Peer Effects in Consumption in India: An Instrumental Variables Approach Using Negative Idiosyncratic Shocks

Punarjit Roychowdhury

Indian Institute of Management Indore,

Economics Area,

Prabandh Shikhar, Rau-Pithampur Road,

Indore, Madhya Pradesh 453556, India.

email: punarjitroychowdhury@gmail.com

Abstract

This paper examines peer effects in consumption in context of a less developed country. Specifically, the question that I seek to answer is whether consumption expenditure of a household is influenced by that of its peers in a less developed country. To examine this question, I use newly available household level data from India. I define a household's peer group as other households living in its village/neighborhood. In assessing the influences of peers in this context, there are two key empirical challenges including shared group-level unobservables, and simultaneity of peer influences. I address these issues by using an instrumental variables/fixed effects approach that compares households in the same district but different villages/neighborhoods who are thus exposed to different sets of peers. In particular, I use plausibly exogenous variation in idiosyncratic expenditure shocks – which are accidental and negative in nature – faced by peers as instruments for peers' consumption expenditure. Preferred specification suggests that a one standard deviation increase in average consumption expenditure of a household's peers causes the household's own consumption expenditure to increase by 0.42 standard deviations. Falsification tests and robustness checks support the validity of my results. My findings suggest that policies that influence a household's consumption will also affect the consumption of the household's peers through social interactions. This implies traditional analyses of consumption intervention programs that do not take into account such spillover effects will understate the total social impact of the programs, and hence lead to inaccurate evaluation of cost-effectiveness of such programs.

Peer Effects in Consumption: Evidence from India

Abstract

This paper examines peer effects in consumption in context of a less developed country. Specifically, the question that I seek to answer is whether consumption expenditure of a household is influenced by that of its peers in a less developed country. To examine this question, I use newly available household level data from India. I define a household's peer group as other households living in its village/neighborhood. In assessing the influences of peers in this context, there are two key empirical challenges including shared group-level unobservables, and simultaneity of peer influences. I address these issues by using an instrumental variables/fixed effects approach that compares households in the same district but different villages/neighborhoods who are thus exposed to different sets of peers. In particular, I use plausibly exogenous variation in idiosyncratic expenditure shocks (which are accidental and negative in nature) faced by peers as instruments for peers' consumption expenditure. Preferred specification suggests that a 1 Indian Rupee increase in consumption expenditure of a household's peers causes the household's own consumption expenditure to increase by 0.7 Indian Rupee. Falsification tests and robustness checks support the validity of my results. My findings suggest that policies that influence a household's consumption will also affect the consumption of the household's peers through social interactions. This implies traditional analyses of consumption intervention programs that do not take into account such spillover effects will understate the total social impact of the programs, and hence lead to inaccurate evaluation of cost-effectiveness of such programs.

1 Introduction

Consumption of households is traditionally modeled as being a function of their own income and preferences. However, it is often thought that a household's peers' consumption expenditure plays a major role in determining its own consumption expenditure. As famously noted by Duesenberry (1949), the strength of any household's desire to increase its consumption expenditure is a function of the ratio of its expenditure to some weighted average of the expenditure of others with whom it comes in contact. In this paper, I empirically study peer effects in household consumption specifically in context of a less developed country. More precisely, I attempt to address the following question: Does consumption expenditure of households' peers affect their own consumption expenditure in a less developed country? I examine this question using newly available household level data from India – which is home to one-third of the world's poor.[1]

There are at least two reasons why a careful analysis of peer effects in consumption in context of low income countries is crucial. First, understanding the magnitude and nature of peer effects is imperative for accurate evaluation of consumption intervention programs (e.g., consumption tax policies, conditional cash transfer programs, etc.) that are used massively in low income countries as developmental policies. This is because, if there are non-negligible peer effects, such policies, in addition to having a direct effect, would have an indirect spillover effect.[2] If this is not taken into account, the total 'social effect' of the policies would be underestimated.

Secondly, if consumption of households' peers affect their own consumption in a positive way, ceteris paribus, this would imply that the households must be lowering their savings or taking up loans to finance their increase in consumption when peer consumption rises. This is likely to magnify the risk of poor and middle income households (who make up the majority of the population in any low income country) of getting trapped in poverty (or severely hamper their ability to move out of poverty traps) and increase their economic vulnerability in the long run owing to 'under-saving' or 'over-borrowing' (Moav and Neeman, 2012). Thus finding evidence of positive peer effects would highlight the importance of formulating innovative social policies that represses social pressure and using such policies in conjunction with traditional anti-poverty and redistributive policies in order to augment social welfare.[3]

I begin by constructing a simple model of social interactions in consumption to guide my empirics. The model allows for both endogenous peer effects (peer effects due to peer consumption) as well as exogenous peer effects (peer effects due to peer characteristics). The model is constructed by adjusting a standard model of consumer choice with a social utility component that captures the satisfaction that a household

---

[1] http://time.com/2999550/india-home-to-most-poor-people/

[2] Also referred to as 'social multiplier effect' (Akerlof, 1997; Glaeser et al., 2003).

[3] As suggested by Putnam (2007), a policy that would repress social pressure might be promotion of a broad sense of 'we' among members of the same community or reference groups through popular culture, education and common experiences.

gets from 'status-seeking behavior'. In other words, it is the utility that a household obtains from getting ahead of others in its peer group in terms of consumption spending. My model predicts a positive monotonic relationship between own and average peer consumption expenditure. I empirically test this prediction using data from the 2012 Indian Human Development Survey (Desai et al., 2015). I define a household's peer group as other households living in its village (for rural areas) or neighborhood (for urban areas) since "people almost certainly compare themselves to their immediate geographical neighbors" (Deaton, 2001, p. 21). Fafchamps and Shilpi (2008), in fact, note that since social mobility is very low in low income countries as a result of which people live along people they grew up with, immediate neighbors constitute almost a 'natural' peer group for people in living in these countries.

As noted by Manski (1993), identification of endogenous peer effects is 'notoriously difficult'. More specifically, there are two econometric problems that hamper inferences about peer influences on household behavior. The first problem that arises is due to a simultaneity bias. This bias is generated by a 'reflection problem' – the simultaneous determination of own and peer outcome (which is consumption in the present case). The second factor complicating identification and estimation of peer influences is an omitted variables bias. In the absence of a perfect set of controls, one cannot validate a peer influence on consumption by observing that a group of neighbors spend similar amounts of resources on consumption. Inferences will be biased whenever there are group level unobservables that are correlated with consumption expenditure of all those belonging to the group (i.e., correlated unobservables).[4]

In the present paper, I tackle these problems based on a strategy of instrumental variables (IV)/fixed effects. I create my instruments based on assumptions invoked in my theoretical model. Specifically, my model assumes that a household faces various idiosyncratic shocks that affects its own consumption, and some of these shocks are *observable* (to the econometrician). I further assume that these shocks are 'household specific'. In other words, it is only the own idiosyncratic shocks that affect own consumption and peer idiosyncratic shocks do not have any influence on own consumption. As argued by Helmers and Patnam (2014, p. 95), "this is a credible assumption given the idiosyncratic nature of the shocks".

The above assumptions allow me to instrument average peer consumption (which is the source of the simultaneity bias) by average peer observable idiosyncratic shocks, and thankfully, the IHDS 2012 has data on some such shocks (e.g., death of a household member, accident/injury, job loss, incident of crime, etc). The intuition is that since own idiosyncratic shocks affect own consumption and do not contain any information about consumption of other households, to this extent, average peer idiosyncratic shocks should affect average

---

[4]Another source of correlated unobservables is non-random sorting of households into peer groups since this would imply that unobservable characteristics of households are correlated with the characteristics of the group. However, as I argue in section 4, this is unlikely to be a cause of concern in the present case since social mobility at the household level is very low in India.

peer consumption and that there should not be any effect of average peer idiosyncratic shocks on the target household's consumption after conditioning on own idiosyncratic shocks (to show that my instruments are plausibly exogenous, I carry out balancing tests (Bifulco et al., 2011; Lavy and Schlosser, 2011) and other standard IV diagnostic tests). I also include a full set of district fixed effects to control for group level unobservables that might be correlated with peer characteristics and/or peer idiosyncratic shocks.

My IV/fixed effects strategy is similar to the 'spatial IV' method used widely in empirical spatial literature (for an overview see Gibbons and Overman, 2012; Gibbons et al. 2015). The strategy requires some/all exogenous characteristics of neighboring spatial units (the spatial units being households in my case) to be used as instruments for spatial interaction term which is endogenous. The major requirements for this strategy, thus, is to carefully justify why such characteristics might potentially affect a spatial unit's own outcome but not the outcome of its neighbors. This approach has recently been used in some papers in the peer effects literature including Gaviria and Raphael (2001), Goux and Maurin (2007), Fletcher (2010, 2012, 2015), Helmers and Patnam (2014) and McVicar and Polanski (2014).

The IV/fixed effects-based identification strategy used in this paper not only allows one to clearly identify endogenous peer effects, but also has the advantage of being fairly flexible in terms of data requirement. Alternative methods, although novel and unique, are either unable to isolate endogenous peer effects from exogenous ones (e.g., Sacerdote, 2001; Graham, 2008; Ammermueller and Pischke, 2009), or do so at the cost of being extremely restrictive in terms of data requirement. For instance, the method proposed by Lee (2007) and developed later by Bramoulle et al. (2009) allows one to identify endogenous peer effects but requires peer groups to be 'small' on average and that there should be sufficient variation in peer group size. Again, using the empirical strategies proposed by Bramoulle et al. (2009), Calvo-Armengol et al. (2009) and Lee et al. (2010) to isolate endogenous peer effects from exogenous ones requires a researcher to be able to observe *all* social interaction links in the data (that is, the researcher must have very detailed network data). Such requirements are not met by most micro datasets, including the dataset used in the present paper.

My results are striking. In consonance with the prediction of the theory of status-seeking behavior, I find robust evidence that average peer consumption has a significant positive impact on households' own consumption. More specifically, I find that an increase in average peer consumption expenditure by 1 Indian Rupee causes an average household to increase its own spending by 0.7 Indian Rupee. This translates into a social multiplier of about 3. Further, I find that the increase in own consumption due to a one standard deviation increase in average peer consumption is exactly equal to the increase in own consumption due to a one standard deviation increase in a household's own income. Thus, the estimated endogenous peer effect is not only economically significant in absolute terms, but is also remarkably strong compared to other major determinants of consumption. Additionally, results of my subsample analysis suggest that the

3

point estimates of the endogenous peer effect is significantly higher for non-poor households compared to the households who fall below the poverty line.

Over the last few years, a sizable body of literature on peer influences in consumption has emerged. However, much of the existing research on this topic focuses on developed countries. One of the earliest empirical studies in this literature is that by Grinblatt et al. (2008). Using Finnish data, they provide evidence of endogenous peer effects in automobile purchases. Kuhn et al. (2011), using data from the Dutch Postcode Lottery (PCL), also find significant peer effects in car consumption. Moretti (2011), in an interesting paper, estimate peer effects and social multiplier in consumption of movies using high quality box-office data from the United States. De Giorgi et al. (2016) exploit detailed network data in a recent working paper to show the existence of non-negligible peer effects in aggregate consumption in context of Denmark.

The only two studies that I am aware of which focus on understanding neighborhood and peer effects in context of low income countries are those by Angelucci and De Giorgi (2009), and Angelucci et al. (2017).[5] Both these studies, using data from Progresa, which is a program that targets poor households in small rural communities in Mexico with bimonthly conditional grants to improve living standards, find strong evidence of positive program externalities on non-eligible households. They, however, estimate only a composite social interaction effect. Unlike these papers, in this study, I disentangle the endogenous peer effects from the exogenous ones. From a policy perspective, being able to isolate the endogenous peer effects the from exogenous effects is likely to be important "since endogenous effect is the source of a social multiplier" (Boucher et al., 2014, p. 91). This is the first contribution of this paper to the existing literature on peer effects in consumption.

The second contribution of my work has to do with the instruments constructed for identification of the endogenous peer effects. More specifically, the novelty of the dataset (i.e., the availability of information on household specific idiosyncratic expenditure shocks) used in this paper allows me to construct instruments for peer consumption which are credibly excluded from the regression equation in order to identify endogenous peer effects without having to assume zero exogenous peer effects.[6] This is a substantial improvement over many prior papers in the peer effects literature employing an IV strategy to estimate endogenous peer effects in absence of high dimensional social network data. Generally, these papers simply assume away all

---

[5] Of course, there is separate body of literature that looks at how relative consumption and income affects subjective welfare of households and individuals in low income countries (see for e.g. Fafchamps and Shilpi, 2008; Ravallion and Lokshin, 2010). Also, in a recent paper Roychowdhury (2017), looks at the psychosocial impact of inequality on conspicuous consumption using data from India. These papers, however, are not concerned with quantifying or evaluating endogenous peer effects in consumption.

[6] In fact, being able to control for a host of exogenous peer effects also means that the concern of omitted variable bias emanating from group level unobservables (which may remain despite the inclusion of district fixed effects) gets considerably reduced.

exogenous peer effects in the main equation predicting the outcome and use these as instruments for peer outcome variable (e.g., Gaviria and Raphael, 2001; McVicar and Polanski, 2014). However, as noted by Fletcher (2010), it is not very clear if any of their instruments are appropriately excluded from the main equation.

The paper unfolds as follows. In section 2, I set out the theoretical model. Section 3 discusses the data. Section 4 discusses the identification issues and presents the empirical strategy. The main results and results of robustness checks are presented in Section 5. The last section concludes.

2   A Simple Model of Social Interactions in Consumption

In this section, I develop a theoretical framework of peer effects in consumption to motivate my empirics. I adjust a standard model of consumer choice with a social utility component that captures the satisfaction that a household gets from status-seeking behavior. In other words, it is utility that a household obtains from getting ahead of others in the peer group. My model predicts that a household's consumption is positively related with that of its peers'.

Basic Setup. Suppose a finite set of households $\{i = 1, 2, ..., M\}$ is partitioned into non-overlapping social groups indexed by $r = 1, 2, ..., \bar{r}$. Let M_r be the rth social group of size m_r. I define household i's peer group as all other members in its social group. Let $M_{i,r}$ be household i's group of peers of size $m_r - 1$. A peer is any other household whose consumption decision and personal characteristics may affect i's consumption decision.

Households in my framework live for one period and have a given endowment of income; let $y_{i,r}$ denote household i's income. Household i uses its income to finance its consumption $c_{i,r}$ and leave bequests $b_{i,r}$ for its future generations which is motivated by joy-of-giving. Even though in a narrow sense $b_{i,r}$ captures financial bequests, it can be interpreted as any investment (e.g., human capital) from current income that enhances the productive capacity of children (e.g., health, education).

Preferences. I assume households have common preferences given by the utility function

$$U_{i,r} = u_1(c_{i,r}, b_{i,r}; \mathbf{\Psi}_{i,r}) + v(c_{i,r} - \bar{c}_{-i,r}) \tag{1}$$

where $u(.)$ is a 'conventional' utility function; $v(.)$ captures social utility or utility from social comparisons which measures the utility attributable to the deviation of own consumption from average peer consumption given by $\bar{c}_{-i,r} = \frac{1}{\hat{m}_r - 1} \sum_{j \in M_{i,r}} c_{j,r}$. The component $\mathbf{\Psi}_{i,r}$ introduces the exogenous heterogeneity that

captures the differences between households. Sources of such heterogeneity are household i's demographic characteristics, $\mathbf{x}_{i,r} = [x_{i,r}^1, x_{i,r}^2, ..., x_{i,r}^K]$, idiosyncratic preference shocks that affect wellbeing of household i, $\epsilon_{i,r}$, as well as average income and non-income demographic characteristics of household i's peers given by $\overline{y}_{-i,r}$ and $\overline{\mathbf{x}}_{-i,r} = [\overline{x}_{-i,r}^1, \overline{x}_{-i,r}^2, ..., \overline{x}_{-i,r}^K]$ respectively, where $\overline{y}_{-i,r} = \frac{1}{\tilde{m}_r - 1} \sum_{j \in M_{i,r}} y_{j,r}$ and $\overline{x}_{-i,r}^k = \frac{1}{\tilde{m}_r - 1} \sum_{j \in M_{i,r}} x_{j,r}^k$, $k = 1, 2, ..., K$. I assume that the idiosyncratic preference shocks that affect decisions of household i (and which are a part of $\Psi_{i,r}$) can be decomposed into two components as follows:

$$\epsilon_{i,r} = \mathbf{z}_{i,r} + \xi_{i,r} \tag{2}$$

where $\mathbf{z}_{i,r}$ denotes a $J \times 1$ vector of 'observed' (or, reported) idiosyncratic shocks $z_{i,r}^j$, $j = 1, 2, ..., J$, and $\xi_{i,r}$ denote the 'unobserved' idiosyncratic shock. I will come back to this assumption later.

