IARIW 2025

IARIW 2025

Thursday, October 2 & Friday, October 3

Two-Sample Cross-Tabulation: Application to poverty and undernutrition

Tomoki Fujii
(Singapore Management University)

Roy van der Weide
(World Bank)

Paper prepared for the IARIW–World Bank–UEB/VNU Conference on "Improving Well-being Measurement in Data-challenged Environments in Developing Countries for Better Evidence-based Policies" October 2-3, 2025

Session 2A: Measurement Errors
Time Slot: Thursday, October 2, 1:15- 3:15 PM

Two-Sample Cross-Tabulation: Application to poverty and undernutrition

Tomoki Fujii Roy van der Weide

March 25, 2025

Abstract

Conventional one-sample cross-tabulation of two categorical variables assumes that the data are available for the same individuals, households, plots, or firms. This requirement is however not always met. One way to proceed in that case is to cross-tabulate exact observations with imputed values. The resulting two-sample cross-tabulation will be subject to both imputation error and sampling error. The added imputation error increases the standard errors relative to the one-sample analogue, and may introduce a bias. We put forward a bias-corrected estimator and derive its asymptotic properties. This gives the applied user analytic standard errors that are easily implemented without the use of bootstrapping. Monte-Carlo simulations confirm that: (a) the analytic standard errors coincide perfectly with bootstrap standard errors for all sample sizes considered, (b) the contribution of imputation error to total error is not negligible, and (c) the bias-correction works remarkably well. An empirical example is also provided.

1 Introduction

Cross-tabulation is a useful descriptive tool that is routinely used in economics, sociology, political science and other disciplines. It is easy to produce and helpful for understanding the relationship between two discrete variables of interest. Typical units of observation are individuals, households, firms, agricultural plots etc. For ease of exposition we will refer to observations as "households". The only requirement for computing a cross-tabulation is that these two variables are available for the same households, which either means that the variables come from the same survey or that households have been re-visited for different surveys. If the above requirement does not apply, cross-tabulations cannot be produced, at least not by conventional one-sample methods.

It is not uncommon that variables of interest are divided over different samples that cannot be linked with certainty at the household level. Surveys tend to be specialized and are undertaken by different parties. It is often not in the interest of those commissioning the survey to collect data beyond that what is relevant for their program or project. Even if funding and coordination are not an issue, grouping multiple specialized surveys into one so that all data are collected for the same sample of households will be demanding both on those being interviewed and on the enumerators. This may compromise the quality of the data collected. For a recent review study on combining survey data, see e.g. Ridder and Moffitt (2007).

A popular approach, when no one sample contains both variables of interest, is to pick one sample and impute the variable that is missing. It is assumed that the samples have a set of variables in common in addition to the variables that are unique to each sample. The shared variables then serve as instruments in the model used for imputing the missing variable. This means that exact observations (for variable one) will be cross-tabulated with imputed values (for variable two). We will refer to this as a two-sample cross-tabulation.

Two-sample cross-tabulations are subject to two types of error: imputation error and sampling error. The added imputation error will add to the standard error of the estimates. It will generally also introduce a bias, as the imputation error is likely to be correlated with the variable with which it is cross-tabulated. The dual error structure is often ignored in empirical applications leading to an over-estimation of statistical precision of cross-tabulations that feature imputed values (examples will be given below).

This paper puts forward a bias-corrected estimator and derives its asymptotic distribution. The asymptotic distribution takes into account both the imputation- and the sampling error. These analytic results enable the user to compute standard errors without having to resort to bootstrap simulations. To ensure exact standard errors for any sample size and sampling design we also offer a bootstrap procedure. Monte Carlo simulations suggest that the asymptotic- and bootstrap standard errors are equally accurate also for small sample sizes; both match the true standard errors. Having analytic standard errors that are both easy to compute and accurate makes two-sample cross-tabulation a user-friendly tool for a wide audience of applied users.

The contexts in which two-sample cross-tabulations can be applied are wide-ranging. Two prominent examples of specialized surveys are the household budget survey (HBS) and the demographic health survey (DHS). The HBS collects detailed household expenditure data that is commonly used to determine a household's poverty status. The DHS collects detailed health and health-care data that includes anthropometric indicators, whether the individual is HIV positive, and useage of health services. While basic demographics, education, asset ownership and dwelling unit characteristics are typically available in both types of surveys, the detailed expenditure and health data are unique to the corresponding surveys. A cross-tabulation of health with poverty would then require a combination of the two surveys. More specifically, one may wish to examine the relationship between socio-economic status and child malnutrition (see e.g. Sahn and Stifel, 2003), child health care (see e.g. Schellenberg et al., 2003), and useage of health services more generally (see e.g. Lindelow, 2006).[1]

Breaking down estimates by region, province, and district too denotes an example of a two-sample cross-tabulation when there is no one sample that both contains the variable of interest and is representative at the desired subnational level. A frequently adopted solution is to bring in a larger survey (or census) that provides the necessary coverage, but is lacking the variable one is interested in. See the small area estimation method put forward by Elbers et al. (2003), which inspired Demombynes et al. (2007), Elbers et al. (2008), Tarozzi and Deaton (2009), Tarozzi (2011), and Viet-Cuong et al. (2010). In these studies poverty derived from a household expenditure survey is imputed into the larger survey (or census), which is then aggregated at the small area level.[2] Ivaschenko and Lanjouw (2010) apply the same approach to impute the probability of being tested HIV positive into a DHS (that does not collect HIV data) to obtain subnational estimates of HIV prevalence. Here the imputation model is estimated with sentinel survey data which contains HIV data for a sample of individuals that is not representative at the stratum or even national level. Fujii (2011) combines the DHS with census data to estimate child malnutrition at the small area level.

The tracking of welfare over time offers another example. It denotes an important tool in monitoring progress and identifying leading and lagging areas. In two scenarios there is a need for imputed data in which case the cross-tabulation of poverty say versus time becomes a two-sample cross-tabulation: (1) the household expenditure survey is outdated and/or conducted infrequently. A viable solution may then be to impute household poverty into an alternative survey that is conducted more regularly such as a DHS, see e.g. Stifel and Christiaensen (2007) and Christiaensen et al. (2011); (2) even if household expenditure surveys are up-to-date, the comparison of consumption

[1]These studies measure socio-economic status by means of an asset index. The latter is often computed as predicted consumption in which case it coincides with imputed consumption. By treating the variable as an exact asset index however rather than imputed consumption, one circumvents the problem of having to account for imputation error. See for example Filmer and Scott (2008). By the same token, had consumption and health data been available in the same sample, consumption would likely have been preferred over the asset index to determine the individual's socio-economic status. The approach advocated in this paper enables one to obtain estimates of the latter with easy to implement standard errors that does not do away with the error due to imputation.

[2]If a census is used, then there is no sampling error only imputation error.

poverty between time periods may be hampered by a lack of comparability due to significant changes in the questionnaires. The Indian poverty debate denotes a well-known example, see e.g. Deaton and Dreze (2002), Deaton (2003), Kijima and Lanjouw (2003), and Tarozzi (2007), where the offered solution is to work with imputed consumption that is comparable over time.

The transition matrix is also an example of a cross-tabulation which is conventionally only available if one has panel data. The two-sample analogue is obtained by combining repeated cross-sections to create a so-called synthetic-panel. Consider a table that shows the percentage of households falling into and out of poverty which helps address the question of whether poverty is of a chronic or a transient nature. Lanjouw et al. (2011) estimate this poverty transition matrix by imputing time t consumption into the time $t+1$ survey and vice versa. Similarly one may study the transitions between levels of food security and malnutrition, sectors of employment, political preferences, location of residence (urban or rural; country of origin or abroad) etc. Note that transition matrices can also be estimated from repeated cross-sections without the use of survey-to-survey imputation. With the help of additional assumptions one could estimate the transition probabilities directly by means of maximum-likelihood, see e.g. Moffitt (1993) for the methodology and Pelzer et al. (2001) for an empirical application.

Few empirical studies implement correct standard errors that acknowledge the dual error structure of estimates based on imputed data. It is common practice to either treat the imputed data as observed data or ignore the sampling error, or omit the standard errors all together. This includes the empirical applications listed above. While there are bootstrap procedures available to deal with imputed survey data (see e.g. Shao and Sitter, 1996), to the best of our knowledge they have not yet been adopted in the context of cross-tabulations and made available in standard statistical packages. This leaves to user to having to program it him- or herself. The analytic standard errors presented in this paper are easily implemented, without the use of bootstrapping, making the two-sample cross-tabulation almost as user-friendly as its one-sample counterpart.

The paper is organized as follows. Section 2 reviews the commonly used cross-tabulations, both the one-sample and the two-sample version. We offer a bias-corrected alternative to the naive two-sample cross-tabulation in Section 3, which includes a discussion of its asymptotic properties. Sections 4 and 5 provide respectively a Monte-Carlo simulation study and an empirical application. Section 6 concludes.

2 Conventional cross-tabulation

A cross-tabulation presents in a tabular form estimates of the proportion of the population that belongs to combinations of two categories. For example a the share of children that are both poor and malnoursihed denotes one of four cells of a cross-tabulation of poverty and/with child malnutrition. To simplify the presentation, we focus on the point estimate and the standard error for a particular cell in the cross-tabulation table.

We start with a general framework and make progressively more restrictive (but still fairly general) assumptions to obtain sharper analytical results. Throughout the paper, we maintain the following assumption:[3]

Assumption 1 (x_i, y_i) *denotes an independently and identically distributed sequence of dummy (binary) variables.*

The subscript i is the observation unit in the data. To be consistent with our empirical application, we shall refer to i as "household" in this paper. Our objective is to derive a consistent estimator of $\mu_{xy} \equiv \Pr(x_i = 1, y_i = 1) = E[x_i y_i]$ and find its asymptotic properties. The results are easily extended to the other cells of the cross-tabulation: $E[x_i(1 - y_i)]$, $E[(1 - x_i)y_i]$, and $E[(1 - x_i)(1 - y_i)]$.

Remark 2 *Under the assumption of binary x_i and y_i the cross-tabulation takes the form of a 2×2 table. Our approach is equally applicable however to the case of two categorical variables. Assume that the variables x_i and y_i can take on the values $\kappa_x \in \{v_1^x, \cdots, v_{d_x}^x\}$ and $\kappa_y \in \{v_1^y, \cdots, v_{d_y}^y\}$, which may represent more than two values, and suppose that we want to estimate $\Pr(\kappa_x = v_a^x, \kappa_y = v_b^y)$. Then, by letting $x_i = \mathrm{Ind}(\kappa_x = v_a^x)$ and $y_i = \mathrm{Ind}(\kappa_y = v_b^y)$, we have $\mu_{xy} = \Pr(\kappa_x = v_a^x, \kappa_y = v_b^y)$.*

2.1 One-sample cross-tabulation

Suppose that x_i and y_i are observed for the first n households in the sequence. Then, the (standard) one-sample cross-tabulation estimator μ_{xy}^O is simply the following sample average:

$$\hat{\mu}_{xy}^O \equiv n^{-1} \sum_{i=1}^{n} x_i y_i,$$

where n is the size of the sample.

Under Assumption 1 we have that $\hat{\mu}_{xy}^O$ is unbiased and consistent, and has a variance of $n^{-1} Var[x_i y_i] = n^{-1} \mu_{xy}(1 - \mu_{xy})$. The variance can be consistently estimated by

[3] When more restrictive assumptions are made subsequently, this assumption becomes redundant. However, to keep the presentation simple, we shall not explicitly drop the assumptions implied by other assumptions.

the $n^{-1}svar[x_iy_i]$ (i.e. the sample variance of x_iy_i divided by the sample size). As sampling denotes the sole source of randomness, the error in the one-sample estimator is commonly referred to as *sampling error*.

2.2 Naïve two-sample cross-tabulation

To create a cross-tabulation from two independently-drawn samples that cannot be linked at the household level we need a set of covariates $z \in \mathbb{R}^L$ that are available in both samples, where L is the dimension of the covariates. z serves as a set of instruments that will be used to impute y in the sample that holds x (or impute x in the sample that holds y). In what follows, we shall maintain the following set of assumptions, which is slightly more restrictive than Assumption 1:

Assumption 3 *The triple (x_i, y_i, z_i) is independently and identically distributed across i.*

Without loss of generality, let the index set for Sample 1 and Sample 2 be $\mathcal{S}_1 \equiv \{1, \cdots, n_1\}$ and $\mathcal{S}_2 \equiv \{n_1+1, \cdots, n_1+n_2\}$, respectively. We assume that the pair (x_i, z_i) is observed for all $i \in \mathcal{S}_1$ and the pair (y_i, z_i) is observed for all $i \in \mathcal{S}_2$. With some abuse of notation, we shall also use $\mathcal{S}_1$ and $\mathcal{S}_2$ to refer to Sample 1 and Sample 2, respectively.

It is possible that some of the factors in z are related only to x but not to y, and that others are related only to y, but not to x. In this case, we may set the coefficient on "irrelevant" covariates identically equal to zero in the estimation of model parameters discussed below. In general, z does not perfectly determine x and y so that x and y are stochastic conditional on z.

We denote conditional moments by bar variables, and assume that they can be described by the following parametric models:

$$\bar{w}_i \equiv E[w_i|z_i] = G_{1w}(z_i, \theta_w) \quad \text{for} \quad w \in \{x, y\}, \quad \text{and} \quad \overline{x_iy_i} \equiv E[x_iy_i|z_i] = G_2(z_i, \theta),$$

where θ includes θ_x, θ_y, and possibly some other parameters.

In previous studies cross-tabulation tables have been constructed from two samples in the following manner: First, the parameter θ_y is estimated using Sample 2. Then, using the estimate $\hat{\theta}_y$, y_i is imputed by $\hat{y}_i (\equiv G_{1y}(z_i, \hat{\theta}_y))$ for each $i \in \mathcal{S}_1$. Taking the sample average of $x_i\hat{y}_i$ over $i \in \mathcal{S}_1$, we obtain the following naïve two-sample estimator of μ_{xy}:

$$\hat{\mu}^N_{xy} = n_1^{-1} \sum_{i \in \mathcal{S}_1} x_i \hat{y}_i,$$

where $n_s \equiv \#(\mathcal{S}_s)$ is the sample size $s \in \{1, 2\}$.

Remark 4 *It is in general possible to swap the sample for estimation and the sample for imputation. That is, we can estimate the parameter* θ_x *using Sample* 1 *and impute* $\hat{x}_i (\equiv G_{1x}(z_i, \theta_x))$ *for each observation in Sample* 2.

We refer to $\hat{\mu}_{xy}^N$ as a naïve two-sample estimator as it is in general biased even when both n_1 and n_2 tend to infinity. This is because the estimator does not take into account the correlation between x and y conditional on z. This point can be easily seen by considering the special case where θ_y is estimated without error (all the proofs are provided in the appendix):

Proposition 5 *Suppose that* $\hat{\theta}_y = \theta_y$. *Then, the bias* B *of the naïve estimator* $\hat{\mu}_{xy}^N$ *solves* $B = E_z[\bar{x}_i \bar{y}_i - \overline{x_i y_i}] = -E_z[Cov[x_i|z_i, y_i|z_i]]$.

Remark 6 *The bias expression given in Proposition* 5 *also applies to the case where* x_i *is imputed for each record in sample* 2 *where* θ_x *is estimated without error. This can be readily seen from the symmetry of the bias expression. In practice,* $\hat{\theta}_w \neq \theta_w$ *for* $w \in \{x, y\}$, *in which case the symmetry is broken, i.e. it will matter for the bias whether we impute in Sample* 1 *or Sample* 2.

In the current form, it is difficult to see how important the bias could be. Therefore, we maintain the following assumption, which is slightly more restrictive than Assumption 3, for the remainder of this paper:

Assumption 7 *The triple* $(z_i, \varepsilon_{xi}, \varepsilon_{yi})$ *is independently and identically distributed across* i. *Further, the variables* x_i *and* y_i *are jointly described by:*

$$x_i = \text{Ind}(z_i^T \theta_x + \varepsilon_{xi} \leq 0) \quad (1)$$

$$y_i = \text{Ind}(z_i^T \theta_y + \varepsilon_{yi} \leq 0), \quad (2)$$

where Ind$(\cdot)$ *is the indicator function that equals* 1 *if the argument is true and* 0 *otherwise. The idiosyncratic error terms* ε_{xi} *and* ε_{yi} *are independent of* z_i, *and have mean zero, unit variance, and correlation* ρ.

Note that we may normalize ε_{xi} and ε_{yi} to have zero mean and unit variance without loss of generality by including a constant term in z_i. We denote the joint distribution function of ε_{xi} and ε_{yi} by F_2 and their marginal distribution functions by F_{1x} and F_{1y}. We shall denote the probability density functions corresponding to F_{1x} and F_{1y} by f_{1x} and f_{1y} respectively. We assume that the probability density functions are at least once differentiable. We define $s(w) \equiv \text{Ind}(w = x) + 2\text{Ind}(w = y)$, which is the index for the sample that contains $w \in \{x, y\}$ (e.g., $s(x) = 1$ as x is contained in $\mathcal{S}_1$).

Two-sample estimation requires estimation of the model parameters. In what follows, we make the following assumption:

Assumption 8 *For $w \in \{x, y\}$, the estimator $\hat{\theta}_w$ of θ_w satisfies the following properties as $n_w \to \infty$:*

$$\begin{cases} \hat{\theta}_w & \xrightarrow{p} & \theta_w \\ \sqrt{n_{s(w)}}(\hat{\theta}_w - \theta_w) & \xrightarrow{d} & \mathcal{N}(0, \Omega_w) \\ \hat{\theta}_w & \perp & (z_i, \varepsilon_{xi}, \varepsilon_{yi}) \quad \text{for} \quad i \in \mathcal{S}_{3-s(w)}, \end{cases} \tag{3}$$

where Ω_w is the asymptotic variance of $\hat{\theta}_w$.

Remark 9 *Assumption 8 is very general. For a wide range of F_{1x} and F_{1y}, $\sqrt{n}$-consistent estimators of θ_x and θ_y are available. An obvious example is the maximum likelihood (ML) estimator when the ML regularity conditions are satisfied. Furthermore, the independence condition is typically satisfied because $\mathcal{S}_2$ [$\mathcal{S}_1$] contains y_i [x_i] and z_i and thus θ_y [θ_x] can be estimated with $\mathcal{S}_2$ [$\mathcal{S}_1$] alone.*

Under Assumption 7, we have $\bar{w}_i = \Pr(\varepsilon_{wi} \leq -z_i^T\theta_w | z_i) = F_{1w}(-z_i^T\theta_w)$ for $w \in \{x, y\}$ and $\overline{x_i y_i} = F_2(-z_i^T\theta_x, -z_i^T\theta_y, \rho) = \Pr(\varepsilon_{xi} \leq -z_i^T\theta_x, \varepsilon_{yi} \leq -z_i^T\theta_y | z_i)$. Therefore, the functions F and G are related by $G_{1w}(z_i, \theta_w) = F_{1w}(-z_i^T\theta_w)$ and $G_{2w}(z_i, \theta) = F_{2w}(-z_i^T\theta_x, -z_i^T\theta_y, \rho)$. The imputed values can be written as $\hat{w}_i = F_{1w}(-z_i^T\hat{\theta}_w)$. With a slight abuse of notation, we denote the derivative of F_2 with respect to the first and second argument by F_{2x} and F_{2y}, respectively.

