WPS4514

POLICY RESEARCH WORKING PAPER 4514

Competition and Demographics

Mohammad Amin

The World Bank
Financial and Private Sector Development Vice-Presidency
Enterprise Analysis Unit
February 2008

Policy Research Working Paper 4514

Abstract

Mainstream economics views demographic changes in the structure of households as of little relevance for the behavior of firms or the functioning of markets. The present paper dispels this view by arguing that changes in the number of non-workers could affect the intensity with which consumers search for best prices and therefore the level of competition. The author also analyzes the relationship between income and competition, which some studies suggest is negative. The author argues that the negative relationship is most likely due to the demographic factors discussed.

This paper—a product of the Financial and Private Sector Development Vice-Presidency, Enterprise Analysis Unit—is part of a larger effort in the department to better understand the functioning of the private sector. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at mamin@worldbank.org.

Competition and Demographics

Mohammad Amin*
World Bank

Keywords: Competition, Demographics
JEL: D24, L11, L81

*Enterprise Analysis Unit, the World Bank, Washington DC, 20433. Email: mamin@worldbank.org. Phone: (202)-473-1915. I would like to thank Simeon Djankov, Alvaro Gonzalez, Ernesto Lopez-Cordova and Siddharth Sharma for helpful comments. All errors are my own.

1. Introduction

Traditionally, mainstream economic policies have been largely immune to social and demographic changes in the structure of households. Perhaps the thinking is that these factors have little to do with the behavior of firms or the functioning of markets. The present paper dispels such thinking by arguing that changes in the number of adult non-workers per household (henceforth, non-workers) could have a significant effect on the level of competition in the retail sector. More non-workers imply lower household shopping time opportunity-cost leading to more intensive search for best prices and deals and consequently more competition.

While our main focus is on the relationship between non-workers and competition, we also look at other household attributes such as income levels, number of children per household and gender composition. Below we argue why these factors may be important for the level of competition drawing on the broader literature on search intensity and competition.

Our results strongly support the predicted relationship between non-workers and competition. Estimates suggest that a decrease in non-workers (average value at the city level) from its 25^{th} percentile to its lowest value leads to a 15.9 percentage point decline in the proportion of stores in the city that face significant competition, a large effect given that only 38.2 percent of the stores in the sample face significant competition. The relationship is robust to a number of controls including the ones for per capita income, literacy rate, financial and physical infrastructure availability, number of retailers in the city, various store characteristics, etc.

We contrast the effect on competition of non-workers and the number of children per household to shed light on the mechanism through which non-workers affect the level of competition. Search theory predicts that more children are likely to increase household shopping time opportunity-cost lowering search intensity and thereby (reducing) the level of competition. Our informal test is based on two observations. First, children and non-workers are positively correlated (correlation of .488 in our data) and yet we expect their effects on competition to be in opposite directions. A direct implication of this is that controlling for children should increase the estimated effect of non-workers on competition. Second, children and non-workers are likely to be correlated with income levels and overall development of cities in the same direction (evidence provided below). Hence, if non-workers is spuriously picking up the effect of income levels or overall development on competition, then controlling for children should (partly) eliminate some of this spurious correlation causing the estimated effect of non-workers on competition to decrease. Our results below easily pass this test. That is, controlling for children causes the estimated coefficient of non-workers to increase by over 62%.

There is no previous work on the relationship between non-workers or other socio-demographic consumer-household attributes and competition. Goldman et al (2002) argue that the number of non-working adults in the household is a good surrogate for the "household shopping time opportunity-cost". However, the authors do not examine the implications of this for the level of competition in the market.[1] There is some hint in the literature that household income may affect competition through search intensity. Lower income households have higher marginal utility of income (savings from

[1] 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.

more intensive search) and also lower opportunity cost of time. Hence, they are likely to search more intensively implying a negative correlation between income and the level of competition (Marvel 1976, Hoch et al 1995).[2] Findings in the present paper do show a strong negative relationship between income and competition but this relationship is completely wiped out once we control for non-workers. Hence, our results cast doubt on the results in the literature that attribute the income-competition relationship to differences in the marginal utility of income across rich and poor households. It could be spuriously driven by differences in opportunity cost of shopping time due to difference in the number of non-workers per household. We discuss this issue in detail below.

At a broad level, the paper contributes to the small but growing literature that links the level of competition in product markets to consumer attributes (Ausubel 1991, Giulietti et al 2005, Prendergast 2002, Waterson 2003). Studies show that consumer's income level, past experience with switching suppliers (Giulietti et al), commuting distance to shops (Baron et al 2004, Claycombe 2000), perceptions about search and switching costs and whether expected benefit (from comparison shopping) are long-term or short-term (Giulietti et al), competition policy and the cost of switching suppliers (Lalive and Schmutzler 2007, Giulietti et al, Calem and Mester 1995, Knittel 1997) are some of the factors that affect consumer behavior and thereby the level of competition in the market.

Our motivation in exploring the link between competition and non-workers is two-fold. First, it provides a better understanding of the determinants of competition in retailing. The importance of search cost for market competition has been empirically

[2] The counterargument here is that poorer households consumer less and this could lower their propensity to search if there are fixed costs in searching.

established but only for the case of developed countries. The present paper makes a first attempt at extending this literature to a developing country. Second, our findings are important for the appropriate design and targeting of competition policy. As economic development spreads to the relatively poorer countries, household structure in these countries is likely to change as it is happening in India. Additional policy measures that factor in the structure of households could play an important role in ensuring that markets remain truly competitive. Tying of competition policy to specific consumer attributes (different from ours) is also recommended by Giulietti et al based on their finding that pensioner households, low-income households, rural dwellers and consumers with disabilities benefited differently from the recent deregulation of the U.K. energy markets.

