

Can Census Data Alone Signal Heterogeneity in the Estimation of Poverty Maps?

Alessandro Tarozzi
Duke University

May 2010*

Abstract

Methodologies now commonly used for the construction of poverty maps assume a substantial degree of homogeneity within geographical areas in the relationship between income and its predictors. However, local labor and rental markets and other local environmental differences are likely to generate heterogeneity in such relationships, at least to some extent. The purpose of this paper is to argue that useful if only indirect and suggestive evidence on the extent of area heterogeneity is readily available in virtually any census. Such indirect evidence is provided by non-monetary indicators—such as literacy, asset ownership or access to sanitation—which are routinely included in censuses. These indicators can be used to perform validation exercises to gauge the extent of heterogeneity in their distribution conditional on predictors analogous to those commonly used in poverty mapping. We argue that the same factors which are likely to generate area heterogeneity in poverty mapping are also likely to generate heterogeneity in such kind of validation exercises. We construct a very simple model to illustrate this point formally. Finally, we evaluate empirically the argument using data from Mexico. In our empirical illustrations, the performance of imputation methodologies to construct maps of indicators typically feasible with census data alone is indeed informative about how effectively such methodologies can produce correct inference in poverty mapping. **JEL: I32, C31**

Key words: Census, Heterogeneity, Poverty Mapping, Small Area Estimation

*I am grateful to Manuela Angelucci, Angus Deaton, Barbara Rossi and especially two referees for useful comments and discussions and to Maria Genoni for outstanding research assistance. I am also grateful to IPUMS for granting access to the 2000 Mexican Census extract and to the International Food Policy Research Institute for the Evaluation of PROGRESA data set. I am solely responsible for all errors and omissions. Address: Dept of Economics, Duke University, Social Sciences Building, PO Box 90097, Durham, NC 27708, taroz@econ.duke.edu.

1 Introduction

Estimates of poverty for virtually all countries in the world are calculated using income or expenditure data recorded in household surveys. The sample size and design of these surveys are often such that relatively precise poverty estimates can also be obtained for sub-national administrative units such as states or provinces. However, the estimation of welfare indicators becomes usually impossible or very imprecise if one is interested in poverty for small geographical areas such as towns or districts. For such small areas, a nationally representative household survey will often provide a handful of observations, or none at all. On the other hand, socio-economic indicators such as literacy or access to sanitation can usually be evaluated even for very small areas during census years, when the whole population is surveyed. However, in most countries, the census does not collect income or expenditure information, so that small area poverty estimates are typically not available even in census years. This is unfortunate, because such estimates may facilitate considerably the targeting of welfare programs towards communities with a higher marginal benefit.

To address this shortcoming, statistical agencies in several developing countries, in collaboration with the World Bank, have constructed maps of welfare measures using versions of a simulation-based imputation procedure described in [Elbers et al. \(2003\)](#).¹ [Tarozzi and Deaton \(2009\)](#) describe a simpler alternative projection-based estimator and they show that its performance is similar to that of the methodology in [Elbers et al. \(2003\)](#) under a variety of conditions.² Both estimators share a common intuitive framework: suppose that a survey contains a “large” number of observations from a given *region* for which both income/expenditure (“income” hereafter) and some predictors such as asset ownership or literacy are measured; suppose also that the same predictors are similarly measured in a census. Then it may be possible to recover precise—that is, low variance—estimates of welfare measures for areas significantly smaller than the region by using imputation procedures that merge survey information on the conditional distribution of income given the predictors with census information on the distribution of the predictors.

¹For a comprehensive description of the methodology used by the Bank, as well as for reference to the numerous applications, see www.worldbank.org/poverty.

²While the focus of this paper is on the poverty mapping techniques described in [Elbers et al. 2003](#) and [Tarozzi and Deaton 2009](#), Small Area Estimation has been for decades an important field of research for a broader statistical literature. Much of this literature has focused on the estimation of small area statistics (especially means and totals) using data from the United States Census. For recent reviews see [Pfeffermann \(2002\)](#), [Rao \(2003a\)](#) and [Rao \(2003b\)](#). [National Research Council \(2000\)](#) provides an overview that is less technical but is more specifically related to small area poverty estimation.

The methodology introduced by [Elbers et al. \(2003\)](#) first uses survey data to estimate a linear projection of (log) income on a set of predictors and to estimate the properties of the residuals. Simulations are then used to estimate local welfare measures making use of the parameter estimates from the first step and census information on the predictors at the local level. When the object of interest are small area estimates of poverty head counts, the estimator proposed in [Tarozzi and Deaton \(2009\)](#) first uses survey data to estimate the probability of being poor conditional on the predictors and then estimates head counts as the mean predicted probability of being poor for individuals in the small area.

Clearly, in order for either of these methodologies to allow correct statistical inference, one must assume a substantial degree of homogeneity within the region in the relationship between income and its predictors. Whether such homogeneity assumption holds is ultimately an empirical matter. [Tarozzi and Deaton \(2009\)](#) argue that the presence of local labor and rental markets and other local environmental differences are likely to generate heterogeneity which will violate this assumption, at least to some extent. Using data from the 2000 Mexican census, they use the projection-based estimator to show that disregarding heterogeneity leads to the estimation of confidence intervals for poverty headcount rates that do not have correct coverage. In other words, confidence intervals constructed for a poverty count H_0 using a “nominal” 95% coverage included in fact the true value H_0 less than 95% of the times. The validation exercise was possible because the Mexican census, unlike most census data in developing countries, includes measures of income. It was then possible to calculate “true” poverty counts at the local level, which were compared to estimates obtained through small area estimation techniques using data from synthetic “household surveys” generated from the census. A similar validation methodology is used in [Elbers et al. \(2008\)](#) who, using data from the Brazilian state of Minas Gerais, find that the coverage of confidence intervals constructed for small areas is close to nominal.³

The extent of area heterogeneity is likely to differ in different countries and regions. As a consequence, the incorrect coverage rates found in Mexico do not *imply* that incorrect coverage should be expected elsewhere, and in the same way correct coverage found somewhere does not imply that correct coverage should be expected everywhere. Unfortunately, direct assessments of the reliability

³Note, however, that [Elbers et al. \(2008\)](#) base their results on only 41 Monte Carlo replications, completed by pooling together samples drawn according to three separate sampling schemes. Most of the samples are drawn using a peculiar stratification scheme which is not clearly taken into account in the calculation of the standard errors. Also, the authors do not document how the predictors are chosen and include, in their model to predict income, mean values of income itself. Income means at the local level would clearly not be available in a typical poverty mapping exercise.

of poverty maps using validations are not generally possible, because they require the availability of income measures in the census whose absence is the very reason why small area estimation techniques are necessary. The purpose of this paper is to argue that useful, if only indirect, evidence on the extent of area heterogeneity is readily available in virtually any census. Such indirect evidence is provided by non-monetary indicators—such as literacy, asset ownership or work status—which are routinely included in a census. Validation exercises can be completed using such indicators to gauge the extent of area heterogeneity in their conditional distribution given predictors analogous to those typically used in poverty mapping. It is important to stress that such evidence will in any case only be indirect, and *cannot and should not* be used as a formal test on the accuracy of poverty mapping. To illustrate, the existence of area homogeneity of the conditional probability of being poor given, say, literacy and employment status *does not imply and is not implied by* homogeneity across areas of the conditional probability of being literate given employment status.

We argue, however, that differences across areas in relative prices, relative returns and other location-specific characteristics are likely to generate heterogeneity in the distribution of any wealth-related indicator conditional on other such indicators. To illustrate this point formally, we construct a simple model where individuals allocate an exogenously determined income between three different goods, with individual-specific preferences and area heterogeneity in relative prices. The solution of the model shows that the same sources of heterogeneity appear in both the conditional expectation of (log) income given expenditure in two of the consumed goods and in the conditional expectation of (log) expenditure on one good given consumption in the second good. Obviously, the model does not pretend to be realistic. However, it formalizes the idea that because most variables used in poverty mapping (e.g. income, asset ownership, education, dwelling characteristics etc.) are the result of choices made by households within the constraints of local prices and returns and other local unobservable conditions, the resulting heterogeneity, when present, is likely to characterize most or all conditional distributions that involve such variables. A series of Monte Carlo simulations where the data generating process is consistent with the simple model shows that ignoring heterogeneity leads to confidence intervals with coverage below the nominal level, regardless of whether the simulation-based estimator of [Elbers et al. \(2003\)](#) or the projection-based one in [Tarozzi and Deaton \(2009\)](#) is used.

Next, we evaluate the empirical relevance of our argument using two empirical illustration. The first, strictly related to the validation exercise in [Tarozzi and Deaton \(2009\)](#), uses data from an extract of the 2000 Mexican census. This data set also recorded measures of income, which allows to replicate a poverty mapping exercise, by assuming that one has access only to a synthetic “survey” sampled

from the census, and then comparing the estimated maps with the values reconstructed from the full census extract. [Tarozzi and Deaton \(2009\)](#) find that coverage rates of confidence intervals are below the nominal level for a large number of small areas and interpret the results as evidence of area heterogeneity. Here, we employ analogous simulation techniques to evaluate the performance of mapping when the outcomes of interest are based on variables which are routinely measured in a census. Namely, we estimate “literacy” and “sanitation” maps. In the former case the statistic of interest is the fraction of the population in an area who live in households with a literate head. In the latter case, we estimate the fraction of individuals with access to a toilet. According to the argument put forth in this paper, incorrect coverage rates in the estimation of poverty maps would also be reflected in similarly incorrect coverage in the estimation of maps of other indicators associated to poverty. We find that this is indeed the case. In the context of this empirical application, the performance of non-income maps constructed using only information usually available in a census *is* informative about the performance of poverty mapping.

While informative, the validation exercise described above does not replicate exactly the framework usually faced by researchers in poverty mapping. In a typical application, poverty maps are calculated over areas formed by a large number of smaller clusters which usually coincide with census enumeration areas and only include about 100 households. Because the census extract that we use does not include cluster identifiers, the same geographical unit (the *municipio*) represents both the cluster and the area. This is not ideal, because knowledge of cluster identifiers may allow to reduce concerns about area heterogeneity, by including cluster-specific means as predictors in the projection estimated in the first stage of the procedure. To address this drawback, we move to a second empirical validation where we use data from a complete census carried out in 1997-98 in a large number of rural and poor Mexican communities eligible for the well known welfare program PROGRESA. We therefore perform a validation exercise analogous to the one described earlier, but in this case we use state-specific aggregates as areas while using *localidades*, who on average include fewer than 50 households, as primary stage units. Coverage rates for poverty maps are overall better than in the application with 2000 Census data, but once again they remain below the nominal 95 percent. Also, as predicted by the main argument of this paper, incorrect inference in poverty mapping is also reflected in incorrect inference in alternative maps for measures of literacy, employment and access to sanitation.⁴

⁴The idea of using this data set for a poverty mapping validation is borrowed from [Demombynes et al. \(2008\)](#). However, [Demombynes et al. \(2008\)](#) use only 10 Monte Carlo replications and, above all, do so after generating small areas by randomly aggregating groups of *localidades*. By construction, this approach generates area homogeneity so it is

The rest of the paper is organized as follows. The next section briefly summarizes the statistical problem and provides details on estimation. Section 3 describes a simple model to illustrate why heterogeneity in some structural relations among household-level economic variables is likely to signal the presence of heterogeneity in alternative structural relations. Section 4 describes the two experiments with data from Mexico, and Section 5 concludes.

2 Statistical Background and Estimation

The purpose of this section is to describe the statistical framework of the construction of “maps” of welfare measures. Because our main purpose here is not to describe techniques for small area estimation, we only provide a simplified description for the case where the welfare measure is the fraction of units (e.g. households or individuals) in a given area with a given characteristic. The most common example is perhaps the poverty head count, that is, the fraction of the population with income per head below a given threshold (the poverty line). More generally, the same framework could be applied to the estimation of other statistics. For instance, a policy maker could be interested in evaluating the extent of illiteracy, unemployment or malnutrition across different areas (see for instance Fujii 2007). A formal and more general framework which also encompasses other poverty or inequality measures is described in Tarozzi and Deaton (2009) (TD hereafter), from which this section borrows.

Suppose that the object of interest is the head count ratio H for a “small area” A , where $A \subset R$ denotes a small area such as a town or district included in a larger “region” R such as a state. In a census, each small area will be further divided into a number of smaller units or clusters which are usually referred to as census “tracts” or enumeration areas (EAs), typically containing about 100 households. Let H_i denote a binary variable equal to one if individual i has a given characteristic, such as being poor, unemployed or illiterate. Let also $\Theta(A)$ denote the set of individuals who live in area A . The parameter of interest is then

$$H_A = \frac{1}{N_A} \sum_{i \in \Theta(A)} H_i, \tag{1}$$

where N_A is the number of elements in $\Theta(A)$.⁵ When data on H_i are recorded in a survey, the head count H_A could in principle be trivially estimated by its sample analogue $\bar{H}_A = n_A^{-1} \sum_{i \in \Theta_n(A)} H_i$,

not surprising that the authors find satisfactory coverage rates.

