8-2023

Spatial disaggregation of poverty and disability: Application to Tanzania

Tomoki FUJII
Singapore Management University, tfujii@smu.edu.sg

Empirical Economics
https://doi.org/10.1007/s00181-023-02478-8

Spatial disaggregation of poverty and disability: application to Tanzania

Tomoki Fujii[1]

Received: 4 February 2020 / Accepted: 20 July 2023

Abstract
Estimating poverty measures for disabled people in developing countries is often difficult, partly because relevant data are not readily available. We extend the small-area estimation developed by Elbers, Lanjouw and Lanjouw (2002, 2003) to estimate poverty by the disability status of the household head, when the disability status is unavailable in the survey. We propose two alternative approaches to this extension: Aggregation and Instrumental Variables Approaches. We apply these approaches to data from Tanzania and show that both approaches work. Our estimation results show that disability is indeed positively associated with poverty in every region of mainland Tanzania.

Keywords Poverty · Disability · Tanzania · Aggregation · Two-sample instrumental variables estimation

JEL Classification C20 · I10 · I32

1 Introduction

Studies over the last two decades have shown interlinkages between poverty and disability. Such interlinkages emerge from the bidirectional causality between them. On the one hand, disability may lead to poverty for at least four reasons. First, disabled people may suffer from loss of income due to productivity loss and social exclusion.

This research grew out of my consulting work for the World Bank. I benefited from discussion with Chris Elbers, Hans Hoogeveen, Denis Leung, Wietze Lindeboom, Roy van der Weide, and an anonymous referee. An earlier version of this study was supported by SMU Research Grant (07-C208-SMU-007). Usual caveats apply.

✉ Tomoki Fujii
tfujii@smu.edu.sg

1 School of Economics, Singapore Management University, 90 Stamford Road, 178903 Singapore, Singapore

Second, household incomes from non-disabled members may also be reduced as a result of higher demand for time to take care of disabled members. Third, disabled people may incur additional costs such as medical expenses, equipment, adaptations to housing and specialized services (Elwan 1999). As a result, disabled people may require higher expenditures than non-disabled people to maintain a given standard of living (Zaidi and Burchardi 2005). Fourth, disabled people may have limited access to services—including, but not limited to, healthcare and education services—because of, for example, physical or social barriers (Mitra 2004). On the other hand, poverty may cause disability. Poor people are more likely to suffer not only from the lack of adequate food and water but from the lack of adequate and timely healthcare. They may also have to accept more hazardous working conditions and less safe living environment than non-poor people (Yeo and Moore 2003).

Disentangling the bidirectional causality between poverty and disability is methodologically challenging, because it is generally difficult to find exogenous variations in disability or poverty (Grech 2016). However, it is still useful to understand the strength of the correlation between poverty and disability, as disability is a significant issue in scale and closely related to poverty (World Health Organization 2011; United Nations 2019), given that there is a dearth of empirical evidence on the relationship between poverty and disability, particularly at spatially disaggregated level.[1]

This limitation primarily comes from the lack of relevant data on poverty and disability. Poverty analysis normally relies on a socioeconomic survey that is representative at a highly aggregated level such as regions and provinces. Such survey data typically contain, if any, only a small number of people with disability, because the proportion of disabled people to the total population can be as low as a few percent.[2] This makes it difficult to derive a reliable estimate of poverty for disabled people. As a result, it is difficult to clearly see the correlation between poverty and disability within a country.

Nevertheless, existing cross-country studies tend to find a positive correlation between poverty and disability (Braithwaite and Mont 2009; Masset and White 2004; Filmer 2008). Likewise, a small number of within-country studies such as Mont and Cuong (2011) and Loyalka et al. (2014) also suggest a positive relationship between poverty and disability, particularly when the cost of disability is taken into account.[3]

To our knowledge, Tanzania—the country studied in this paper—has only one published study that directly addresses the correlation between poverty and disability, which is Mitra (2018), who compares poverty across different levels of functional disabilities at the national level and find that more severely disabled people tend to be poorer. Besides Mitra (2018), there are at least two studies in Tanzania relevant to this study. Using data from the Kagera region, those with disability at the baseline in 1993 were found to have lower asset values in 2004 than those without disability at the baseline. The former is also less likely to experience welfare improvement than

[1] See Elwan (1999); Haveman and Wolfe (2000), and Yeo and Moore (2003) and studies cited therein for earlier studies on this and related topics.