The bias is seen to satisfy:

$$B = E_z[F_{1x}(-z_i^T\hat{\theta}_x)F_{1y}(-z_i^T\hat{\theta}_y) - F_2(-z_i^T\hat{\theta}_x, -z_i^T\hat{\theta}_y, \rho)]. \tag{4}$$

Note that the naive estimator is implicitlty built on the assumption that $\rho = 0$. If that assumption were to hold, there would be no bias as can be seen from eq. (4). In practice ρ will rarely equal zero. In general we have that the more ρ deviates from zero the larger the bias. [Include Corollary that shows exactly that?]

To derive the asymptotic distribution of the naïve estimator, we need to make an assumption about how fast the two sample sizes tend to infinity. Therefore, we make the following assumption:

Assumption 10 *The sample sizes n_1 and n_2 are non-decreasing functions of t. Further, $n_1 \to \infty$ and $n_2 \to \infty$ as $t \to \infty$. We also assume that $\lim_{t\to\infty} n(t)/n_1(t) = r_1 \in [0, 1]$ and $\lim_{t\to\infty} n(t)/n_2(t) = r_2 \in [0, 1]$, where $n(t) = \min(n_1(t), n_2(t))$.*

We assume the convergence of n/n_1 and n/n_2 for the convenience of argument. Later we shall discuss the consequence of relaxing this assumption. Under this assumption we obtain the following proposition:

Proposition 11 *Let* $V^N \equiv r_1 Var[x_i\bar{y}_i] + r_2 E[m_i^T]\Omega_y E[m_i]$*, where* $m_i \equiv x_i f_{1y}(-z_i^T\theta_y)z_i$*. Then, under Assumptions* 7*,* 8*, and* 10*, and under some regularity conditions, the naïve estimator* $\hat{\mu}_{xy}^N$ *has the following asymptotic properties as* $t \to \infty$*:*

$$\begin{aligned}\hat{\mu}_{xy}^N - \bar{\mu}_{xy}^N &\xrightarrow{p} 0\\ \sqrt{n}(\hat{\mu}_{xy}^N - \bar{\mu}_{xy}^N) &\xrightarrow{d} \mathcal{N}\left(0, V^N\right),\end{aligned}$$

where $\bar{\mu}_{xy}^N \equiv E[x_i\bar{y}_i]$ *which is generally different from* μ_{xy}*.*

Proposition 11 shows that there are two sources of error that contribute to the variance of the naïve estimator. The first component, $r_1 Var[x_i\bar{y}_i]$, is due to sampling error from the first sample. The second component is due to *model error*, i.e. the error in the estimation of the model parameters. In this case, the model error is $\hat{\theta}_y - \theta_y$. When we specifically refer to the model error for θ_w for $w \in \{x, y\}$, we call it *w-model error.* The w-model error leads to the *imputation error*, $\hat{w}_i - \bar{w}_i$.

From Proposition 11 we also learn that if we ignore the fact that $\hat{y}_i$ is an imputed value and not an observation, we will underestimate the variance. Therefore, even if $\rho = 0$ in which case the naïve estimator has no bias asymptotically, conclusions drawn from the naïve estimator may still be misleading unless the standard errors are appropriately calculated. In other words, conventional one-sample cross-tabulation standard errors where imputed data is treated as observed data (i.e. $svar_{\mathcal{S}_1}[x_i\hat{y}_i]^{0.5}$) cannot be used as an estimate of the naïve estimator's standard errors because they do not take into account the imputation error (i.e., $\hat{y}_i - \bar{y}_i$).

It is useful to have an idea about the order of magnitude of the bias. To this end we shall make two normality assumptions for the remainder of this subsection. Results derived under these assumptions reveal how the bias is shaped by the underlying model parameters. In particular, we can see how sensitive the bias is with respect to ρ.

Assumption 12 *The pair* $(\varepsilon_{xi}, \varepsilon_{yi})$ *jointly follows a bi-variate standard normal distribution for all* i*.*

Note that under Assumption 12, x and y are described by a bi-variate probit model. We denote the bi-variate and univariate standard normal distribution functions by Φ_2 and Φ_1, respectively. Their probability density functions are denoted by ϕ_2 and ϕ_1.

To obtain an (approximate) analytical closed form expression for the bias we additionally make the following normality assumption:

Assumption 13 *Let the systematic components be* $Z_{xi} \equiv -z_i^T\theta_x$ *and* $Z_{yi} \equiv -z_i^T\theta_y$*.* Z_{xi} *and* Z_{yi} *jointly have the following distribution for all* i*:*

$$\begin{bmatrix} Z_{xi} \\ Z_{yi} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{zx} \\ \mu_{zy} \end{bmatrix}, \begin{bmatrix} \sigma_{zx}^2 & \lambda\sigma_{zx}\sigma_{zy} \\ \lambda\sigma_{zx}\sigma_{zy} & \sigma_{zy}^2 \end{bmatrix}\right).$$

Note that if z is normally distributed, this assumption is satisfied. Also, it should be noted that the variance of the systematic component serves as a measure of the goodness-of-fit because the variance of the idiosyncratic component is fixed.

Under Assumptions 12 and 13, we obtain the following closed-form expression for the bias:

Theorem 14 *Let* $W_0 \equiv (1-\lambda^2)\sigma_{zx}^2\sigma_{zy}^2+\sigma_{zx}^2+\sigma_{zy}^2+1$, $W_1 = (\sigma_{zy}^2+1)\mu_{zx}^2+(\sigma_{zx}^2+1)\mu_{zy}^2$, $W_{2x} \equiv \lambda\sigma_{zx}\sigma_{zy}-(\sigma_{zy}^2+1)\mu_{zx}$, $W_{2y} \equiv \lambda\sigma_{zx}\sigma_{zy}-(\sigma_{zx}^2+1)\mu_{zy}$, $W_{3x} \equiv (1-\lambda^2)^2W_{2x}^2/W_0^2 - (\sigma_{zx}^2+1)W_0$, *and* $W_{3y} \equiv (1-\lambda^2)^2W_{2y}^2/W_0^2 - (\sigma_{zy}^2+1)W_0$. *Then, under Assumptions* 12 *and* 13 *and using a third-order Taylor approximation of* Φ_2 *with respect to* ρ *around* $\rho = 0$*, the bias solves:*

$$\begin{aligned} B &= E_z[\Phi_1(Z_{xi})\Phi_1(Z_{yi}) - \Phi_2(Z_{xi}, Z_{yi}, \rho)] && (5) \\ &\simeq -c_0\left(\rho + \frac{\rho^2}{2}c_2 + \frac{\rho^3}{6}c_3\right), && (6) \end{aligned}$$

where:

$$\begin{aligned} c_0 &= (2\pi\sqrt{W_0})^{-1}\exp\left[(2W_0)^{-1}(2\lambda\sigma_{zx}\sigma_{zy}\mu_{zx}\mu_{zy} - W_1)\right] \\ c_2 &= W_0^{-2}\left(\lambda(W_0 - W_1)\sigma_{zx}\sigma_{zy} + (W_0 + 2\lambda^2\sigma_{zx}^2\sigma_{zy}^2)\mu_{zx}\mu_{zy}\right) \\ c_3 &= W_{3x}W_{3y} + 2\lambda\sigma_{zx}\sigma_{zy}W_0^{-2}(2c_2W_0 - \lambda\sigma_{zx}\sigma_{zy}). \end{aligned}$$

For most practical purposes this approximation will suffice. As can be seen in Figure ?, the third-order approximation is indeed very accurate. [ADD SOME FIGURES HERE].

As eq. (6) shows an intricate function of the model parameters, let us focus on the first-order term ($-c_0\rho$) to get a sense of what determines the bias. Due to $W_0 > 0$, and the fact that the following relationship holds true:

$$\begin{aligned} & 2\lambda\sigma_{zx}\sigma_{zy}\mu_{zx}\mu_{zy} - W_{1x} - W_{1y} \\ = & -\lambda(\sigma_{zx}\mu_{zx} + \sigma_{zy}\mu_{zy})^2 + (1+\lambda)(\sigma_{zx}^2\mu_{zx}^2 + \sigma_{zy}^2\mu_{zy}^2) + \mu_{zx}^2 + \mu_{zy}^2 \\ \geq & \ 0, && (7) \end{aligned}$$

we have that $|-c_0\rho| \leq (2\pi\sqrt{W_0})^{-1}\rho$. Therefore, $(2\pi\sqrt{W_0})^{-1}\rho$ can be considered as a first-order approximation to the upper bound on the bias when μ_{zx} and μ_{zy} are allowed to vary. Since eq. (7) is held with equality if and only if $\mu_{zx} = \mu_{zy} = 0$, this upper bound is the largest when the center of distribution of (Z_{xi}, Z_{yi}) is at the origin. This is where the binary outcomes (x_i, y_i) are most sensitive to the error terms. Furthermore, when the center of distribution of (Z_{xi}, Z_{yi}) is at the origin, the bound is found to be the largest when $|\lambda| = 1$.

Because W_0 is a monotonically increasing function of σ^2_{zx} and σ^2_{zy}, the upper bound tends to be smaller when the variance of Z_{xi} and Z_{yi} is larger. This is because when the variance of the systematic component is large, the goodness-of-fit of the model is good. As a result, the idiosyncratic component will be relatively unimportant in determining the value of x_i or y_i.

Note that ρ cannot be estimated from the data, which has motivated the use of the naïve estimator. One could however guesstimate the value of ρ, and use this 'estimate' to correct for the bias. A natural choice is to borrow estimates of ρ from other data sets that contain both x_i and y_i (e.g., data taken at different time periods or in different countries). We shall show in subsequent sections that even a crude guesstimate can substantially improve the estimate.

3 Bias-corrected two-sample cross-tabulation

If we know the value of ρ, we can estimate the bias. This in turn means that we can adopt a bias correction. The bias is estimated by replacing in Proposition 5 the expected values of the conditional moments (i.e. the bar variables) by their sample averages [analogues] (i.e. the hat variables):

$$\hat{B} = n_1^{-1} \sum_{i \in \mathcal{S}_1} \hat{x}_i \hat{y}_i - \widehat{x_i y_i},$$

where $\widehat{x_i y_i} \equiv F_2(-z_i\hat{\theta}_x, -z_i\hat{\theta}_y, \rho)$.

It will be convenient to define $q_i \equiv (x_i - \bar{x}_i)\bar{y}_i + \overline{x_i y_i}$ and its estimate by $\hat{q}_i \equiv (x_i - \hat{x}_i)\hat{y}_i + \widehat{x_i y_i}$. Subtracting $\hat{B}$ from $\hat{\mu}^C_{xy}$ then yields the following bias-corrected estimator $\hat{\mu}^C_{xy}$:

$$\hat{\mu}^C_{xy} = \hat{\mu}^N_{xy} - \hat{B} = \frac{1}{n_1} \sum_{i \in \mathcal{S}_1} \hat{q}_i \tag{8}$$

Assuming that the true value of ρ is inserted, it is straightforward to establish the consistency of the bias-corrected estimator $\hat{\mu}^C_{xy}$.

Proposition 15 *Under Assumptions* 7 *and* 8 *and under some regularity conditions, the bias-corrected estimator* $\hat{\mu}^C_{xy}$ *satisfies the following asymptotic property as* $t \to \infty$*:*

$$\hat{\mu}^C_{xy} - \mu_{xy} \ \xrightarrow{p} \ 0.$$

Notice that the bias-corrected estimator features both $\hat{y}_i$ and $\hat{x}_i$. This means that we now have two different sources of model error; $\hat{\mu}^C_{xy}$ is a function of $\hat{\theta}_y - \theta_y$ as well as $\hat{\theta}_x - \theta_x$. The former model error originates from $\mathcal{S}_2$ and is thus independent from the sampling error which comes from $\mathcal{S}_1$. As $\hat{\theta}_x$ is estimated using data from $\mathcal{S}_1$ the latter model error will generally not be independent from the sampling error.

To derive the asymptotic distribution of $\hat{\mu}^C$, taking into account the correlation between model error and sampling error, it is necessary to impose some structure on $\hat{\theta}_x$. We make the following assumption:

Assumption 16 $\hat{\theta}_w$ *for* $w \in \{x, y\}$ *is an ML estimator:*

$$\sqrt{n}(\hat{\theta}_w - \theta_w) \ \xrightarrow{d} \ \mathcal{N}\left(0, \Omega_w\right),$$

where:

$$\Omega_w = E[\frac{f_{1w}(-z_i^T\theta_w)^2}{\bar{w}_i(1-\bar{w}_i)} z_i z_i^T]. \tag{9}$$

We are now ready to establish asymptotic normality of $\hat{\mu}^C_{xy}$:

Proposition 17 *Let us define* $m_{1i} \equiv \left(\bar{y}_i f_{1x}(-z_i^T\theta_x) - F_{2x}(-z_i^T\theta_x, -z_i^T\theta_y, \rho)\right) z_i$*,* $m_{2i} \equiv \left((\bar{x}_i - x_i) f_{1y}(-z_i^T\theta_y) - F_{2y}(-z_i^T\theta_x, -z_i^T\theta_y, \rho)\right) z_i$*,* $d_{xi} = (f_{1x}(-z_i^T\theta_x)(\bar{x}_i - x_i)/(\bar{x}_i(1 - \bar{x}_i)))z_i$*, and* $a_i \equiv q_i + E[m_{1i}^T]\Omega_x d_{xi}$*. Then, under Assumptions* 7*,* 8*,* 10 *and* 16*, and under some regularity conditions, the bias-corrected estimator* $\hat{\mu}^C_{xy}$ *has the following asymptotic property as* $t \to \infty$*:*

$$\sqrt{n}(\hat{\mu}^C_{xy} - \mu_{xy}) \ \xrightarrow{d} \ \mathcal{N}\left(0, V^C\right),$$

where:

$$V^C \equiv r_1 Var[a_i] + r_2 E[m_{2i}^T]\Omega_y E[m_{2i}]. \tag{10}$$

The rate of convergence is clearly determined by the sample that increases at the slowest rate. If n_1 grows at a slower rate than n_2 as $t \to \infty$, the y-model error component is less important than the sampling error and the x-model error. The opposite is true when n_1 grows faster than n_2. Notice that n/n_1 and n/n_2 do not necessarily converge. In this case, the following interpretation is possible:

Remark 18 *Define* $\bar{r}_1 \equiv \limsup_{t\to\infty} n/n_1$ *and* $\bar{r}_2 \equiv \limsup_{t\to\infty} n/n_2$. *By replacing* r_1 *and* r_2 *by* $\bar{r}_1$ *and* $\bar{r}_2$ *in Propositions 11 and 17, we have the upper bound on the asymptotic variance.*

A natural estimator for the variance of $\hat{\mu}^C_{xy}$ is obtained by replacing θ_w by $\hat{\theta}_w$, Ω_w by $\hat{\Omega}_w$, and expected values by sample averages:

$$n^{-1}\hat{V}^C = \frac{1}{n_1} svar_{\mathcal{S}_1}[\hat{a}_i] + \frac{1}{n_2}\hat{E}[m_{2i}^T]\hat{\Omega}_y\hat{E}[m_{2i}] \tag{11}$$

where $\hat{a}_i = \hat{q}_i + \hat{E}[m_{1i}^T]\hat{\Omega}_x\hat{d}_{xi}$, $\hat{E}[m_{ki}] = n_1^{-1}\sum_{i\in\mathcal{S}_1}\hat{m}_{ki}$ for $k \in \{1,2\}$, and where $svar_{\mathcal{S}_1}$ denotes the sample variance operator. Note that for the maximum-likelihood estimator, the estimate for Ω_w is given by $\hat{\Omega}_w = n_{s(w)}^{-1}\sum_{i\in\mathcal{S}_{s(w)}}(f_{1w}(-z_i^T\hat{\theta}_w)^2)(\hat{w}_i(1-\hat{w}_i))^{-1}z_iz_i^T$. It is straightforward to establish consistency of $\hat{V}^C$:

Proposition 19 *Under Assumptions* 7, 8, 10, *and* 16 *and under some regularity conditions,* $\hat{V}^C \xrightarrow{p} V^C$ *as* $t \to \infty$.

Assumption 16 is motivated by the fact that the maximum-likelihood estimator denotes a popular choice for estimating the parameters of the probit model (see Assumption 7). However, the proofs of Propositions 17 and 19 essentially only rely on the consistency of $\hat{\theta}_w$, $\hat{\Omega}_w$ and some relevant sample moments. Therefore, replacing the maximum-likelihood estimator with an alternative such as a GMM estimator requires only minor modifications to the derivation of the asymptotic distribution of $\hat{\mu}^C_{xy}$.

To see how standard errors are shaped by the underlying parameters we derive a variance decomposition of the total estimation error. These results show when the estimation error can be expected to be either large or small. By taking the first-order Taylor expansion of $\widehat{x_iy_i}$ and $\hat{w}_i$ for $w\{x,y\}$ around (θ_x, θ_y) we obtain:

$$\begin{aligned}
\hat{w}_i &= \bar{w}_i - f_1(-z_i^T\theta_w)z_i^T(\hat{\theta}_w - \theta_w) + h_{iw} \\
\widehat{x_iy_i} &= \overline{x_iy_i} - \sum_{w\in\{x,y\}} F_{2w}(-z^T\theta_x, -z^T\theta_y, \rho)z_i^T(\hat{\theta}_w - \theta_w) + h_{i2},
\end{aligned}$$

where h_{iw} and h_{i2} are functions involving second- or higher-order terms of $\hat{\theta}_x - \theta_x$, $\hat{\theta}_y - \theta_y$, or both. Noting that $\mu_{xy} = E[\overline{x_iy_i}]$ yields:

$$\hat{\mu}^C_{xy} - \mu_{xy} = \frac{1}{n_1}\sum_{i\in\mathcal{S}_1} q_i - E[q_i] + m_{1i}^T(\hat{\theta}_x - \theta_x) + m_{2i}^T(\hat{\theta}_y - \theta) + h_{ic}, \tag{12}$$

where h_{ic} is a function involving second- or higher-order terms of $\hat{\theta}_x - \theta_x$, $\hat{\theta}_y - \theta_y$, or both.

