Calculating Comparable Statistics from Incomparable Surveys, with an Application to Poverty in India

Alessandro Tarozzi
Duke University*

February 2004

Abstract

Applied economists are often interested in studying changes over time of important economic indicators, such as inequality or poverty, but such comparisons can be made impossible by changes in data collection methodology. We describe an easily implemented procedure to recover comparability that can be adopted whenever the statistic of interest satisfies a moment condition, when the researcher has available a set of auxiliary variables whose reports are not affected by the different survey design, and whose relation with the main variable of interest is stable over time. We analyze the asymptotic properties of the estimator taking into account the presence of clustering, stratification, and sampling weights, which characterize most household surveys. We use the 1999-2000 Round of the Indian National Sample Survey as an empirical illustration. Due to important changes in the adopted questionnaire, the unadjusted figures are likely to understate poverty relative to the previous rounds. We use previous waves of the same survey to provide evidence supporting the plausibility of the identifying assumptions and conclude that most of the very large reduction in poverty implied by the unadjusted figures is real.

JEL: C13, C42, I32, O53

Key words: Poverty, Inequality, India, Method of Moments, Survey Methods.

Address: Department of Economics, Duke University. Social Sciences Building, PO Box 90097, E-mail taroz@econ.duke.edu.

1 Introduction

Applied economists and policy makers are often interested in studying changes over time of important economic indicators, such as inequality, poverty or consumption, which are routinely evaluated using data from household surveys. However, comparisons are only meaningful insofar as the necessary data are collected consistently over time, and it is common to observe changes in the questionnaire adopted by the statistical agencies. The survey literature convincingly shows that such revisions can affect respondents' reports in important ways, so that changes in observed economic indicators sometimes reflect changes in the survey, rather than real transformations of the economic environment.[1]

The difficulties of data collection have been part of the survey literature for decades (see, for example, Mahalanobis (1954), and Neter and Waksberg (1964)), but a growing attention is now developing within economics for the important consequences that survey design can have on the estimation of consumption, income, poverty and inequality. Browning, Crossley, and Weber (2003) and Battistin, Miniaci, and Weber (2003) discuss the use of recall diary expenditure data for the estimation of expenditure, income, and savings in different household surveys. Battistin (2003) shows how different data collection methodologies within the US Consumer Expenditure Survey can lead to very different conclusions when testing the permanent income hypothesis, and when evaluating the evolution of inequality in consumption within the Unites States. The latter topic is also analyzed in Attanasio, Battistin, and Ichimura (2003). Gibson (1999) uses an experiment carried out in Papua New Guinea to study the effects on poverty estimation of collecting expenditure data using diaries instead of recall interviews. Gibson, Huang & Rozelle (2001, 2003) note that changing the reference period in the Chinese Household Income and Expenditure Survey would have dramatic effects on the estimation of poverty and inequality. Jolliffe (2001) studies the large changes in poverty estimates for El Salvador that arise when the list of items included in the expenditure questionnaire is changed. Lanjouw and Lanjouw (2001) perform a similar analysis for Ecuador, Nepal, and Brazil. Others have analyzed the effect of the survey design in expenditure surveys on the estimation of elasticities (Ghose and Bhattacharya, 1995) and economies of scale at the household level (Gibson 2002).

In this paper we develop a simple two-step adjustment procedure that reestablishes comparabil-

---

[1] See Deaton and Grosh (2000, part II) and references therein for an overview of the methodological issues involved in collecting expenditure data.

ity over time for statistics estimated using surveys of different design. Suppose that a researcher is interested in studying the evolution over time of a parameter ϕ_0 that is identified by the population moment condition $E\left[m\left(y; \phi_0\right)\right] = 0$, but that data on y are no longer available due to a change in survey design. Identification can be achieved if there exists of a set of *auxiliary variables* whose reports are not affected by the change in survey design, and whose relation with y is stable over time. Intuitively, information on the evolution over time of the distribution of the auxiliary variables can be used to identify changes in the distribution of y, and therefore in the value of the parameter of interest. We study the asymptotic properties of the estimator in a method of moments framework, taking into account the presence of clustering, stratification, and sampling weights that typically characterize household surveys.

This paper derives its main empirical motivation from a change in the methodology of data collection that took place recently in India, a country which accounts for a significant share of the world poor, and whose poverty numbers are central to many economic and political debates, not only in India, but also within the World Bank. The Indian official poverty numbers are historically based on expenditure data, collected approximately every five years by the Indian National Sample Survey (NSS). The 55th round of the NSS, carried out between July 1999 and June 2000, was awaited by many, with the expectation that it would at least partly dispel the mixed evidence on poverty reduction during the nineties, which apparently conflicted with the high rates of economic growth that followed a process of economic liberalization started in 1991.[2] The official poverty counts show a large reduction in poverty over a short period of time, but there are a priori arguments suggesting that the unadjusted figures are likely to *understate* poverty relative to the previous NSS rounds. The reasons and consequences of the non-comparability of the 1999-2000 survey with previous NSS rounds have already been analyzed elsewhere, and in this paper we will only summarize the main issues. The interested reader is referred to Datt, Ravallion and Kozel (2003), Deaton (2001, 2003a, 2003b), Deaton and Drèze (2002), Sen (2000), Sundaram and Tendulkar (2002, 2003) and Visaria (2000). In particular, Deaton and Drèze (2002) calculate adjusted poverty estimates making use of an identification strategy analogous to the one formalized here, but they use only one auxiliary variable. The major comparability concerns stem from the decision of the statistical agency to adopt a revised questionnaire, introducing changes in the *recall period* adopted for many expenditure items. One major departure from the previous waves was the presence of separate

---

[2] The liberalization process started after a Balance of Payment crisis in the summer of 1991. See Sachs, Varshney, and Bajpai (1999), and references therein.

questions on household expenditure during the last seven *and* thirty days for food items. Previous rounds of the NSS made use only of the latter recall period. We will show that there are convincing arguments for expecting such changes to lead to an *overstatement* of total household expenditure with respect to previous waves.

The usefulness of our adjustment relies crucially on the identifying assumptions being appropriate. Since such assumptions involve variables that are *not* observed, they cannot be formally tested. However, we provide *indirect* empirical support to the assumptions, by making use of smaller experimental waves carried out in years preceding the 55th round. From the 51st to 54th wave, the NSS contains data obtained from *two* different expenditure questionnaires: the standard one, and an experimental one characterized by a modified set of recall periods. The change in the questionnaire produces huge differences in estimated poverty. Because the two questionnaire types were assigned at random, one can examine whether respondents' reports on household characteristics, or expenditure in items for which the recall did *not* change, are left unaffected by the changed recall period for the other expenditure categories. If this is the case, then such unaffected reports are good candidates to be used as auxiliary variables. We also use these smaller waves to analyze the stability over time of the relation between total expenditure and the auxiliary variables. Finally, we use these experimental surveys to test the performance of our adjustment procedure. In a given wave, one can estimate poverty using the standard questionnaires, or using our adjustment procedure to recover a "comparable" estimate from the experimental questionnaires. If the estimator performs well, one should expect the two poverty counts to differ only due to sampling error. Overall, the evidence suggests that our estimates are useful to recover comparability across NSS rounds of different design. Surprisingly, our results show that most of the reduction in poverty shown by the unadjusted figures is real, even if the adjustment suggests that the change in the questionnaire caused, as expected, a relative underestimation of poverty rates, especially in rural areas.

The rest of the paper is organized as follows. Section 2 describes the case study that will form the core of the empirical illustration. Section 3 delineates the general theoretical problem, linking it to the empirical application, and discussing links with previous literature. Section 4 introduces the estimator and describes its asymptotic properties. Section 5 covers the empirical application. In Section 6 we conclude and discuss alternative applications of the results developed in the paper.

2 The Empirical Framework: Poverty in India

For decades, the Planning Commission of the Government of India has regularly published "official" headcount poverty ratios, separately for rural and urban areas of every Indian state and Union Territory. The poverty counts are computed as the fraction of the population living in households with consumption per head below a poverty line. The poverty lines have been calculated to represent the minimum monthly expenditure per head associated, on average, with a sector-specific minimum calorie intake, recommended by the Indian National Institute of Nutrition. The lines are kept constant in real terms by using two different state-specific price indexes: the Consumer Price Index for Agricultural Labourers (CPIAL) for rural areas, and the Consumer Price Index for Industrial Workers (CPIIW) for the urban sector.[3] Expenditure data are collected by the Indian National Sample Survey Organization (NSSO) approximately every five years, from a large sample of Indian households interviewed over a one-year period. Each NSS round contains information on a wide spectrum of socioeconomic variables, but the largest section of the database consists of records of household consumption of a very detailed list of items.

Until the 50th round, carried out in 1993-94, all NSS surveys adopted a 30-day recall period for all expenditure items. This choice of recall period is unusual, as most statistical agencies use a shorter reporting period for items that are typically purchased frequently, like food, and a longer period for infrequent expenditures like clothing, footwear, educational expenses, and durables. Several experimental studies find that expenditure reports for frequently purchased items are on average proportionally *lower* when the recall period becomes *longer*.[4] According to some, a switch to more standard recall periods would have helped reconciling the high rate of growth measured by the National Accounts Statistics in the nineties, and the unimpressive rates of poverty reduction estimated using NSS data over the same period. To explore further the issue, the NSSO experimented with different recall periods using the smaller ('thin') NSS rounds that followed the 1993-94 survey. Such surveys were not specifically designed for poverty monitoring, so some doubts remain on the comparability of their sampling frames, but each wave did also gather information on expenditure

---

[3] For a detailed overview of the issues related to the choice of poverty lines in India see GOI, Planning Commission (1993), or Deaton and Tarozzi (2000), who also criticize the appropriateness of the indexes used to price inflate the lines, and propose alternatives.

[4] See, in particular, Scott and Amenuvegbe (1990). The 30-day recall period was adopted because of an early experimental study by Mahalanobis and Sen (1954) who found that reports based on a 7-day recall period were *too* high. A new study carried out by the NSSO (2003) suggests instead that the shorter recall period is more appropriate for many high-frequency expenditures.

for a list of items basically identical to the standard one.[5] However, the NSSO adopted *two* different expenditure questionnaires and all households living in a given primary stage unit of the survey were assigned the same type, chosen at random. One questionnaire type—*schedule 1*—was the standard one, with a 30-day recall period for all items, while in the other type—*schedule 2*—the period was set equal to the 7 days before the interview for food, beverages and some other items generally bought frequently, and to 365 days for durables, clothing, footwear and some other low-frequency purchases. For what follows, it should be kept in mind that the standard 30-day recall was instead kept *in both schedules* for a list of items accounting for a substantial share of the budget. This list included fuel and light, miscellaneous goods and services, rents and consumer taxes, and certain medical expenses. We will generally refer to this list of items as '30-day' or 'miscellaneous' items.

The experimental surveys once again confirmed the finding that reported expenditure in food is significantly higher if the recall period is shortened. The opposite result was found for most durables: even if more households report some purchases when a one year recall period is adopted, average monthly expenditure is lower than the one observed using a 30-day reference period.[6] Given the large average fraction of the budget spent in food, the net effect in all surveys and in both sectors is a larger estimate of total per capita expenditure (*pce* hereafter) when the experimental questionnaire is used. Table 1 contains summary statistics for the major Indian states.[7] In all surveys and both sectors, average total monthly pce is systematically 10-20 percent higher for households in the experimental group, and the differences are always statistically significant at standard statistical levels. Row (5) shows that, if one keeps the poverty line constant, this gap would translate into a *fifty percent drop in poverty* 'achieved' through a change in the survey methodology.[8] However, the results in row (3) show that the differences in reports on miscellaneous items are always very

---

[5] Expenditure is not the main focus of most NSS waves. The focus, for example, was on education and fertility in the 52nd round, and on informal sector enterprises in the 51st.

[6] For details, see Deaton (2001) and Sen (2000).

[7] Such states include Andhra Pradesh, Assam, Bihar, Gujarat, Haryana (urban sector only), Jammu & Kashmir, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, West Bengal, and Delhi. These states account for more than 95 percent of the total Indian population.

[8] The fact that the poverty lines are based on predicted calorie consumption given total pce might suggest that if the revised questionnaire had to be used for poverty monitoring, the lines should be recalculated. Deaton (2001) notes that doing so would, if anything, further *increase* the gap. Since a shorter recall period causes a larger proportion of the reported total budget to be spent in food, then a lower level of pce would be necessary to achieve the recommended minimum calorie intake. This would further reduce poverty. However, after analyzing the Engel curves for food, Deaton (2001) argues that the poverty lines should remain the same.

small, and in most cases not statistically significant. This is crucial for our purposes, since it gives some preliminary support to the claim that reports on a set of items are not influenced by reports in a different set of items. This observation makes expenditure in miscellaneous items an excellent candidate for inclusion in the vector of auxiliary variables. Moreover, the figures in row (5) show that 30-day items are likely to be good predictor of total pce, since the corresponding average budget share is above 20 percent in the rural sector—using standard questionnaires—and above 25 percent in urban areas. In each one of the surveys represented in Table 1, a simple log-linear regression of total pce on pce in 30-day expenditure produces an R^2 above 0.65.