[2] The proportion of disabled people would also depend on the way it is defined and measured. See Online Appendix for a related discussion on this issue.

[3] See Mitra et al. (2017) for a review of literature on extra cost of disability (i.e., additional income that disabled people will need to achieve a given level of standards of living).

the latter by either peer assessment or self-assessment (de Weerdt 2010). Further, the risk of death for visually impaired people over the age of 40 is found to be 3.33 times higher than individuals with normal eyesight, after controlling for age, sex, and village of residence (Taylor et al. 1991). These findings are consistent with the positive correlation between poverty and disability in Tanzania and serve as a locally-relevant motivation for the current study.

The goal of this study is to provide a spatial disaggregation of the estimates of poverty and disability in Tanzania. Our contributions are twofold. The first contribution is methodological. We extend the small area estimation (SAE) developed by (Elbers et al. (2002, 2003), hereafter ELL) which combines a survey and a census to produce spatially disaggregated estimates of poverty and other welfare measures. As elaborated later, the ELL SAE method requires all covariates in the consumption model to be included both in the census and survey, or in auxiliary datasets that can be merged into both the census and survey.[4] Hence, unless we naïvely assume that one single consumption model applies to both disabled and non-disabled groups, the ELL method does not work in the absence of disability information in either the survey or census data.

We resolve this issue by extending the ELL method to allow for the presence of regressors unobserved in the survey, such as the disability status of the household head. We propose two alternative approaches to this extension. The first approach uses aggregation in a spirit similar to Feige and Watts (1972). This method is useful because disability information from another source can often be merged into the survey at an aggregated level. In the second approach, we adopt the two-sample two-stage least squares regression (Angrist and Krueger 1992; Inoue and Solon 2010). By choosing an appropriate instrument, we can estimate the consumption model for both disabled and non-disabled groups even when the survey lacks disability information. The empirical results from Tanzania suggest that both methods work reasonably well. To our knowledge, this is the first study that combines the ELL method with two-sample instrumental variables estimation.

The second contribution is empirical. We show that people residing in households with disabled heads are significantly poorer than the rest of the population in every region of Tanzania. This finding is empirically relevant, because there is no published study that quantitatively explores the regional-level relationship between poverty and disability in Tanzania. Further, this is among a small number of existing studies that describe the relationship between poverty and disability with some spatial disaggregation. The level of disaggregation we provide is finer than existing studies.[5]

It should be noted that this is *not* the first study to apply the ELL method to the analysis of poverty and disability. Hoogeveen (2005) applied the ELL method to find the

---

[4] For example, community-level indicators are often merged into the census and survey. Once covariates from auxiliary datasets are merged into the census and survey, we regard these covariates to be part of the census and survey datasets for our purpose.

[5] Both Mont and Cuong (2011) and Loyalka et al. (2014) provide poverty estimates by disability status only for urban and rural areas in Vietnam and China, respectively. Hoogeveen (2005) discussed below reports poverty rates by the disability status only for four regions. Mont and Nguyen (2018) provide poverty measures at the level of six regional estimates. We report poverty rates by the disability status of the household heads for 21 regions.

poverty rates for disabled households—or households headed by disabled people—and non-disabled households in urban Uganda. As with this study, Hoogeveen (2005) did not observe the disability status in the survey. Thus, he just used a single consumption model for disabled and non-disabled households, but he introduced interaction terms between several household-level covariates and the fraction of disabled people in the community in the consumption model to account for the potential heterogeneity due to disability. His analysis shows that poverty rates for disabled households are 7.4−12.6 percentage points higher than those for non-disabled households, depending on the region (Hoogeveen 2005, Table 3).

While Hoogeveen (2005) made an important first step towards deriving poverty statistics for disabled households in the absence of reliable disability data in the survey, there are two important limitations in his study. First, as acknowledged in the study, the introduction of interaction terms does not eliminate the bias in the estimation. Second, the standard errors reported in his study are also likely problematic. This is because Hoogeveen (2005) essentially takes the community-level prevalence of disability as a proxy for the household-level disability status, but no adjustment is made to account for the fact that the community-level prevalence is just a proxy. The current study addresses both of these issues.[6]

This paper is organized as follows. In Sect. 2, we develop the methodology, followed by the description of the data and discussion on measurement issues in Sect. 3. Section 4 presents the empirical results. We offer some discussion in Sect. 5.