The rest of the paper is as follows. In section 2 we describe the data and our main variables. In section 3 we present our main empirical findings. Robustness checks are reported in section 4 while in section 5 we focus on the income-competition relationship. A summary of the main findings is provided in the concluding section.

2. Data and Main Variables

A formal definition of all the variables used in the regressions is provided in Table 1. Summary statistics for the main variables are reported in Table 2. Correlation coefficients for the main variables are provided in Table 3 and discussed in detail below. We use store level data for India's retail sector collected by the World Bank's Enterprise Surveys in

2006.[3] The data are a cross section of 1,948 retail stores spread over 16 states and 41 cities of India.[4]

We choose to study India because the country is witnessing significant demographic changes of the kind discussed above and, according to our results, these changes are likely to have significant effects on functioning of the retail sector in the near future. Between 1991 and 2001, a majority of states in India showed a significant decline in the number of adult non-workers per household. For the fourteen major states of India, the decline averaged 7.2 percent with a high of 18.6 percent in the state of Haryana, 12.7 percent in Kerala and 12.6 percent in Punjab (Figure 1). The downward trend is likely to continue in the future being part of a larger socio-economic-demographic churning in the country characterized by greater participation of men and women in the labor market, increasing proportion of young in total adult population, the emergence of nuclear family as a social norm, break-up of the joint family system due to increased mobility of labor in search of economic opportunities and lower fertility rates.

2.1 *Dependent variable*

In one survey question stores were asked how important the pressure from domestic competitors is over prices of the store's main products.[5] The response was recorded on a 1-4 scale defined as not at all important (1), slightly important (2), fairly important (3) and important (4). Our dependent variable, *Competition*, is a dummy variable equal to 1

[3] The survey and methodology for data collection are available at www.enterprisesurveys.org.
[4] For the regressions reported below, we leave out the city of Kozhikode from the sample (1.8% of the sample) because it is a gross outlier. Including the city in the sample does not have much impact on the sign or the significance level of estimated coefficients of our main variables although their magnitude is lower. Amin (2007) uses the same data to estimate the effect of computer usage and labor regulation on employment. This study also finds Kozhikode to be a gross outlier.
[5] There is no competition from foreign retailers in India.

if a store reported facing significant competition (fairly important or important) and 0 otherwise. The mean value of the dependent variable is .38 and the standard deviation equals .48.

We note that our dependent variable is based on the perception of stores (about competition) and it is not an objective measure (number of competitors) usually used in the literature. There is some concern in the literature that such perception based measures may be too colored by firm (store) characteristics to reflect objective reality.[6] Hence, an additional contribution of the paper lies in showing that a substantial variation in the subjective measure is due to objective factors such as non-working adults, retailer density and physical infrastructure. Another concern with the dependent variable could be that it relates to price-competition alone and therefore may fail to capture the broader competitive environment. 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 by providing a greater range of products and better quality of service. While this problem cannot be ruled out completely, we provide some evidence which suggests that it is unlikely to be severe. Specifically, in another survey question, stores were asked how important the influence of domestic competitors is in their decision to introduce new product lines. Responses were recorded on the same 1-4 scale as above. The correlation between the response of stores on this question and the one on price-competition above equals .811. The high correlation is reassuring in that it suggests that our measure of competition captures the broader competitive environment rather than the narrow specifics of price setting.[7]

[6] For more details on this point, see for example, Pierre and Scarpetta (2006).
[7] Our main results are roughly similar irrespective of which of the two competition measures is used.

2.2 *Explanatory variables*

Our main explanatory variable is the number of adult non-workers per household in the city. The variable equals total number of adult non-workers in the city divided by the total number of households in the city (*Non workers*). We use (lagged) 1991 values of the variable taken from Census of India (1991).[8] Shopping for many products such as grocery items is typically done for the entire household which motivates our preference for non-workers per household over, for example, non-workers as a proportion of city population.[9] We follow the Census definition of adults and household. The former include all agents above seven years of age and the latter is defined as the set of individuals living in a common house and sharing a common kitchen.[10] For the cities in our sample, *Non workers* varies between 2 and 3.9 with a mean value of 2.9 and a standard deviation of 0.38.

Reverse causality from the level of competition to non-workers is unlikely given that the latter is lagged by 15 years. A relatively more serious problem with our results could be a bias due to omitted variables.

It is natural to expect some correlation between non-workers and income levels of households and overall development of cities. This is both, an advantage and a

[8] Data on non-workers is available every ten years with the latest year being 2001. Our results do not change much if we use 2001 values of the variable. The correlation coefficient between 1991 and 2001 values of non-workers per household is .875.

[9] For the cities in our sample, the correlation coefficient between *Non workers* as defined above and non-workers as a proportion of total city population equals .768. The high correlation between the two variables makes it difficult to infer whether the distribution of non-workers across households within cities (as opposed to the distribution of non-workers across cities) matters for the level of competition. However, in some specifications where our results are particularly strong such as the sample of traditional stores (selling grocery items), the effect of Non workers on competition remains significant even if we control for the proportion of non-workers in total city population.

[10] Data on non-workers by other age groups are not available.

disadvantage for our estimation. The advantage is that higher values of non-workers are likely to be associated less development such as poorer quality of power supply or frequent power outages, less access to finance, more burdensome regulation etc. (see, Table 3), which are likely to have a direct negative effect on competition (confirmed below). Since we predict a positive effect of non-workers on competition, the omitted variable bias here should lower the estimated coefficient of non-workers towards zero (downward bias). We provide one example to illustrate the point further. In the survey, stores reported irregular power supply as the single most important bottleneck for their business. The correlation between hours of power outage per day reported by stores (averaged at the city level) and non-workers is positive and equal to .132. Our estimation results clearly show that higher values of power outage have a statistically significant negative effect on the level of competition and failure to control for power outages lowers the estimated coefficient of non-workers towards zero, although by a small amount.