I assume that the social utility function $v(.)$ measures utility due to status-seeking behavior or utility from getting ahead of others. Specifically, I assume $v'(.) > 0$ and $v''(.) < 0$. These assumptions have been used by Corneo (2002) and Clark and Oswald (1998) among many others, and implies that people enjoy surpassing others' consumption, and that comparison utility is concave (i.e., diminishing marginal comparison utility holds or higher is a household's consumption relative to the mean, lesser is its increase in utility due to increase in status). As noted by Clark and Oswald (1998, p. 137), "...individuals constantly compare themselves to others. The most natural interpretation is that they feel in competition with human beings like themselves, and enjoy a sense of well-being when they out-perform their peers, or obtain a sense of failure and envy when they rank poorly besides those peers. Much of social psychology rests upon notions of this kind."

Maximization Problem. The problem of household i is to choose $c_{i,r}$ and $b_{i,r}$ by maximizing

$$U_{i,r} = u(c_{i,r}, b_{i,r}; \mathbf{x}_{i,r}, \mathbf{z}_{i,r}, \xi_{i,r}, \overline{y}_{-i,r}, \overline{\mathbf{x}}_{-i,r}) + v(c_{i,r} - \overline{c}_{-i,r}) \tag{3}$$

subject to a income budget constraint given by:

$$c_{i,r} + b_{i,r} \leq y_{i,r}; c_{i,r}, b_{i,r} \geq 0 \tag{4}$$

Solving the first order conditions of the above maximization problem, one can obtain

$$c_{i,r} = \Phi(\overline{c}_{-i,r}, y_{i,r}, \mathbf{x}_{i,r}, \mathbf{z}_{i,r}, \xi_{i,r}, \overline{y}_{-i,r}, \overline{\mathbf{x}}_{-i,r}) \tag{5}$$

which is the consumption best-response function of household i in peer reference group r.

Comparative Static. In equation (5), my right hand side variable of interest is $\bar{c}_{-i,r}$. Specifically, I am interested to find how consumption of household i responds to a change in its average peer consumption, which in the peer effects literature is referred to as the *endogenous* peer effect. To sign this comparative static effect in absence of any explicit functional form of the utility function, note that for an interior maximum the following condition must hold:

$$u_1'(c_{i,r}, y_{i,r} - c_{i,r}; \mathbf{x}_{i,r}, \mathbf{z}_{i,r}, \xi_{i,r}, \bar{y}_{-i,r}, \overline{\mathbf{X}}_{-i,r}) - u_2'(c_{i,r}, y_{i,r} - c_{i,r}; \mathbf{x}_{i,r}, \mathbf{z}_{i,r}, \xi_{i,r}, \bar{y}_{-i,r}, \overline{\mathbf{X}}_{-i,r}) + v'(c_{i,r} - \bar{c}_{-i,r}) = 0$$

$$(6)$$

where $u_1' = \partial u / \partial c_{i,r}$ and $u_2' = \partial u / \partial (y_{i,r} - c_{i,r})$.

Differentiating (6) implicitly and simplifying I get:

$$\frac{\partial c_{i,r}}{\partial \bar{c}_{-i,r}}$$

$$= \frac{v''(c_{i,r} - \bar{c}_{-i,r})}{u_1''(c_{i,r}, y_{i,r} - c_{i,r}; \mathbf{x}_{i,r}, \mathbf{z}_{i,r}, e_{i,r}, \bar{y}_{-i,r}, \overline{\mathbf{X}}_{-i,r}) + u_2''(c_{i,r}, y_{i,r} - c_{i,r}; \mathbf{x}_{i,r}, \mathbf{z}_{i,r}, e_{i,r}, \bar{y}_{-i,r}, \overline{\mathbf{X}}_{-i,r}) - v''(c_{i,r} - \bar{c}_{-i,r})}$$

$$(7)$$

The numerator on the right-hand side of equation (7) is negative. The denominator is also negative by the requirement of the maximization problem to be concave. Thus equation (7) implies that

$$\frac{\partial c_{i,r}}{\partial \bar{c}_{-i,r}} > 0. \tag{8}$$

In other words, an increase in average consumption causes household i to increase its own consumption. The focus of this paper is to test this prediction. To do so, I formulate an econometric model in Section 4 to estimate the consumption best-response function given by equation (5). The estimation of equation (5) however faces several challenges. First, average peer consumption is endogenous due to simultaneity. Second, there might be unobserved aggregate shocks that are correlated with average peer consumption and possibly with other variables in equation (5) and/or there might be non-random sorting of households into social groups – both of which would bias the estimates of peer effects. Nonetheless, I leave the discussion of these issues for Section 4 and now I turn to describe my data.

3 Data

3.1 The Indian Human Development Survey 2012

The Data. The Indian Human Development Survey (IHDS) 2012 is a nationally representative multitopic household survey conducted by the National Council for Applied Economic Research (NCAER) in New Delhi and University of Maryland (Desai et al., 2015). It was designed to complement existing Indian household surveys by bringing together a wide range of topics in a single survey. This breadth permits analyses of associations across a range of social and economic conditions. The sample was drawn using stratified random sampling.

The IHDS 2012, conducted between November 2011 and October 2012, covers 42,152 households in 1420 villages and 1042 census-defined urban neighborhoods located throughout India.[7] The survey covered all the states and union territories of India except Andaman and Nicobar, and Lakshadweep. These two account for less than 0.05 percent of India's population. The data is publicly available from the Data Sharing for Demographic Research program of the Inter-university Consortium for Political and Social Research (ICPSR).[8]

The IHDS 2012 is based on two one-hour interviews with a knowledgeable informant in each household. The interviews covered health, education, employment, economic status, marriage, fertility, gender relations and social capital of the households. The survey instruments were translated into 13 Indian languages and were administered by local interviewers. The main advantage of using the IHDS is that it includes many questions that are not asked in the larger and more commonly used Indian household survey, the National Sample Survey (NSS). In particular, detailed questions on income, consumption expenditure, and expenditure shocks are asked in the IHDS 2012. It also has a broad array of demographic and family questions which permits me to include a variety of control variables. Moreover, the survey provides the precise geographic location of households (i.e., one can identify a household's village or neighborhood from the survey). All these features make the use of IHDS 2012 attractable for the present research.

Peer Group. Like most other household surveys, IHDS does not have precise social interactions data. Consequently, I need to define peer/reference groups of the households before carrying out my analysis. As noted by Cojocaru (2014), in laboratory experiments, the relevant reference group is obvious in games

---

[7]According to the Indian National Census (2011a), the definition of urban area is as follows: (1) All places with a municipality, corporation, cantonment board or notified town area committee, etc.; (2) All other places which satisfied the following criteria: (i) A minimum population of 5,000; (ii) At least 75% of the male main working population engaged in non-agricultural pursuits; and (iii) A density of population of at least 400 persons per sq. km. The urban neighborhoods are defined as per the Indian Census. Each decade, the Census draws neighborhoods of approximately equal size (about 200 households) in all Indian towns and cities (urban areas).

[8]http://www.icpsr.umich.edu/icpsrweb/DSDR/studies/36151

involving two subjects, and most theories of other-regarding preferences in n-person games assume the remaining $(n-1)$ actors to form the relevant reference group (Fehr and Schmidt, 2006). It is much less clear what the relevant reference groups are in the general population for purposes of social comparisons, and there is little consensus in the literature on this issue. The social comparison theory proposed by Festinger (1954) suggested that individuals seek to compare their abilities/opinions with others who are perceived to be similar in relevant dimensions. Duesenberry (1949), in formulating the relative income hypothesis, took the neighbors as the group against which relative status is assessed. Deaton (2001) also suggests that, in absence of precise social interactions data, the most sensible reference group of households are those that live in the immediate geographic location. Fafchamps and Shilpi (2008), in fact, suggest that in context of low income countries immediate neighbors is likely to constitute a 'natural' reference group.[9] This is because social-psychologists have shown that, when making relative consumption assessments, people compare themselves with a peer group composed of people who started from the same conditions, e.g., those they grew up with. Since in India social mobility at the household level is remarkably low (Munshi and Rosenzweig, 2009; Munshi and Rosenzweig, 2016) meaning that most people live along people they grew up with, "people almost certainly compare themselves to their immediate geographical neighbors" (Deaton, 2001, p. 21).

Due to these reasons, I assume that the villages (in rural areas) and neighborhoods (in urban areas), which basically are small geographic units[10] populated by households who are similar in many dimensions and are exposed to similar geographic and institutional conditions, form the relevant social groups.[11] The peer group or the reference group of a household is, therefore, comprised of all other households in its village or neighborhood. In the section on robustness checks, however, I consider various alternative definitions of peer group to check the sensitivity of my baseline results.

3.2 Variables and Sample Summary Statistics

Outcome Variable. The aim of the paper is to examine social interactions in consumption. The key outcome variable is, therefore, household consumption expenditure. There are fifty-two consumption categories in the IHDS 2012 (see Table A1 in the Supplementary Appendix for a complete list of consumption categories). Thirty of the consumption categories, which are frequently purchased items, use a thirty day time frame while the other twenty-two use a three hundred and sixty five day time frame. I convert all

---

[9] That, following the suggestions of Deaton (2001) and Fafchamps and Shilpi (2008), "most empirical studies use geographic limiters" in defining reference groups has recently been noted by Bhuiyan (2017, p. 2). Cojocaru (2014) also provides a detailed review of how reference groups are defined in non-experimental literature on social comparisons, and reaches a similar conclusion.

[10] The average area of villages included in the IHDS is approximately 3.3 sq. miles. While the average area of urban neighborhoods are not available from the survey, these are also likely to be reasonably small since, as mentioned previously, the average number of households residing in a census-defined urban neighborhood is roughly 200.

[11] Implicitly, the assumption here is that neighborhoods are urban counterparts of rural villages.

expenditures to the annual time frame. The sum of the expenditure of households on all these consumption categories form the household consumption expenditure.

Demographics. Estimation of the household consumption function given by equation (5) requires me to control for household characteristics that includes household current income and other demographic variables. Fortunately, unlike most other household surveys, IHDS 2012 has very good data on current household income. The current income of households reported in the survey is the sum total (for each household) of wages and salaries, non-farm business income, net agricultural income, remittances, property and other income and public benefits. Each of these incomes are in turn constructed from more than fifty different sources of income queried in the survey.[12]

The set of demographic variables can be classified into two categories: characteristics of household heads and socioeconomic features of households. Characteristics of household heads include age, gender, marital status, and literacy status (whether literate or not). Also a set of dummy variables indicating caste affiliation of the household head are included: Brahmin, non-Brahmin forward caste, other backward caste (OBC), Scheduled Castes (SC), Scheduled Tribes (ST) and others. The socioeconomic features of households that are used as controls are: household size, number of years they have been living in their current village or neighborhood, proportion of children, adolescents and adults in the household, binary variables indicating the number of married household members and a binary variable indicating whether the geographical location of the household is classified as urban or rural as per the Indian National Census (2011a).

Idiosyncratic Shocks. Additionally, I also need to control for idiosyncratic shocks that potentially might impact household consumption expenditure. As noted before, IHDS 2012 has several questions on expenditure shocks, i.e., events that have impacted households' consumption expenditure in the recent past. Most of these events are likely to be unforeseen and hence might be considered random. This is an unique feature of the IHDS 2012 compared to other similar household surveys. As I will show later, the availability of data on expenditure shocks in the IHDS 2012 is extremely important for the present study not only because my theoretical model requires me to control for idiosyncratic shocks, but also because my identification strategy (which I shall discuss later) crucially hinges on being able to find data on these. My vector of observable idiosyncratic household specific consumption shocks consist of the following events that have been reported by the knowledgeable informant of the households as to have impacted their consumption expenditure: death of a household member, loss of job in the household, incident of theft and incident of break-in. Here, I take care to include only those shocks that are specific to any given household and are not correlated with the

---

[12]Although the IHDS takes considerable care in designing a questionnaire that enumerates several sources of income, measurement errors in income cannot be ruled out. I address this issue further in the section on robustness checks.

occurrence of the same shock in other households.[13]

Analytic Sample. My estimation sample consists of 40,980 households from 1,411 rural villages and 1,039 urban neighborhoods located across 375 districts: these are households in the IHDS 2012 where I have individual level information for household heads and for which the household head is above 18 years of age, household current income is more than zero but less than Rs. 1,000,000 (equivalent to \$15,385), total consumption expenditure is more than zero, information on other household characteristics is non-missing and finally the household lives in a village or neighborhood with not less than three (sampled) members.

Table 1 presents the description and summary statistics of all the variables used in this study.

[INSERT TABLE 1 ABOUT HERE]

4 Empirics

4.1 Econometric Model

Guided by the consumption best-response function presented in equation (5), I assume that my baseline econometric model is given by

$$c_{i,r} = \alpha + \beta \overline{c}_{-i,r} + \gamma y_{i,r} + \mathbf{x}'_{i,r}\boldsymbol{\varpi} + \overline{\mathbf{x}}'_{-i,r}\boldsymbol{\delta} + \theta \overline{y}_{-i,r} + \mathbf{z}'_{i,r}\boldsymbol{\rho} + \xi_{i,r} \tag{9}$$

where all the variables are defined in section 2. The coefficient β captures the endogenous peer effect which measures the effect of average peer consumption on households' own consumption. The *exogenous peer effects* or *contextual effects* are captured by $\boldsymbol{\delta}$ and θ. These measure the impact of peer characteristics and peer income on household consumption. The *household effects*, or the impact of own income and own characteristics, are captured by γ and $\boldsymbol{\varpi}$. Finally, ρ measures the effect of idiosyncratic shocks that a household faces on its own consumption. The coefficient of interest in equation (9) is β. If households exhibit status-seeking behavior, I should obtain $\beta > 0$.

---

[13]By definition idiosyncratic shocks are different from community shocks. While community shocks generally affect most or all members of the households within a particular community, idiosyncratic shock hits only a few of them. In other words, while the proportion of households who are likely to be hit by a community level shock is close to 1, this figure is likely to be much smaller for idiosyncratic shocks. My sample reveals that, on an average, the proportion of households in a village/neighborhood in which there is a death is 0.19, the proportion of households that suffer a job-loss is 0.03, the proportion of households that report an incident of theft is 0.04, and the proportion of household that report an incident of break-in is 0.02. This is in line with the notion of idiosyncratic shocks

4.2 Identification Issues

There are two main challenges in identifying and estimating peer effects based on my basic empirical specification given by equation (9). Firstly, since household i's consumption affects its peers' mean consumption and vice versa, one cannot distinguish if a group member's action is the cause or the effect of peers' influence. This in turn will imply that β is subject to endogeneity bias. Manski (1993) labeled this the 'reflection problem'.

Secondly, even though I have a comprehensive vector of group level controls (i.e., peer characteristics), there might still be certain unobservable environmental attributes that are specific to social groups and/or common to all members of a particular group. That is, the unobserved error term might be of the following form:

$$\xi_{i,r} = \upsilon_r + \varepsilon_{i,r}, \tag{10}$$

where υ_r is the vector of group specific unobserved characteristics common to all group members (e.g., same motivation towards consumption or similar credit constraints), and $\varepsilon_{i,r}$'s are innovations. It is evident that υ_r's are correlated with mean peer consumption expenditure. In fact, the village/neighborhood level unobserved effects may be also correlated with household characteristics, and peer characteristics. Econometrically, this would imply existence of a non-zero correlation between the group unobservables υ_r and one or more regressors in equation (9). If there are such unobserved heterogeneity across social groups, then estimates of peer effects will be biased.

Additionally, the problem of correlated unobservables could arise if households self-select into social groups with specific objectives (Falk and Knell, 2004). One way of doing this is typically via migration or residential relocation (Stark and Taylor, 1991). For instance, a poor household living in a prosperous neighborhood, to reduce its feeling of relative deprivation, might want to relocate to a less prosperous neighborhood. Frequently there is such positive selection in which 'similar' people join or are assigned to the same group (Sacerdote, 2011). This non-random sorting potentially implies that unobserved households characteristics are correlated with the characteristics of the group which could cause substantial upward bias in the estimated magnitude of the endogenous and contextual peer effects. However, this is unlikely to be a cause of concern in my case, given that the spatial mobility at the household level is extremely low in India (Munshi and Rosenzweig, 2009; Munshi and Rosenzweig, 2016). In fact, the data that I use in this paper also shows that around 97% of the sampled households have been living in the same place for more than 10

years, while 77% report to have been living in the same place 'forever'.[14][15]

4.3 Empirical Strategy

To identify endogenous peer effects, I implement an approach based on IV/fixed effects. My IV strategy, in essence, is similar to the method of 'spatial IV' used widely in empirical spatial literature (see Gibbons and Overman (2012) and Gibbons et al. (2015) for an overview). Not only does this strategy offers the advantage of computation ease, IV estimation remains consistent even in the presence of spatially correlated error terms (Kelejian and Prucha, 1998; Brueckner, 2003). In addition, I try to address the potential problem of unobserved heterogeneity across social groups – that is not taken care by my IV strategy – by including district fixed effects.

