Supplemental Appendix

The Missing Poor

Torsten Figueiredo Walter, Niclas Moneke and Ana Radu

A.1 Census enumeration

A.1.1 Census enumeration forms

Figure A1 presents an example of a census enumeration form. It shows the form used for the 2012 Tanzania Population Census. Enumeration forms for most of the other censuses in our sample are available from the IPUMS website.

A.1.2 Enumerator instructions

More household members than pre-printed slots. If the number of household members in a household exceeds the number of pre-printed slots in the household roster, enumerators are instructed to use a second enumeration form. Below, we provide five illustrative examples of such instructions from enumerator manuals:

1. Chile 2002: "If there are more than 6 people in the household, use a second questionnaire."
2. Philippines 2010: "If there are more than eight (8) members, you will need to use additional booklet of CPH Form 2."
3. Russia 2010: "If you find out that the dwelling accommodates more than ten permanent occupants or more than four temporary occupants you will have to rearrange the questionnaires and to complete separate Forms S and P per each dwelling separately."
4. South Africa 1996: "If there are more than nine (9) people, please make use of additional questionnaires."

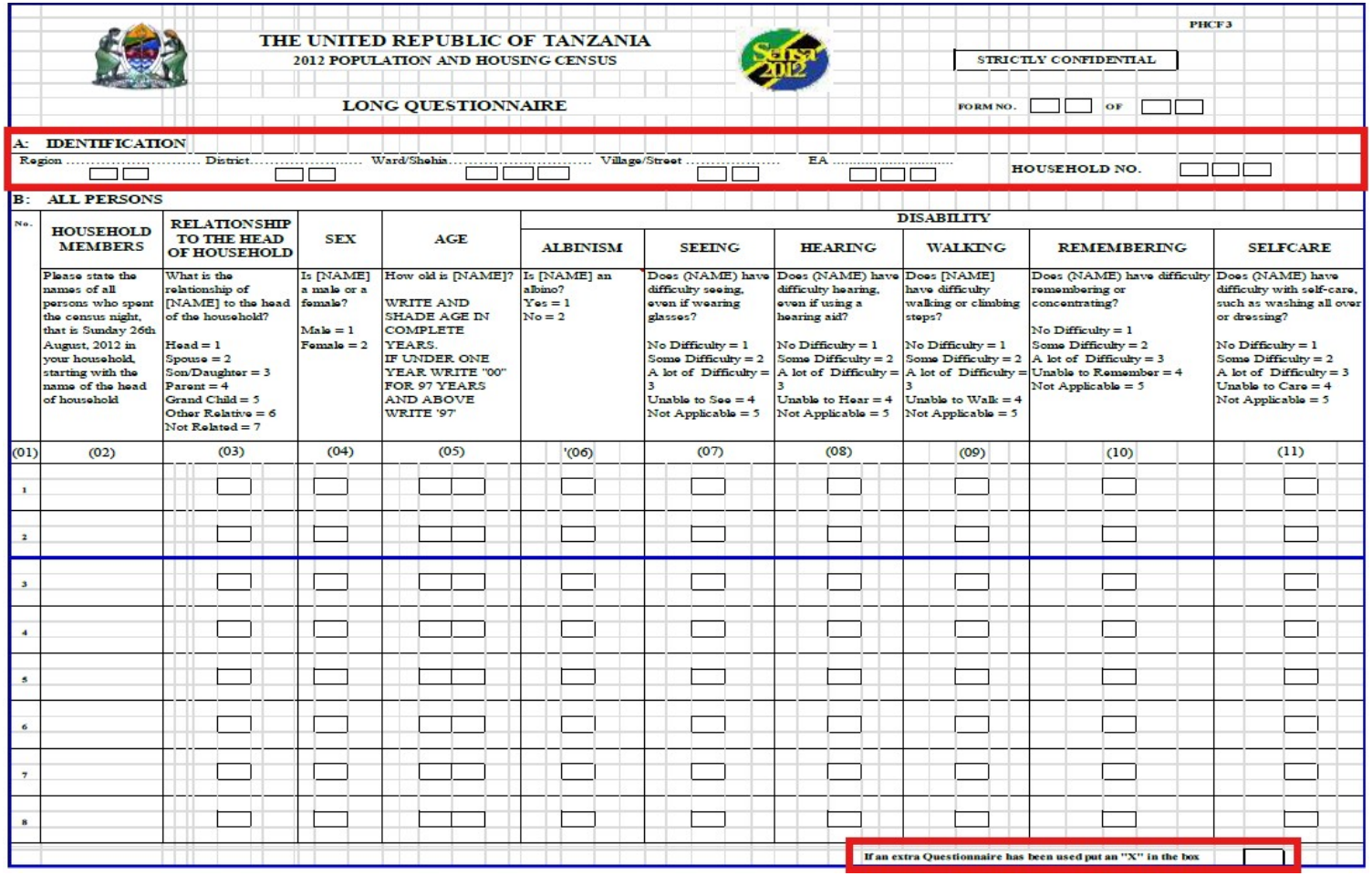
5. Zambia 2010: "The household listing has enough space for 8 persons. If a household has more than 8 persons, use a continuation questionnaire."

Some manuals provide explicit instructions in case enumerators run out of enumeration forms. For example:

1. Bolivia 2012: "If the number of census forms is not sufficient for all the dwellings in the enumeration area, inform your supervisor as soon as possible."

2. Zambia 2000: "Should you run out of questionnaires, quickly inform your supervisor who should immediately supply you with more."

Other manuals point to the possibility of enumerators borrowing spare forms from each other: "It's possible that some of your colleagues might run out of forms. They could then use the ones you have left over." (Colombia 1993)



THE UNITED REPUBLIC OF TANZANIA
2012 POPULATION AND HOUSING CENSUS

PHCF3

STRICTLY CONFIDENTIAL

LONG QUESTIONNAIRE

FORM NO. ☐☐ OF ☐☐

A: IDENTIFICATION

Region ☐☐ District ☐☐ Ward/Shehia ☐☐☐ Village/Street ☐☐ EA ☐☐☐ HOUSEHOLD NO. ☐☐☐

B: ALL PERSONS

No.	HOUSEHOLD MEMBERS	RELATIONSHIP TO THE HEAD OF HOUSEHOLD	SEX	AGE	DISABILITY					
					ALBINISM	SEEING	HEARING	WALKING	REMEMBERING	SELFCARE
	Please state the names of all persons who spent the census night, that is Sunday 26th August, 2012 in your household, starting with the name of the head of household	What is the relationship of [NAME] to the head of the household? Head = 1 Spouse = 2 Son/Daughter = 3 Parent = 4 Grand Child = 5 Other Relative = 6 Not Related = 7	Is [NAME] a male or a female? Male = 1 Female = 2	How old is [NAME]? WRITE AND SHADE AGE IN COMPLETE YEARS. IF UNDER ONE YEAR WRITE "00" FOR 97 YEARS AND ABOVE WRITE '97'	Is [NAME] an albino? Yes = 1 No = 2	Does (NAME) have difficulty seeing, even if wearing glasses? No Difficulty = 1 Some Difficulty = 2 A lot of Difficulty = 3 Unable to See = 4 Not Applicable = 5	Does (NAME) have difficulty hearing, even if using a hearing aid? No Difficulty = 1 Some Difficulty = 2 A lot of Difficulty = 3 Unable to Hear = 4 Not Applicable = 5	Does [NAME] have difficulty walking or climbing steps? No Difficulty = 1 Some Difficulty = 2 A lot of Difficulty = 3 Unable to Walk = 4 Not Applicable = 5	Does (NAME) have difficulty remembering or concentrating? No Difficulty = 1 Some Difficulty = 2 A lot of Difficulty = 3 Unable to Remember = 4 Not Applicable = 5	Does (NAME) have difficulty with self-care, such as washing all over or dressing? No Difficulty = 1 Some Difficulty = 2 A lot of Difficulty = 3 Unable to Care = 4 Not Applicable = 5
(01)	(02)	(03)	(04)	(05)	'(06)	(07)	(08)	(09)	(10)	(11)
1		☐	☐	☐☐	☐	☐	☐	☐	☐	☐
2		☐	☐	☐☐	☐	☐	☐	☐	☐	☐
3		☐	☐	☐☐	☐	☐	☐	☐	☐	☐
4		☐	☐	☐☐	☐	☐	☐	☐	☐	☐
5		☐	☐	☐☐	☐	☐	☐	☐	☐	☐
6		☐	☐	☐☐	☐	☐	☐	☐	☐	☐
7		☐	☐	☐☐	☐	☐	☐	☐	☐	☐
8		☐	☐	☐☐	☐	☐	☐	☐	☐	☐

If an extra Questionnaire has been used put an "X" in the box ☐

Figure A1: Tanzania Population Census 2012: Enumeration form

This figure shows the enumeration form for the 2012 Tanzanian Population Census. The red box on the bottom highlights the checkbox that the enumerator has to fill if an extra form is used. The red box on the top highlights the household identification field which the enumerator has to fill on every form they use, including any extra form(s) for the same household.

Ordering of household members. Enumerator manuals typically provide instructions on the order in which household members are to be listed. Below, we provide five illustrative examples:

1. Brazil 2000: "In each household, members shall be recorded in the following order: household head, spouse or partner, children and step children (in decreasing order of age), parents and parents-in-law, grandchildren, siblings, other relatives, non-relatives, domestic employees and relatives of domestic employees."

2. Fiji 1996: "Start with the head and his wife and unmarried children, beginning with the eldest and working down to the youngest. Then enter married children and their spouses and children. Then list other relatives and their wives and children who were in the household on census night. Finally list those who are not related to the head or anyone else in the household."

3. Myanmar 2014: "List members of the household by nuclear family, starting with the head and husband/wife and unmarried children, beginning with the eldest and working down to the youngest. Then list the names of each married child with spouse and their children. This must be done for every married child/couple in the household. Next, list the names of other relatives and non-relatives, in that order including visitors."

4. Uganda 1991: "Start with the head and his wife and unmarried children, beginning with the eldest and working down to the youngest. If a man has more than one wife and if all live and eat together, list each wife and her unmarried children in turn. Then enter married children and their spouses and children who spent census night with the household. Then list other relatives and their wives and children who were in the household on census night. Finally list those who are not related to the head or anyone else who spent census night with the household."

5. Ukraine 2001: "Husband and wife are recorded one after one; children (including grown up not married children) are recorded after

the parents; if there are several married couples in the household firstly one married couple and their children are recorded and then another married couple and their children are recorded, etc.; household members that are not in relative or law relations with other household members are recorded ultimately."

A.1.3 Case studies of enumerator contracting

Census enumerations are large data collection exercises that aim to count every person in a given country on a given date. Fieldwork is typically completed within a few days to avoid double-counting as well as undercounting that could result from the mobility of people. To accomplish this, a large number of enumerators is required for a short period of time. These enumerators are commonly recruited through large recruitment drives by national statistical offices, offering short-term contracts with fixed wages. Below, we present examples from three censuses:

1. Chile 2017: The National Statistical Office aimed to recruit 538,000 volunteers for enumeration. On the census day, each enumerator had to enumerate the inhabitants of 16-20 dwellings in urban areas and 6-8 dwellings in rural areas. Enumerators received a map of their census tract as well as census enumeration forms in the morning of the census day, enumerated the assigned dwellings and then returned all the material to their supervisor. Supervisors were to verify the completeness of the enumeration forms. The compensation for enumerators amounted to 15,000 Chilean pesos for the day's work.[1]

2. Kenya 2019: The Kenyan National Bureau of Statistics advertised 135,000 enumerator positions, offering fixed term contract for a period of 16 days. The gross remuneration amounted to 1,500 Kenyan shilling per day, yielding a total of 23,000 Kenyan shilling for the contract period to be paid after completion of the work.

[1] Based on Jiménez (2017).

This period included 7 days of training, 2 days of household listing and 7 days of enumeration.[2]

3. Zimbabwe 2012: The Zimbabwe National Statistics Agency recruited approximately 31,000 enumerators for the 2012 population census. Enumerators were contracted for a period of 10 days with a fixed pay of 90 dollars per day.[3]

Generally speaking, enumerators are typically either paid in installments, with a first payment after a brief training and a second payment upon completion of the enumeration, or entirely paid after the completion of their work to the satisfaction of their supervisor. We are not aware of any explicit rewards for accuracy for enumerators. However, supervisors can ask enumerators to return to households in case of errors or omissions.

[2]Based on Kenya National Bureau of Statistics (2019) and Nzekwe (2019).
[3]Based on The Herald (2012) and The Zimbabwean (2012).

A.2 Census data

IPUMS-International provides access to microdata, typically a 10% random sample, from 354 population censuses conducted between 1960 and today. We do not include historical censuses collected before the World Wars that are also available on IPUMS. In addition, we exclude all surveys available on IPUMS and focus only on censuses. In the end, we use 246 of these censuses for our analysis. We disregard 108 censuses because they are not suitable for our analysis for the following reasons:

1. In 26 datasets, we cannot measure household size. This is either because persons are not organized into households or because the data does not contain all members for each sampled household.

2. 63 censuses are either register-based or self-administered.

3. For 10 censuses the documentation is insufficient to determine the number of pre-printed slots on the household roster. This includes one case for which there is variation in the number of slots on the household roster across different copies of enumeration forms.

4. For 9 censuses, the microdata does not match the listed population. We can separate two types of cases:

   (i) the census lists both the de jure and the de facto population, but the microdata only contains one of these two populations;

   (ii) the census lists present usual members, absent usual members and/or visitors separately, but the microdata does not contain information on residential status.

   For both cases, we cannot observe the relevant population affected by the finite number of pre-printed roster slots in the data.

We complement the resulting database with microdata from eight additional recent population censuses. These were obtained directly from National Statistical Offices as detailed in Table A1 (Departamento Administrativo Nacional de Estadística (DANE), 2023; Oficina Nacional de Estadística (ONE), 2023–2024; Instituto Nacional de

Estadística y Censos (INEC), 2023; Ghana Statistical Service, 2022; National Institute of Statistics of Rwanda (NISR), 2024; Agence Nationale de la Statistique et de la Démographie (ANSD), 2023; National Bureau of Statistics (NBS), 2024; Statistics South Africa, 2024).