The disadvantage stems from a possible negative correlation between household income levels and competition (discussed above). Higher values of non-workers are likely to be associated with lower income levels. If lower income has a positive effect on competition, as suggested in the literature, then failure to properly control for household income could bias the estimated coefficient of non-workers upwards.

Another possibility is that firm characteristics may be systematically correlated with non-workers. For example, we find that store-size measured by the floor area of the shop is smaller (on an average) in cities with more non-workers although the correlation between the two is small (-.027, Table 3). Thus, failure to control for size could bias our results for non-workers upwards (downwards) if size and competition are inversely

(positively) correlated. Much like store-size, most store characteristics reported in the survey are only weakly correlated with non-workers (Table 3) which suggests that the omitted variable bias problem here is unlikely to be severe. Our empirical results confirm this in that the estimated coefficient of non-workers does not change much in magnitude due to controls for various store characteristics.

We address the potential problems discussed above by directly controlling for a number of city and store characteristics and by contrasting the effect on competition of non-workers and (1991 values of) the number of children per household averaged at the city level (*Children*). Formally, *Children* equals total number of individuals below 7 years of age in the city divided by total number of households in the city. Data source for the variable is Census of India (1991).

The motivation for comparing the effects of non-workers and children on competition has been discussed above. Our claim there was that non-workers and children are likely to be correlated with income and overall development of cities in the same direction. Table 3 reports on various correlations to support this claim. We note that the correlation between non-workers and children is high (.488) but not too high to preclude the estimation of their independent effects on competition. Further, we show that our result for the effect of non-workers on competition holds with and without controlling for children.

Other than *Children*, controls in our main specification include a measure of household income which we proxy by the mean per capita expenditure of the district population (*Expenditure*) and two proxy measures of overall development of the cities which are *Population* and *Metro*. Data on GDP or income levels are not available at the

city level for India. However, the National Sample Survey Organization (NSSO) routinely conducts representative surveys of household expenditures which are the closest proxy measure to household income levels. These data are available at the district level which is bigger than the city.[11] Our *Expenditure* variable is taken from the 50th (1991-92) NSSO round.[12] *Population* equals the total adult city population in 1991 and *Metro* is a dummy variable equal to 1 if a store is located in the metropolitan cities of Delhi, Mumbai, Chennai, Hyderabad, Kolkatta and Bangalore and 0 otherwise. Larger cities (in terms of population) and the metropolitan cities are known to be more developed and richer and also the main beneficiaries of the ongoing retail boom.

We also experimented with the number of workers per household (*Workers*) and literacy rates as alternative proxy measures of city level incomes. However, both these variables are very poorly correlated with non-workers (Table 3), they showed no significant effect on competition, and controlling for them made no difference to our main results (discussed in the section on robustness).

In the robustness section, we show that our main results are robust to a number of additional controls which include literacy rate, sex ratio, number of workers per household, store-size, duration of power outage faced by stores, number of retailers in the city, etc.

The correlation coefficients between our main variables are as follows: *Non workers* and *Expenditure* (-.290), *Non workers* and *Population* (-.096), *Children* and

[11] The NSSO reports data separately for the rural and urban population of the district. The urban part of the district is perhaps a better approximation of the cities in our sample. Our results do not change much if we use expenditure values for the urban or total district population which are correlated upto .916.

[12] We use NSSO data from the 50th round because it relates to the year 1991-92, closest to the year for our main variable, *Non workers*. Similarly, for *Population*, we 1991 values using Census data. We also experimented with 2001 values of *Expenditure* (NSSO Round 55) and *Population* but this did not change our results much. The correlation between 1991 and 2001 values of *Population* equals .901 and .808 for *Expenditure*.

Expenditure (-.126), *Children* and *Population* (-.160), *Population* and *Expenditure* (.548) and *Non workers* and *Children* (.488). Although *Expenditure* shows a high correlation with *Non workers* (and *Children*), it is not too high to cause any significant multicollinearity problem. Further, our results show that, irrespective of the set of controls, *Expenditure* has no significant effect on competition and it has virtually no effect on the estimated coefficients of *Non workers* and *Children*. The correlation between non-workers and children has already been discussed above.

3. Estimation

We use a logit specification with all standard errors clustered on the city.[13] Estimated marginal effects (at the mean value of the explanatory variables) are reported in Table 4. In all the specifications, the estimated effect of *Non workers* on *Competition* is positive and significant at less than 5% level. Without any additional controls, a unit increase in *Non workers* raises the probability of a store facing significant competition by 17 percentage points with a p-value of .014 (column 1, Table 4). The estimate implies that moving from the lowest (city of Noida) to the 25th percentile value (city of Bhubaneswar) of *Non workers* raises the probability of a store facing significant competition by 9.8 percentage points, a large effect given that only 38.2 % of the stores in the sample face significant competition. Alternatively, a one standard deviation increase in *Non workers* raises the stated probability by 6.4 percentage points. As we show below, these magnitudes are on the conservative side.

[13] Our main results are slightly stronger if we do not adjust standard errors for clustering.

In columns 2 and 3 of Table 4 we report the independent effects of children and expenditure on competition (without any additional controls). Both these effects are negative but neither of them is significant at 10% or less. The estimated coefficient of *Expenditure* is particularly small in magnitude. A move from the lowest (city of Patna) to the median (city of Jaipur) value of *Expenditure* lowers the probability of a store facing significant competition by only 2.24 percentage points (discussed in more detail below). The corresponding change for children here equals 5.6 percentage points.