However, close examination of the figures in row (3) shows that in the rural sector mean pce in miscellaneous items is systematically higher when computed using the standard questionnaire, even if the differences are always small. The sign of the differences is reversed in the urban sector. This empirical regularity is likely to be related to differences in consumption patterns and in household characteristics between the two sectors. The survey literature shows that the cognitive processes adopted to remember expenditure in a given item is generally associated with the characteristics of both item and respondent.[9] This suggests that it might be important to include household characteristics among the auxiliary variables. Here we will make use of information on household size, completed education of the household head, and categorical variables for land holdings, main economic activity of the household, and whether the household belongs to special social groups (called 'Scheduled Castes and Tribes' in the NSS). We will generally refer to these variables as (household-specific) *controls*.

Finally, notice that the figures in Table 1 show no apparent trend in poverty reduction over the examined period. However, the thin rounds were not specifically designed as expenditure surveys. The relatively small samples, coupled with the choice of sampling frames more suited to the different main purpose of these surveys, induced many observers to look at these poverty figures with some suspicion, and to wait for the next quinquennial expenditure survey, that is, for the 55th wave of the NSS.

However, in the 55th round of the survey, the NSSO decided to adopt a questionnaire combining *both* sets of recall periods used in the thin rounds. Expenditure in food was to be recorded using *both* the 30-day *and* the 7-day recall periods for *all* households, while for durables and other infrequently purchased items only a 365-day recall was to be used. Once again, a 30-day recall was kept for items with intermediate frequency of purchase. The unadjusted results estimated by the Indian

---

[9] See Deaton and Grosh (2000, part II) for an overview.

Planning Commission showed an impressive reduction in poverty with respect to the early nineties: using 30-day reports for food, in the rural sector the counts dropped from 37.2 in 1993-94 to 27.1 six years later, while in urban areas the proportion dropped from 32.6 percent to 23.6. In both cases this amounts to a reduction of one third in poverty rates, in less than a decade. However, the changes in the questionnaire cast serious doubts on the comparability of the more recent figures with previous poverty estimates, especially if one considers the results of the thin experimental rounds.

On the one hand, the thin rounds showed that reports on durables are on average lower when a 1-year recall period is used, so that the new questionnaire would overstate poverty. At the same time, more respondents reported *some* expenditure in durables, with the consequence that the corresponding distribution is much more spread out when the shorter recall period is used. Keeping the average report constant, this would cause the opposite result of *lower* poverty estimates when the experimental questionnaire is used. The two conflicting effects combine with the fact that durables typically account for a small share of the total budget, especially among poor households, making it unlikely that important comparability issues arise as a consequence.

On the other hand, the new questionnaire recorded the two separate reports on food expenditure in two parallel columns printed next to each other. One can therefore expect that this format prompted the respondents (or the interviewers) to reconcile the two different reports. So, consumption of food reported with the traditional 30-day recall period would be disproportionately *high* (since the respondent would tend to avoid large discrepancies with the 7-day reports, which are typically higher), *and/or* the corresponding reports based on a 7-day recall period would be disproportionately *low* (by a symmetric argument). The plausibility of this argument is strengthened by the fact that in the 55th round average pce in food as estimated with a 7-day recall period exceeded the corresponding figure calculated using the 30-day recall period by about 6 percent, while in all the thin rounds the gap was consistently above 30 percent. Since for most Indian households food accounts for a very large share of the total budget, these arguments lead to the expectation that the unadjusted figures might significantly overstate total expenditure, and therefore significantly understate poverty.

3　The Theoretical Problem

In this section, we describe the general theoretical problem, drawing explicit links to the empirical illustration that we will develop later in the paper. We refer to the population sampled using a revised methodology as the *target population*, while an *auxiliary population* is one that has been sampled using a standard questionnaire. Target and auxiliary *surveys* are analogously defined. In the empirical application, the 55th NSS wave will be the target survey, while we will use different previous rounds of the same survey as potential auxiliary surveys.

Let τ be a binary variable equal to one when an observation is drawn from the target population, and zero otherwise. The researcher is interested in estimating the value of a parameter ϕ_0 in a target population, where ϕ_0 satisfies the following population moment condition:

$$E\left[nm\left(y; \phi_0\right) \mid \tau = 1\right] = 0 \tag{1}$$

where n is household size, and y is the main variable of interest *as measured in a standard questionnaire*. In poverty or inequality measurement, y typically measures expenditure or income per head. The population moment condition (1) explicitly refers to the frequent situation in which the parameter of interest is defined in terms of *individuals*, but data are sampled at the *household* level. If the sampling unit is the same as the unit in terms of which ϕ_0 is defined, all the results that follow can be obtained as a straightforward special case with $n = 1$. In (1), we abstract from issues of intrahousehold allocation of resources, so that each individual within a household is treated equally.[10] The moment condition (1) encompasses a broad set of commonly used poverty and inequality measures.[11] For example, if ϕ_0 represents a Foster-Greer-Thorbecke poverty index, and z is a fixed poverty line, then $m\left(y; \phi_0\right) = 1(y < z)\left(1 - \frac{y}{z}\right)^{\alpha} - \phi_0$, where $\alpha \geq 0$, and $1(E)$ is an indicator function equal to one when event E is true. When $\alpha = 0$, the index becomes the headcount poverty ratio, while $\alpha = 1$ characterizes the poverty gap ratio. A higher parameter α indicates that large *poverty gaps* $(1 - y/z)$ are given a larger weight in the calculation, so that the poverty index becomes more sensitive to the distribution of y among the poor. Equation (1) also identifies well-known inequality measures like the Atkinson index, if $m\left(y; \boldsymbol{\phi_0}\right) = \left[\left(\frac{y}{\phi_{02}}\right)^{1-\epsilon} - \phi_{01}, y - \phi_{02}\right]^{T}$,

---

[10] See Deaton (1997, Ch. 4), and Case and Deaton (2002) for an overview of the issues involved in welfare evaluation when household scale economies and equivalence scales are taken into account.

[11] For an introduction to the theory and practice of poverty measurement see Deaton (1997, Ch. 3).

where $\phi_0 = [\phi_{01}, \phi_{02}]^T$, or the Theil index, with $m(y; \phi_0) = \left[\frac{y}{\phi_{02}} \ln \frac{y}{\phi_{02}} - \phi_{01}, y - \phi_{02}\right]^T$. The superscript T will indicate transpose throughout the paper. Note that in these latter cases one needs also to estimate the mean of y, which is generally not known.

Clearly, if y is not measured in the data collected from the target population, the estimation of the parameter ϕ_0 through the sample analogue of (1) is infeasible. This is precisely the case if the survey questionnaire changed in such a way that the respondents' reports are no longer comparable with those from previous surveys, so that the researcher can only observe a different variable, say $\tilde{y}$, but not y. In our empirical setting, y is total expenditure per head when a 30-day recall is used for all items, while $\tilde{y}$ is the expenditure observed when a revised questionnaire is adopted.

Let $\mathbf{v}$ denote a set of other *auxiliary* variables that are recorded by a standard methodology. In analogy with the notation introduced for the main variable of interest, $\tilde{\mathbf{v}}$ denotes the set of auxiliary variables when they are measured using a revised methodology. The set of auxiliary variables $\mathbf{v}$ can include other variables recorded in a questionnaire, like household size, schooling of household members etc., but it can also include *sub-components* of y. For example, an important auxiliary variable for our empirical application will be reported expenditure in "30-day items", for which the recall period was *not* revised.

Each observation is then characterized by the set of variables $(y, \tilde{y}, \mathbf{v}, \tilde{\mathbf{v}}, \tau)$, but the econometrician only observes either $(y, \mathbf{v})$, when $\tau = 0$, or $(\tilde{y}, \tilde{\mathbf{v}})$, if $\tau = 1$. The problem of estimating ϕ_0, then, is clearly a case of estimation in presence of missing data, and it cannot be solved without further assumptions. To recover identification, we make the assumption that the distribution of the auxiliary variables is not affected by the change in survey design—so that the distribution of $\mathbf{v}$ *is* identified in the target survey—and that the expectation of $m(y; \phi_0)$ conditional on the auxiliary variable is stable over time—so that it can be estimated from an auxiliary survey. The following proposition describes formally the argument:

Proposition 1 - Suppose that there exist a set of auxiliary variables $\mathbf{v}$, including household size n, distributed according to $dP(\mathbf{v})$, and assume that the following conditions hold: (A1) $dP(\tilde{\mathbf{v}} \mid \tau = 1) = dP(\mathbf{v} \mid \tau = 1)$ a.s. (A2) $E[m(y; \phi_0) \mid \mathbf{v}, \tau = 1] = E[m(y; \phi_0) \mid \mathbf{v}, \tau = 0]$ a.s.; (A3) $dP(\mathbf{v} \mid \tau = 1)$ is absolutely continuous w.r.t. $dP(\mathbf{v} \mid \tau = 0)$ (A4) $Supp(\mathbf{v} \mid \tau = 1) \subseteq Supp(\mathbf{v} \mid \tau = 0)$. Then ϕ_0 satisfies the following modified population moment condition

$$E[nR(\mathbf{v}) m(y; \phi_0) \mid \tau = 0] = 0 \tag{2}$$

where $R(\mathbf{v})$ is the *reweighting function* defined as

$$R(\mathbf{v}) = \frac{dP(\mathbf{v} \mid \tau = 1)}{dP(\mathbf{v} \mid \tau = 0)} = \frac{P(\tau = 1 \mid \mathbf{v}) P(\tau = 0)}{P(\tau = 0 \mid \mathbf{v}) P(\tau = 1)} \tag{3}$$

moreover, $R(\mathbf{v})$ is nonparametrically identified by the sampling process.

Proof: see appendix.

The conditional probability $P(\tau = 1 \mid \mathbf{v})$ in (3) is the probability that a *household* belongs to the target population conditional on observing $\mathbf{v}$, and the other probabilities are defined accordingly. The proof proceeds through simple manipulations by repeated application of Bayes' rule, and is reported in the appendix. Notice that the reweighting function $R(\mathbf{v})$ *is* identified by the sampling process even if the researcher only observes $\hat{\mathbf{v}} = \tau \tilde{\mathbf{v}} + (1 - \tau) \mathbf{v}$. Indeed, in the appendix we prove that the assumptions of Proposition 1 are sufficient to ensure that $P(\tau = 1 \mid \hat{\mathbf{v}}) = P(\tau = 1 \mid \mathbf{v})$ a.s..

Assumption A1 requires that the econometrician has access to a set of *auxiliary* variables $\mathbf{v}$ whose marginal distribution is identified by the sampling process both in the auxiliary and in the target population. In other words, $\mathbf{v}$ includes variables whose *distribution* of reports are left unaffected by the change in survey design (note, instead, that we do *not* require $\mathbf{v} = \tilde{\mathbf{v}}$). Such variables are likely to be available in most empirical settings, since questionnaire revisions generally leave several questions unchanged. In our empirical application, good candidates for inclusion in $\mathbf{v}$ are household size and other household characteristics, whose definition remained the same across different NSS waves, and also reported expenditure in items for which the 30-day recall was retained. Assumption A2 is the most crucial assumption, and the one who should be scrutinized more closely in every empirical application of our methodology. It requires that the conditional expectation of the function $m(y; \boldsymbol{\phi_0})$ is the same in the target and the auxiliary surveys. When one is interested in estimating a headcount poverty ratio, A2 amounts to assuming that the fraction of households to be counted as poor *conditional on* $\mathbf{v}$ remains constant across the two surveys. Finally, assumptions A3 and A4 ensure that $R(\mathbf{v})$ exists and is bounded for each value of $\mathbf{v}$.[12]

The function $R(\mathbf{v})$ transforms the conditional expectation of $m(y; \boldsymbol{\phi_0})$ from the auxiliary survey into the unconditional expectation in the target survey, down-weighting (up-weighting) households whose auxiliary variables have a relatively high (low) density in the auxiliary survey. This form of

---

[12] Note that even if A4 does not hold one can still estimate bounds for the parameter of interest treating observations with $\mathbf{v}$ outside the common support as missing values, using the setting described in Horowitz and Manski (1995).

reweighting is closely related to an estimation procedure frequently referred to as *propensity score* or *inverse probability weighting*, where observations are weighted with the inverse of the conditional expectation of a binary variable—the *propensity score* (Rosenbaum and Rubin (1973))—whose meaning depends upon the context.[13] Horvitz and Thompson (1952) introduced weighing to account for the different probability of selection of different observations within a sample. Wooldridge (1999) studies inverse probability weighting for stratified samples. Several authors have used this framework to solve missing data problems in the context of cross sectional or longitudinal data (see, among the others, Robins, Rotnitzy and Zhao (1994, 1995), Wooldridge (2002b)). In such framework, $P(\tau = 1 \mid \mathbf{v})$ represents the probability that a unit is not missing, conditional on some observed covariates. Propensity score reweighting is also used in the program evaluation literature (see, for example, Abadie (2003), Hirano, Imbens, and Ridder (2003)). In such context, the propensity score is the probability that an individual participates to a program, conditional on observed covariates.

