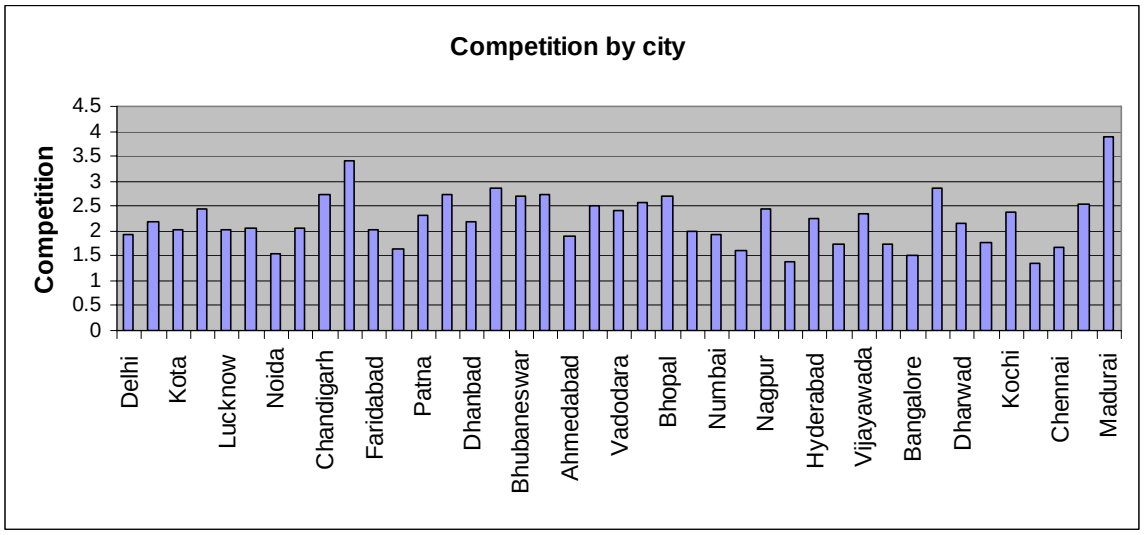

Competition by city
Competition
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Delhi
Kota
Lucknow
Noida
Chandigarh
Faridabad
Patna
Dhanbad
Bhubaneswar
Ahmedabad
Vadodara
Bhopal
Numbai
Nagpur
Hyderabad
Vijayawada
Bangalore
Dharwad
Kochi
Chennai
Madurai