IV Strategy. I begin by describing my IV strategy. To start with, I assume that village/neighborhood unobserved effects (if present) are uncorrelated with household characteristics and observable shocks. I shall relax this assumption later.

To see how my approach works, I express my baseline model given by equation (9) in matrix notations. For expository purposes, I assume that own income and peer income do not affect consumption (i.e., $\gamma = \theta = 0$). Also assume that households have unique characteristics and face unique observable idiosyncratic shocks (i. e., $K = J = 1$). Recall that in my framework, I have assumed that households are affected by all other households in their social group and by none outside it. This means that the observed social interactions can be modeled as an $M \times M$ block-diagonal matrix $\mathbf{G}$ specified as

$$\mathbf{G} = diag(\mathbf{G}_1, \mathbf{G}_2, ..., \mathbf{G}_{\overline{r}}) \tag{11}$$

$$\mathbf{G}_r = \frac{1}{m_r - 1} \left(\boldsymbol{\iota}_{m_r} \boldsymbol{\iota}'_{m_r} - \mathbf{I}_{m_r} \right), \ r = 1, 2, .., \overline{r} \tag{12}$$

where M denotes total number of households, $\overline{r}$ is the number of social groups, m_r is the group size of group r, $\boldsymbol{\iota}_{m_r}$ is the m_r-dimensional vector of ones, and $\mathbf{I}_{m_r}$ is the m_r-dimensional identity matrix. Thus equation (9) can be rewritten in matrix form as follows:

---

[14]In the Supplementary Appendix, I exclude the households who report to have migrated to their current village/neighborhood in the last ten years, and re-estimate the baseline regression model to check the sensitivity of my baseline results. I repeat the same exercise also by excluding the households who have been living in their current village/neighborhood for less than twenty years.

[15]It is worth noting however that although household level migration between villages/neighborhoods is rare, households often send out individuals/migrants for work who send remittances back home. I explore this issue in more detail in the section of robustness checks.

$$\mathbf{C}_r = \alpha\boldsymbol{\iota}_{m_r} + \beta\mathbf{G}_r\mathbf{C}_r + \mathbf{X}_r\varpi + \delta\mathbf{G}_r\mathbf{X}_r + \mathbf{Z}_r\rho + \boldsymbol{\xi}_r, \qquad E(\boldsymbol{\xi}_r|\mathbf{X}_r, \mathbf{Z}_r, \mathbf{G}_r) = 0 \qquad (13)$$

where $\mathbf{C}_r$, $\mathbf{X}_r$ and $\mathbf{Z}_r$ are $m_r \times 1$ vectors of consumption, household characteristics and idiosyncratic shocks respectively, and $\boldsymbol{\xi}_r$ is the vector of the composite error term $\boldsymbol{\iota}_{m_r}\upsilon_r + \boldsymbol{\varepsilon}_r$. Note that equation (13) is similar to a spatial autoregressive (SAR) model (e.g., Cliff and Ord, 1981).

Concatenating 3quation (13) over all groups yields

$$\mathbf{C} = \alpha\boldsymbol{\iota} + \beta\mathbf{GC} + \mathbf{X}\varpi + \delta\mathbf{GX} + \mathbf{Z}\rho + \boldsymbol{\xi} \qquad (14)$$

where $\mathbf{C}$ (respectively $\boldsymbol{\iota}$, $\mathbf{X}$ and $\mathbf{Z}$) is obtained by stacking vectors $\mathbf{C}_r$ (respectively $\boldsymbol{\iota}_{m_r}$, $\mathbf{X}_r$ and $\mathbf{Z}_r$) for $r = 1, 2, ..., \overline{r}$. Assuming $\beta < 1$, $\mathbf{I} - \beta\mathbf{G}$ is invertible. I can, thus, write equation (14) as:

$$\mathbf{C} = \alpha\left(\mathbf{I} - \beta\mathbf{G}\right)^{-1}\boldsymbol{\iota} + \left(\mathbf{I} - \beta\mathbf{G}\right)^{-1}\left(\mathbf{X}\varpi + \delta\mathbf{GX}\right) + \left(\mathbf{I} - \beta\mathbf{G}\right)^{-1}\mathbf{Z}\rho + \left(\mathbf{I} - \beta\mathbf{G}\right)^{-1}\boldsymbol{\xi} \qquad (15)$$

Note that since $\left(\mathbf{I} - \beta\mathbf{G}\right)^{-1} = \sum_{k=0}^{\infty} \beta^k \mathbf{G}^k$, I can rewrite equation (15) as

$$\mathbf{C} = \alpha/(1-\beta)\boldsymbol{\iota} + \varpi\sum_{k=0}^{\infty}\beta^k\mathbf{G}^k\mathbf{X} + \delta\sum_{k=0}^{\infty}\beta^k\mathbf{G}^{k+1}\mathbf{X} + \rho\sum_{k=0}^{\infty}\beta^k\mathbf{G}^k\mathbf{Z} + \sum_{k=0}^{\infty}\beta^k\mathbf{G}^k\boldsymbol{\xi} \qquad (16)$$

Further, notice that from equation (16) the expected mean peer consumption can be written as:

$$\mathbf{E}[\mathbf{GC}\,|\mathbf{X}, \mathbf{Z}] = \alpha/(1-\beta)\boldsymbol{\iota} + \rho\mathbf{GZ} + \rho\beta\sum_{k=0}^{\infty}\beta^k\mathbf{G}^{k+2}\mathbf{Z} + \varpi\mathbf{GX} + (\varpi\beta + \delta)\sum_{k=0}^{\infty}\beta^k\mathbf{G}^{k+2}\mathbf{X} \qquad (17)$$

This implies that one can use $\mathbf{GZ}$ to instrument for $\mathbf{GC}$.[16] In other words, average peer observable idiosyncratic shocks serve as valid instruments for average peer consumption. Note that my IV strategy exploits the fact that in my theoretical model, the idiosyncratic shocks are assumed to be household specific and that they *do not* contain any information about the consumption of other households, even those located in the same village/neighborhood. As argued by Helmers and Patnam (2014, p. 95), "this is a credible assumption given the idiosyncratic nature of the shocks". Since this assumption implies that average peer idiosyncratic shocks affect average peer consumption and that average peer idiosyncratic shocks have no direct impact on households' own consumption (other than through peer consumption), these could potentially be used as instruments for average peer consumption. My first stage regression, therefore, becomes:

---

[16] In principle, the higher order spatial lags of $\mathbf{Z}$ and $\mathbf{X}$ – $\mathbf{G}^2\mathbf{Z}, \mathbf{G}^3\mathbf{Z},...., \mathbf{G}^2\mathbf{X}, \mathbf{G}^3\mathbf{X}, ...$ – can also be used instruments along with $\mathbf{GZ}$. However, as noted by Gibbons and Overman (2012), the higher order lags do not work well as instruments in practice and gives rise to 'weak instruments/identification' problem. This is because there is little independent variation (and hence little additional information) in the higher order spatial lags of $\mathbf{Z}$ and $\mathbf{X}$, conditional on $\mathbf{GZ}$ and $\mathbf{GX}$.

$$\overline{c}_{-i,r} = \Pi_o + \Pi_1 y_{i,r} + \mathbf{x}'_{i,r}\mathbf{\Pi}_2 + \overline{\mathbf{x}}'_{-i,r}\mathbf{\Pi}_3 + \Pi_4 \overline{y}_{-i,r} + \mathbf{z}'_{i,r}\mathbf{\Pi}_5 + \overline{\mathbf{z}}'_{-i,r}\Pi_6 + u_{i,r} \qquad (18)$$

In equation (18), $\overline{\mathbf{z}}'_{-i,r}$ denotes the vector of average observable peer idiosyncratic shocks where the idiosyncratic shocks are chosen from the events reported by the knowledgeable informant of the households in 2012 IHDS to have had affected their consumption expenditure.

Fixed Effects. The IV strategy described above is sufficient to produce consistent parameter estimates of my baseline econometric model under the null hypothesis of no correlated effects (i.e., $v_r = 0$) or when there are group specific unobserved traits correlated with only consumption expenditure of the households (assuming that the IVs are valid). But what happens if village or neighborhood level unobservables are allowed to be correlated with households' own characteristics? Unfortunately, the above described IV strategy may not produce consistent estimates of the endogenous peer effect when there are such correlations.[17]

To address this concern, I incorporate a full set of district fixed effects. Districts, which represent administrative divisions of an Indian state, are clusters of several villages/neighborhoods located in the same geographical area.[18] The logic behind including the district fixed effects is that these would capture the unobserved heterogeneity at the level of districts. Since villages/neighborhoods within a particular district are likely to be 'similar',[19] the incorporation of the district fixed effects along with the fact that I have a comprehensive vector of group level controls (peer characteristics) is likely to be sufficient for the above discussed IV strategy to produce consistent parameter estimates.[20] The assumption is that, given that I have a comprehensive vector of group level attributes (i.e., contextual effects) in my baseline econometric equation, the remaining group-level heterogeneity (if any) does not vary across villages/neighborhoods within a district.

Validity of Identification Strategy. While in principle the above described IV/fixed effects strategy allows me to consistently estimate the parameters of the baseline econometric model given by equation (9), in practice the success of it depends on whether the set of instruments, $\overline{\mathbf{z}}_{-i,r}$, constructed using the expenditure shocks chosen from IHDS 2012 satisfy two crucial conditions. First, a testable condition is that

---

[17] To see this, suppose that income of every household is correlated with unobserved village/neighborhood level characteristics as well as with household specific idiosyncratic shocks. This means average peer income will be correlated with group level unobservables as well as with average observed peer shocks. This, in turn, is likely to imply that average observed peer idiosyncratic shocks will be correlated with group level unobservables which violates the exogeneity requirement of the instrument.

[18] In India, as of 2014, there are 29 states and, on an average, there are 23 districts in each state. The average area of a district is 1775 sq. miles.

[19] In India, districts are divided on the basis of ethical, cultural and social interaction rather on the basis of easiness or prosperity (Indian National Census, 2011b). As such, villages/neighborhoods within a particular district are likely to be similar along observable and unobservable ethical and cultural dimensions.

[20] My identification strategy ensures that it is not the 'price effect' that is driving my result. This is because my IV strategy will produce consistent parameter estimates even when there are unobserved differences in prices of goods across villages/neighborhoods. In fact, if prices are the *only* omitted village/neighborhood level characteristics, I do not even need to use district fixed effects since prices, presumably, are uncorrelated with the demographic characteristics (and shocks) of households.

average peer idiosyncratic shocks must be correlated with average peer consumption (which should hold as long as the shocks affect household consumption expenditure). Second, the idiosyncratic shocks hitting the peers of household i should affect the household's own consumption *only* indirectly through their impact on its peers' consumption. This is tantamount to saying that average peer idiosyncratic shocks must be uncorrelated with the error term. This second criterion is an untestable maintained assumption.

To examine whether the chosen shocks satisfy the first condition, I carry out the Kleibergen-Paap rk LM test (2006). The LM test seeks to test whether that the excluded instruments are correlated with the endogenous regressors. The null hypothesis of this test is that the minimum canonical correlation between the endogenous variables and the instruments is not statistically different from zero. Rejection of the null hypothesis indicates that the model is identified. Further, since IV estimates based on 'weak' instruments are biased towards OLS estimates (Bound et al., 1995; Staiger and Stock, 1997; Stock et al., 2002) I report the F-statistic from the first stage regressions which is the test to examine strength of instruments. According to Staiger and Stock (1997) and Stock and Yogo (2005), an F-statistic value of 10 (or higher) implies rejection of the null hypothesis of weak instruments.

While, as noted above, it is in general not possible to test whether the instruments are uncorrelated with the error term, I carry out *balancing tests* (Bifulco et al., 2011; Lavy and Schlosser, 2011) to assess the likelihood of my instruments being uncorrelated with the unobserved error term. This test seeks to assess the correlation between the instruments and unobserved error term based on the correlation between the instruments and observed characteristics of households.[21] If the instruments are found to be uncorrelated with observable household characteristics related to consumption (more than what would be expected by chance), they may also be uncorrelated with the unobservable factors related to consumption (that are in the error term), following the logic of Altonji et al. (2005). As argued by Fletcher (2010, 2012, 2015), this is suggestive, but not conclusive, that the instruments can be treated as plausibly exogenous (or random 'shocks').

Further, since my model is over identified, I carry out the Hansen J test (1982), which is an overidentification test designed to examine the validity of the instruments (in terms of satisfying the exogeneity restriction). This test seeks to obtain multiple estimates of the treatment effect (which in the present case are the endogenous peer effect) based on various subsets of the instrumental variables and tests whether the obtained treatment effects are same. If all instruments are uncorrelated with the error term, all subsets should (asymptotically) return the same estimate of the treatment effect. The joint null hypothesis of this test is that the instruments are valid instruments (i.e., uncorrelated with the error term) and that the ex-

---

[21]I perform this test by regressing households' own background characteristics (income, household size, etc.) on peer shocks (instruments) controlling for district fixed effects, plus own shocks, and peer background characteristics.

cluded instruments are correctly excluded from the estimating equation. A rejection of the null hypothesis casts doubt on the validity of the instruments. However, failure to reject the null does not necessarily mean that the exclusion restriction holds.

Model Estimation. I estimate my baseline model by the technique of Generalized Method of Moments (GMM). Since my model is overidentified, I report the two-step GMM estimates or optimal GMM estimates, which is the most efficient GMM estimator for overidentified models with heteroscedastic errors of unknown form (for a detailed overview of the two-step GMM see Cameron and Trivedi, 2005 and Baum et al., 2007).

4.4 Threats to Identification

As noted previously, both the balancing tests as well as the Hansen J test provide only suggestive evidence that the IV/fixed effects approach is valid. In practice, however, one can think of several situations where my identification strategy might not yield consistent estimates of the endogenous peer effect. First, as discussed above, the district fixed effects would take care of the group level omitted variable bias if villages/neighborhoods within a district do not differ along unobservable ways. If, even after conditioning on the contextual effects, villages/neighborhoods within a district differ along unobservable ways, then the IV/district fixed effects-based strategy would not be sufficient to address the endogeneity due to group level unobservables.

Second, the instruments used in this study are created based upon idiosyncratic shocks that are all necessarily 'negative' and 'accidental' in nature. Thus, the instruments are exploiting the variation in average peer consumptions related to this particular type of shock. Whether the results would hold if instead I had considered other types of shocks (e.g. 'positive' idiosyncratic shocks like winning a lottery) is not readily apparent.

Third, since the peer idiosyncratic shocks used to construct the instruments are all 'negative', it may be the case that they are correlated with a household's feeling of reciprocity or empathy which I am not able to control for in the baseline model. For instance, a household hears that its relevant neighbor just met with an accident, and this may cause that household to be more empathetic or augments its feelings of reciprocity. This may in turn cause it to exhibit less of a status-seeking behavior. If this is the case, not controlling for such social feelings will cause my instruments to be correlated with the error term. However, it is extremely difficult to find data on variables that would capture social attitudes, feelings and emotions from non-experimental household surveys (especially those carried out in developing countries) that would allow me to perfectly control for these effects.

Finally, the social interaction structure considered in this paper may not be very realistic. For instance,

instead of being equally affected by all others in the same village/neighborhood, a household may be influenced more significantly by some people in the village/neighborhood, especially by its 'friends'. On the other hand, a household that does not send out migrants for work, might not be affected at all by households in its village/neighborhood that does so. This is because the households that send out migrants for work might be very different in their wealth levels compared to the households that do not.

In the section of robustness checks, I try to address some of these limitations in more detail and check the sensitivity of my baseline results to various sorts of biases. For now, I turn to the discussion of my baseline results.

5   Results

5.1   Main Results

OLS Results. Table 2 reports the naive OLS results. These estimates provide a useful benchmark with which to compare the results from the IV method. Column (1) regresses household consumption on average peer consumption, household income, household characteristics and household specific idiosyncratic shocks. Column (2) extends the set of regressors to include contextual effects. Column (3) adds district fixed effects. Note that, I report only the point estimates of endogenous peer effect for every specification. Full results are relegated to the Supplementary Appendix due to space constraints (see Table A2).

[INSERT TABLE 2 ABOUT HERE]

Across all specifications, I find a strong association between own and peer consumption with the coefficient of average peer consumption varying between 0.2 and 0.6. Moreover all the coefficients are statistically significant at 1% level of significance. These estimates are not, however, what I would wish to take seriously, given the identification issues discussed above, which OLS fails to resolve.

IV Estimates. Before presenting results using instrumental variables, I present the results from the balancing tests (Bifulco et al., 2011; Lavy and Schlosser, 2011) to show that the preferred instrumental variables are plausibly exogenous. In Table 3, I show that the instrumental variables are conditionally uncorrelated with all household characteristics. This evidence is suggestive, but not conclusive, that the instruments may also be uncorrelated with unobservable factors related to household consumption (which are in the error term), following the logic of Altonji et al. (2005), which is a maintained untestable assumption for the analysis to be valid. Overall, these results add confidence that the instruments are quasi-random within districts.