The final database includes more than 178 million households from 254 censuses across 82 countries, enumerated between 1960 and 2023 (Figueiredo Walter et al., 2026). As illustrated in Figure A2, these censuses cover all continents, but coverage is lower in North America, Europe and Oceania, where self-administered and register-based censuses are more common. A range of African, Arab and Central Asian countries is not included because census microdata was not accessible.

The number of slots pre-printed on household rosters varies between three and 33 across censuses. Ten slots are most common, observed in one third of censuses. Enumeration forms with eight, six and nine pre-printed slots are also common, as shown in Figure A3.

Our source of GDP per capita data are the Penn World Tables 11.0 (Feenstra et al., 2025).

Country	**Year**	**Source**
COL	2018	Departamento Administrativo Nacional de Estadística; retrieved from Archivo Nacional de Datos on 17/07/2025
DOM	2022	Oficina Nacional de Estadistica; retrieved from El Portal de las Estadisticas Dominicanas on 17/07/2025
ECU	2022	National Institute of Statistics and Censuses; retrieved from CensoEcuador on 17/07/2025
GHA	2021	Ghana Statistical Service; retrieved from GSS Microdata Catalog on 16/07/2025
RWA	2022	National Institute of Statistics; retrieved from NISR Central Data Catalog on 16/07/2025
SEN	2023	Agence Nationale de la Statistique et de la Demographie; retrieved from Archivage National des Donnees on 24/07/2025
TZA	2022	National Bureau of Statistics; retrieved from NBS Microdata Catalog on 10/07/2025
ZAF	2022	Statistics South Africa; retrieved from DataFirst Data Portal on 16/07/2025

Table A1: Data sources for non-IPUMS population censuses

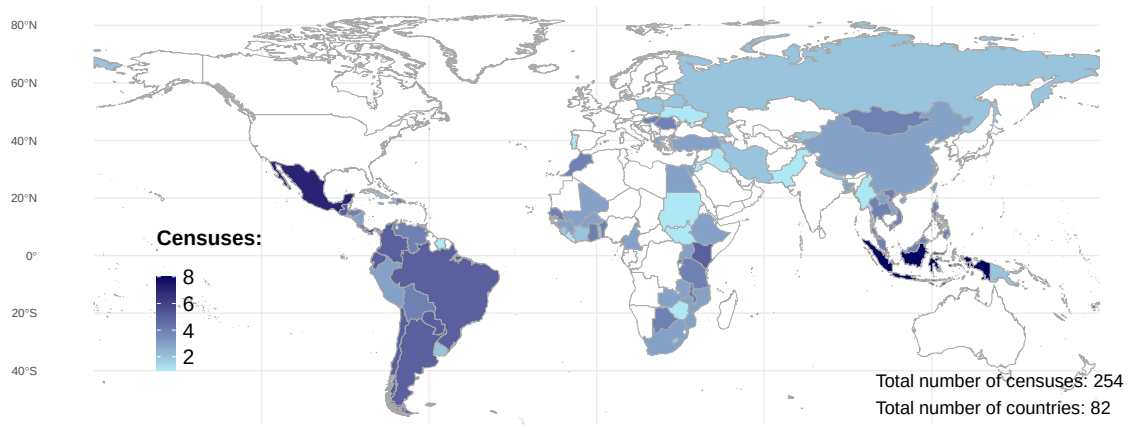


Figure A2: Geographic coverage of the data

This figure illustrates the geographic coverage of the census data (World Bank, 2024). Darker shades of blue indicate the inclusion of more censuses from a given country. Countries for which no censuses are included are shown in white. Overall, the data comprises 254 population censuses conducted across 82 countries between 1960 and 2023.

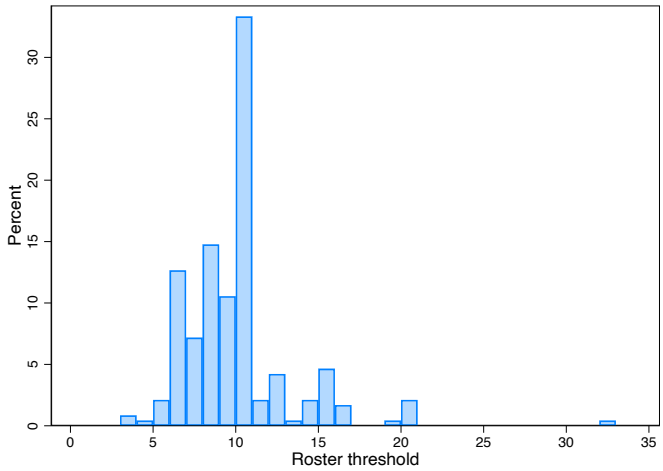


Figure A3: Distribution of pre-printed slots on enumeration forms across censuses

This figure displays the distribution of pre-printed slots for household members on the enumeration form across all censuses in our sample. More than 30% of censuses feature ten pre-printed slots. Forms with six, eight or nine pre-printed slots are also common.

A.3 Alternative mechanisms

An alternative explanation for the bunching of household sizes at the threshold that does not imply missing population is that all members are recorded but enumeration forms are not linked at the time of data entry (perhaps because identifiers are not entered correctly). If this were the case, one would expect to observe a significant fraction of households without household heads in censuses for which we estimate a large share of missing people (because household heads are listed first on the first form). We test this in the data. We find few households without heads. In 76% of censuses, all households have a household head. In 94% of censuses, more than 99% of households have a household head. Across censuses, there is no statistically significant relationship between the share of households without a head and our estimate of missing population.

Another alternative explanation for the observed bunching at the threshold is that the second enumeration form gets lost rather than never being filled. In this case, the population would still be undercounted, but the underlying mechanism would not be enumerator shirking. It is not clear, however, why second forms should be any more likely to be lost than first forms. If the probability of losing them is equal, on the other hand, this probability should be equal to the share of households without heads - which is very small empirically, as discussed above.

A.4 Estimation approach

A.4.1 Estimator

Assumptions about the data generating process. Let x be the number of pre-printed slots on the enumeration form, also referred to as the threshold below. We impose the following assumptions on the data generating process:

1. The true household size distribution exhibits exponential decay at rate $\alpha \in (0,1)$ in the window $[x, x+c]$ where $c > 0$. Let H_s be the true share of households of size s. Then, for all $s \in [x, x+c]$, we have:

$$H_s = H_{x-1}\alpha^{s-x+1} \tag{1}$$

2. Households can be displaced to any multiple i of the threshold x that is below their true size, including displacements from higher multiples of the threshold to lower multiples:

$$\tilde{H}_{ix} = H_{ix} + \sum_{i \leq j < c/x} \sum_{s=jx+1}^{\max\{(j+1)x-1, c\}} D_s^{ix} + \sum_{i < j \leq c/x} D_{jx}^{ix} - \sum_{0 < j < i} D_{ix}^{jx} + \epsilon_x \tag{2}$$

where $D_l^m \geq 0$ is the share of households displaced from size l to size m. The first term captures displacements from non-threshold multiples to the right of a given threshold multiple ix to the given threshold multiple ix. The second term captures displacement from all larger threshold multiples ($jx > ix$) to the given threshold multiple ix, and the third term captures displacements away from the given threshold multiple ix to lower threshold multiples ($jx < ix$).

The observed share of households of sizes that are above the threshold and do not correspond to a multiple of the threshold is the true number of households of this size minus the displacement to

all lower multiples of the threshold:

$$\tilde{H}_s = H_s - \sum_{i<s/x} D_s^{ix} + \epsilon_s \tag{3}$$

Finally, the observed share of households of size $x-1$ is simply:

$$\tilde{H}_{x-1} = H_{x-1} + \epsilon_{x-1} \tag{4}$$

Moment conditions. We use two moments to estimate the decay parameter α:

1. In expectation, the sum of observed households in $[x, x+c]$ equals the true number of households in this window. This is because displacements across household sizes within the interval cancel out and the sampling error is zero in expectation:

$$E\left[\sum_{s=x}^{x+c} \tilde{H}_s\right] = E\left[\sum_{s=x}^{x+c} (H_s + \epsilon_s)\right] \tag{5}$$

$$= \sum_{s=x}^{x+c} H_s + \sum_{s=x}^{x+c} E\left[\epsilon_s\right] \tag{6}$$

$$= \sum_{s=x}^{x+c} H_{x-1}\alpha^{s-x+1} \tag{7}$$

$$= H_{x-1}\frac{\alpha(1-\alpha^{(c+1)})}{1-\alpha} \tag{8}$$

2. In expectation, the number of observed households of size $x-1$ equals the true number of households of this size:

$$E\left[\tilde{H}_{x-1}\right] = H_{x-1} + E\left[\epsilon_{x-1}\right] = H_{x-1} \tag{9}$$

We define our estimator of α as the unique solution $\hat{\alpha}$ that satisfies both of the conditions:

$$\frac{\hat{\alpha}(1-\hat{\alpha}^{(c+1)})}{1-\hat{\alpha}} = \frac{1}{\tilde{H}_{x-1}} \sum_{s=x}^{x+c} \tilde{H}_s \tag{10}$$

Uniqueness. There exists a unique solution $\hat{\alpha}$ to equation (10), as shown below.

Define the function $g(\alpha)$ as

$$g(\alpha) \;:=\; \frac{\alpha\,(1-\alpha^{(c+1)})}{1-\alpha} \;=\; \alpha + \alpha^2 + \cdots + \alpha^{(c+1)} \tag{11}$$

and denote

$$\tilde{S} := \sum_{s=x}^{x+c} \tilde{H}_s \tag{12}$$

Note that g is strictly increasing in the interval $(0, 1)$:

$$g'(\alpha) \;=\; \sum_{k=1}^{c+1} k\,\alpha^{\,k-1} \;>\; 0 \qquad \text{for all } \alpha \in (0,1) \tag{13}$$

Additionally, we have that

$$\lim_{\alpha \to 0} g(\alpha) = 0 \qquad \text{and} \qquad \lim_{\alpha \to 1} g(\alpha) = c+1 \tag{14}$$

Therefore, by the Intermediate Value Theorem, there exists a unique $\hat{\alpha} \in (0,1)$ solving equation (10) for each $\tilde{S}/\tilde{H}_{x-1} \in (0, c+1)$.

Consistency. We assume that draws from the population of recorded (yet manipulated) household sizes are independent and identically distributed. Under this assumption, the sampling error converges to its expectation (zero) as the sample size goes to infinity. Additionally, we assume that both the observed and the true number of households of

size $x-1$ is strictly positive: $\tilde{H}_{x-1} > 0$ and $H_{x-1} > 0$. Then, we have

$$\lim_{n\to\infty} \frac{1}{\tilde{H}_{x-1}} \sum_{s=x}^{x+c} \tilde{H}_s = \frac{1}{\lim_{n\to\infty} \tilde{H}_{x-1}} \lim_{n\to\infty} \left[\sum_{s=x}^{x+c} \tilde{H}_s \right] \tag{15}$$

Note that the observed share of households of size $x-1$ converges to the true share:

$$\begin{aligned} \lim \tilde{H}_{x-1} &= \lim(H_{x-1} + \epsilon_{x-1}) && (16)\\ &= H_{x-1} + \lim \epsilon_{x-1} && (17)\\ &= H_{x-1} + E[\epsilon_{x-1}] && (18)\\ &= H_{x-1} && (19) \end{aligned}$$

Similarly, the observed sum of households of sizes $[x, x+c]$ converges to the true sum because all displacements within the interval cancel out:

$$\begin{aligned} \lim \left[\sum_{s=x}^{x+c} \tilde{H}_s \right] &= \lim \sum_{s=x}^{x+c} (H_s + \epsilon_s) && (20)\\ &= \sum_{s=x}^{x+c} H_s + \sum_{s=x}^{x+c} \lim \epsilon_s && (21)\\ &= \sum_{s=x}^{x+c} H_s + \sum_{s=x}^{x+c} E[\epsilon_s] && (22)\\ &= \sum_{s=x}^{x+c} H_s && (23) \end{aligned}$$

Hence, the right-hand side of equation (10) converges to the truth, and thus, so does our estimator $\hat{\alpha}$.

A.4.2 Estimation

Numerical procedure. We use an iterative numerical procedure to solve equation (10) for $\hat{\alpha}$.

First, we define the function

$$f(\alpha) \;=\; \tilde{H}_{x-1}\,\frac{\alpha(1-\alpha^{(c+1)})}{1-\alpha} \;-\; \tilde{S} \tag{24}$$

Its first derivative is

$$f'(\alpha) \;=\; \tilde{H}_{x-1}\sum_{k=1}^{c+1} k\,\alpha^{k-1} \;>\; 0 \tag{25}$$

The code then employs a Newton-Raphson update procedure where

$$\alpha_{k+1} = \alpha_k - \frac{f(\alpha_k)}{f'(\alpha_k)},$$

with an initial guess of $\alpha_0 = 0.27$.[4] We then iterate until the algorithm converges to a solution with a tolerance of 10^{-8}:

$$|\alpha_{k+1} - \alpha_k| < 10^{-8} \tag{26}$$

We impose a maximum of 1000 iterations. The converged solution $\widehat{\alpha}$ is then used to generate the counterfactual distribution in the prediction range:

$$P_i = H_{x-1}\,\widehat{\alpha}^{\,i-x+1}, \qquad i \in [x, x+c] \tag{27}$$

Outside the interval $[x, x+c]$, we set

$$P_i = \tilde{H}_i \qquad i \notin [x, x+c] \tag{28}$$

Finally, the share of missing people M is computed as follows:

$$M = 1 - \frac{\sum_{s=1}^{\tilde{s}_{max}} s \cdot \tilde{H}_s \cdot T}{\sum_{s=1}^{\tilde{s}_{max}} s \cdot P_s \cdot T} \tag{29}$$

where T represents the total number of households and $\tilde{s}_{max}$ represents the maximum observed household size.

[4] Starting with different initial guesses yields the same results.

Implementation. In our main estimation approach, we set $c = x - 1$. The uniqueness of the estimated α solution then relies on the condition that $\tilde{S}/\tilde{H}_{x-1} \in (0, 2x-1)$. Three cases must be considered in the data. First, $\tilde{S} = 0$, i.e., that no mass lies inside the displacement interval to the right of the threshold. This condition is violated in one case in our sample: in the 2001 Armenian census, the maximum household size observed in the raw data lies below the threshold x. We set the missing population to zero in this case, assuming the design issue we study does not apply. Second, $\tilde{H}_{x-1} = 0$, i.e., that the household size just to the left of the first threshold has zero mass. We do not find any cases of this in our data. Third, $\tilde{S} \leq \tilde{H}_{x-1} \cdot x$. In our data, this condition is never violated either, with the fraction $\tilde{S}/\tilde{H}_{x-1}$ ranging between 0.45 and 6.12 across censuses, whereas the lowest threshold included in our main estimation is 7. The resulting distribution of the estimated $\widehat{\alpha}$ values is shown in Figure A4.