Figure 5

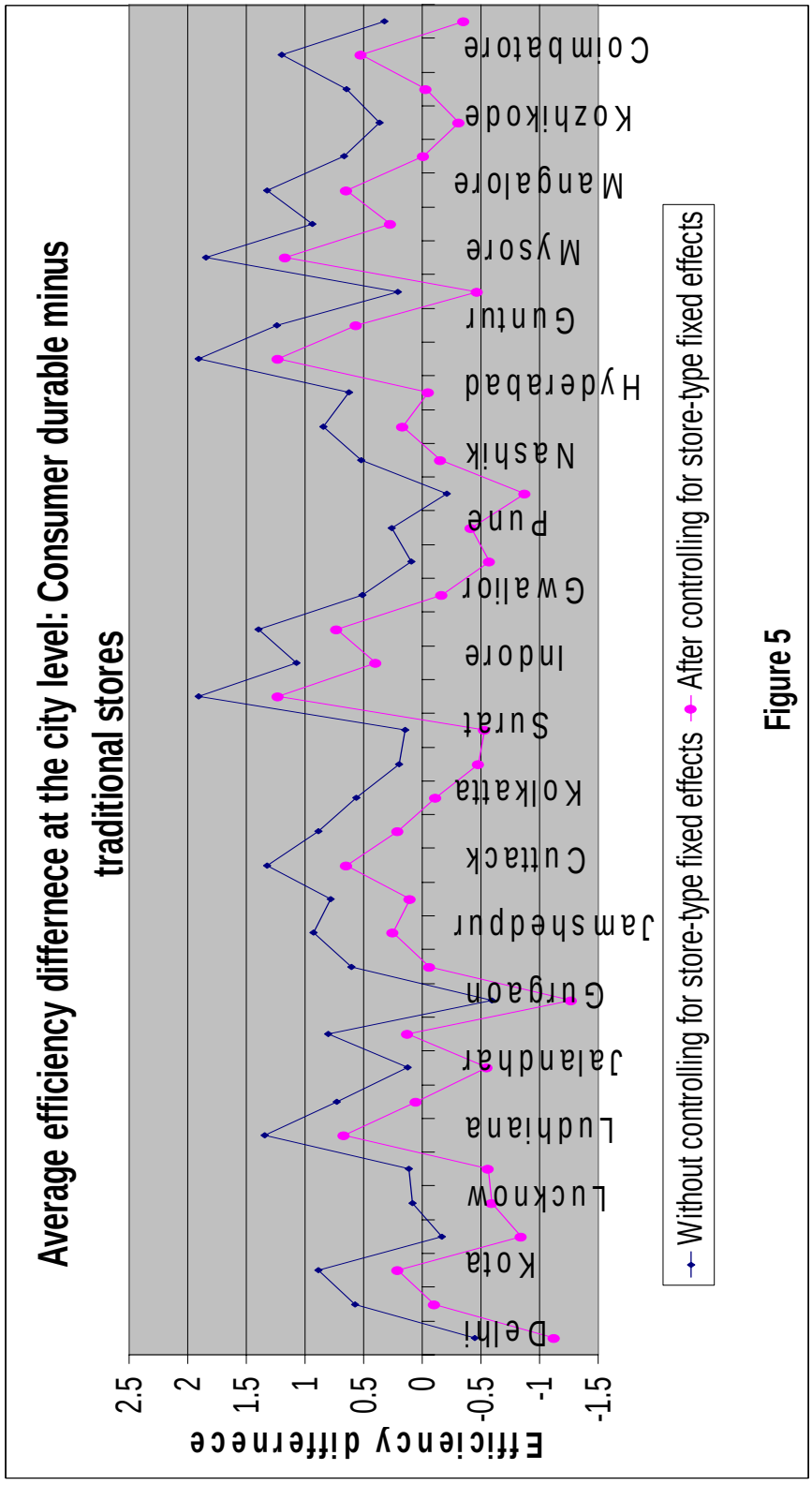

Average efficiency differnece at the city level: Consumer durable minus traditional stores
Efficiency difference
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Delhi
Kota
Lucknow
Ludhiana
Jalandhar
Gurgaon
Jamshedpur
Cuttack
Kolkatta
Surat
Indore
Gwalior
Pune
Nashik
Hyderabad
Guntur
Mysore
Mangalore
Kozhikode
Coimbatore
Without controlling for store-type fixed effects
After controlling for store-type fixed effects
Figure 5

Table 1: Description of Main Variables

Variable	Description
"Last fiscal year" below means fiscal year 2005-06.	
Efficiency	Log of total sales (in Rs.) divided by total employment during the last fiscal year. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Competition	Average score at the city-store type level reported by the stores on the following question asked in the survey: For this store, how important are each of the following influences over prices of its main products? a. Pressure/Influence from domestic competitors Not at all important (1), Slightly important (2), Fairly important (3) and Important (4). Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Size	Total selling area of the store measured in square feet (log values). Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
CFE (City fixed effects)	A set of 41 dummy variables for the cities in which the Enterprise survey was conducted. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
SFE (Store-type fixed effects)	A set of three dummy variables: *Traditional*, *Consumer Durable* and *Modern Format* as defined below. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Traditional	A dummy variable equal to 1 if a store is a "traditional store" and 0 otherwise. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Consumer Durable	Dummy equal to 1 is store is a "consumer durable" store and 0 otherwise. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Modern Format	A dummy variable equal to 1 is a store is a "modern format" store and 0 otherwise. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Non-workers	Log of total adult non-workers in the city divided by total number of households in the city (1991 vales). Source: Census of India (1991).
Children	Log of total number of children (below 7 years) divided by the total number of households in the city in 1991. Source: Census of India (1991)
Literacy	Log of total number of adults in the city who are literate divided by total city population (1991 values). Source: Census of India (1991)
Retailer density	Log of total employment in retail and distribution sector divided by adult city population in 1991. Source: Census of India (1991).

Table 1: Description of Other Variables

Variable	Description
"Last fiscal year" below means fiscal year 2005-06.	
Employment (Current employment at the store level)	Total number of workers working at the store in the last fiscal year. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Age	2006 minus the year the store began operation. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Outage	Hours of power failure in a typical day. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Overdraft	A dummy variable which equals 1 if a store has overdraft facility and 0 otherwise. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Computers	A dummy variable which equals 1 if a store uses a computer for running its business and 0 otherwise. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Informal_Competition	Average score at the city-store type level on the response of stores to the following question: For this store, how important is the following influences over prices of its main products? Pressure/Influence from unorganized trade (hawkers, traders sitting on pavement, people selling from home, people selling spurious good)" Not at all important (1), Slightly important (2), Fairly important (3), Important (4). Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Non-payment	Percentage of store's annual sales in 2005-06 that were not paid for. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Inventory	Number of days of inventory currently maintained by the store. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Female	Dummy equal to 1 if a store has a female as a principal owner and 0 otherwise. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Audited	Dummy equal to 1 if a store had its accounts audited by an external auditor in 2005-06 and 0 otherwise. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)