The explicit requirement of Assumption A1 is in general not necessary in the literature described above, as the conditioning variables are usually available both when $\tau = 1$ and when $\tau = 0$. However, such condition is not necessarily satisfied here, as changes in some parts of a survey questionnaire might, in principle, affect the pattern of reports for variables recorded in unchanged sections. So, for example, a longer list of expenditure items included in a revised questionnaire might affect measurement error in other reported household characteristics, because it can increase weariness in the subject interviewed.

It is relatively straightforward to describe sufficient conditions analogous to the ones in Proposition 1 that allow to identify the density $f(y \mid \tau = 1)$. This is clearly of interest if, for example, y is expenditure per head, and one is interested in changes in distribution over time. Note also that the identification of a comparable density is sufficient for the identification of most commonly used poverty and inequality measures, as long as they are defined in terms of the variable y. Below we consider again a context where the object of interest is the density of a *per capita* quantity y, while data are collected at the *household* level. This implies that the individual-based density of y, which we denote by $f_n(y \mid \tau = 1)$, is described in the population by the following expression:

$$f_n(y \mid \tau = 1) = \frac{E[nf(y \mid n, \tau = 1)]}{E[n \mid \tau = 1]}$$

where $f(y \mid n, \tau = 1)$ is the density defined over *households*.[14] The following proposition formalizes

---

[13] For a textbook treatment of inverse probability weighting see Wooldridge (2002a)

[14] It is easy to check that $f_n(y \mid \tau = 1)$ actually integrates to one.

conditions for identification.

Proposition 2 - Suppose that there exist a set of auxiliary variables $\mathbf{v}$, including household size n, distributed according to $dP(\mathbf{v})$, such that A1, A3, and A4 hold. Let $\mathbf{v}_{-n}$ be a vector of observed variables including all the variables in $\mathbf{v}$ except n. Suppose also that (A2b) $f(y \mid \mathbf{v}, \tau = 1) = f(y \mid \mathbf{v}, \tau = 0)$. Then

$$f(y \mid n, \tau = 1) = f(y \mid n, \tau = 0) E[R(\mathbf{v}_{-n}) \mid y, n, \tau = 0] \tag{4}$$

where the reweighting function is now defined as $R(\mathbf{v}_{-n}) = \frac{P(\tau=1|\mathbf{v})P(\tau=0|n)}{P(\tau=0|\mathbf{v})P(\tau=1|n)}$.[15]

Proof: see appendix.

We do not proceed to analyze the estimation of (4), since we are mostly interested in the estimation of parameters identified by a moment condition such as the one described in (2). We turn now to analyze the estimation and the asymptotic properties of a two-step estimator for such parameters.

4 Estimation

From the modified moment condition (2) above, it is clear that the estimation of ϕ_0 requires the estimation, in a first step, of the reweighting function $R(\mathbf{v})$, which is a function of the unknown probabilities $P(\tau = 1 \mid \mathbf{v})$ and $P(\tau = 0)$. The latter probability represents the proportion of *households* belonging to the target population, while the propensity score can be interpreted as the fraction of households whose covariates are equal to $\mathbf{v}$ that belongs to the target survey. We emphasize that both probabilities refer to the distribution of households, and not individuals, so that they have to be estimated without inflating observations by household size.

In what follows we proceed assuming that the sample is drawn from a population encompassing both the target and the auxiliary population. We use a parametric model for the propensity score. Specifically, we use a logit model.[16] A flexible functional form can be achieved using polynomials,

---

[15] For an example of reweighting in the context of density estimation see DiNardo, Fortin, and Lemieux (1996).

[16] An analogous strategy is adopted in Wooldridge (2002b), Robins, Rotnitzky, and Zhao (1995), Abadie (2003). The latter also analyzes nonparametric first-step estimators, as in Hirano, Imbens, and Ridder (2003)

and in any case we are only interested in obtaining good predictions for the conditional probabilities, while the parameters estimated in the binary variable model will be of little or no interest per se. Also, it is well known that the choice of functional form in a binary dependent variable model has rarely important consequences for the predicted probabilities.[17] All the asymptotic results can be adapted in a straightforward way to a more general case where the propensity score can be written as a known function $P\left(\tau = 1 \mid \mathbf{v}; \boldsymbol{\theta}^1\right)$, where $\boldsymbol{\theta}^1$ is a (column) vector of parameters, estimated with maximum likelihood. The unconditional probability $P\left(\tau = 0\right)$, which we denote by θ^0, can be easily estimated as the fraction of households in the sample that belongs to the auxiliary population. Once $\hat{\boldsymbol{\theta}}^1$ and $\hat{\theta}^0$ have been calculated, $\hat{\phi}$ is obtained by using the sample analogue of the moment conditions in (2).

If the observations are a simple random sample, standard errors can be estimated using the standard asymptotic theory for method of moments estimators (see, for example, Newey and Mc-Fadden (1994)). However, in our framework the asymptotics is complicated by the fact that most surveys adopt a stratified and clustered design, which makes the assumption of i.i.d. observations untenable.[18] This is the case for most widely used surveys, like the World Bank's LSMS, or the CPS and the PSID in the United States, or the Indian NSS. In stratified and clustered surveys, the population is first divided into a fixed number of *strata*, which are usually defined following geographical and/or socioeconomic criteria. Then a predetermined number of *clusters* (typically villages, or urban blocks) are sampled independently from each stratum. Finally, households are selected independently within each cluster or, like in the NSS, from separate second-stage strata created within cluster. The use of stratification in survey design typically leads to *lower* standard errors, as all possible samples become by construction more similar, because a fixed proportion of observations are selected from different areas. On the contrary, clustering frequently leads to standard errors that are considerably *higher* than the ones calculated assuming simple random sampling. This is a consequence of the positive correlation that is common for variables recorded in the same cluster. In most cases, the net effect of clustering and stratification is an *increase* in standard errors, so that ignoring the multi-stage design of a survey can lead to seriously misleading inference. Also, in most surveys the sampling scheme is such that the *ex-ante* probability of selection is not the same for each households, so that consistent estimation of population parameters requires the use of sampling weights.

---

[17] See, for example, Amemiya, Ch. 9.

[18] For an overview of the issues involved in estimation and inference in multi-stage surveys see Deaton (1997, Ch. 1).

We estimate correct standard errors that take into account the presence of a complex survey design, and we derive the asymptotic results letting the total number of *clusters* grow to infinity, while both the number of households selected in a cluster and the proportion of clusters selected in each stratum are kept constant. This setting is appropriate for our purposes, as in the NSS the number of clusters is much higher than the number of households selected per cluster. Since we use observations sampled from two different databases, one has to be explicit about how the sampling from the two populations is done. Here we assume that the first stage strata are the same across the two subpopulations. Related results can be found in Bhattacharya (2003), who studies asymptotic properties of GMM estimators in presence of multi-stage surveys, in the standard situation where all observations belong to the same population.

Let $\boldsymbol{\beta}_0$ denote the (column) vector that contains the true value of the parameters to be estimated (including the ones estimated in the first step), and let $\mathbf{m}(y, \mathbf{v}, \tau; \boldsymbol{\beta}_0)$ denote the set of moments that identify all such parameters. These moments, that also include the first order conditions from the maximum likelihood estimation of the logit coefficients, are described in the appendix. Suppose that the population is divided into S strata, and that stratum s contains a *mass* of H_s clusters, indexed by c. Each cluster contains $M(s, c)$ households, and household h has n_{sch} members. Then, the population moment condition can be written as

$$\sum_{s=1}^{S} H_s E_c \left\{ \sum_{h=1}^{M(s,c)} n_{sch} \mathbf{m}\left(y_{sch}, \mathbf{v}_{sch}, \tau_{sih}; \boldsymbol{\beta}_0\right) \mid s \right\} = 0 \tag{5}$$

where the expectation is taken with respect to the distribution of clusters in stratum s. A total of n clusters is sampled from the population of clusters, and a constant fraction a_s of them, for a total of n_s clusters, is selected from the sth stratum. Then m_{sc} households are selected from cluster c belonging to stratum s. So, the sample analogue of (5) is

$$0 = \sum_{s=1}^{S} \frac{H_s}{n_s} \sum_{c=1}^{n_s} \frac{M(s,c)}{m_{sc}} \sum_{h=1}^{m_{sc}} n_{sch} \mathbf{m}\left(y_{sch}, \mathbf{v}_{sch}, \tau_{sih}; \hat{\boldsymbol{\beta}}\right) \tag{6}$$

where $\omega_{sch} = (n_s m_{sc})^{-1} H_s M(s, c) n_{sch}$ represents the sampling weight. Asymptotic properties can be derived rewriting (6) as the average of n independent cluster-specific terms. So, re-indexing clusters by i :

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{m}_i\left(\hat{\boldsymbol{\beta}}\right) = 0 \quad \text{where} \quad \mathbf{m}_i\left(\hat{\boldsymbol{\beta}}\right) = \sum_{s=1}^{S} 1\left(i \in s\right) \sum_{h=1}^{m_{si}} \omega_{sih} \mathbf{m}\left(y_{sih}, \mathbf{v}_{sih}, \tau_{sih}; \hat{\boldsymbol{\beta}}\right) \tag{7}$$

In the appendix we describe regularity conditions that are sufficient for consistency and asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}$. In particular, such conditions lead to the following:

Proposition 3: If assumptions A0-A8 described in the appendix hold, then

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{d} N\left(0, \Gamma^{-1}W\left(\Gamma^{-1}\right)^T\right) \tag{8}$$

where $\xrightarrow{d}$ indicates convergence in distribution, and

$$\Gamma = p\lim \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}E\left(\mathbf{m}_i\left(\boldsymbol{\beta}_0\right)\right), \quad W = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}Var\left(\mathbf{m}_i\left(\boldsymbol{\beta}_0\right)\right)$$

Proof: see appendix.

Letting m^d indicate the d-th row of the vector $\mathbf{m}\left(.\right)$, and β_l indicate the l-th element of $\boldsymbol{\beta}$, the elements of Γ and W can be estimated consistently by using

$$\hat{\Gamma}_{dl} = \sum_{s=1}^{S}\sum_{i=1}^{n_s}\sum_{h=1}^{m_{si}}\omega_{sih}\frac{\partial m^d\left(y_{sih}, \mathbf{v}_{sih}, \tau_{sih}; \hat{\boldsymbol{\beta}}\right)}{\partial\beta_l} \tag{9}$$

$$\hat{W}_{lj} = \sum_{s=1}^{S}\sum_{i=1}^{n_s}m_{si}^l m_{si}^j - \sum_{s=1}^{S}\frac{1}{n_s}\left[\sum_{i=1}^{n_s}m_{si}^l\right]\left[\sum_{i=1}^{n_s}m_{si}^j\right] \tag{10}$$

where $m_{si}^l = \sum_{h=1}^{m_{si}}\omega_{sih}m^l\left(y_{sih}, \mathbf{v}_{sih}, \tau_{sih}; \hat{\boldsymbol{\beta}}\right)$.

In the empirical illustration, we will use the expressions described in this section to estimate standard errors for adjusted poverty estimates, taking into account the multi-stage design of the Indian National Sample Surveys.

5 Empirical Application

Here, we turn to the empirical illustration of the methodology described in the previous section. Before discussing the adjusted poverty rates from the last—otherwise non-comparable—wave of the NSS, we use the smaller ("thin") NSS waves described in Section 2 to provide indirect evidence supporting the validity of the assumptions necessary for identification, and to test the performance of the adjustment procedure. This is possible because in these experimental waves two different questionnaire types were used, and assigned at random at respondents living in different clusters. So, using the notation laid out in Section 3, for each respondent we observe either $(y, \mathbf{v})$ or $(\tilde{y}, \tilde{\mathbf{v}})$. However, random assignment of questionnaires to households selected from the same population ensures that the distribution of both $(y, \mathbf{v})$ and $(\tilde{y}, \tilde{\mathbf{v}})$ are identified in each survey. In the rest of the paper, we will not use data from the 54th round. This survey was carried out over a six-month

period, while all other rounds are carried out over a whole year, so seasonality issues might cause comparability problems.