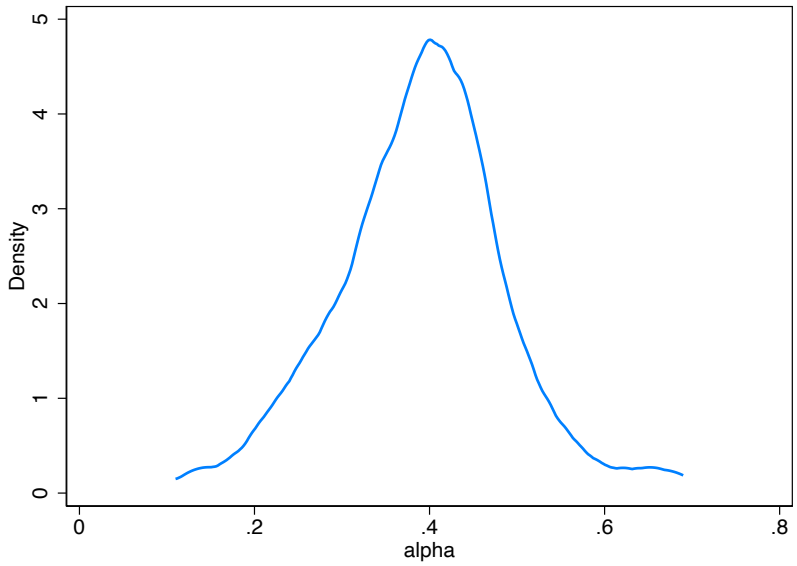
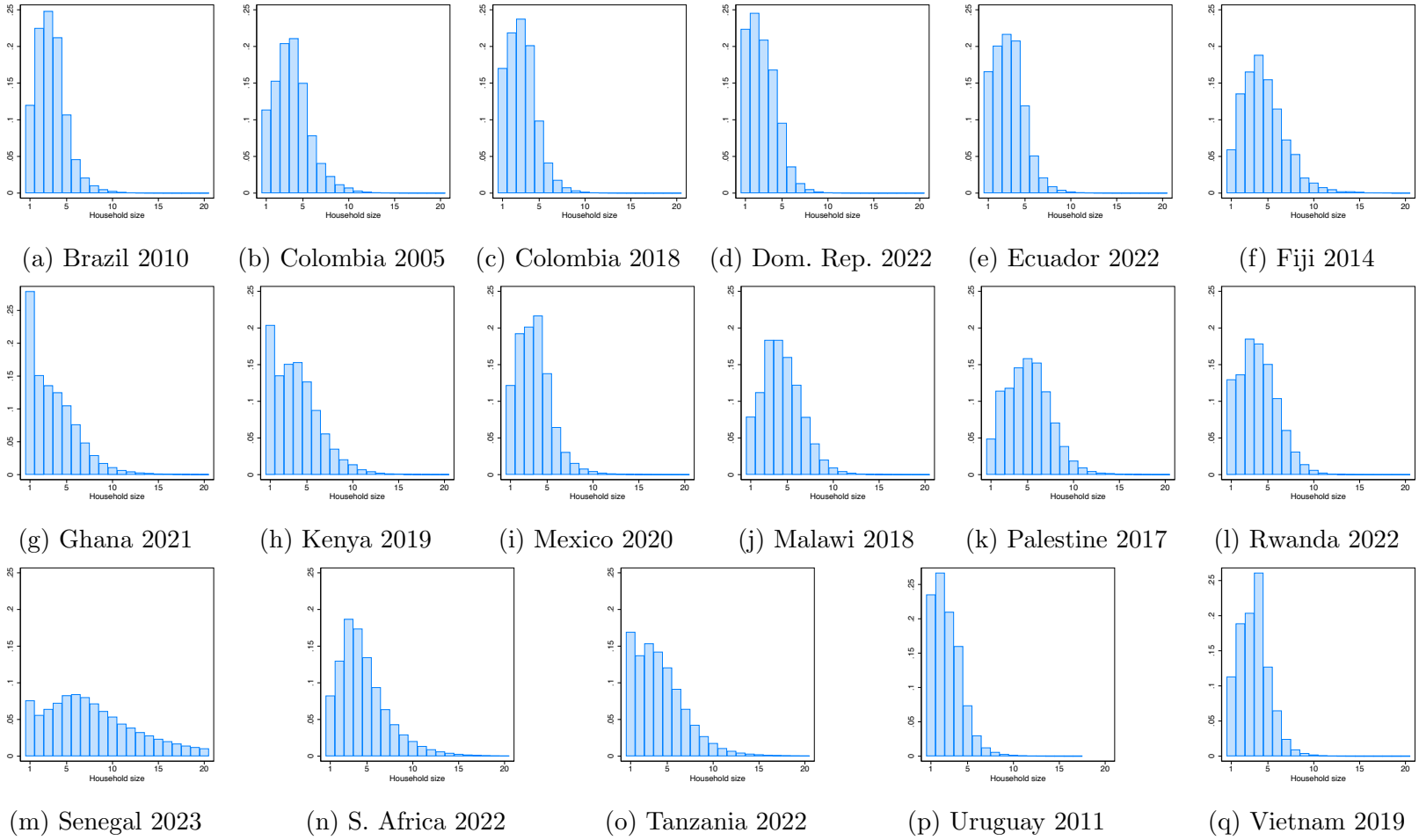


Figure A4: Distribution of decay rate estimates ($\widehat{\alpha}$)

This figure displays the distribution of estimated values for $\widehat{\alpha}$ in the main estimation for 188 censuses.

A.4.3 Supporting evidence

Household size distributions in tablet-based censuses. We show household size distributions in tablet-based censuses in Figure A5 for each of the 17 censuses below to highlight the absence of obvious bunching at common numbers of pre-printed enumeration forms. To motivate our assumption of exponential decay, Figure A6 displays the corresponding decay rates of the household size distributions for the same 17 tablet-based censuses, plotted from the respective household size distribution peaks onward. The parts of each sub-figure highlighted in bold indicate the part of the household size distribution for which we assume decay rates to be constant in our main estimation approach. Note that observed decay rates for larger household sizes are noisy in some censuses because the number of observations in this part of the distributions is very small. Overall, decay is fairly constant within census.



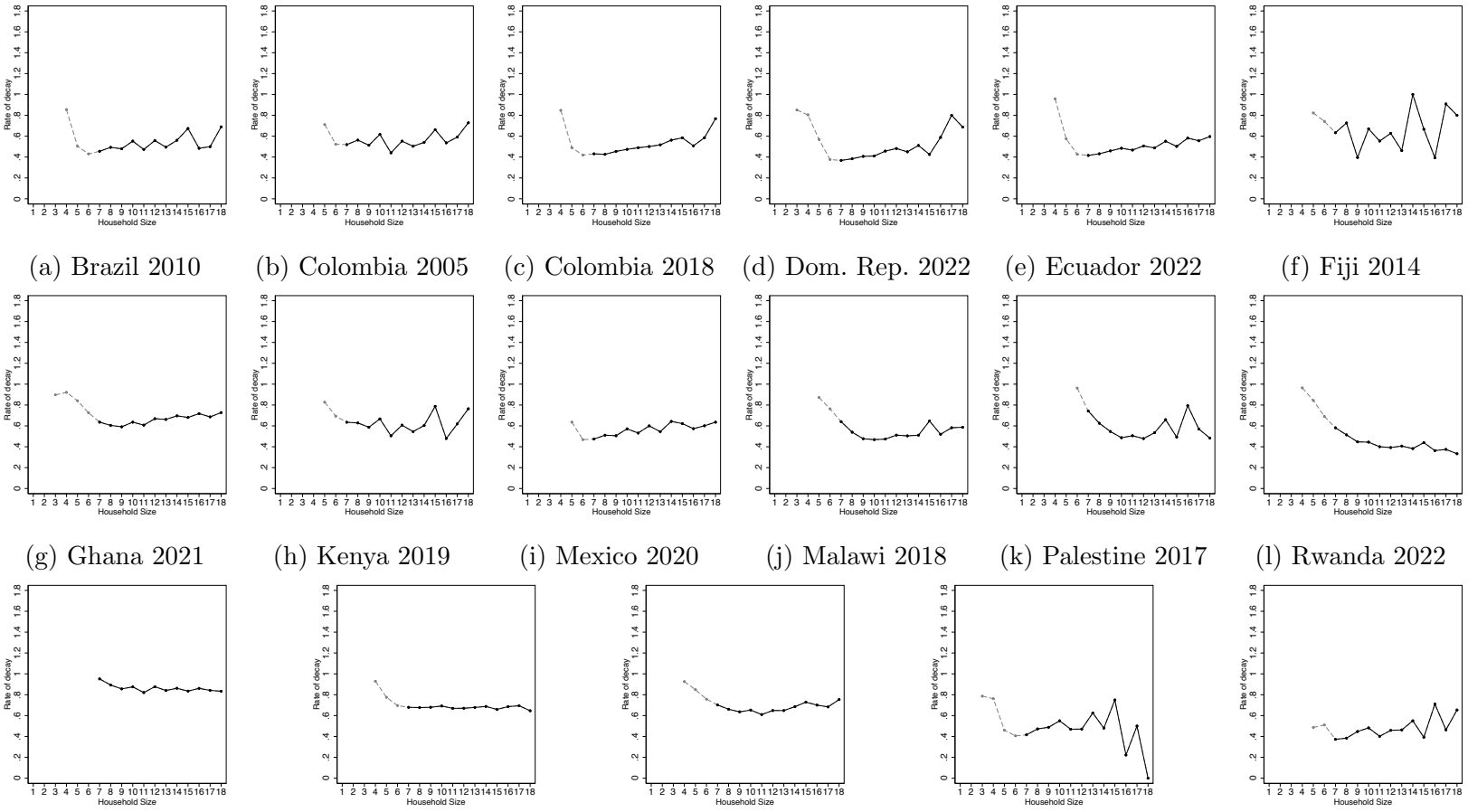
(a) Brazil 2010 (b) Colombia 2005 (c) Colombia 2018 (d) Dom. Rep. 2022 (e) Ecuador 2022 (f) Fiji 2014

(g) Ghana 2021 (h) Kenya 2019 (i) Mexico 2020 (j) Malawi 2018 (k) Palestine 2017 (l) Rwanda 2022

(m) Senegal 2023 (n) S. Africa 2022 (o) Tanzania 2022 (p) Uruguay 2011 (q) Vietnam 2019

Figure A5: Absence of bunching in censuses using tablets

This figure shows the household size distribution for every of the 17 censuses in our data that were conducted on tablets, commonly referred to as computer-assisted personal interviewing (CAPI).



(a) Brazil 2010 (b) Colombia 2005 (c) Colombia 2018 (d) Dom. Rep. 2022 (e) Ecuador 2022 (f) Fiji 2014

(g) Ghana 2021 (h) Kenya 2019 (i) Mexico 2020 (j) Malawi 2018 (k) Palestine 2017 (l) Rwanda 2022

(m) Senegal 2023 (n) S. Africa 2022 (o) Tanzania 2022 (p) Uruguay 2011 (q) Vietnam 2019

Figure A6: Decay in censuses using tablets

This figure shows the decay rate of the household size distribution from its peak onward for every of the 17 censuses in our data that were conducted on tablets. The parts highlighted in bold indicate the part of the household size distribution for which we assume decay rates to be constant in our estimation. Note that observed decay rates for larger household sizes are noisy in some censuses because the number of observations in this part of the distributions is very small.

Placebo estimates from tablet-based censuses. We estimate missing people for hypothetical numbers of pre-printed slots between 7 and 15 in 17 tablet-based censuses. As Figure A7 shows, our estimates are overwhelmingly close to zero. More specifically, 78% of our estimates fall between -0.1% and 0.1%, with the majority of these (94 out of 120) slightly negative. Another 18 estimates (12%) are more negative, and only 15 out of 153 estimates (10%) exceed 0.1% (shown in yellow, orange and red) and 1 exceeds 1% (shown in red).

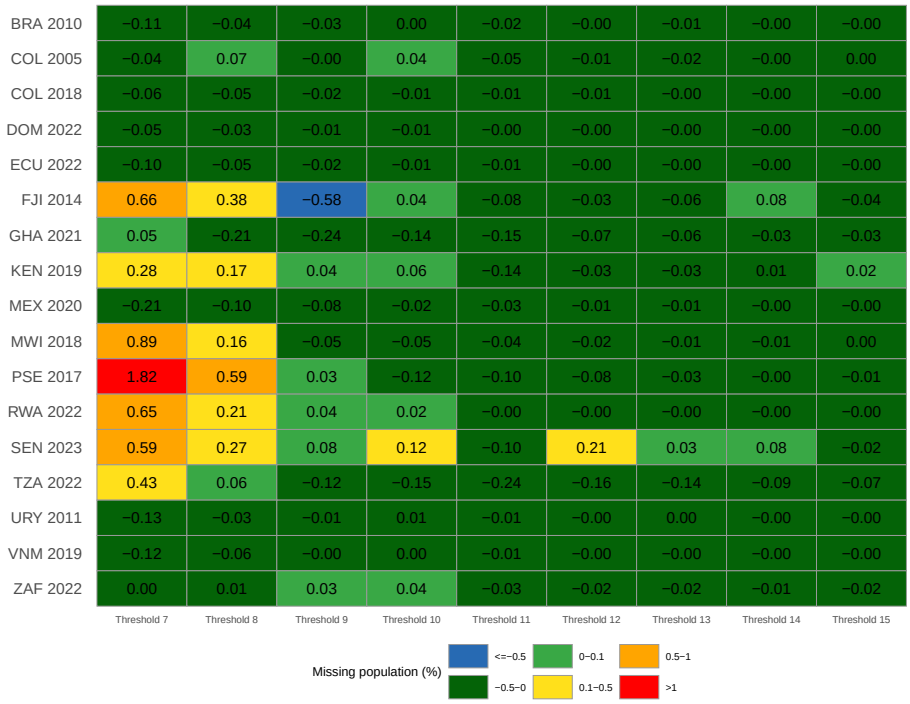


Figure A7: Placebo estimates of missing people in tablet censuses (main approach)

This figure displays estimates of missing people in tablet-based (CAPI) censuses across different hypothetical thresholds. Every cell represents a counterfactual estimate of missing population for every combination of 17 tablet-conducted censuses and nine hypothetical numbers of pre-printed slots that we include in our main estimation for paper-conducted censuses. The estimation employs our main estimation approach explained in Section IIA of the paper.