Eq. (12) shows that the estimation error consists of three important components. The first component, $n_1^{-1}\sum_{i\in\mathcal{S}_1} q_i - E[q_i]$, is due to the sampling error. The second component, $n_1^{-1}\sum_{i\in\mathcal{S}_1} m_{1i}^T(\hat{\theta}_x - \theta_x)$, is due to the x-model error. The third component, $n_1^{-1}\sum_{i\in\mathcal{S}_1} m_{2i}^T(\hat{\theta}_y - \theta_y)$, is due to the y-model error.

It follows that the first two components are in general correlated because both components originate from $\mathcal{S}_1$. The correlation between the third component and the first two on the other hand can be ignored asymptotically as the third component is primarely driven by $\mathcal{S}_2$.

Proposition 20 *Let us define* $m_{4i} \equiv f_{1x}(-z_i^T\theta_x)\bar{y}_i z_i$. *Then, under Assumptions* 7, 8, 10, *and* 16 *and under some regularity conditions, the asymptotic variance of the bias-corrected estimator* $\hat{\mu}_{xy}^C$ *can be decomposed as follows:*

$$
\begin{aligned}
V^C &= r_1 E[\bar{x}_i(1-\bar{x}_i)\bar{y}_i^2] + r_1 Var[\overline{x_i y_i}] - 2r_1 E[m_{1i}^T]\Omega_x E[m_{4i}] \\
&+ r_1 E[m_{1i}^T]\Omega_x E[m_{1i}] + r_2 E[m_{2i}^T]\Omega_y E[m_{2i}].
\end{aligned}
$$

Proposition 20 decomposes the asymptotic variance V^C into five components. The first two components are both due to sampling error from $\mathcal{S}_1$. The first component ($r_1 E[\bar{x}_i(1-\bar{x}_i)\bar{y}_i^2]$) increases with the variation in the idiosyncratic error term ε_{xi} (and decreases when the systematic part increases). The second component ($r_1 Var[\overline{x_i y_i}]$) increases with the variation in z_i. The third component ($2r_1 E[m_{1i}^T]\Omega_x E[m_{4i}]$) measures the covariance between the x-model error and the sampling error. The remaining two components ($r_1 E[m_{1i}^T]\Omega_x E[m_{1i}]$ and $r_2 E[m_{2i}^T]\Omega_y E[m_{2i}]$) are both due to model error.

To highlight the factors that determine the variance we next evaluate each of the components of the variance decomposition as a function of the underlying model parameters and moments of z. Because $E[(\bar{x}_i - x_i)f_{1y}(-z_i^T\theta_y)] = 0$, we have $E[m_{2i}] = E[-F_{2y}(-z_i^T\theta_x)z_i]$. Under Assumption 12 (the bivariate probit model) we have: holds.

$$
\begin{aligned}
\hat{E}[m_{1i}] &= n_1^{-1}\sum_{i\in\mathcal{S}_1} \phi_1(-z_i^T\hat{\theta}_x)\left(\Phi_1(-z_i^T\hat{\theta}_y) - \Phi_1(\psi(\rho z_i^T\hat{\theta}_x - z_i^T\hat{\theta}_y))\right) z_i \\
\hat{E}[m_{2i}] &= -n_1^{-1}\sum_{i\in\mathcal{S}_1} \phi_1(-z_i^T\hat{\theta}_y)\Phi_1\left(\psi(\rho z_i^T\hat{\theta}_y - z_i^T\hat{\theta}_x)\right) z_i,
\end{aligned}
$$

where $\psi \equiv (1-\rho^2)^{-1/2}$.

Applying a variant of the delta method to V^C with respect to z yields the following proposition:

Proposition 21 *Suppose that Assumptions 7, 12, and 16 hold and that* $\hat{\theta}_x$ *and* $\hat{\theta}_y$ *are maximum-likelihood estimates of* θ_x *and* θ_y*, respectively. Then, by taking the second-*

order approximation around $z_i = 0$ of each function involving z_i before the expectation is taken, each term in the variance decomposition for V^C can be approximated by:

$$
\begin{aligned}
E[\bar{x}_i(1-\bar{x}_i)\bar{y}_i^2] &\simeq \frac{1}{16} - \frac{1}{4\sqrt{2\pi}}\theta_y^T\mu_z \\
&\quad + \frac{1}{8\pi}\left(\theta_y^T(\Sigma_z + \mu_z\mu_z^T)\theta_y - \theta_x^T(\Sigma_z + \mu_z\mu_z^T)\theta_x\right) && (13) \\
Var[\overline{x_i y_i}] &\simeq \frac{1}{8\pi}(\theta_x + \theta_y)^T\Sigma_z(\theta_x + \theta_y) && (14) \\
2E^T[m_{1i}]\Omega_x E[m_{4i}] &\simeq \frac{1}{4\sqrt{2\pi}}\Theta^T\mu_z - \frac{1}{4\pi}\Theta^T\left(\Sigma_z + \mu_z\mu_z^T\right)\theta_y && (15) \\
E[m_{1i}^T]\Omega_x E[m_{1i}] &\simeq \frac{1}{8\pi}\Theta^T\left(\Sigma_z + \mu_z\mu_z^T\right)\Theta && (16) \\
E[m_{2i}^T]\Omega_y E[m_{2i}] &\simeq \frac{1}{16}\mu_z^T\left(\Sigma_z + \mu_z\mu_z^T\right)^{-1}\mu_z + \frac{\psi}{4\sqrt{2\pi}}\mu_z^T(\rho\theta_y - \theta_x) \\
&\quad + \frac{\psi^2}{8\pi}(\rho\theta_y - \theta_x)^T\left(\Sigma_z + \mu_z\mu_z^T\right)(\rho\theta_y - \theta_x), && (17)
\end{aligned}
$$

where $\Sigma_z \equiv Var[z_i]$, $\mu_z \equiv E[z_i]$, and $\Theta \equiv (\psi - 1)\theta_y - \psi\rho\theta_x$.

Unlike the standard delta method, the first-order term does not disappear unless $E[z_i] = 0$. It can be verified however that the approximations in Proposition 21 remain accurate as long as Σ_z and μ_z are not too large in absolute value.

Proposition 21 shows what factors play an important role in determining the magnitude of the variance.

- When $\mu_z = 0_L$, where 0_L is a L-vector of zeros, all components except for the covariance term tend to increase with the variance of z (i.e. the diagonal elements of Σ_z). (Increases in μ_z, when μ_z slightly deviates from 0_L, tend to have a similar effect on all error components.)

- When $\mu_z \neq 0_L$, the variance of z still increases the x-model error component but the same is not necessarily true for the y-model error component. Simulation results presented later confirm that total variance can be both an increasing and a decreasing function of the variance of z (largely depending on the value of μ_z).

- It follows that the y-model error component is generally larger than the x-model error component when the two sample sizes are comparable. The third and last term of eq. (17) (y-model term) is found to be of a similar order of magnitude as the corresponding (and sole) term of eq. (16) (x-model term). The extra terms in eq. (17) is what typically makes the y-model error component larger than the x-model error component.

- The covariance component is generally small. For $|\rho| << 1$ we have that $\psi - 1$ is small, which in turn tends to make the terms featuring Θ small.
- The parameter vectors θ_x and θ_y have a similar effect on the sampling error components. This is not the case for the model error components. For the latter $\rho\theta_y - \theta_x$ appears to be an important vector from which it follows that θ_y generally has a smaller role in the model error than θ_x, especially when ρ is small.

Remark 22 *For ease of exposition all results derived in this section are based on the iid assumption. This assumption is however not crucial and can be relaxed. In many practical applications sampling is carried out in two stages. The first stage randomly draws a specified number of primary sampling units (clusters). The second stage then randomly draws a specified number of households from each of the clusters selected in the first stage. By considering the asymptotics with respect to the number of clusters, our propositions will hold without major modifications. All relevant population moments, assuming they exist, can be estimated by their sample analogs. The main difference is that the sample variance as well as the Ω matrices must now be computed in a way consistent with the sampling design.*

Finally note that the bias-corrected estimator $\hat{\mu}_{xy}^C$ requires knowledge of ρ. In general ρ cannot be estimated from two separate samples that cannot be linked at the unit record level. One therefore needs to use a guesstimate $\tilde{\rho}$ instead. As we shall show in the next two sections, working with reasonable guestimates of ρ generally yields superior estimates compared to estimates that ignore this correlation (i.e. effectively assuming $\rho = 0$).

4 Monte-Carlo

We use Monte-Carlo simulations to illustrate: (a) how standard errors of two-sample cross-tabulations are shaped by both sampling error and model error, (b) that the analytic (asymptotic) standard errors are accurate also for small sample sizes, and (c) that the bias-corrected estimator provides a significant improvement over the naive estimator even if the guestimate of ρ used represents a crude estimate.

In all simulations we impute $E[y|z]$ into sample 1, which contains the other dummy variable x as well as the regressor z. The parameter values shown in Table 4 form the basis for the simulations. Specifications 1 and 2 correspond to models where the systematic part is small relative to the idiosyncratic error (i.e. a modest goodness-of-fit). With specification 3 the systematic part plays a more dominant role. We condider centered regressors ($\mu_{z_1} = \mu_{z_2} = 0$) as well as non-centered regressors ($\mu_{z_1} = \mu_{z_2} = 0.5$). By shifting the center of the systematic part of the model away from zero, the joint probability μ_{xy}, the parameter we are estimating, will move closer to either 0 or 1, which will reduce the sampling error component but may increase the model error component of the standard error. We will examine in more detail how the statistical precision changes with θ, σ_z^2, and the sample sizes n_1 and n_2 by varying these parameters while keeping the other parameters fixed.

parameter	θ_x	θ_y	μ_z	σ_z^2	$cov[z_1, z_2]$	ρ
1	$(-0.5, 0)$	$(0, 0.5)$	$(0, 0)$, $(0.5, 0.5)$	$(0.5, 0.5)$	0	0.5
2	$(0.5, 0)$	$(0, 0.5)$	$(0, 0)$, $(0.5, 0.5)$	$(0.5, 0.5)$	0	0.5
3	$(0.5, 1)$	$(1, 0.5)$	$(0, 0)$, $(0.5, 0.5)$	$(0.5, 0.5)$	0	0.5

The number of Monte-Carlo simulations is set at $K = 10,000$. Standard errors obtained by bootstrapping are based on $R = 200$ replications. With each simulation $k = 1, ..., K$, we go through the following steps:

1. Generate samples $S_1^k = \{x_i^k, z_i^k\}_{i=1}^{i=n_1}$ and $S_2^k = \{y_i^k, z_i^k\}_{i=1}^{i=n_2}$, which will take on the role of 'orginal data'.
2. Compute $\hat{\mu}_{xy}^{N(k)}$ (the naive estimator) and $\hat{\mu}_{xy}^{C(k)}$ (the bias-corrected estimator) by using samples S_1^k and S_2^k.
3. Compute the asymptotic standard errors for $\hat{\mu}_{xy}^{N(k)}$ and $\hat{\mu}_{xy}^{C(k)}$ (see Propositions ...), denoted by $se_X^k(\hat{\mu}_{xy}^N)$ and $se_X^k(\hat{\mu}_{xy}^C)$.
4. Call the bootstrapping procedure. For replications $r = 1, ..., R$, go through the steps:
   a. Bootstrap sample 1 to obtain sample $S_1^{(k,r)}$.
   b. Bootstrap sample 2 to obtain sample $S_2^{(k,r)}$.
   c. Construct the estimator $\hat{\theta}_x^{(k,r)}$ using sample $S_1^{(k,r)}$.
   d. Construct the estimator $\hat{\theta}_y^{(k,r)}$ using sample $S_2^{(k,r)}$.
   e. Derive imputed values $\hat{y}_i^{(k,r)} = F_1(-z_i^T \hat{\theta}_y^{(k,r)})$ using sample $S_1^{(k,r)}$.
   f. Compute the naive estimate $\hat{\mu}_{xy}^{N(k,r)} = \frac{1}{n_1} \sum_i x_i \hat{y}_i^{(k,r)}$ using sample $S_1^{(k,r)}$.
   g. Compute the bias-correction term $BC^{(k,r)} = B(z, \hat{\theta}_x^{(k,r)}, \hat{\theta}_x^{(k,r)}, \rho)$ (see Proposition ...) using sample $S_1^{(k,r)}$.
   h. Compute the bias-corrected estimate $\hat{\mu}_{xy}^{C(k,r)} = \hat{\mu}_{xy}^{N(k,r)} - B^{(k,r)}$.

i. Take the sample standard deviations over the R replications of $\hat{\mu}_{xy}^{N(k,r)}$ and $\hat{\mu}_{xy}^{C(k,r)}$ as estimates of the bootstrapped standard errors of respectively the naive and bias-corrected estimator, denoted by $se^k(\hat{\mu}^N)$ and $se^k(\hat{\mu}^C)$.

5. Compute the average standard errors over the K Monte-Carlo simulations.

6. Take the sample standard deviations over the K Monte-Carlo simulated values of $\hat{\mu}_{xy}^{N(k,r)}$ and $\hat{\mu}_{xy}^{C(k,r)}$, which we will refer to as the 'true' standard errors for respectively the naïve and the bias-corrected estimators.

4.1 Analytical approximation of standard errors

This sub-section considers the bias-corrected estimator only. We compare estimates of the standard errors obtained by: (a) asymptotics (analytic standard errors), (b) bootstrapping, and (c) Monte-Carlo simulation. The latter measures the standard deviation over the K Monte-Carlo simulations of the bias-corrected point estimates which is considered to be closest to the true standard errors. The objective is to see how accurate the asymptotic standard errors are for finite samples with a realistic number of observations.

Figures 1 to 4 examine the sensitivity of the estimates to values of the sample size $n = n_1 = n_2$, the correlation parameter ρ, and the variance of the regressors relative to the idiosyncratic errors σ_z^2, respectively. We included Figure 2, which shows the estimated standard errors multiplied by $\sqrt{n}$ plotted against n, to facilitate the comparison between the different estimates for larger n.

We find that the asymptotic standard errors are remarkably accurate, for both large and small sample sizes (Figures 1 and 2), and for all values considered of ρ and σ_z^2 (Figures 3 and 4 respectively). That is good news as the use of analytical standard errors makes the two-sample cross-tabulation particularly easy to implement as well as fast.

The reported results are for parameter specification 1 (see Table 4). Estimates not included here confirm that the results are similar for other choices of model parameters.

4.2 Model- and and sampling error vs. sample size

How significant is the model error when compared to the sampling error? The simulations presented in Figure 5 show the magnitude of both types of error as we change sample sizes n_1 and n_2. As theory predicts, sampling error declines as we increase n_1, the size of the sample in which we are computing the sample means. In contrast to the conventional one-sample cross-tabulation, the overall standard error does not vanish unless the size of the second sample n_2 too tends to infinity. The latter sample has an important stake in the model error. It is used to identify the model used to impute y into the first sample. The simulations confirm the significance of both sample sizes n_1 and n_2. The simulations also show that model error does not play a secondary role, its magnitude is comparable to that of sampling error.

4.3 Model- and sampling error vs. model parameters

The simulations presented here examine the determinants of the standard error. What makes the errors large or small? We center our attention on the roles played by the model parameters θ, σ_z^2, and μ_z. (The sample variance is expected to be smallest when μ_{xy} tends to either 0 or 1, which is when the systematic part of the model is not centered around zero.)

The analytic approximation of the variance decomposition (see Proposition ...) predicts that:

- θ: The parameter θ impacts the model error by means of the vector $\gamma(\theta) = \rho\theta_y - \theta_x$. Whether increasing γ increases or decreases the model error largely depends on the distribution of z. The term that is linear in γ can make a positive and a negative contribution depending on the sign of μ_z. The term that is quadratic in γ is strictly a non-decreasing function of γ, yet its magnitude depends also on the magnitude of σ_z^2 and μ_z. Note that the effect of θ is non-linear; increasing θ may first decrease (increase) γ before it increases (decreases).

- ρ: The correlation parameter ρ largely controls the non-linearity of θ's effect on the model error. Note that it also determines the significance of the θ_y vector. When ρ tends to zero, θ_y drops out. Unless ρ tends to one, θ_x will generally have a larger impact on the model error than θ_y.

- σ_z^2: Model error can be both an increasing and a decreasing function of σ_z^2. For $\mu_z = 0$, increasing σ_z^2 strictly increases the model error. For non-zero μ_z, the model error tends to decline when increasing σ_z^2. In the latter case the model error has two terms that are a function of σ_z^2. One is increasing while the other decreasing. For a wide range of parameter values we find that the decreasing function dominates.

The simulation results confirm the theoretical predictions. Figures 6 and 7 plot the standard errors for both the sampling error and the model error as functions of θ, σ_z^2, respectively. We observe that: (a) both θ and σ_z^2 have a sizeable impact on the model error; by changing these parameters, the model error can both be made larger and smaller than the sampling error, (b) θ has a non-linear effect on the model error; you can tell when the vector $\rho\theta_y - \theta_x$ is minimized, (c) σ_z^2 increases the model error when $\mu_z = 0$, but decreases the model error when $\mu_z = 0.5$, as predicted.

Note that the results are for selected parameters. Simulations not included show similar findings for different parameters.

4.4 Bias-corrected vs. naive vs. one-sample estimator

Here we focus on the root-means-squared-error (RMSE) as a measure to compare the performance of the bias-corrected to the naive estimator. The conventional one-sample cross-tabulation (based on sample 1) is included as a benchmark, as this denotes the best possible outcome, i.e. no two-sample estimator can do better than the one-sample estimator.

The RMSE as a function of ρ for two different choices of parameter values (specifications 1 and 3, see Table 4) is shown in Figure 8. Since the true value of ρ is not known in practice, we also included bias-corrected estimates that have been obtained with distorted values of ρ; the true ρ plus and minus 0.15 which represents a realistic margin of error for guestimated values used in practice. It is found that the bias-corrected estimator outperforms the naive estimator (build on the naive assumption of $\rho = 0$) for practically all values of ρ. Only when the true ρ is particularly close to zero does the naive estimator outperform the bias-corrected estimator that is based on an imprecise guestimate of ρ. Not surprisingly, the estimators do equally well when the guestimated ρ used by the bias-corrected estimator is as far from the true ρ as the ρ assumed by the naive estimator is ($\rho = 0$), i.e. when ρ equals the distortion imposed on the guestimated value of ρ.