2 Methodology

In this section, we formally develop the methodology. We first introduce the ELL SAE method. Then, we discuss the issues we would face when applying the standard ELL method in the absence of disability information in the survey. Finally, we extend the ELL method by two alternative approaches.

2.1 ELL SAE method

The essence of the ELL SAE method is simple. It combines a survey and a census through a consumption regression model. The model parameters are estimated with a survey, and the estimated parameters are used to predict the consumption for each household in the census. The predicted consumption is then aggregated to calculate welfare indicators (e.g., poverty and inequality measures) for small areas. Monte-Carlo simulation is carried out to account for the model error—error associated with the estimation of model parameters—and the idiosyncratic error. The model error systematically affects all predictions. Unlike the model error, the idiosyncratic error tends to cancel out with each other when we aggregate (Elbers et al. 2002, 2003).

[6] Mont and Nguyen (2018) also apply the ELL method to derive poverty measures for people with and without disability at a spatially disaggregated level. However, it does not allow for the difference in model parameters between these two groups.

By combining a survey with a census, the ELL method addresses a particular type of missing data problem. That is, a typical consumption survey has no observation for a majority of small administrative units such as towns and villages. Even for some small administrative units for which there are observations, the number of observations can be limited (only 20 households or so per village is quite common). On the other hand, a census typically covers all or nearly all households in each small administrative unit in a country. ELL method combines the two data sources to address the missingness of the consumption data for small administrative units.

We adopt the ELL method in this study for a similar reason. A typical survey contains only a small number of disabled households, even if the disability status of individuals is known. Therefore, it is generally difficult to estimate poverty for disabled households. Further, even if it is possible to estimate poverty for disabled households with a survey (because of oversampling of disabled households), it is typically not possible to produce meaningful spatial disaggregation. We overcome this problem by combining a survey with a census, the latter of which contains a sufficiently large number of disabled households even at a spatially disaggregated level.

However, we are unable to apply the ELL method directly, because the survey data do not contain individual status of disability in our application. Therefore, we propose an extension of the ELL method to include a regressor that is not observed in the survey. We will discuss this extension in the subsequent subsections.

The ELL method has been applied to dozens of countries. The spatially disaggregated poverty estimates are typically presented in the form of maps, which are called poverty maps. Besides producing poverty maps, the ELL method and its variants have been used to analyze geographic targeting (Elbers et al. 2007; Fujii 2008), inequality (Elbers et al. 2004; Demombynes and Özler 2005; Araujo et al. 2008), and regression analysis at aggregated levels (Elbers et al. 2005), among others. It has also been applied to health-related issues such as the estimations of child undernutrition (Fujii 2010; Sohnesen et al. 2017), health worker absenteeism (Fujii 2019), and HIV prevalence (Ivaschenko and Lanjouw 2010). As noted earlier, Hoogeveen (2005) directly used the ELL method to derive poverty estimates for disabled households.

Let us now formally describe the ELL method. Suppose that the set of clusters (e.g., villages) in a sample is $\mathcal{K} = \{1, \cdots, K\}$, where K is the number of clusters. We denote the (non-empty) set of households in cluster $k \in \mathcal{K}$ by $\mathcal{H}_k$ and denote the number of households in cluster k by $H_k \equiv \#\{\mathcal{H}_k\}$, where $\#\{\cdot\}$ is the counting measure. The set of all households in the sample is denoted by $\mathcal{H} \equiv \cup_{k\in\mathcal{K}}\mathcal{H}_k$ and the total number of households by $N \equiv \#\{\mathcal{H}\}(= \sum_k H_k)$ in the sample. The cluster membership function $\kappa : \mathcal{H} \to \mathcal{K}$ maps each household to the cluster it belongs to, so that $\kappa(h) = k \Leftrightarrow h \in \mathcal{H}_k$. We denote the measure of standard of living (e.g., logarithmic consumption per capita) and the weight for household $h(\in \mathcal{H})$ by y_h and w_h, respectively. In our empirical application, they are respectively logarithmic consumption per adult equivalent and the population expansion factor. When necessary, we use the dataset subscript $D \in \{C, S\}$ with a slight abuse of notation to specify the data source, where C and S respectively denote the census and survey, respectively. For example, $H_{k,D}$ and $\mathcal{H}_D$ respectively denote the number of households in cluster k and the set of all households in the dataset D.