In the introduction we argued that if elements of search cost are indeed driving the relationship between competition and non-workers then controlling for children should increase the estimated coefficient of non-workers. We had also argued that if our results suffer from spurious correlation, then the opposite should hold. Of course, one could extend a similar argument for the estimated coefficient of children as well.[14]

Regression results in column 4 of Table 4 show that simultaneously controlling for non-workers and children causes the estimated coefficient of non-workers to rise sharply from .169 above to .274 (significant at less than 1% level) and the same for children increases (in absolute value) from -.204 to -.517 (p-value of .011). Thus, our results clearly pass the informal test stated above.

Based on the results from the previous estimation, our main variables show large effects on competition. A move from the lowest to the 25th percentile value of non-workers raises the probability of a store facing significant competition by 15.9 percentage points. The same for children is a decrease by 7.5 percentage points.

[14] That is, children and non-workers have common covariates such as income levels, power outage, etc (table 3). Hence, if the relationship between children and competition is spuriously driven by these common covariates, then controlling for non-workers should lower the estimated effect of children on competition towards zero.

Controlling for expenditure, population and metropolitan city fixed effect (column 5, Table 4) does not change our results much from above although the estimated coefficient of non-workers rises further from .274 to .330. For the remaining variables, we find no significant effect of either *Expenditure* or *Population* on competition. [15] However, the metro fixed effect is significant at close to 5% level (p-value of .051). Relative to stores in other cities, the probability that a store located in a metropolitan city faces significant competition is lower by a large 17.5 percentage points. This could be due to differences in income levels or the structure of the retail markets between metropolitan and non-metropolitan cities.

One concern with the results so far could be that *Expenditure* does not show any significant effect on competition and this holds even if we do not control for population and the metro fixed effect. If this is because the variable does not properly capture household income levels, then our results for the effect of non-workers on competition could be biased upwards (as discussed in the introduction). We discuss this issue in detail in section 5 and show exactly why the expenditure-competition relationship estimated above is weak.

4. Robustness

Results from the robustness checks are reported in Table 5. We begin by first controlling for adult literacy rate (*Literacy*) and the ratio of females to males (*Sex ratio*). Both the variables are defined at the city level and we use their 1991 values taken from Census of

[15] We note that the estimated coefficients of *Expenditure* and *Population* are insignificant at 10% or less even if we do not control for the metropolitan city fixed effect in column 5 of Table 4.

India (1991).[16] These controls further guard against the possibility of our main variables spuriously picking up the effect of overall development and income on competition. Literacy rate is a useful proxy for wage rate and income levels. Some care in required in interpreting what sex ratio captures. In addition to higher income levels and greater gender parity (favoring women), higher values of sex ratio may be associated with less involvement of women in traditional roles including household shopping. If this is indeed true, then we might expect a negative relationship between sex ratio and competition. While our empirical results confirm such a negative relationship, we treat our suggested interpretation with extreme caution fully realizing that more work is required to ascertain its validity.

Regression results controlling for literacy and sex ratio are reported in column 1 of Table 5. Sex ratio shows a strong negative effect on competition which is significant at less than 5% level. Like for expenditure and population, we do not find any significant effect of literacy on competition. Lastly, the estimated coefficients of non-workers and children maintain their signs and significance levels with little change in their values.

As mentioned above, stores reported irregular power supply as the single most important obstacle to doing business followed by access to finance. We control for two measures of power supply using data from the Enterprise survey. These are the duration of power outage reported by a store on an average per day during 2005-06 (*Outage – own store*) and the average duration of power outage reported by all *other* stores in the city (*Outage*). The motivation here is that power outage in one's own store is likely to deflect customers to the neighboring stores implying greater competitive pressure for the store in

[16] We use 1991 values because data for non-workers and children are also for the same year. However, 1991 and 2001 values of these variables are almost perfectly correlated (correlation of over .90) and our results do not change much if we use 2001 values instead.

question. By the same logic, the opposite holds for power outages in the other (neighboring) stores. For financial development of the cities, we use two separate measures constructed from the survey. These are the proportion of stores in the city that have a checking or savings account (*Checking*) and the proportion of stores in the city that have either a line of credit or overdraft facility (*Financial Access*). The correlation between the two finance variables equals .435.[17]

Regression results reported in column 2 of Table 5 show that both the two power outage variables have significant effects on competition (at less than 1% level) and these effects are in opposite directions as predicted. However, we do not find any significant effect on competition of the two finance related variables either individually or jointly. Results for the remaining variables do not change much from above although we do find a slight increase in the estimated coefficient of non-workers from .312 to .327.

Next, we introduce measures to capture store-size, number of retailers in the city and the extent of regulatory burden which could affect industry structure and thereby the level of competition. We expect that the level of competition should be higher in cities with higher density of retail shops or retailers. Excessive regulatory burden could reduce the threat of new entry lowering market competition (less contestable markets). We do not have any apriori prediction for the direction of the relationship between store-size and competition and we treat this as an empirical question.

[17] The survey also provides information on whether a store has a checking/savings account and a line of credit. Our results do not change much if we control for these measures also (discussed towards the end of the section).

For store size, we use (log of) floor area of the shop as reported in the survey (*Size*).[18] Since direct measures of entry barriers are not available, we follow Sevensson and Fisman (2007) in using the percentage of stores' senior management's time (averaged at the city level) that is spent in dealing with business regulations as reported in the survey (*Regulation*). Lastly, for retailer density, we use total employment in retail and distribution sector as a proportion of total adult population at the city level (*Retailer density*). Data for retailer density is from Census of India (1991). We are not aware of any data on the number of retail shops in India.

Column 3 of Table 5 reports regression results with the additional controls mentioned above. We find that retailer density has a positive effect on competition significant at less 5% level (p-value of .042). The estimated effect is quite large in magnitude. An increase in retailer density from its 25^{th} to the 75^{th} percentile value increases the probability of a store facing significant competition by 10.4 percentage points. As expected, more burdensome regulation has a negative effect on the level of competition, significant at close to 10% level (p-value of .104). Lastly, size has a positive effect on competition significant at less than 5% level. Below we show that the significant size-competition relationship is probably a result of different products sold by stores of varying size. Lastly, there is not much change in the estimated coefficients of the remaining variables from above.