4.5 Sensitivity of bias-corrected estimator to guesstimated value of ρ

We include this subsection to examine in more detail how sensitive the precision of the bias-corrected estimator is to the precision of the guestimated value of ρ. For each true value of ρ we ask: What are the worst possible guestimates of ρ for which the true μ_{xy} (the parameter being estimated) is still included in the 95 percent confidence interval belonging to the bias-corrected estimator? We numerically solve for ρ^1 and ρ^2 using the analytic expression for the asymptotic distribution of the bias-corrected estimator: $\hat{\mu}_{xy}^C(\rho^1) + 1.96\hat{\sigma}^C(\rho^1) = \mu_{xy}$, and $\hat{\mu}_{xy}^C(\rho^2) - 1.96\hat{\sigma}^C(\rho^2) = \mu_{xy}$. The results are shown in Figure 9. We find that the permitted margin of error for the guestimate of ρ for which the bias-corrected estimator still yields accurate estimates is reasonably large,

large enough to accommodate the margin of error we expect to find in practice (see also the empirical section in this paper).

5 Empirical application

5.1 Data and Measurement

We use the following five living standards measurement surveys (LSMS) for our empirical analysis: the Albania Living Standards Measurement Survey for 2002 and 2005, the Tajikistan Living Standards Measurement Survey for 2003 and 2007, and the Malawi Second Integrated Household Survey for 2004-2005. In what follows, we shall refer to these as ALB02, ALB05, TJK03, TJK07 and MLW04, respectively.

All of these surveys are collected by the national statistics office in the respective country in collaboration with the World Bank.[4] They contain a wide array of household-level variables including demographic characteristics, education, employment, housing conditions, detailed consumption expenditure, asset holdings and subjective assessment of poverty. In the upper part of Table 1, some basic summary statistics for these data are provided.

The three countries we study are obviously very different, among other things, in culture, ethnicity, religion, and geographic location. The three countries are also very different in the level of economic development. The Gross Domestic Product per capita on a purchasing power parity basis in 2005 is about USD 6,000, USD 1,500 and USD 600 in Albania, Tajikistan and Malawi, respectively. Table 1 also shows that Albanian households tend to possess more assets than Tajikistan and Malawi households. While our empirical findings may not be directly applicable to other countries, we do not have compelling reasons to believe that our results are driven by the choice of countries.

In this study, we use the consumption poverty (CP) and subjective poverty (SP) measures as dependent variables. We choose this example for two reasons. First, consumption measure has been widely used to define poverty, but it is not obvious how well this "objective" measure of poverty compares with the subjective assessment of poverty. If they are similar, measuring either one of them is sufficient, and fighting CP would be almost the same thing as fighting SP. We can determine whether this is indeed the case by looking at a cross-tabulation table between CP and SP.

Second, the consumption and subjective poverty measures are likely to be correlated even after controlling for a number of observable indicators. This is because some

[4]All the data used in this study as well as the supporting documents are available from the World Bank's LSMS website (`http://www.worldbank.org/lsms/`).

Table 1: Key summary statistics

Variable	ALB02	ALB05	TJK03	TJK07	MLW04
Household Size	4.324	4.272	6.284	6.239	4.547
Ratio of children (-15)	0.254	0.217	0.351	0.323	0.391
Ratio of elderly (64+)	0.113	0.125	0.064	0.063	0.063
HH has a SP	0.862	0.867	0.771	0.775	0.715
Age of HH	50.79	51.91	48.96	51.08	42.46
HH away for 12+ mths	0.001	0.015	0.040	0.001	0.004
HH employed	0.621	0.637	0.649	0.630	0.960
HH completed 8+ yrs schooling	0.846	0.859	0.837	0.893	0.238
HH completed sec.school	0.496	0.515	0.736	0.776	0.137
HH completed college educ	0.173	0.163	0.186	0.194	0.007
Age of spouse	44.52	45.74	42.67	45.17	34.37
SP employed	0.469	0.476	0.489	0.343	0.880
SP completed 8+ yrs schooling	0.868	0.864	0.894	0.900	0.132
SP completed sec. school	0.432	0.407	0.746	0.698	0.068
SP completed college educ	0.137	0.108	0.062	0.078	0.002
Single house	0.632	0.670	0.740	0.761	0.801
Elec. meter within premise	0.704	0.883	0.826	0.836	0.056
Flush toilet	0.764	0.817	0.313	0.207	0.028
Drinking water from pipe	0.699	0.681	0.507	0.516	0.218
Have a telephone	0.295	0.307	0.179	0.244	0.008
Have a mobile phone	0.367	0.776	0.008	0.376	0.030
Have an air conditioner	0.020	0.048	0.052	0.066	0.002
Have a bicycle	0.148	0.123	0.007	0.136	0.362
Have a car	0.099	0.130	0.106	0.158	0.012
have a refrigerator	0.825	0.891	0.333	0.360	0.020
Have a tape or CD player	0.500	0.497	0.280	0.262	0.164
Have a truck	0.025	0.020	0.015	0.028	0.002
Have a washing machine	0.533	0.644	0.125	0.110	0.002
Rural area	0.456	0.451	0.635	0.648	0.872
Capital	0.167	0.176	0.159	0.185	0.043
# Clusters	450	455	208	270	564
# Obs	3599	3638	4160	4860	11257
CP line (PPP USD)	3.616	3.616	3.601	4.220	1.253
ECP line (PPP USD)	2.253	2.253	1.868	2.702	0.778
# Steps	10	10	10	6	6
SP line	3	3	3	2	2
Extreme SP line	2	2	2	1	1

Note: HH and SP refer to household head and spouse, respectively.

unobservable characteristics, such as the inane ability of household members and the presence of relatives in the neighborhoods who provide assistance to the household, are likely to affect both the consumption and subjective well-being of the household. Therefore, we have a strong prior belief that our method would be essential if a CP-SP cross-tabulation table is to be created from two samples.

We adopt the definition of CP widely used in empirical studies. A person is deemed (consumption) poor if the consumption per capita in the person's household falls below the poverty line. We simply use the national poverty lines to make our figures comparable with official poverty figures. These poverty lines are set at the level of consumption that is required to meet the basic food requirement and some non-food consumption.

As shown in Table 1, the poverty lines for CP ("CP line") are not set at the same level across surveys; they range between 1.253 and 3.616 dollars per day per capita. Therefore, poverty figures we report in this study are not comparable across surveys, except for ALB02 and ALB05 whose consumption figures are based on constant prices for year 2002. While the choice of poverty line is not important for the illustration of our methodology, we also check the robustness of our results by using the extreme consumption poverty line ("ECP line" in Table 1), which is set at the consumption level just enough to satisfy the basic food needs.

The subjective measure of poverty is based on the subjective assessment of their position on the economic ladder. Each survey asks a question like the following: "imagine a k-step ladder where on the bottom, the first step, stand the poorest people, and on the highest step, the kth, stand the rich. On which step are you today?" The number of steps k varies from survey to survey as shown in Table 1. We set the subjective poverty line ("SP line" in Table 1) at around the bottom 30th percentile. As with the CP, we use the extreme subjective poverty line ("ESP line"), which is set at around the bottom 20th percentile, to check the robustness of our results. It would be reasonable say that those under the SP or ESP line think of themselves poor at least in relative terms, though alternative choice of poverty lines would be possible. Hereafter, we shall simply refer to the combination of CP and SP poverty lines as "normal" poverty lines, whereas the combination of ECP and ESP poverty lines as "extreme" poverty lines.

5.2 Bivariate Probit Estimation of ρ

Since our results depend on the educated guess ("guesstimate") of the unknown parameter ρ, it is important to assess how stable ρ is across countries and over time. Therefore, we estimate a bivariate probit regression model using an identical set of regressors for each of the five data sets to make the estimates comparable, which is all the variables above the double line in Table 1. These variables are relatively easy to observe and included in many LSMS data sets.

The first column ("Baseline") in Table 2 reports the estimates of ρ for each combination of poverty lines and for each data set.[5] As can be seen from the estimates, the point estimates do vary across countries, but the range of estimates is only 0.177. From these estimates, we can take a guesstimate of $\tilde{\rho} = 0.28$ for our model. None of the estimates are different from this guesstimate by more than 0.1.

Our estimates also indicate that the value of ρ is reasonably stable over time and with respect to the choice of poverty lines. Therefore, if ρ can be estimated using a past data set in a given country, the estimate of ρ is likely to be accurate enough for the purpose bias correction.

Since the estimates of ρ depends on the choice of regressor, it is important check how the choice of regressors matters. To this end, we have also run bivariate regressions with only the constant term, which is reported in the second column ("Const. only") in Table 2. As can be seen from this table, the value of ρ in this column is generally higher than the baseline values of ρ in the first column. This clearly indicates that the guesstimate of ρ should not be taken from a completely different model.

A natural question that arises here is how much ρ changes due to the model selection. To address this, we have randomly selected $l \in \{5, 10, 15, 20, 25\}$ variables out of the thirty variables in the baseline model and estimated ρ for 400 times.[6] We find that the point estimates of ρ tend to be higher when l is smaller, and that the standard errors due to model selection never exceeds 0.045.

It should be noted that the estimates of ρ for the constant-only model do not have a straightforward relationship with the unconditional correlation between consumption and subjective poverty, which is reported in the third column ("Uncond.") in Table 2. This is because unconditional correlation depends not only on ρ but on the coefficients on the constant terms in the bivariate probit model. As a result, the unconditional correlation *per se* is not particularly informative of the value of ρ.

The results presented here are limited both in scale and scope; we are only using five data sets and the application is made only for the cross tabulation of consumption and subjective poverty. Whether ρ is stable in other data sets or for other problems is an empirical question we are unable to answer here. However, our results do indicate that, with sufficient prior information, having a guesstimate that is off the true parameter value by 0.15 is not an unreasonable requirement.

[5] The full regression results for the normal poverty lines are reported in Tables 5 to 9 in the Appendix.
[6] The details are reported in Table 10 in the Appendix.

Table 2: Estimation of conditional and unconditional correlations

Data	Poverty Lines	Baseline		Const. only		Uncond.
ALB02	Normal	0.212	(0.047)	0.501	(0.032)	0.301
	Extreme	0.333	(0.075)	0.518	(0.055)	0.219
ALB05	Normal	0.362	(0.048)	0.562	(0.033)	0.312
	Extreme	0.334	(0.087)	0.565	(0.060)	0.218
TJK03	Normal	0.198	(0.037)	0.305	(0.030)	0.166
	Extreme	0.212	(0.033)	0.297	(0.029)	0.176
TJK07	Normal	0.185	(0.030)	0.271	(0.027)	0.168
	Extreme	0.282	(0.047)	0.321	(0.043)	0.142
MLW04	Normal	0.285	(0.024)	0.438	(0.018)	0.240
	Extreme	0.276	(0.021)	0.350	(0.018)	0.206

Note: Figures in parentheses are the standard errors.

5.3 Evaluation of bias correction

The usefulness of our methodology rests on the validity of the assumptions about the underlying model and whether a guesstimate is good enough to allow a reasonable bias correction. Therefore, it is important to verify that our estimation procedure works in a practical setting. To this end, we use a simulation procedure described below to compare the following three types of two-sample estimators with the standard one-sample estimator: best, guesstimate, and naïve two-sample estimators.

The "best" two-sample estimator is the bias-corrected two-sample estimator that uses the point estimate of ρ in the first column of Table 2. This is a benchmark case where we have a very accurate estimate of ρ. The second is the guesstimate two-sample estimator, which is the bias-corrected estimator that uses $\tilde{\rho}$ regardless of the survey or poverty line used for estimation. This is a realistic case where we have some prior information about ρ, but the information is not as good as the best case. Finally, we consider the naïve two-sample estimator, in which no bias correction is made.

To focus on the applicability of two-sample estimation, let us for now treat the original sample as a simple random sample. With this simplifying assumption, we independently draw with replacement two bootstrap samples of the original sample size in each of the 400 rounds of simulations. In each round, we compute the point estimates and analytic standard errors (ASE) for all the cross-tabulation estimators.[7] The reported point estimate and the bootstrap standard errors (BSE) are given by the mean and standard deviation of the point estimates taken over simulations, whereas

[7] For the analytic standard error of one-sample estimation, we simply adopt the formula for a binomial distribution (i.e., $n^{-1}\hat{\mu}^N_{xy}(1-\hat{\mu}^N_{xy})$).

the reported analytic standard errors are the average of the analytic standard errors over simulations.

Two points should be noted here. First, it is inappropriate to randomly split the original sample into two samples. Random splitting violates the assumption of two independence samples. It can also result in misleadingly small standard errors when ρ is positive. To see this point, suppose that Sample 1 happens to contain a large fraction of positive ϵ_{yi} and Sample 2 a large fraction of negative ϵ_{yi}. In this case, $\hat{y}_i$ tends to underestimate the true value of $\bar{y}_i$ because θ_y is estimated only with Sample 2. On the other hand, since ρ is positive, a large fraction of ϵ_{xi} is likely to be positive in Sample 1. Therefore, the value of x_i tends to be larger than $\bar{x}_i$ on average in Sample 1, offsetting the effect of the underestimation of $\bar{y}_i$. As a result, the two-sample estimates tend to be relatively stable under random splitting when ρ is positive.

Second, while using two independently-drawn samples does increase the total number of observations used for estimation, the number of observations used in any part of the two-sample estimation never exceed the original sample size. Therefore, we are in fact making a fair comparison between the one-sample and two-sample estimators.

Let us now turn to the main estimation results reported in Table 3. NC and NS stand for not consumption poor and not subjectively poor, respectively. The first column ("One smpl") presents the one-sample estimates. For example, the proportion of people who are both consumption poor and subjectively poor in Albania in 2002 is 19.21 percent. The ASE and BSE for this point estimate are 0.66 percent and 0.75 percent respectively. The second column ("BC(best)") reports the best two-sample estimator, whereas the third column ("BC(guest.)") reports the guesstimate two-sample estimator. The fourth column ("Naïve)") reports the naïve estimator that does not correct for the correlation between unobservable error terms.

Table 3 shows that the point estimates in the second column are very close to those in the first column, suggesting that our bias-correction method works extremely well with a direct estimate of ρ. However, direct estimation of ρ is typically not available in a situation where two-sample estimation is required.[8]

To make a more realistic assessment, it is important to look at the results for the guesstimate two-sample estimator reported in the third column. As expected, these estimates are generally not as close to the first column as the second column. However, they are still close enough to the one sample estimate. In fact, when we carry out a

[8] Note that two-sample estimation may be better than one-sample estimation even when one-sample estimation (and direct estimation of ρ) is possible. Such a situation arises when there are two large cross-section data sets with a very small number of overlapping observations. By combining the two large samples, both the sampling and model errors could be made very small so that the resulting estimates are more accurate than the one-sample estimate.

Table 3: Cross-tabulation tables with the normal poverty lines.

Data			One smpl		BC (best)		BC (guest.)		Naïve	
			NS	SP	NS	SP	NS	SP	NS	SP
ALB02	NC	Est.	43.99	30.62	43.67	30.89	44.19	30.47	42.29	32.33
		ASE	(0.83)	(0.77)	(0.95)	(0.90)	(0.95)	(0.90)	(0.95)	(0.89)
		BSE	[1.09]	[1.05]	[0.98]	[0.90]	[1.00]	[0.91]	[0.91]	[0.88]
	CP	Est.	6.18	19.21	6.39	19.06	5.82	19.52	7.85	17.52
		ASE	(0.40)	(0.66)	(0.48)	(0.78)	(0.47)	(0.78)	(0.50)	(0.78)
		BSE	[0.40]	[0.75]	[0.45]	[0.80]	[0.46]	[0.79]	[0.47]	[0.78]
ALB05	NC	Est.	47.70	33.80	45.86	34.91	45.28	35.45	45.61	35.96
		ASE	(0.83)	(0.78)	(0.99)	(0.97)	(0.99)	(0.96)	(0.97)	(0.92)
		BSE	[1.01]	[1.01]	[1.22]	[1.06]	[1.04]	[0.95]	[0.97]	[0.91]
	CP	Est.	3.39	15.12	3.61	15.62	4.13	15.14	5.61	12.82
		ASE	(0.30)	(0.59)	(0.39)	(0.72)	(0.40)	(0.72)	(0.41)	(0.69)
		BSE	[0.31]	[0.65]	[0.33]	[0.79]	[0.36]	[0.76]	[0.39]	[0.69]
TJK03	NC	Est.	11.03	5.96	10.91	6.05	11.59	5.40	9.40	7.56
		ASE	(0.49)	(0.37)	(0.42)	(0.34)	(0.42)	(0.33)	(0.42)	(0.34)
		BSE	[0.60]	[0.50]	[0.47]	[0.31]	[0.45]	[0.30]	[0.44]	[0.35]
	CP	Est.	35.63	47.38	35.69	47.35	35.08	47.93	37.20	45.84
		ASE	(0.74)	(0.77)	(0.86)	(0.88)	(0.86)	(0.88)	(0.86)	(0.88)
		BSE	[0.83]	[0.95]	[0.86]	[0.84]	[0.94]	[0.91]	[0.84]	[0.86]
TJK07	NC	Est.	34.73	12.18	34.65	12.26	35.82	11.08	32.46	14.45
		ASE	(0.68)	(0.47)	(0.71)	(0.45)	(0.70)	(0.45)	(0.72)	(0.47)
		BSE	[0.82]	[0.62]	[0.76]	[0.45]	[0.74]	[0.43]	[0.76]	[0.48]
	CP	Est.	30.79	22.29	30.84	22.25	29.71	23.39	33.03	20.06
		ASE	(0.66)	(0.60)	(0.78)	(0.64)	(0.78)	(0.64)	(0.76)	(0.62)
		BSE	[0.80]	[0.68]	[0.81]	[0.64]	[0.79]	[0.68]	[0.79]	[0.63]
MLW04	NC	Est.	12.86	34.75	11.93	35.04	11.94	35.06	10.86	36.70
		ASE	(0.32)	(0.45)	(0.31)	(0.49)	(0.31)	(0.49)	(0.31)	(0.48)
		BSE	[0.37]	[0.51]	[0.66]	[0.52]	[0.65]	[0.48]	[0.31]	[0.50]
	CP	Est.	4.58	47.81	4.69	48.34	4.73	48.26	6.43	46.01
		ASE	(0.20)	(0.47)	(0.21)	(0.51)	(0.21)	(0.51)	(0.23)	(0.52)
		BSE	[0.20]	[0.51]	[0.21]	[0.67]	[0.20]	[0.69]	[0.23]	[0.51]

Note: All figures in percentages. ASE in parentheses and BSE in square brackets.

piecewise z-test of equality between the one-sample, the null hypothesis could not be rejected at a 5 percent significance level regardless of whether the direct estimation or the guesstimate is used.[9] This indicates that our guess is good enough for practical purposes. However, the one-sapmle estimates and the naïve estimates are statistically different from each other at least in one cell for each of the five surveys. Therefore, while the naïve estimates do not appear to be completely unreasonable, statistical inferences based on the naïve estimates are likely to be misleading.