As noted above, the ELL method combines two datasets, a census and a survey. We assume that the census covers all clusters and households in the population of interest. The survey households are assumed to be selected in the following two steps. In the first step, a given fraction of clusters is randomly drawn from the set of all clusters. In the second step, a given fraction of survey households are randomly drawn from each of the chosen clusters. We hereafter assume that the latter fraction is sufficiently large so that the cluster averages in the survey and census are sufficiently close.

The goal of the ELL method is to produce an estimate of aggregate welfare measure $P_J \equiv P(\{y_h\}_{h\in J}, \{w_h\}_{h\in J})$ for each set J of households that is of interest. Typically, each set J contains a set of households located in a given administrative unit. In our case, J is determined by both the disability status of the household head and region in which the household is located. To simplify our presentation, we focus on the case where P_J is poverty rate, which is the proportion of the people under a given poverty line ζ and can be written as: $P_J = \sum_{h\in J} w_h \mathbf{1}(y_h \leq \zeta)/\sum_{h\in J} w_h$. Poverty rate, also called head count index, is the most widely used measure in poverty analysis including common applications of the ELL method. It is simple and a special case of the Foster-Greer-Thorbecke (FGT) class of poverty measures (Foster et al. 1984).

It should be reiterated here that the survey may not have any observation of y_h in some J, in which case P_J cannot be calculated from the survey alone. However, because the census covers the households in J, we can use a predicted value of y_h for each census household $h \in J$ to estimate P_J. To predict y_h in the census, we use a row L-vector of household-level covariates $\boldsymbol{x}_h = (x_{h,1}, \cdots, x_{h,L})$, which is contained both in the census and survey. Covariates that are routinely used for $\boldsymbol{x}_h$ include demographic characteristics, housing conditions, education, and asset holdings of the household. Some community-level variables are often included as well.

We assume that the conditional expectation $E[y_h|\boldsymbol{x}_h]$ is a linear combination of $\boldsymbol{x}_h$ so that $E[y_h|\boldsymbol{x}_h] = \boldsymbol{x}_h\boldsymbol{\beta}$. Further, we also assume that the error term $u_h \equiv y_h - E[y_h|\boldsymbol{x}_h]$ can be expressed as a sum of a cluster-specific random effect $\eta_{\kappa(h)}$ and a household-specific random effect ϵ_h. Thus, we have the following model:

$$\begin{aligned} y_h &= E[y_h|\boldsymbol{x}_h] + u_h \\ &= \boldsymbol{x}_h\boldsymbol{\beta} + \eta_{\kappa(h)} + \epsilon_h. \end{aligned} \tag{1}$$

Both components of the error term have a zero expectation such that $E[\eta_{\kappa(h)}] = E[\epsilon_h] = 0$ for each h. Further, η_k and ϵ_h are assumed to be independently and identically distributed across k and h, respectively. They are also independent with each other and uncorrelated with $\boldsymbol{x}_h$ and w_h. In a standard application of the ELL method, the heteroskedasticity of ϵ_h is typically modelled with a logistic regression and η_k is assumed to be homoskedastic. The regression coefficients and their associated covariance matrix are estimated through a (feasible) generalized least-squares (GLS) regression using the survey. In addition, the distributions of η_k and ϵ_h are also estimated. Note that eq. (1) simply describes the conditional expectation and should not be interpreted as a causal model.

Once all relevant parameters are estimated, y_h is repeatedly imputed to each census household $h \in \mathcal{H}_C$ in a Monte-Carlo simulation. Let R be the total rounds of the simulation. The regression coefficient $\tilde{\boldsymbol{\beta}}^{(r)}$ for the rth round of the simulation for $r \in \{1, \cdots, R\}$ is randomly drawn from the estimated distribution of the GLS estimator $\hat{\boldsymbol{\beta}}_{GLS}$. This random draw captures the model error. In addition, the cluster-specific random effect $\tilde{\eta}_k^{(r)}$ and the household-specific random effect $\tilde{\epsilon}_h^{(r)}$ are drawn, respectively, for each cluster k and each household h from their estimated distributions. These draws capture the idiosyncratic error. With all these random draws, the imputed measure of the standard of living for household h in the rth round of the simulation is calculated as: $\tilde{y}_h^{(r)} = \boldsymbol{x}_h \tilde{\boldsymbol{\beta}}^{(r)} + \tilde{\eta}_{\kappa(h)}^{(r)} + \tilde{\epsilon}_h^{(r)}$. The poverty rate in J for the rth round is then estimated by $\tilde{P}_J^{(r)} = P(\{\tilde{y}_h^{(r)}\}_{h \in J}, \{w_h\}_{h \in J})$. The point estimate $\hat{P}_J$ and its estimated standard error $\widehat{\text{s.e.}}(\hat{P}_J)$ are obtained by taking the average and standard deviation of $\tilde{P}_J^{(r)}$ over r. Other aggregate poverty measures can be estimated in a similar manner.