Truthful recording of household sizes away from the threshold. We leverage the particular design of the 2000 Papua New Guinean population census to provide evidence consistent with the underrecording of household members in full enumeration forms, but no such behavior at other household sizes. In particular, we build on two questions asked in the census at hand. First, the census records for each individual whether or not their mother is in the same household (if alive). Second, it asks each woman above the age of 14 to report the number of their biological children living with them in their household. We compare the total number of children reported by mothers in a given household to the total number of children in the roster whose mother is indicated to reside in the same household. As Figure A8 shows, the average ratio of children reported by mothers over children recorded in the roster exceeds 1 at all household sizes, but does so disproportionately at the number of pre-printed slots (7). Importantly, the ratio varies little across all other household sizes (even if noisily estimated for some). We interpret this as suggestive evidence that household sizes that differ from the number of pre-printed slots are truthfully reported (or equally distorted) while rosters of households with full enumeration forms (disproprtionately) under-record household members. The increase in the ratio at the number of pre-printed slots can be explained by the disproportionate omission of children relative to mothers when enumerators do not make use of a second enumeration form in large households. This arises because, by instruction, children are recorded later in the roster and thus more susceptible to omission.

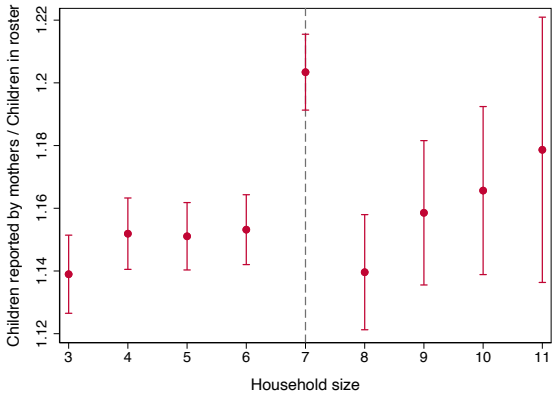


Figure A8: Mothers and children in the Papua New Guinean population census 2000

This figure displays the mean of the ratio of biological children reported by mothers in a given household and the total number of children in the roster whose mother is indicated to reside in the same household by household size. Red dots indicate point estimates and bars indicate 95% confidence intervals. The number of pre-printed slots in the household roster is seven, as indicated by the vertical black dashed line.

Consistency of counterfactual with stipulated displacement process. Our estimator does not constrain displacement of mass within the displacement interval to go from $[x+1, 2x-1]$ to the threshold x. To ensure that our counterfactual household size distributions are consistent with the stipulated displacement process, we empirically examine how our counterfactual distributions relate to the observed distributions.

In 95% of censuses, we detect weakly positive excess mass at the threshold. In the household size bins to the right of the threshold between $x+1$ and $2x-1$, we detect missing mass in 69% of bins (1,139 out of 1,649). In the remaining 31% of bins (510 out of 1,649), spread across 138 censuses, we find some excess mass. However, the excess is typically very small. In fact, it falls within the 95% confidence interval of the observed household share in 61% of these bins (313 out of 510). The remaining 39% of bins (197 out of 510) are spread across 71 censuses.

Further analysis of these cases of predicted excess mass to the right of the threshold shows that in 46 out of these 71 censuses, the underprediction occurs exclusively in the rightmost part of the interval $[x+1, 2x-1]$ where the data are noisy due to the low number of observations in the tails of the distributions. In all of these cases, our prediction is redistributing mass from large household sizes in this interval to smaller ones, thereby leading to an underestimate of missing population.

In the remaining 25 censuses out of the above 71, underpredicted mass is spread across different household sizes in the interval $[x+1, 2x-1]$. Only in 3 of them is the underprediction exclusively found in the leftmost part of the interval, implying that we are unequivocally overestimating the missing population in these 3 cases. In all other cases, it remains ambiguous whether we are over- or underestimating. However, the set of these 25 censuses is characterized by below-average estimates of missing population overall. In fact, if we exclude them from our analysis altogether, we obtain a larger estimate of average missing population in low- and lower-middle-income countries than reported in the paper, 0.63% missing population rather than 0.58%.

A.5 Missing people across countries

A.5.1 Main results

We find that the omission of members of large households is a widespread phenomenon. Figure A9 presents aggregated household size distributions by the number of pre-printed slots on census enumeration forms, uncovering prima facie evidence of this omission across the most common census form designs. Our estimation approach delivers the counterfactual household size distribution for every census in the estimation sample. Figure A10 provides a graphical representation of this counterfactual for the Tanzanian population censuses in 2012 and 2022. The solid red line represents the estimated counterfactual household size distribution, whereas the dashed red line represents the observed household sizes outside of the prediction range.

We present our estimates of missing population for each census in Table A2 and plot these estimates against GDP per capita in Figure A11. Note that for two of the poorest country-years in our data (Papua New Guinea 1990 and 2000), the share of missing is 2.39% and 2.28%, respectively, but GDP data is not available. Hence, the reported share of missing population in poor countries likely represents an underestimate.

Table A2: Missing population and regional percentiles

Country	Year	Slots	Overall	Poor	Rich	Urban	Rural	Regional 10th pctl	Regional median	Regional 90th pctl
ARG	1970	7	0.20					-0.12	0.41	2.07
ARG	1980	7	0.02	-0.17	1.14	0.01	0.89	-0.26	0.38	1.11
ARG	2001	7	0.05	-0.01	0.82	0.04	0.36	-0.06	0.20	0.59
ARM	2001	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ARM	2011	10	0.08	-0.02	0.14	0.04	0.15	-0.02	0.05	0.27
BEN	1979	10	-0.20	-0.64	0.00			-0.53	-0.24	0.58
BEN	1992	8	0.13	0.41	0.04	0.22	0.07	-0.20	0.25	0.56
BEN	2002	10	-0.16	-0.09	0.18	-0.23	-0.10	-0.43	-0.10	0.53
BEN	2013	10	0.01	0.07	0.06	0.08	-0.02	-0.10	-0.02	0.64
BFA	1985	10	0.20					-0.06	0.25	0.53
BFA	1996	10	0.35	-0.05	0.53			0.01	0.49	0.94
BFA	2006	10	0.34	0.05	0.55	0.09	0.46	0.17	0.33	1.06
BGD	1991	11	0.13	0.24	0.04	0.39	0.08	0.03	0.16	0.22
BGD	2001	10	0.15			0.21	0.13	0.08	0.17	0.39
BGD	2011	10	0.05			0.06	0.05	0.01	0.05	0.15
BLR	1999	8	-0.01	-0.01	-0.01	-0.01	-0.02	-0.03	-0.01	-0.00
BLR	2009	7	-0.03	-0.02	0.00	-0.03	-0.01	-0.04	-0.02	-0.00
BOL	1976	9	0.27	0.43	0.20	0.27	0.30	0.13	0.38	2.36
BOL	1992	9	0.17	-0.04	0.30	0.10	0.31	0.01	0.27	0.47
BOL	2001	9	0.17	0.12	0.34	0.10	0.30	0.10	0.27	0.92
BOL	2012	8	0.82	0.29	1.31	0.59	1.33	0.40	1.13	2.14
BRA	1960	10	0.64	0.21	0.99	0.31	0.91	0.24	0.83	1.11
BRA	1970	10	0.55	0.12	0.70	0.34	0.84	0.40	0.73	1.02

Table A2: Missing population and regional percentiles

Country	Year	Slots	Overall	Poor	Rich	Urban	Rural	Regional 10th pctl	Regional median	Regional 90th pctl
BWA	1981	8	0.93	0.17	1.12			-0.49	0.79	3.13
BWA	1991	10	-0.05	-0.13	-0.02	-0.11	-0.00	-0.95	0.11	1.86
BWA	2001	11	-0.31	-0.05	-0.44			-0.78	-0.21	0.40
BWA	2011	10	0.02	0.00	-0.25			-0.43	0.06	0.60
CHL	1970	9	0.79	0.22	1.11	0.59	1.45	0.35	0.95	1.50
CHL	2017	8	-0.02	0.00	-0.02	-0.03	-0.01	-0.07	0.00	0.09
CIV	1988	7	0.81	1.36	0.86			0.25	0.73	1.42
CIV	1998	8	0.80	1.28	0.88	0.84	0.76	0.30	0.84	1.89
CMR	1976	10	0.29	0.22	0.24			-0.11	0.19	1.90
CMR	1987	10	0.22	0.22	0.02	0.20	0.23	-0.08	0.32	0.90
COL	1973	9	1.19	0.75	1.49	0.92	1.63	0.69	1.27	2.30
COL	1985	14	-0.01	0.00	-0.02	-0.00	-0.02	-0.05	0.02	0.12
COL	1993	10	0.28	0.12	0.62	0.18	0.56	0.05	0.29	0.87
CRI	1973	9	1.09	0.25	1.85	0.43	1.61	0.83	1.18	1.53
CRI	1984	10	0.09	-0.01	0.21	-0.02	0.21	-0.02	0.09	0.32
CRI	2000	10	-0.00	-0.01	0.05	-0.01	0.02	-0.07	0.01	0.03
CRI	2011	12	-0.00	-0.01	-0.00	-0.01	-0.00	-0.05	-0.01	0.01
DOM	1981	8	1.58	0.74	2.47			0.93	1.99	3.17
ECU	1982	9	0.44	-0.05	0.72			0.15	0.66	1.32
ECU	1990	9	0.41	0.09	0.85	0.24	0.75	0.26	0.55	0.89
ECU	2001	7	0.56	0.12	1.44	0.44	0.95	0.36	0.89	1.40
ECU	2010	10	0.09	0.05	0.19	0.06	0.15	0.04	0.11	0.46
EGY	1986	7	1.98	0.81	2.72	1.54	2.79	1.38	2.38	3.07
EGY	1996	7	1.73	0.44	3.24	0.92	2.57	0.79	1.95	3.36
EGY	2006	7	0.52	0.11	1.36	0.20	0.85	0.05	0.42	2.10
ETH	1984	8	0.90	1.03	1.42	0.90		0.18	0.62	1.42
ETH	1994	10	0.27	0.37	0.26	0.27	0.18	0.06	0.22	1.52
ETH	2007	8	2.31			1.69	2.36	1.39	2.40	6.63
FJI	1966	10	0.58					0.21	0.76	1.64
FJI	1976	10	0.20					-0.26	0.46	0.68
FJI	1986	10	0.50	0.38	0.66	1.09	-0.02	-1.27	0.57	0.79
FJI	1996	10	0.11	0.05	0.22			-0.30	-0.01	0.33
FJI	2007	8	0.07	0.46	0.78	0.17	0.01	-0.12	-0.01	0.34
FRA	2011	8	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.00	0.00
GHA	1984	10	0.08					0.01	0.19	0.77
GHA	2000	10	0.95	0.54	1.59	0.81	1.08	0.61	0.95	2.81
GHA	2010	10	0.59	0.20	1.12	0.39	0.79	0.23	0.59	2.22
GIN	1983	10	0.45	1.44	0.21	1.31	0.22	-0.69	0.42	1.42
GIN	1996	9	0.58	0.52	0.50	0.59	0.59	-0.02	0.72	1.20
GIN	2014	10	0.86			0.69	0.95	0.46	1.10	1.84
GRC	1971	8	-0.04	-0.01	-0.01			-0.11	-0.02	0.09
GTM	1964	12	0.02	0.20	-0.06	0.05	0.06	-0.05	0.06	0.20
GTM	1973	9	0.59			0.53	0.62	0.24	0.62	1.23
GTM	1981	9	0.44	0.23	0.48	0.21	0.55	0.05	0.52	0.91
GTM	1994	10	0.22	0.04	0.24	0.11	0.28	0.07	0.20	0.45
GTM	2002	10	0.29	0.03	0.53	0.12	0.45	0.13	0.25	0.64
HND	1974	9	0.87			0.65	0.99	0.35	0.94	2.03
HND	1988	9	0.81	0.16	1.14	0.35	1.11	-0.00	0.82	1.71
HND	2001	10	0.25	0.01	0.49	0.09	0.39	0.05	0.42	1.08
HND	2013	9	0.13	-0.02	0.42	0.08	0.21	-0.02	0.17	0.53
HTI	1971	10	0.17	0.62	0.12	0.62	0.17	0.07	0.21	0.28
HTI	2003	8	1.22	0.96	1.30	1.00	1.38	0.81	1.29	1.99
HUN	1990	8	-0.03	-0.04	-0.04	-0.01	-0.06			
HUN	2001	10	0.01	-0.00	0.07					
IDN	1971	15	-0.01	-0.04	-0.01	-0.03	-0.01	-0.15	-0.02	0.08
IDN	1976	15	-0.01	-0.03	-0.01	0.01	-0.01	-0.12	0.02	0.21
IDN	1980	15	-0.00	-0.02	0.00	-0.01	-0.00	-0.02	-0.00	0.03
IDN	1985	15	0.00	-0.01	-0.02	-0.01	-0.00	-0.02	0.00	0.05
IDN	1990	15	-0.00	-0.01	-0.00	-0.01	0.00	-0.01	0.00	0.07
IDN	1995	15	0.00	-0.00	-0.00	-0.01	-0.00	-0.10	0.00	0.00
IDN	2000	8	0.25			0.23	0.27	0.09	0.32	0.66
IDN	2005	10	0.02	-0.01	0.01	0.01	0.02	-0.02	0.02	0.09
IRN	2006	11	-0.02	-0.00	-0.04	-0.01	-0.01	-0.03	0.01	0.02
IRQ	1997	10	0.19	0.35	0.13	0.23	0.18	-0.26	0.22	0.41
JAM	1991	10	-0.01	0.02	0.17			-0.30	0.11	0.55
JAM	2001	10	-0.02	0.03	0.04	-0.10	0.07	-0.11	0.05	0.19
JOR	2004	10	0.29	0.14	0.67	0.21	0.76	0.08	0.23	1.33
KEN	1969	10	0.66					-0.04	0.65	1.99
KEN	1989	10	0.28	0.24	0.13	0.26	0.29	0.17	0.29	0.50
KEN	1999	10	0.29	0.09	0.22	0.19	0.32	0.06	0.28	1.22
KEN	2009	10	0.55	0.08	1.17	0.25	0.68	0.03	0.25	4.64
KGZ	1999	12	0.03	0.01	0.04	0.02	0.03	-0.05	0.03	0.09
KGZ	2009	14	0.10			0.15	0.07	0.01	0.11	0.31
LAO	1995	10	0.19	0.35	0.13	0.23	0.18	-0.26	0.22	0.41
LAO	2005	10	0.21	0.02	-0.06	-0.06	-0.00	-0.25	-0.04	0.16
LAO	2015	10	0.12	0.02	0.11	0.06	0.15	-0.17	0.08	0.36
LBR	2008	10	1.28	0.14	2.31	0.54	1.92	0.38	1.11	6.52
LSO	1996	10	0.13	-0.11	0.15	-0.15	0.20	-0.30	0.18	0.41
LSO	2006	8	1.06	0.45	1.46	0.39	1.31	0.59	1.23	2.28
MAR	1982	10	0.23	0.34	-0.07			-0.06	0.29	5.62
MAR	1994	10	0.06	0.12	0.13			-0.35	0.19	0.52