⁵For simplicity, we also abstract from the difference between individual and household specific welfare measures.

where $\Theta_n(A)$ indicates the set of individuals in the survey sampled from area A , and n_A indicates their number.⁶ However, a survey may include few or no observations at all on H_i if the area A is sufficiently small. Suppose instead that the survey sample from the broader region R includes a relatively large number of individuals for which (H_i, \mathbf{x}_i) are observed, where \mathbf{x}_i indicates a vector of household or community characteristics correlated with H_i . We also assume that \mathbf{x}_i is measured in the census as well, so that it is observed for all $i \in \Theta(A)$. The set of predictors \mathbf{x}_i can also include community characteristics or location-specific means of household-level variables recorded *only* in the census, as long as the survey includes detailed geographical identifiers which allow to merge such census information into the survey (Elbers et al. 2003). When instead census and survey provide separate measurements of the same variables, care should be taken to make sure that the information has been collected consistently across the two data sources, because it is well-known that data collection can be heavily influenced by the methodology adopted to elicit information (Deaton and Grosh 2000).

To fix ideas, it is now useful to adopt a superpopulation approach, assuming that the small area A includes an infinite number of individuals, so that using the law of iterated expectations we can write

$$H_A = E(H_i | i \in \Theta(A)) = E[P(H_i = 1 | \mathbf{x}_i, i \in \Theta(A)) | i \in \Theta(A)]. \quad (2)$$

Even if the survey sample contains no observations on H_i , the head count can be estimated using auxiliary information from the census if one can assume $P(H_i = 1 | \mathbf{x}_i, i \in \Theta(R)) = P(H_i = 1 | \mathbf{x}_i, i \in \Theta(A))$. Such “stability” assumption, labeled *area homogeneity* in TD, is closely related to the conditional independence assumptions which has been used extensively in statistics and econometrics to address missing data problems due to non-response, attrition, measurement error or unobserved counterfactuals in program evaluation.⁷ As emphasized in TD, area homogeneity is a demanding assumption in the estimation of small area statistics. Suppose for instance that H_A is a poverty head count and \mathbf{x} includes schooling or occupation variables. Heterogeneity in the conditional probability of being poor given \mathbf{x} is likely to arise due to differences across areas in the local rates of return. The presence in \mathbf{x} of measures of asset ownership will likely capture some of the variation in the rates of return, but assets are subject to similar concerns because their relative prices and rates of return may vary across areas. In general, area homogeneity will fail when the relevant conditional probability is a function of unobserved factors such as relative prices or returns, tastes or other location-specific characteristics

⁶If the sampling design implies unequal probability of selection into the sample, consistent estimates can still be obtained by using sampling weights.

⁷For extensive references see, for instance, Carroll et al. (1995), Heckman et al. (1999), Little and Rubin (2002), Rubin (1996), Todd (2007), Chen et al. (2008).

which vary across different areas. Concerns are also likely to become more severe when the census and the survey used for poverty mapping are completed in different years. This is a common circumstance, because while a census is usually completed only once every ten years, household surveys are often completed at shorter intervals. In such cases, the construction of poverty maps has to rely on information merged from two data sets collected possibly years apart from each other. This is problematic, especially for developing countries whose economies are often growing and changing rapidly.

Keeping all these *caveats* in mind, area homogeneity, coupled with the availability of predictors \mathbf{x}_i in both census and survey, provides the basis for the estimation of H_A using a two-step imputation procedure. TD suggest a simple approach based on the sample analogue of (2). First, the parameters of a binary dependent variable model $P(H_i = 1 \mid \mathbf{x}_i; \boldsymbol{\theta})$ are estimated using survey data from region R . Second, the estimated coefficients $\hat{\boldsymbol{\theta}}$ are used to form imputed values of the conditional probabilities for all census observations from the small area, and the head count for the small area is finally calculated as

$$\hat{H}_A = \frac{1}{N_A} \sum_{i \in \Theta(A)} P(H_i = 1 \mid \mathbf{x}_i; \hat{\boldsymbol{\theta}}). \quad (3)$$

In this paper, as in TD, we model the conditional probability $P(\cdot)$ using logit.⁸ In such case, the standard errors of \hat{H}_A can be estimated in a straightforward way using the delta method, recalling that because the census population is kept fixed and is therefore non-random, the only source of sampling error derives from parameter estimation. Hence:

$$\widehat{Var}(\hat{H}_A) = \hat{\mathbf{G}} \widehat{Var}(\hat{\boldsymbol{\theta}}) \hat{\mathbf{G}}' \quad (4)$$

where

$$\hat{\mathbf{G}} \equiv \frac{1}{N_A} \sum_{i \in \Theta(A)} F(\tilde{\mathbf{x}}_i' \hat{\boldsymbol{\theta}}) (1 - F(\tilde{\mathbf{x}}_i' \hat{\boldsymbol{\theta}})) \tilde{\mathbf{x}}_i',$$

and where $\tilde{\mathbf{x}}_i$ is the vector of functions of \mathbf{x}_i chosen as regressors and $F(\tilde{\mathbf{x}}_i' \hat{\boldsymbol{\theta}}) = e^{\tilde{\mathbf{x}}_i' \hat{\boldsymbol{\theta}}} \left(1 + e^{\tilde{\mathbf{x}}_i' \hat{\boldsymbol{\theta}}}\right)^{-1}$.

In the common circumstance that the survey data used for the estimation of $\hat{\boldsymbol{\theta}}$ are collected using a two-stage clustered design, the standard errors of the parameters should be adjusted taking intra-cluster correlation into account.⁹ It is important to note, however, that the expression in (4) may substantially underestimate the mean squared error (MSE) of the estimator if the number of

⁸Note that this approach can be made robust to parametric misspecification by estimating the conditional probability using a non-parametric estimator such as sieve-logit (see [Hirano et al. 2003](#) or [Chen et al. 2008](#)). However, non-parametric estimation is complicated by the necessity of choosing smoothing parameters and the degree of flexibility may be limited by sample size.

⁹More specifically, if $\boldsymbol{\theta}$ is estimated using logit, robust standard errors can be easily produced using built-in commands

observations in the small area A is not large. In fact, while \hat{H}_A is a consistent estimator for the area-specific mean value of $P(H_i = 1 \mid \mathbf{x}_i; \boldsymbol{\theta})$, the true object of interest is (1), that is, the mean value of H_i in the area. In a superpopulation approach the two quantities are identical, but in real empirical applications their difference (that is, the *bias* of the proposed estimator) may be different enough to make inference based on (4) incorrect. Intuitively, because poverty mapping is essentially an exercise in forecasting, confidence intervals for \hat{H}_A should be calculated taking into account not just the variance of the estimator, but rather its MSE, which also takes into account the (squared) bias. Letting $p_i \equiv P(H_i = 1 \mid \mathbf{x}_i; \boldsymbol{\theta})$, TD show that, when the area and the cluster coincide, the squared bias $b^2(\hat{H}_A)$ can be approximated by the following expression

$$\hat{b}^2(\hat{H}_A) = \frac{\hat{E}[(p_i - H_i)^2]}{N_A} + \frac{N_A - 1}{N_A} \hat{E}[(p_i - H_i)(p_{i'} - H_{i'})], \quad (5)$$

where both expectations are estimated using their respective sample analogues. A confidence interval for \hat{H}_A with nominal coverage $(1 - \tau)$ can then be constructed as

$$\hat{H}_A \pm \Phi^{-1}(1 - \tau/2) \times \left[\widehat{Var}(\hat{H}_A) + \hat{b}^2(\hat{H}_A) \right], \quad (6)$$

where $\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution function of a standard normal. In a series of Monte Carlo experiments, TD show that, when area homogeneity holds, confidence intervals constructed in this way have approximately correct coverage with a variety of data generating processes, even when the number of observations within the area A is as small as 100 individuals. In the common situation where each area is composed of a number of different clusters, and under the assumption that the deviations $(p_i - H_i)$ are uncorrelated across different clusters, the expression in (5) becomes

$$\hat{b}^2(\hat{H}_A) = \frac{\hat{E}[(p_i - H_i)^2]}{N_A} + \frac{\sum_{c \in \Lambda(A)} N_c(N_c - 1)}{\left(\sum_{c \in \Lambda(A)} N_c\right)^2} \hat{E}[(p_i - H_i)(p_{i'} - H_{i'})], \quad (7)$$

where $\Lambda(A)$ is the set of all clusters in area A and N_c is the number of individuals in the population residing in cluster c within area A , so that $\sum_{c \in \Lambda(A)} N_c = N_A$.¹⁰ The first term on the right-hand side of this expression is associated with the variance of the deviation between p_i and the binary indicator H_i and it decreases quickly to zero with the size of the area population. The second term on the right-hand side will also, in general, decline towards zero when area population size increases, because

in most statistical packages. For instance, in Stata this can be done using the `svylogit` command, or the `cluster` option in the standard `logit` command.

¹⁰This expression results from adapting the result in Appendix B of TD to the case where each area includes a number of different clusters and assuming no inter-cluster correlation.

the numerator of the ratio that multiplies the estimated expectation is the number of within-cluster cross-products, while the denominator is the number of within-area cross-products.¹¹

To conclude, let $\Lambda_n(R)$ indicate the set of all clusters c in region R included in the sample and let $\Theta_n(c)$ indicate the set of the n_c individuals sampled from cluster c (and define $\Theta_n(R)$ analogously). Let also $\hat{p}_i \equiv P(H_i = 1 \mid \mathbf{x}_i; \hat{\boldsymbol{\theta}})$. Then

$$\begin{aligned}\hat{E}[(p_i - H_i)^2] &= \frac{1}{\sum_{c \in \Lambda_n(R)} n_c} \sum_{c \in \Lambda_n(R)} \sum_{i \in \Theta_n(c)} (\hat{p}_i - H_i)^2 \\ \hat{E}[(p_i - H_i)(p_{i'} - H_{i'})] &= \frac{\sum_{c \in \Lambda_n(R)} \sum_{i \in \Theta_n(c)} \sum_{i' \in \Theta_n(c), i' \neq i} (\hat{p}_i - H_i)(\hat{p}_{i'} - H_{i'})}{\sum_{c \in \Lambda_n(R)} n_c (n_c - 1)}.\end{aligned}$$

These expressions can then be plugged into (7) and confidence intervals can finally be calculated using (6) together with the estimated variance calculated using (4). This projection based estimator will be used throughout most of this paper.

Elbers et al. (2003) (ELL hereafter) describe a class of alternative, simulation-based estimators.¹² These more complex estimators start from a first-step estimation of the projection of the logarithm of income on the predictors \mathbf{x} . Because poverty measures are typically functions of the distribution of income (and not just of its mean) ELL need to make further explicit functional form assumptions about aspects of the distribution of the projection residuals such as the form of heteroskedasticity. The projection estimates, together with the estimated structure of the residuals, are then fed into a simulation procedure reminiscent of multiple imputation (Rubin 1996). We postpone a detailed description of the simulation procedure to Section 3.1, where we use both the projection method and ELL in a Monte Carlo experiment to evaluate the performance of both estimators under different degrees of area heterogeneity. We will show that both estimators perform well and produce confidence intervals of approximately correct coverage when the area homogeneity assumption holds, while neither does when the assumption fails.

¹¹For instance, if each area is composed of p equally-sized clusters of m individuals, the fraction will be $pm(m - 1)/p^2m^2 = p^{-1}(m - 1)/m$, which approaches zero when the number of clusters become large. Note that instead the covariance-related term in (5) did not converge to zero with the increase in area size, because in that case area and cluster coincided so that p was always equal to one.

¹²See also Elbers et al. (2002) and Demombynes et al. (2008) for details.

3 A Formal Illustration

In this section we describe a very simple model to illustrate formally the idea that if factors such as relative price variation across areas lead to area heterogeneity, such heterogeneity is likely to be reflected not only in area heterogeneity in the distribution of income given predictors, but also in heterogeneity in the distribution of each predictor given the remaining ones. We assume that individuals allocate an exogenously determined income between three different goods, with heterogeneity in preferences across individuals and a varying degree of heterogeneity in prices across areas. In order to allow for the existence of intra-cluster correlation within each area, we also assume that each area is composed of a number of synthetic enumeration areas or clusters, and that individuals who live within the same cluster have correlated tastes. Within the boundaries of such oversimplified model, we show that heterogeneity in the distribution of expenditure in one of the goods conditional on expenditure in a second good is reflected also in heterogeneity of the distribution of income given expenditure in both goods.