Table 3 also shows that ASE and BSE are generally very close, though there are some gaps. The gaps may arise for a number of reasons, including the misspecification of the underlying model and poor finite-sample propertes. However, we do not have evidence that suggests the ASE is systematically biased. In the subsequent discussions, we shall use the bootstrap standard errors. However, our conclusion will not be affected even if analytic standard errors are used. In Table 11 in the appendix, we also show the results of variance decomposition. While the patterns vary across data sets and cells, the idiosyncratic error, z-samling error and y-model error tend to be larger than x-model error and covariance terms.

The problem with the naïve estimates is even more pronounced when we derive the conditional distribution of poverty. Table 4 presents four conditional probabilities. For example, CP|SP is the probability that the person is consumption poor given that he is subjectively poor. CP|NS, on the other hand, is the probability that the person is consumption poor given that he is not subjectively poor. As expected from the discussion above, the best two-sample estimates are very close to one-sample estimates. Table 4 also shows that the guesstimate two-sample estimates also do well, whereas the naïve estimator can be very different from the one-sample estimator. As with before, the bias-corrected estimators never failed to pass the z-test of equality at a 5 percent significance level, whereas the equality was rejected for the naïve estimator in fourteen out of twenty piecewise comparisons at the same significance level.

To investigate the robustness of our main results, we have first checked our results with the extreme poverty lines. The estimates of the joint distribution and conditional distribution are reported in Tables 12 and 13 in the appendix. While there are a few cases where the naïve estimates are closer to the one-sample estimates, the two-sample estimates are much closer to the one-sample estimates on average.

We have also checked the robustness of our results with respect to the sampling

[9] Let $\hat{\mu}_O$ and $\hat{\mu}_T$ be the one-sample and two-sample estimates, respectively, and $\hat{V}_B[\hat{\mu}]$ be the bootstrap estimate of $\hat{\mu}$. The two-sided z-statistic we used here is $Z = \frac{|\hat{\mu}_O - \hat{\mu}_T|}{\sqrt{\hat{V}[\hat{\mu}_O] + \hat{V}[\hat{\mu}_T]}}$. We assume the independence of the two estimates, which is justified because the bootstrap simulations for the two estimators are carried out independently.

Table 4: Conditional distribution of poverty for normal poverty lines.

		One sample				Two-sample (best)				Two-sample (guest)				Naïve			
		CP|SP	CP|NS	SP|CP	SP|NC	CP|SP	CP|NS	SP|CP	SP|NC	CP|SP	CP|NS	SP|CP	SP|NC	CP|SP	CP|NS	SP|CP	SP|NC
ALB02	Est.	38.54	12.31	75.66	41.04	38.15	12.76	74.90	41.43	39.04	11.63	77.04	40.82	35.15	15.66	69.05	43.33
	BSE	[1.38]	[0.80]	[1.75]	[1.24]	[1.33]	[0.83]	[1.42]	[1.08]	[1.29]	[0.84]	[1.44]	[1.10]	[1.32]	[0.85]	[1.54]	[1.02]
ALB05	Est.	30.90	6.64	81.68	41.47	30.91	7.30	81.23	43.22	29.92	8.36	78.57	43.91	26.27	10.96	69.53	44.08
	BSE	[1.27]	[0.60]	[2.06]	[1.10]	[1.30]	[0.66]	[1.41]	[1.28]	[1.29]	[0.71]	[1.38]	[1.08]	[1.23]	[0.72]	[1.70]	[1.03]
TJK03	Est.	88.83	76.36	57.08	35.06	88.67	76.58	57.03	35.66	89.87	75.17	57.74	31.79	85.84	79.83	55.20	44.59
	BSE	[0.91]	[1.14]	[1.03]	[1.93]	[0.55]	[0.96]	[0.94]	[1.10]	[0.52]	[0.94]	[1.05]	[1.14]	[0.64]	[0.88]	[0.94]	[1.19]
TJK07	Est.	64.66	46.99	41.99	25.97	64.48	47.09	41.91	26.13	67.87	45.34	44.05	23.62	58.13	50.43	37.79	30.80
	BSE	[1.49]	[1.06]	[1.35]	[1.03]	[1.00]	[1.02]	[1.03]	[0.78]	[0.91]	[0.96]	[1.11]	[0.79]	[1.05]	[1.00]	[0.99]	[0.83]
MLW04	Est.	57.91	26.27	91.26	73.00	57.97	28.25	91.16	74.61	57.92	28.41	91.07	74.60	55.63	37.19	87.74	77.16
	BSE	[0.56]	[1.14]	[0.45]	[0.61]	[0.57]	[1.49]	[0.36]	[1.17]	[0.57]	[1.50]	[0.34]	[1.12]	[0.55]	[0.95]	[0.42]	[0.60]

Note: All figures in percentage. Bootstrap standard errors in square brackets.

design. To this end, we have estimated θ allowing for clustering. Further, we have carried out two-stage bootstrapping to account for the sampling design. The standard errors tend to get larger for all the estimators, but the advantage of the bias-corrected estimators over the naïve estimator still hold.

5.4 Tips for practitioners

As discussed in Remark 4, one can always create cross-tabulation tables using the two-sample estimators in two different ways; one way is to estimate θ_y using Sample 2 and impute x_i in Sample 1. The other way is to estimate θ_x using Sample 1 and impute y_i in Sample 2.

The estimates produced in these two different ways coincide with each other asymptotically. In practice, however, the two estimates are not identical. To see how much the direction of imputation affects the results, we have estimated the cross-tabulation table by imputing x_i instead of y_i. As Table 14 in the appendix shows, the differences between the two estimates are remarkably small.

It should also be noted that the standard errors are the asymptotic variance is in general different. Since the variance from the smaller sample tends to dominate, the better choice of the direction of imputation would depend on whether the sampling error component of the variance dominates the model error component or vice versa. The former is likely to dominate when z has a high variance. On the other hand, the latter would dominate when the estimate of θ is poor, which is likely to be the case when z has a small variance. Therefore, it is likely to be better to use the smaller sample for estimation [imputation] if z has a high [low] variance in that sample.

In practical applications, the underlying populations for the two samples may not be identical. For example, it is common to combine two samples collected at different points in time. In this case, we need an additional assumption that the parameter value is time invariant. With this assumption, we can make two-sample tabulations in two different ways. However, they may not be the same because the distribution of z may change over time. In fact, the two-sample estimators produces the estimates of μ_{xy} for the underlying population of the sample for imputation.

If the cross-tabulation tables produced from two-way imputation are very different, there may be misspecification of the model or the violation of the constancy of model parameters over time. Therefore, it is generally useful to produce the cross-tabulation tables by the bi-directional imputation. This exercise serves as a robust check of the estimates.

6 Concluding remarks

Standard errors associated with conventional cross-tabulations are easily computed and tend to be small. The sole source of estimation error is sampling error. When no one sample contains both variables, two-sample cross-tabulations denote a viable alternative where one of the two variables is imputed. This means that imputation error is added as a second source of error. While it is straightforward to cross-tabulate observed data with imputed data, conventional standard errors no longer apply. It is tempting to treat the imputed data as observed data but this ignores the imputation error and thereby underestimates the standard error. Similarly, treating the sample as a census, focussing solely on the imputation error, ignores the samping error (see e.g. Elbers et al., 2003 and Ivaschenko and Lanjouw, 2010). Accounting for both types of error can be achieved by means of a double bootstrap (one bootstrap for each type of error) which provides precise standard errors regardless of sample size. Implementing this bootstrap procedure however will likely pose a hurdle for many applied users.

To provide an alternative to bootstrapping we derived the asymptotic distribution for the two-sample cross-tabulation. This gives the applied user analytic standard errors that are easily implemented without the need for simulation, and account for both imputation- and sampling error. While the analytic standard errors are only valid asymptotically, Monte Carlo simulations confirm that they perfectly coincide with the boostrap standard errors also for small sample sizes.

The ultimate objective is to make two-sample cross-tabulations both precise and user-friendly. To this end, we implemented both the asymptotic and bootstrap versions in STATA; the ado-files are available upon request by the authors. Suggestions for further research include extensive testing of the approach using a variety of empirical data.

A appendix

A.1 Lemmas

In this section, we shall prove some useful lemmas.

Lemma 23 *Under Assumption 12,* $\Phi_{2x}(x,y,\rho) = \phi_1(x)\Phi_1(\frac{y-\rho x}{\sqrt{1-\rho^2}})$ *and* $\Phi_{2y}(x,y,\rho) = \phi_1(y)\Phi_1(\frac{x-\rho y}{\sqrt{1-\rho^2}})$.

Proof of Lemma 23 Because the proof for Φ_{2y} is similar, we only provide the proof for Φ_{2x}.

$$\begin{aligned}
\Phi_{2x}(x,y,\rho) &= \int_{-\infty}^{y} \phi_2(x,s,\rho)\mathrm{d}s \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{y} \exp\left(-\frac{1}{2}\frac{(s-\rho x)^2}{1-\rho^2} - \frac{x^2}{2}\right) \mathrm{d}s \\
&= \phi_1(x)\frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{\frac{y-\rho x}{\sqrt{1-\rho^2}}} \exp\left(-\frac{s'}{2}\frac{\mathrm{d}s}{\mathrm{d}s'}\right) \mathrm{d}s' \\
&= \phi_1(x)\Phi_1\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right),
\end{aligned}$$

where $s' \equiv (s-\rho x)/\sqrt{1-\rho^2}$. □

Lemma 24 *Let $k \geq 0$. The k-th order derivative of Φ_2 with respect to ρ evaluated at $\rho = 0$ satisfies:*

$$\left.\frac{\partial^k \Phi_2(x,y,\rho)}{\partial \rho^k}\right|_{(x,y,0)} = \Phi_1^{(k)}(x)\Phi_1^{(k)}(y), \tag{18}$$

where $\Phi_1^{(k)}(x)$ denotes the k-th order derivative of the univariate standard normal cumulative distribution function $\Phi_1(x)$.

Proof of Lemma 24 First, eq. (18) holds for $k = 0$:

$$\Phi_2(x,y,0) = \int_{-\infty}^{x}\int_{-\infty}^{y} \phi_2(r,s,0)\mathrm{d}r\mathrm{d}s = \left(\int_{-\infty}^{x} \phi_1(r)\mathrm{d}r\right)\cdot\left(\int_{-\infty}^{y} \phi_1(s)\mathrm{d}s\right) = \Phi_1(x)\Phi_1(y).$$

Now suppose $k \geq 1$. Letting $l \equiv k-1$, we can rewrite Eq.(18) as follows:

$$\phi_2^{(l)}(x,y,0) = \phi_1^{(l)}(x)\phi_1^{(l)}(y), \tag{19}$$

where $\phi_1^{(l)}(w)$ is the l-th order derivative of $\phi_1(w)$ for $w \in \{x,y\}$, and $\phi_2^{(l)}(x,y,\rho)$ is the l-th order derivative of $\phi_2(x,y,\rho)$ with respect to ρ.

Any probability density function may be written as a unique transform of its characteristic function. For ϕ_1 and ϕ_2, we have:

$$\begin{aligned}
\phi_1(w) &= (2\pi)^{-1}\int_{-\infty}^{\infty} \exp\left(-itw - \frac{t^2}{2}\right)\mathrm{d}t \\
\phi_2(x,y,\rho) &= (2\pi)^{-2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-i(t_1x+t_2y) - \frac{t_1^2+2\rho t_1t_2+t_2^2}{2}\right)\mathrm{d}t_1\mathrm{d}t_2.
\end{aligned}$$

Therefore, for $w \in \{x, y\}$, we have:

$$\phi_1^{(l)}(w) = (2\pi)^{-1} \int_{-\infty}^{\infty} \frac{\partial^l}{\partial x^l} \left[\exp\left(-itw - \frac{t^2}{2} \right) \right] \mathrm{d}t = (2\pi)^{-1} \int_{-\infty}^{\infty} (-it)^l \exp\left(-itw - \frac{t^2}{2} \right) \mathrm{d}t.$$

Using this, we can prove Eq.(19) as follows:

$$\begin{aligned}
\phi_2^{(l)}(x, y, 0) &= (2\pi)^{-2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^l}{\partial \rho^l} \left[\exp\left(-i(t_1 x + t_2 y) - \frac{t_1^2 + 2\rho t_1 t_2 + t_2^2}{2} \right) \right] \Bigg|_{\rho=0} \mathrm{d}t_1 \mathrm{d}t_2 \\
&= (2\pi)^{-2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (-t_1 t_2)^l \exp\left(-i(t_1 x + t_2 y) + \frac{t_1^2 + t_2^2}{2} \right) \mathrm{d}t_1 \mathrm{d}t_2 \\
&= \left[(2\pi)^{-1} \int_{-\infty}^{\infty} (-it_1)^l \exp\left(-it_1 x + \frac{t_1^2}{2} \right) \mathrm{d}t_1 \right] \\
&\quad \times \left[(2\pi)^{-1} \int_{-\infty}^{\infty} (-it_2)^l \exp\left(-it_2 y + \frac{t_2^2}{2} \right) \mathrm{d}t_2 \right] \\
&= \phi_1^{(l)}(x) \phi_1^{(l)}(y).
\end{aligned}$$

Hence, eq.(18) holds for $k \geq 1$ as well, completing the proof of the lemma. □

Lemma 25 *Let $D \equiv a_{xx}x^2 + a_x x + a_{yy}y^2 + a_y y - 2a_{xy}xy$, where $a_{xx} > 0$, $a_{yy} > 0$ and $a_{xx}a_{yy} - a_{xy}^2 > 0$. Then, we have:*

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{D}{2}} \mathrm{d}x \mathrm{d}y = \frac{2\pi \mathrm{e}^{\frac{a_x^2 a_{yy} + a_y^2 a_{xx} + 2a_x a_y a_{xy}}{8(a_{xx}a_{yy} - a_{xy}^2)}}}{\sqrt{a_{xx}a_{yy} - a_{xy}^2}} \equiv \tilde{A} \tag{20}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w \mathrm{e}^{-\frac{D}{2}} \mathrm{d}x \mathrm{d}y = \tilde{A} \tilde{\mu}_w \quad \text{for} \quad w \in \{x, y\} \tag{21}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \mathrm{e}^{-\frac{D}{2}} \mathrm{d}x \mathrm{d}y = \tilde{A}(\tilde{\mu}_x \tilde{\mu}_y + \tilde{\rho} \tilde{\sigma}_x \tilde{\sigma}_y) \tag{22}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w^2 \mathrm{e}^{-\frac{D}{2}} \mathrm{d}x \mathrm{d}y = \tilde{A}(\tilde{\mu}_w^2 + \tilde{\sigma}_w^2) \quad \text{for} \quad w \in \{x, y\} \tag{23}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 y^2 \mathrm{e}^{-\frac{D}{2}} \mathrm{d}x \mathrm{d}y = \tilde{A} \left((\tilde{\mu}_x^2 + \tilde{\sigma}_x^2)(\tilde{\mu}_y^2 + \tilde{\sigma}_y^2) + 2\tilde{\sigma}_x \tilde{\sigma}_y \tilde{\rho} (2\tilde{\mu}_x \tilde{\mu}_y + \tilde{\rho} \tilde{\sigma}_x \tilde{\sigma}_y) \right), \tag{24}$$

where $\tilde{\mu}_x \equiv \frac{a_x a_{yy} + a_y a_{xy}}{2(a_{xy}^2 - a_{xx}a_{yy})}$, $\tilde{\mu}_y \equiv \frac{a_y a_{xx} + a_x a_{xy}}{2(a_{xy}^2 - a_{xx}a_{yy})}$, $\tilde{\sigma}_x \equiv \sqrt{\frac{a_{yy}}{a_{xx}a_{yy} - a_{xy}^2}}$, $\tilde{\sigma}_y \equiv \sqrt{\frac{a_{xx}}{a_{xx}a_{yy} - a_{xy}^2}}$, and $\tilde{\rho} \equiv \frac{a_{xy}}{\sqrt{a_{xx}a_{yy}}}$.

Proof of Lemma 25 It is straightforward to verify that the following relationship holds:

$$D = \frac{1}{1-\tilde{\rho}^2}\left[\left(\frac{x-\tilde{\mu}_x}{\tilde{\sigma}_x}\right)^2 + \left(\frac{y-\tilde{\mu}_y}{\tilde{\sigma}_y}\right)^2 - 2\tilde{\rho}\left(\frac{x-\tilde{\mu}_x}{\tilde{\sigma}_x}\right)\left(\frac{y-\tilde{\mu}_y}{\tilde{\sigma}_y}\right)\right] + \frac{a_x^2 a_{yy} + a_y^2 a_{xx} + 2a_x a_y a_{xy}}{4(a_{xy}^2 - a_{xx}a_{yy})}.$$

Therefore, letting $\tilde{x} = \frac{x-\tilde{\mu}_x}{\tilde{\sigma}_x}$ and $\tilde{y} = \frac{y-\tilde{\mu}_y}{\tilde{\sigma}_y}$, we have:

$$\mathrm{e}^{-\frac{D}{2}}\mathrm{d}x\mathrm{d}y = \tilde{A}\phi_2(\tilde{x},\tilde{y},\tilde{\rho})\mathrm{d}\tilde{x}\mathrm{d}\tilde{y}.$$

Eq. (20) immediately follows from this. We arrive at Eqs. (21)-(24) by integrating the following equations over the $\mathbb{R}^2$ space.