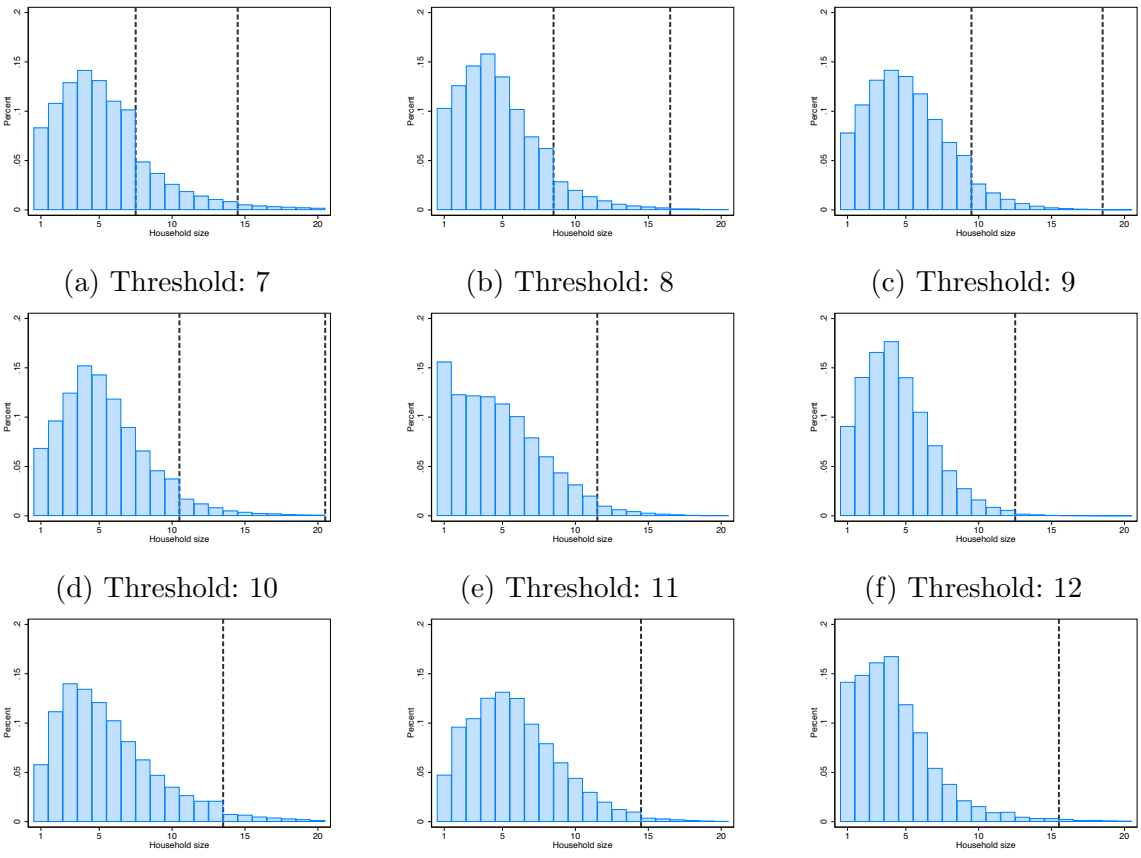
Table A2: Missing population and regional percentiles

Country	Year	Slots	Overall	Poor	Rich	Urban	Rural	Regional 10th pctl	Regional median	Regional 90th pctl
MAR	2004	9	0.15	-0.08	0.62			-0.26	0.04	1.41
MAR	2014	8	-0.19	-0.25	0.05	-0.15	-0.06	-0.28	-0.18	0.20
MEX	1970	14	0.01	0.08	0.05	0.02	0.02	-0.14	0.08	0.54
MEX	1990	10	0.15	0.46	-0.02	0.02	0.24	-0.11	0.09	0.43
MEX	1995	12	-0.02	-0.11	-0.06	0.01	-0.05	-0.12	-0.00	0.15
MLI	1987	13	0.11	-0.02	0.08			-0.50	0.16	0.50
MLI	1998	14	0.10	0.14	0.07	0.05	0.14	-0.02	0.08	0.24
MLI	2009	15	0.04	0.20	-0.04	0.23	0.01	-0.14	0.01	0.30
MMR	2014	8	0.15	0.10	0.15	0.12	0.18	0.04	0.13	0.53
MOZ	1997	9	-0.00	0.36	-0.01	0.22	-0.06	0.01	0.10	0.19
MOZ	2007	9	-0.08	0.30	-0.02	0.16	-0.14	-0.06	0.06	0.15
MUS	1990	10	0.08			0.14	0.06	-0.21	0.18	0.27
MUS	2000	10	-0.05			-0.01	-0.07	-0.17	-0.01	0.27
MWI	1987	10	0.08	0.31	0.07	0.37	0.05	-0.07	0.09	0.38
MWI	1998	10	0.14	0.19	0.06	0.19	0.14	0.04	0.17	0.34
MWI	2008	8	0.55	0.90	0.25	0.64	0.53	0.28	0.63	1.17
MYS	2000	7	0.10	0.09	0.03	-0.04	0.51	-0.79	0.38	1.14
NIC	1971	9	1.48	0.50	1.99			0.73	1.52	2.91
NIC	1995	10	0.33	0.01	0.32			-0.02	0.40	0.97
NIC	2005	10	-0.06	-0.20	0.33	-0.17	0.06	-0.34	0.05	0.39
NPL	2001	7	0.92	0.07	1.06	0.13	0.81	0.42	0.68	1.04
PAK	1998	10	-0.16	0.02	-0.24	-0.01	-0.21	-0.44	-0.16	0.15
PAN	1960	9	0.49			0.41	0.54			
PAN	1970	9	0.36			0.28	0.44			
PAN	1980	7	1.20	0.03	2.77	0.54	2.13			
PAN	1990	8	0.31	-0.27	1.50					
PAN	2000	8	0.08	0.02	0.59	0.05	0.25			
PAN	2010	8	0.08	0.14	0.65	0.06	0.23			
PER	2007	10	0.05	0.04	0.11	0.04	0.11	-0.03	0.06	0.18
PER	2017	10	0.00	0.00	0.02	0.01	0.01	-0.07	-0.00	0.09
PHL	1990	8	1.26	0.72	1.51	0.95	1.55	0.70	1.34	1.94
PHL	1995	8	0.70					0.31	0.87	1.22
PHL	2000	8	0.96	0.50	1.46			0.49	1.15	1.93
PHL	2010	8	0.50	0.26	0.88			0.29	0.60	1.02
PNG	1990	7	2.40	1.70	2.58	1.71	2.58	0.69	2.27	4.04
PNG	2000	7	2.28			2.14	2.35	0.97	2.09	3.74
PRT	1991	12	0.00			-0.01	0.01	-0.04	-0.00	0.02
PRY	1962	11	0.04			0.09	-0.00	-0.57	-0.01	1.10
PRY	1972	9	1.32	0.55	1.63	0.56	1.83	1.11	1.45	2.42
PRY	1982	9	1.26	0.51	2.18	0.66	1.76	0.47	1.56	2.05
PRY	1992	9	0.65	0.02	1.29	0.17	1.22	-0.05	0.73	1.43
PRY	2002	10	0.21	-0.01	0.56	0.10	0.36	0.03	0.15	0.53
PSE	1997	14	-0.06	-0.09	-0.08	-0.09	-0.00	-0.22	-0.04	0.27
ROU	1992	12	0.00			-0.00	0.00	-0.01	-0.00	0.02
ROU	2002	12	0.01	0.00	0.01	0.01	0.00	-0.01	0.00	0.04
RUS	2002	10	0.05			0.06	0.04	-0.01	0.00	0.12
RUS	2010	10	0.03			0.04	0.02	-0.02	0.01	0.12
RWA	1991	11	0.02	0.22	-0.01			-0.05	0.03	0.10
RWA	2002	10	0.07	0.23	0.05	0.07	0.09	-0.02	0.11	0.24
RWA	2012	12	-0.00	-0.00	-0.00	0.01	-0.01	-0.01	0.00	0.04
SDN	2008	8	2.88	2.04	2.88	2.37	3.78	1.70	2.38	5.87
SEN	1988	7	0.55	0.53	0.54			0.21	0.60	0.83
SEN	2002	10	0.53	0.31	0.89	0.40	0.64	0.22	0.41	0.82
SEN	2013	15	0.05	-0.21	-0.19	-0.10	0.18	-0.36	0.09	0.41
SLE	2004	10	1.25	0.27	1.76	0.48	1.70	0.27	1.08	2.69
SLE	2015	10	1.81	0.76	2.42	1.30	2.19	0.83	1.66	5.13
SLV	1992	8	0.82	0.10	1.61			0.23	1.28	2.24
SLV	2007	8	0.13	-0.02	0.61	-0.00	0.46	0.03	0.27	0.43
SSD	2008	8	5.47	4.55	5.51	5.35	5.45	1.60	3.43	11.14
SUR	2012	15	0.22	-1.31	-0.05			-2.36	-0.17	0.77
TGO	2010	10	0.16	-0.08	0.21	0.06	0.26	-0.06	0.14	1.03
THA	1970	12	-0.08			-0.06	-0.04	-0.28	-0.00	0.27
THA	1980	10	0.07	-0.01	0.09	-0.03	0.11	-0.16	0.11	0.67
THA	1990	10	0.07	0.14	0.09	0.07	0.08	-0.12	0.07	0.45
TUR	1985	9	0.31					-0.11	0.29	1.80
TUR	1990	7	0.84					-0.02	1.28	3.64
TUR	2000	15	-0.32					-0.93	-0.18	-0.01
TZA	1988	10	0.08	0.07	0.07			-0.12	0.14	0.51
TZA	2012	8	1.39	1.13	1.41	1.15	1.49	0.51	1.23	2.76
UGA	1991	10	0.07	0.31	0.06	0.04	0.09	-0.18	0.17	0.37
UGA	2002	10	0.18	0.15	0.39	0.16	0.20	-0.06	0.15	0.59
UGA	2014	10	0.31	0.08	0.41	0.14	0.36	0.03	0.27	0.81
UKR	2001	10	0.00			0.00	0.00	-0.01	0.00	0.02
VEN	1971	12	0.10	0.12	0.09			-0.07	0.11	0.29
VEN	1981	10	0.33	0.02	0.55	0.28	0.58	0.06	0.34	0.89
VEN	1990	10	0.18	0.10	0.34	0.12	0.60	0.04	0.37	0.57
VEN	2001	7	3.48	1.16	4.99	3.25	5.33	1.31	3.76	5.14
VNM	1989	8	0.46	0.18	0.52	0.31	0.58	0.05	0.71	1.24
VNM	1999	8	0.20	0.18	0.37	0.09	0.29	-0.03	0.18	0.69
ZAF	1996	9	0.84	0.12	1.37	0.49	1.30	0.30	0.68	1.36
ZAF	2001	10	0.19	0.06	0.28	0.11	0.30	0.02	0.13	0.53

Table A2: Missing population and regional percentiles

Country	Year	Slots	Overall	Poor	Rich	Urban	Rural	Regional 10th pctl	Regional median	Regional 90th pctl
ZMB	2000	15	-0.03	-0.03	-0.05	-0.03	-0.03	-0.11	-0.00	0.04
ZMB	2010	8	2.08	1.88	1.73			1.59	2.19	2.81
ZWE	2012	8	2.75	-0.05	1.05	0.07	3.11	2.05	2.61	4.03

This table reports the share of population missing from household rosters, overall and by subgroup, across census years. Slots denotes the number of pre-printed slots in the census. Poor and Rich refer to the bottom and top income quintiles, respectively. Urban and Rural indicate the share missing within each area type. The last three columns indicate the 10th percentile, median and 90th percentline missing estimates across regions.



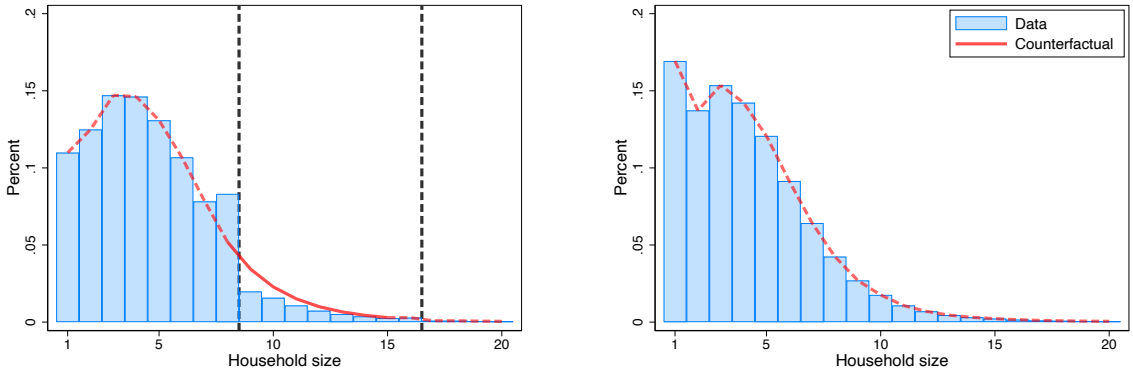
(a) Threshold: 7 (b) Threshold: 8 (c) Threshold: 9

(d) Threshold: 10 (e) Threshold: 11 (f) Threshold: 12

(g) Threshold: 13 (h) Threshold: 14 (i) Threshold: 15

Figure A9: Household size distribution by number of pre-printed slots

Each sub-figure plots aggregated household size distributions from population censuses that share a common number of pre-printed slots. For the aggregation, the number of households in each bin in each census is weighted by the inverse of the total number of households in the census. Dashed vertical lines indicate multiples of the number of pre-printed slots in the household roster.