5.1 Validating the Assumptions

Assumption A1 requires that the reports on the variables included in $\mathbf{v}$ be independent of the questionnaire type. This assumption is easily tested for the discrete household specific controls that we include in $\mathbf{v}$, all of which are discrete. For each sector and survey we test the hypothesis that the distribution of each of these variables is independent upon the questionnaire types, by using a Pearson χ^2 statistic modified to take into account stratification and clustering.[19] Under the null hypothesis, once we tabulate observations across values and schedules, the 'joint' proportion of observations in the cell related to the the i-th value and the k-th schedule, should be the same as the product of the 'marginal' proportion of observations having the i-th value, and the 'marginal' proportion of observations in the k-th schedule. The test rejects the null when a normalized sum of the differences between joints and products of marginals is large. We report the p-values of each test in Table 2. The results strongly support the null. Only 3 out of 30 tests reject the null, and even in these cases the proportions are very similar across the schedules, as one can see from the cross tabulations reported in the lower part of Table 2. These results are hardly surprising, since the two different questionnaires are assigned randomly, and there is no obvious reason why differences in the recall periods should affect respondents' reports on the household controls we use here.

Let m and $\tilde{m}$ denote the (log of) pce in miscellaneous items, as measured respectively in a standard and experimental questionnaire. Assumption A1 requires the equality of the densities of m and $\tilde{m}$ in the same round and sector. In Figure 1 we draw nonparametric kernel estimates of the densities of $m, \tilde{m}, y$ and $\tilde{y}$ for each survey and sector.[20] It is apparent that in all cases the distribution of $\tilde{y}$ is shifted to the right with respect to the distribution of y. This is consistent with the results in Table 1, which showed that mean total pce is systematically higher for households in the experimental group. However, there is no such large and systematic gap between the distributions of m and $\tilde{m}$. Only in the urban sector of the 51st round does the difference between the two curves appear visually not negligible, but even in this case the two densities closely coincide for low

---

[19] More precisely, we use a second-order corrected Pearson statistic, as in Rao and Scott (1984).

[20] We use the robust bandwidth proposed by Silverman (1986) for the estimation of approximately normal densities with a biweight kernel.

values of expenditure, which is the relevant range when one is interested in poverty counts. We perform a simple test for the equality of the distributions of m and $\tilde{m}$ using the Pearson χ^2 statistic described above.[21] First we divide the range of m into bins of equal length, and then we test the null hypothesis that the distribution of the observations across bins is independent upon the schedule used to measure m. To avoid the presence of many empty cells, we consider only observations included between the first and the last percentile of the round-sector distribution. Table 3 shows the estimated p-values computed using 10, 15, or 20 bins, as a robustness check. In most cases the null hypothesis cannot be rejected, and the p-values are above 0.2. The null is never rejected if we use a one percent significance level, while using a five percent level we reject in the rural sector of the 52nd Round, with 10 or 15 bins, and in both sectors of the 51st Round with 10 bins.

Overall, assumption A1 appears to hold well for the variables we plan to include in $\mathbf{v}$, so we move to analyze assumption A2, which in this context requires the stability over different rounds of the probability of being poor, conditional on the observed $\mathbf{v}$. In Figure 2 we plot the estimated probabilities conditional on m only. Each line represents a nonparametric locally weighted regression on m of a dummy variable equal to one when a household's pce is below the poverty line z.[22] Because we are interested in the stability of $P\left(y < z \mid m\right)$, all the lines are constructed using only observations from households that received the standard questionnaire. Even if the lines do present some systematic gaps in some areas of their range, they look extremely similar overall, suggesting that the assumption of a constant conditional probability is at the very least a sensible working hypothesis. Note also that there is no apparent time trend in the way the curves differ from each other, suggesting that no important gradual change is affecting the relation between m and y.

In our context, the availability of empirical evidence supporting the stability of the conditional probability is extremely important, since there are several reasons why such assumption might fail.

---

[21] We do not use a Kolmogorov-Smirnov test for the equality of two densities estimated nonparametrically since this test does not allow for the presence of a complex survey design.

[22] In a locally weighted regression a "local" OLS regression is run at every point where the conditional expectation is evaluated (see Fan, 1992). The regression is local since at every point we use only observations for which the regressor is inside a neighborhood of the point itself, defined by a bandwidth. The observations are weighted by using a kernel, so that observations closer to the point have more weight in the regression itself. We prefer locally weighted regressions to the traditional Nadaraya-Watson estimators since the former tends to reduce the bias arising in the estimates when the density of the regressor is not flat. For a clear treatment of the Nadaraya-Watson estimator see Pagan and Ullah (1999, Ch. 3). Deaton (1997, Ch. 3) provides an intuitive treatment of locally weighted regressions.

The major concern is that movements in the relative price of m should be expected to affect the conditional probability of being poor, whose stability ultimately depends upon the stability of the Engel curve linking y to m. Also, changes in tastes—or other demand shocks—might change the relation between one survey and another. All these concerns are likely to be less pressing if the auxiliary and the target survey are carried out in consecutive years. In any case, note that our procedure is flexible enough to accommodate many of these factors, as long as they are observable, and therefore can be included in the vector $\mathbf{v}$.

As a further check, we impose a logit functional form to the conditional probability $P\left(y < z \mid \mathbf{v}\right)$, and we test for the equality of the coefficients across surveys. In Table 4 we report robust tests for the equality of coefficients across surveys for each pair of rounds, for the rural and urban sectors separately, and using different sets of variables included in $\mathbf{v}$. Over a total of 60 tests, the joint null hypothesis of equal coefficients is rejected in 21 cases using a five percent significance level, and in 11 cases using a one percent level. Overall, the conditional probabilities appear remarkably stable, especially taking into account the very large size of the samples (each tests uses a minimum of 18153 observations, and a maximum of 72528), and the fact that in most cases the null imposes more than twenty restrictions.

5.2   Performance of the Estimator

The previous section shows that the assumptions needed for the good performance of our adjustment procedure are reasonable, in our empirical context. As a further check, we use the thin rounds to perform an empirical exercise in order to evaluate the performance of the estimator. For each thin round we can estimate a 'comparable' poverty count using data on total pce from the subset of households who received a standard questionnaire. Then we can attempt to replicate these 'benchmark' poverty rates, using data on the auxiliary variables collected from households who received an experimental questionnaire, while deriving information on $P\left(y < z \mid \mathbf{v}\right)$ from any of the available standard surveys. If the reweighting procedure performs well, one should not be able to reject the hypothesis that the benchmark and the adjusted estimates are equal. In the first step of the estimation we use a logit model for the propensity score $P\left(\tau = 1 \mid \mathbf{v}\right)$. To construct the tests we estimate robust standard errors by using expressions (9) and (10) described in Section 4 of the paper. In the appendix we characterize in detail the form taken by such expressions when the object of interest is a headcount poverty ratio, and the propensity score is modeled using a logistic

distribution, and estimated using maximum likelihood.

We report the results of the tests in Table 5. Columns 1 and 2 report the benchmark and the unadjusted poverty count for each round. As we already pointed out, the poverty counts are approximately halved when the experimental surveys are used. In every case the differences are statistically significant at any standard significance level. Columns 3 to 6 report the adjusted poverty ratios and their standard errors, while asterisks denote the cases in which the null of no difference between benchmark and adjusted ratios is rejected. Each test is simply based on the ratio between the difference and its standard error. Since in each cell the benchmark and the adjusted ratio are computed using data from independent surveys, such standard errors are trivial to compute, once one has the standard errors for each poverty count. For each sector in each thin round, ordered along the rows of the tables, we experiment using auxiliary data from the same sector, in either the 50th round, or from the standard survey in each of the thin rounds. For each target-auxiliary pair we use three different sets of auxiliary variables: a polynomial in m and household size, a polynomial in household size and dummies for household controls, or all of the above.

The results are mixed, but overall encouraging. Using a five percent significance level, the null is not rejected in 48 out of 72 tests. The null is never rejected when we use the standard survey in a thin round as auxiliary survey for the experimental survey in the same round. This can be interpreted as further indirect evidence of the validity of assumption A1, since the random assignment of questionnaire type should guarantee that, within the same round, assumption A2 holds. In all cases, the absolute differences between the benchmark and the adjusted figures are much lower than the differences across questionnaire types in the same round. However, in the rural sector, when m is included among the auxiliary variables, the adjusted figures are systematically above the corresponding benchmark, by 3-5 percentage points. However, this should not suggest that excluding m from the auxiliary variables would improve the performance of the estimator. In fact, careful examination of the adjusted estimates obtained excluding m reveals that the adjustment does little more than reproducing the poverty ratio from the corresponding auxiliary survey. For example, the headcount ratio for the rural sector of the 51st round is 41.8, and adjusted ratios calculated using this as auxiliary survey are 41.6 in the 52nd round, and 40.9 in the 53rd round. If the auxiliary survey is the 53rd round, in which the benchmark poverty ratio is 35.7, the adjusted headcounts are 35.3 for the 51st round, and 35.6 for the 52nd round. One plausible explanation is the fact that the household controls do not explain a large fraction of the total variance of y, so that

the reweighting function $R(\mathbf{v})$ is everywhere close to one, and as a consequence the adjustment merely reproduces the poverty count in the auxiliary survey. For this reason, when applying our adjustment procedure to the 55th NSS round, we always include m among the auxiliary variables.

As one would expect, there is a close correspondence between the performance of the estimator and the results of the tests for the validity of the identifying assumptions that we described in the previous sections. For example, in the urban sector of the 53rd round the hypothesis that the benchmark and the adjusted ratios are the same is rejected in all specifications, when the 52nd round is used as auxiliary survey. At the same time, the figures in the bottom panel of the last column in Table 4 show that, for this pair of rounds, the assumption of equal conditional probability in the same urban sector always fails. We observe the opposite if we use the 51st round as auxiliary survey for the 53rd (or vice-versa). These results clearly stress that in empirical applications much care should be taken in evaluating the credibility of the identifying assumptions.

5.3 Adjusted Poverty Estimates from the 55th Round of the National Sample Survey

The questionnaire adopted in the 55th round of the NSS is different from any previously adopted one. We already mentioned that the questionnaire asked all respondents to report consumption in food using two different recall periods, while a 365-day recall period was introduced for consumption on durables and some other items. However, expenditure in miscellaneous items, a good predictor of total pce, was reported only using the standard 30-day recall period, so we can use this variable, together with other controls, to implement our adjustment procedure. In the previous section we provided indirect evidence supporting the stability of the conditional probability of being in poverty given the auxiliary variables, and we showed that changes in some sections of the questionnaire seem to have only mild consequences on reports recorded in unchanged sections.

Table 6 contains the sector-specific adjusted estimates for the poverty counts. As a robustness check, we use all the NSS surveys between the 50th and the 53rd Round as auxiliary surveys. As usual, we make use of the official poverty lines for all India in 93-94 Rupees. All monetary values from subsequent Rounds are deflated using state and sector specific official Consumer Price Indexes. Here we exclude Jammu and Kashmir from our estimates since we do not have information on the relevant price indexes for the latest survey period. We report two sets of estimates. In Column 1, the auxiliary variables include a cubic in m and household size. In Column 2 we also include the

other household-specific controls mentioned earlier in the paper. We obtain all adjusted poverty ratios using the expression in (6), while the standard errors are computed as the first element along the diagonal of the robust covariance matrix in (8).

The unadjusted poverty figures for the states included in our analysis are 28.4 percent in rural areas and 24.5 percent in urban areas. In all cases but one, our adjustment procedure produces higher estimates of poverty, as expected. The one exception is the point estimate for urban areas when the 52nd round is used as auxiliary survey, and household specific controls are not included among the auxiliary variables. In the rural sector, adjusted poverty counts range from 30.4 percent to 32.5 percent. In urban areas, the adjusted figures range from 24 to 27.3 percent. In all cases the increase in point estimates is more pronounced if the controls are included. The inclusion of such controls also slightly increases the estimated standard deviations.

Expenditure data from the previous quinquennial round, carried out in 1993-94, showed that the proportion of the population in poverty in the states considered here was 33.4 percent in urban areas, and 38.2 percent in the countryside. So, even if the adjustment delivers—as expected—higher headcount ratios than the official ones, our estimates confirm a very large poverty reduction in India during the nineties. This conclusion is consistent with the results obtained by Deaton (2003b) and Sundaram and Tendulkar (2003). The latter authors, together with Deaton and Drèze (2002), also argue that such large decline in poverty is consistent with evidence from employment surveys, the National Accounts and from data on agricultural wages. So, our conclusion is that most of the poverty reduction measured by using NSS data is real, and not simply a statistical artifact due to a change in the survey design.