Suppose that each area a is populated by individuals i , grouped into a number of clusters c which can be interpreted as representing different geographical subdivisions of the area.¹³ Assume also that individuals are endowed with different nominal income Y_i and that the budget is allocated among three goods, whose relative prices are potentially different across different areas. We assume that preferences are Cobb-Douglas and that each individual solves the following constrained maximization problem:

$$\begin{aligned} \max_{x_{i1}, x_{i2}, x_{i3}} \quad & U(x_{i1}, x_{i2}, x_{i3}) = x_{i1}^{\alpha_{i1}} x_{i2}^{\alpha_{i2}} x_{i3}^{\alpha_{i3}}, \quad \alpha_{i1} + \alpha_{i2} + \alpha_{i3} = 1 \\ \text{s.to} \quad & p_1^a x_{i1} + p_2^a x_{i2} + p_3^a x_{i3} \leq Y_i. \end{aligned} \quad (8)$$

The existence of heterogeneity in relative prices across areas implicitly requires that markets are not perfectly integrated, and that each area can be seen as a separate economy. In some cases this may be a very strong assumption, but in reality such differences in relative prices are often observed, especially in developing countries where market integration is limited. The solution of the utility maximization problem leads to the following standard demand functions:

$$x_{ij} = Y_i \frac{\alpha_{ij}}{p_j^a}, \quad j = 1, 2, 3. \quad (9)$$

¹³We only denote individuals by the subscript i , instead of aci , to avoid notational clutter.

From (9), it follows through simple algebraic manipulations that

$$\ln Y_i = 0.5 \ln \left(\frac{p_1^a p_2^a}{\alpha_{i1} \alpha_{i2}} \right) + 0.5 \ln x_{i1} + 0.5 \ln x_{i2} \quad (10)$$

$$\ln x_{i1} = \ln \left(\frac{\alpha_{i1} p_2^a}{p_1^a \alpha_{i2}} \right) + \ln x_{i2}. \quad (11)$$

Suppose now that there is heterogeneity in the budget shares and that such heterogeneity is correlated within clusters. Specifically, we assume that $\alpha_{ij} = \alpha_j e^{\eta_{jc} + \varepsilon_{ji}}$, $j = 1, 2, 3$, where α_j is a scalar and $\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}$ are random variable with variance $\sigma_{\varepsilon_j}^2$, $j = 1, 2, 3$ whose distribution guarantees that $\alpha_{i1} + \alpha_{i2} + \alpha_{i3} = 1$. The spatial correlation of tastes is reflected by the cluster-specific component η_{jc} . Assume finally that the area heterogeneity in prices takes the specific functional form $p_j^a = p_j e^{\phi_j^a}$, $j = 1, 2, 3$, where p_j is a constant and ϕ_j^a is an area-specific random variable with variance $\sigma_{\phi_j}^2$, $j = 1, 2, 3$.

Under these conditions, equations (10) and (11) can be re-written as

$$\ln Y_i = 0.5 \ln \left(\frac{p_1 p_2}{\alpha_1 \alpha_2} \right) + 0.5 \ln x_{i1} + 0.5 \ln x_{i2} + 0.5 u_i^Y \quad (12)$$

and

$$\ln x_{i1} = \ln \left(\frac{p_2 \alpha_1}{p_1 \alpha_2} \right) + \ln x_{i2} + u_i^{x_1}, \quad (13)$$

where $u_i^Y \equiv \phi_1^a + \phi_2^a - \eta_{1c} - \eta_{2c} - \varepsilon_{1i} - \varepsilon_{2i}$ and $u_i^{x_1} \equiv \varepsilon_{1i} + \eta_{1c} - \phi_1^a - \varepsilon_{2i} - \eta_{2c} + \phi_2^a$. Both equation (12) and (13) can be estimated if data on Y_i , x_{i1} and x_{i2} are available, although it is easy to see that the unobserved components u_i^Y and $u_i^{x_1}$ are by construction correlated with the observable right-hand side variables in their respective equations. However, the correlation between the residual and the regressors is not a concern here, because poverty mapping is an exercise in forecasting which does not attempt to estimate causal relationships. What does, instead, generate difficulties is that the presence of area-specific relative prices generates area heterogeneity in the projection coefficients. Specifically, consider the univariate model (13).¹⁴ The Appendix shows that under standard regularity conditions the Best Linear Projection (BLP) of $\ln x_{i1}$ on $\ln x_{i2}$ estimated by Ordinary Least Squares (OLS) using data from the larger region R will be different from the area-specific BLP. Specifically, the Appendix shows that

$$BLP(\ln x_{i1} \mid \ln x_{i2}; R) = \beta_0 + \beta_1 \ln x_{i2} \quad (14)$$

$$BLP(\ln x_{i1} \mid \ln x_{i2}; a) = \beta_0^a + \beta_1^a \ln x_{i2}, \quad (15)$$

¹⁴A similar argument could be extended to the income model (12). We choose to focus on model (13) because the presence of a single regressor simplifies the algebra.

where

$$\begin{aligned}\beta_0 &= \ln\left(\frac{p_2\alpha_1}{p_1\alpha_2}\right) + (1 - \beta_1)(\mu_{\ln Y} + \ln\alpha_2 - \ln p_2) \\ \beta_1 &= \frac{\sigma_{\ln Y}^2}{\sigma_{\ln Y}^2 + \sigma_{\phi_2}^2 + \sigma_{\varepsilon_2}^2 + \sigma_{\eta_2}^2} \\ \beta_0^a &= \ln\left(\frac{p_2\alpha_1}{p_1\alpha_2}\right) - \phi_1^a + \beta_1^a\phi_2^a + (1 - \beta_1^a)(\mu_{\ln Y} + \ln\alpha_2 - \ln p_2) \\ \beta_1^a &= \frac{\sigma_{\ln Y}^2}{\sigma_{\ln Y}^2 + \sigma_{\varepsilon_2}^2 + \sigma_{\eta_2}^2}.\end{aligned}$$

The discrepancy between area-specific and region-wide BLPs will generate obvious difficulties in poverty mapping exercises, because the available (survey) data will in general only allow consistent estimation of β_0 and β_1 , both of which will in general be different from the area-specific coefficients that would be necessary for correct imputation. Given that the BLP coefficients are a fundamental ingredient in poverty mapping, assuming area heterogeneity should be expected to lead to bias in the estimation of poverty measures based on consumption of good 1, regardless of whether estimation is done using a simulation-based method such as ELL or a projection-based method such as TD. Similar arguments would lead to incorrect inference on income-based poverty measures.¹⁵ Note also that without area heterogeneity in prices we would have $\phi_1^a = \phi_2^a = \sigma_{\phi_2} = \sigma_{\phi_1} = 0$ so that the BLPs in the region and in each area would coincide and both the poverty mapping methodologies considered in this paper should be expected to produce correct inference. The next subsection uses a Monte Carlo simulation to further illustrate this point.

3.1 A Monte Carlo Experiment

We assume now that an economy is described by a Data Generating Process (DGP) consistent with the model described in the previous section. Income Y_i is exogenous and generated from a normal distribution with mean 1000 and standard deviation 200. We generate individual-specific preference parameters assuming that $\alpha_1 = \alpha_2 = 0.25$. The cluster-specific components (η_{1c}, η_{2c}) are drawn

¹⁵In our model, income is predicted with continuous component of expenditure. However, typical poverty applications mostly rely on categorical predictors, such as literacy, asset ownership or dwelling characteristics. Even so, the model's insights remain unchanged, because the association between such indicators and income would remain a function of relative prices. Also, one could construct a similar model where, for instance, demand for assets is determined by discrete choice equations where purchase depends on relative prices. We have chosen to focus on a model with continuous components of expenditure because such model is simpler to solve and the econometric implications are more transparent and have relatively simple closed form solutions.

independently from a uniform distribution in the interval $[-0.05, 0.05]$, while the individual-specific preference parameters $(\varepsilon_{1i}, \varepsilon_{2i})$ are drawn independently from a uniform distribution in the interval $[-0.25, 0.25]$. This implies that all random preference components have zero mean, while $\sigma_{\eta,j} = \sqrt{0.1/12}$, $\sigma_{\varepsilon,j} = \sqrt{0.5/12}$, $j = 1, 2$.¹⁶ Prices are generated as random draws from a tri-variate normal distribution with mean 50, standard deviation σ_p and correlations equal to 0.5. The parameter σ_p is chosen in such a way that prices are always positive. Once income, prices and preferences are determined, consumption levels of goods 1 and 2 are calculated using the equilibrium demand equations (9). The most crucial parameter in the model is the standard deviation σ_p (or, equivalently, σ_ϕ), which measures the extent of area heterogeneity. The case where $\sigma_p = 0$ corresponds to a situation where there is no area heterogeneity in relative prices. As we will describe later, this DGP produces models where the first stage projections have very good explanatory power and the intra-cluster correlation of the unexplained residuals remains small.

Each Monte Carlo experiment proceeds as follows. First we generate data as described above for the synthetic “census” of a region composed of 20 small areas composed of 100 equally-sized clusters of 100 individuals each. We assume that the object of interest is either the fraction of individuals in each area with income below a poverty line $z_y = 1000$ (which we indicate with H_y^a), or the fraction of individuals in each area with consumption of x_1 below a threshold $z_{x_1} = 5$ (denoted $H_{x_1}^a$). The census is then kept fixed and treated as the actual population. In each of 200 Monte Carlo replications we draw a sample of 10 individuals from 50 randomly selected clusters from the synthetic census. Each synthetic sample is therefore composed of 500 units. Both individuals and areas are selected using simple random sampling without replacement. Once the synthetic survey has been generated, we calculate the estimates \hat{H}_y^a and $\hat{H}_{x_1}^a$ for each area as well as the corresponding 95 percent confidence intervals either using the projection method described in Section 2 or using the more widely used simulation method proposed in ELL. Note that both estimators should be expected to perform similarly regardless of whether area homogeneity holds. More specifically, both estimators should produce confidence interval of approximately correct coverage when the assumption holds and should instead overestimate coverage when it does not, as also suggested by the Monte Carlo simulations in TD.

When using the projection method, we estimate the conditional probabilities $P(Y_i < z_y \mid x_{i1}, x_{i2})$ and $P(x_{i1} < z_{x_1} \mid x_{i2})$ using a logit model, and point estimates and MSE of head counts are estimated

¹⁶Recall that the variance of a continuous random variable having a uniform distribution over the interval $[a, b]$ is equal to $(b-a)/12$. This DGP implies that both α_{1i} and α_{2i} are bounded between $0.25e^{-0.3} = 0.1852$ and $0.25e^{0.3} = 0.3375$. The parameter α_{3i} is calculated using the homogeneity condition $\alpha_{1i} + \alpha_{2i} + \alpha_{3i} = 1$, which implies that $\alpha_{3i} \in [0.325, 0.6296]$.

as described in (3) and (6). There are several variants of ELL and the one we use here is closer to Demombynes et al. (2008). In describing the details of ELL we also borrow from TD. Here we describe in detail the procedure for the estimation of $H_{x_1}^a$, but the estimation of H_y^a proceeds in analogous fashion with straightforward modifications.¹⁷ The starting point is the estimation of model (13) with OLS, under the assumption that the model can be written as $\ln x_{i1} = \beta_0 + \beta_1 \ln x_{i2} + u_c + v_i$, where u_c is a cluster fixed effect and v_i is an individual-level error. The estimates $(\hat{\beta}_0 + \hat{\beta}_1)$ and their estimated sampling covariance matrix are then stored away. Next, the cluster fixed effects u_c are estimated as the mean value of the estimated BLP residuals over all the observations from the same cluster c . Then estimates of the individual-specific errors \hat{v}_i are calculated as deviations of the BLP residuals from the corresponding estimated cluster-specific fixed effects. The cluster-specific effects are assumed to be *i.i.d.* and homoskedastic, while the variance of the idiosyncratic errors is estimated assuming the following model:

$$v_i^2 = \frac{Ae^{\pi_0 + \pi_1 \ln x_{i2} + s_i}}{1 + e^{\pi_0 + \pi_1 \ln x_{i2}}}. \quad (16)$$

The parameter A is set to $A^* = (1.05) \max\{\hat{v}_i^2\}$ and $(\hat{\pi}_0, \hat{\pi}_1)$ and their standard errors are obtained by estimating with OLS the sample analogue of the following model which derives directly from (16) above:

$$\ln \left(\frac{v_i^2}{A - v_i^2} \right) = \pi_0 + \pi_1 \ln x_{i2} + s_i. \quad (17)$$

Next, the variance of the idiosyncratic errors is calculated using the following expression, obtained taking the expectation of a Taylor approximation of (16) around $s_i = 0$:

$$\hat{\sigma}_{i,v}^2 = \frac{A^* B_i}{1 + B_i} + 0.5 \times \widehat{var}(s_i) \frac{A B_i (1 - B_i)}{(1 + B_i)^3}, \quad (18)$$

where $B_i \equiv e^{\hat{\pi}_0 + \hat{\pi}_1 \ln x_{i2}}$. In the last step of the first stage, standardized idiosyncratic residuals \hat{v}_i^* are calculated by dividing the residuals \hat{v}_i by $\hat{\sigma}_{i,v}$ and subtracting from each the mean of these ratios calculated over the whole sample.