In column 4 of Table 5 we control for some additional store characteristics. The motivation here is that if store characteristics vary systematically with non-workers and they have a direct effect on competition also then our results above could still suffer from

[18] Our results do not change much if we use labor employment or annual store sales as measures of store-size.

an omitted variable bias problem. Since theory provides little guidance on which store characteristics may be important here, we control for basic store features which include age of the store (*Age*), a dummy variable equal to 1 if a store is part of a larger chain and 0 otherwise (*Chain*), percentage of the firm (store) held by the largest shareholder (*Ownership concentration*) and store-type fixed effects.[19] The survey classifies all stores into traditional stores (selling grocery items), consumer durable stores (selling consumer durables) and modern format stores (large stores part of a shopping complex). Store-type fixed effects are dummy variables that capture factors common to all stores within each of these categories (a formal definition of these fixed effects is provided in Table 1).

Regression results (column 4 of Table 5) show that these additional controls do not change any of the results discussed above except for the *Size*. The estimated coefficient of non-workers remains positive and significant at less than 1% level while children continues to show a strong negative effect on competition with no significant change in their magnitudes from above. Age has no significant effect on competition while being part of a larger chain and more concentrated ownership show negative effects on competition (significant at less than 10% level). For store-type fixed effects, we find that, relative to consumer durable stores, the probability that a traditional store faces significant competition is lower by 14.9 percentage points and this difference is significant at less than 1% level (not reported). We do not find any significant difference between consumer durable and modern format stores in the level of competition they face. We note that the estimated coefficient of *Size* is now insignificant due to the controls for store-type fixed effects suggesting that size was earlier picking up store-type

[19] Our motivation in controlling for ownership concentration is that it could affect store efficiency which may in turn determine the competitive pressure faced by a store. Excluding the variable form the regressions does not change any of our results significantly.

specific factors.[20] We suspect that these store-specific factors could be different product lines carried by these stores (groceries vs. consumer durables) although we cannot ascertain this due to data limitations.[21]

We performed a number of additional robustness checks and found that our main results discussed above remained intact. We provide a brief outline of these checks.

About 30% of the stores in our sample are located in the metropolitan cities. Since the structure of retailing could be different across metropolitan and the remaining cities, we dropped all the metropolitan cities from the sample to check if our results hold for the majority of cities in the country. Regression results for the sub-sample of non-metropolitan cities are reported in column 5 of Table 5 and these are roughly similar to the ones discussed above.

Next, we controlled for additional store characteristics which include years of store manager's experience in retailing, initial level of employment when the store first started operations, number of day's of inventory maintained by the store, two dummy variables indicating whether or not has a line of credit and a checking/savings account, percentage of a store's annual sales (in 2005-06) that were never paid for, losses due to theft (as percentage of sales in 2005-06) averaged at the city level, percentage of the firm (store) held by the largest shareholder, a measure of crime at the city level derived from store's perception of crime as an obstacle to their business, the number of workers per household (1991 values, at the city level) and measures of vehicle availability which include the proportion of households that have a four-wheeler (car, jeep, van, etc) and

[20] Size varies sharply across store-types. The average value of *Size* for traditional stores equals 4.78 (traditional stores), 5.34 (consumer durable stores) and 7.03 (modern format stores).
[21] Data on the specific products carried by stores are not available.

proportion of households that have a two-wheeler (scooter, motor cycle, moped, etc).[22] There was virtually no effect of these additional controls on our main results discussed above with the estimated coefficients of non-workers and children remaining almost unchanged from above.

5. Household Income-Competition Relationship

In this section we focus on why our results fail to show any significant effect of expenditure on competition. This is important to address any concerns that the variable may be a poor proxy for household income in which case the estimated effect of non-workers on competition discussed above could still be biased upwards.

There are three reasons for why expenditure does not show much effect on competition. First, population absorbs some of the effect of expenditure on competition.[23] However, this is at best a contributory factor because even without controlling for population, expenditure has at best a weak effect on competition in all the specifications discussed above. Second, higher values of expenditure are expected to have a negative effect on competition. However, these higher values are likely to pick the positive effect on competition of various covariates such as fewer children, less burdensome regulation, fewer power outages, etc.[24] We show that this is the main reason why expenditure shows a weak effect on competition without any controls (as in column 3 of Table 4). The natural question to ask here is that if this is indeed true then controlling for these covariates (as we did in Table 5) should have resulted in a significant effect of

[22] Data on two-wheelers and four-wheelers was first collected in the 2001 Census and it is available at the district level. We use these data for the urban part of the district population.
[23] The correlation between *Expenditure* and *Population* is .548 for the cities in our sample.
[24] For the cities in our sample, the correlation between *Expenditure* and *Children*, *Outage* and *Regulation* equals, respectively, -.290, -.065 and -.143.

expenditure on competition. This brings us to our third point. That is, non-workers and expenditure are inversely correlated (correlation of -.290) and the prediction is that they have opposite effects on competition. This implies that failure to control for non-workers (as in the literature) could give rise to a spurious relationship between expenditure and competition. Our specification controls for non-workers and therefore wipes out this spurious relationship. Below we show that is indeed the case.

To address the first of the three points in the previous paragraph, we drop population from all the specifications discussed below. To illustrate the second point, we report regression results controlling for the various covariates except non-workers as discussed. Regression results reported in columns 1-3 of Table 6 clearly show that expenditure has a statistically significant negative effect (at close to 5% level) on competition with just a few basic controls (column 1). The strong negative effect here survives even when we control for all the variables (except non-workers) discussed above (columns 2 and 3). For the specifications in columns 1-3 of Table 6, we now add non-workers. Regression results reported in columns 4-6 of Table 6 clearly show that controlling for non-workers leads to a sharp drop in the estimated coefficient of expenditure and it is no longer significant at 10% level or less. We note that these results hold whether we control for the metro city fixed effect or not which could be a close substitute of *Expenditure*.