$$\begin{aligned}
x\mathrm{e}^{-\frac{D}{2}}\mathrm{d}x\mathrm{d}y &= \tilde{A}(\tilde{\mu}_x + \tilde{\sigma}_x\tilde{x})\phi_2(\tilde{x},\tilde{y},\tilde{\rho})\mathrm{d}\tilde{x}\mathrm{d}\tilde{y} \\
xy\mathrm{e}^{-\frac{D}{2}}\mathrm{d}x\mathrm{d}y &= \tilde{A}(\tilde{\mu}_x + \tilde{\sigma}_x\tilde{x})(\tilde{\mu}_y + \tilde{\sigma}_y\tilde{y})\phi_2(\tilde{x},\tilde{y},\tilde{\rho})\mathrm{d}\tilde{x}\mathrm{d}\tilde{y} \\
x^2\mathrm{e}^{-\frac{D}{2}}\mathrm{d}x\mathrm{d}y &= \tilde{A}(\tilde{\mu}_x + \tilde{\sigma}_x\tilde{x})^2\phi_2(\tilde{x},\tilde{y},\tilde{\rho})\mathrm{d}\tilde{x}\mathrm{d}\tilde{y} \\
x^2y^2\mathrm{e}^{-\frac{D}{2}}\mathrm{d}x\mathrm{d}y &= \tilde{A}(\tilde{\mu}_x + \tilde{\sigma}_x\tilde{x})^2(\tilde{\mu}_y + \tilde{\sigma}_y\tilde{y})^2\phi_2(\tilde{x},\tilde{y},\tilde{\rho})\mathrm{d}\tilde{x}\mathrm{d}\tilde{y} \\
&= \left(\tilde{\mu}_x^2\tilde{\mu}_y^2 + \tilde{\mu}_x^2\tilde{\sigma}_y^2\tilde{y}^2 + \tilde{\mu}_y^2\tilde{\sigma}_x^2\tilde{x}^2 + 4\tilde{\mu}_x\tilde{\mu}_y\tilde{\sigma}_x\tilde{\sigma}_y\tilde{x}\tilde{y} + \tilde{\sigma}_x^2\tilde{\sigma}_y^2\tilde{x}^2\tilde{y}^2\right)\phi_2(\tilde{x},\tilde{y},\tilde{\rho})\mathrm{d}\tilde{x}\mathrm{d}\tilde{y},
\end{aligned}$$

where we let $w = x$ in eqs. (21) and (23). The proof for $w = y$ is similar. □

Lemma 26 *Let us define the following:*

$$\begin{aligned}
d_{xi}(\hat{\theta}_x) &\equiv \frac{f_{1x}(-z_i^T\hat{\theta}_x)(x_i - \bar{x}_i)z_i}{\hat{x}_i(1-\hat{x}_i)} \\
m_{3xi}(\hat{\theta}_x,\tilde{\theta}_x) &\equiv \frac{f_{1x}(-z_i^T\hat{\theta}_x)f_{1x}(-z_i^T\tilde{\theta}_x)z_i z_i^T}{\bar{x}_i(1-\bar{x}_i)} \\
m_{3x}(\hat{\theta}_x,\tilde{\theta}_x) &\equiv n_1^{-1}\sum_{i\in\mathcal{S}_1} m_{3xi}(\hat{\theta}_x,\tilde{\theta}_x).
\end{aligned}$$

Then, under Assumptions 7 and 16 and some regularity conditions, there exist θ_x^+ such that each of its component is between the corresponding components of θ_x and $\hat{\theta}_x$ that satisfies eq. (25) below:

$$\begin{aligned}
\sqrt{n_1}(\hat{\theta}_x - \theta_x) &= m_{3x}^{-1}(\hat{\theta}_x,\theta_x^+)n_1^{-1/2}\sum_i d_{xi}(\hat{\theta}_x) && (25) \\
&\xrightarrow{d} \mathcal{N}(0, E^{-1}[m_{3xi}(\theta_x,\theta_x)]) \quad \text{as } n_1 \to \infty. && (26)
\end{aligned}$$

Proof of Lemma 26 Let us define the log likelihood function as follows:

$$L(\theta) \equiv \sum_i x_i \ln F_{1x}(-z_i^T\theta) + (1-x_i)\ln(1-F_{1x}(-z_i^T\theta))$$

By the definition of the maximum likelihod estimator, we have:

$$\hat{\theta}_x = \arg\max_\theta L(\theta).$$

Taking the first order condition and applying the Mean Value Theorem, we have:

$$\begin{aligned} 0 = \nabla_\theta L(\hat{\theta}_x) &= \sum_{i\in\mathcal{S}_1} \frac{(\hat{x}_i - x_i) f_{1x}(-z_i^T\hat{\theta}_x) z_i}{\hat{x}_i(1-\hat{x}_i)} \\ &= \sum_{i\in\mathcal{S}_1} \left[\bar{x}_i + f_{1x}(-z_i^T\theta_x^+) z_i^T(\hat{\theta}_x - \theta_x) - x_i\right] \frac{f_{1x}(-z_i^T\hat{\theta}_x) z_i}{\hat{x}_i(1-\hat{x}_i)} \\ &= m_{3x}(\hat{\theta}_x, \theta_x^+)(\hat{\theta}_x - \theta_x) - \frac{1}{n_1}\sum_{i\in\mathcal{S}_1} d_{xi}(\hat{\theta}_x), \end{aligned}$$

where θ^+ satisfies the condition in Lemma 26. Eq. (25) follows directly from this.

By the law of large numbers and the consistency of $\hat{\theta}_x$, we have:

$$m_{3x}(\hat{\theta}_x, \theta_x^+) \xrightarrow{p} E[m_{3xi}(\theta_x, \theta_x)].$$

With this and by the law of large numbers and the consistency of $\hat{\theta}_x$, we have:

$$\sqrt{n}(\hat{\theta}_x - \theta_x) \xrightarrow{d} (0, E^{-1}[m_{3xi}(\theta_x, \theta_x)] Var[d_{xi}(\theta)] E^{-1}[m_{3xi}(\theta_x, \theta_x)]).$$

Eq. (26) holds because

$$Var[d_{xi}(\theta)] = E\left[\frac{f_{1x}^2(-z_i^T\theta_x) z_i z_i^T}{\bar{x}_i^2(1-\bar{x}_i)^2}\left((\bar{x}_i - 1)^2\bar{x}_i + \bar{x}_i^2(1-\bar{x}_i)\right)\right] = E[m_{3x}(\theta_x, \theta_x)].$$

□

A.2 Proofs

Proof of Proposition 5 Since $\hat{\theta}_y = \theta_y$, we have $\hat{y}_i = \bar{y}_i$. Therefore, $B = E[\hat{\mu}_{xy}^N - \mu_{xy}] = E_z[E[x_i\bar{y}_i - x_i y_i | z_i]] = E_z[\bar{x}_i\bar{y}_i - \overline{x_i y_i}] = -E_z[Cov[x_i|z_i, y_i|z_i]]$. □

Proof of Theorem 14 Applying Lemma 24 for $k = 1, 2, 3$ provides the derivatives of

$\Phi_2(Z_x, Z_y, \rho)$ with respect to ρ evaluated at $\rho = 0$.[10] Therefore, the third-order Taylor expansion of $\Phi_2(Z_x, Z_y, \rho)$ with respect to ρ around $\rho = 0$ is:

$$\Phi_2(Z_x, Z_y, \rho) \simeq \Phi_1(Z_x)\Phi_1(Z_y) + \left(\rho + \frac{\rho^2}{2} Z_x Z_y + \frac{\rho^3}{6}(Z_x^2 - 1)(Z_y^2 - 1)\right) \phi_1(Z_x)\phi_1(Z_y).$$

Substituting this approximation for $\Phi_2(Z_x, Z_y, \rho)$ into eq. (5) and rearranging terms, we have:

$$B \simeq -E_z \left[\left(\rho + \frac{\rho^2}{2} Z_x Z_y + \frac{\rho^3}{6}(Z_x^2 - 1)(Z_y^2 - 1)\right) \phi_1(Z_x)\phi_1(Z_y)\right]. \tag{27}$$

Let $g(Z_x, Z_y)$ be the probability density function for (Z_x, Z_y). Then, we can write:

$$\begin{aligned} & \phi_1(Z_x)\phi_1(Z_y) g(Z_x, Z_y) \\ = \; & \frac{\mathrm{e}^{-\frac{1}{2}\left[Z_x^2 + Z_y^2 + \frac{1}{1-\lambda^2}\left[\left(\frac{Z_x - \mu_{zx}}{\sigma_{zx}}\right)^2 + \left(\frac{Z_y - \mu_{zy}}{\sigma_{zy}}\right)^2 - \left(2\lambda \frac{(Z_x - \mu_{zx})(Z_y - \mu_{zy})}{\sigma_{zx}\sigma_{zy}}\right)\right]\right]}}{4\pi^2\sqrt{1-\lambda^2}\sigma_{zx}\sigma_{zy}} \\ = \; & A_0 \mathrm{e}^{-\frac{a_{xx} Z_x^2 + a_x Z_x + a_{yy} Z_y^2 + a_y Z_y - 2a_{xy} Z_x Z_y}{2}}, \end{aligned}$$

where

$$\begin{cases} a_{xx} = 1 + \frac{1}{(1-\lambda^2)\sigma_{zx}^2}, & a_x = \frac{2}{(1-\lambda^2)\sigma_{zx}}\left[\frac{\lambda\mu_{zy}}{\sigma_{zy}} - \frac{\mu_{zx}}{\sigma_{zx}}\right] \\ a_{yy} = 1 + \frac{1}{(1-\lambda^2)\sigma_{zy}^2}, & a_y = \frac{2}{(1-\lambda^2)\sigma_{zy}}\left[\frac{\lambda\mu_{zx}}{\sigma_{zx}} - \frac{\mu_{zy}}{\sigma_{zy}}\right] \\ a_{xy} = \frac{\lambda}{(1-\lambda^2)\sigma_{zx}\sigma_{zy}}, & A_0 = \frac{\mathrm{e}^{-\frac{1}{2(1-\lambda^2)}\left[\frac{\mu_{zx}^2}{\sigma_{zx}^2} + \frac{\mu_{zy}^2}{\sigma_{zy}^2} - \frac{2\lambda\mu_{zx}\mu_{zy}}{\sigma_{zx}\sigma_{zy}}\right]}}{4\pi^2\sqrt{1-\lambda^2}\sigma_{zx}\sigma_{zy}}. \end{cases} \tag{28}$$

Thus, applying Lemma 25 to eq. (27), we have:

$$\begin{aligned} B \; & \simeq \; -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\rho + \frac{\rho^2}{2} Z_x Z_y + \frac{\rho^3}{6}(Z_x^2 - 1)(Z_y^2 - 1)\right] \phi_1(Z_x)\phi_1(Z_y) g(Z_x, Z_y) \mathrm{d}Z_x \mathrm{d}Z_y \\ & = \; -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\rho + \frac{\rho^2}{2} Z_x Z_y + \frac{\rho^3}{6}(Z_x^2 Z_y^2 - Z_x^2 - Z_y^2 + 1)\right] \phi_1(Z_x)\phi_1(Z_y) g(Z_x, Z_y) \mathrm{d}Z_x \mathrm{d}Z_y \\ & = \; -A_0 \tilde{A} \Bigg[\rho + (\tilde{\mu}_x \tilde{\mu}_y + \tilde{\sigma}_x \tilde{\sigma}_y \tilde{\rho}) \frac{\rho^2}{2} \\ & \qquad + \left((\tilde{\mu}_x^2 + \tilde{\sigma}_x^2 - 1)(\tilde{\mu}_y^2 + \tilde{\sigma}_y^2 - 1) + 2\tilde{\sigma}_x \tilde{\sigma}_y \tilde{\rho}(\tilde{\sigma}_x \tilde{\sigma}_y \tilde{\rho} + 2\tilde{\mu}_x \tilde{\mu}_y)\right) \frac{\rho^3}{6}\Bigg]. \end{aligned}$$

It is straightforward to verify $c_0 = A_0 \tilde{A}$ and $c_2 = \tilde{\mu}_x \tilde{\mu}_y + \tilde{\sigma}_x \tilde{\sigma}_y \tilde{\rho}$. Noting $\tilde{\mu}_x = (1 - \lambda) W_{2x}/W_0$, $\tilde{\mu}_y = (1-\lambda) W_{2y}/W_0$, $\tilde{\sigma}_x^2 - 1 = -(\sigma_{zy}^2 + 1)/W_0$, and $\tilde{\sigma}_y^2 - 1 = -(\sigma_{zx}^2 + 1)/W_0$,

[10] The lemmas used in this proof are proved in Section A.1.

we can also verify $c_3 = \big((\tilde{\mu}_x^2 + \tilde{\sigma}_x^2 - 1)(\tilde{\mu}_y^2 + \tilde{\sigma}_y^2 - 1) + 2\tilde{\sigma}_x\tilde{\sigma}_y\tilde{\rho}(\tilde{\sigma}_x\tilde{\sigma}_y\tilde{\rho} + 2\tilde{\mu}_x\tilde{\mu}_y)\big)$. □

Proof of Proposition 11 Using the first-order approximation of $\hat{y}_i$ around $\hat{\theta}_y$, we have:

$$\hat{y}_i = \bar{y}_i - f_1(-z_i^T\theta_y)z_i^T(\hat{\theta}_y - \theta_y) + h_i^N,$$

where h_i^N involves second- and higher-order terms of $\hat{\theta}_y - \theta_y$.

Therefore, by Assumption 8 and the law of large numbers, we have:

$$\hat{\mu}_{xy}^N = \frac{1}{n_1}\sum_{i\in\mathcal{S}_1} x_i\bar{y}_i - m_i^T(\hat{\theta}_y - \theta_y) + x_ih_i^N \xrightarrow{p} E[x_i\bar{y}_i] = \bar{\mu}_{xy}^N \quad \text{as} \quad t \to \infty,$$

because the second and third terms go to zero.

By the law of large numbers and the central limit theorem we have the following as $t \to \infty$:

$$\frac{1}{n_1}\sum_{i\in\mathcal{S}_1} m_i \xrightarrow{p} E[m_i]. \tag{29}$$

$$n_1^{-1/2}\sum_{i\in\mathcal{S}_1} x_i\bar{y}_i - E[x_i\bar{y}_i] \xrightarrow{d} \mathcal{N}(0, Var[x_i\bar{y}_i]) \tag{30}$$

Note that independence between $\hat{\theta}_y$ and $x_i\bar{y}_i$ for all $i \in \mathcal{S}_1$ follows from Assumption 8. Therefore, noting that $E[x_i\bar{y}_i] = \bar{\mu}_{xy}^N$ holds, we have:

$$\begin{aligned}
& \sqrt{n}(\hat{\mu}_{xy}^N - \bar{\mu}_{xy}^N) \\
= & \sqrt{n}\left[\frac{1}{n_1}\sum_{i\in\mathcal{S}_1} x_i\hat{y}_i - x_i\bar{y}_i + x_i\bar{y}_i - E[x_i\bar{y}_i]\right] \\
= & \sqrt{\frac{n}{n_1}}\left[n_1^{-1/2}\sum_{i\in\mathcal{S}_1} x_i\bar{y}_i - E[x_i\bar{y}_i]\right] - \sqrt{\frac{n}{n_2}}\left[\frac{1}{n_1}\sum_{i\in\mathcal{S}_1} m_i^T\right]\sqrt{n_2}(\hat{\theta}_y - \theta_y) \\
& + \sqrt{\frac{n}{n_1}}\left[n_1^{-1/2}\sum_{i\in\mathcal{S}_1} x_ih_i^N\right] \\
\xrightarrow{d} & \mathcal{N}\left(0, r_1Var[x_i\bar{y}_i] + r_2E^T[m_i]\Omega_yE[m_i]\right),
\end{aligned}$$

where the convergence in distribution follows from Assumption 8, eqs. (29), and (30). □

Proof of Proposition 15 By the law of large numbers, we have $n_1^{-1}\sum_{i\in\mathcal{S}_1} m_{1i} \xrightarrow{p} E[m_{1i}]$. By this and Assumption 8, we have the desired result. □

Proof of Proposition 17 Let $m_1 \equiv n_1^{-1}\sum_{i\in\mathcal{S}_1} m_{1i}$ and $m_2 \equiv n_1^{-1}\sum_{i\in\mathcal{S}_1} m_{2i}$. Then, by eq. (8) and Lemma 26, we have:

$$
\begin{aligned}
& \sqrt{n}(\hat{\mu}_{xy}^C - \mu_{xy}) \\
= \;& \sqrt{\frac{n}{n_1}}\sqrt{n_1}\left[\frac{1}{n_1}\sum_{i\in\mathcal{S}_1} q_i - E[q_i]\right] + \sqrt{\frac{n}{n_1}} m_1^T \sqrt{n_1}(\hat{\theta}_x - \theta_x) \\
& \sqrt{\frac{n}{n_2}} m_2^T \sqrt{n_2}(\hat{\theta}_y - \theta_y) + n_1^{-1}\sum_{i\in\mathcal{S}_1} h_{ic} \\
= \;& \sqrt{\frac{n}{n_1}}\sqrt{n_1}\left[\frac{1}{n_1}\sum_{i\in\mathcal{S}_1} \tilde{q}_i - E[\tilde{q}_i]\right] + \sqrt{\frac{n}{n_2}} m_2^T \sqrt{n_2}(\hat{\theta}_y - \theta_y) \\
& + \sqrt{\frac{n}{n_1}}\frac{1}{\sqrt{n_1}} E^T[m_{1i}]E^{-1}[m_{3xi}(\theta_x,\theta_x)]\sum_{i\in\mathcal{S}_1}\left[d_{xi}(\hat{\theta}_x) - d_{xi}(\theta_x)\right] \\
& + \sqrt{\frac{n}{n_1}}\left(m_1^T m_{3x}^{-1}(\hat{\theta}_x, \theta_x^+) - E^T[m_1]E^{-1}[m_{3xi}(\theta_x,\theta_x)]\right)\frac{1}{\sqrt{n_1}}\sum_{i\in\mathcal{S}_1} d_{xi}(\hat{\theta}_x) + n_1^{-1}\sum h_{iq} \\
\xrightarrow{d} \;& \mathcal{N}\left(0, r_1 Var[\tilde{q}_i] + r_2 E[m_{2i}^T]\Omega_y E[m_{2i}]\right),
\end{aligned}
$$

where $\tilde{q}_i \equiv q_i + E^T[m_{1i}]E^{-1}[m_{3xi}(\theta_x,\theta_x)]d_{xi}(\theta_x)$ and the convergence in distribution follows from the law of large numbers and the independence between $\tilde{q}_i$ and $\hat{\theta}_y$.

By Assumption 16, we have $\Omega_x = E^{-1}[m_{3xi}(\theta_x,\theta_x)]$. Therefore, we have the following relationship:

$$
\begin{aligned}
& Var[\tilde{q}_i] \\
= \;& Var[q_i] + E[m_{1i}^T]\Omega_x E[m_{1i}^T] + 2E[m_{1i}^T]\Omega_x E[q_i d_{xi}(\theta_x)] \\
= \;& E[\bar{x}_i(1-\bar{x}_i)\bar{y}_i^2] + Var[\overline{x_i y_i}] + E[m_{1i}^T]\Omega_x E[m_{1i}] - 2E[m_{1i}^T]\Omega_x E[f_{1x}(-z_i^T\theta_x)\bar{y}_i z_i],
\end{aligned}
$$

where we used the law of total variance to obtain the last equality. The proposition immediately follows from this. □

Proof of Proposition 19 Notice first that $\hat{\theta}_w \xrightarrow{p} \theta_w$ under the ML regularity conditions. Therefore, by this and the law of large numbers, $\hat{E}[m_{ai}] \xrightarrow{p} E[m_{ai}]$ for $a \in \{1,2,4\}$. The consistency of $\hat{V}^C$ follows from this, $\hat{\Omega}_w \xrightarrow{p} \Omega_w$ for $w\{x,y\}$ and $svar[\widehat{\overline{x_i y_i}}] \xrightarrow{p} Var[\overline{x_i y_i}]$. □

Proof of Proposition 21 We shall drop the subscript i for the simplicity of the proof. Let the gradient operator for z be $\nabla_z \equiv \left(\partial/\partial z^1, \partial/\partial z^2, \cdots, \partial/\partial z^L\right)^T$ and the Hessian operator be ∇_z^2. Then, for $w \in \{x, y\}$, we have:

$$\bar{w} = \Phi(Z_w), \quad \nabla_z \bar{w} = -\phi_1(Z_w)\theta_w, \quad \text{and} \quad \nabla_z^2 \bar{w} = -Z_w \phi_1(Z_w)\theta_w \theta_w^T.$$

Therefore, $\bar{w}$, $\nabla_z \bar{w}$ and $\nabla_z^2 \bar{w}$ evaluated at $z = 0$ are $1/2$, $-\phi_1(0)\theta_w$, and O_L, respectively, where O_L is a $L \times L$ matrix of zeros.