2.2 Applying the ELL SAE method to disability

Now, let us consider an application of the ELL SAE method to a situation where the model parameters may differ across different groups. These groups can be based on some household characteristics such as the disability status of the household head. We focus on a situation where the group membership is observed in the census but not in the survey. We do not consider the case where the group membership is observed both in the census and survey, because the standard ELL method is directly applicable in this case. We also do not consider the case where the group membership is missing in the census, because the predicted consumption evidently cannot reflect the heterogeneity across groups.

When the survey does not contain the group information, a naïve approach would be to ignore the heterogeneity of model parameters. However, this approach is likely to result in inconsistent estimates. In this subsection, we discuss this issue and the main idea behind the two alternative approaches we propose to address the issue. The details of these approaches are presented in the next subsection.

Suppose that there are G groups, and let a_{hg} be the group membership indicator, which takes one if household h belongs to group $g \in \{1, \cdots, G\}$ and zero otherwise. Each household belongs to exactly one group. We denote the row G-vector of group membership indicators for household h by $\boldsymbol{a}_h = (a_{h1}, \cdots, a_{hg})$, and the $(N \times G)$-matrix of group membership indicators for all households by $\boldsymbol{A} (\equiv (a_{hg}))$.

Let us assume that a row L^e-vector of ("common") regressors $\boldsymbol{x}_h^e$ has a coefficient vector $\boldsymbol{\beta}^e$ that is equal across groups, whereas a row L^d-vector of ("uncommon") regressors $\boldsymbol{x}_h^d$ has a group-specific coefficient vector $\boldsymbol{\beta}_g^d$. Then, by letting $\boldsymbol{x}_h \equiv (\boldsymbol{x}_h^e, \boldsymbol{a}_h \otimes \boldsymbol{x}_h^d)$, the regression coefficient $\boldsymbol{\beta}$ for $\boldsymbol{x}_h$ is a column L-vector for $L \equiv L^e + L^d G$, which stacks $\boldsymbol{\beta}^e, \boldsymbol{\beta}_1^d, \cdots, \boldsymbol{\beta}_G^d$ into one vector. Hence, if $\boldsymbol{x}_h^e$, $\boldsymbol{a}_h$ and $\boldsymbol{x}_h^d$ are all observed in both the census and the survey, $\boldsymbol{x}_h$ exists in both samples so that the standard ELL method can be used.

However, it is not the case when the group membership vector $\boldsymbol{a}_h$ is not observed in the survey. One quick fix is to simply assume that the coefficients are equal across groups so that $\boldsymbol{\beta}_1^d = \cdots = \boldsymbol{\beta}_G^d$. This assumption is clearly problematic when the coefficients are indeed different among different groups. The estimates of $\boldsymbol{\beta}^d$ under this assumption would only reflect the relationship between $\boldsymbol{x}^d$ and y_h averaged over different groups. Hence, unless the coefficients are indeed equal across groups, the predicted consumption given the group indicators will be inconsistent.

Hoogeveen (2005) attempts to mitigate the bias by including a number of interaction terms between the community-level prevalence of disability (taken from the census) and some household-level variables. This is in effect equivalent to replacing $\boldsymbol{a}_h$ with its cluster average $\bar{\boldsymbol{a}}_{\kappa(h)}$ (and its interactions), where $\bar{\boldsymbol{a}}_{\kappa(h)}$ is defined as follows:

$$\bar{\boldsymbol{a}}_{\kappa(h)} \equiv \frac{\sum_{h' \in \mathcal{H}_{\kappa(h)}} a_{h'} w_{h'}}{\sum_{h' \in \mathcal{H}_{\kappa(h)}} w_{h'}}.$$

His approach would capture some of the variations across groups since $\bar{\boldsymbol{a}}_{\kappa(h)}$ serves as a proxy for $\boldsymbol{a}_h$. However, his method does not account for the fact that $\bar{\boldsymbol{a}}_{\kappa(h)}$ is just a proxy. Further, the potential heteroskedasticity across groups are ignored, which in turn leads to inconsistent estimate of standard errors. These issues are particularly important when we are interested in comparisons across groups.