$Courts_{cs}$	Average score at the city-store type level reported on the following question in the survey: Do you think that the functioning of courts is No obstacle (0), a Minor obstacle (1), a Moderate obstacle (2), Major obstacle (3) or a Very Severe obstacle (4) to the current functioning of this store. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
$Theft_{cs}$	Percentage of stores in each city-store type cell that report incidence of theft in the last fiscal year. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
$Skill\_shortage_{cs}$	Average score at the city-store type level reported on the following question in the survey: Is an inadequately educated workforce No obstacle (0), a Minor obstacle (1), a Moderate obstacle (2), Major obstacle (3) or a Very Severe obstacle (4) to the current operation of this store? Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
$Tax\ rates_{cs}$	Average score at the city-store type level reported by stores on the following question in the survey: Is/Are tax rates No obstacle (0), minor obstacle (1), Moderate obstacle (2), Major obstacle (3) or Very Severe obstacle (4) to the current operations and opportunities for growth of this store? Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
$Corruption_{cs}$	Average score at the city-store type level reported by stores on the following question in the survey: Is/Are corruption No obstacle (0), minor obstacle (1), Moderate obstacle (2), Major obstacle (3) or Very Severe obstacle (4) to the current operations and opportunities for growth of this store? Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
$Permits_{cs}$	Average score at the city-store type level reported by stores on the following question in the survey: Is/Are business licensing and permits No obstacle (0), minor obstacle (1), Moderate obstacle (2), Major obstacle (3) or Very Severe obstacle (4) to the current operations and opportunities for growth of this store? Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Sex ratio	Ratio of females to males in the city in 2001. Source: Census of India (2001)
Metro	Dummy variable equal to 1 for stores located in the metropolitan cities of Delhi, Mumbai, Hyderabad, Kolkatta, Bangalore and Chennai and 0 otherwise. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)

Tax rates$_c$	Average score at the city level reported by stores on the following question in the survey: Is/Are tax rates No obstacle (0), minor obstacle (1), Moderate obstacle (2), Major obstacle (3) or Very Severe obstacle (4) to the current operations and opportunities for growth of this store? Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Corruption$_c$	Average score at the city level reported by stores on the following question in the survey: Is/Are corruption No obstacle (0), minor obstacle (1), Moderate obstacle (2), Major obstacle (3) or Very Severe obstacle (4) to the current operations and opportunities for growth of this store? Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Permits$_c$	Average score at the city level reported by stores on the following question in the survey: Is/Are business licensing and permits No obstacle (0), minor obstacle (1), Moderate obstacle (2), Major obstacle (3) or Very Severe obstacle (4) to the current operations and opportunities for growth of this store? Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Courts$_c$	Same as *Courts*$_{cs}$ defined above except that the average is taken at the city level. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Skill_shortage$_c$	Same as *Skill_shortage*$_{cs}$ defined above except that the average is taken at the city level. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)

Table 2: OLS, Dependent variable: Log (Sales/Employment)

	(1)	(2)	(3)	(4)
Competition	.143*	.141**	.391***	.343***
	(.063)	(.044)	(.001)	(.002)
Size		.216***	.181***	1.33***
		(.000)	(.000)	(.000)
CFE (City fixed effects)			Yes	Yes
SFE (Store-type fixed effects):			Yes	Yes
Traditional			-.026	-.317**
			(.861)	(.016)
Consumer Durable			.340**	.002
			(.022)	(.991)
$Size^2$				-.106***
				(.000)
R^2 (Adjusted)	.005	.050	.165	.212
Sample size	1907	1897	1897	1897

p-values in parentheses; all regressions use Huber-White correction for heteroskedasticity allowing for clustering by city-store type. Sample size varies across columns due to missing observations on *Size*.

Table 3: OLS Dependent variable: Log (Sales/Employment)

	(1)	(2)	(3)	(4)
Competition	.307***	.282***	.269**	.245**
	(.004)	(.010)	(.011)	(.018)
Size	.930***	.925***	.968***	.914***
	(.000)	(.000)	(.000)	(.000)
$Size^2$	-.063***	-.065***	-.073***	-.070***
	(.000)	(.000)	(.000)	(.000)
City & store-type fixed effects	Yes	Yes	Yes	Yes
Employment	-.010***	-.010***	-.010***	-.010***
	(.001)	(.001)	(.001)	(.001)
Age	.006***	.005***	.006***	.005**
	(.003)	(.008)	(.004)	(.028)
Outage		-.009	-.011	-.011
		(.481)	(.413)	(.409)
Overdraft		.266***	.218***	.169**
		(.001)	(.008)	(.028)
Computers			.376***	.357***
			(.001)	(.003)
Inventory			.005***	.005***
			(.004)	(.004)
Informal_Competition			-.071	-.048
			(.547)	(.691)
Non-Payment				-.019**
				(.026)
Audited				.246***
				(.008)
Female				-.218*
				(.060)
$Theft_{cs}$				-.584
				(.121)
R^2 (Adjusted)	.233	.239	.251	.259
Sample size	1897	1869	1869	1844

p-values in brackets; all standard errors are Huber-White robust and clustered by city-store type.