Letting $v_1(z) \equiv \bar{x}(1-\bar{x})\bar{y}^2$, we have:

$$\begin{aligned} \nabla_z v_1 &= (1-2\bar{x})\bar{y}^2 \nabla_z \bar{x} + 2\bar{x}(1-\bar{x})\bar{y}\nabla_z \bar{y} && (31) \\ \nabla_z^2 v_1 &= \bar{y}^2((1-2\bar{x})\nabla_z^2 \bar{x} - 2\nabla_z \bar{x} \nabla_z^T \bar{x}) + 2(1-2\bar{x})\bar{y}(\nabla_z \bar{x} \nabla_z^T \bar{y} + \nabla_z \bar{y} \nabla_z^T \bar{x}) \\ &\quad + 2\bar{x}(1-\bar{x})(\bar{y}\nabla_z^2 \bar{y} + \nabla_z \bar{y} \nabla_z^T \bar{y}) && (32) \end{aligned}$$

By taking the second-order approximation of v_1 around $z = 0$, we have:

$$\begin{aligned} E[v_1(z)] &\simeq E\left[v_1(0) + \nabla_z^T v_1(0) z + \frac{1}{2} z^T \nabla_z^2 v_1(0) z\right] && (33) \\ &= \frac{1}{16} - \frac{1}{4\sqrt{2\pi}} \theta_y^T \mu_z + \frac{1}{8\pi}\left(\theta_y^T(\Sigma_z + \mu_z \mu_z^T)\theta_y - \theta_x^T(\Sigma_z + \mu_z \mu_z^T)\theta_x\right) && (34) \end{aligned}$$

Let $v_2(z) \equiv \overline{xy} = \Phi_2(Z_x, Z_y, \rho)$. Then, we have:

$$\nabla_z v_2(z) = -\phi_1(Z_x)\Phi_1(\psi(\rho Z_x - Z_y))\theta_x - \Phi_1(Z_y)\phi_1(\psi(\rho Z_y - Z_x))\theta_y,$$

and thus $\nabla_z v_2(0) = -\phi_1(0)\Phi_1(0)(\theta_x + \theta_y)$. Therefore,

$$Var[v_2(z)] \simeq \nabla_z^T v_2(0) Var[z] \nabla_z v_2(0) = \frac{1}{8\pi}(\theta_x + \theta_y)^T \Sigma_z (\theta_x + \theta_y).$$

With some slight abuse of notation, we shall write $m_1(z)$ to emphasize that it is a function of z. Then, we can show:

$$\begin{aligned} E[m_1(z)] &\simeq E\left[m_1(0) + \sum_{l=1}^{L} \frac{\partial m_1(0)}{\partial z^l} z^l + \frac{1}{2}\sum_{l=1}^{L}\sum_{l'=1}^{L} \frac{\partial^2 m_1(0)}{\partial z^l \partial z^{l'}} z^l z^{l'}\right] \\ &= E\left[\frac{zz^T}{2\pi}\right]\left[(\psi - 1)\theta_y - \psi\rho\theta_x\right]. && (35) \\ &= \frac{1}{2\pi}\left[\Sigma_z + \mu_z \mu_z^T\right]\left[(\psi-1)\theta_y - \psi\rho\theta_x\right] && (36) \end{aligned}$$

Similarly, we have:

$$E[m_2(z)] \simeq \frac{1}{2\sqrt{2\pi}}\mu_z + \frac{\psi}{2\pi}\left[\Sigma_z + \mu\mu^T\right](\rho\theta_y - \theta_x)$$

When θ_w for $w \in \{x, y\}$ is a maximum likelihood estimator, its asymptotic variance $\Omega_w(z)$ can be approximated in the following manner:

$$\Omega_w(z) = E^{-1}\left[\frac{\phi^2(Z_w)}{\Phi(Z_w)(1-\Phi(Z_w))}zz^T\right] \simeq E^{-1}[4\phi^2(0)zz^T] = \frac{\pi}{2}[\Sigma_z + \mu\mu^T]^{-1}.$$

Finally, we have:

$$E[m_4(z)] \simeq E\left[\frac{z}{2\sqrt{2\pi}} - \frac{1}{2\pi}zz^T\theta_y\right] = \frac{\mu_z}{2\sqrt{2\pi}} - \frac{1}{2\pi}\left[\Sigma_z + \mu\mu_z^T\right]\theta_y.$$

Therefore, using the approximation of m_1, m_2, Ω_x, and Ω_y, we have eqs. (16) and (17). □

A.3 Additional Tables and Figures

Table 5: Biprobit regression results for Albania 2002.

Dependent variable	Consumption Poverty		Subjective Poverty	
Household Size	0.345**	(0.027)	0.079**	(0.020)
Ratio of children (-14)	0.716**	(0.230)	-0.126	(0.185)
Ratio of elderly (65+)	0.363	(0.253)	0.132	(0.165)
HH has a SP	0.579	(0.391)	0.196	(0.318)
Age of HH	-0.005	(0.006)	-0.006	(0.005)
HH away for 12+ mths	1.148	(0.727)	6.167**	(0.210)
HH employed	-0.362**	(0.092)	-0.397**	(0.074)
HH completed 8+ yrs schooling	-0.008	(0.123)	-0.019	(0.100)
HH completed sec school	-0.184*	(0.090)	-0.185*	(0.076)
HH completed college educ	-0.293*	(0.148)	-0.125	(0.096)
Age of spouse	-0.006	(0.007)	-0.004	(0.005)
SP employed	-0.096	(0.087)	-0.120	(0.070)
SP completed 8+ yrs schooling	-0.076	(0.136)	0.084	(0.114)
SP completed sec school	-0.246*	(0.104)	-0.119	(0.080)
SP completed college educ	0.178	(0.161)	-0.105	(0.112)
Single house	-0.054	(0.090)	-0.071	(0.069)
Elec. meter within premise	0.203**	(0.079)	-0.021	(0.065)
Flush toilet	-0.337**	(0.095)	-0.454**	(0.082)
Drinking water from pipe	0.115	(0.084)	0.087	(0.066)
Have a telephone	-0.620**	(0.115)	-0.351**	(0.079)
Have a mobile phone	-0.660**	(0.090)	-0.494**	(0.064)
Have an air conditioner	0.337	(0.348)	-1.043**	(0.247)
Have a bicycle	-0.085	(0.105)	-0.295**	(0.078)
Have a car	-0.913**	(0.258)	-0.649**	(0.114)
have a refrigerator	-0.385**	(0.094)	-0.384**	(0.094)
Have a tape or CD player	-0.058	(0.076)	-0.085	(0.060)
Have a truck	-0.617	(0.386)	-0.299	(0.202)
Have a washing machine	-0.211*	(0.096)	-0.205**	(0.073)
Rural area	-0.463**	(0.113)	-0.069	(0.085)
Capital	0.270*	(0.112)	0.478**	(0.081)
Constant	-1.235**	(0.424)	1.526**	(0.328)
rho	0.212		(0.047)	
# obs	3599			
chi-sq	4344.45			

Note: Standard errors in parenthesis.
* : Significant at a 5 percent level.
**: Significant at a 1 percent level.

Table 6: Biprobit regression results for Albania 2005.

Dependent variable	Consumption Poverty		Subjective Poverty	
Household Size	0.270**	(0.031)	0.073**	(0.021)
Ratio of children (-14)	0.291	(0.224)	-0.056	(0.177)
Ratio of elderly (65+)	-0.157	(0.257)	-0.122	(0.156)
HH has a SP	0.779	(0.452)	-0.266	(0.319)
Age of HH	-0.003	(0.007)	-0.003	(0.005)
HH away for 12+ mths	-0.046	(0.312)	-0.232	(0.276)
HH employed	-0.126	(0.102)	-0.148	(0.078)
HH completed 8+ yrs schooling	-0.126	(0.137)	-0.197	(0.104)
HH completed sec school	-0.185	(0.096)	-0.073	(0.070)
HH completed college educ	0.109	(0.153)	-0.187	(0.101)
Age of spouse	-0.010	(0.008)	0.002	(0.005)
SP employed	-0.245**	(0.088)	-0.120	(0.066)
SP completed 8+ yrs schooling	0.126	(0.150)	0.230*	(0.113)
SP completed sec school	-0.229*	(0.108)	-0.025	(0.079)
SP completed college educ	0.294	(0.171)	-0.284*	(0.120)
Single house	-0.260*	(0.101)	-0.089	(0.069)
Elec. meter within premise	0.107	(0.116)	-0.213*	(0.090)
Flush toilet	-0.139	(0.096)	-0.436**	(0.083)
Drinking water from pipe	0.211	(0.166)	0.260	(0.105)
Have a telephone	-0.357**	(0.112)	-0.322**	(0.075)
Have a mobile phone	-0.367**	(0.091)	-0.509**	(0.079)
Have an air conditioner	-5.155**	(0.137)	-0.744**	(0.175)
Have a bicycle	-0.123	(0.119)	-0.297**	(0.083)
Have a car	-1.432**	(0.327)	-0.621**	(0.095)
have a refrigerator	-0.492**	(0.107)	-0.521**	(0.108)
Have a tape or CD player	-0.014	(0.081)	-0.067	(0.059)
Have a truck	-0.818**	(0.307)	-0.943**	(0.210)
Have a washing machine	-0.434**	(0.096)	-0.350**	(0.075)
Rural area	-0.137	(0.106)	-0.367**	(0.082)
Capital	-0.391*	(0.147)	-0.080	(0.082)
Constant	-0.844	(0.473)	2.341**	(0.345)
rho	0.362		(0.048)	
# obs	3638			
chi-sq	4585.45			

Note: Standard errors in parenthesis.
* : Significant at a 5 percent level.
**: Significant at a 1 percent level.

Table 7: Biprobit regression results for Tajikistan 2003.

Dependent variable	Consumption Poverty		Subjective Poverty	
Household Size	0.091**	(0.016)	-0.009	(0.010)
Ratio of children (-14)	0.516**	(0.148)	-0.048	(0.132)
Ratio of elderly (65+)	-0.194	(0.229)	0.349	(0.223)
HH has a SP	0.052	(0.273)	-0.085	(0.238)
Age of HH	0.000	(0.004)	-0.004	(0.003)
HH away for 12+ mths	-0.104	(0.176)	-0.471**	(0.151)
HH employed	-0.046	(0.078)	-0.097	(0.066)
HH completed 8+ yrs schooling	0.069	(0.134)	-0.163	(0.108)
HH completed sec school	-0.161	(0.109)	-0.020	(0.083)
HH completed college educ	-0.120	(0.080)	-0.203**	(0.069)
Age of spouse	0.000	(0.005)	0.001	(0.004)
SP employed	-0.088	(0.068)	-0.033	(0.056)
SP completed 8+ yrs schooling	0.140	(0.137)	0.015	(0.115)
SP completed sec school	-0.054	(0.094)	-0.104	(0.077)
SP completed college educ	-0.187	(0.124)	-0.284*	(0.128)
Single house	0.121	(0.081)	-0.113	(0.076)
Elec. meter within premise	-0.141*	(0.071)	0.000	(0.064)
Flush toilet	0.257**	(0.076)	-0.143*	(0.065)
Drinking water from pipe	-0.146*	(0.059)	0.011	(0.052)
Have a telephone	-0.187*	(0.093)	-0.160*	(0.081)
Have a mobile phone	-0.446	(0.234)	-0.200	(0.258)
Have an air conditioner	-0.160	(0.132)	-0.231	(0.118)
Have a bicycle	0.251	(0.305)	0.036	(0.313)
Have a car	-0.470**	(0.088)	-0.528**	(0.083)
have a refrigerator	-0.179**	(0.069)	-0.284**	(0.061)
Have a tape or CD player	-0.432**	(0.063)	-0.298**	(0.057)
Have a truck	-0.549**	(0.210)	-0.082	(0.182)
Have a washing machine	-0.315**	(0.086)	-0.302**	(0.083)
Rural area	-0.403**	(0.088)	-0.075	(0.073)
Capital	-0.311**	(0.093)	0.000	(0.084)
Constant	0.887**	(0.292)	1.181**	(0.256)
rho	0.198		(0.037)	
# obs	4160			
chi-sq	748.48			

Note: Standard errors in parenthesis.
* : Significant at a 5 percent level.
**: Significant at a 1 percent level.

Table 8: Biprobit regression results for Tajikistan 2007.

Dependent variable	Consumption Poverty		Subjective Poverty	
Household Size	0.104**	(0.012)	-0.013	(0.011)
Ratio of children (-14)	0.693**	(0.126)	0.128	(0.126)
Ratio of elderly (65+)	0.729**	(0.221)	0.252	(0.221)
HH has a SP	-0.175	(0.232)	-0.159	(0.233)
Age of HH	-0.007*	(0.003)	-0.001	(0.003)
HH away for 12+ mths	-0.408	(0.543)	-0.109	(0.644)
HH employed	-0.176**	(0.058)	-0.146*	(0.058)
HH completed 8+ yrs schooling	-0.058	(0.108)	0.168	(0.102)
HH completed sec school	-0.057	(0.080)	0.081	(0.079)
HH completed college educ	-0.277**	(0.069)	-0.225**	(0.070)
Age of spouse	0.002	(0.004)	0.002	(0.004)
SP employed	0.010	(0.057)	-0.019	(0.057)
SP completed 8+ yrs schooling	0.052	(0.109)	0.066	(0.103)
SP completed sec school	0.044	(0.069)	-0.118	(0.070)
SP completed college educ	-0.129	(0.123)	-0.157	(0.135)
Single house	0.268**	(0.084)	-0.038	(0.089)
Elec. meter within premise	-0.096	(0.066)	0.045	(0.064)
Flush toilet	0.028	(0.098)	-0.218*	(0.105)
Drinking water from pipe	0.232**	(0.056)	-0.051	(0.056)
Have a telephone	-0.150*	(0.072)	0.017	(0.075)
Have a mobile phone	-0.252**	(0.051)	-0.280**	(0.052)
Have an air conditioner	-0.689**	(0.116)	-0.495**	(0.123)
Have a bicycle	-0.012	(0.065)	-0.074	(0.067)
Have a car	-0.405**	(0.069)	-0.315**	(0.074)
have a refrigerator	-0.029	(0.060)	-0.414**	(0.061)
Have a tape or CD player	-0.183**	(0.059)	-0.161**	(0.060)
Have a truck	-0.767**	(0.154)	-0.247	(0.164)
Have a washing machine	-0.248**	(0.088)	-0.185*	(0.092)
Rural area	-0.197**	(0.076)	-0.119	(0.077)
Capital	0.100	(0.084)	0.209*	(0.088)
Constant	-0.187	(0.253)	0.171	(0.262)
rho	0.185		(0.030)	
# obs	4860			
chi-sq	769.16			

Note: Standard errors in parenthesis.
* : Significant at a 5 percent level.
**: Significant at a 1 percent level.

Table 9: Biprobit regression results for Malawi 2004-2005.

Dependent variable	Consumption Poverty		Subjective Poverty	
Household Size	0.147**	(0.013)	-0.022*	(0.010)
Ratio of children (-14)	1.413**	(0.093)	0.166	(0.098)
Ratio of elderly (65+)	0.122	(0.111)	0.241	(0.148)
HH has a SP	-0.010	(0.099)	0.212	(0.111)
Age of HH	0.006**	(0.002)	0.002	(0.002)
HH away for 12+ mths	0.330	(0.219)	-0.182	(0.244)
HH employed	-0.216**	(0.082)	0.075	(0.090)
HH completed 8+ yrs schooling	-0.348**	(0.054)	-0.299**	(0.055)
HH completed sec school	-0.155*	(0.074)	-0.241**	(0.069)
HH completed college educ	-1.820**	(0.559)	-0.481	(0.291)
Age of spouse	0.000	(0.002)	-0.005*	(0.002)
SP employed	0.157*	(0.064)	-0.080	(0.063)
SP completed 8+ yrs schooling	-0.125	(0.082)	-0.256**	(0.076)
SP completed sec school	-0.251*	(0.124)	0.037	(0.102)
SP completed college educ	2.555**	(0.548)	-0.043	(0.635)
Single house	-0.071	(0.040)	-0.066	(0.045)
Elec. meter within premise	-0.997**	(0.137)	-0.347**	(0.096)
Flush toilet	-0.109	(0.165)	0.043	(0.126)
Drinking water from pipe	0.032	(0.045)	0.035	(0.052)
Have a telephone	0.405	(0.400)	-0.170	(0.217)
Have a mobile phone	-1.082**	(0.275)	-0.474**	(0.116)
Have an air conditioner	-0.303	(0.278)	-0.253	(0.320)
Have a bicycle	-0.418**	(0.035)	-0.369**	(0.038)
Have a car	-1.927**	(0.411)	-0.465*	(0.187)
have a refrigerator	-0.792*	(0.362)	0.049	(0.164)
Have a tape or CD player	-0.592**	(0.050)	-0.403**	(0.047)
Have a truck	-6.193**	(0.229)	-0.637	(0.438)
Have a washing machine	0.324	(0.361)	0.063	(0.392)
Rural area	0.182*	(0.078)	0.145	(0.076)
Capital	-0.126	(0.116)	-0.328**	(0.096)
Constant	-1.226**	(0.147)	1.293**	(0.153)
rho	0.285		(0.024)	
# obs	11257			
chi-sq	8257.52			

Note: Standard errors in parenthesis.
* : Significant at a 5 percent level.
**: Significant at a 1 percent level.

Table 10: The robustness of the estimates of ρ due to model selection.