The relatively small difference between adjusted and unadjusted estimates suggests that most of the reconciliation between 7-day and 30-day reports comes from a bias of the 7-day reports towards the 30-day ones, and not vice versa.

6 Conclusions

In this paper we have described how an inverse probability weighting procedure can be used to recover comparability over time for statistics made otherwise incomparable due to changes in data collection methodology. The adjustment can be applied to estimators of parameters identified by a population moment condition of the form $E\left[m\left(y, \phi_0\right)\right] = 0$, where y is the variable measured in a

non-comparable way in the different surveys. This framework encompasses a broad set of poverty and inequality measures. The estimator requires the existence of a set of auxiliary variables whose reports are not affected by the different survey design, and whose relation with the main variable of interest is stable across the surveys. The reliability of the adjusted estimates depends crucially on the reliability of the necessary identifying assumptions, which should be carefully evaluated by the researcher on a case by case basis. With this caveat, the procedure introduced here should be a very useful tool for the researcher interested in the evolution over time of welfare or aggregate economic indicators, since changes in survey methodology are frequent, and can easily lead to non-comparability issues.

We estimate adjusted poverty counts from the 55th round of the Indian National Sample Survey, a large expenditure survey carried out in 1999-2000, for which comparability issues arose due to changes in the adopted questionnaire. The identifying assumptions needed for the good performance of our estimator involve unobserved variables, and therefore cannot be directly tested. However, by using previous waves of the NSS, we provide indirect evidence supporting to a large extent their validity. According to our estimates, in 1999-2000 the poor accounted for approximately 30 percent of the rural population, and 25 percent of the urban one. Even if these figures are slightly higher than the unadjusted ones, they still show an impressive poverty decline in the nineties, since the previous estimates—from the 1993-94 round of the NSS—were approximately 30 percent higher.

As a *caveat* to the empirical application, we stress that our results should not be interpreted to suggest the superiority of one set of recall period for expenditure over the others. When the identifying assumptions hold, the reweighting procedure will recover statistics that are *comparable* with others calculated previously using a different methodology. To ascertain *which* questionnaire type is more appropriate is an important task, but one that we did not address in this paper.

Another cautionary note is the fact the identifying assumptions will typically *not* hold if the target and the auxiliary surveys refer to very different populations, so, for example, our methodology is unlikely to be useful if one needs to make cross-country comparisons of welfare indicators, or if the two surveys are separated by a wide temporal gap.

Even if our emphasis on comparability issues over time is justified by the empirical application, the inverse probability weighting procedure described here can be fruitfully applied to several other data missing problems. The asymptotic results will be particularly useful when such problems arise in the context of data collected with multi-stage surveys.

One important application is the estimation of welfare measures for small areas, like villages

or towns. Household surveys seldom contain enough observations from small areas to allow the estimation of poverty or inequality measures with acceptable precision. To recover precise estimates of welfare measures, Elbers, Lanjouw and Lanjouw (2003) propose to merge information on auxiliary variables from a census (which commonly does not record expenditure) with data on the same variables *and* expenditure from a household survey. However, their estimator requires fairly complicated simulation techniques. An inverse probability weighting estimator provides a simpler alternative, without requiring stronger identifying assumptions. One can easily apply the methodology described here, using the census as the target population, and household surveys as auxiliary data.

Inverse probability weighting can also be used to analyze method of moments models in presence of measurement error, when auxiliary data are present. So, suppose the researcher has to estimate ϕ in the moment condition $E\left[\mathbf{m}\left(\mathbf{y}, \phi\right)\right] = \mathbf{0}$, but $\mathbf{y}$ is measured with error. Suppose that an auxiliary database is available where both $\mathbf{y}$ and a set of auxiliary variables $\mathbf{v}$ are measured without error. If the conditional expectation of $\mathbf{m}\left(\mathbf{y}, \phi\right)$ given $\mathbf{v}$ is the same in the the primary and in the auxiliary databases, then ϕ can be recovered using a reweighting procedure, even in the context of *arbitrary* correlation between $\mathbf{y}$ and its mismeasured counterpart in the primary sample. A similar framework is studied in Chen, Hong and Tamer (2003), who propose an alternative, semi-parametric sieve based estimator.

References

[1] Abadie, Alberto (2003). "Semiparametric difference-in-differences estimators". *Review of Economic Studies*, forthcoming.

[2] Amemiya, Takeshi (1985), *Advanced Econometrics* (Cambridge, MA, Harvard University Press).

[3] Attanasio, Orazio, Erich Battistin, and Hidehiko Ichimura (2003), "What really happened to consumption inequality in the US?" (Mimeo, University College London).

[4] Battistin, Erich (2003), "Errors in survey reports of consumption expenditures" (Mimeo, Institute for Fiscal Studies. London).

[5] Battistin, E., Miniaci, R. and Weber, G. (2003), "What do we learn from recall consumption data?", *Journal of Human Resources*, **38** (2), 354–85.

[6] Bhattacharya, Debopam (2003), "Asymptotic Inference from multi-stage samples", *Journal of Econometrics*, forthcoming.

[7] Browning, Martin, Thomas F. Crossley, and Guglielmo Weber (2003), "Asking consumption questions in general purpose surveys", *Economic Journal*, **113**, F540-F567.

[8] Case, Anne, and Angus S. Deaton (2002), "Consumption, health, gender, and poverty", (mimeo, Research Program in Development Studies, Princeton University).

[9] Chen, Xiaohong, Han Hong, and Elie Tamer (2003), "Measurement error models with auxiliary data" (Mimeo).

[10] Datt, Gaurav, Martin Ravallion, and Valerie Kozel (2003), "A model based assessment on India's progress in reducing poverty in the 1990s", *Economic and Political Weekly*, January 25, 355-361.

[11] Deaton, Angus S. (1997), *The analysis of household surveys: a microeconometric approach to development policy* (Baltimore, MD, Johns Hopkins University Press).

[12] Deaton, Angus S. (2001), "Survey design and poverty monitoring in India", (mimeo, Research Program in Development Studies, Princeton University).

[13] Deaton, Angus S. (2003a), "Prices and poverty in India, 1987-2000", *Economic and Political Weekly*, January 25, 362-368.

[14] Deaton, Angus S. (2003b), "Adjusted Indian poverty estimates for 1999-2000", *Economic and Political Weekly*, January 25, 322-326.

[15] Deaton, Angus S., and Jean Drèze (2002), "Poverty and inequality in India, a re-examination", *Economic and Political Weekly*, September 7, 3729-3748.

[16] Deaton, Angus S., and Margaret Grosh (2000), "Consumption," Chapter 5 in Margaret Grosh and Paul Glewwe, eds., *Designing household survey questionnaires for developing countries: lessons from 15 years of the Living Standards Measurement Study* (Oxford University Press for the World Bank, Vol 1., 91–133).

[17] Deaton, Angus S., and Alessandro Tarozzi (2000), "Prices and poverty in India," (Research Program in Development Studies Working Paper no. 196, Princeton University).

[18] Di Nardo, John, Nicole M. Fortin, and Thomas Lemieux (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: a Semiparametric Approach", *Econometrica*, **64** (5), 1001-1044.

[19] Elbers, Chris, Jean O. Lanjouw, and Peter Lanjouw (2003) "Micro-level estimation of poverty and inequality", *Econometrica*, **71** (1), 355-364.

[20] Fan, Jianquing (1992), "Design-adaptive Nonparametric Regression", *Journal of the American Statistical Association*, **87**, 998-1004.

[21] Ghose and Bhattacharya (1995), "Effects of Reference Period on Engel Elasticities of Clothing and Other Items: Further Results", *The Indian Journal of Statistics*, **57**, series B, 3, 433-449.

[22] Gibson, John (1999), "How Robust are Poverty Comparisons to Changes in Household Survey Methods? A Test Using Papua New Guinea Data" (Mimeo, University of Waikato).

[23] Gibson, John (2002), "Why Does the Engel Method Work? Food Demand, Economies of Size, and Household Survey Methods", *Oxford Bulletin of Economics and Statistics*, **64** (4), 341-359.

[24] Gibson, John, Jikun Huang and Scott Rozelle (2001), "Why is income inequality so low in China compared to other countries? The effect of household survey methods", *Economics Letters*, **71** (3), 329-333.

[25] Gibson, John, Jikun Huang and Scott Rozelle (2003), "Improving estimates of inequality and poverty from urban China's Household Income and Expenditure Survey", *Review of Income and Wealth*, forthcoming.

[26] Government of India (1993), *Report of the Expert Group on the estimation of proportion and number of poor* (Delhi, Planning Commission).

[27] Hirano, Keisuke, Guido Imbens, and Geert Ridder (2003), "Efficient estimation of average treatment effects using the estimated propensity score", *Econometrica*, **71** (4), 1161-1189.

[28] Horowitz, Joel L., and Charles F. Manski (1995), "Identification and Robustness with Contaminated and Corrupted Data", *Econometrica*, **63** (2), 281-302.

[29] Jolliffe, Dean (2001), "Measuring absolute and relative poverty: the sensitivity of estimated household consumption to survey design", *Journal of Economic and Social Measurement*, **27**, 1-23.

[30] Lanjouw, Jean Olson, and Peter Lanjouw (2001), "How to Compare Apples and Oranges: Poverty Measurement based on Different Definition of Consumption", *Review of Income and Wealth*, **47** (1), pp. 25-42.

[31] Mahalanobis, P. C., and S. B. Sen (1954), "On some aspects of the Indian National Sample Survey," *Bulletin of the International Statistical Institute*, **34**, part II.

[32] National Sample Survey Organization (2003), "Stability of different recall period for measuring household consumption: results of a pilot study", *Economic and Political Weekly*, January 25, 307-321.

[33] Newey, Whitney, and Daniel McFadden (1994), "Large Sample Estimation and Hypothesis Testing", in *Handbook of Econometrics*, Volume IV, R.F. Engle and D.L. McFadden eds. (Elsevier).

[34] Pagan, Adrian, and Aman Ullah (1999), *Nonparametric Econometrics* (Cambridge, Cambridge University Press).

[35] Rao J. N. K., and A. J. Scott (1984), "On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data", Annals of Statistics, **12**, 46-60.

[36] Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao (1994). "Estimation of regression coefficients when some regressors are not always observed". *Journal of the American Statistical Association*, **89**, 427, 846-866.

[37] Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao (1995). "Analysis of semiparametric regression models for repeated outcomes in the presence of missing data". *Journal of the American Statistical Association*, **90**, 429, 106-121.

[38] Rosenbaum, Paul R., and Donald B. Rubin (1983). "The central role of the propensity score in observational studies for causal effects". *Biometrika*, **70** (1), 41-55.

[39] Sachs, Jeffrey D., Ashutosh Varshney, and Nirupam Bajpai (1999), *India in the Era of Economic Reforms* (Oxford, Oxford University Press).

[40] Scott, Chris, and Ben Amenuvegbe (1990), "Effects of Recall Duration on Reporting of Household Expenditures: An Experimental Study in Ghana", Social Dimensions of Adjustment in Sub-Saharan Africa Working Paper No. 6 (Washington, DC, The World Bank).

[41] Sen, Abhijit (2000), "Estimates of Consumer Expenditure and Its Distribution: Statistical Priorities after NSS 55th Round", *Economic and Political Weekly*, December 16, 4499-4518.

[42] Silverman, Bernard W. (1986), *Density Estimation for Statistics and Data Analysis* (London, Chapman and Hall).

[43] Sundaram, K and Suresh Tendulkar (2002), "Recent Debates on Data Base for Measurement of Poverty in India", (Mimeo, Delhi School of Economics). Available at http://www.worldbank.org/indiapovertyworkshop.

[44] Sundaram, K and Suresh D Tendulkar (2003), "Poverty *Has* Declined in the 1990s: A Resolution of Comparability Problems of NSS on Consumer Expenditure", *Economic and Political Weekly*, January 25, 322-326.

[45] Visaria, Pravin (2000), "Poverty in India during 1994–98: alternative estimates," (Mimeo, Institute for Economic Growth, New Delhi).

[46] Wooldridge, Jeffrey M. (1999). "Asymptotic properties of weighted M-estimators for variable probability samples", *Econometrica*. **67**, 1385-1406.

[47] Wooldridge, Jeffrey M. (2002a). *Econometric analysis of cross-section and panel data* (Cambridge, MA, MIT Press).