[INSERT TABLE 3 ABOUT HERE]

The IV results are reported in Table 4. As in the case of OLS, I report only the estimates of endogenous peer effects for each specification. Full IV results (including the first stage) are presented in Table A3 in the Supplementary Appendix.

I start by describing the performance of IV/fixed effect models in terms of the diagnostic tests. First, notice that all the specifications perform remarkably well in terms of the Hansen (1982) overidentification test, Kleibergen-Paap (2006) rk LM test for underidentification as well as the F-test for excluded instruments to assess the strength of the instruments. Specifically, for all the specifications, based on the Hansen J statistic, I am strongly unable to reject the joint null hypothesis that the instruments are uncorrelated with the error term and that the excluded instruments are correctly excluded from the estimated equation. Next, across all the specifications reported, the estimated Kleibergen-Paap rk LM statistic allows me to clearly reject the null hypothesis that the instruments are uncorrelated with the endogenous regressors and that the model is not identified. Finally, the first stage F-statistic for excluded instruments lies well above 10 across all the specifications in each panel, which clearly indicates that none of the specifications suffer from the weak instrument problem.[22]

[INSERT TABLE 4 ABOUT HERE]

Next, I turn to the actual two-step GMM estimates of the endogenous peer effects in consumption. Strikingly, across all specifications, I find strong evidence of endogenous peer effects in consumption. This is in line with the OLS results. Specifically, column (1) shows that the estimated impact of peers' consumption on own consumption based on the regression that neither includes contextual effects nor district fixed effects is 0.78 (s.e. = 0.147). Column (2) shows that the estimated impact of endogenous peer effects based on the regression that adds contextual effects (but not district fixed effects) to the set of regressors is 0.85 (s.e. = 0.128). Both these coefficients are also statistically significant at 1% level of significance. These results indicate that under the null hypothesis of no correlated unobservables, there is clear evidence of strong endogenous peer effects in consumption. However, if there are village/neighborhood level unobservables correlated with household characteristics and/or household specific idiosyncratic shocks, then these estimates are not consistent.

---

[22]The first stage IV results reported in Table A3 in the Supplementary Appendix also show that the individual instruments (peer idiosyncratic shocks) are significantly correlated with the endogenous variable (peer consumption expenditure). That, for my preferred specification (column 5), these correlations are positive – conditional on the fact the peer income is held constant in the first stage – is not surprising: a death in the household in India is followed by a huge social ceremony that entails a huge expenditure. Similarly, a theft or breakin may cause households to replace the stolen goods or enhance the security of their homes – both of which are likely to cause expenditure to rise. Finally, a job loss in the household may lead to increase in consumption expenditure due to at least two reasons. First, a job loss might be followed by the person losing job to enroll in a job-training program which might increase total household expenditure. Second, there is ample evidence that suggests that job loss is associated with health compromising behaviors like alcohol abuse, which might also drive up total household expenditure.

To account for the potential problem of correlated unobservables, I report the results of my preferred specification reported in column (3) which regresses own consumption on average peer consumption controlling for own income, own characteristics, own idiosyncratic shocks, average peer income, average peer characteristics and district fixed effects. I find the estimate of endogenous peer effect to be 0.69 (s.e. = 0.223) which is statistically significant at 1% level of significance. This means that an increase in (mean) peer consumption expenditure by 1 Indian Rupee causes households to increase their own consumption expenditure by roughly 0.7 Indian Rupee, on average. In terms of standard deviations, this translates into a 0.34 standard deviations (roughly 37,637 Indian Rupees/year) increase in own consumption in response to a one standard deviation (roughly 54,546 Indian Rupees/year) increase in peer consumption. Note that this effect is not only large in absolute terms but is in fact exactly equal to the impact of an equal increase (in terms of standard deviation) in own income. Overall, my results indicate the presence of strong peer effects in consumption.

To my knowledge, the only other study that seeks to estimate peer effects in aggregate consumption is that by De Giorgi et al. (2016). Using network data from Denmark they find that the elasticity of own consumption with respect to peers consumption is 0.3. How do my results compare to the findings of De Giorgi et al. (2016)? To answer this question, I estimate a double log version of my baseline specification (i.e., I use both households own consumption and average peer consumption in logs). My IV strategy delivers an estimate of the elasticity of roughly 0.5 (s.e. = 0.124).[23] This indicates that not only there are substantial peer effects in consumption in a low income country, but these effects are larger than those obtained for high income countries.

Social Multiplier. Peer effects potentially imply that small exogenous shock at the household level is magnified through the social interactions process to deliver larger aggregate level social effects. Glaeser and Scheinkman (2001) and Glaeser et al. (2003) define the *social multiplier* as the ratio of the household effect from an *exogenous* shock to the aggregate effect from the same shock. As argued by Sacerdote (2011), social multiplier is useful because it delivers the parameter of direct interest to policy makers, namely if a policy can exogenously induce one additional person to take action A, how many total people will take action A in *equilibrium*? In the Supplementary Appendix, I show that the magnitude of the social multiplier based on my baseline regression equation (9) is $1/(1 - \beta)$. Thus, based on my preferred specification (i.e., $\beta \approx 0.7$) the value of the social multiplier in the present work is roughly equal to 3. This means that the equilibrium response to a shock that induces an exogenous variation in mean household consumption spending is about 3 times the initial average response.

---

[23] Results of this specification are not reported, but are available from the author on request.

Heterogeneity Analysis. To check for heterogeneity in endogenous peer effects, I report results of sub-sample analysis in Table 5. Specifically, I cut the sample along three dimensions: household poverty status, caste affiliation, and area of residence.

[INSERT TABLE 5 ABOUT HERE]

It turns out that the point estimates of the endogenous peer effect is higher for the non-poor households compared to the households who fall below the poverty line, for the high caste households compared to the low caste ones, as well as for the households living in urban areas compared to the rural households. Note, however, the difference in endogenous peer effect between the subsamples is statistically significant only for the first case, and not for the second or the third cases. This perhaps indicates that those households who are just trying to survive perhaps do not look at what others consume. Once households cross a minimum economic threshold they start looking at their peers consumption and consume some goods accordingly.[24] This implies that the problem of under-saving or over-borrowing owing to status-seeking consumption behavior is likely to be more salient for the non-poor households compared to households who are officially classified as poor.

5.2   Robustness Checks

In this section, I conduct various robustness checks to assess the robustness of my estimates to different sorts of biases.

Falsification Test. I begin by providing a falsification test. All my results indicate the presence of a positive and significant endogenous peer effects in consumption of a household where peer groups are assumed to comprise of all other people living in the same village or neighborhood. I now show that such a result is not obtained from considering *just any random* peer group. In essence, I validate the strength and significance of the actual observed peer group by ruling out the presence of peer effects within randomly generated peer groups. Put differently, this is a test for my identifying assumption that geographical proximity mediates peer effects.

To test this, I *randomly* assign peers to each household from across the entire sample. The total number of peers assigned to each household is equal the average size of peer groups, which is equal to 16. Subsequently, I estimate equation (9) controlling for district fixed effect. I repeat this exercise 100 times, each time randomly assigning peers to every household.

---

[24]Whether the differences in endogenous peer effects across subsamples are statistically significant or not is judged based on the results of the z-test (Paternoster et al., 1998). A value of z-statistic below 2 generally implies that the difference between the estimated endogenous peer effects is not statistically significant.

The histogram in Figure 1 shows the empirical distribution of t-statistics corresponding to the endogenous peer effects obtained from 100 replications. The mean t-statistic is 1.06. This means that I cannot reject the null hypothesis that endogenous peer effect is equal to zero. Moreover, the majority of point estimates of endogenous peer effects are statistically insignificant as indicated by the distribution of the t-statistics. Specifically, for peer effects associated with each of the 100 iterations, I find that I am unable to reject the null that the coefficient (point estimate) is equal to zero for 89 coefficients out of 100. This shows that repeated experiments with different randomized social groups produce statistically insignificant endogenous peer effects on average. Hence, this exercise lends support to my approach of constructing peer groups based on geographical proximity.

Inclusion of controls for covariate shocks. Despite having a comprehensive set of group level controls, as well as district fixed effects in my preferred specification of the baseline regression model given by equation (9), there might still be some omitted factors that vary between groups within districts. If this is the case, the estimates of the endogenous peer effects may not be consistent. To examine this, I augment my set of regressors by including three important covariate (community) shocks, namely, crop failure, droughts/floods and conflict. By definition, covariate shocks affect almost all households within a neighborhood/village. As such, it would be interesting to check the sensitivity of my baseline results to inclusion of these group level shocks.

The results are reported in columns (1) - (3) of Table 6. I find that inclusion of the covariate shocks makes almost no difference to the estimates of the endogenous peer effects.[25] This suggests that group level shocks are unlikely to be responsible for the co-movement in household consumption within peer groups reported in my baseline estimation.

Inclusion of correlates of social emotions as controls. In this paper, I identify the endogenous peer effects by exploiting the variation in average peer idiosyncratic shocks. However, since the idiosyncratic shocks on which I have data are all negative, it may be the case that they are correlated with an individual's feeling of reciprocity or empathy which I do not control for in the baseline model. As noted previously, a household might hear that its relevant neighbors just met with an accident. It may cause that household

---

[25] In fact, the coefficient of the covariate shocks themselves turn out to be statistically insignificant. This means that the covariate shocks, conditional on the other household and group level regressors and district fixed effects, have no impact on households' own consumption expenditure.

to be more empathetic or augment its feelings of reciprocity, and this causes it to exhibit less status-seeking behavior. If this is the case, not controlling for such social feelings will cause my instruments to be correlated with the error term of the baseline regression model. However, it is extremely difficult to find data on variables that would capture social attitudes, feelings and emotions from non-experimental household surveys (especially those carried out in developing countries) that would allow me to perfectly control for these effects. However, since I believe this issue is important, I explore the dataset that I use, and also the literature on behavioral issues like empathy and reciprocity to see if I could examine this issue in more detail.

It turns out that according to several studies in the behavioral sciences, feelings of reciprocity and empathy are strongly correlated with prosocial behavior in general (see Telle and Pfister, 2015 for a recent review of the literature on this topic). Now, one way in which prosocial behavior is displayed is of course via participation in social groups/festivals and engaging in informal social networks. This implies that a measure of prosocial behavior is likely to be participation of individuals in social groups/festivals, youth clubs and sports groups. Thus, there is likely to exist a strong correlation between feelings of empathy and reciprocity and participation in social groups/festivals. Therefore, including variables capturing an individual's participation in social groups/festivals as well as in sports groups and informal clubs might serve as useful controls for feelings of empathy/reciprocity. Thankfully, IHDS has information on individuals' participation in social groups/festivals and whether they belong to any sports groups or youth clubs. Specifically, the IHDS asks the household heads: "(1) Does anybody in the household belong to a social group or festival society? (2) Does anybody in the household belong to a youth club, sports group or reading room?" I include these binary variables and re-estimate my baseline model to check the sensitivity of my results to the inclusion of proxies for feelings of reciprocity/empathy. Results are reported in columns (4) and (5) of Table 6. It turns out that the point estimate of my endogenous peer effect remains almost unchanged. Thus, although this is a crude method to control for feelings or emotions of reciprocity/ empathy, nonetheless, this provides some evidence that my results are not driven by omitted variables like social emotions or feelings of empathy.

Local Public Goods. One concern regarding the instruments is that if the instruments are correlated with average peer income, and average peer income is related to a household's consumption via the public goods (assuming there is some substitutability between public and private good) and this is not accounted for, then the instruments would be correlated with the error term. One way to examine this problem is perhaps to include some direct measure(s) of public goods at the local level as additional regressors in the baseline model, and test the sensitivity of the baseline results to the inclusion of these variables. If the estimates of the endogenous peer effects do not change significantly after the inclusion of the public good variables, it

is unlikely to be the case that the instruments are correlated to the error term due to the non-inclusion of measures of local public good provision in the baseline analysis.

To examine this issue, I look for variables in the IHDS that may be potentially used as measures of local public goods. IHDS has a separate village level survey, which, fortunately, has information on government schools and government health facilities (hospitals, clinics, etc.) in villages, which potentially might be used as measures of local public goods (given that public schools and public health facilities are generally considered as important public goods). In particular, IHDS has information on whether there are such public facilities in the villages, as well as the number of government schools and number of government hospitals, clinics, etc. I re-estimate my baseline model including these variables as controls for local public goods. If average village income is correlated with these public goods, then explicitly controlling for these will not cause my instruments to be correlated with the error term (even if the instruments are correlated with average village income). However, note that IHDS has information at the local level only for rural areas (i.e. villages) and not urban areas (i.e., the local survey questionnaire is administered only in rural areas). In other words, I do not have information about local public good provision for urban areas. As such, my sensitivity analysis is restricted only to the rural sample.[26] I estimate two specifications. In the first specification, I include village level controls for number of government primary schools, number of government middle schools, number of government secondary schools, number of government higher secondary schools, and number of government colleges. In the second, I control for number of health subcenters, number of government dispensaries, number of primary health centers, and number of community health centers. For both the specifications, I find that the results, reported in Table 6 (columns (6) and (7)), are similar to my baseline results. Thus, at least, for the rural sample I can conclude that public goods are not a source of omitted variable bias in the baseline analysis.

Alternative Peer Reference Group. In my baseline analysis, I assume that a household's reference group consists of all other households in its village/neighborhood. However, one concern could be that I might care only about 'some' people (and not about everyone) in my village/neighborhood. In other words, it might be useful to define reference groups in one or two additional dimensions, for example by caste and by whether the household has sent out any migrants who send remittances back home, and check the sensitivity of my results. I am particularly interested in defining peer groups based on caste because the importance of the caste system as a regulator of social interactions in India has been highlighted in various studies (e.g., Munshi, 2016). This could potentially imply that self-identification is stronger among households of the same caste living in the same region than among households living in the same region but belonging to different

---

[26] This is the best I could do to address this concern, given the available data.

castes. On the other hand, households who regularly send out migrants who send remittances back home can be very different in their wealth levels and hence may not serve as a reference peer group for those who do not send out migrants.

[INSERT TABLE 7 ABOUT HERE]

One problem in defining reference groups in terms of multiple dimensions in the present case, given that average number of households sampled from villages/neighborhoods is roughly 20, is that I am likely to end up with many households having no other households in their reference group. As I would need to exclude such households, this will reduce the size of my analytical sample. Nonetheless, I re-estimate my baseline econometric model using three alternative definitions of reference groups: (1) village/neighborhood of residence and caste (i.e. households in the same village and same caste group form a peer reference group), and (2) village of residence and migrant status (i.e., households within a village who have not sent out migrants for work form a peer reference group, whereas households within a village who have sent out migrants form a different peer group), and (3) village of residence, caste, and migrant status. Results are reported in Table 7. Reassuringly, the estimates of endogenous peer effect from all the three specifications remains similar to that obtained from the baseline regression model.

5.3 Additional Robustness Checks and Disaggregate Analysis

In addition to the falsification test and the robustness checks presented above, I conduct several other robustness checks. First, I run my baseline regressions transforming consumption expenditure and income in logarithms. Second, I instrument household income by literacy status of the father of the household head (or father of household head's husband). Third, I present IV results of additional specifications using different subsets of the vector of peer idiosyncratic shocks as instruments in the first stage and including the remaining peer shocks as covariates in the second stage. Fourth, I examine whether the 'type' of idiosyncratic shock hitting the household matters. Fifth, I estimate my baseline regression equation based on a sample that excludes households who have migrated to their current area of residence relatively recently. Sixth, I create a household level price index based on prices of nontradable goods. Then I use this price index as an additional control in my regression model. Also, I re-estimate my baseline regression model by deflating total consumption expenditure using this price index. Seventh, I exclude those households who regularly send out migrants from my analytical sample, as these households might come from a different data generating process (DGP). Finally, I use sampling weights and re-estimate my baseline regression model. My baseline results remain robust throughout these exercises.

Additionally, I carry out a disaggregate analysis and find that the increase in own consumption expenditure due to endogenous peer effect is driven by an increase in own expenditure on temptation goods and food. Detailed results of all these additional robustness checks and the disaggregate analysis can be found in the Supplementary Appendix.

6 Conclusion

In this paper, I examine whether consumption of households' peers affect their own consumption in a less developed country. I define a household's peer group as other households living in its village/neighborhood. In assessing the influences of peers in this context, there are two key empirical challenges including shared group-level unobservables, and simultaneity of peer influences. I address these issues by using an instrumental variables/fixed effects approach that compares households in the same district but different villages/neighborhoods who are thus exposed to different sets of peers. In particular, I use plausibly exogenous variation in idiosyncratic expenditure shocks faced by peers as instruments for peers' consumption expenditure. Preferred specification suggests that a 1 Indian Rupee increase in consumption expenditure of a household's peers causes the household's own consumption expenditure to increase by 0.7 Indian Rupee which translates into a social multiplier of about 3. This means that the equilibrium response to a shock that induces an exogenous variation in mean household consumption spending is about 3 times the initial average response. Falsification tests and robustness checks support the validity of the results.