Our results in this section suggest that the negative income-competition relationship found in the literature could be due to the failure to control for non-workers. They also raise our confidence in the *Expenditure* variable as far as measurement errors with the variable are concerned.

6. Conclusion

The structure of Indian households is witnessing a significant change with the emergence of smaller families and greater participation in the labor market. One consequence of these changes is fewer non-workers per household. So far, these changes have been of interest only to demographers and sociologists with mainstream economic policy being largely immune to them. The present paper shows that these changes are likely to have a significant effect on the level of competition in the retail sector of the country and these changes are independent of potentially correlated changes in the household income levels.

Our results clearly cast doubt on some of the findings on the negative income-competition relationship based on the idea that lower income households search more intensively because of their higher marginal utility of income. A more plausible explanation (confirmed in our data) is that lower income households have more non-workers and therefore lower shopping time opportunity cost. Hence, they tend to search more intensively causing the level of competition to rise.

We have also shown that number of children per household and gender specific factors are also important drivers of competition in the sector. Due to data limitations we were unable to explore in more the gender-competition relationship but this could be a fruitful area for future work. Lastly, based on our findings, tying of competition policy to these socio-demographic-economic factors is recommended to ensure that the retail sector remains truly competitive.

References

[1] Amin, M. (2007), "Computer Usage and Labor Regulation," mimeograph, http://www.enterprisesurveys.org.

[2] Ausubel, Lawrence M. (1991), "The Failure of Competition in Credit Market," *American Economic Review*, 81(1), pp. 50-81.

[3] Baron, Jon M., Beck A. Taylor and J. R. Umbeck (2004), "Number of Sellers, Average Prices and Price Dispersion," *International Journal of Industrial Organization*, 22, pp. 1041-1066.

[4] 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.

[5] Claycombe, Richard J. (2000), "The Effects of Market Structure on Prices of Clothing and Household Furnishings," *International Journal of Industrial Organization*, 28, pp. 827-841.

[6] Department of Trade and Industry (2000), **Switching Suppliers**, London: The Stationery Office.

[7] Giulietti, Monica, Catherine W. Price and Michael Waterson (2005), "Consumer Choice and Competition Policy: A Study of UK Entergy Markets," *Economic Journal*, 115, pp. 949-68.

[8] 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. 281-95.

[9] Hoch, S., B. Kim, A. Montgomery and P. Rossi (1995), "Determinants of Store-Level Price Elasticity," *Journal of Marketing Research*, 32(1), pp. 17-29.

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

[11] Lalive, Rafael and Armin Schmutzler (2007), "Exploring the Effects of Competition for Railway Markets," *International Journal of Industrial Organization*, forthcoming.

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

[13] Pierre, Gaelle and Stefano Scarpetta (2006), "Employment Protection: Do Firm's Perceptions Match with Legislation?" *Economics Letters*, 90, pp. 328-334.

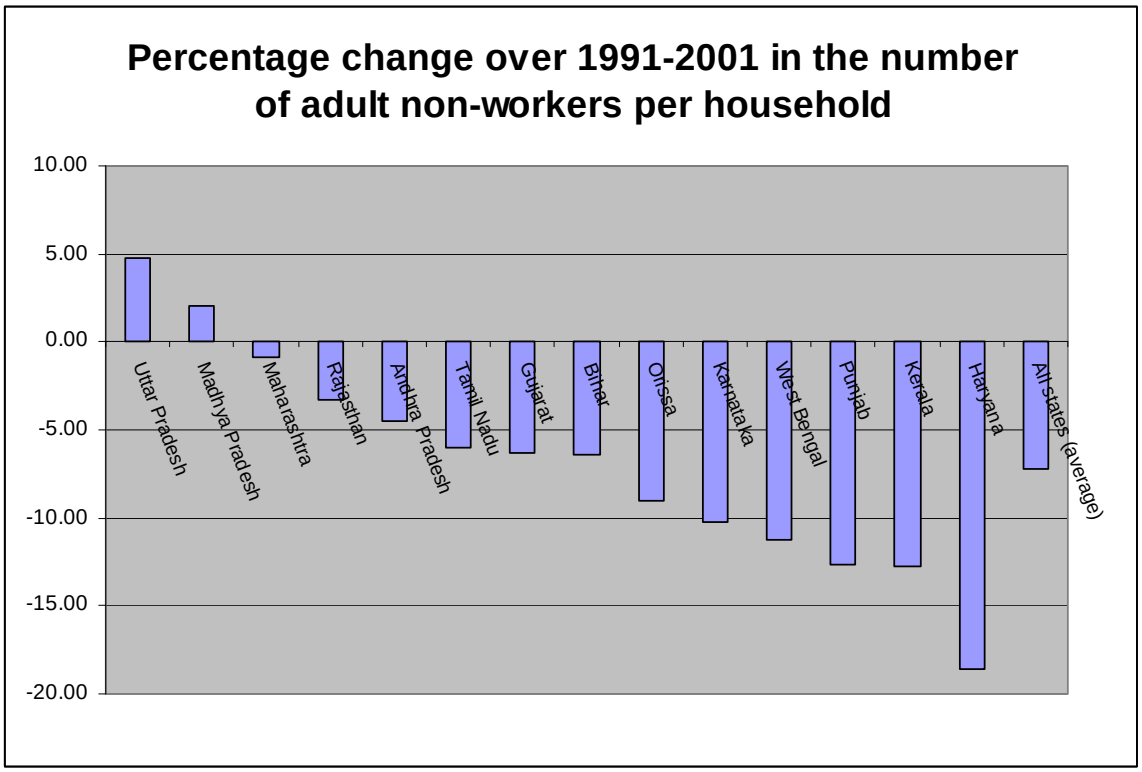
[14] Prendergast, C. (2002), "Consumers and Agency Problems," *Economic Journal*, 112 (478), pp. C34-51.