In order to highlight these issues and present the ideas for possible solutions, let us consider a simplified version of the ELL estimation in the remainder of this subsection. We assume that there are only two groups (i.e., $G = 2$): disabled and non-disabled households. We also maintain this assumption in our empirical analysis, even though we present our approaches in a more general setting in the next subsection. We let a_{h1} and $a_{h2}(= 1 - a_{h1})$ denote the indicator variables for non-disabled and disabled households, respectively.

Furthermore, we assume that there is no common regressor (i.e., $L^e = 0$), and the uncommon regressor contains only a constant term (i.e., $L^d = 1$ and $\boldsymbol{x}_h^d = 1$). Each household has a unit weight. There is no cluster-specific random effect, and the heteroskedasticity of ϵ_h depends only on the group household h belongs to. Hence, we have $\mathrm{var}[u_h] = \mathrm{var}[\epsilon_h] = a_{h1}\sigma^2_{\epsilon,1} + a_{h2}\sigma^2_{\epsilon,2}$, where $\sigma^2_{\epsilon,g} \equiv \mathrm{var}[\epsilon_h | a_{hg} = 1]$ is the group-specific variance of ϵ_h for $g \in \{1, 2\}$. Under these assumptions, eq. (1) reduces to:

$$y_h = a_{h1}\beta_1 + a_{h2}\beta_2 + \epsilon_h. \tag{2}$$

If we observe $\boldsymbol{a}_h$ for every household h in the survey sample $\mathcal{H}_S$, we simply need to regress y_h on $\boldsymbol{a}_h$ to obtain an unbiased estimate of $\beta_g = E[y_h | a_{hg} = 1]$ for $g \in \{1, 2\}$. This is, of course, equivalent to taking the sample mean of y_h for group g (i.e., $\sum_{h \in \mathcal{H}_S} y_h a_{hg} / \sum_{h \in \mathcal{H}_S} a_{hg}$). This operation is not feasible when the group indicators $\boldsymbol{a}_h$ are not observed in the survey.

Suppose that we observe the cluster-level average of group indicators $\bar{\boldsymbol{a}}_{\kappa(h)}$ for each household h in the survey and that we apply the method used by Hoogeveen (2005) to our simplified setup. Then, y_h would be regressed on $\bar{\boldsymbol{a}}_{\kappa(h)}$ using the survey and predicted for each census household with $\bar{\boldsymbol{a}}_{\kappa(h)}$. In this case, $\hat{\boldsymbol{\beta}}_{GLS}$ is in general inconsistent. Furthermore, even if the true $\boldsymbol{\beta}$ were known, the predicted value $\hat{y}_h =$

$\bar{a}_{\kappa(h),1}\beta_1 + \bar{a}_{\kappa(h),2}\beta_2$ would still be biased, because we have $\hat{y}_h \neq E[y_h|a_{h1}, a_{h2}] = a_{h1}\boldsymbol{\beta}_1 + a_{h2}\boldsymbol{\beta}_2$ in general.

Besides, Hoogeveen (2005) also failed to take into account the potential heteroskedasticity across different groups. In our simplified model, this is equivalent to assuming that the estimated variance $\widehat{\text{var}}[\hat{y}_h]$ is identical across groups. However, this is unlikely to hold in practice. Consider, for example, the presence of local supporters for disabled households, which would positively affects the standards of living of disabled households. If the presence of the local supporters is unobservable, we are forced to treat it as a household-specific random effect. Since this effect affects only disabled households, ϵ_h is likely to be heteroskedastic across groups.

The importance of the above-mentioned issues depends on several factors in practice. First, if each cluster is dominated by one group so that $\bar{a}_{kg}$ is approximately equal to zero or one for each k and g, then the bias in the estimate of $\boldsymbol{\beta}$ is likely to be small. This is because $\bar{a}_{\kappa(h)g}$ is a very good proxy for a_{hg}. However, the bias in the predicted value of y_h for small minority groups may be still severe, because the prediction is based on $\bar{\boldsymbol{a}}_k$. Second, if $\boldsymbol{x}_h^e$ captures most of the variations of y_h, the bias in the predicted value of y_h is likely small, because the heterogeneity across groups is relatively unimportant. Third, when the variance of η_k is large relative to that of ϵ_h, a large portion of the idiosyncratic errors for $\hat{P}_J$ would come from the cluster-specific random effects. This is because cluster-specific random effects diminish only slowly by aggregation in comparison with household-specific random effects. As a result, ignoring the heteroskedasticity of ϵ_h across groups would do little harm, because the household-specific random effect is a relatively unimportant source of errors. Otherwise, failure to account for the heteroskedasticity due to groups may lead to a serious underestimation of standard errors for certain groups.