[48] Wooldridge, Jeffrey M. (2002b). "Inverse probability weighted M-estimators for sample selection, attrition, and stratification", *Portuguese Economic Journal*, **1**, 117-139.

Appendix

Proof of Proposition 1

Using the law of iterated expectations, and A2, we can rewrite the initial moment condition as

$$E\left[nm\left(y,\phi_0\right)\mid\tau=1\right]=\int nE\left[m\left(y,\phi_0\right)\mid\mathbf{v},\tau=0\right]dP\left(\mathbf{v}\mid\tau=1\right)$$

Then, using A3 and A4,

$$
\begin{aligned}
E\left[nm\left(y,\phi_0\right)\mid\tau=1\right] &= \int nE\left[m\left(y,\phi_0\right)\mid\mathbf{v},\tau=0\right]R\left(\mathbf{v}\right)dP\left(\mathbf{v}\mid\tau=0\right)\\
&= E\left[nR\left(\mathbf{v}\right)m\left(y,\phi_0\right)\mid\tau=0\right] \qquad \blacksquare
\end{aligned}
$$

Note that if $\mathbf{v}$ does not include n, A2 should be modified as

$$E\left[nm\left(y,\phi_0\right)\mid\mathbf{v},\tau=0\right]=E\left[nm\left(y,\phi_0\right)\mid\mathbf{v},\tau=1\right].$$

For the identification of $R\left(\mathbf{v}\right)$ it is sufficient to show that $P\left(\tau=1\mid\mathbf{v}\right)$ is identified by the sampling process. First, note that the econometrician observes $\hat{\mathbf{v}}=\tau\tilde{\mathbf{v}}+\left(1-\tau\right)\mathbf{v}$. So, the sampling process identifies $P\left(\tau=1\mid\hat{\mathbf{v}}\right)=P\left(\tau=1,\tilde{\mathbf{v}}\right)/P\left(\hat{\mathbf{v}}\right)$. Applying Bayes Theorem,

$$
\begin{aligned}
P\left(\tau=1\mid\hat{\mathbf{v}}\right) &= \frac{P\left(\tilde{\mathbf{v}}\mid\tau=1\right)P\left(\tau=1\right)}{P\left(\tilde{\mathbf{v}}\mid\tau=1\right)P\left(\tau=1\right)+P\left(\mathbf{v}\mid\tau=1\right)P\left(\tau=0\right)}\\
&= P\left(\tau=1\mid\mathbf{v}\right)\quad\text{a.s.}
\end{aligned}
$$

where the last equality follows from A1.

Proof of Proposition 2

Let $\mathbf{v}=\begin{bmatrix} n & \mathbf{v}_{-n}^{T} \end{bmatrix}^{T}$.

$$
\begin{aligned}
f\left(y\mid n,\tau=1\right) &= \int_{\mathbf{v}_{-n}}dP\left(y,\mathbf{v}_{-n}\mid n,\tau=1\right)=\int_{\mathbf{v}_{-n}}f\left(y\mid\mathbf{v}_{-n},n,\tau=1\right)dP\left(\mathbf{v}_{-n}\mid n,\tau=1\right)\\
\text{by A2b} &= \int_{\mathbf{v}_{-n}}f\left(y\mid\mathbf{v}_{-n},n,\tau=0\right)dP\left(\mathbf{v}_{-n}\mid n,\tau=1\right)\\
&= \int_{\mathbf{v}_{-n}}R\left(\mathbf{v}_{-n}\right)f\left(y\mid\mathbf{v}_{-n},n,\tau=0\right)dP\left(\mathbf{v}_{-n}\mid n,\tau=0\right)\\
&= \int_{\mathbf{v}_{-n}}R\left(\mathbf{v}_{-n}\right)dP\left(y,\mathbf{v}_{-n}\mid n,\tau=0\right)\\
&= f\left(y\mid n,\tau=0\right)E\left[R\left(\mathbf{v}_{-n}\right)\mid y,n,\tau=0\right] \qquad \blacksquare
\end{aligned}
$$

Note that here the reweighting function has a different form, since all probabilities are now conditional. So

$$
\begin{aligned}
R\left(\mathbf{v}_{-n}\right) &= \frac{dP\left(\mathbf{v}_{-n},n,\tau=1\right)}{dP\left(n,\tau=1\right)}\frac{dP\left(n,\tau=0\right)}{dP\left(\mathbf{v}_{-n},n,\tau=0\right)}\\
&= \frac{P\left(\tau=1\mid\mathbf{v}_{-n},n\right)dP\left(\mathbf{v}_{-n},n\right)}{P\left(\tau=0\mid\mathbf{v}_{-n},n\right)dP\left(\mathbf{v}_{-n},n\right)}\frac{P\left(\tau=0\mid n\right)dP\left(n\right)}{P\left(\tau=1\mid n\right)dP\left(n\right)}=\frac{P\left(\tau=1\mid\mathbf{v}\right)P\left(\tau=0\mid n\right)}{P\left(\tau=0\mid\mathbf{v}\right)P\left(\tau=1\mid n\right)}
\end{aligned}
$$

The complete Proposition 3, and proof.

First note that the sampling is done from a population that encompasses both the auxiliary and the target survey. The population is divided into strata, which are common to both sub-populations. A random sample of clusters is then drawn from each stratum. This allows to treat the variable τ_{sih}, the dummy equal to one when a household belongs to the target population, in the same way as all other variables involved in the estimation. Define $\mathbf{z}_{sch} = (y_{sih}, \mathbf{v}_{sih}, \tau_{sih}, n_{sch})$. Let $\mathbf{m}_i(z; \boldsymbol{\beta})$ be defined as in (7):

$$\mathbf{m}_i(\boldsymbol{\beta}) = \sum_{s=1}^{S} 1(i \in s) \sum_{h=1}^{m_{si}} \omega_{sih} \mathbf{m}(y_{sih}, \mathbf{v}_{sih}, \tau_{sih}; \boldsymbol{\beta})$$

A0 For $s, s' = 1, ..., S$, $(\mathbf{z}_{sch}, \mathbf{z}_{s'c'h'})$ are independent unless $s = s'$ and $c = c'$. For each s, $(\mathbf{z}_{sch})$ are identically distributed. For $s \neq s'$, $\mathbf{z}_s$ and $\mathbf{z}_{s'}$ are independent (but not necessarily identically distributed), where $\mathbf{z}_s \equiv \{\mathbf{z}_{sch}\}_{c=1,...,n_s; h=1,...,m_{sc}}$.

A1 $m_i^j(\mathbf{z}; \boldsymbol{\beta})$ is continuous at each $\boldsymbol{\beta} \in \boldsymbol{\Theta}$ w.p. 1, for each $j = 1, ..., J$.

A2 $\exists d(.)$ with $E[d(.)] < \infty$ such that $\left\| m_i^j(\mathbf{t}; \boldsymbol{\beta}) \right\| \leq d(\mathbf{t})$ for each $j = 1, ..., J$ for all $\mathbf{t}$.

A3 The parameter space $\boldsymbol{\Theta}$ is compact.

A4 $\boldsymbol{\beta}_0 \in int(\boldsymbol{\Theta})$, and $\boldsymbol{\beta}_0$ solves (5)—in the text—uniquely.

A5 $E[\mathbf{m}_i(z; \boldsymbol{\beta})]$ is continuously differentiable at $\boldsymbol{\beta}_0$ and Γ is non-singular, where

$$\Gamma = p\lim \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}^T} E(\mathbf{m}_i(\boldsymbol{\beta}_0))$$

A6 The sequence $\nu_n(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\mathbf{m}_i(\boldsymbol{\beta}) - E(\mathbf{m}_i(\boldsymbol{\beta}))]$ is stochastically equicontinuous.

A7 $\sup_{\boldsymbol{\beta} \in \boldsymbol{\Theta}} E|\mathbf{m}_i(\boldsymbol{\beta})|^3 < \infty$

A8 $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Var(\mathbf{m}_i(\boldsymbol{\beta})) = W < \infty$

Under assumptions A0-A8,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\left(0, \Gamma^{-1}W(\Gamma^{-1})^T\right)$$

Proof: It is sufficient to note that assumptions A1-A8 correspond to assumptions A0-A7 and A8b in Bhattacharya (2003). Then the conclusion follows from his Proposition 2, for the case where the number of moments is equal to the number of estimated parameters.

The robust asymptotic variance for a headcount poverty ratio

In the estimation of poverty counts, the expression in (2) specializes to

$$E[nR(\mathbf{v})(1(y < z) - \phi_0) \mid \tau = 0] = 0$$

where $R(\mathbf{v}) = \left[P(\tau = 0 \mid \mathbf{v})(1 - \theta_0^0)\right]^{-1} P(\tau = 1 \mid \mathbf{v})\theta_0^0$, and $\theta_0^0 = P(\tau = 0)$. The above expression is also equivalent to

$$E\left[nR(\mathbf{v})\mathbf{1}(y < z) - \phi_0\eta^1 \mid \tau = 0\right] = 0$$

where $\eta^1 = E[n \mid \tau = 1]$ is average household size in the target survey. We assume a parametric single index model for the propensity score. Let $\bar{\mathbf{v}}$ indicate the column vector of all auxiliary variables and their powers and interactions, as entered in the first step regression (including the constant). Then $P(\tau = 1 \mid \mathbf{v}) = P(\bar{\mathbf{v}}^T\boldsymbol{\theta}_0^1)$. We estimate the parameters $\boldsymbol{\theta}_0^1$ using maximum likelihood. Let $\boldsymbol{\beta} \equiv [\phi \ \boldsymbol{\theta}^{1T} \ \theta^0 \ \eta^1]^T$. Then $\mathbf{m}(y, \mathbf{v}, \tau; \boldsymbol{\beta}_0)$ in (6) becomes

$$\mathbf{m}(y, \mathbf{v}, \tau; \boldsymbol{\beta}_0) = \begin{bmatrix} \mathbf{1}(\tau = 0)\left[n\dfrac{P(\bar{\mathbf{v}}^T\boldsymbol{\theta}_0^1)\theta_0^0}{\left[1 - P(\bar{\mathbf{v}}^T\boldsymbol{\theta}_0^1)\right](1 - \theta_0^0)}\mathbf{1}(y < z) - \phi_0\eta_0^1\right] \\ \dfrac{\tau - P(\bar{\mathbf{v}}^T\boldsymbol{\theta}_0^1)}{P(\bar{\mathbf{v}}^T\boldsymbol{\theta}_0^1)\left[1 - P(\bar{\mathbf{v}}^T\boldsymbol{\theta}_0^1)\right]}\dfrac{\partial P(\bar{\mathbf{v}}^T\boldsymbol{\theta}_0^1)}{\partial(\bar{\mathbf{v}}^T\boldsymbol{\theta}_0^1)}\bar{\mathbf{v}} \\ \left[(1 - \tau) - \theta_0^0\right] \\ \mathbf{1}(\tau = 1)\left[n - \eta_0^1\right] \end{bmatrix} \quad (11)$$