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

[16] Svensson, Jakob and Ray Fisman (2007), "Are Corruption and Taxation Really Harmful to Growth? Firm Level Evidence," *Journal of Development Economics*, 83 (1), pp. 63-75.

[17] Waterson, M. (2003), "The Role of Consumers in Competition and Competition Policy," *International Journal of Industrial Organization*, 21(2), pp. 129-50.

Figure 1



Source: Census of India, 1991 and 2001. Percentage changes in the figure above equal the number of adult non-workers per household in 2001 minus the same in 1991 and expressed as a percentage of the 1991 values of the variable. These values are shown in the table below.

	Adult Non-workers per household		
State	1991 values	2001 values	% change
Uttar Pradesh	3.36	3.52	4.79
Madhya Pradesh	2.45	2.50	2.03
Maharashtra	2.39	2.37	-0.89
Rajasthan	2.84	2.74	-3.29
Andhra Pradesh	2.14	2.04	-4.52
Tamil Nadu	2.14	2.01	-6.05
Gujarat	2.67	2.50	-6.35
Bihar	3.32	3.11	-6.40
Orissa	2.70	2.46	-9.05
Karnataka	2.62	2.35	-10.23
West Bengal	3.06	2.72	-11.19
Punjab	3.46	3.02	-12.61
Kerala	3.21	2.80	-12.72
Haryana	3.48	2.83	-18.59
All states (average)	2.84	2.64	-7.18

Source: Census of India, 1991 and 2001.

Table 1: Description of Main Variables

Variable	Description
"Last fiscal year" below means fiscal year 2005-06.	
Competition	A dummy variable equal to 1 if a store reported "Fairly important" or "Important" and 0 otherwise 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)
Non-workers	Total adult non-workers in the city divided by total number of households in the city (1991 values). Source: Census of India (1991).
Children	Total number of children (below 7 years) divided by the total number of households in the city in 1991. Source: Census of India (1991)
Expenditure (in thousand Indian Rupees)	Per capita expenditure in the urban part of the district where the cities in our sample are located. Source: National Sample Survey Organization (50th Round, 1991-92), Government of India.
Population (in millions)	Total (adult) population of the city in 1991. Source: Census of India (1991).
Metro	A dummy variable equal to 1 for a store located in a metropolitan city and 0 otherwise. The metropolitan cities are Delhi, Mumbai, Kolkatta, Chennai, Bangalore and Hyderabad. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)

Table 1: Description of Other Variables

Variable	Description
Literacy	Number of adult literates in the city per 1000 adult population of the city (1991 values). Source: Census of India (1991).
Retailer density	Total employment in retail and distribution in the city divided by adult city population (1991 values). Source: Census of India (1991).
Sex ratio	Ratio of females to males in the city in 1991. Source: Census of India (1991)
Size	Total selling area of the store measured in square feet (log values). Source: World Bank Enterprise Surveys

	(www.enterprisesurveys.org)
Traditional store	A dummy variable equal to 1 if a store is a "traditional store" and 0 otherwise. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Modern Format store	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)
Store-type fixed effects	Two dummy variables indicating whether a store is a traditional store or a modern format store (as defined above). The omitted category is that of consumer durable stores. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Outage-own store	Hours of power failure faced by the store on a typical day during the last fiscal year. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Outage	Hours of power outage on a typical day faced on an average by all other stores in the city during the last fiscal year. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Financial Access	Proportion of stores in the city that have a line of credit or overdraft facility. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Checking	Proportion of stores in the city that have a checking or savings account. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Regulation	Percentage of store's senior management's time spent in dealing with business regulations during the last fiscal year (average values at the city level). Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Chain	A dummy variable equal to 1 if a store is part of a larger chain and 0 otherwise. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Ownership concentration	Percentage of the firm (store) held by the largest shareholder. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Age	Age of the store equal to 2006 minus the year store was established. Source: World Bank Enterprise Surveys (www.enterprisesurveys.org)
Workers	Total number of workers in the city divided by the total number of households in the city (1991 values). Source: Census of India (1991).

Table 2: Summary statistics of the main variables

Variable	Mean	Standard deviation
Competition	.382	.486
Non workers	2.85	.381
Children	.766	.144
Expenditure (in thousand Indian Rupees)	.423	.115
Population (in millions)	1.31	1.68
Retailer density	.082	.017
Size (square feet, log values)	5.13	1.17
Sex ratio	.877	.065

Table 3: Correlation coefficients for *Non workers* and *Children* with various indicators of overall development of the cities

	Non workers	*Children*
Non workers	1	.488
Children	.488	1
Expenditure (per capita expenditure in 1991- district level values)	-.290	-.126
Population (adult city population in millions in 1991)	-.096	-.160
Retailer density (total employment in the retail sector in the city as a proportion of adult population in the city, 1991 values)	-.020	-.257
Sex ratio (ratio of females to males in the city, 1991 values)	-.036	-.321
Size (Log of floor area of the retail stores in the sample measured in square feet– city level average)	-.027	-.139
Power outage (hours of power outage in a typical day faced by stores during 2005-06)	.132	.368
Financial Access (proportion of stores in the city that have a line or credit or overdraft facility)	-.047	-.135
Checking (proportion of stores in the city that have a checking or savings account)	-.007	-.138
Regulation (amount of time spent by senior management of stores in dealing with business regulation during 2005-06 and averaged at the city level)	-.037	-.151
Literacy rate (proportion of adults in the city that are literate, 1991 values taken from Census of India.)	-.013	-.496
Workers (total number of workers divided by total number of households in the city, 1991 values taken from Census of India.)	.028	.026

1. All variables are defined at the city level. Detailed description of the variables along with data sources is provided in Table 1.