Data	Poverty Lines	25 Vars		20 Vars		15 Vars		10 Vars		5 Vars	
ALB02	Normal	0.246	[0.025]	0.281	[0.032]	0.321	[0.037]	0.367	[0.039]	0.421	[0.039]
	Extreme	0.345	[0.016]	0.362	[0.024]	0.385	[0.030]	0.415	[0.036]	0.456	[0.036]
ALB05	Normal	0.380	[0.016]	0.402	[0.022]	0.426	[0.027]	0.457	[0.031]	0.498	[0.033]
	Extreme	0.361	[0.023]	0.385	[0.033]	0.414	[0.039]	0.452	[0.044]	0.499	[0.041]
TJK03	Normal	0.207	[0.014]	0.219	[0.019]	0.233	[0.024]	0.248	[0.026]	0.271	[0.025]
	Extreme	0.218	[0.012]	0.225	[0.018]	0.236	[0.022]	0.249	[0.023]	0.268	[0.022]
TJK07	Normal	0.191	[0.010]	0.199	[0.014]	0.208	[0.018]	0.221	[0.019]	0.239	[0.019]
	Extreme	0.283	[0.010]	0.287	[0.014]	0.291	[0.017]	0.296	[0.017]	0.305	[0.016]
MLW04	Normal	0.292	[0.017]	0.302	[0.025]	0.314	[0.032]	0.330	[0.038]	0.362	[0.040]
	Extreme	0.280	[0.017]	0.286	[0.024]	0.293	[0.029]	0.302	[0.033]	0.317	[0.031]

Note: Figures in squared brackets are the model errors.

Table 11: Variance decomposition.

Data		Idiosync.		z-smpl		x-model		y-model		Covariance		5:2 ratio	
		Est.	BSE.	Est.	BSE.	Est.	BSE.	Est.	BSE.	Est.	BSE.	Est.	BSE.
ALB02	NC & NS	12.27	[1.07]	38.24	[1.94]	0.17	[0.01]	51.07	[2.08]	-1.75	[0.18]	101.62	[0.46]
	NC & SP	31.89	[1.55]	11.75	[1.14]	0.19	[0.01]	54.82	[1.79]	1.35	[0.26]	98.50	[0.82]
	CP & NS	49.42	[3.83]	10.96	[1.51]	0.70	[0.09]	46.05	[3.08]	-7.13	[1.25]	105.57	[2.72]
	CP & SP	42.54	[2.33]	38.93	[2.18]	0.25	[0.02]	16.46	[2.14]	1.82	[0.42]	98.48	[1.41]
ALB05	NC & NS	9.48	[0.99]	30.59	[1.99]	0.62	[0.05]	63.33	[2.26]	-4.01	[0.37]	104.89	[0.83]
	NC & SP	23.60	[1.36]	9.80	[1.02]	0.60	[0.04]	61.93	[1.68]	4.06	[0.37]	97.46	[1.18]
	CP & NS	79.49	[6.61]	8.55	[1.63]	5.25	[0.95]	40.76	[3.54]	-34.04	[5.99]	137.31	[10.03]
	CP & SP	43.60	[2.24]	38.91	[2.69]	1.12	[0.12]	8.83	[1.32]	7.54	[0.96]	99.21	[2.64]
TJK03	NC & NS	55.59	[1.43]	21.29	[1.75]	0.99	[0.08]	11.26	[0.94]	10.86	[0.75]	90.78	[1.90]
	NC & SP	93.36	[2.83]	6.06	[0.76]	2.11	[0.19]	23.79	[1.42]	-25.32	[1.67]	122.30	[2.30]
	CP & NS	14.41	[1.06]	6.65	[1.04]	0.26	[0.02]	75.89	[1.48]	2.81	[0.17]	96.94	[0.28]
	CP & SP	11.69	[0.95]	13.68	[1.18]	0.26	[0.02]	77.53	[1.56]	-3.16	[0.22]	103.21	[0.26]
TJK07	NC & NS	55.40	[1.53]	21.17	[1.43]	0.13	[0.01]	20.97	[1.21]	2.32	[0.27]	97.15	[0.59]
	NC & SP	43.88	[1.84]	5.91	[0.73]	0.33	[0.04]	54.47	[1.52]	-4.59	[0.55]	103.57	[0.87]
	CP & NS	46.89	[1.84]	7.79	[1.04]	0.11	[0.01]	43.25	[1.81]	1.96	[0.18]	98.18	[0.43]
	CP & SP	22.20	[1.35]	12.77	[1.08]	0.17	[0.01]	67.17	[1.79]	-2.31	[0.17]	102.29	[0.28]
MLW04	NC & NS	7.45	[0.74]	33.94	[4.46]	0.38	[0.05]	58.58	[3.98]	-0.34	[0.18]	100.38	[0.32]
	NC & SP	62.64	[0.99]	15.23	[0.81]	0.15	[0.02]	23.21	[0.93]	-1.24	[0.24]	102.14	[1.33]
	CP & NS	16.17	[1.12]	6.35	[0.94]	0.82	[0.07]	77.43	[1.31]	-0.77	[0.43]	99.62	[0.73]
	CP & SP	59.24	[1.23]	28.38	[1.19]	0.14	[0.01]	13.40	[1.08]	-1.17	[0.22]	102.73	[1.28]

Note: All the figures are in percentage. ASE in parenthesis and BSE in square brackets.

Table 12: Cross-tabulation tables with the extreme poverty lines.

Data			One smpl NS	One smpl SP	BC (best) NS	BC (best) SP	BC (guest.) NS	BC (guest.) SP	Naïve NS	Naïve SP
ALB02	NC	Est.	72.38	22.88	68.68	25.82	68.67	25.86	71.50	23.73
		ASE	(0.75)	(0.70)	(1.03)	(0.97)	(1.03)	(0.96)	(0.94)	(0.87)
		BSE	[0.90]	[0.82]	[1.05]	[0.98]	[1.05]	[0.99]	[0.93]	[0.85]
	CP	Est.	1.46	3.28	1.72	3.78	1.88	3.59	2.25	2.51
		ASE	(0.20)	(0.30)	(0.31)	(0.36)	(0.31)	(0.36)	(0.28)	(0.30)
		BSE	[0.22]	[0.33]	[0.26]	[0.40]	[0.26]	[0.41]	[0.27]	[0.30]
ALB05	NC	Est.	73.32	23.21	68.30	27.40	68.14	27.55	72.55	23.96
		ASE	(0.73)	(0.70)	(1.06)	(1.01)	(1.07)	(1.01)	(0.92)	(0.86)
		BSE	[0.89]	[0.87]	[1.13]	[1.09]	[1.16]	[1.08]	[0.91]	[0.86]
	CP	Est.	0.83	2.64	1.18	3.11	1.32	2.99	1.54	1.95
		ASE	(0.15)	(0.27)	(0.28)	(0.38)	(0.28)	(0.38)	(0.23)	(0.30)
		BSE	[0.14]	[0.25]	[0.21]	[0.41]	[0.25]	[0.42]	[0.22]	[0.29]
TJK03	NC	Est.	47.84	10.15	47.49	10.30	48.18	9.61	45.34	12.45
		ASE	(0.77)	(0.47)	(0.82)	(0.44)	(0.82)	(0.43)	(0.84)	(0.47)
		BSE	[0.92]	[0.55]	[0.81]	[0.44]	[0.81]	[0.44]	[0.83]	[0.48]
	CP	Est.	28.27	13.73	28.44	13.78	27.77	14.44	30.59	11.63
		ASE	(0.70)	(0.53)	(0.81)	(0.52)	(0.81)	(0.53)	(0.79)	(0.49)
		BSE	[0.78]	[0.61]	[0.84]	[0.52]	[0.87]	[0.53]	[0.81]	[0.49]
TJK07	NC	Est.	77.88	4.72	77.72	4.80	77.66	4.80	76.69	5.85
		ASE	(0.60)	(0.30)	(0.74)	(0.32)	(0.74)	(0.32)	(0.76)	(0.36)
		BSE	[0.64]	[0.40]	[0.75]	[0.33]	[0.72]	[0.32]	[0.78]	[0.38]
	CP	Est.	14.69	2.70	14.82	2.66	14.87	2.66	15.86	1.60
		ASE	(0.51)	(0.23)	(0.68)	(0.18)	(0.68)	(0.18)	(0.67)	(0.14)
		BSE	[0.54]	[0.25]	[0.71]	[0.19]	[0.67]	[0.18]	[0.71]	[0.14]
MLW04	NC	Est.	48.10	29.50	45.94	30.70	45.99	30.66	45.13	32.36
		ASE	(0.47)	(0.43)	(0.49)	(0.46)	(0.49)	(0.46)	(0.50)	(0.45)
		BSE	[0.51]	[0.50]	[0.66]	[0.51]	[0.68]	[0.54]	[0.51]	[0.46]
	CP	Est.	8.41	14.00	8.83	14.53	8.81	14.54	11.06	11.46
		ASE	(0.26)	(0.33)	(0.31)	(0.32)	(0.31)	(0.31)	(0.31)	(0.30)
		BSE	[0.26]	[0.36]	[0.30]	[0.34]	[0.31]	[0.37]	[0.30]	[0.28]

Note: All the figures are in percentage. ASE in parenthesis and BSE in square brackets.

Table 13: Conditional distribution of poverty for normal poverty lines.

		One sample				Two-sample (best)				Two-sample (guest)				Naïve			
		$CP \mid SP$	$CP \mid NS$	$SP \mid CP$	$SP \mid NC$	$CP \mid SP$	$CP \mid NS$	$SP \mid CP$	$SP \mid NC$	$CP \mid SP$	$CP \mid NS$	$SP \mid CP$	$SP \mid NC$	$CP \mid SP$	$CP \mid NS$	$SP \mid CP$	$SP \mid NC$
ALB02	Est.	12.55	1.97	69.26	24.02	12.76	2.44	68.76	27.32	12.19	2.66	65.72	27.36	9.57	3.05	52.72	24.92
	BSE	[1.19]	[0.30]	[4.99]	[0.85]	[1.27]	[0.36]	[2.62]	[1.02]	[1.29]	[0.37]	[2.67]	[1.03]	[1.07]	[0.37]	[2.73]	[0.88]
ALB05	Est.	10.23	1.12	76.06	24.04	10.20	1.70	72.57	28.63	9.78	1.90	69.39	28.79	7.54	2.07	55.95	24.83
	BSE	[0.92]	[0.20]	[6.14]	[0.89]	[1.27]	[0.30]	[3.17]	[1.13]	[1.27]	[0.35]	[3.17]	[1.12]	[1.06]	[0.30]	[3.77]	[0.88]
TJK03	Est.	57.49	37.14	32.70	17.51	57.23	37.45	32.64	17.82	60.05	36.56	34.22	16.62	48.29	40.29	27.54	21.54
	BSE	[1.72]	[0.99]	[1.41]	[0.87]	[1.16]	[0.97]	[1.12]	[0.68]	[1.11]	[1.00]	[1.20]	[0.69]	[1.19]	[0.96]	[1.02]	[0.74]
TJK07	Est.	36.41	15.87	15.54	5.71	35.66	16.02	15.21	5.81	35.66	16.07	15.19	5.82	21.48	17.14	9.17	7.09
	BSE	[3.08]	[0.58]	[1.78]	[0.46]	[1.53]	[0.75]	[1.06]	[0.39]	[1.37]	[0.71]	[1.01]	[0.39]	[1.23]	[0.75]	[0.75]	[0.45]
MLW04	Est.	32.18	14.88	62.47	38.02	32.12	16.12	62.19	40.06	32.17	16.08	62.27	40.00	26.15	19.68	50.88	41.76
	BSE	[0.74]	[0.46]	[1.34]	[0.56]	[0.59]	[0.55]	[0.81]	[0.67]	[0.65]	[0.56]	[0.82]	[0.70]	[0.54]	[0.49]	[0.79]	[0.55]

Note: All figures in percentage. Bootstrap standard errors in square brackets.

Table 14: Two-way imputation.

Data			BC (best)		BC (best), rev		Naïve		Naïve, rev	
			NS	SP	NS	SP	NS	SP	NS	SP
ALB02	NC	Est.	43.67	30.89	44.01	30.69	44.19	30.47	42.52	32.17
		ASE	(0.95)	(0.90)	(0.94)	(0.90)	(0.95)	(0.90)	(0.96)	(0.90)
		BSE	[0.98]	[0.90]	[1.00]	[0.91]	[1.00]	[0.91]	[1.00]	[0.90]
	CP	Est.	6.39	19.06	6.19	19.11	5.82	19.52	7.68	17.64
		ASE	(0.48)	(0.78)	(0.50)	(0.80)	(0.47)	(0.78)	(0.52)	(0.79)
		BSE	[0.45]	[0.80]	[0.48]	[0.83]	[0.46]	[0.79]	[0.52]	[0.82]
ALB05	NC	Est.	45.86	34.91	45.80	35.08	45.28	35.45	43.56	37.33
		ASE	(0.99)	(0.97)	(0.96)	(0.99)	(0.99)	(0.96)	(0.99)	(0.97)
		BSE	[1.22]	[1.06]	[1.23]	[1.09]	[1.04]	[0.95]	[1.26]	[1.06]
	CP	Est.	3.61	15.62	3.52	15.59	4.13	15.14	5.77	13.35
		ASE	(0.39)	(0.72)	(0.38)	(0.75)	(0.40)	(0.72)	(0.43)	(0.71)
		BSE	[0.33]	[0.79]	[0.37]	[0.83]	[0.36]	[0.76]	[0.45]	[0.79]
TJK03	NC	Est.	10.91	6.05	10.99	5.98	11.59	5.40	9.48	7.50
		ASE	(0.42)	(0.34)	(0.46)	(0.31)	(0.42)	(0.33)	(0.43)	(0.34)
		BSE	[0.47]	[0.31]	[0.45]	[0.32]	[0.45]	[0.30]	[0.42]	[0.37]
	CP	Est.	35.69	47.35	35.70	47.33	35.08	47.93	37.21	45.82
		ASE	(0.86)	(0.88)	(0.87)	(0.85)	(0.86)	(0.88)	(0.85)	(0.87)
		BSE	[0.86]	[0.84]	[0.91]	[0.87]	[0.94]	[0.91]	[0.89]	[0.88]
TJK07	NC	Est.	34.65	12.26	34.79	12.29	35.82	11.08	32.61	14.47
		ASE	(0.71)	(0.45)	(0.73)	(0.46)	(0.70)	(0.45)	(0.73)	(0.47)
		BSE	[0.76]	[0.45]	[0.73]	[0.44]	[0.74]	[0.43]	[0.73]	[0.47]
	CP	Est.	30.84	22.25	30.72	22.20	29.71	23.39	32.91	20.01
		ASE	(0.78)	(0.64)	(0.78)	(0.62)	(0.78)	(0.64)	(0.77)	(0.62)
		BSE	[0.81]	[0.64]	[0.76]	[0.65]	[0.79]	[0.68]	[0.74]	[0.64]
MLW04	NC	Est.	11.93	35.04	12.01	35.06	11.94	35.06	10.23	36.85
		ASE	(0.31)	(0.49)	(0.29)	(0.48)	(0.31)	(0.49)	(0.29)	(0.49)
		BSE	[0.66]	[0.52]	[0.68]	[0.49]	[0.65]	[0.48]	[0.71]	[0.49]
	CP	Est.	4.69	48.34	4.65	48.29	4.73	48.26	6.46	46.46
		ASE	(0.21)	(0.51)	(0.24)	(0.52)	(0.21)	(0.51)	(0.24)	(0.52)
		BSE	[0.21]	[0.67]	[0.21]	[0.70]	[0.20]	[0.69]	[0.25]	[0.70]

Note: All the figures are in percentage. ASE in parenthesis and BSE in square brackets.

References

Christiaensen, L., Lanjouw, P., Luoto, J. and Stifel, D. (2011). Small area estimation-based prediction methods to track poverty. Policy Research Working Paper 5683. The World Bank.

Deaton, A. (2003). Adjusted indian poverty estimates for 1999-2000. *Economic and Political Weekly*, **January 25, 2003,** 322–326.

Deaton, A. and Dreze, J. (2002). Poverty and inequality in India: A re-examination. *Economic and Political Weekly*, **Semtember 7,2002,** 3729–3748.

Demombynes, G., Elbers, C., Lanjouw, J.O. and Lanjouw, P. (2007). How good a map? putting small area estimation to the test. Policy Research Working Paper 4155. The World Bank.

Elbers, C., Lanjouw, J.O. and Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, **71,** number 1, 355–364.

Elbers, C., Lanjouw, P. and Leite, P.G. (2008). Brazil within Brazil: Testing the poverty map methodology in Minas Gerais. Policy Research Working Paper 4513. The World Bank.

Filmer, D. and Scott, K. (2008). Assessing asset indices. Policy Research Working Paper 4605. The World Bank.

Fujii, T. (2011). Micro-level estimation of child undernutrition indicators in Cambodia. *World Bank Economic Review* number forthcoming.

Ivaschenko, O. and Lanjouw, P. (2010). A new approach to producing geographic profiles of HIV prevalence: An application to Malawi. *World Medical and Health Policy*, **2,** number 1, 235–266.

Kijima, Y. and Lanjouw, P. (2003). Poverty in India during the 1990s: A regional perspective. Policy Research Working Paper 3141. The World Bank.

Lanjouw, P., Luoto, J. and McKenzie, D. (2011). Using repeated cross-sections to explore movements in and out of poverty. Policy Research Working Paper 5550. The World Bank.

Lindelow, M. (2006). Sometimes more equal than others: how health inequalities depend on the choice of welfare indicator. *Health economics*, **15,** 263–279.

Moffitt, R. (1993). Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics*, **59,** 99–123.

Pelzer, B., Eisinga, R. and Franses, P.H. (2001). Estimating transition probabilities from a time series of independent cross sections. *Statistica Neerlandica*, **55,** number 2, 249–262.

Ridder, G. and Moffitt, R. (2007). The econometrics of data combination. *Handbook of econometrics*, **6B,** number 75, 5469–5547.

Sahn, D.E. and Stifel, D. (2003). Exploring alternative measures of welfare in the absence of expenditure data. *Review of Income and Wealth*, **49,** number 4, 463–489.

Schellenberg, J.A., Victora, C.G., Mushi, A., de Savigny, D., Schellenberg, D., Mshinda, H. and Bryce, J. (2003). Inequities among the very poor: health care for children in rural southern Tanzania. *The Lancet*, **361,** 561–566.

Shao, J. and Sitter, R.R. (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association*, **91,** number 435, 1278–1288.

Stifel, D. and Christiaensen, L. (2007). Tracking poverty over time in the absence of comparable consumption data. *The World Bank Economic Review*, **21,** number 2, 317–341.

Tarozzi, A. (2007). Calculating comparable statistics from incomparable surveys, with an application to poverty in India. *Journal of Business and Economic Statistics*, **25,** number 3, 314–336.

Tarozzi, A. (2011). Can census data alone signal heterogeneity in the estimation of poverty maps? *Journal of Development Economics*, **95,** 170–185.

Tarozzi, A. and Deaton, A. (2009). Using census and survey data to estimate poverty and inequality for small areas. *The Review of Economics and Statistics*, **91,** number 4, 773–792.

Viet Cuong, N., Ngoc Truong, T. and van der Weide, R. (2010). Poverty and inequality maps for rural Vietnam. Policy Research Working Paper 5443. The World Bank.