(a) Population census 2012 (b) Population census 2022

Figure A10: Counterfactual household size distributions for Tanzania

This figure reproduces the household size distributions for the 2012 and 2022 population censuses in Tanzania, as shown in panels B and C of Figure 1 in the paper, respectively. The solid red line indicates the estimated counterfactual household size distribution over the permissible part of the distribution. For the 2022 counterfactual, we assume a hypothetical of eight pre-printed slots, equal to the actual number in 2012.

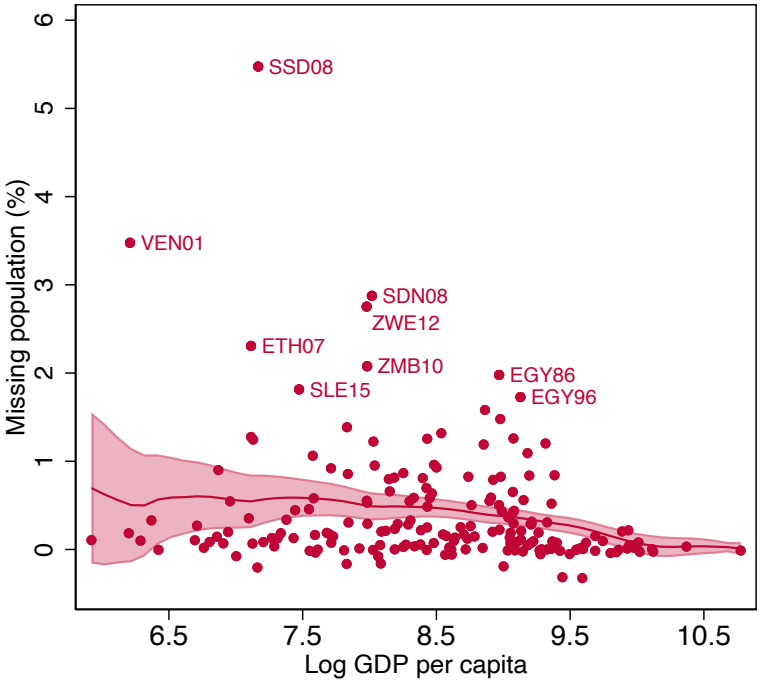


Figure A11: Missing people across censuses

This figure plots missing people against GDP per capita across censuses. Each dot represents a population census, with selected outliers labeled by a three-letter country code and year. The red line indicates a local polynomial fit and the shaded area represents the 95% confidence interval.

A.5.2 Robustness to alternative sample selection criteria

In this section, we examine the robustness of our results with respect to the sample selection criteria employed. In particular, we expand the sample of censuses included in the estimation beyond those that have pre-printed numbers of slots between 7 and 15 (188 censuses) and include all censuses with thresholds greater or equal 7 (199 censuses). As Figure A12 shows, the negative correlation between missing population and log GDP per capita is robust to this sample expansion. Moreover, the average missing population estimated in low- and middle-income countries (0.56% vs. 0.58%) as well as upper-middle countries (0.39% vs. 0.41%) and high-income countries (0.01 % vs. 0.01%) is very similar to the main approach.

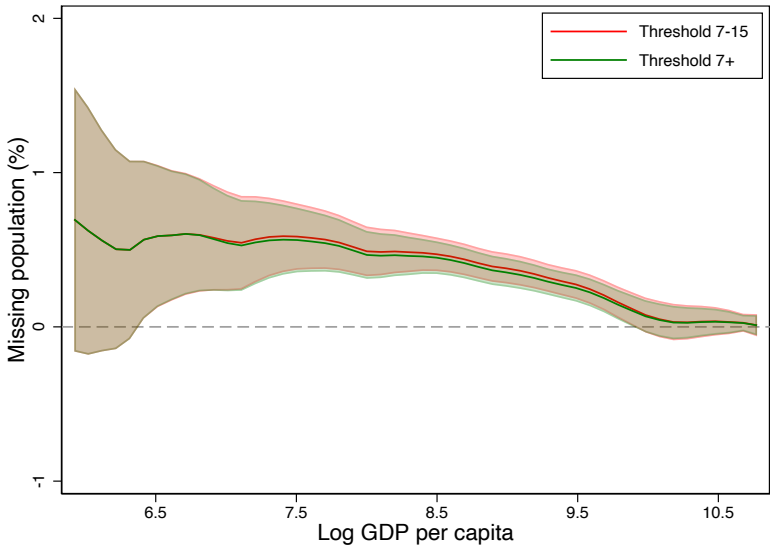


Figure A12: Missing people by GDP per capita across alternative samples

This figure plots missing people against GDP per capita across censuses, overlaying the results from the main sample used in the paper (thresholds between 7 and 15) and an alternative larger sample including all censuses with thresholds of 7 and larger. The lines indicate local polynomial fits, the shaded areas the 95% confidence intervals, respectively.

A.5.3 Robustness to alternative assumptions on data generating process

Allowing for larger upper bound to displacement (alternative 1). We relax the assumption that $c = x - 1$. Instead, we set c equal to the 90th percentile of the observed household size distribution above $2x - 1$, allowing for displacement to the threshold from larger household sizes. We maintain all other assumptions underlying our preferred estimation approach.

We validate this approach by estimating the missing population in the 17 censuses conducted using tablets for different possible thresholds from 7 to 15. As shown in Figure A13, we find that the approach yields estimates between -0.5 and 0.1 in 86% of cases. In the remaining cases, concentrated at the lower thresholds 7 and 8, estimates mainly vary between 0.1% and 1% (12 cases shown in yellow and orange). Additionally, we obtain large positive estimates above 1% in four cases (shown in red) and large negative estimates below -0.5% in five cases (shown in blue).

Household size distribution following a power law (alternative 2). Motivated by research on firm size distributions in economics (Axtell, 2001; Hernández-Pérez et al., 2006), we assume that the true household size distribution in the interval $[x, x + c]$ follows a power law with exponent $\beta > 0$ of the following form (rather than decaying exponentially):

$$H_s = H_{x-1} \left(\frac{s}{x-1} \right)^{-\beta} \tag{30}$$

We maintain all other assumptions about the data generating process laid out in section A.4.1.

The resulting moment equations can be written as:

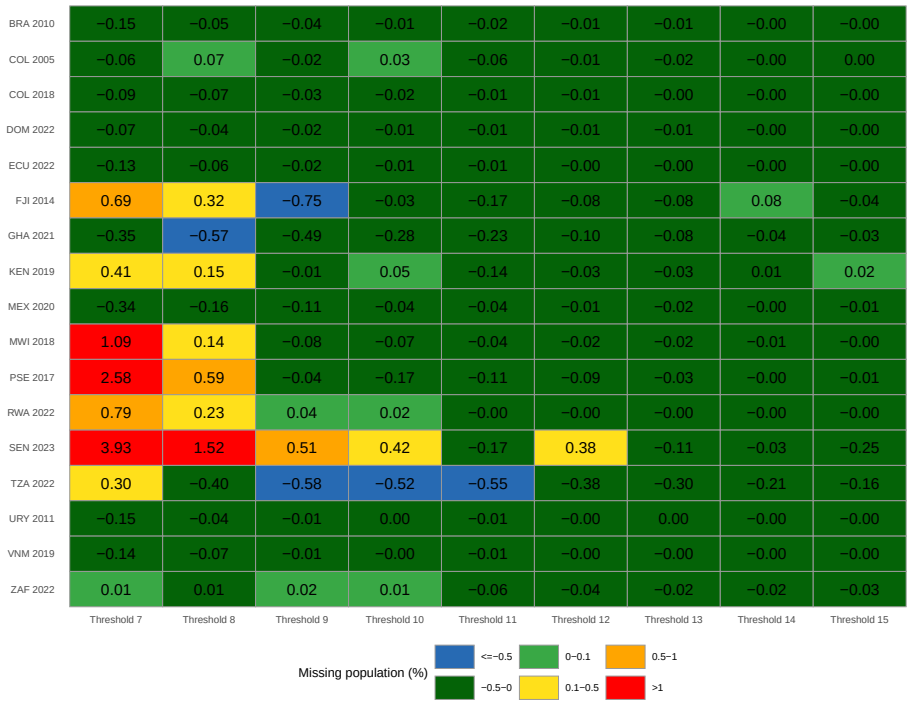


Figure A13: Placebo estimates of missing people in tablet censuses (alternative 1)

This figure displays estimates of missing people in tablet-based censuses across different hypothetical thresholds. Every cell represents a counterfactual estimate of missing population for every combination of 17 tablet-conducted censuses and nine hypothetical numbers of pre-printed slots that we include in our main estimation for paper-conducted censuses. The estimation employs our first alternative estimation approach explained in Appendix A.5.3.

$$E\left[\sum_{s=x}^{x+c}\tilde{H}_s\right] = E\left[\sum_{s=x}^{x+c}(H_s+\epsilon_s)\right] \tag{31}$$

$$= \sum_{s=x}^{x+c} H_s + \sum_{s=x}^{x+c} E\left[\epsilon_s\right] \tag{32}$$

$$= \sum_{s=x}^{x+c} H_{x-1}\left(\frac{s}{x-1}\right)^{-\beta} \tag{33}$$

and

$$E\left[\tilde{H}_{x-1}\right] = H_{x-1} \tag{34}$$

We define our estimator of the exponent β as the unique solution $\hat{\beta}$ to the below equation:

$$\frac{1}{\tilde{H}_{x-1}} \sum_{s=x}^{x+c} \tilde{H}_s = \sum_{s=x}^{x+c} \left(\frac{s}{x-1} \right)^{-\hat{\beta}} \tag{35}$$

To establish the uniqueness of $\hat{\beta}$, we rewrite the above equation as follows:

$$\frac{1}{\tilde{H}_{x-1}} \sum_{s=x}^{x+c} \tilde{H}_s = \sum_{s=x}^{x+c} \left(\frac{x-1}{s} \right)^{\hat{\beta}} \tag{36}$$

Note that the left-hand side is a constant and each term in the sum on the right-hand side is between zero and 1:

$$0 < \frac{x-1}{s} < 1 \qquad \forall s \in [x, x+c] \tag{37}$$

Then define

$$g(\beta) := \sum_{s=x}^{x+c} \left(\frac{x-1}{s} \right)^{\beta} \tag{38}$$

This function is strictly decreasing

$$g'(\beta) = \sum_{s=x}^{x+c} \left(\frac{x-1}{s} \right)^{\beta} ln \left(\frac{x-1}{s} \right) \tag{39}$$

because $ln(k) < 0$ for all $k \in (0, 1)$. Further, we have

$$\lim_{\beta \to \infty} g(\beta) = 0 \tag{40}$$

and

$$\lim_{\beta \to 0} g(\beta) = c + 1 \tag{41}$$

Therefore, by the Intermediate Value Theorem, there exists a unique $\hat{\beta} > 0$ solving equation (35) for each $\tilde{S}/\tilde{H}_{x-1} \in (0, c+1)$.

Consistency follows from section A.4.1, and the resulting distribution of the estimated $\hat{\beta}$ is shown in Figure A14, where $c = x - 1$.

Figure A15 reveals that this approach performs reasonably well, producing estimates between -0.5% and 0.09% in 83% of cases (shown dark green and light green). At the same time, it yields moderate (shown

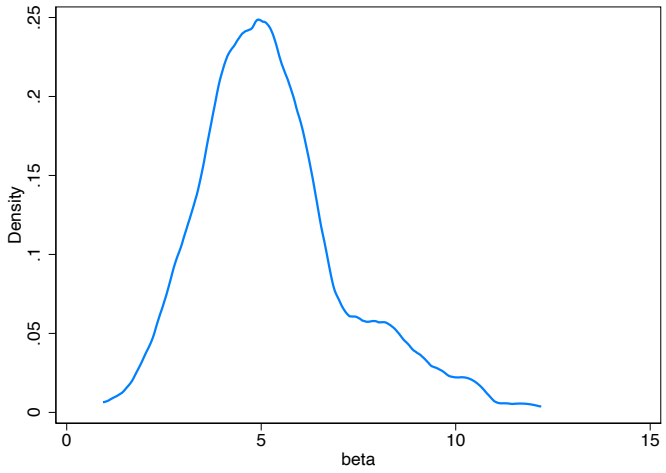


Figure A14: Distribution of power law parameter estimates ($\widehat{\beta}$)

This figure displays the distribution of estimated values for the exponent $\widehat{\beta}$ in the alternative for 188 censuses.

in yellow) estimates in 3.3% of the cases. Additionally, we obtain large negative estimates below -0.5% in 13.7% of the cases (21 out of 153). These are particularly concentrated at low thresholds and in censuses with thick household size distribution tails, such as Senegal 2023.

Comparison of results across estimation approaches. We find that the negative correlation between missing population and GDP per capita is robust to the use of the alternative estimation approaches described above. Figure A16 overlays local polynomial fits of missing population on log GDP per capita from our main approach and the two alternative approaches. Across all approaches, we observe substantial missing population in low- and lower-middle-income countries and hardly any missing at the upper end of the income distribution. Table A3 shows that the average missing population by income group are quantitatively very similar in the main approach and the first alternative approach. The second alternative approach yields somewhat larger estimates throughout.