2. Main point of the table is to show that *Non workers* and *Children* are correlated in the same direction with various proxy measures of overall development of the cities.

Table 4: Marginal effects from Logit Regressions					
Dependent variable: *Competition*					
	(1)	(2)	(3)	(4)	(5)
Non workers	.169**			.274***	.330***
	(.014)			(.000)	(.000)
Children		-.204		-.517**	-.717***
		(.372)		(.011)	(.002)
Expenditure			-.249		.375
			(.233)		(.248)
Population					-.002
					(.919)
Metro					-.175*
					(.051)
Predicted probability	.380	.381	.381	.378	.375
Sample Size	1866	1866	1866	1866	1866

p-values in brackets; all standard errors clustered on the city; significance level is denoted by *** (1% or less), ** (5% or less) and * (10% or less). All standard errors are clustered on the city.

Table 5: Marginal effects from logit regressions: Robustness

Dependent variable: *Competition*					
	(1)	(2)	(3)	(4)	(5)
Non workers	.312***	.327***	.308***	.305***	.345***
	(.000)	(.000)	(.000)	(.000)	(.000)
Children	-.866***	-.753***	-.634***	-.666***	-.879***
	(.000)	(.000)	(.003)	(.001)	(.003)
Expenditure	.321	.314	.167	.161	-.082
	(.291)	(.344)	(.563)	(.554)	(.817)
Population	-.008	-.013	-.022	-.020	-.088*
	(.645)	(.516)	(.159)	(.198)	(.080)
Metro	-.178**	-.187**	-.218***	-.217***	
	(.046)	(.024)	(.006)	(.006)	
Literacy	-4.47	-5.43	-5.45	-5.81	-7.24
	(.302)	(.213)	(.256)	(.208)	(.219)
Sex ratio	-.880**	-1.22***	-1.24***	-1.24***	-1.27***
	(.020)	(.000)	(.000)	(.000)	(.005)
Outage		-.042***	-.042***	-.042***	-.048***
		(.000)	(.000)	(.000)	(.001)
Outage–own store		.015***	.014***	.013**	.018***
		(.004)	(.007)	(.011)	(.003)
Financial Access		.140	.071	.077	.237
		(.591)	(.783)	(.762)	(.430)
Checking		-.059	.093	.104	.094
		(.822)	(.620)	(.570)	(.669)
Retailer density			4.76**	4.62**	4.64**
			(.042)	(.045)	(.031)
Regulation			-.011	-.011*	-.015**
			(.104)	(.096)	(.026)
Size			.031**	.015	.007
			(.017)	(.334)	(.730)
Store type fixed effects				Yes	Yes
Age				.001	.0002
				(.606)	(.905)
Chain				-.093*	-.072
				(.088)	(.344)
Ownership concentration				-.002*	-.002
				(.078)	(.126)
Predicted probability	.374	.372	.368	.366	.397
Sample Size	1866	1859	1850	1850	1279

p-values in brackets; all standard errors are clustered on the city; significance level is denoted by ***(1% or less), ** (5% or less) and * (10% or less).The sample in column 5 excludes the metropolitan cities of Chennai, Delhi, Mumbai, Kolkatta, Bangalore and Hyderabad. Sample size in columns 1-4 varies due to missing obsevations.

Table 6: Marginal effects from logit regressions: Robustness

Dependent variable: *Competition*

	(1)	(2)	(3)	(4)	(5)	(6)
Non workers				.248***	.278***	.313***
				(.002)	(.000)	(.000)
Children	-.155	-.222	-.225	-.399***	-.521***	-.695***
	(.489)	(.232)	(.305)	(.040)	(.000)	(.000)
Expenditure	**-.405***	**-.453****	**-.426***	**-.168**	**-.006**	**.091**
	(.068)	**(.039)**	**(.052)**	**(.512)**	**(.982)**	**(.717)**
Population						
Metro		-.226***	-.240***		-.270***	-.275***
		(.001)	(.001)		(.000)	(.000)
Literacy			-.198			-6.29
			(.968)			(.192)
Sex ratio		-1.49***	-1.52***		-1.30***	-1.17***
		(.000)	(.000)		(.000)	(.001)
Outage	-.034**	-.043***	-.045***	-.034***	-.045***	-.043***
	(.013)	(.000)	(.001)	(.003)	(.000)	(.000)
Outage–own store	.014***	.014***	.013**	.014***	.015***	.013**
	(.006)	(.005)	(.012)	(.005)	(.004)	(.012)
Financial Access			-.187			.038
			(.484)			(.883)
Checking			.087			.070
			(.646)			(.689)
Retailer density		5.52**	5.87**		4.73**	4.47*
		(.042)	(.035)		(.039)	(.054)
Regulation	-.015**	-.007	-.006	-.012*	-.006	-.010
	(.022)	(.264)	(.404)	(.073)	(.198)	(.120)
Size			.007			.014
			(.659)			(.350)
Store type fixed effects			Yes			Yes
Age			.001			.001
			(.304)			(.611)
Chain			-.096			-.095*
			(.101)			(.080)
Ownership concentration			-.001			-.002*
			(.126)			(.099)
Predicted probability	.378	.374	.372	.376	.369	.366
Sample Size	1859	1859	1850	1859	1859	1850

1. p-values in brackets; all standard errors are clustered on the city; significance level is denoted by ***(1% or less), ** (5% or less) and * (10% or less). Sample size varies due to missing observations.

2. Main point of the table is to show that the significant effect of *Expenditure* (columns 1-3) disappears when we control for *Non workers* (columns 4-6).