While this paper is able to address many of the relevant econometric issues in estimating the importance of peer influences on consumption, there may be a few limitations as discussed previously. However, even with such limitations of the study, I find robust evidence that peer consumption expenditure provide a strong influence on Indian households' own consumption expenditure. These findings suggest that policies that influence a household's consumption expenditure will also affect the consumption decisions of the household's peers through social interactions. This implies that traditional analyses of consumption intervention programs that do not take into account such spillover effects likely understate the total social impact of the programs. Moreover, the findings also suggest that it might be worthwhile to use innovative social policies that represses social pressure together with traditional anti-poverty policies for raising social welfare in a less developed country.

References

Akerlof, G.A. (1997). Social Distance and Social Decisions. *Econometrica*, 65, 1005–1027.

Angelucci, M., & De Giorgi, G. (2009). Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption? *American Economic Review*, 99, 486–508.

Angelucci, M., De Giorgi, G., & Rasul, I. (2017). Consumption and Investment in Resource Pooling Family Networks. *Economic Journal* (forthcoming).

Altonji, J. G., Elder, T. E., & Tabor, C. R. (2005). Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113, 151–184.

Ammermueller, A., & Pischke, J-S. (2009). Peer effects in European primary schools: evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*, 27, 315–348.

Baum, C.F., Schaffer, M.E., & Stillman, S. (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal*, 7, 465–506.

Bhuiyan, M.F. (2017). Life Satisfaction and Economic Position Relative to Neighbors: Perceptions Versus Reality. *Journal of Happiness Studies*, forthcoming, 1-30.

Bifulco, R., Fletcher, J., & Ross, S. (2011). The effect of classmate characteristics on individual outcomes: evidence from the add health. *American Economic Journal: Economic Policy*, 3, 25–53.

Boucher, V., Bramoulle, Y., Djebbari, H., & Fortin, B. (2014). Do peers affect student achievement? Evidence from Canada Using Group Size Variation. *Journal of Applied Econometrics*, 29, 91–109.

Bound, J., Jaeger, D., & Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of American Statistical Association*, 90, 443–450.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150, 41–55.

Brueckner, J.K. (2003). Strategic interaction among governments: an overview of empirical studies. *International Regional Science Review*, 26, 175–188.

Calvo-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer effects and social networks in education. *Review of Economic Studies*, 76, 1239–1267.

Cameron, A.C., & Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Clark, A.E., & Oswald, A.J. (1998). Comparison-concave utility and following behaviour in social and economic settings. *Journal of Public Economics*, 70, 133–155.

Cliff, A.D., & Ord, J.K. (1981). *Spatial Processes: Models and Applications*. London: Pion.

Cojocaru, A. (2014). Fairness and inequality tolerance: Evidence from the Life in Transition Survey. *Journal of Comparative Economics*, 42, 590–608.

Corneo, G. (2002). The efficient side of progressive income taxation. *European Economic Review*. 46, 1359–1368.

De Giorgi, G., Frederiksen, A., & Pistaferri, L. (2016). *Consumption Network Effects*. NBER Working Paper 22357.

Deaton, A. (2001). *Relative deprivation, inequality and mortality*. NBER Working Paper 8099.

Desai, S., Dubey, A., & Vanneman, R. (2015). India Human Development Survey-II (IHDS-II) (Computer file), University of Maryland and National Council of Applied Economic Research, New Delhi (producers), 2015. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributor).

Duesenberry, J.S. (1949). *Income, Saving and the Theory of Consumer Behavior*. Cambridge: Harvard University Press.

Fafchamps, M., & Shilpi, F. (2008). Subjective welfare, isolation, and relative consumption. *Journal of Development Economics*, 86, 43–60.

Falk, A., & Knell, M. (2004). Choosing the joneses: endogenous goals and reference standards. *Scandinavian Journal of Economics*, 106, 417–435.

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, 7, 117–140.

Fehr, E., & Schmidt, K.M. (2006). The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories. In S.-C. Kolm and J.M. Ythier (Eds.), *Handbook on the Economics of Giving, Reciprocity and Altruism*, 1, p. 615–691, Amsterdam: North Holland.

Fletcher, J. (2010). Social Interactions and Smoking: Evidence using Multiple Student Cohorts, Instrumental Variables, and School Fixed Effects. *Health Economics*, 19, 466–484.

Fletcher, J. (2012). Peer influences on adolescent alcohol consumption: evidence using an instrumental variables/fixed effect approach. *Journal of Population Economics*, 25,1265–1286.

Fletcher, J. (2015). Social interactions and college enrollment: A combined school fixed effects/instrumental variables approach. *Social Science Research*, 52, 494–507.

Gaviria, A., & Raphael, S. (2001). School-based Peer Effects and Juvenile Behavior. *Review of Economics and Statistics*, 83, 257–268.

Goux, D., & Maurin, E. (2007). Close Neighbours Matter: Neighbourhood Effects on Early Performance at School. *Economic Journal*, 117, 1193–1215.

Gibbons, S., & Overman, H.G. (2012). Mostly Pointless Spatial Econometrics. *Journal of Regional Science*, 52, 172–191.

Gibbons, S. Overman, H. G., & Patacchini, E. (2015). Spatial Methods. In: G. Duranton, J. V. Henderson and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, 5A, 115–168, Amsterdam: North Holland.

Glaeser, E.L., & Scheinkman, J. (2001). Measuring social interactions. In: S. Durlauf, S., H.P. Young (Eds.), *Social Dynamics*, 83–131, Cambridge: MIT Press.

Glaeser, E.L., Scheinkman, J. A., & Sacerdote, B. I. (2003). The Social Multiplier. *Journal of the European Economic Association*, 1, 345–353.

Graham, B.S. (2008). Identifying Social Interactions through Conditional Variance Restrictions. *Econometrica*, 76, 643–660.

Grinblatt, M., Keloharju, M., & Ikäheimo, S. (2008). 'Social Influence and Consumption: Evidence from the Automobile Purchases of Neighbors. *Review of Economics and Statistics*, 90, 735–753.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50, 1029–1054.

Helmers, C., & M. Patnam (2015). Does the rotten child spoil his companion? Spatial peer effects among children in rural India. *Quantitative Economics*, 5, 67–121.

Indian National Census (2011a). Provisional Population Totals: Urban Agglomerations and Cities. Government of India, http://censusindia.gov.in/2011-prov-results/paper2/data_files/India2/1.%20Data%20Highlight.pdf (last accessed September 30, 2016).

Indian National Census (2011b). Top 10 Smallest Districts of India by Total Area. Government of India, http://www.census2011.co.in/facts/smalldistricts.html (last accessed September 30, 2016).

Kelejian, H.H., & Prucha, I.R. (1998). A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics*, 17, 99–121.

Kleibergen, F., & Paap, R. (2006). Generalized Reduced Rank Tests Using the Singular Value Decomposition. *Journal of Econometrics*, 133, 97–126.

Kuhn, P., Kooreman, P., Soetevent, A., & Kapteyn, A. (2011). The Effects of Lottery Prizes on Winners and Their Neighbors: Evidence from the Dutch Postcode Lottery. *American Economic Review*, 101, 2226–2247.

Lavy, V., & Schlosser, A. (2011). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3, 1–33.

Lee, L.F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140, 333–374.

Lee, L.F., Liu, X., & Lin, X. (2010). Specification and Estimation of Social Interaction Models with Network Structures. *Econometrics Journal*, 13, 145–176.

Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60, 531–542.

McVicar. D., & Polanski, A. (2014). Peer Effects in UK Adolescent Substance Use: Never Mind the Classmates? *Oxford Bulletin of Economics and Statistics*, 76, 589–604.

Moav, O., & Neeman, Z. (2012). Saving Rates and Poverty: The Role of Conspicuous Consumption and Human Capital. *Economic Journal*, 122, 933–956.

Moretti, E. (2011). Social Learning and Peer Effects in Consumption: Evidence from Movie Sales. *Review of Economic Studies*, 78, 356–393.

Munshi, K. (2016). Caste Networks in the Modern Indian Economy. In: S. Mahendra Dev and P. G. Babu (Eds.), *Development in India: Micro and Macro Perspectives*, 13-37, New Delhi: Springer.

Munshi, K., & Rosenzweig, M. (2009). *Why is Mobility in India so Low? Social Insurance, Inequality, and Growth.* NBER Working Paper No. 14850.

Munshi, K., & Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. American Economic Review, 106, 46–98.

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36, 859–866.

Putnam, R.D. (2007). E Pluribus Unum: diversity and community in the twenty-first century (The 2006 Johan Skytte Prize Lecture). Scandinavian Political Studies, 30, 137–74

Ravallion, M., & Lokshin, M. (2010). Who cares about relative deprivation? *Journal of Economic Behavior & Organization*, 73, 171-185.

Roychowdhury, P. (2016). Visible inequality, status competition and conspicuous consumption: evidence from rural India. *Oxford Economic Papers*, 63, 36–54.

Sacerdote, B. (2001). Peer effects with random assignment: results for Dartmouth roommates. *Quarterly Journal of Economics*, 116, 681–704.

Sacerdote, B. (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In: E.A. Hanushek, S. Machin and L. Woessmann (Eds.), *Handbook of Economics of Education*, 3, 249–277, Amsterdam: North Holland.

Staiger, D., & Stock, J.H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65, 557–586.

Stark, O., & Taylor, J.E. (1991). Migration incentives, migration types: the role of relative deprivation. *Economic Journal*, 101, 1163–1178.

Stock, J., Wright, J., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20, 518–529.

Stock, J.H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In: D.W. Andrews and J.H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80–108, New York: Cambridge University Press.

Telle, N.T., & Pfister, H.R. (2016). Positive empathy and prosocial behavior: a neglected link. *Emotion Review*, 8, 154-163.

Table 1. Variable Definitions and Summary Statistics

Variables	Definition	N	Own Characteristics		Peers' Mean Characteristics	
			Mean	SD	Mean	SD
Consumption	Annual total household consumption expenditure (Indian Rupees)	40980	115239	111528	115239	54545
Income	Annual total household income (Indian Rupees)	40980	118130	130239	118130	71653
Household Size	Total number of people in the household	40980	4.85	2.29	4.85	0.91
Age	Age of household head	40980	49.62	13.55	49.62	4.7
Male	= 1 if household head is male; = 0 otherwise	40980	0.86	0.35	0.86	0.11
Literate	= 1 if household head is literate; = 0 otherwise	40980	0.68	0.47	0.68	0.22
Married	= 1 if household head is married; = 0 otherwise	40980	0.81	0.39	0.81	0.12
(Children Proportion)	Number of children/Total number of people in the household	40980	0.24	0.22	0.24	0.09
Teenage Proportion	Number of teens/Total number of people in the household	40980	0.11	0.16	0.11	0.05
Adult Proportion	Number of adults/Total number of people in the household	40980	0.65	0.23	0.65	0.1
(No Married)	= 1 if there are no married people in the household; = 0 otherwise	40980	0.09	0.28	0.09	0.08
One to Five Married	= 1 if there are more than zero and less than five married people in the household; = 0 otherwise	40980	0.87	0.33	0.87	0.09
More than Five Married	= 1 if there are more than five married people in the household; 0 otherwise	40980	0.04	0.19	0.04	0.05
Years in Place	= 1 if the household has been living in the same place for more than 10 years; 0 otherwise	40980	0.97	0.18	0.97	0.08
Urban	= 1 if the household's place of residence is categorized as Urban as per the 2011 Census; 0 otherwise	40980	0.35	0.48	0.35	0.48
Brahmin	= 1 if household head's caste is Brahmin; 0 otherwise	40980	0.05	0.22	0.05	0.11
Forward Caste	= 1 if household head's caste is Non-Brahmin Forward Caste; = 0 otherwise	40980	0.23	0.42	0.23	0.26
OBC	= 1 if household head's caste is Other backward Classes (OBC); = 0 otherwise	40980	0.41	0.49	0.41	0.3
SC	= 1 if household head's caste is Scheduled Caste (SC); = 0 otherwise	40980	0.22	0.41	0.22	0.23
ST	= 1 if household head's caste is Scheduled Tribe (ST); = 0 otherwise	40980	0.09	0.28	0.09	0.21
(Other Caste)	= 1 if household head is a member of some other caste; = 0 otherwise	40980	0.01	0.11	0.01	0.06
Job Loss	= 1 if a household member has lost job in the recent past; = 0 otherwise	40980	0.03	0.16	0.03	0.08
Death	= 1 if there has been a death in the recent past; 0 = otherwise	40980	0.19	0.39	0.19	0.13
Theft	= 1 if there has been an incident of theft in the household in the recent past; = 0 otherwise	40980	0.04	0.19	0.04	0.07
Break-in	= 1 if the household faced an incident of break-in in the recent past; = 0 otherwise	40980	0.01	0.1	0.01	0.03

Notes: The variables in the parentheses are the omitted categories in the following estimations.

Table 2. Peer Effects in Consumption: OLS Estimation

Variables	[1]	[2]	[3]
Endogenous Peer Effect	**0.461***	**0.568***	**0.207***
	(0.013)	**(0.017)**	**(0.026)**
Household Effects	YES	YES	YES
Idiosyncratic Shocks	YES	YES	YES
Contextual Effects	NO	YES	YES
District Fixed Effects	NO	NO	YES
Observations	40980	40980	40980
Adjusted R-squared	0.324	0.329	0.338

Notes: Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. Household effects are controlled by own demographic characteristics of households that include Income, Household Size, Age, Male, Literate, Married, Teenage Proportion, Adult Proportion, One to Five Married, More than Five Married, Years in Place, Urban, Brahmin, Forward Caste, OBC, SC and ST. Idiosyncratic shocks include Job Loss, Death, Theft and Break-in. Contextual effects are controlled by average peer demographic characteristics. For definition of variables see Table 1. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3. Validity of Instruments: Balancing Test

Dependent Variable	Peer Job Loss	Peer Death	Peer Theft	Peer Break-in	F-statistic	Observations
Income	11716	8246	6630	-15261	0.88	40980
	(13460)	(5877)	(12505)	(23283)	[p=0.474]	
Household Size	0.077	0.156	0.082	-0.589	1.21	40980
	(0.215)	(0.103)	(0.217)	(0.401)	[p=0.305]	
Age	-0.478	1.734***	-1.087	-1.417	2.31	40980
	(1.415)	(0.626)	(1.301)	(2.458)	[p=0.055]	
Male	0.0273	0.023	0.080**	-0.156**	2.99	40980
	(0.035)	(0.017)	(0.033)	(0.064)	[p=0.018]	
Literate	0.011	0.023	0.064	0.010	1.13	40980
	(0.045)	(0.020)	(0.043)	(0.077)	[p=0.340]	
Married	0.047	0.035*	0.093	-0.095	3.12	40980
	(0.039)	(0.018)	(0.037)	(0.071)	[p=0.014]	
Children Proportion	0.013	0.006	-0.002	0.049	0.60	40980
	(0.022)	(0.010)	(0.021)	(0.040)	[p=0.663]	
Teenage Proportion	-0.012	-0.012	0.012	-0.014	0.85	40980
	(0.017)	(0.008)	(0.016)	(0.031)	[p=0.492]	
Adult Proportion	-0.000	0.006	-0.011	-0.035	0.45	40980
	(0.024)	(0.010)	(0.021)	(0.041)	[p=0.772]	
No Married	-0.077**	-0.021	-0.036	0.081	3.39	40980
	(0.030)	(0.013)	(0.026)	(0.054)	[p=0.009]	
One to Five Married	0.077**	-0.002	0.051	-0.120*	2.48	40980
	(0.035)	(0.016)	(0.031)	(0.063)	[p=0.042]	
More than Five Married	0.000	0.023**	-0.015	0.0388	1.92	40980
	(0.019)	(0.009)	(0.018)	(0.034)	[p=0.105]	
Years in Place	-0.008	0.006	-0.016	-0.000	0.43	40980
	(0.019)	(0.007)	(0.018)	(0.037)	[p=0.786]	
Brahmin	-0.014	0.002	-0.010	0.027	0.21	40980
	(0.021)	(0.009)	(0.023)	(0.045)	[p=0.934]	
Forward Caste	0.016	-0.007	-0.027	0.046	0.26	40980
	(0.037)	(0.017)	(0.035)	(0.076)	[p=0.903]	
OBC	-0.023	-0.016	0.033	-0.071	0.49	40980
	(0.042)	(0.019)	(0.041)	(0.080)	[p=0.744]	
SC	0.026	0.022	0.005	-0.023	0.60	40980
	(0.039)	(0.017)	(0.036)	(0.071)	[p=0.663]	
ST	-0.015	-0.000	-0.002	0.025	0.32	40980
	(0.018)	(0.008)	(0.019)	(0.034)	[p=0.868]	
Other Caste	0.009	-0.000	0.000	-0.004	0.13	40980
	(0.015)	(0.0051)	(0.008)	(0.016)	[p=0.973]	