The second row of (11) represents the F.O.C. from the maximum likelihood estimation of $\boldsymbol{\theta}_0^1$. Note that in (11) household size appears *within* $\mathbf{m}(.)$, so that when calculating (9) and (10) one has to use sampling weights for households, and not individuals. In the empirical application we use a logistic model for the propensity score, and then

$$\mathbf{m}\left(y_{sih}, \mathbf{v}_{sih}, \tau_{sih}; \hat{\boldsymbol{\beta}}\right) = \begin{bmatrix} \mathbf{1}(\tau_{sih} = 0)\left[n_{sih}R\left(\mathbf{v}_{sih}; \hat{\boldsymbol{\theta}}^1, \hat{\theta}^0\right)\mathbf{1}(y_{sih} < z) - \hat{\phi}\hat{\eta}^1\right] \\ \left[\tau_{sih} - P\left(\bar{\mathbf{v}}_{sih}^T\hat{\boldsymbol{\theta}}^1\right)\right]\bar{\mathbf{v}}_{sih} \\ \left[(1 - \tau_{sih}) - \hat{\theta}^0\right] \\ \mathbf{1}(\tau_{sih} = 1)\left[n_{sih} - \hat{\eta}^1\right] \end{bmatrix}$$

where $R\left(\mathbf{v}_{sih}; \hat{\boldsymbol{\theta}}^1, \hat{\theta}^0\right) = \left[\left(1 - P\left(\bar{\mathbf{v}}_{sih}^T\hat{\boldsymbol{\theta}}^1\right)\right)\left(1 - \hat{\theta}^0\right)\right]^{-1} P\left(\bar{\mathbf{v}}_{sih}^T\hat{\boldsymbol{\theta}}^1\right)\hat{\theta}^0$. These expressions are used directly to estimate the matrix (10). An estimate for the matrix Γ in (9) is obtained using the following:

$$\hat{\Gamma} = \begin{bmatrix} \hat{\gamma}_{11} & \hat{\boldsymbol{\gamma}}_{12} & \hat{\gamma}_{13} & \hat{\gamma}_{14} \\ \mathbf{0} & \hat{\boldsymbol{\gamma}}_{22} & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{0} & \hat{\gamma}_{33} & 0 \\ 0 & \mathbf{0} & 0 & \hat{\gamma}_{44} \end{bmatrix}$$

where (omitting, for simplicity, the subscripts)

$$\hat{\gamma}_{11} = -\mathbf{1}(\tau = 0)\hat{\eta}_1 \qquad\qquad \hat{\boldsymbol{\gamma}}_{22} = -P\left(\bar{\mathbf{v}}^T\hat{\boldsymbol{\theta}}_1\right)\left[1 - P\left(\bar{\mathbf{v}}^T\hat{\boldsymbol{\theta}}_1\right)\right]\bar{\mathbf{v}}\bar{\mathbf{v}}^T$$

$$\hat{\boldsymbol{\gamma}}_{12} = \mathbf{1}(\tau = 0)nR\left(\mathbf{v}; \hat{\boldsymbol{\theta}}^1, \hat{\theta}^0\right)\mathbf{1}(y < z)\bar{\mathbf{v}}^T \quad \hat{\gamma}_{33} = -1$$

$$\hat{\gamma}_{13} = \mathbf{1}(\tau = 0)n\dfrac{R\left(\mathbf{v}; \hat{\boldsymbol{\theta}}^1, \hat{\theta}^0\right)}{\hat{\theta}_0(1 - \hat{\theta}_0)}\mathbf{1}(y < z) \qquad \hat{\gamma}_{44} = -1(\tau = 1)$$

$$\hat{\gamma}_{14} = -\mathbf{1}(\tau = 0)\hat{\phi}$$

Finally, in the empirical application we take the states to represent different strata. In the NSS, strata are actually defined at a finer level (typically, districts in the rural sector sectors, and town, or sections of large towns, in urban areas).
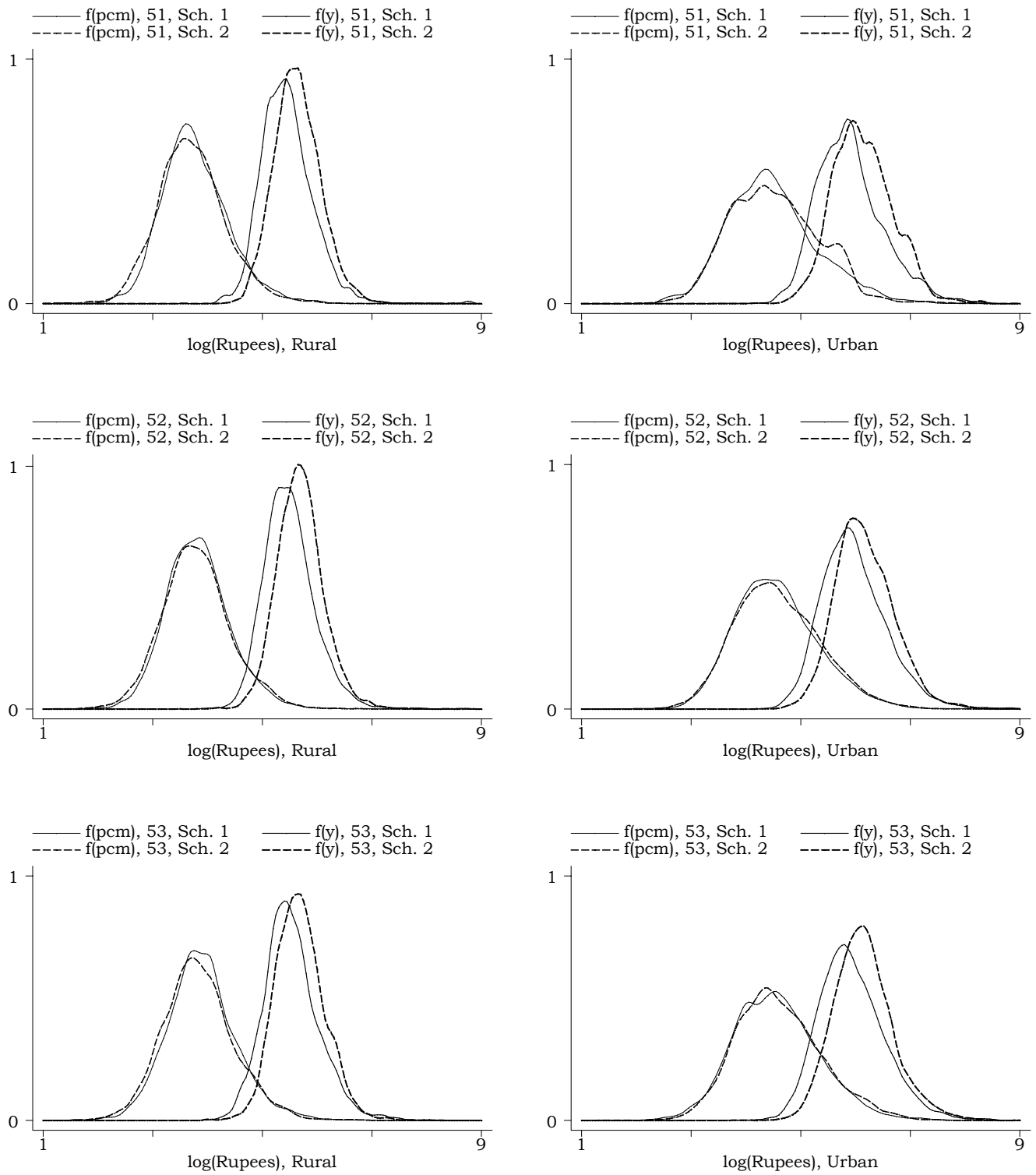
Figure 1 - Source: author's computation from NSS - All major Indian states. In each graph the two rightmost lines represent kernel estimates of the densities of total pce as measured with the two different schedules. The other two lines represent kernel estimates of the densities of pce in 30-day items. All values are in 1993-94 Rupees, deflated using CPIAL for the rural sector, and CPIIW for the urban sector. All densities are estimated using Silverman's robust bandwidth adapted for a biweight kernel, which is the one we use here.
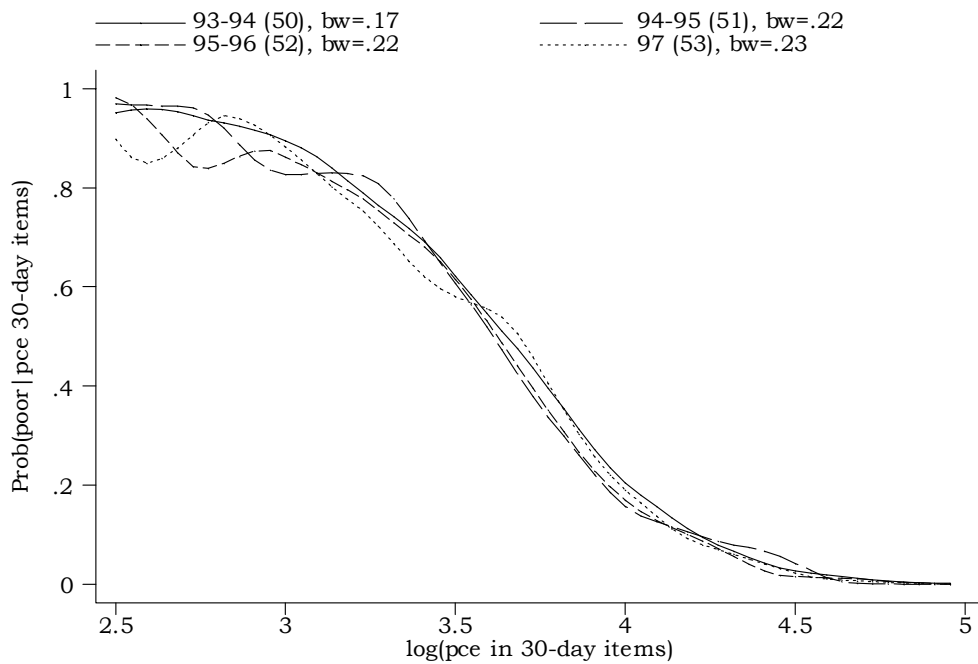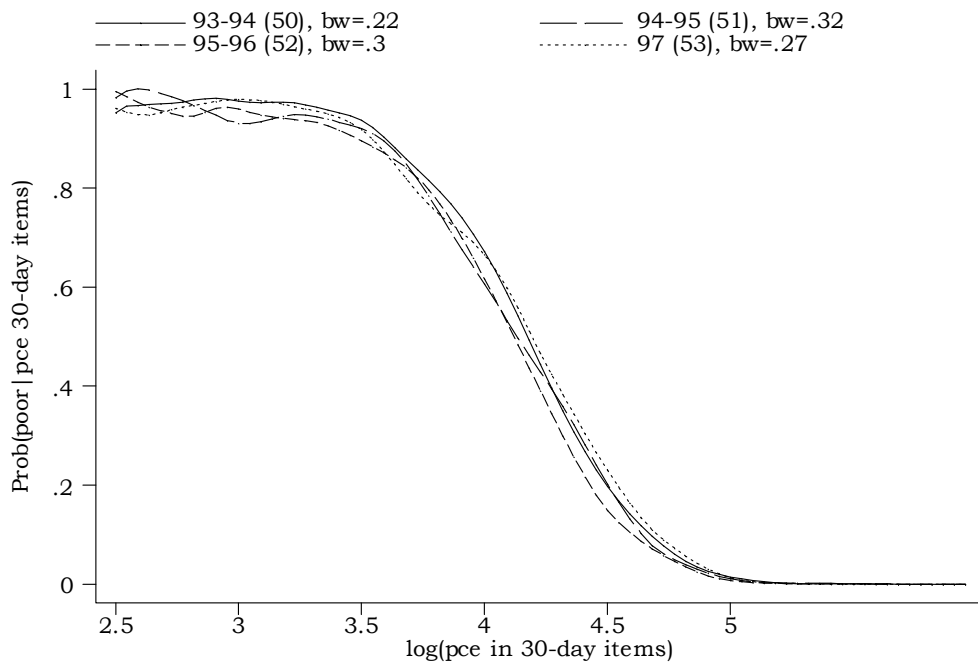
Figure 2 - Conditional probability of being poor given log(PCE in 30-day items)
Source: author's computations from NSS, rounds 50-53, all major Indian States. Locally weighted regressions. The dependent variable is a dummy variable equal to one when the household's per capita monthly expenditure is below the sector-specific poverty line. The bandwidth used in each line is indicated in the corresponding label.

Table 1 - Summary statistics, Indian NSS Rounds 51-54

		NSS 51				NSS 52				NSS 53				NSS 54			
Deflator is CPIIW for urban sector and CPIAL for rural sector		July 1994 - June 1995				July 1995 - June 1996				January-December 1997				January-June 1998			
		rural		urban		rural		urban		rural		urban		rural		urban	
Schedule (Questionnaire Type)		S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
(1)	Sample size (households)	13606	13415	9283	9214	12253	12047	8870	8749	12313	9214	16418	10555	8676	8545	2946	2911
(2)	Mean per capita total monthly expenditure	273.4	310.5	461.6	545.9	278.8	328.1	495.3	560.1	294.1	328.2	323.16	377.43	265.83	316.22	457.27	520.77
	s.e.	(13.17)	(3.45)	(18.78)	(20.05)	(3.91)	(3.35)	(10.84)	(7.54)	(3.25)	(3.21)	(5.21)	(5.86)	(2.98)	(2.78)	(14.44)	(21.35)
	t-ratio for equality across sch.	2.73		3.07		9.57		4.91		7.46		6.92		12.37		2.46	
(3)	Mean per capita expenditure in 30-day items (in 1993-94 Rs.)	54.4	50.8	131.4	133.4	57.9	56.6	130.4	138.0	61.9	58.9	91.8	96.3	58.0	57.3	131.8	143.2
	s.e.	(1.53)	(0.80)	(7.75)	(7.64)	(2.46)	(1.07)	(2.28)	(2.81)	(0.79)	(0.93)	(2.49)	(2.32)	(0.78)	(0.83)	(4.67)	(17.79)
	t-ratio for equality across sch.	2.09		0.18		0.48		2.10		2.46		1.34		0.56		0.62	
(4)	Mean budget share of 30-day items	20.1	16.0	26.3	21.7	20.3	16.4	26.1	22.4	21.1	17.3	27.3	23.4	21.6	17.4	28.2	22.8
	s.e.	(0.22)	(0.17)	(0.40)	(0.34)	(0.16)	(0.14)	(0.17)	(0.18)	(0.23)	(0.22)	(0.21)	(0.21)	(0.16)	(0.13)	(0.41)	(0.36)
(5)	Headcount Poverty Ratio	41.8	22.7	36.3	18.5	38.2	18.4	30.7	15.4	35.7	21.1	33.1	17.5	41.8	22.4	35.3	21.3

Source: author's computations from NSS. Robust standard errors in parenthesis. All values in 1993-94 Rupees. Only the major Indian states are included. All statistics are weighted using inflation factors. S1 is the standard questionnaire, with a 30-day reference period for all items. S2 is the experimental questionnaire, with a 7-day recall period for food and other high frequency items, and a 365day recall period for durables, clothing, footwear, educational and institutional medical expenses. The category "30-day items" includes: fuel and light, miscellaneous consumer goods and services, rents, consumer taxes and non-institutional medical expenses. The mean budget shares are averages of household-specific ratios between expenditure in 30-day items and total expenditure. The poverty counts are the proportion of individuals living in households where per capita expenditure is below the poverty line. The poverty lines are the official ones published by the Planning Commission for 1993-94, and are expressed in 1993-94 Rupees (the exchange rate at the time was approximately (1US$/32 Rs). The poverty line for All India is Rs 205.7 for the rural sector, and Rs 283.4 for the urban sector.