The estimated sampling distributions of $(\hat{\beta}_0, \hat{\beta}_1, \hat{\pi}_0, \hat{\pi}_1)$, together with the empirical distribution of the estimated cluster-specific and idiosyncratic errors, form the inputs that can now be fed into the simulation procedure, which now proceeds making use also of the synthetic census. The structure of each simulation step r is as follows. First, a set of parameters $(\hat{\beta}_0^r, \hat{\beta}_1^r, \hat{\pi}_0^r, \hat{\pi}_1^r)$ is drawn from their estimated sampling distribution. Second, each cluster in the census is assigned a cluster-specific error

¹⁷In estimating the first-stage projection for H_y^a , we use $\ln x_{i1}$ and $\ln x_{i2}$ as predictors, consistently with equation (10). In a real application where the correct functional form is not known and continuous predictors are available, the researcher would likely also experiment with the inclusion of powers and interactions.

\hat{u}_c^r drawn randomly from the empirical distribution of all \hat{u}_c 's.¹⁸ Third, each observation in the census is assigned a normalized idiosyncratic error \hat{v}_i^{*r} drawn at random from the corresponding empirical distribution. Fourth, heteroskedastic errors \hat{v}_i^r are calculated by using the parametric model in (18) evaluated at the simulated parameter values. Lastly, simulated values for $\ln x_{i1}^r$ are generated as $\hat{\beta}_0^r + \hat{\beta}_1^r \ln x_{i2} + \hat{u}_c^r + \hat{v}_i^r$, and the point estimate $H_{x_1}^{ar}$ is calculated as the fraction of individuals in area a with $\ln x_{i1}^r$ below the threshold z_{x_1} . The mean and the variance over a large number of replications are finally used as an estimate of $H_{x_1}^a$ and $Var(H_{x_1}^a)$.

After each Monte Carlo simulation, we determine whether the value of each area-specific estimate lies within the 95% confidence interval calculated using either the projection method or the simulation-based ELL estimator.

The resulting coverage rates for different degrees of area heterogeneity are displayed in Figure 1, while we report selected summary statistics for the Monte Carlo results in Table 1. Each histogram in Figure 1 represents the distribution of coverage rates over the 20 synthetic areas. The extent of area heterogeneity, measured by the standard deviation of prices, is kept constant in each row while it increases from top to bottom. The graphs in the first two columns show coverage rates for the “income-based” head counts H_y^a , while those in the third and fourth columns refer to the head counts $H_{x_1}^a$ based on consumption of good 1. Coverage rates calculated using ELL are shown in columns 1 and 3, while those produced with the projection method are shown in columns 2 and 4. If the estimator is asymptotically normal and the MSEs are estimated correctly, coverage for all areas should be approximately equal to the nominal size of the confidence intervals, that is, 0.95. In such case, the histograms would show a spike at or close to 0.95. With no differences in relative prices, we expect to see correct coverage using either ELL or the projection method, while we expect both methodologies to produce increasingly poorer coverage rates the larger the extent of area heterogeneity measured by σ_p . The simulation results are consistent with these priors.

When $\sigma_p = 0$, the DGP produces head counts H_y^a and $H_{x_1}^a$ approximately equal to 0.5 in all areas. For both sets of head counts, both ELL and the projection method produce coverage rates of approximately correct size for all areas, as shown by the peaks around 0.95 in all four graphs along the first row of Figure 1. The statistics in columns 1 and 2 of Table 1 also show that both

¹⁸Note that the ELL algorithm can be easily adapted to assign the *same* cluster-specific error to all clusters within an area. If the only source of heterogeneity is due to the presence of area fixed effects, this modification would likely lead to conservative (that is, *too wide*) confidence interval. Throughout the paper we only focus on the standard version of ELL described in this section, which is what most researchers use.

methods produce precise estimates, with confidence intervals whose average width range from 2.5 percent (for H_y^a estimated with ELL) to 5.1 percent (for $H_{x_1}^a$ estimated with the projection method). The simulation method generates smaller standard errors (likely due to the parametric assumptions adopted at different stages of the estimation) but coverage rates appear to be closer to nominal ones when we use the projection method, although only marginally so.

The results change considerably when heterogeneity in prices is introduced. The graphs along the second row of Figure 1 show coverage rates when the standard deviation of prices σ_p is equal to 0.5. Summary statistics for the Monte Carlo results are in columns 3 and 4 of Table 1. This modification produces variation in the true values of $H_{x_1}^a$ (which now range from 0.47 to 0.56) because demand for good 1 now changes depending on its relative price. Note that the increased value of σ_p still leads to fairly homogeneous relative prices. For instance, the ratio p_1^a/p_2^a in the 20 synthetic areas ranges from a minimum of .98 to a maximum of 1.02. Even so, coverage rates for both parameters now worsen considerably, and both estimators produce average coverage rates around 75 percent for H_y^a and around 85 percent for $H_{x_1}^a$. In most cases area-specific coverage rates remain above 60 percent, although ELL also produces a long left tail in the histogram for both parameters, with a handful of areas whose coverage rate is 50 percent or below. Note that, as also pointed out in TD, the incorrect coverage arises despite the fact that location-specific unobservables explain a very small fraction of the total variance of the BLP residual. In this DGP, such fraction (that is, the intra-cluster correlation of the residual) remains around 4 percent of the total variance or below.¹⁹

The bottom graphs of Figure 1 and columns 5 and 6 of Table 1 contain the results of the Monte Carlo experiment when σ_p is increased to 1. This produces more heterogeneous relative prices, now ranging from 0.96 to 1.03, which in turn leads to area-specific good 1 head counts ranging from 45 to 60 percent. Coverage rates worsen further, to the point of becoming on average below 60 percent for H_y^a and around 72 percent for $H_{x_1}^a$. The distribution of coverage rates also become more skewed, so that in this case a large number of areas have coverage rates of 50 percent and below.

To summarize, at least in the context of this very simple model, poor coverage in “good 1 head counts” signals poor coverage for income poverty maps, because heterogeneity in the underlying conditional probabilities is due to a common cause, that is, heterogeneity in relative prices. The model is

¹⁹The fact that area heterogeneity can cause incorrect inference even in situation where location effects are very small is also consistent with the Monte Carlo results in Table 3 of TD. TD show that the concerns only disappear if, absent any heterogeneity in the BLP slopes, the variance of the area-specific location effect is negligible relative to the variance of the cluster-specific location effect.

obviously a naive representation of a true economy, but it should still provide a useful exemplification of the main argument put forth in this paper: that is, that unobserved heterogeneity in local characteristics such as prices is likely to be reflected both in heterogeneity in the distribution of income given some of its correlates and in heterogeneity in the distribution of such correlates conditional on the others. In the next sections, we use data from Mexico to study whether such an argument is consistent with evidence from a real empirical setting.

4 Empirical Illustrations

This section evaluates the merit of the arguments put forth in the previous section by using data from Mexico to construct two alternative validation exercises. The first uses data from the 10.6% random extract of the 2000 Mexican Census from the Integrated Public Use Micro Sample (IPUMS) (Ruggles and Sobek 1997, Minnesota Population Center 2007) while the second uses information from a complete census carried out in 1997 and 1998 in communities that were considered eligible for participation in the well-known PROGRESA welfare program (ENCASEH).²⁰ Both validation exercises have a common rationale: both datasets include measures of income or expenditure which could be used to directly construct poverty maps for very small areas. We therefore use each data set as the population of interest and we perform with each a simulation exercise analogous to the one described in Section 3.1, that is, we compare coverage rates for poverty counts with rates for other indicators and we evaluate whether the quality of inference in the latter is informative about the quality of inference in the former. The results from both Census 2000 and ENCASEH confirm that this is the case. Because both the Monte Carlo results in Section 3.1 and those in TD show that the ELL simulation method and the projection method perform similarly under a broad set of scenarios and given that the latter estimator is significantly simpler and faster we only use the projection method in our exercise.

4.1 A Quasi-validation Using Data from the 2000 Mexican Census

As customary for most census micro-data, the 2000 Mexican Census includes a long list of correlates of income. Among the others, such covariates include housing characteristics, household composition, asset ownership, occupation and schooling indicators of each household member. Unlike most census data sets, however, Mexico 2000 also includes a measure of individual income during the previous 30 days. This allows TD to compare the poverty rates calculated with the *actual* income data with those

²⁰For documentation and reports see <http://www.ifpri.org/themes/progresas.htm>.

estimated simulating a poverty mapping procedure carried out assuming that only a random sample of observations is available from the census. The results in TD suggest the existence of substantial heterogeneity in the areas considered. In fact, while poverty mapping produces overall informative results, the estimated MSEs appear to be biased downwards for a large number of small areas. Here we show that an analogous under-coverage of confidence intervals can be observed if poverty mapping methodologies are applied to the calculation of maps for alternative economic indicators, where both the indicators and the predictors are variables typically measured in census data. The indicators we choose are the fraction of individuals living in households where the head can read and write in at least one language (“literacy head counts”) and the fraction of individuals with access to a toilet (“sanitation head counts”). As in TD, we use data from three of the largest Mexican states, namely Chiapas, Oaxaca and Veracruz. We treat each state as a separate region (R), and we adopt smaller administrative units called *municipios* as small areas (A). Because the IPUMS data set only includes a 10.6% extract of the complete micro-data, we first generate a complete “pseudo-census” with a number of observations equal to actual census population. For this purpose, we replace each observation in the extract with identical replicates in number identical to the (integer) weight provided in the data set. The pseudo-census created in this fashion is then treated as the actual (non-random) population of interest.²¹ Because the census extract includes only a handful of observations from some *municipios*, we only use observations from *municipios* with at least 500 households. Table 2 shows that this leads to a number of small areas equal to 108 in Chiapas, 240 in Oaxaca and 182 in Veracruz.

All the three states have high poverty rates. The mean area-specific head count is 0.66 in Chiapas, 0.65 in Oaxaca and 0.44 in Veracruz. The mean fraction of literate household heads is similar across states and close to 70 percent, while the mean proportion of individuals with access to a toilet ranges from 71 percent in Chiapas to 78 percent in Veracruz. Figure 2 shows that there is a substantial degree of variation in the heterogeneity of these statistics across states. In all states, the *municipio*-level distribution of poverty head counts covers most of the unit interval and is relatively uniform, with more evidence of left-skew in Chiapas and Oaxaca, the two poorest states. The range of the sanitation head counts distributions is similarly wide, and all distributions are left-skewed, including in Veracruz. The distributions of the literacy head counts are left-skewed too, but appear to be somewhat more concentrated around their mean.

²¹As suggested by a referee, an alternative approach would have been to use the extract itself as the population. We do not choose to do so because, although neither approach would allow to work with a real census, using the inflation factors at least allows us to use a data set where population size is the same as the one of the real census.

Table 3 contains the list of variables included as predictors of poverty status. Individuals are considered to be poor when income per head is below 200 Pesos per month.²² Note that the list of predictors also includes location means. As noted in ELL, the inclusion of location means typically reduces the correlation of the residuals in the first-stage regression, which in turn reduces the bias of the estimator described in Section 2. In many circumstances, such location means can be included in the mapping procedure even though they refer to variables which are *only* recorded in the census, as long as the survey includes location identifiers which allow to link them to the observations included in the survey.

When the imputation methodology is applied to the construction of “literacy maps”, we replace the poverty indicator with a binary variable equal to one if the household head is literate. In this case the list of predictors used is the same as in Table 3, with the omission of all information on schooling. Finally, when the object of interest is a map of *municipio*-specific sanitation head counts, the dummy variable for access to toilet is dropped from the predictors (together with its mean) and is used as dependent variable.