Table 4: IV Base regressions

	(1)	(2)	(3)	(4)	(5)
Panel A: Second stage IV regressions					
Dependent variable: Log (Sales/Employment)					
Competition	.244	.402	.604*	.620**	.417**
	(.642)	(.212)	(.071)	(.050)	(.040)
Literacy		2.14**	1.81*	1.84*	1.03
		(.024)	(.059)	(.053)	(.190)
Children		.054	.141	.246	.044
		(.878)	(.695)	(.501)	(.880)
Size			1.34***	1.33***	1.36***
			(.000)	(.000)	(.000)
$Size^2$			-.101***	-.101***	-.103***
			(.000)	(.000)	(.000)
Retailer density				.239	.318
				(.493)	(.253)
Sample size	1907	1907	1897	1897	1864
Panel B: First stage IV regressions					
Dependent variable: *Competition*					
Excluded instrument					
Non-workers	.740*	1.19***	1.22***	1.25***	1.69***
	(.066)	(.002)	(.001)	(.001)	(.000)
Included instruments					
Literacy		-1.32	-1.37	-1.35	-.398
		(.156)	(.146)	(.148)	(.637)
Children		-.932***	-.939***	-.803**	-.733**
		(.004)	(.004)	(.016)	(.021)
Size			.176**	.176**	.192**
			(.028)	(.026)	(.015)
$Size^2$			-.015**	-.015**	-.015**
			(.039)	(.034)	(.029)
Retailer density				.359	.348
				(.183)	(.157)
F statistic for the significance of the excluded instrument	3.45*	10.2***	10.8***	10.9***	24.2***
	(.066)	(.002)	(.001)	(.001)	(.000)

p-values in parentheses; all regressions use Huber-White correction for heteroskedasticity allowing for clustering by city-store type.

*** denotes significant at 1% or less, ** denotes significant at 5% or less and * denotes significant at 10% or less. Sample size in columns 1-4 varies due to missing observations. The sample used for column 5 excludes all stores located in the city of Kozhikode.

Table 5: IV Robustness

	(1)	(2)
Panel A: Second stage IV regressions		
Dependent variable: *Efficiency*		
Competition	1.53** (.042)	1.79** (.045)
Size	.891*** (.000)	.847*** (.000)
Size^2	-.066*** (.000)	-.061*** (.000)
Retailer density	-1.41 (.246)	-1.48 (.205)
Children	.377 (.462)	.720 (.294)
Literacy	.480 (.665)	.639 (.618)
Sex ratio	1.83 (.264)	2.19 (.234)
Metro	1.03* (.078)	1.35* (.064)
Store-type fixed effects	Yes	Yes
Employment	-.011*** (.000)	-.011*** (.000)
Age	.005* (.062)	.004 (.142)
Outage	.001 (.937)	.016 (.511)
Overdraft	.159* (.095)	.162*** (.086)
Computers	.395*** (.005)	.369*** (.006)
Inventory	.002 (.355)	.002 (.476)
Informal_Competition	-1.24** (.034)	-1.34** (.043)
Non-Payment	-.021* (.055)	.023** (.045)
Audited	.174 (.105)	.101 (.386)
Female	-.230* (.087)	-.278* (.054)
Theft$_{cs}$	.816 (.352)	.215 (.796)
Skill_shortage$_c$		.397 (.235)
Courts$_c$		.452 (.250)
Tax rates$_c$		-.121 (.698)
Corruption$_c$		.098 (.567)
Land laws$_c$		-.190 (.315)
Permits$_c$		-.364 (.341)
Sample size	1844	1844
Panel B: First stage IV regressions		
Dependent variable: *Competition*		
Non-workers	.690** (.011)	.734** (.013)
F statistic for significance of the excluded instrument	6.71** (.011)	6.43** (.013)

p-values in parentheses; all regressions use Huber-White correction for heteroskedasticity allowing for clustering by city-store type. *** denotes significant at 1% or less, ** denotes significant at 5% or less and * denotes significant at 10% or less. Sample size varies due to missing observations.