Notes: Each row shows results of regression of a specific household demographic characteristic on different types of average peer shocks. Additional controls in each regression include own idiosyncratic shocks, average peer characteristics and district fixed effects. For definition of variables see Table 1. All regressions are estimated by OLS and include a constant. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4. Peer Effects in Consumption: IV Estimation

Variables	[1]	[2]	[3]
Endogenous Peer Effect	**0.782***	**0.848***	**0.697***
	(0.147)	**(0.128)**	**(0.223)**
Household Effects	YES	YES	YES
Idiosyncratic Shocks	YES	YES	YES
Contextual Effects	NO	YES	YES
District Fixed Effects	NO	NO	YES
Observations	40980	40980	40980
Adjusted R-squared	0.305	0.321	0.325
Hansen J statistic	6.096	4.355	1.829
	[p=0.107]	[p=0.226]	[p=0.609]
Kleibergen-Paap rk LM statistic	147.5	339.9	154.8
	[p=0.000]	[p=0.000]	[p=0.000]
First stage F-statistic	38.03	89.97	40.01

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristic, idiosyncratic shocks and definition of variables see Table 1 and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5. Peer Effects in Consumption: Subsample Analysis

Variables	Non-Poor	Poor	Upper Caste	Lower Caste	Urban	Rural
			Subsamples			
Endogenous Peer Effect	0.727***	0.024	0.887**	0.565**	0.873***	0.754***
	(0.240)	(0.079)	(0.360)	(0.240)	(0.304)	(0.227)
Observations	34,168	6,812	11,465	29,515	14,171	26,809
Adjusted R-squared	0.218	0.735	0.190	0.199	0.185	0.195
Hansen J statistic	3.597	7.773	0.416	2.815	1.493	0.766
	[p=0.308]	[p=0.0509]	[p=0.937]	[p=0.421]	[p=0.684]	[p=0.858]
Kleibergen-Paap rk LM statistic	133.6	39.37	68.70	121.2	95.99	191.5
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
First stage F-statistic	34.49	10.61	17.26	32.30	25.23	53.03
z-statistic	2.786		0.744		0.314	

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristic, idiosyncratic shocks, definition of variables and definition of subsamples see Table 1, and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6. Peer Effects in Consumption: Investigating Omitted Variable Bias

Variables	[1]	[2]	[3]	[4]	[5]	[6]	[7]
	Covariate Shock Controls			**Controls for Correlates of Empathy/Reciprocity**		**Local Public Good Controls**	
	Crop Failure	Droughts/Floods	Conflict	Member of Social Group/Festival Society	Member of Youth Club/ Reading Group/Sports Society	Government Schools and Colleges	Government Health Centers, Hospitals and Dispensaries
Endogenous Peer Effect	0.684***	0.681***	0.681***	0.673***	0.662***	0.707***	0.783***
	(0.238)	(0.236)	(0.289)	(0.244)	(0.245)	(0.250)	(0.257)
Observations	40,980	40,980	40,980	40,948	40,951	25,772	25,659
Adjusted R-squared	0.326	0.326	0.332	0.334	0.335	0.3077	0.3048
Hansen J statistic	1.926	1.834	1.997	1.771	1.927	0.420	0.519
	[p=0.588]	[p=0.608]	[p=0.573]	[p=0.621]	[p=0.588]	[p=0.936]	[p= 0.915]
Kleibergen-Paap rk LM statistic	135.6	137.5	129.5	129.44	128.99	164.673	170.685
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
First stage F-statistic	34.90	35.45	33.28	33.25	33.14	44.55	46.56

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Additionally, the regressions reported in columns (1) - (3) include controls for covariate shocks, regressions in columns (4) and (5) include controls for variables that are strong correlates of social emotions like empathy and reciprocity, and the regressions reported in the last two columns include controls for local public goods. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristics, idiosyncratic shocks and definition of variables see Table 1 and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7. Peer Effects in Consumption Based on Alternative Definitions of Reference Groups

Variables	[1]	[2]	[3]
	Village/Neighborhood and Caste	Village/Neighborhood and Migrant Status	Village/Neighborhood, Caste and Migrant Status
Endogenous Peer Effect	0.593***	0.644***	0.583***
	(0.216)	(0.215)	(0.206)
Observations	40,081	39,414	38286
Adjusted R-squared	0.328	0.332	0.3274
Hansen J statistic	3.482	0.915	1.896
	[p=0.323]	[p=0.8219]	[p=0.594]
Kleibergen-Paap rk LM statistic	117.464	146.433	119.297
	[p=0.000]	[p=0.000]	[p=0.000]
First stage F-statistic	29.23	37.58	29.57

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. First column report results of regression when peer reference group is defined by village/neighborhood of residence and caste (i.e. households in the same village and same caste group form a peer reference group). Second column report results of regression when peer reference group is defined by village of residence and migrant status (i.e., households within a village who have not sent out migrants for work form a peer reference group, whereas households within a village who have sent out migrants form a different peer group). Third column report results of regression when peer reference group is defined by village of residence, caste, and migrant status. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristics, idiosyncratic shocks and definition of variables see Table 1 and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
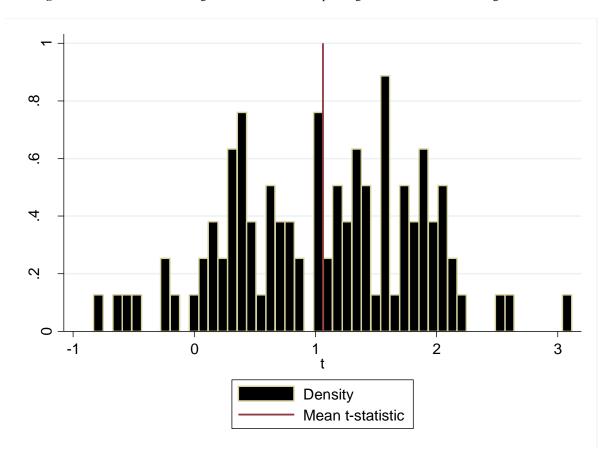
Figure 1. Falsification Test: Histogram of t-statistics corresponding to Point Estimates of Endogenous Peer Effects

Supplementary Appendix

Peer Effects in Consumption: Evidence from India

FOR ONLINE PUBLICATION ONLY

1 Social Multiplier

Endogenous peer effects imply that small exogenous shock at the individual level is magnified through the social interactions process to deliver larger aggregate level social effects. Glaeser and Scheinkman (2001) and Glaeser et al. (2003) define the social multiplier as the ratio of the individual effect from an *exogenous* shock to the aggregate effect from the same shock. In what follows, I derive the social multiplier for my baseline econometric model:

$$c_{i,r} = \alpha + \beta \bar{c}_{-i,r} + \gamma y_{i,r} + \mathbf{x}'_{i,r} \boldsymbol{\varpi} + \bar{\mathbf{x}}'_{-i,r} \boldsymbol{\delta} + \theta \bar{y}_{-i,r} + \mathbf{z}'_{i,r} \boldsymbol{\rho} + \xi_{i,r} \tag{1}$$

Let

$$\Theta_{i,r} = \alpha + \gamma y_{i,r} + \mathbf{x}'_{i,r} \boldsymbol{\varpi} + \bar{\mathbf{x}}'_{-i,r} \boldsymbol{\delta} + \theta \bar{y}_{-i,r} + \mathbf{z}'_{i,r} \boldsymbol{\rho} + \xi_{i,r} \tag{2}$$

Thus, equation (1) can be rewritten as

$$
\begin{aligned}
c_{i,r} &= \Theta_{i,r} + \beta \bar{c}_{-i,r} \\
&= \Theta_{i,r} + \frac{\beta}{m_r - 1} \sum_{j \in M_{i,r}} c_{j,r}
\end{aligned}
\tag{3}
$$

where $M_{i,r}$ denotes peer group of household i and $(m_r - 1)$ denotes size of the peer group.

Based on equation (3), it can show that

$$\frac{1}{m_r} \sum_{j \in M_r} c_{j,r} = \frac{1}{1-\beta} \frac{1}{m_r} \sum_{j \in M_r} \Theta_{j,r} \tag{4}$$

$$\text{or, } \bar{c}_r = \frac{1}{1-\beta} \bar{\Theta}_r \tag{5}$$

where M_r denotes the social group to which household i belongs to. This is the aggregate level equation.

And also we can show that in equilibrium,

$$c_{i,r} = \Theta_{i,r} \left(1 + \frac{\beta^2}{(1-\beta)(m_r - 1 + \beta)} \right) + \frac{\beta}{m_r - 1 + \beta} \sum_{j \in M_{i,r}} \Theta_{j,r} \tag{6}$$

Equation (6) is the individual/household level equation. To simplify the analysis, I consider a three-person social group (i.e., $m_r = 3$). Let the individuals be indexed by $i = 1, 2, 3$. Further, I assume $\gamma = \theta = \rho = 0$ and there is a unique exogenous characteristic x that affects household consumption. Thus I have:

$$c_{1,r} = \Theta_{1,r} + \beta\frac{c_{2,r} + c_{3,r}}{2} \tag{7}$$

$$c_{2,r} = \Theta_{2,r} + \beta\frac{c_{1,r} + c_{3,r}}{2} \tag{8}$$

$$c_{3,r} = \Theta_{3,r} + \beta\frac{c_{1,r} + c_{2,r}}{2} \tag{9}$$

$$\Theta_{1,r} = \varpi x_{1,r} + \delta\frac{x_{2,r} + x_{3,r}}{2} + \xi_{1,r} \tag{10}$$

$$\Theta_{2,r} = \varpi x_{2,r} + \delta\frac{x_{1,r} + x_{3,r}}{2} + \xi_{2,r} \tag{11}$$

$$\Theta_{3,r} = \varpi x_{3,r} + \delta\frac{x_{1,r} + x_{2,r}}{2} + \xi_{3,r} \tag{12}$$

In this three-person social group setup, equation (4) yields

$$
\begin{aligned}
\frac{1}{m_r}\sum_{j\in M_r} c_{j,r} &= \frac{1}{1-\beta}\frac{1}{3}\left[\left(\varpi x_{1,r} + \delta\frac{x_{2,r} + x_{3,r}}{2} + \xi_{1,r}\right) + \left(\varpi x_{2,r} + \delta\frac{x_{1,r} + x_{3,r}}{2} + \xi_{2,r}\right)\right.\\
&\quad \left. + \left(\varpi x_{3,r} + \delta\frac{x_{1,r} + x_{2,r}}{2} + \xi_{3,r}\right)\right]\\
&= \frac{1}{1-\beta}\left[\varpi\frac{(x_{1,r} + x_{2,r} + x_{3,r})}{3} + \delta\frac{(x_{1,r} + x_{2,r} + x_{3,r})}{3} + \frac{(\xi_{1,r} + \xi_{2,r} + \xi_{3,r})}{3}\right] \tag{13}\\
or, \ \bar{c}_r &= \frac{1}{1-\beta}\left[(\varpi + \delta)\bar{x}_r + \bar{\xi}_r\right] \tag{14}
\end{aligned}
$$

and equation (6) yields

$$
\begin{aligned}
c_{1,r} &= \left(\varpi x_{1,r} + \delta\frac{x_{2,r} + x_{3,r}}{2} + \xi_{1,r}\right)\left(1 + \frac{\beta^2}{(1-\beta)(m_r - 1 + \beta)}\right)\\
&\quad + \frac{\beta}{m_r - 1 + \beta}\left[\left(\varpi x_{2,r} + \delta\frac{x_{1,r} + x_{3,r}}{2} + \xi_{2,r}\right) + \left(\varpi x_{3,r} + \delta\frac{x_{1,r} + x_{2,r}}{2} + \xi_{3,r}\right)\right] \tag{15}
\end{aligned}
$$

Thus, from the aggregate level equation (14), a change in $\bar{c}_r$ due to a change in $\bar{\xi}_r$ is given by

$$\frac{\partial\bar{c}_r}{\partial\bar{\xi}_r} = \frac{1}{1-\beta} \tag{16}$$

and from the individual level equation (15), a change in $c_{1,r}$ due to a change in $\xi_{1,r}$ is given by

$$\frac{\partial c_{1,r}}{\partial \xi_{1,r}} = \left(1 + \frac{\beta^2}{(1-\beta)(m_r - 1 + \beta)}\right) \tag{17}$$

For any household i, thus, I get:

$$\frac{\partial c_{i,r}}{\partial \xi_{i,r}} = \left(1 + \frac{\beta^2}{(1-\beta)(m_r - 1 + \beta)}\right) \tag{18}$$

Hence, the social multiplier Δ is given by:

$$\Delta = \frac{\frac{\partial \bar{c}_r}{\partial \bar{\xi}_r}}{\frac{\partial c_{i,r}}{\partial \xi_{i,r}}} = \frac{\frac{1}{1-\beta}}{\left(1 + \frac{\beta^2}{(1-\beta)(m_r - 1 + \beta)}\right)} \tag{19}$$

For sufficiently large social groups $(m_r \longrightarrow \infty)$:

$$\Delta = \frac{1}{1 - \beta} \tag{20}$$

2   Additional Robustness Checks

In this section, I conduct various robustness checks to assess the robustness of my estimates to different sorts of bias.

Using Consumption and Income in Logs. As noted by Suri (2004), a question that arises in the empirical work relating to consumption is whether to look at logs or levels of consumption, since in theory, consumption is assumed to follow a log normal distribution (Battistin et al., 2009). Similar question also arises in context of income. Although, given that I have a large enough sample, using consumption and income in logs is not necessary, it is worthwhile to check whether there is any qualitative change in my baseline results when I take a log-transformation of the variables in question.

I present the results in column (1) of Table A4. Reassuringly, I find that the results remain qualitatively unaltered.

Measurement Error in Income. Measurement error in household income is always a concern in survey data. To show that this issue does not influence my main results, I re-estimate my baseline econometric model now instrumenting household's own income by literacy status of household head's father (or household head's husband's father, in case the household head is a woman). I also instrument average peer income by average literacy status of peer household head's father (or peer household head's husband's father). The assumption here is that whether a household head's father was literate or not affects earnings of the household, but

whether a peer household head's father was literate or not has no direct influence on the household's own earnings. Column (2) of Table A4 reports the results from this exercise. As it turns out, the point estimate of endogenous peer effect remains almost unchanged compared to the baseline results.

Exclusion of IVs. One potential critique of my IV strategy is that peer idiosyncratic shocks may have direct impact on own consumption. If this is true, the exclusion restriction for the instruments to work well is violated. To examine this, I report results of additional specifications in Table A5.

In each of the specification, I use two out of four instruments to instrument the endogenous peer effect and include the remaining two instruments as additional regressors in the second stage. I do this for all possible combinations of my instruments (hence, I report results from six specifications). I then examine whether the coefficients of instrument included as regressors in the second stage are statistically significant or not. I find that, across all the specifications, the coefficients of instruments in question are statistically indistinguishable from zero. This implies that, conditional on *any* two instruments being valid, the other instruments are excluded from the second stage. In other words, if one is willing to believe that only two instruments are valid (and is not sure about the validity of the remaining ones), this exercise shows that conditional on that belief, the remaining instruments are likely to be valid as well. This increases my confidence in the overall set of instruments used to identify my baseline model.

'Type' of Idiosyncratic Shock matters? Note that the instruments used is in this study are all constructed based upon idiosyncratic shocks that are 'negative' and 'accidental' in nature (or due to bad luck). The consumption expenditure adjustment due to status seeking behavior may be very different when a neighbor wins a lottery (compared to the case when there is a death in the family of my neighbors) for instance. However, testing this empirically is difficult in the present context, because all the questions that are specifically on expenditure shocks in the IHDS are essentially negative and accidental in nature. However, there is one question in the survey questionnaire which could be uses to address this concern. This question is regarding the sources of household income. Specifically, there is question in the IHDS that asks households whether a part of their earnings come from lottery, dowry or chit fund. I create a dummy variable that takes a value equal to 1 if the household has received some earnings from lottery, dowry or chit funds, and zero otherwise. I treat this as a shock to household income (and expenditure) that is due to the household trying to be 'greedy'. The nature of this shock is likely to be very different from that of the shocks that I have used in my baseline analysis. Based on this shock, I now create an additional instrument (the procedure of constructing the instrument remains the same as before, i.e., I take the average of the shocks over all the households in my reference group, leaving out myself). I, then, use this average peer 'lottery/chit fund/dowry' shock as an instrument for average peer consumption. I report two specifications. First, in which I use only this

instrument. Second, in which I use this new instrument along with the other instruments that I had already been using. The results are reported in Table A6 (however, note that I now use state fixed effects, since within a district the variation in the 'lottery/chit fund/dowry' shock is extremely low). It turns out the estimate of the endogenous peer effect falls slightly (compared to the baseline estimates) when only the new instrument is used, and remains almost unchanged when the new instrument is used along with the other instruments. This is reassuring, and provides at least some suggestive evidence that the specific 'type' of shock is unlikely to be driving my estimates of the endogenous peer effects.