For each of the three economic indicators (poverty head counts, literacy and sanitation rates), we evaluate the coverage of 95% nominal confidence intervals using 250 Monte Carlo simulations. We carry out independent simulations for the three states of Chiapas, Oaxaca and Veracruz. In each replication we assume that the object of interest is a map of *municipio*-specific indicators, but that the researcher has access only to a synthetic “survey”. The survey is generated by drawing from the state pseudo-census clusters of ten households from 50 *municipios* selected at random without replacement.²³

In each Monte Carlo replication, we first use the synthetic survey sample to estimate the parameters in the conditional probability in (3) using a logit model. These parameters are assumed to be common to all observations within the same state, and are used together with information on the predictors for all households in the state pseudo-census to estimate the indicator for each *municipio*. The confidence intervals are calculated as in (6). Finally, we calculate coverage rates as the fraction of Monte Carlo simulations where the nominal 95% confidence intervals include the “pseudo-true” *municipio*-specific indicators, calculated using the actual data included in the pseudo-census.

The results of the simulations are summarized in Figure 3. Each graph shows a histogram of the

²²In 2000, the PPP exchange rate between USD and Mexican Peso was 6.79, so that 200 Pesos correspond to approximately one PPP dollar per person per day (Heston et al. 2006).

²³All estimations are done using sampling weights, because this sampling design generates differences in the probability of selection across households.

coverage rates for a state-head count pair. Each histogram is constructed using the coverage rate for a given small area (that is, a *municipio*) as an observation.²⁴ If the confidence intervals had correct coverage, the whole mass in each histogram would be centered on 0.95. The first conclusion one can draw from these results is that clearly this is not the case. Although coverage rates remain acceptably large and above 0.75 for most small areas, there is a substantial fraction of *municipios* where coverage remains much below the nominal value. Coverage rates for poverty head count ratios remain below 0.75 in 35 percent of *municipios* in Chiapas, 38 percent in Oaxaca and 39 percent in Veracruz (see Table 4, columns 1-3), while they lie below 50% in approximately 11 percent of areas. Although the estimated confidence intervals appear to systematically overstate the precision of the estimator, they are relatively wide. For instance, the mean width of a confidence interval in the estimation of poverty head counts is 0.33 in Chiapas (minimum 0.19 and maximum 0.57), 0.38 in Oaxaca (minimum 0.23 and maximum 0.71) and 0.36 in Veracruz (minimum 0.21 and maximum 0.70). The second important conclusion, which is central for the argument put forth in this paper, is that in all states the distributions of coverage rates appear to be remarkably similar among different indicators. The only exceptions are the sanitation indices in Chiapas, where coverage rates are substantively better, although at the expense of very large confidence intervals. Of course, these results cannot be generalized to other contexts without scrutiny. However, they are consistent with the idea that, first, census data alone always offer the opportunity to complete validations to evaluate the performance of small area estimation techniques in constructing maps of non-poverty indices and, second, that such validations may provide important signals about how effectively such methodologies will be able to produce poverty maps with the correct precision.

4.1.1 Model Selection

Our interpretation of the results presented in Figure 3 relies crucially on the fact that discrepancies between nominal and actual coverage rates are evidence of area heterogeneity. In real poverty mapping applications, discrepancies could also be due to predictors being measured differently in the census and in the survey (for instance, because of differences in how occupations or schooling are coded). However, this possibility does not arise within our simulation, because the synthetic survey data are drawn directly from the census, so comparability of predictors between the two data sources holds by construction. Another possibility is that the poor performance of mapping is due to an incorrect choice

²⁴Note that the histograms of the coverage rates for poverty head count are very similar but not identical to those in TD, because in this paper we have only included in the analysis *municipios* with at least 500 households.

of the conditional model. On the one hand, the choice of the predictors is necessarily atheoretical, both because in a typical application the selection is limited by the variables that can be matched between the census and the survey, and because the first-stage projections are not structural relationships. On the other hand, it is conceivable that the use of a statistical criterion to select which of the predictors in Table 3 to include in the first stage may improve the performance of the mapping exercise. We experiment with two alternative selection procedures, one based on an information criterion and another based on the statistical significance of the first stage projection coefficients. In each of the two cases, the model is selected based on the specific synthetic survey drawn from the pseudo-census. In other words, these procedures allow the selection of a different model in each Monte Carlo replication. In both cases we also include, among the candidate predictors, a list of interactions between household-specific covariates and *municipio*-level means in Table 3. The rationale is that such interactions may help to some extent in controlling for heterogeneity in the projection coefficients across areas.

With the first selection approach, the first-stage model is chosen by using a Bayesian Information Criterion (BIC, Schwarz 1978). First, we estimate a series of univariate logit models for the dependent variable H_i where each of the K candidate predictors is entered separately. In our simulations, H_i is equal to one if the household is poor, or if it has access to sanitation, or if the head is literate. Second, we sort the models in decreasing order of pseudo- R^2 . Let $k = 1, \dots, K$ denote the index of the sorted predictors. In the third step, we calculate the BIC for K models, where each model includes the first k sorted predictors. Finally, we identify the model that maximizes the BIC. If the selected model is the j th in the ranking, all variables with a ranking not larger than j are included as predictors. Intuitively, this criterion trades off the increase in the pseudo- R^2 that results from the inclusion of more predictors with a penalization that has the purpose of limiting the number of regressors, to avoid overfitting. We show the resulting coverage rates in Figure 4, while columns 4-6 of Table 4 summarize the results of the simulations.

Visual inspection of the histograms in Figure 4 is sufficient to see that the accuracy of the poverty mapping declines for both poverty and literacy head counts, when compared to the results obtained with the inclusion of all predictors shown in Figure 3. In the estimation of poverty head counts (left-most figures), the fraction of *municipios* for which coverage is below 50 percent increases from 12 to 21 percent in Chiapas, from 11 to 15 percent in Oaxaca and from 11 to 29 percent in Veracruz. In the prediction of literacy head counts, the increase is from 11 to 17 percent in Chiapas, from 14 to 26 in Oaxaca and from 9 to 23 percent in Veracruz. The three right-most graphs in Figure 4 show instead that confidence intervals for the sanitation head counts are somewhat better than when all

predictors are included, and in a substantial fraction of cases they are conservative. This is especially true for Chiapas, where coverage is larger than 0.95 in approximately two-thirds of the *municipios*. This finding, however, is the result of very large estimated MSEs, which reflect relatively large values of the covariance component in expression (7). In this state, the mean value of the root MSE is 0.17, so that on average the width of the confidence intervals is 0.69.

In Figure 5 we display the distribution of coverage rates estimated using a second selection criterion. In this case, the model selection proceeds by first estimating the most inclusive model, and then dropping from the prediction exercise all regressors not significant at the 10 percent level. We refer to this second t-based criterion as “tC”. The coverage rates in Figure 5 show that, despite this criterion leading to more parsimonious models, the distributions of coverage rates across *municipios* remain relatively similar to those estimated when all predictors are included, although in most cases the latter still performs better.²⁵

Overall, these results suggest that the main cause for incorrect coverage is the presence of area heterogeneity, and not the use of an inappropriate model in the first-stage. On the other hand, the validation described here has limitations. First, the population used throughout the simulation is not a real census, but rather it is a census extract inflated to the size of the census by using sampling weights. Second, the income measures included in the census were measured using a very short questionnaire section, so that they are likely to be less accurate than estimates of income or expenditure that could be obtained using purposely designed household survey such as, for instance, the Living Standard Measurement Surveys, the Indonesian Family Life Surveys or the Mexican Family Life Surveys. Third, the census extract does not include geographical identifiers below the *municipio*, so that our mapping exercise adopts the area itself as the primary stage unit. In reality, a researcher with access to a complete census may also have access to identifiers for the smaller census enumeration areas (EAs), which usually include around 100 households only. This allows the use of EA-specific (rather than area-specific) household means as predictors in the first-stage model, which may increase

²⁵As a referee suggested, there may also be useful information in the comparison of coverage rates achieved by different selection models for non-poverty indicators. That is, if it turns out that a given criterion always outperforms all others, a researcher may want to use such criterion when constructing a poverty map. The choice of an optimal algorithm for model selection is however beyond the scope of this paper, so we do not pursue this idea further. Note also that in this empirical setting no such clear ranking exists. For instance, the statistics in columns 1-2, 4-5 and 7-8 of Table 4 show that the most inclusive model performs best for literacy in all states, while for the sanitation index BIC achieves the best coverage rates in Oaxaca and Veracruz but not in Chiapas, where no selection model dominates. The results in the alternative quasi-validation exercise discussed later in the paper are similarly inconclusive, see Section 4.2.

the explanatory power of the predictors and reduce the extent of spatial correlation of the first-stage residuals.²⁶

4.2 A Quasi-validation Using Data from ENCASEH

To address the shortcomings described above, we perform one further analogous validation exercise using an alternative data source, the *Encuesta de Características Socioeconómicas de los Hogares* (ENCASEH), which represent a complete enumeration of households residing in communities deemed eligible to participate to the well-known Mexican welfare program PROGRESA . This strategy is largely inspired by the similar validation exercise in [Demombynes et al. \(2008\)](#). As we will explain, however, our approach is different along fundamental dimensions which also likely explain the different results.

In November 1997, data on a broad spectrum of socio-economic and demographic indicators were collected from a census of 24,077 households, with a total of about 125,000 individuals. These households resided in a total of 506 communities (*localidades*) spread over the seven states of Guerrero, Hidalgo, Michoacán, Puebla, Querétaro, San Luis Potosí and Veracruz. Income information was collected as well, but not unlike in the 2000 Mexican census, the income module was very simple, leading to income measures of dubious reliability, as also suggested by the large number of respondents reporting no income at all. In order to construct poverty measures, we use instead expenditure data collected using a fairly detailed consumption module in March 1998, before the implementation of PROGRESA. Expenditure and other socio-economic information were then collected at two different points in time, but the time gap is very limited. The household characteristics used in the analysis are likely to be very persistent over such a short time frame, so that the presence of a gap between the two data sources should be of little concern. The consumption module was not as comprehensive as other household surveys such as the World Bank Living Standard Measurement Surveys, but it was still fairly detailed, with information on cash expenditure on about 50 item categories including food and beverages, transportation, alcohol and tobacco, expenditure for personal health and hygiene, clothing and footwear.

²⁶Note, however, that in this empirical application the intra-area correlation of first-stage residuals is relatively low, so that the coverage rates are unlikely to be simply the result of the high level of aggregation of the clusters. For instance, when we predict poverty status with logit using all predictors in Table 3 and the whole pseudo-census, the resulting intra-area correlation of the residuals is 8.5 percent in Veracruz, 5.3 percent in Chiapas and 4.5 percent in Oaxaca. Recall also that the Monte Carlo results in Section 3.1 as well as those in TD show that incorrect coverage can arise even in situations where the intra-cluster correlation is very small.

Using information on total household expenditure, we construct a measure of total weekly expenditure per head. The other household characteristics used in the analysis are listed in Table 5. The validation exercise proceeds as before, that is, we draw repeatedly a large number of synthetic surveys from the census, we estimate next head counts and the corresponding MSEs for a series of areas and finally we evaluate coverage rates for each area. Demombynes et al. (2008) use the same data to perform a similar validation (albeit with only ten replications) and find that coverage rates produced using the simulation-based ELL estimator are approximately correct. However, in their simulation areas are constructed by aggregating randomly groups of *localidades*, therefore biasing the results *by construction* in favor of finding area homogeneity. To avoid this problem, we choose to use the seven *states* as areas, while using *localidades* as clusters. In the ENCASEH data set, states include between 1,325 (Quéretaro) and 5,915 households (Veracruz), while *localidades* include a very small number of households, ranging from 2 to 236, with a mean of 48 and a median of 39. Our choice of areas and clusters allows us to overcome one limitation of our validation exercise using the 2000 Mexico, that is, the large size of the clusters and their coincidence with the areas. However, it should be kept in mind that this approach still does not replicate perfectly the typical poverty mapping exercise. In a typical application, both the first and the second stage of the imputation procedure are completed using data from a region which should be kept as small as allowed by the survey sample size. For instance, like in our previous application with 2000 census data, the first stage model would be estimated perhaps using survey data from a Mexican state, while small areas would be represented by state subdivisions such as *municipios*. This may reduce concerns about area heterogeneity, if there are reasons to expect projection coefficients to be more stable within a state than across states. In our case, the first stage is estimated using data from seven different states, while each state is used as a separate area. On the other hand, this concern is significantly limited here by the observation—also noted by Demombynes et al. 2008—that the communities included in ENCASEH were purposely selected to be relatively homogeneous, as they were all poor rural communities eligible for inclusions in PROGRESA.