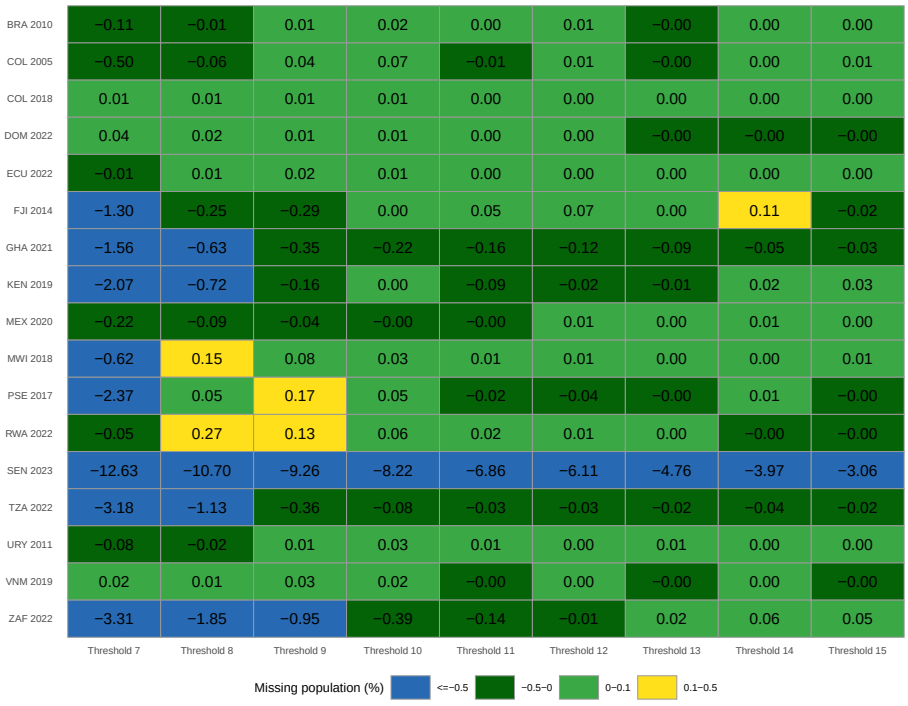


Figure A15: Placebo estimates of missing people in in tablet censuses (alternative 2)

This figure displays estimates of missing people in tablet-based censuses across different hypothetical thresholds. Every cell represents a counterfactual estimate of missing population for every combination of 17 tablet-conducted censuses and nine hypothetical numbers of pre-printed slots that we include in our main estimation for paper-conducted censuses. The estimation employs our second alternative estimation approach explained in Appendix A.5.3.

Table A3: Estimates of missing population across estimation approaches

	Main estimation	Alternative 1	Alternative 2
Censuses with missing population	60.1%	49.5%	81.4%
Average missing in low- and lower-middle income	0.58%	0.54%	1.07%
Average missing in upper-middle income	0.41%	0.45%	0.86%
Average missing in high income	0.01%	-0.02%	0.19%

This table compares results of our estimates of missing populations across our main and alternative estimation approaches. The first row shows the share of censuses that have any missing population, defined as 0.1% missing. The second row shows the equivalent share for censuses from low- and lower-middle income countries. Rows three and four show the equivalent shares for upper-middle and high income countries, respectively.

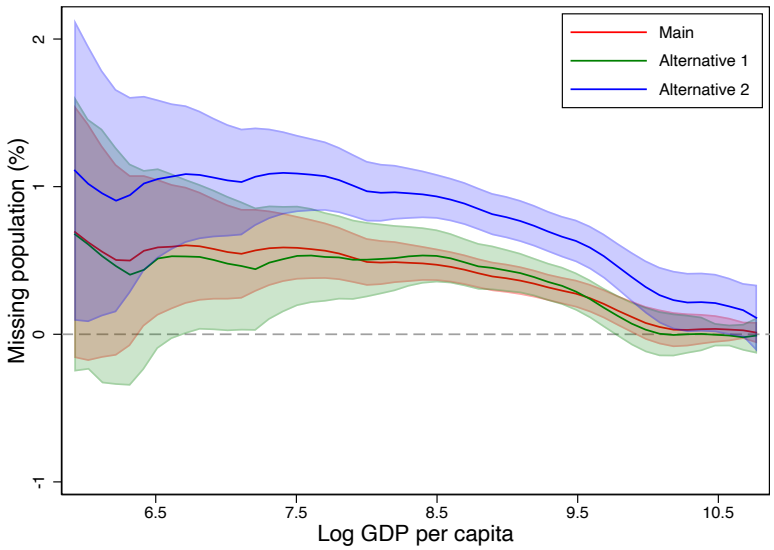


Figure A16: Missing people by GDP per capita across estimation approaches

This figure plots missing people against GDP per capita across censuses, overlaying the main estimation approach with alternative ones. The lines indicate local polynomial fits, the shaded areas the 95% confidence intervals, respectively.

A.5.4 Design of enumeration forms and missing population

Table A4 examines the extent to which the design of enumeration forms can account for the observed correlation between missing population and GDP per capita, as discussed in Section II of the paper.

Table A4: Design of enumeration forms and missing population

	(1) No. of slots	(2) Share of HH below no. of slots	(3) Missing population	(4) Missing population
log GDP per capita	-0.270	2.211***	-0.160*	-0.071
	(0.151)	(0.547)	(0.065)	(0.051)
Share of HH below no. of slots				-0.040***
				(0.012)
N	186	186	186	186
R2	0.018	0.065	0.050	0.274

No. of slots indicates the number of pre-printed slots per enumeration form. Share of HH below no. of slots indicates the share of households that is strictly smaller than the number of pre-printed slots. Missing population indicates the share of missing population. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

A.6 Missing people within countries

A.6.1 Subnational poverty rates

We use data from the Global Multidimensional Poverty Index 2013-2024 (Alkire et al., 2024) to measure the poverty headcount ratio at the subnational region level. This poverty measure is generated from Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS). It is updated whenever a new survey is conducted in a given country. For each census in our data, we match each subnational region with the corresponding subnational region in the last preceding MPI update. This way, we obtain regional poverty information for 178 regions across 16 countries. We cannot match more regions and censuses because either there was no preceding MPI data (the first

surveys underlying the MPI were conducted in 2003) or because the definitions of regions vary between censuses and the DHS/MICS.

A.6.2 Classification of households into rich and poor

We leverage household-level information on asset ownership to categorize households into poor and rich. Following the approach used by the Demographic and Health Survey, we build a wealth index using principal component analysis and group households into wealth quintiles based on the resulting factor score. We restrict our sample to censuses that report ownership of at least three assets, corresponding to a total of 157 censuses.

A.6.3 Missing urban vs. rural within countries

In addition to the heterogeneity in missing populations reported for household wealth, we find that undercounting is more serious in rural populations by comparing the share of missing people between urban and rural households – in line with the higher prevalence of poverty among rural households.

Figure A17 shows that the share of missing rural exceeds the share of missing urban in the 113 out of 148 censuses for which we observe the locality of households. On average, members of rural households are 1.8 times as likely to be missing as urban ones – implying an upward bias in urbanisation rates derived from population censuses.

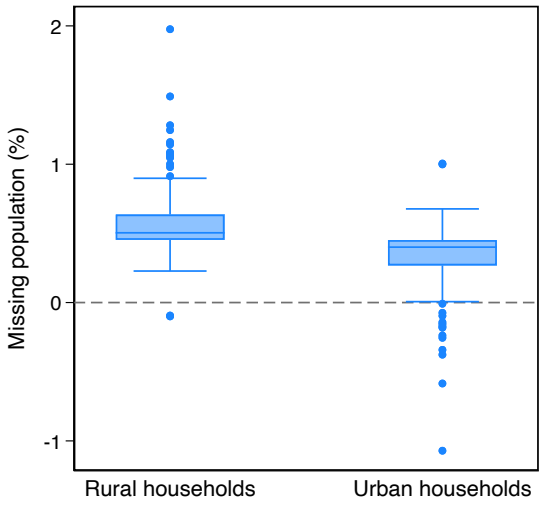


Figure A17: Missing people in rural vs. urban households

This figure illustrates the within-census variation in missing population between urban and rural areas. To isolate the within-census variation, the x and the y variable are regressed on a set of census fixed effects. Then the mean of each variable is added to the residuals from these regressions to generate the x and the y variable for the plot. The figure shows a box plot for 148 out of 188 censuses for which the locality (rural/urban) of the household is observed.

A.7 Implications of census undercounting for household surveys

Selective undercounting also undermines the representation of large households in household surveys. These surveys typically employ a two-stage clustered random sampling approach to generate a representative sample of the population. In a first stage, census enumeration areas are drawn with probability proportional to the population recorded in the last census. In a second stage, a given number of households is randomly drawn in each selected enumeration area. The disproportionate omission of members from larger households in the census thus leads to an underrepresentation of large households in surveys for two reasons. First, the probability of drawing enumeration areas with many large households is downward biased because their population is underestimated by more. Second, the sampling weights for selected enumeration areas with many large households are likewise downward biased.

Underrepresentation in household surveys matters because these surveys provide a key input to national planning – especially in poorer countries where few alternative data sources exist. The disproportionate omission of members from larger household may thus mean that their needs are not adequately reflected in policy design.

A.8 Resource allocation

A.8.1 Local government funding in Tanzania

To quantify the implied funding changes we proceed as follows. First, we estimate the missing population separately for each local government authority (LGA) to construct counterfactual LGA populations. Figure A18 illustrates the resulting estimates across LGAs. Second, we determine the counterfactual population-based budget share of each LGA by dividing the counterfactual LGA population by the counterfactual total population in the country. Finally, we take the difference between the counterfactual and the actual budget share in each LGA and divide this by the actual budget share to determine the relative change in the budget share.

We correlate the relative change in the budget with the poverty headcount across LGAs (Figure 4, panel B in the paper). To this end, we compile LGA-level poverty headcount ratio data from Research and Analysis Working Group (2005).

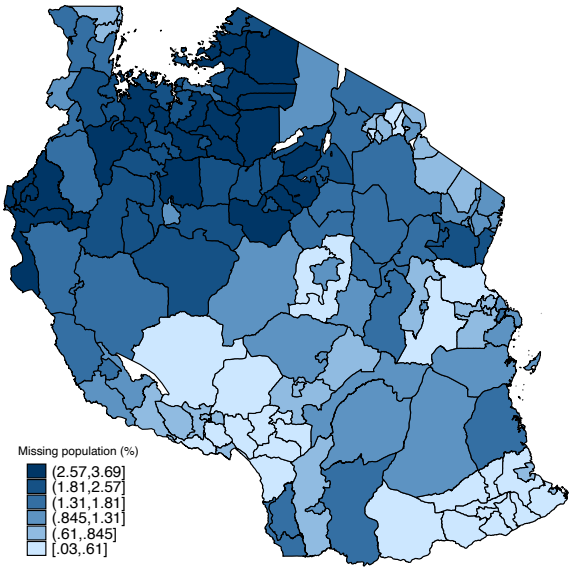


Figure A18: Missing population across local government authorities in Tanzania

This figure shows the estimated missing population by Local Government Authority (LGA) in Tanzania (2012 Census). (Ruggles et al., 2024)

A.8.2 Targeting of infrastructure in Bihar, India

We provide evidence of a direct link between population count and resource allocation at the local level using the example of Bihar, the poorest and third-most populous Indian state with 104m people in 2011.

Similar to many other Indian states and territories, the allocation of several centrally designed signature infrastructure projects in Bihar is determined by village-level population eligibility thresholds. Hence, the omission of household members in Indian population censuses directly affects the number of eligible villages that receive a given project or source of funding.

To assess how missing household members affect village-level infrastructure allocation in Bihar, we estimate the missing population in each of Bihar's 38 districts in a 5% sample of the 2011 population census where the number of pre-printed slots was 8. We find that on average 0.49% of the population is missing, ranging from virtually no missing to 1.32% across districts, with higher missingness in less developed districts with lower literacy rates and less urban population.

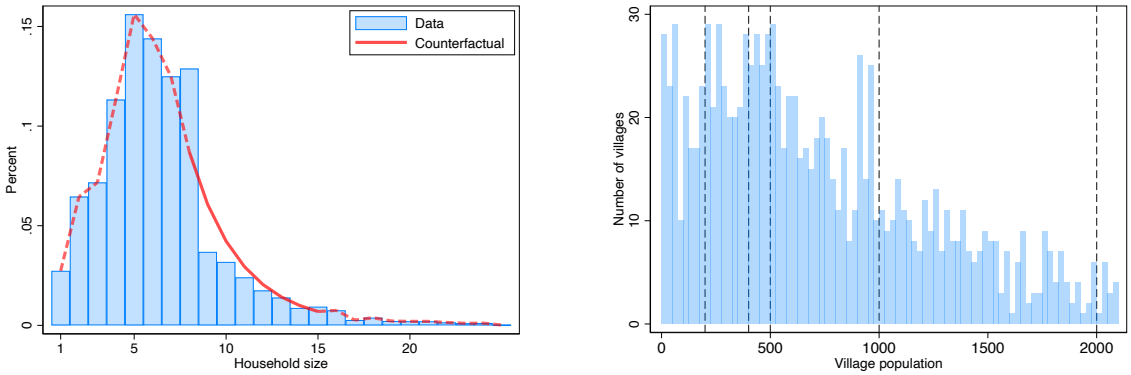
Figure A19a provides the household size distribution and estimated counterfactual for one exemplary district, Kaimur-Bhabua, in Bihar. We estimate the counterfactual district population to be 1% larger than the reported one. As a result, if every village was missing on average 1% of population in this district, Figure A19b shows how many villages had reported populations just below program eligibility thresholds, indicated by dashed lines. Hence, these villages, would have likely received infrastructure if all household members had been recorded.

For example, under the assumption that missing people at the village-level are uncorrelated with village size, at least three more villages should have received a primary school and electricity in Kaimur-Bhabua district, and eight more villages across all 38 districts of Bihar. Once corrected for the district's estimate of missing population, these villages would have reached or crossed the eligibility threshold.

Across all districts in Bihar, 18 more villages should have received a child health centre at 400 people, 30 villages a rural road under Phase II at 500 people, 54 villages a second child health centre at 800 people,

48 villages a rural road under Phase I at 1,000 people, another 66 villages double payments for drinking water supplies at 2,000 people and 40 villages should have received a health sub-centre, as well as higher wastewater treatment payments at 5,000 people. Overall, we estimate that 264 villages in the entire state missed out on key infrastructure investments due to the disproportionate omission of members from larger households.