Excluding (Relatively) Recent Migrant-Households. As noted in the main text, household-level migration is remarkably low in India. Consequently, selection of households into villages/neighborhoods is unlikely to result in biased coefficients of endogenous peer effects. However, to provide further evidence that selection into villages or neighborhoods is unlikely to be an issue of concern, I exclude the households who report to have migrated in the last 10 years from the sample, and re-estimate the baseline model. The results are reported in Table A7. As evident, the estimate of the endogenous peer effect remains almost altered compared to the baseline results. Specifically, the coefficient of endogenous peer effect now is 0.73, whereas in the baseline case it was 0.69. As an additional sensitivity check, I delete all those households who report to have been migrated in the last 20 years, and I repeat the same exercise. Again, the point estimates are similar to that obtained from the baseline regression. Thus, my baseline estimate of the endogenous peer effect is unlikely to be inconsistent due to selection of households into villages/neighborhoods.

Using Prices as Additional Controls. Next, I examine whether my results are driven by prices of goods that are consumed/purchased by the households in my sample. Although, IHDS does not have information about prices at the household level of all goods and services, it has information on prices of most (specifically 11) non-tradable goods, which according to the FAO, are goods that are mainly sold locally and produced in rural villages (rather than in urban cities).[1] These goods are basically staple goods including rice, wheat, pulses, sugar, etc. Households are asked the market price of these 11 staple goods, and then are asked whether they have purchased these goods or whether the goods that they use are homegrown. I create a household level price index based on this information. Specifically, for a particular good, I multiply the market price of the good (which is reported by the household) by a dummy taking a value of one if the household has purchased that particular good, and then repeat this for all the 11 goods, and finally take the average of these. Although, this is a crude measure of prices that the households face, I use this to check the sensitivity of my baseline results by including this as an additional control. Note, in this specification, I also control for the local price level. I define this as the average level of prices computed

---

[1]http://www.fao.org/docrep/w9500e/w9500e12.htm

at the village/neighborhood level that I calculate based on the prices reported by the households living in a particular village/neighborhood (excluding the focal household). I also report results of an additional specification: I deflate total consumption expenditure by the price index that I have computed, and estimate the baseline regression model using deflated (own and peer) consumption expenditure. Results are reported in Table A8. Reassuringly, the estimate of the endogenous peer effect, across both the specifications, roughly remains unchanged compared to initial estimate of the endogenous peer effect obtained from the baseline regression model (i.e., the endogenous peer effect remains close to 0.7, which is what it was in the baseline case).

Issue of Individual-level Migration. As noted in the main text, it may be the case that households who send out migrants (who send remittances back home) may be very different in their wealth levels from the households who do not, and hence they may not serve as a reference peer group. To examine this issue, I compute the summary statistics relating to migration from the IHDS data. Specifically, the IHDS asks household heads: "Have you or any member of your household left to find seasonal/short term work during last five years?" It turns out that out of 41,000 (approximately) households surveyed by the IHDS, only 2,600 households answer "yes" to this question (of which 2,343 are rural households). Since these households can potentially be very different and not serve as a reference peer group, I drop these households from my analytical sample (under the assumption that this sample of households come from a different data generating process (DGP)), and re-estimate the baseline regression model. Results are reported in column (1) of Table A9. I find that even after putting this additional cut on the sample, the IV estimates of the endogenous peer effect remain statistically and economically significant. The magnitude of the estimate is also remarkably similar to the baseline results.[2]

Using Sampling Weights. Survey weights are often a critical part of representativeness of a survey. However, as noted by Gelman (2007), Kott (2007), Winship and Radbill (1994), and Young and Johnson (2012), there is little consensus on whether weights should be routinely used in multivariate models, such as regression. In general, the literature concludes that including weights in a regression model is likely to make the estimates less efficient. Consequently, it is often suggested that, instead of using weights, adopting a *model-based method* might be preferable. A model-based strategy does not use the weights but includes the variables used to construct the weights as variables in the regression model as regressor. As noted by Young and Johnson (2012), this will provide unbiased and consistent parameter estimates even without explicitly using the sample weights. A similar recommendation is also made by Arellano (2014). He recommends that,

---

[2]In addition to this, I also estimate a specification where I do not exclude households that send out migrants, but instead include a dummy regressor in the right-hand side that takes a value of 1 if the household answers "yes" to the migration question. Results remain unaltered. I do not report this regression; however the results are available upon request.

when using data from surveys that are done using stratified random sampling, there is no need to use sample weights in estimation if stratification is based exclusively on the variables that are included as controls in the regression (call them x) (i.e. the stratification variables are functions of x). The standard unweighted estimator remains consistent in the stratified sample.

Now, it is often the case that one does not have full knowledge about how weights are constructed or how stratification is done. As such, even if one uses a sufficiently large number of relevant covariates as right-hand side variables, one can never sure whether he/she has included all the relevant variables based on which stratification is done. In that case, to decide whether one should use weights or not, it is recommended that one should carry out a sensitivity analysis and compare the coefficients of interest from analysis with and without weights. As noted by Young and Johnson (2012), if the coefficients do not significantly differ, then weights are unnecessary.

In my baseline analysis, since I have a large set of control variables including individual level controls, village/neighborhood level controls, plus district fixed effects, I have not used survey weights under the assumption that stratification variables (or the weights) are likely to functions of the controls included in my model (in other words, I have followed the model-based approach). However, as noted previously, the validity of such an assumption is questionable when one does not have perfect knowledge about how exactly weights are created. As such, to check whether my assumption is likely to be valid, and whether my unweighted parameter estimates are consistent, I follow the general recommendation in the literature and check the sensitivity of my baseline results to inclusion of survey weights provided by the IHDS. Results are reported in column (2) of Table A9. I find that the point estimates of the endogenous peer effects remain statistically and economically significant as before, and that it is not significantly different from the baseline estimate (based on the results of the z-test (Paternoster, 1998)). As such, I conclude that my baseline unweighted estimates of the endogenous peer effects are consistent.

3   Extension: Disaggregate Analysis

My data distinguishes between several different categories of consumption expenditure. This allows me to investigate more specifically the composition of the increase in own consumption expenditure in response to an increase in average peer consumption. I report the estimation results in Table A10.

I find that an increase in average peer consumption has a huge impact on own expenditure on temptation goods,[3] and a moderate impact on own food expenditure. More specifically, a 1 Indian Rupee increase in average peer consumption leads to a 0.48 Indian Rupee increase in own expenditure on temptation goods

---

[3]Expenditure on temptation goods include spending on entertainment, vacation, jewelry, clothing/bedding, footwear, etc. See Table A1 for the full list of temptation goods.

and 0.13 Indian Rupee increase in own expenditure on food items. These effects are statistically as well as economically significant. The impacts of an increase in average peer consumption on own health expenditure and own education expenditure, although positive, are much smaller compared to what I have found for the cases of temptation goods and food. Moreover, these effects are also imprecisely estimated. Overall, the results of the disaggregate analysis suggests that the increase in own consumption due to endogenous peer effect is driven by increase in own expenditure on temptation goods and food.[4]

4   Alternative Mechanism: Can the Results be Explained by Risk Sharing?

An alternative mechanism driving my results could be partial risk-sharing against idiosyncratic income shocks (Townsend, 1994).[5] According to the partial risk sharing hypothesis, own consumption co-moves with average peer consumption, albeit imperfectly. This means that the endogenous peer effects obtained in this paper may be due to partial risk sharing rather than due to status seeking behavior. So, in order to establish the validity of status seeking behavior as the plausible mechanism driving my findings, it becomes critical to rule out partial risk sharing (although it is apparent that distinguishing status seeking behavior from partial risk sharing is extremely difficult in practice since the predictions of the two models are virtually identical).

One way to assess (although somewhat adhoc) whether the model of partial risk sharing can explain my results is to look at the results of the heterogeneity/subsample analysis. The central idea of the analysis is as follows. According to Townsend (1994), mutual coinsurance is likely to be more relevant for people living in a high risk environment with (almost) no formal risk reduction mechanism. As such, it is likely to be the case that, under partial risk sharing, the co-movement in consumption is likely to be greater for the people who are economically more vulnerable and who are exposed to high degrees of risk compared to the people who are economically less vulnerable and who are exposed to low degrees of risk (in fact, the risk sharing model may actually be irrelevant for the latter group of people). If I assume people who are economically more vulnerable and live in a high risk environment are more likely to be living in rural areas than in urban areas, are members of low caste than of high caste, and/or live below the poverty line, then

---

[4]The disaggregate analysis is subject to a caveat. Note that my baseline model given by Equation (9) implies that if I use a expenditure on a certain type of goods as my dependent variable, expenditure on all other consumption goods ends up in the error term on the right hand side (unless they are explicitly controlled for). This means that the error term becomes correlated with all the regressors in the model. Now, if my instruments are correlated with some of these regressors in addition to being correlated with average peer consumption, then the estimate of the coefficient of average peer consumption will no longer be consistent.

[5]Partial risk sharing implies that individuals coinsure income fluctuations 'imperfectly' (i.e., they pool only a part of their income). Instead, the individuals could be 'perfectly' coinsuring income fluctuations. However, as noted by Schulhofer-Wohl (2011), full risk sharing has almost always been rejected in data.

it should be the case that the magnitude of the co-movement in consumption (in other words, the estimate of the endogenous peer effect) is greater for those living in rural areas, are members of low caste than of high caste, and living below the poverty line than their counterparts. This, however, is not what I find in my heterogeneity analysis reported in the main paper (Table 6). Specifically, for all the cases I either find that the difference between the estimated endogenous peer effects for two subsamples is not statistically significant or that the estimate of the endogenous effect is significantly lower for the subsample which is economically more vulnerable. This provides suggestive evidence that my baseline results are unlikely to be explained by the mechanism of partial risk sharing.[6]

References

Arellano, M. (2014). *Econometrics of Survey Data*. Lecture Notes, CMFI. Available at http://www.cemfi.es/~arellano/survey-metrics-sep2014-tr.pdf

Battistin, E., Blundell, R., & Lewbel, A. (2009). Why is consumption more log normal than income? Gibrat's law revisited. *Journal of Political Economy*, 117, 1140-1154.

Gelman, A. (2007). Struggles with Survey Weights and Regression Modeling. *Statistical Science*, 22, 153-164.

Glaeser, E. L., & Scheinkman, J. (2001). Measuring social interactions. In: S. Durlauf and H. P. Young (Eds.), *Social Dynamics*, 83–131, Cambridge: MIT Press.

Glaeser, E. L., Scheinkman, J. A., & Sacerdote, B. I. (2003). The Social Multiplier. *Journal of the European Economic Association*, 1, 345–353.

Kott, P.S. (2007). Clarifying some Issues in the Regression Analysis of Survey Data. *Survey Research Methods*, 1, 11-18.

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36, 859–866.

Schulhofer-Wohl, S. (2011). Heterogeneity and tests of risk sharing. *Journal of Political Economy*, 119, 925–958

Suri, T. (2004). *Spillovers in Village Consumption: Testing the Extent of Partial Insurance.* Working Paper, Sloan School of Management, Massachusetts Institute of Technology. Available at http://economie.esg.uqam.ca/upload/files/seminaires/suri_spillovers.pdf

Townsend, R. M. (1994). Risk and insurance in village India. *Econometrica*, 62, 539–591.

Young, R., & Johnson, D.R. (2012). To Weight, or Not to Weight, That is the Question: Survey Weights and Multivariate Analysis. The American Association for Public Opinion Research (AAPOR) 67th Annual Conference, 2012. Available at http://www.aapor.org/AAPOR_Main/media/AnnualMeetingProceedings/2012/03_-Young-Johnson_A2_Weighting-paper_aapor-2012-ry.pdf

Winship, C., & Radbill, L. (1994). Sampling Weights and Regression Analysis. *Sociological Methods & Research*, 23, 230-257.

---

[6]In Tables A11 and A12, I provide additional evidence that partial risk sharing is unlikely to be driving my baseline results. Specifically, I show that the estimated endogenous peer effect is not significantly higher for those households who belong to relatively more vulnerable social groups compared to those who belong to relatively less vulnerable social groups. I also show that the estimated endogenous peer effect is higher for those households belonging to more heterogenous social groups compared to those who belong to less heterogenous social groups. These findings are exactly opposite to what one would expect to find if the mechanism driving my baseline results is partial risk sharing.

Table A1. Categories of Consumption: IHDS 2012

Food	Vegetables	Soap/detergent	Therapeutic appliances
Rice	Salts/spices	Conveyance	Personal care
Wheat	Tea/Coffee	Diesel/Petrol/CNG	Other personal expenditure
Sugar	Processed food	House/other rent/ loans	Repair/maintenance
Kerosene	Paan/Tobacco	Consumer tax/fees	Insurance premiums
Other Cereal	Fruits/Nuts	Services/servants	Vacations
Pulses	**Temptation Good**	Clothing/bedding	Social Functions
Meat	Eating out	Footwear	**Health**
Sweeteners	Fuel	Furnitures/fixtures	Medical out-patient
Edible Oil	Light	Crockery/utensils	Medical in-patient
Eggs	Entertainment	Household appliances	**Education**
Milk	Telephone	Recreation goods	School/college fees
Milk products	Cosmetics/Toiletries	Jewelry	Private tuition
Cereal Products	Household items	Transport equipment	School books

Table A2. Peer Effects in Consumption: OLS Estimation (Full Results)

Variables	[1]	[2]	[3]
Endogenous Peer Effect	**0.461***	**0.568***	**0.208***
	(0.0132)	**(0.0173)**	**(0.0256)**
Household Effects			
Income	0.283***	0.291***	0.291***
	(0.00838)	(0.00892)	(0.00893)
Household Size	8327***	8564***	8414***
	(352.4)	(357.5)	(355.7)
Age	60.11	83.41**	104.9**
	(40.62)	(40.94)	(41.33)
Male	-4531**	-4776**	-5317**
	(2156)	(2142)	(2147)
Literate	15047***	16211***	17011***
	(901.6)	(929.8)	(946.0)
Married	-1380	-1704	-2080
	(2329)	(2335)	(2341)
Teenage Proportion	53137***	52975***	52022***
	(3274)	(3264)	(3250)
Adult Proportion	34547***	33183***	31926***
	(3169)	(3161)	(3185)
Zero to Five Married	20652***	20505***	21472***
	(2127)	(2109)	(2124)
More than Five Married	34309***	33236***	33475***
	(4677)	(4657)	(4566)
Years in Place	-6196**	-3461	-3497
	(2767)	(2908)	(2928)
Urban	-48.03	1997	5144***
	(1191)	(1230)	(1713)
Brahmin	-11827*	-9402	-10226*
	(6206)	(6207)	(6133)
Forward Caste	-13286**	-6932	-7412
	(5823)	(5786)	(5697)
OBC	-17472***	-15945***	-16533***
	(5751)	(5728)	(5656)
SC	-29420***	-29953***	-30424***
	(5748)	(5682)	(5604)
ST	-28145***	-27050***	-27395***
	(5827)	(5925)	(5870)
Individual Idiosyncratic Shocks			
Job Loss	1035	1292	4282
	(2538)	(2552)	(2829)
Death	4767***	4579***	4015***
	(1174)	(1168)	(1171)
Theft	14865***	14721***	14807***
	(2741)	(2748)	(2789)

Table A2. Peer Effects in Consumption, OLS Full Results *(Continued)*

Variables	[1]	[2]	[3]
Break-in	130.0	549.5	923.7
	(5869)	(5837)	(5835)
Contextual Effects			
Income		-0.0953***	0.00343
		(0.0131)	(0.0154)
Household Size		-3805***	-3163***
		(978.5)	(1,184)
Age		-254.3*	119.4
		(149.9)	(162.1)
Male		-159.4	-10,961
		(8420)	(9229)
Literate		-12699***	7032*
		(2734)	(3704)
Married		864.4	-3,242
		(8066)	(9088)
Teenage Proportion		-5659	-3478
		(11852)	(12677)
Adult Proportion		7392	-1414
		(10709)	(12168)
Zero to Five Married		-9865	13214
		(8496)	(9302)
More than Five Married		-23304*	-6461
		(12862)	(14629)
Years in Place		-23773***	-25290***
		(7872)	(9472)
Brahmin		2890	-11266
		(11421)	(14217)
Forward Caste		-10770	-19472
		(10609)	(12849)
OBC		-1717	-15194
		(10398)	(12771)
SC		8984	-7720
		(10404)	(12553)
ST		628.2	-12260
		(10460)	(12583)
Constant	-43477***	15876	8212
	(7009)	(15756)	(19702)
District Fixed Effects	NO	NO	YES
Observations	40980	40980	40980
Adjusted R-squared	0.324	0.329	0.338

Notes: Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. For definition of variables see Table 1. Heteroscedasticity robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A3. Peer Effects in Consumption: IV Estimation (First Stage and Second Stage Full Results)