The validation exercise proceeds in a fashion similar to the previously described ones. In each replication we draw a random sample of 10 households from each of 50 randomly selected *localidades*, so that each “survey” has sample size equal to 500. We exclude from the estimation observations with missing values and households who live in *localidades* with fewer than 10 households. The final population includes 21,406 households from 498 *localidades*. We then construct maps of poverty as well as of four other different indicators. Poverty head counts are calculated assuming that the poverty line is equal to the median of the distribution of weekly expenditure per head over the population,

which we estimate to be equal to 34 Pesos. The other indicators are the proportion of individuals who live in a household with a literate household head, that is, a head who can read and write in at least one language (“literacy” head count), or who live in a dwelling with a latrine (“sanitation”) or where the head worked in the previous week (“employment”).

The state-specific head counts and their corresponding confidence intervals are again estimated using the projection estimator. Given that, unlike the previous validation with data from the 2000 Mexico census, in this case the primary stage units are small subsets of the areas, we estimate MSEs using the expression in (7), while the standard errors of the first stage BLP are clustered at the *localidad* level. In each replication, predictors for poverty counts are chosen from the list in Table 5 as well as from their means at the *localidad* level. Recall that this was not possible in the previous validation, because the 2000 census did not include identifiers below the *municipio* level. In estimating literacy counts we drop all predictors (including the location means) related to schooling or language, while we drop the predictors related to in-house sanitation facilities when estimating sanitation head counts, and all work-related information when estimating employment counts. Overall, there are about 60 predictors and given their large number we select predictors using either the BIC or the tC selection models described in Section 4.1.1. Coverage rates of 95 percent confidence intervals are finally calculated using 250 Monte Carlo replications.

The results of the validation exercise are shown in Table 6. Coverage rates for the poverty head counts in each of the seven states are below the nominal 95 percent, although in almost all cases they remain close to or above 70 percent. When regressors are chosen using BIC (panel A in the table), coverage is actually close to 90 percent in the three states of Hidalgo, Puebla and Veracruz. With model selection based on BIC, we estimate that the seven coverage rates would range from 88 to 98 percent if all root-MSEs were increased by 50 percent, while they would all be approximately one if they were doubled. The outcome is similar when model selection relies on tC (panel B). In this case scaling up the size of the confidence intervals by a factor of 1.5 lead to coverage rates ranging from 0.81 to 0.92, while doubling the size leads to rates ranging from 91 to 98 percent.

Coverage rates for the three non-poverty indices confirm that this kind of exercise is informative about the reliability of poverty mapping estimates. When we use BIC, out of 21 coverage rates, two-thirds are above 65 percent, although in four cases coverage remains below 25 percent. Even for these indicators, with the exception of the sanitation indicator in Qu er etar and all indicators in the state of Guerrero, doubling the estimated MSEs would lead to approximately correct or even conservative confidence intervals, with coverage ranging from 89.6 to 99.6 percent. Note, however, that

in all but four cases coverage rates for literacy, sanitation and employment indicators remain below the corresponding state-specific coverage for poverty, although in several instances the differences are relatively small. The results are overall similar when the model is chosen based on tC, with the main difference being perhaps a considerable improvement in the still low coverage for non-poverty indices in the state of Guerrero. Note, however, that (like in the previous quasi-validation with data from Census 2000) a clear ranking between BIC and tC in terms of performance cannot be found. The tC selection model leads to higher coverage rates than BIC in seven of 21 cases. In particular, tC performs relatively well with the literacy index, where coverage is actually closer to 0.95 than for the corresponding poverty count in four of seven cases. When we look at the sanitation and employment index, BIC performs instead better than tC in most cases.

5 Caveats and Conclusions

Poverty mapping methodologies which are now routinely used require a substantial degree of geographic homogeneity in the relationship between income and its predictors. However, heterogeneity in local unobserved factors such as relative prices and returns is likely to generate heterogeneity in such relationships, at least to some extent. The purpose of this paper is to argue that useful although indirect and suggestive evidence on the extent of area heterogeneity is readily available in virtually any census. Such indirect evidence is provided by non-monetary indicators such as, for instance, literacy, asset ownership, work status or access to sanitation. These variables are routinely recorded in a census, and validation exercises like those described in this paper can then be completed to gauge the extent of area heterogeneity. We argue that the same factors which may generate incorrect inference in poverty mapping will also likely generate heterogeneity, and hence incorrect inference, in the estimation of maps for non-monetary indicators such as those mentioned above.

We have described a simple model which formalizes this intuition, showing that heterogeneity in relative prices will generate not only area heterogeneity in the projection of (log) income on a list of predictors, but also in the projection of the predictors on the remaining ones. The associated Monte Carlo experiment confirms that area heterogeneity causes both the simulation-based estimator of [Elbers et al. \(2003\)](#) and the projection-based estimator of [Tarozzi and Deaton \(2009\)](#) to produce incorrect inference when heterogeneity in relative prices exist. Importantly, incorrect coverage rates of confidence intervals constructed for poverty maps is also reflected in incorrect coverage for maps based on the predictors.

Finally, we have evaluated empirically the merit of our argument using two alternative data sets from Mexico. In both cases, we find that poverty maps produce useful information, but coverage rates remain usually below the nominal level of the confidence intervals, therefore producing incorrect inference. Consistently with the main argument put forth in the paper, we also find that similar results hold for other indicators, such as literacy or access to sanitation, which are usually available in a census. On the other hand, we have emphasized that none of the two simulation exercises replicates perfectly the kind of validation exercise that would be possible using a real census, to which we do not have access.

One further *caveat* is that even this form of validation exercise will be of little help if the census and the survey are completed in different years. This may be a common occurrence, because while census data are usually decadal, surveys are in most cases carried out at shorter time intervals. But in such situation, much more than area homogeneity is required for poverty mapping to produce valid inference. Namely, one also need stability *over time* of the distribution of the predictors, because such variables are only recorded for the whole population in the census. Such assumption is clearly problematic, especially in the context of fast-growing countries where the underlying structure of the economy is likely to change rapidly. So, even if a validation suggested the presence of area homogeneity in a cross-section, such evidence would be silent about the plausibility of the identifying assumptions required by poverty mapping when census and survey data have been collected in different time periods.

Finally, we emphasize again that, from a statistical point of view, the good or poor performance of a mapping methodology in constructing, say, literacy maps should *not* be used to *imply* that poverty mapping will be equally successful or unsuccessful in producing reliable inference. However, we argue that such similarity in performance should be expected. This observation, together with the relative simplicity of the validation exercises described in this paper, suggest that analogous validation exercises should be performed routinely by researchers or statistical agency who wish to proceed with a poverty mapping exercise, regardless of whether one expects the assumptions underlying poverty mapping to fail or to hold.

APPENDIX

Consider first the BLP of $\ln x_1$ given $\ln x_2$ in the larger region, that is, not conditional on a given area. Using standard formulae, OLS estimates of the intercept and the slope will have the following usual forms:

$$\hat{\beta}_1 = \frac{\widehat{cov}(\ln x_1, \ln x_2)}{\widehat{var}(\ln x_2)}, \quad \hat{\beta}_0 = \hat{E}(\ln x_1) - \hat{\beta}_1 \hat{E}(\ln x_2),$$

which under standard regularity conditions will have the following probability limits:

$$p \lim \hat{\beta}_1 = \frac{cov(\ln x_1, \ln x_2)}{var(\ln x_2)} \equiv \beta_1, \quad p \lim \hat{\beta}_0 = E(\ln x_1) - \beta_1 E(\ln x_2) \equiv \beta_0.$$

Consider first the probability limit of the slope. Rewriting each element in terms $\ln Y$ and unobserved components, some simple manipulations show that

$$\begin{aligned} \beta_1 &= 1 + \frac{cov(\eta_{2c} + \varepsilon_{2i} - \phi_2^a, \varepsilon_{1i} + \eta_{1c} - \phi_1^a - \varepsilon_{2i} - \eta_{2c} + \phi_2^a)}{var(\ln Y_i + \eta_{2c} + \varepsilon_{2i} - \phi_2^a)} \\ &= 1 - \frac{\sigma_{\phi_2}^2 + \sigma_{\varepsilon_2}^2 + \sigma_{\eta_2}^2}{\sigma_Y^2 + \sigma_{\phi_2}^2 + \sigma_{\varepsilon_2}^2 + \sigma_{\eta_2}^2} = \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_{\phi_2}^2 + \sigma_{\varepsilon_2}^2 + \sigma_{\eta_2}^2}. \end{aligned}$$

Similarly, and using the fact that all unobserved components have zero mean, we have

$$\begin{aligned} \beta_0 &= \ln\left(\frac{p_2 \alpha_1}{p_1 \alpha_2}\right) + E(\ln x_{i2}) - \beta_1 E(\ln x_{i2}) \\ &= \ln\left(\frac{p_2 \alpha_1}{p_1 \alpha_2}\right) + (1 - \beta_1) E(\ln x_{i2}) \\ &= \ln\left(\frac{p_2 \alpha_1}{p_1 \alpha_2}\right) + (1 - \beta_1) (\mu_{\ln Y} + \ln \alpha_2 - \ln p_2), \end{aligned}$$

which completes the proof of (14).

If we move now to the BLP for a given area, the components ϕ_1^a and ϕ_2^a are no longer random but have to be treated as constant. Hence

$$\begin{aligned} p \lim \hat{\beta}_1^a \equiv \beta_1^a &= p \lim \frac{\widehat{cov}(\ln x_1, \ln x_2 | a)}{\widehat{var}(\ln x_2 | a)} \\ &= 1 + \frac{cov(\eta_{2c} + \varepsilon_{2i}, \varepsilon_{1i} + \eta_{1c} - \varepsilon_{2i} - \eta_{2c})}{var(\ln Y_i + \eta_{2c} + \varepsilon_{2i})} \\ &= 1 - \frac{\sigma_{\varepsilon_2}^2 + \sigma_{\eta_2}^2}{\sigma_Y^2 + \sigma_{\varepsilon_2}^2 + \sigma_{\eta_2}^2} = \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_{\varepsilon_2}^2 + \sigma_{\eta_2}^2}, \end{aligned}$$

and

$$\begin{aligned} p \lim \hat{\beta}_0 \equiv \beta_0^a &= p \lim \left(\hat{E}(\ln x_1 | a) - \hat{\beta}_1^a \hat{E}(\ln x_2 | a) \right) \\ &= \ln\left(\frac{p_2 \alpha_1}{p_1 \alpha_2}\right) - \phi_1^a + \phi_2^a + E(\ln x_{i2}|a) - \beta_1^a E(\ln x_{i2}|a) \\ &= \ln\left(\frac{p_2 \alpha_1}{p_1 \alpha_2}\right) - \phi_1^a + \phi_2^a + (1 - \beta_1^a) (\mu_{\ln Y} + \ln \alpha_2 - \ln p_2 - \phi_2^a) \\ &= \ln\left(\frac{p_2 \alpha_1}{p_1 \alpha_2}\right) - \phi_1^a + \beta_1^a \phi_2^a + (1 - \beta_1^a) (\mu_{\ln Y} + \ln \alpha_2 - \ln p_2), \end{aligned}$$

which completes the proof of (15).

References

- Carroll, R., D. Ruppert, and L. Stefanski (1995). *Measurement Error in Non-Linear Models*. Chapman and Hall.
- Chen, X., H. Hong, and A. Tarozzi (2008). Semiparametric efficiency in GMM models with auxiliary data. *Annals of Statistics* 36(2), 808–843.
- Deaton, A. and M. Grosh (2000). Consumption. In M. Grosh and P. Glewwe (Eds.), *Designing household survey questionnaires for developing countries: lessons from 15 years of the Living Standards Measurement Study*, Volume 1, Chapter 5, pp. 91–133. Oxford University Press for the World Bank.
- Demombynes, G., C. Elbers, J. O. Lanjouw, and P. Lanjouw (2008). How good a map? Putting small area estimation to the test. *Rivista Internazionale di Scienze Sociali* (4), 465–493.
- Elbers, C., J. Lanjouw, and P. Lanjouw (2003). Micro-level estimation of poverty and inequality. *Econometrica* 71(1), 355–364.
- Elbers, C., J. O. Lanjouw, and P. Lanjouw (2002). Micro-level estimation of welfare. Policy Research Working Paper 2911, The World Bank, Washington, DC.
- Elbers, C., P. Lanjouw, and P. G. Leite (2008). Brazil within Brazil: Testing the poverty map methodology in Minas Gerais. World Bank Policy Research Working Paper 4513.
- Fujii, T. (2007). Micro-level estimation of child undernutrition indicators and its application to targeting in Cambodia. Unpublished Manuscript, Singapore Management University.
- Heckman, J., R. LaLonde, and J. Smith (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics, Vol. 3A*. Amsterdam, The Netherlands: Elsevier Science.
- Heston, A., R. Summers, and B. Aten (2006). Penn World Table version 6.2. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania. http://pwt.econ.upenn.edu/php_site/pwt_index.php.
- Hirano, K., G. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71(4), 1161–1189.
- Little, R. and D. Rubin (2002). *Statistical Analysis with Missing Data* (Second ed.). New York: John Wiley & Sons.
- Minnesota Population Center (2007). Integrated Public Use Microdata Series International: Version 3.0. Minneapolis: University of Minnesota.
- National Research Council (2000). *Small-area income and poverty estimates. Priorities for 2000 and beyond*. Washington, DC.: National Academy Press.
- Pfeffermann, D. (2002). Small area estimation: new developments and directions. *International Statistical Review* 70(1), 125–143.
- Rao, J. N. K. (2003a). *Small Area Estimation*. New York: Wiley.
- Rao, J. N. K. (2003b). Some new development in Small Area Estimation. *Journal XXX* 2(2), 145–169.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of The American Statistical Association* 91, 473–489.
- Ruggles, S. and M. Sobek (1997). Integrated public use microdata series: Version 2.0. Historical Census Projects, University of Minnesota. <http://www.ipums.umn.edu>.

- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6(2), 461–464.
- Tarozzi, A. and A. Deaton (2009). Using census and survey data to estimate poverty and inequality for small areas. *Review of Economics and Statistics* 91(4), 773–792.
- Todd, P. E. (2007). Evaluating social programs with endogenous program placement and selection of the treated. In T. P. Schultz and J. Strauss (Eds.), *Handbook of Development Economics*, Volume IV, Chapter 60. Amsterdam: Elsevier Science.

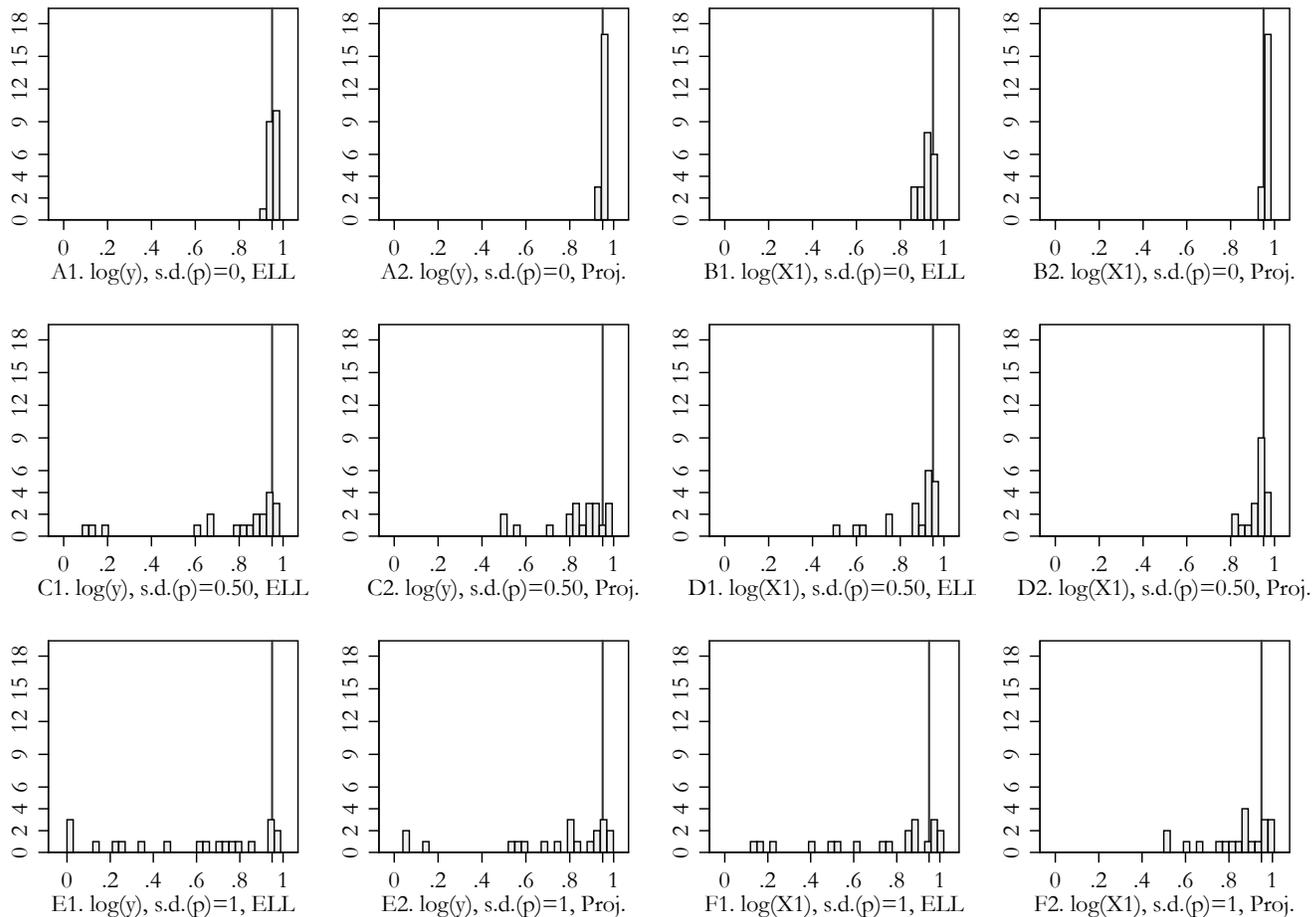


Figure 1: Histograms of coverage rates, based on 200 Monte Carlo replications. Each graph shows simulated coverage rates for 20 synthetic areas of 10,000 individuals grouped in 100 clusters of 100 units. The vertical axis indicates the number of synthetic areas in each bin. The labels below each graph indicate whether the coverage rates refer to the estimation of income head counts H_y^a (“log(y)”) or “good 1” head counts $H_{x_1}^a$ (“log(x1)”) and also indicate whether they have been estimated using the simulation-based (“ELL”) or the projection-based (“Proj.”) method. The vertical line is drawn at the nominal coverage level, that is, 0.95.

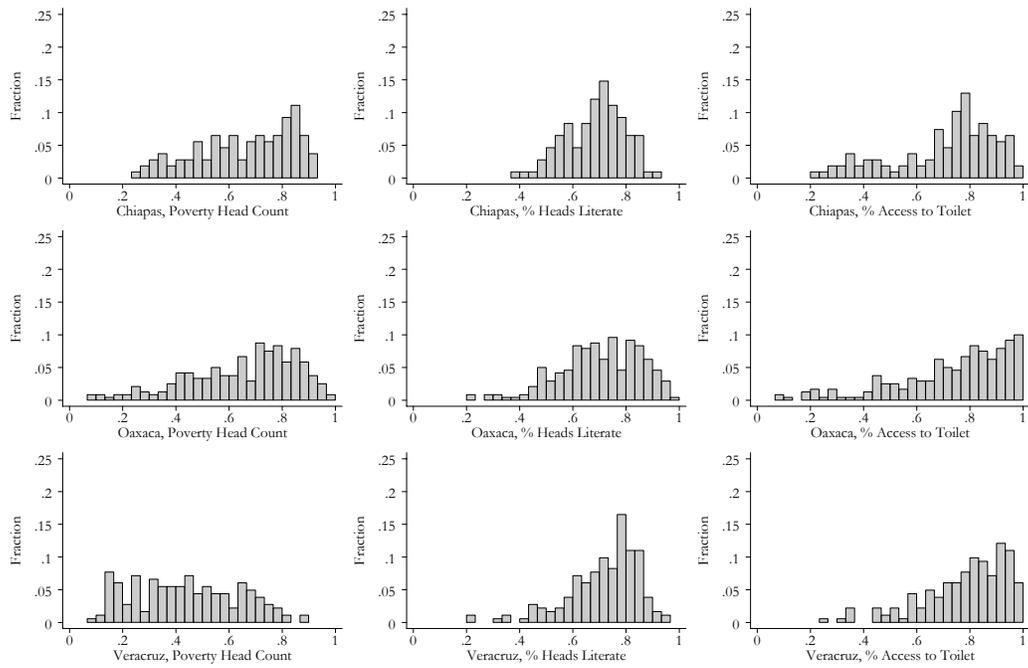


Figure 2: Distributions of *municipio*-specific statistics, by state. Source: Author's calculations from IPUMS Mexico Census 2000.

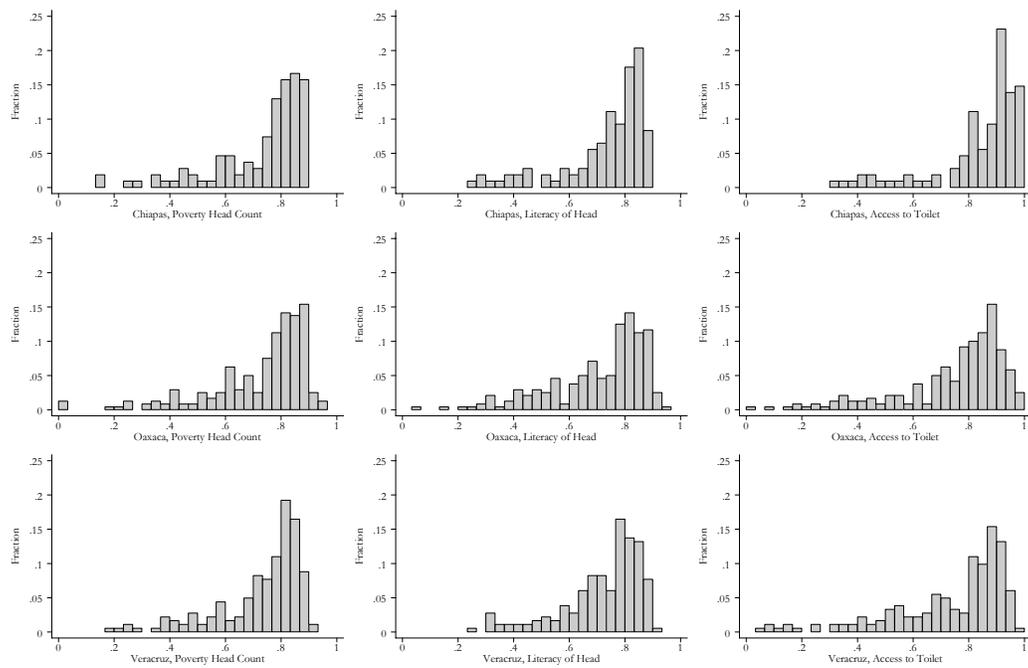


Figure 3: Histograms of coverage rates, based on 250 Monte Carlo replications. Data Source: IPUMS Mexico Census 2000.

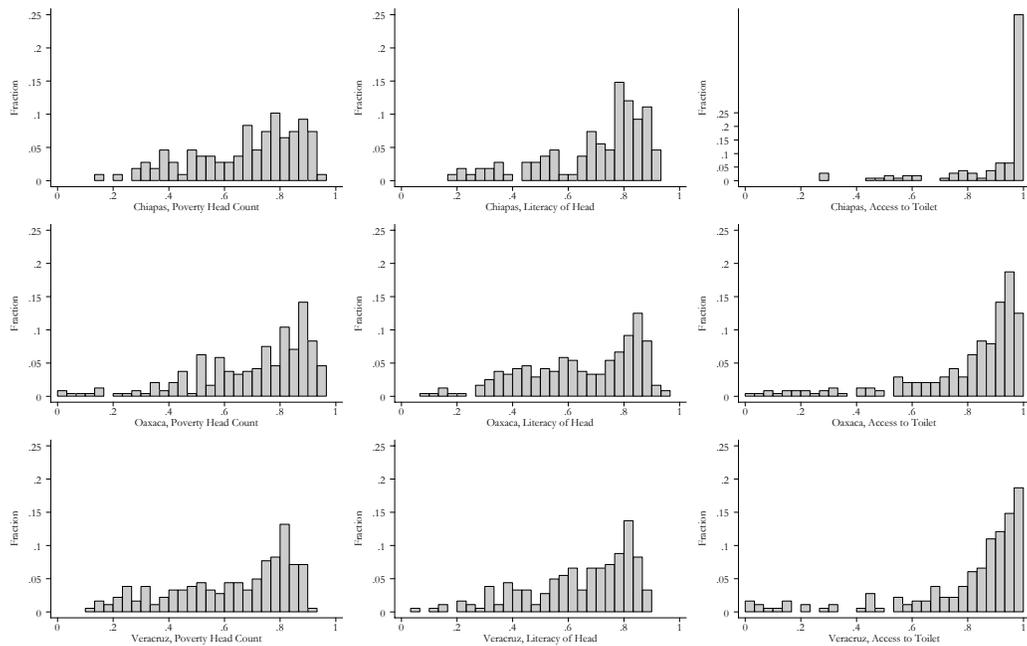


Figure 4: Histograms of coverage rates, based on 250 Monte Carlo replications, with first-stage model selected using a Bayesian Information Criterion. Note the different scale of the vertical axis in the top-right graph. Data Source: IPUMS Mexico Census 2000.