The concentration of missing populations in poorer districts further implies that the poor are most affected by these wrongful disqualifications. In addition, and in line with results reported in Section IIC of the paper, missing populations across Bihar districts correlate negatively and significantly with literacy (as income proxy), and rural districts have more missing people than urban ones. Especially smaller and more rural villages in the left tail and centre of the village size distribution are affected be the village eligibility changes reported above. We would require the full 100% sample of both the 2001 and 2011 Population Censuses complete with village-level identifiers to highlight which exact villages were erroneously deemed ineligible for these various types of public infrastructure. Since we can only estimate average district-level, not village-level missing, our corrections likely underestimate how many villages should have been eligible. However, even our across-district estimates in Bihar demonstrate the large potential for the local and tangible misallocation of public resources due to enumerator incentives.



(a) Household size distribution

(b) Village size distribution

Figure A19: Missing people and village size thresholds in Kaimur-Bhabua, Bihar

This figure shows how missing people in an exemplary district in Bihar, India can lead to villages failing to meet program eligibility criteria in terms of village population. Panel (a) show the household size distribution for all villages in the district of Kaimur-Bhabua, Bihar, using a 5% sample of the 2011 Population Census. This census features eight pre-printed household roster slots per form. The solid red line plots the resulting counterfactual household size distribution. In panel (b), we plot the village size distribution in the district of Kaimur-Bhabua, Bihar, using 2011 Population Census village-level aggregates. The dashed vertical lines indicate program eligibility thresholds at populations of 250, 400, 500, 1,000 and 2,000. Although we cannot link village-level household sizes to village-level aggregate population, the figure indicates the number of villages just to the left of these thresholds that would have been eligible if no people were missing.

A.9 Parliamentary seat allocation

A.9.1 Bolivia 2013

The Bolivian Chamber of Deputies is composed of 130 seats. These are distributed across nine departments following the *Ley de distribucion de escanos entre departamentos, 7 de octubre de 2013* (Ley N^o421, Articulo 2).

We replicate the allocation procedure prescribed by the law. First, we validate that the actual allocation of seats that resulted from the 2013 delimitation is consistent with the prescribed procedure. Second, we compute the counterfactual allocation using the corrected departmental population counts. The computation is illustrated in Table A5. Leveraging the official population counts for each department from the 2012 population census (Instituto Nacional de Estadística de Bolivia, 2013) and our estimates of missing population (columns 2-3), we derive the counterfactual population (column 4). Then we determine which departments have a population below the average departmental population (indicated in column 5). The ratio of the average departmental population in departments above and below the national average departmental population (0.2) is multiplied by the total number of seats (130) and divided by the number of departments with below national average population (6) to determine the number of seats each department with below national average population receives, yielding 4 seats per small department after rounding (column 6). Next, one seat is assigned to the least development department (column 7) following the actual allocation of seats in 2013 (Cori, 2022). Finally, the remaining 105 seats are allocated across departments proportional to population, using the so-called highest-remainder method where seats are first distributed based on the integer of the representational quotient (total population over number of seats) and subsequently remaining seats go to the departments with the highest remainder in the representational quotient (columns 8-13). The resulting total number of counterfactual seats is displayed in column 14 and compared to the actual allocation in column 15. The final two columns show the resulting change in seat

allocation and the poverty headcount ratio (from the Global Multidimensional Poverty Index 2016) for each department. We find that the departments of La Paz and Cochabamba would each lose one seat and the departments of Beni and Potosi would each gain one seat. While the poverty headcount ratio in both of the losing departments is approximately 8%, it is 9% and 26%, respectively, in the gaining ones.

A.9.2 Sierra Leone 2017

In Sierra Leone, the parliament is composed of 146 seats. 132 of these are allocated across 16 districts proportional to population using the highest remainder method, the remaining 14 go to paramount chiefs (National Electoral Commission, 2017).

We replicate the highest-remainder method to compute the counterfactual allocation of seats of the 132 seats allocated across districts. First, we verify that the actual allocation of seats is consistent with the application of the highest remainder method. Second, we apply the highest remainder method to the counterfactual district populations, as illustrated in Table A6. Leveraging the official population counts for each district from the 2015 population census (Statistics Sierra Leone, 2016) and our estimates of missing population (columns 2-3), we derive the counterfactual population (column 4). Using the highest-remainder method, we then distribute seats across districts (columns 5-10). The resulting counterfactual allocation is compared to the actual allocation (column 11) to assess changes in the allocation of seats (column 12). Finally, column 13 shows the poverty headcount ratio (from the Global Multidimensional Poverty Index 2019) for each district. We find that the district with the lowest poverty headcount ratio (Western Urban, 20.2%) would lose one seat and the district with the second highest poverty headcount ratio (Pujehun, 80.6%) would gain one seat.

Departamiento	Population (PHC 2012)	Missing share (%)	Counterfactual population	Below population avg	Small population seats	Low HDI seats	Representational quotient	Integer	Remainder	Ranking of remainders	Highest remainder distribution	Proportional seats	Counterfactual seat allocation	Actual seat allocation	Change in seat allocation	Poverty headcount ratio (%)
La Paz	2706351	0.512	2720292	0	0	0	28.241	28	0.241	5	0	28	28	29	-1	7.6
Santa Cruz	2655084	1.108	2684832	0	0	0	27.873	27	0.873	1	1	28	28	28	0	4.6
Cochabamba	1758143	0.395	1765116	0	0	0	18.325	18	0.325	4	0	18	18	19	-1	8.2
Potosi	823517	1.133	832958	1	4	1	8.647	8	0.647	2	1	9	14	13	1	25.6
Chuquisaca	576153	1.288	583673	1	4	0	6.059	6	0.059	9	0	6	10	10	0	20.8
Oruro	494178	0.426	496294	1	4	0	5.152	5	0.152	7	0	5	9	9	0	7.1
Tarija	482196	1.136	487739	1	4	0	5.064	5	0.064	8	0	5	9	9	0	5.1
Beni	421196	2.141	430132	1	4	0	4.468	4	0.468	3	1	5	9	8	1	9.0
Pando	110436	2.087	112790	1	4	0	1.171	1	0.171	6	0	1	5	5	0	15.6
TOTAL	10027254		10114104		24	1						105	130	130	0	

Table A5: Counterfactual allocation of parliamentary seats in Bolivia 2013

District	Population (PHC 2015)	Missing share (%)	Counterfactual population	Representational quotient	Integer	Remainder	Ranking of remainders	Highest remainder distribution	Counterfactual seat allocation	Actual seat allocation	Change in seat allocation	Poverty headcount ratio (%)
Kailahun	526379	2.322	538890	9.914	9	0.914	5	1	10	10	0	68.0
Kenema	609891	1.025	616208	11.337	11	0.337	15	0	11	11	0	63.4
Kono	455369	2.153	465387	8.562	8	0.562	11	1	9	9	0	60.4
Bombali	422960	0.697	425929	7.836	7	0.836	6	1	8	8	0	59.3
Falaba	205353	5.050	216275	3.979	3	0.979	1	1	4	4	0	84.6
Koinadugu	204019	5.223	215262	3.960	3	0.960	2	1	4	4	0	60.8
Tonkolili	513984	1.693	522836	9.619	9	0.619	10	1	10	10	0	73.2
Kambia	345474	2.284	353548	6.504	6	0.504	13	0	6	6	0	69.8
Karene	285546	2.206	291986	5.372	5	0.372	14	0	5	5	0	77.9
Port Loko	530865	1.768	540420	9.942	9	0.942	3	1	10	10	0	67.8
Bo	575478	1.066	581679	10.701	10	0.701	9	1	11	11	0	58.9
Bonthe	200781	3.158	207328	3.814	3	0.814	7	1	4	4	0	71.5
Moyamba	318588	1.033	321915	5.922	5	0.922	4	1	6	6	0	73.9
Pujehun	346461	5.494	366602	6.745	6	0.745	8	1	7	6	1	80.6
Western Rural	444270	0.826	447970	8.241	8	0.241	16	0	8	8	0	41.1
Western Urban	1055964	0.638	1062745	19.552	19	0.552	12	0	19	20	-1	20.2
TOTAL	7041382		7174980						132	132	0	

Table A6: Counterfactual allocation of parliamentary seats in Sierra Leone 2017

References

Agence Nationale de la Statistique et de la Démographie (ANSD). (2023). *Recensement général de la population et de l'habitat, 2023*. ANSD, Archivage National des Données du Sénégal. Retrieved July 24, 2025, from https://anads.ansd.sn/index.php/catalog/311

Alkire, S., Kanagaratnam, U., & Suppa, N. (2024). *Global multidimensional poverty index (MPI) 2013–2024*. Oxford Poverty & Human Development Initiative. Retrieved April 10, 2025, from https://ophi.org.uk/global-mpi/2024

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science*, *293*(5536), 1818–1820. https://doi.org/10.1126/science.1062081

Cori, J. (2022, November 21). *TSE aclara que aplicará la fórmula de la Ley 421 para la distribución de escaños entre departamentos*. Agencia Boliviana de Información (ABI). Retrieved July 29, 2025, from https://www.abi.bo/index.php/noticias/politica/30254-tse-aplicara-la-formula-de-la-ley-421-para-la-distribucion-de-escanos-entre-departamentos

Departamento Administrativo Nacional de Estadística (DANE). (2023). *Colombia – censo nacional de población y vivienda 2018 (CNPV 2018)*. Bogotá, Colombia, DANE, Archivo Nacional de Datos. Retrieved July 17, 2025, from https://microdatos.dane.gov.co/index.php/catalog/643

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2025). *Penn World Table version 11.0*. Groningen Growth; Development Centre. https://doi.org/10.34894/FABVLR

Figueiredo Walter, T., Moneke, N., & Radu, A. (2026). *Census Form Characteristics*. https://doi.org/10.3886/E244879V1

Ghana Statistical Service. (2022). *2021 Population and Housing Census* (Version 1.0). Retrieved July 16, 2025, from https://microdata.statsghana.gov.gh/index.php/catalog/110

Hernández-Pérez, R., Angulo-Brown, F., & Tun, D. (2006). Company size distribution for developing countries. *Physica A: Statistical Mechanics and its Applications*, *359*, 607–618. https://doi.org/10.1016/j.physa.2005.04.027

Instituto Nacional de Estadística de Bolivia. (2013). *Censo Nacional de Población y Vivienda 2012: Principales Resultados*. Instituto Nacional de Estadística de Bolivia, Estado Plurinacional de Bolivia. https://ibce.org.bo/images/publicaciones/Resultados-Censo-2012.pdf

Instituto Nacional de Estadística y Censos (INEC). (2023). *Ecuador – VIII censo de población, VII de vivienda y I de comunidades 2022 (Censo Ecuador 2022): Base de datos a nivel de manzana*. Quito, Ecuador, INEC. Retrieved July 17, 2025, from https://www.censoecuador.gob.ec/data-censo-ecuador/

Jiménez, A. (2017, March 24). *Inscríbete como voluntario y se parte de este Censo 2017*. Facultad de Comunicación e Imagen, Universidad de Chile. Retrieved November 12, 2025, from https://fcei.uchile.cl/noticias/131605/inscribete-para-ser-voluntario-en-este-censo-2017

Kenya National Bureau of Statistics. (2019, June 11). *Enumerators at Kenya National Bureau of Statistics (KNBS)*. MyJobMag. Retrieved November 12, 2025, from https://www.myjobmag.co.ke/job/86990/enumerators-kenya-national-bureau-of-statistics-knbs

National Bureau of Statistics (NBS). (2024). *Tanzania population and housing census 2022*. Dodoma, Tanzania, NBS, Tanzania National Data Archive. Retrieved July 10, 2025, from https://microdata.nbs.go.tz/index.php/catalog/45

National Electoral Commission. (2017). *Annual report 2017*. Republic of Sierra Leone. https://epoll.ec.gov.sl/pdfviewer/final-annual-report-2017/

National Institute of Statistics of Rwanda (NISR). (2024). *Rwanda population and housing census 2022 (RPHC 2022)*. Kigali, Rwanda, NISR, Rwanda National Data Archive. Retrieved July 16, 2025, from https://microdata.statistics.gov.rw/index.php/catalog/109

Nzekwe, H. (2019, November 6). *Kenya's Population Census Has Come And Gone But Over 13,000 Census Staff Are Still Unpaid 2 Month After*. WeeTracker. Retrieved November 12, 2025, from https://weetracker.com/2019/11/06/kenya-census-has-13000-unpaid-staff/

Oficina Nacional de Estadística (ONE). (2023–2024). *República dominicana – X censo nacional de población y vivienda 2022 (XCNPV 2022)*. Santo Domingo, República Dominicana, ONE. Retrieved July 17, 2025, from https://www.one.gob.do/datos-y-estadisticas/temas/censos/poblacion-y-vivienda/2022/

Research and Analysis Working Group. (2005). *Poverty and Human Development Report 2005*. Poverty Monitoring System, United Republic of Tanzania. Dar es Salaam, Tanzania, Mkuki na Nyota Publishers. https://www.repoa.or.tz/wp-content/uploads/2020/07/PHDR_2005_full_report.pdf

Ruggles, S., Cleveland, L., Lovaton, R., Sarkar, S., Sobek, M., Burk, D., Ehrlich, D., Lee, J., & Merrill, N. (2024). *Year-Specific Second-Level Geography: Tanzania 2012*. Integrated Public Use Microdata Series, International. Retrieved February 13, 2024, from https://international.ipums.org/international/gis_yrspecific_2nd.shtml

Statistics Sierra Leone. (2016). *2015 population and housing census: Summary of final results - planning a better future*. Statistics Sierra Leone, Government of Sierra Leone. https://www.statistics.sl/images/StatisticsSL/Documents/final-results_-2015_population_and_housing_census.pdf

Statistics South Africa. (2024). *South african census 2022, 10% sample* (Version 1). Cape Town, DataFirst. https://doi.org/10.25828/efbb-k295

The Herald. (2012, August 1). *Zimbabwe: Preparations for Census 2012 Well Underway*. allAfrica. Retrieved November 12, 2025, from https://allafrica.com/stories/201208010242.html

The Zimbabwean. (2012, August 21). *Poor logistics hinder census . . . locals won't trust enumerators*. Retrieved November 12, 2025, from https://www.thezimbabwean.co/2012/08/poor-logistics-hinder-census-locals/

World Bank. (2024). *World bank official boundaries: Countries, admin 0, 10m.* Retrieved January 29, 2026, from https://datacatalog.worldbank.org/search/dataset/0038272/world-bank-official-boundaries