Variables	[1]	[2]	[3]	[4]	[5]	[6]
Endogenous Peer Effect		0.782***		0.848***		0.697***
		(0.147)		(0.128)		(0.223)
Household Effects						
Income	0.102***	0.250***	0.0109***	0.289***	0.00547***	0.289***
	(0.00261)	(0.0166)	(0.00207)	(0.00885)	(0.00161)	(0.00887)
Household Size	-125.1	8,422***	130.2	8,528***	-112.1	8,453***
	(154.4)	(353.6)	(116.8)	(356.0)	(88.40)	(359.5)
Age	38.57*	49.91	-29.21*	92.31**	11.05	101.2**
	(21.63)	(42.45)	(16.41)	(40.79)	(12.08)	(41.52)
Male	-4000***	-3285	-474.3	-4802**	-952.1	-4860**
	(1167)	(2230)	(885.6)	(2152)	(679.7)	(2158)
Literate	5346***	13357***	-550.6	16381***	743.8**	16628***
	(514.3)	(1220)	(402.3)	(934.2)	(298.8)	(956.5)
Married	1209	-1530	409.8	-1696	-148.4	-1985
	(1148)	(2349)	(851.2)	(2335)	(650.9)	(2341)
Teenage Proportion	6320***	50861***	3235***	51501***	608.1	51336***
	(1607)	(3405)	(1212)	(3273)	(902.3)	(3201)
Adult Proportion	6138***	32663***	3631***	31987***	413.5	31609***
	(1585)	(3241)	(1215)	(3164)	(905.6)	(3190)
Zero to Five Married	-206.3	20181***	449.5	20124***	1465**	20742***
	(1140)	(2155)	(878.3)	(2116)	(669.4)	(2157)
More than Five Married	-5906***	36090***	-576.2	33230***	-15.54	33315***
	(1960)	(4744)	(1500)	(4666)	(1122)	(4586)
Years in Place	-18049***	-1310	-4382***	-3122	-2360**	-2901
	(1652)	(3906)	(1245)	(2957)	(1004)	(2939)
Urban	34910***	-11428**	4710***	633.3	6347***	2105
	(5639)	(5197)	(492.0)	(1334)	(529.2)	(2175)
Brahmin	4327*	-13135**	-500.9	-9021	-1165	-9415
	(2564)	(6389)	(2204)	(6255)	(1645)	(6196)
Forward Caste	-54.74	-13166**	-2462	-6233	-1824	-6561
	(2288)	(5919)	(2000)	(5815)	(1480)	(5766)
OBC	-9982***	-14373**	-1799	-15603***	-1591	-15789***
	(2246)	(5951)	(1969)	(5771)	(1468)	(5722)
SC	-8592***	-26630***	-1439	-29471***	-1256	-29738***
	(2271)	(5914)	(1986)	(5724)	(1475)	(5670)
ST	-22457***	-20986***	-2375	-26743***	-1589	-26562***
	(2356)	(6596)	(2093)	(5979)	(1540)	(5914)
Individual Idiosyncratic Shocks						
Job Loss	-81.09	-83.47	-1,172	1,532	1,224	3,254
	(1608)	(2546)	(1196)	(2567)	(890.1)	(2905)
Death	1875***	4036***	1605***	3993***	441.5	3713***
	(634.7)	(1253)	(480.3)	(1218)	(358.0)	(1187)

Table A3. IV Full Results *(Continued)*

Variables	[1]	[2]	[3]	[4]	[5]	[6]
Theft	-1181	15037***	60.00	14329***	-342.7	14914***
	(1288)	(2758)	(931.4)	(2762)	(680.8)	(2801)
Break-in	5926**	-2761	3307*	-1328	2580	-1006
	(2711)	(6053)	(1997)	(5898)	(1573)	(5877)
Contextual Effects						
Income			0.445***	-0.221***	0.369***	-0.179**
			(0.00500)	(0.0576)	(0.00539)	(0.0832)
Household Size			10608***	-6887***	6699***	-6436***
			(410.7)	(1637)	(410.1)	(1829)
Age			-357.2***	-167.3	250.4***	-8.880
			(60.46)	(157.0)	(51.40)	(176.3)
Male			-14157***	3879	-21046***	-179.0
			(3387)	(8621)	(3121)	(10325)
Literate			9868***	-15997***	30500***	-8574
			(1099)	(3034)	(1091)	(7942)
Married			7200**	1137	-1857	-866.5
			(3447)	(8066)	(3116)	(9133)
Teenage Proportion			100107***	-30960*	58047***	-30183*
			(4506)	(17768)	(3878)	(17659)
Adult Proportion			85786***	-14571	35349***	-17559
			(4322)	(15082)	(4014)	(13588)
Zero to Five Married			27029***	-17062*	42488***	-7815
			(3402)	(9074)	(3070)	(13101)
More than Five Married			24316***	-29606**	32488***	-23375
			(5930)	(13234)	(5126)	(16598)
Years in Place			-60641***	-7443	-33491***	-9751
			(3207)	(11236)	(3481)	(12605)
Brahmin			-15528***	6391	-24836***	1302
			(4986)	(11525)	(4331)	(14823)
Forward Caste			-39230***	-79.35	-30292***	-3488
			(4647)	(11448)	(3949)	(13967)
OBC			-39395***	8925	-36399***	3759
			(4592)	(11248)	(3872)	(14405)
SC			-47856***	21773*	-45272***	15216
			(4613)	(11668)	(3878)	(15240)
ST			-58802***	17160	-46567***	11444
			(4664)	(12244)	(3921)	(15611)
Peer Idiosyncratic Shocks						
Job Loss	3080		-10012***		24826***	
	(2974)		(2169)		(3059)	
Death	20410***		26768***		8794***	
	(2180)		(1680)		(1553)	

Table A3. IV Full Results *(Continued)*

Variables	[1]	[2]	[3]	[4]	[5]	[6]
Theft	4125		14160***		8228***	
	(3942)		(2898)		(2821)	
Break-in	70519***		51949***		41407***	
	(8989)		(6379)		(6093)	
Constant	104204***	-77597***	32269***	5243	3982	2296
	(3,185)	(17188)	(6497)	(16488)	(6219)	(20185)
District Fixed Effects	NO	NO	NO	NO	YES	YES
Observations	40980	40980	40980	40980	40980	40980
Adjusted R-squared	0.259	0.305	0.576	0.321	0.766	0.325
Hansen J statistic		6.096		4.355		1.829
		[p=0.107]		[p=0.226]		[p=0.609]
Kleibergen-Paap rk LM statistic		147.5		339.9		154.8
		[p=0.000]		[p=0.000]		[p=0.000]
First stage F-statistic		38.03		89.97		40.01

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. For full list of demographic characteristic, idiosyncratic shocks and definition of variables see Table 1 and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A4. Peer Effects in Consumption: Consumption in Logs and Endogenous Income

Variables	[1]	[2]
Endogenous Peer Effect	0.639***	0.634***
	(0.157)	(0.241)
Observations	40980	40980
Adjusted R-squared	0.568	0.155
Hansen J statistic	1.454	1.761
	[p=0.693]	[p=0.623]
Kleibergen-Paap rk LM statistic	264.9	134.4
	[p=0.000]	[p=0.000]
First stage F-statistic		
Mean Peer Consumption	71.29	182.97
Household Income		58.18
Mean Peer Income		135.81

Notes: Estimation via two-step GMM. For the specification reported in column [1], the dependent variable is log of Household Annual Total Consumption Expenditure and the Endogenous Peer Effect represents the estimated coefficient of log of Average Peer Household Annual Total Consumption Expenditure. For the specification reported in column [2], the dependent variable is Household Annual Total Consumption Expenditure and the Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. Both specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. Own and Average Peer Income are instrumented in the specification reported in column (2). For full list of demographic characteristics, idiosyncratic shocks, instruments and definition of variables see Table 1, note below Table 2 and main text. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A5. Peer Effects in Consumption: Investigating IV Exclusion

Variables	[1]	[2]	[3]	[4]	[5]	[6]
Endogenous Peer Effect	**0.667****	**0.921*****	**0.445**	**0.879*****	**0.551**	**0.769****
	(0.329)	**(0.348)**	**(0.493)**	**(0.287)**	**(0.356)**	**(0.317)**
Peer Idiosyncratic shocks (Instruments)						
Job Loss	**1257**	**-3965**	**7170**			
	(13307)	**(13541)**	**(16503)**			
Death	**643.2**			**-1139**	**1973**	
	(5566)			**(5579)**	**(6033)**	
Theft		**-12946**		**-12609**		**-12230**
		(10010)		**(9876)**		**(9546)**
Break-in			**16226**		**12361**	**9378**
			(26993)		**(23571)**	**(21086)**
Observations	40980	40980	40980	40980	40980	40980
Adjusted R-squared	0.327	0.310	0.335	0.313	0.332	0.321
Hansen J statistic	1.821	0.109	1.493	0.153	1.568	3.25e-05
	[p=0.177]	[p=0.742]	[p=0.222]	[p=0.695]	[p=0.211]	[p=0.995]
Kleibergen-Paap rk LM statistic	69.69	66.46	42.22	101.8	80.03	91.62
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
First stage F-statistic	36.11	33.92	20.94	52.34	40.67	47.80

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristic, idiosyncratic shocks and definition of variables see Table 1 and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A6. Peer Effects in Consumption: Alternative Instrument Based on 'Positive' Peer Idiosyncratic Shock

Variables	[1]	[2]
	Instrument based on a binary variable indicating whether peer household faced a positive income shock (due to lottery, dowry or chit fund)	Positive income shock instrument + baseline instruments
Endogenous Peer Effect	0.601***	0.705***
	(0.229)	(0.114)
Observations	40,980	40,980
Adjusted R-squared	0.333	0.329
Hansen J statistic		1.603
		[p= 0.808]
Kleibergen-Paap rk LM statistic	167.625	558.615
	[p=0.000]	[p=0.000]
First stage F-statistic	151.81	113.82

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. First column report results of regression where only one instrument - based on a binary variable indicating whether peer household faced a positive income shock (due to lottery, dowry or chit fund) - is used. Second column report results of regression where all baseline instrumental variables are used along with the new instrumental variable. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and state fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristics, idiosyncratic shocks and definition of variables see Table 1 and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A7. Peer Effects in Consumption: Excluding Recent Migrant-Households

Variables	[1]	[2]
	Excluding households who have been living in their current village/ neighborhood for < 10 years	Excluding households who have been living in their current village/ neighborhood for < 20 years
Endogenous Peer Effect	0.741***	0.801***
	(0.207)	(0.194)
Observations	40,024	38,261
Adjusted R-squared	0.327	0.319
Hansen J statistic	2.312	2.109
	[p=0.510]	[p=0.550]
Kleibergen-Paap rk LM statistic	171.752	174.996
	[p=0.000]	[p=0.000]
First stage F-statistic	44.79	45.64

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristics, idiosyncratic shocks and definition of variables see Table 1 and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A8. Peer Effects in Consumption: Including Controls for Price Index

Variables	[1]	[2]
	With controls for household level and neighborhood level price index	Consumption expenditure deflated by price index
Endogenous Peer Effect	0.722***	0.685***
	(0.216)	(0.198)
Observations	40,554	40,523
Adjusted R-squared	0.3338	0.294
Hansen J statistic	1.687	0.875
	[p=0.6399]	[p= 0.832]
Kleibergen-Paap rk LM statistic	165.908	204.062
	[p=0.000]	[p=0.000]
First stage F-statistic	43.18	52.69

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristics, idiosyncratic shocks and definition of variables see Table 1 and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A9. Peer Effects in Consumption: Excluding Households based on Migration Pattern and Using Sample Weights

Variables	[1]	[2]
	Excluding households who have sent out migrants for work	Regression using sampling weights
Endogenous Peer Effect	0.640***	0.825***
	(0.237)	(0.178)
Observations	37,615	40,980
Adjusted R-squared	0.333	0.333
Hansen J statistic	0.927	0.438
	[p=0.819]	[p= 0.932]
Kleibergen-Paap rk LM statistic	136.213	157.400
	[p=0.000]	[p=0.000]
First stage F-statistic	34.55	52.31

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristics, idiosyncratic shocks and definition of variables see Table 1 and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A10. Peer Effects in Consumption: Disaggregate Analysis

Variables	Food	Temptation Good	Health	Education
		Expenditure Categories		
Mean Peer Total Consumption Expenditure	**0.125*****	**0.482*****	**0.055**	**0.006**
	(0.0418)	**(0.173)**	**(0.071)**	**(0.046)**
Observations	40980	40980	40980	40980
Adjusted R-squared	0.562	0.186	0.0220	0.143
Hansen J statistic	2.472	1.557	3.677	4.417
	[p=0.480]	[p=0.669]	[p=0.299]	[p=0.220]
Kleibergen-Paap rk LM statistic	154.800	154.800	154.800	154.800
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
First stage F-statistic	40.01	40.01	40.01	40.01

Notes: Estimation via two-step GMM. Dependent variable for regression reported in each column is the respective expenditure category. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristic, idiosyncratic shocks and definition of variables see Table 1 and note below Table 2. Heteroscedasticity robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A11. Partial Risk Sharing: Subsample Analysis based on (Relative) Group Vulnerability

	Subsamples					
Variables	More Vulnerable: Literacy	Less Vulnerable: Literacy	More Vulnerable: Caste	Less Vulnerable: Caste	More Vulnerable: Poverty	Less Vulnerable: Poverty
Endogenous Peer Effect	**0.806***	**0.897***	**0.624***	**0.832****	**0.629****	**0.767***
	(0.193)	**(0.275)**	**(0.224)**	**(0.323)**	**(0.261)**	**(0.267)**
Observations	20403	20577	20584	20396	20235	20745
Adjusted R-squared	0.178	0.190	0.170	0.217	0.236	0.177
Hansen J statistic	0.376	1.712	0.923	1.371	0.876	0.883
	[p=0.945]	[p=0.634]	[p=0.820]	[p=0.712]	[p=0.831]	[p=0.830]
Kleibergen-Paap rk LM statistic	308.1	116.6	153.5	77.80	195.7	113.1
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
First stage F-statistic	94.37	30.34	41.26	19.79	50.22	30.25

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. Subsample construction: *More Vulnerable: Literacy* includes those groups (villages/neighborhoods) with proportion of illiterate households above the sample group-median level (where the group median level is calculated as the median of the proportion of illiterate households over all groups in the sample). *Less Vulnerable: Literacy* includes those groups with proportion of illiterate households below the sample group-median level. *More Vulnerable: Caste* includes those groups with proportion of low caste households above the sample group-median level. *Less Vulnerable: Caste* includes those groups with proportion of low caste households below the sample group-median level. *More Vulnerable: Poverty* includes those groups with proportion of below poverty line households above the sample group-median level. *Less Vulnerable: Poverty* includes those groups with proportion of below poverty households below the sample group-median level. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristic, idiosyncratic shocks, definition of variables and definition of each subsample see Table 1, note below Table 2 and main text. Heteroscedasticity robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A12. Partial Risk Sharing: Subsample Analysis based on (Relative) Group Heterogeneity

Variables	Subsamples					
	More Heterogeneous: Literacy	Less Heterogeneous: Literacy	More Heterogeneous: Caste	Less Heterogeneous: Caste	More Heterogeneous: Poverty	Less Heterogeneous: Poverty
Endogenous Peer Effect	0.950***	0.742***	1.140***	0.644***	0.731*	0.737***
	(0.280)	(0.228)	(0.415)	(0.223)	(0.415)	(0.235)
Observations	16938	24042	13586	27394	9516	31464
Adjusted R-squared	0.180	0.187	0.155	0.205	0.204	0.206
Hansen J statistic	1.973	0.159	0.618	2.938	1.620	1.781
	[p=0.578]	[p=0.984]	[p=0.892]	[p=0.401]	[p=0.655]	[p=0.619]
Kleibergen-Paap rk LM statistic	94.92	197.5	50.56	138.5	108.1	136.8
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
First stage F-statistic	25.58	51.44	13.54	36.56	30.06	35.07

Notes: Estimation via two-step GMM. Dependent variable is Household Annual Total Consumption Expenditure. Endogenous Peer Effect represents the estimated coefficient of Average Peer Household Annual Total Consumption Expenditure. Subsample construction: *More Heterogeneous: Literacy* includes those groups (villages/neighborhoods) with proportion of illiterate households between 0.25 and 0.75. *Less Heterogeneous: Literacy* includes those groups with proportion of illiterate households below 0.25 or above 0.75. *More Heterogeneous: Caste* includes those groups with proportion of low caste households between 0.25 and 0.75. *Less Heterogeneous: Caste* includes those groups with proportion of low caste households below 0.25 or above 0.75. *More Heterogeneous: Poverty* includes those groups with proportion of below poverty line households between 0.25 and 0.75. *Less Heterogeneous: Poverty* includes those groups with proportion of below poverty households below 0.25 or above 0.75. All specifications include controls for household effects, own idiosyncratic shocks, contextual effects and district fixed effects. Household effects are controlled by own demographic characteristics of households. Contextual effects are controlled by average peer demographic characteristics. For full list of demographic characteristic, idiosyncratic shocks, definition of variables and definition of each subsample see Table 1, note below Table 2 and main text. Heteroscedasticity robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.