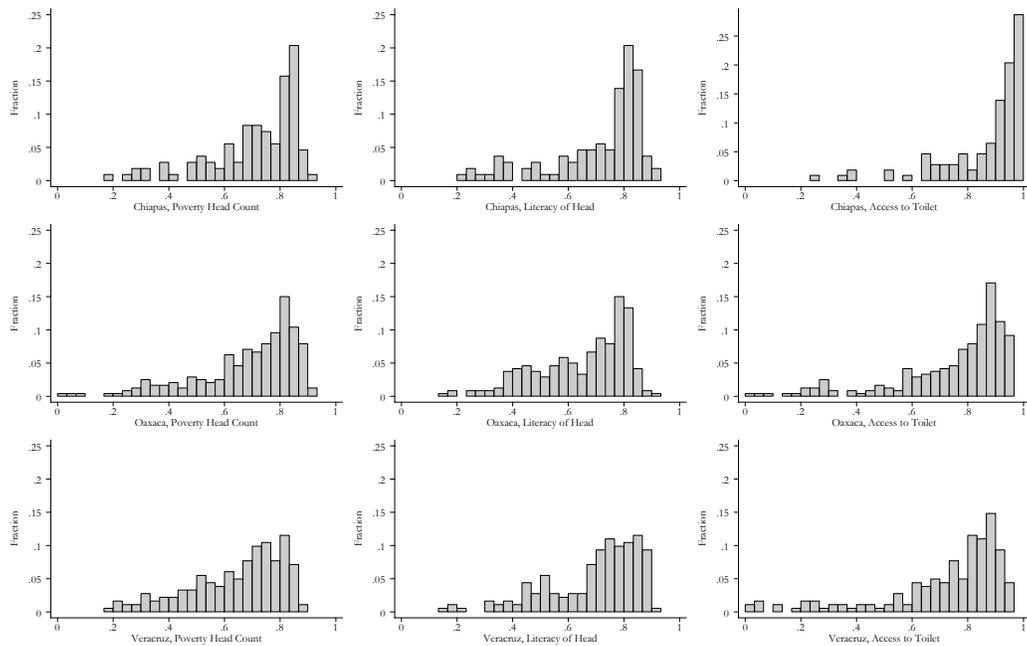


Figure 5: Histograms of coverage rates, based on 250 Monte Carlo replications, with first-stage model selected by including only predictors significant at the 10 percent level. Note the different scale of the vertical axis in the top-right graph. Data Source: IPUMS Mexico Census 2000.

Table 1: Summary of Monte Carlo Results

	$\sigma_p = 0$		$\sigma_p = 0.50$		$\sigma_p = 1$	
	$p_1^a/p_2^a = 1$		$p_1^a/p_2^a \in [0.979, 1.017]$		$p_1^a/p_2^a \in [0.959, 1.034]$	
	(1)	(2)	(3)	(4)	(5)	(6)
	H_y	H_{x_1}	H_y	H_{x_1}	H_y	H_{x_1}
(1) R^2	0.8049	0.4537	0.8029	0.4538	0.7973	0.4533
(2) Intra-cluster correlation of residual	0.0310	0.0326	0.0406	0.0346	0.0662	0.0414
(3) Area-specific Head Count (min)	0.4873	0.5017	0.4873	0.4739	0.4873	0.4484
(4) Area-specific Head Count (max)	0.5059	0.5230	0.5059	0.5605	0.5059	0.6034
(5) Width of Confidence Interval (ELL)	0.0249	0.0380	0.0254	0.0382	0.0276	0.0389
(6) Width of Confidence Interval (Proj.)	0.0350	0.0510	0.0371	0.0524	0.0424	0.0551
(7) Average absolute bias (ELL)	0.0102	0.0168	0.0170	0.0203	0.0293	0.0284
(8) Average absolute bias (Proj.)	0.0132	0.0179	0.0197	0.0218	0.0307	0.0294
(9) Coverage of 95% C.I. (ELL, mean)	0.9532	0.9210	0.7575	0.8532	0.5660	0.7145
(10) Coverage of 95% C.I. (ELL, min)	0.8950	0.8500	0.0850	0.4950	0.0000	0.1200
(11) Coverage of 95% C.I. (ELL, max)	0.9750	0.9600	0.9750	0.9500	0.9650	0.9900
(12) Coverage of 95% C.I. (Proj., mean)	0.9230	0.9150	0.7635	0.8437	0.5917	0.7250
(13) Coverage of 95% C.I. (Proj., min)	0.8750	0.8550	0.3300	0.6900	0.0000	0.3150
(14) Coverage of 95% C.I. (Proj., max)	0.9400	0.9400	0.9500	0.9050	0.9550	0.9650

Notes: figures are summary statistics that describe the results of 200 Monte Carlo replications. The column header indicates if the estimates refer to the income head counts H_y^a or to the “good 1” head counts $H_{x_1}^a$. The R^2 s in row 1 refer to the models in equations (10) and (11) in the text, estimated using the complete synthetic census. The intra-cluster correlations in row 2 are estimated using the residuals from the same regressions used to calculate the R^2 s and are estimated as the fraction of the total residual variance explained by cluster fixed effects. The means in rows 5 and 6 are calculated over all area-specific simulation results (that is, over $20 \times 200 = 4000$ estimates). All other statistics in rows 3 to 14 refer to the distribution over areas.

Table 2: Mexico 2000 Pseudo-census: Summary Statistics

	Chiapas	Oaxaca	Veracruz
Pseudo-census hhs. population size	395078	318701	612942
Pseudo-census individual population size	2035661	1528763	2793627
no. of <i>municipios</i>	108	240	182
Mean no. of hhs. in a pseudo-census <i>municipio</i>	3658	1328	3367
Poverty Head Count Ratio (Pov. line 200 Pesos day/person)	.66	.65	.44
Literacy Head Count	.68	.70	.72
Sanitation Head Count	.71	.74	.78

Source: IPUMS Mexico Census 2000. All figures are calculated using the “pseudo-census” generated as described in the text by using sampling weights and including only *municipios* with at least 500 households in the complete census. These *municipios* account for 99 percent of the population in Chiapas (which includes a total of 118 *municipios*), 81 percent in Oaxaca (562 *municipios*) and 99 percent in Veracruz (210). The “literacy” and the “sanitation” head counts are defined as the fraction of the population living in households with, respectively, a literate household head and access to a toilet.

Table 3: Mexico 2000 Pseudo-census: Variables used as predictors

Head is literate
 Access to electricity
 Owns refrigerator
 Owns TV
 Owns radio
 Number of rooms
 Access to toilet within dwelling
 Age of head
 Head belongs to indigenous group
 Main cooking fuel is wood
 Dwelling has dirt floor
 Primary dwelling material is brick/stone
 Primary roof material is masonry/concrete/tile
 Speaks only indigenous language
 Speaks both indigenous language and Spanish
 Head working previous week
 Head works in Agriculture/Fishery/Forestry/Mining
 # household members ages 0-12 (and its squared)
 # household members older than 65 (and its squared)
 # male members ages 13-65 (and its squared)
 # female members age 13-65 (and its squared)
 Head is a woman

municipio-level means:

Head is literate
 Years of schooling of head
 Access to electricity
 Owns radio
 Access to toilet within dwelling
 Dwelling has dirt floor
 Primary dwelling material is brick/stone
 Primary roof material is masonry/concrete/tile
 Speaks only indigenous language
 Head works in Agriculture/Fishery/Forestry/Mining

Source: IPUMS Mexico Census 2000. List of variables used to predict poverty status. For the predictors of literacy or access to toilet of household head see text.

Table 4: Summary of Results: Mexico Census 2000

	All predictors			BIC			Only significant predictors		
	(1) % covrg < .75	(2) % covrg < .50	(3) C.I. Width	(4) % covrg < .75	(5) % covrg < .50	(6) C.I. Width	(7) % covrg < .75	(8) % covrg < .50	(9) C.I. Width
Chiapas									
Poverty	0.352	0.120	0.329	0.537	0.213	0.314	0.481	0.120	0.361
Literacy	0.361	0.111	0.298	0.444	0.167	0.266	0.407	0.157	0.270
Sanitation	0.176	0.074	0.570	0.148	0.046	0.693	0.185	0.037	0.642
Oaxaca									
Poverty	0.375	0.108	0.375	0.467	0.146	0.399	0.517	0.163	0.388
Literacy	0.438	0.142	0.293	0.579	0.258	0.274	0.608	0.212	0.266
Sanitation	0.342	0.121	0.491	0.271	0.108	0.591	0.346	0.117	0.517
Veracruz									
Poverty	0.385	0.110	0.362	0.615	0.286	0.290	0.659	0.198	0.318
Literacy	0.445	0.093	0.294	0.632	0.225	0.240	0.516	0.148	0.279
Sanitation	0.396	0.132	0.368	0.258	0.121	0.445	0.385	0.137	0.377

Source: Author's calculations with data from IPUMS Mexico Census 2000. The figures show the state-specific fractions of *municipios* for which the coverage rates are below .75 or .50 as well as the average width of the confidence intervals. We calculate each set of three statistics from 250 simulations, separately for each state (Chiapas, Oaxaca or Veracruz) and indicator (poverty, literacy or sanitation) and using three different selection models. The results in columns 1-3 are obtained using all predictors in Table 3. In columns 4-6 we select the predicting model using BIC, while in columns 7-9 the first stage uses only predictors that are individually statistically significant at the 10 percent level in a model where all regressors are included. See main text for additional details.

Table 5: ENCASEH 1997-98: Variables Used as Predictors

Household size (and its square)
Fraction of household members who are males
Household head is female
Age of household head (and its square)
Head speaks only indigenous language
Head speaks both Spanish and an indigenous language
Head can read and write
Head had secundaria schooling (some or complete)
Head has post-Secondaria schooling (some or complete)
Household owns blender
Household owns refrigerator
Household owns gas stove
Household owns water boiler
Household owns radio
Household owns stereo
Household owns TV
Household owns VCR
Household owns washing machine
Household owns fan
Household owns car/truck
no. or room in household (and its square)
Dwelling walls of good quality (brick/concrete etc)
Dwelling roof of good quality (tiles/brick etc)
Latrine in dwelling
Latrine has running water
Access to electricity
Household head worked previous week
Dependency ratio (no. of working members)/(household size)

Source: ENCASEH. The list of available predictors also includes *localidad*-specific means of all variables.

Table 6: ENCASEH 1997-98: Summary of Validation Results

	Poverty		Literacy		Sanitation		Employment	
	C.I. width	Coverage						
(A) - BIC								
Guerrero	0.145	0.828	0.156	0.228	0.235	0.244	0.108	0.180
Hidalgo	0.120	0.872	0.110	0.800	0.172	0.824	0.081	0.908
Michoacá	0.139	0.764	0.112	0.472	0.187	0.668	0.087	0.896
Puebla	0.124	0.874	0.114	0.852	0.167	0.368	0.081	0.728
Quéretar	0.140	0.682	0.122	0.864	0.186	0.080	0.093	0.664
San Luis Potosí	0.123	0.762	0.114	0.572	0.167	0.756	0.086	0.892
Veracruz	0.127	0.874	0.119	0.732	0.178	0.792	0.075	0.792
(B) - tC								
Guerrero	0.189	0.740	0.191	0.468	0.235	0.336	0.164	0.456
Hidalgo	0.116	0.760	0.109	0.792	0.150	0.728	0.095	0.616
Michoacá	0.140	0.724	0.130	0.752	0.176	0.640	0.102	0.700
Puebla	0.119	0.780	0.113	0.796	0.150	0.352	0.093	0.688
Quéretar	0.143	0.624	0.133	0.728	0.181	0.140	0.112	0.644
San Luis Potosí	0.135	0.704	0.120	0.604	0.155	0.660	0.106	0.644
Veracruz	0.131	0.792	0.122	0.736	0.152	0.744	0.082	0.720

Notes: Figures represent the average width of 95 percent confidence intervals estimated using the projection method described in Section 2 and the corresponding coverage rates. All statistics are calculated over 250 simulations. Each replication is completed using a sample of 500 households drawn in the form of 10 households from each of 50 primary stage units (*localidades*). In each draw, the model is selected using BIC (panel A) or tC (panel B). See Section 4.1.1 for details.