Table 2 – Tests for equality of distributions across schedules

(1) p-values	51st round		52nd round		53rd round	
	Rural	Urban	Rural	Urban	Rural	Urban
Household size	0.200	0.486	0.438	0.569	0.350	0.219
Education of head	0.837	0.651	0.577	0.533	0.014	0.208
Main activity	0.007	0.442	0.302	0.479	0.009	0.253
Scheduled caste/tribe	0.235	0.919	0.333	0.698	0.473	0.541
Land holdings	0.738	0.262	0.867	0.491	0.373	0.330

(2) Cross-tabulation of covariates for cells in panel (1) with p-value<0.05

	NSS 51, Rural		NSS 53, Rural			NSS 53, Rural	
	Sch. 1	Sch. 2	Sch. 1	Sch. 2		Sch. 1	Sch. 2
self-employed in non-agriculture	0.156	0.126	0.128	0.113	illiterate	0.490	0.486
agricultural labor	0.266	0.305	0.275	0.319	lit. no schooling	0.050	0.035
other labor	0.051	0.059	0.068	0.059	lit. below primary	0.114	0.095
Self-empl. in agr. and others	0.527	0.510	0.530	0.509	primary	0.131	0.141
					middle	0.111	0.133
					secondary	0.058	0.052
					above secondary	0.046	0.058

Source: author's computations from Indian NSS, rounds 51-53. All major Indian states.

The figures in panel (1) are p-values for Pearson chi-squared statistics corrected for the survey design. The null hypothesis is that, for every sector-round pair, the distribution of the selected variable is the same across the two different schedules. "Education of head" is one of the following: illiterate, literate with no schooling, literate below primary, primary, middle, secondary, above secondary. In the rural sector "Main activity of the household" is one of the following: self-employed in non-agriculture, agricultural labor, other labor, self-employed in agriculture and others; in the urban sector the categories are: self-employed, regular wage/salaried, casual labor, others. "Land holdings" are recorded as a categorical variable, with different codes for different intervals: code 1 is for land holdings below 0.01 acres, code 2 for the interval [0.01,0.2) and so on.

Table 3   -   Tests for equality of distribution of log(pce in 30-day items) across schedules
p-values - design based adjusted F-tests

NSS Round		obs.	number of bins		
			10	15	20
51st - 7/94-6/95	Rural	26339	0.0281	0.0570	0.0931
	Urban	18168	0.0483	0.2698	0.1455
52nd - 7/95-6/96	Rural	23682	0.0426	0.0307	0.0770
	Urban	17224	0.4220	0.2770	0.6345
53rd - 1/97-12/97	Rural	20819	0.3555	0.3922	0.5637
	Urban	26119	0.4352	0.2231	0.2841

Source: author's computations from Indian NSS, rounds 51-53, all major Indian states.
The null hypothesis is that the distribution across bins is the same across the two different schedules. To avoid the presence of cells with very few observation, for every round-sector we use only observations included in the range between the first and the last centile of the round-sector specific distribution. All tests take into account the presence of clustering and stratification, except in the 51st round, for which we have no information on strata.

Table 4 - Tests for equality of conditional probabilities of being poor across different surveys

Surveys compared		50 - 51	50 - 52	50 - 53	51 - 52	51 - 53	52 - 53
		Rural Sector					
	obs.	72528	71175	71235	25859	25919	24566
m, household size[a]	$\chi^2(3)$	2.87 [0.4126]	9.36 [0.0248]	2.46 [0.4830]	0.58 [0.9010]	0.41 [0.9391]	1.41 [0.7031]
Polynomials in m and hh. size	$\chi^2(7)$	48.51 [0.0000]	13.2 [0.0674]	14.32 [0.0458]	20.94 [0.0039]	24 [0.0011]	3.66 [0.8182]
m, hh. size, controls[b]	$\chi^2(23)$	24.73 [0.3642]	36.22 [0.0392]	32.63 [0.0878]	28.54 [0.1962]	24.18 [0.3941]	22.19 [0.5089]
Polynomial in hh. size, controls	$\chi^2(24)$	29.15 [0.2145]	25.16 [0.3971]	35.78 [0.0576]	27.53 [0.2803]	42.53 [0.0113]	25.49 [0.3797]
Polynomial in m and h. size, controls	$\chi^2(27)$	80.77 [0.0000]	37.16 [0.0922]	38.88 [0.0650]	48.04 [0.0076]	53.18 [0.0019]	23.81 [0.6407]
		Urban Sector					
	obs.	47995	47582	55130	18153	25701	25288
m, household size	$\chi^2(3)$	5.06 [0.1672]	25.23 [0.0000]	1.21 [0.7512]	3.26 [0.3536]	2.64 [0.4500]	23.29 [0.0000]
Polynomials in m and hh. size	$\chi^2(7)$	5.91 [0.5508]	24.44 [0.0010]	7.82 [0.3491]	4.82 [0.6822]	6.22 [0.5141]	27.72 [0.0002]
m, hh. size, controls	$\chi^2(23)$	33.04 [0.0804]	44.15 [0.0050]	39.86 [0.0159]	31.18 [0.1185]	28.17 [0.2093]	41.08 [0.0116]
Polynomial in hh. size, controls	$\chi^2(24)$	36.34 [0.0509]	32.12 [0.1241]	39.71 [0.0230]	41.45 [0.0149]	29.48 [0.2025]	36.82 [0.0456]
Polynomial in m and h. size, controls	$\chi^2(27)$	32.54 [0.2127]	38.66 [0.0680]	39.38 [0.0585]	31.09 [0.2675]	28.31 [0.3952]	45.49 [0.0144]

Source: author's computations from NSS, rounds 50-53. Major Indian states only. The tests are robust to the presence of arbitrary heteroskedasticity and correlation within clusters. P-values in parenthesis.

[a] m is (log) per capita expenditure in miscellaneous items, in 93-94 Rs

[b] The controls are categorical variables for education of the household head, main economic activity of the household, whether the household belongs to a "scheduled caste or tribe", and land ownership.

Table 5 - Adjusted and unadjusted poverty rates in thin rounds using a logit first step

			(1)	(2)		(3)	(4)	(5)	(6)
			Schedule 1 (Benchmark)	Schedule 2 Unadjusted		**Auxiliary Survey** - Schedule 2 - adjusted			
			H1	H2		50	51 - Sch. 1	52 - Sch. 1	53 - Sch. 1
Target survey (Sch. 2)	NSS 51 - 94/95	Rural	41.8	22.7	m only	46.4 (2.83)**	45.6 (1.70)	45.3 (1.77)	44.6 (1.22)
		s.e.	(1.42)	(1.19)	all except m	38.0 (2.55)**	41.7 (0.13)	37.8 (2.31)*	35.3 (3.42)**
					all	45.5 (2.27)*	45.3 (1.55)	45.0 (1.65)	43.9 (0.9)
		Urban	36.3	18.5	m only	36.8 (0.27)	33.8 (0.94)	33.4 (1.33)	36.1 (0.06)
		s.e.	(1.88)	(1.34)	all except m	33.4 (1.37)	36.6 (0.14)	32.0 (2.04)*	35.0 (0.56)
					all	36.4 (0.08)	34.7 (0.6)	34.6 (0.77)	36.8 (0.28)
	NSS 52 - 95/96	Rural	38.2	18.4	m only	41.7 (3.43)**	41.0 (1.70)	39.7 (1.31)	39.9 (1.06)
		s.e.	(0.89)	(0.74)	all except m	38.4 (0.23)	41.6 (2.04)*	37.9 (0.16)	35.6 (1.63)
					all	41.7 (3.42)**	41.2 (1.8)	40.4 (1.67)	39.8 (1.02)
		Urban	30.7	15.4	m only	33.0 (2.27)*	30.7 (0.01)	30.4 (0.22)	33.3 (2.19)*
		s.e.	(0.83)	(0.69)	all except m	32.0 (1.31)	35.5 (2.6)**	30.7 (0.01)	33.8 (2.64)**
					all	32.6 (1.88)	31.2 (0.31)	30.8 (0.14)	33.6 (2.46)**
	NSS 53 - 97	Rural	35.7	21.1	m only	40.1 (3.22)**	39.4 (2.06)*	39.0 (2.12)*	38.3 (1.53)
		s.e.	(1.27)	(1.06)	all except m	37.3 (1.3)	40.9 (2.77)**	37.0 (0.95)	34.9 (0.34)
					all	39.9 (3.09)**	39.7 (2.19)*	39.0 (2.1)*	38.4 (1.57)
		Urban	33.1	17.5	m only	32.5 (0.90)	30.3 (1.88)	30.1 (3.01)**	32.9 (0.46)
		s.e.	(0.83)	(0.75)	all except m	31.8 (1.61)	35.1 (0.87)	30.3 (2.8)**	33.9 (0.33)
					all	32.1 (1.3)	30.7 (1.62)	30.3 (2.86)**	33.3 (0.14)

Source: author's calculations from NSS rounds 50-53. Only major Indian states are included. Columns 1 and 2 report robust standard errors in parenthesis. Columns 3 to 6 report in parenthesis the robust t-ratio for the null that the adjusted poverty ratio and the benchmark are the same. ** indicates rejection using 1% significant level, while * indicates rejection using 5% level.

[a] m is (log) per capita expenditure in miscellaneous items, in 93-94 Rs

[b] The controls are categorical variables for education of the household head, main economic activity of the household, whether the household belongs to a "scheduled caste or tribe", and land ownership.

Table 6 - Adjusted Poverty Counts - 55th NSS Round (1999-2000)

Auxiliary survey	(1)	(2)
Rural	m [a]	m and controls[b]
50 - 7/93-6/94	31.8 (0.34)	32.5 (0.35)
51 - 7/94-6/95	30.7 (1.05)	32.0 (1.09)
52 - 7/95-6/96	30.4 (0.77)	31.7 (0.80)
53 - 1/97-12/97	30.5 (1.05)	31.0 (1.07)
Urban		
50 - 7/93-6/94	25.9 (0.47)	26.1 (0.47)
51 - 7/94-6/95	24.9 (1.17)	27.1 (1.27)
52 - 7/95-6/96	24.0 (0.54)	25.0 (0.56)
53 - 1/97-12/97	26.5 (0.67)	27.3 (0.68)

Unadjusted Poverty Counts, only larger Indian States, Standard Questionnaire
Rural 28.4 (0.40), Urban 24.5 (0.58)

Source: author's computation from NSS, Rounds 50-51-52-53-55. Robust standard errors in parenthesis.
The poverty lines are the official ones for the 50th round (205.67 for the rural sector, and 283.44 for the urban sector). All monetary values from subsequent Rounds are deflated using state and sector specific official Consumer Price Indexes (CPIAL for households living in rural areas, and CPIIW for those living in urban areas). All estimates are computed for all major Indian states: Andhra Pradesh, Assam, Bihar, Gujarat, Haryana (urban only), Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, West Bengal, Delhi (urban only). The adjusted poverty counts are estimates using the estimator developed in the paper, using a logit first step. In both columns, the logit first step also includes a polynomial in household size.

[a] m is (log) per capita expenditure in miscellaneous items, in 93-94 Rs

[b] The controls are categorical variables for education of the household head, main economic activity of the household, whether the household belongs to a "scheduled caste or tribe", and land ownership.