The preceding discussion motivates the two approaches formally developed in the next subsection: Aggregation Approach and Instrumental Variables Approach. In the Aggregation Approach, we aggregate all covariates, including the group indicators, to the cluster level, because their cluster-level averages are all observed at this level. As a result, we can run regressions at this level and obtain a consistent estimate of $\boldsymbol{\beta}$. To provide an intuition behind this approach, consider the cluster-level average of Eq. (2):

$$\bar{y}_k = \bar{a}_{k,1}\boldsymbol{\beta}_1 + \bar{a}_{k,2}\boldsymbol{\beta}_2 + \bar{\epsilon}_k, \tag{3}$$

where $\bar{y}_k (\equiv H_k^{-1} \sum_{h\in\mathcal{H}_k} y_h)$ and $\bar{\epsilon}_k (\equiv H_k^{-1} \sum_{h\in\mathcal{H}_k} \epsilon_h)$ are the cluster-level averages of y_h and ϵ_h. Note that we have $\text{var}[\bar{\epsilon}_k] = H_k^{-1}\left(\bar{a}_{k,1}\sigma_{\epsilon,1}^2 + \bar{a}_{k,2}\sigma_{\epsilon,2}^2\right)$. Since we observe $\bar{y}_k$, $\bar{a}_{k,1}$ and $\bar{a}_{k,2}$ for each k, it is straightforward to estimate eq. (3) by a (feasible) GLS regression at the cluster level.

Let us now turn to the Instrumental Variables Approach. In this approach, we treat $\bar{\boldsymbol{a}}_{\kappa(h)}$ as an instrument for $\boldsymbol{a}_h$. To demonstrate the main idea of this approach, let us premultiply eq. (2) by the vector of instruments $(\bar{a}_{\kappa(h),1}, \bar{a}_{\kappa(h),2})^T$, sum over h, and divide by N.

$$\underbrace{\frac{1}{N}\sum_{h\in\mathcal{H}}\begin{pmatrix}\bar{a}_{\kappa(h),1}y_h\\ \bar{a}_{\kappa(h),2}y_h\end{pmatrix}}_{\equiv \boldsymbol{M}_0^N} = \underbrace{\frac{1}{N}\sum_{h\in\mathcal{H}}\begin{pmatrix}\bar{a}_{\kappa(h),1}a_{h1} & \bar{a}_{\kappa(h),1}a_{h2}\\ \bar{a}_{\kappa(h),2}a_{h1} & \bar{a}_{\kappa(h),2}a_{h2}\end{pmatrix}}_{\equiv \boldsymbol{M}_1^N}\boldsymbol{\beta} + \frac{1}{N}\sum_{h\in\mathcal{H}}\begin{pmatrix}\bar{a}_{\kappa(h),1}\epsilon_h\\ \bar{a}_{\kappa(h),2}\epsilon_h\end{pmatrix} \tag{4}$$

Let $\boldsymbol{M}_0^N$ and $\boldsymbol{M}_1^N$ denote, respectively, the sample moment on the left hand side of the equality and the coefficient of $\boldsymbol{\beta}$ on the right hand side of equality when the sample size is N. The moment $\boldsymbol{M}_1^N$ is the sample cross-moment between $\bar{\boldsymbol{a}}_{\kappa(h)}$ and $\boldsymbol{a}_h$. We denote the corresponding population moments by $\boldsymbol{M}_i$ for $i \in \{0, 1\}$ where $\boldsymbol{M}_i^N \xrightarrow{p} \boldsymbol{M}_i$ as $N \to \infty$. Assuming $\boldsymbol{M}_1$ is non-singular, premultiplying $(\boldsymbol{M}_1^N)^{-1}$ and letting $N \to \infty$, we have $(\boldsymbol{M}_1^N)^{-1}\boldsymbol{M}_0^N \xrightarrow{p} \beta$. This suggests that we could use $(\boldsymbol{M}_1^N)^{-1}\boldsymbol{M}_0^N$ as an estimator of $\boldsymbol{\beta}$. However, we cannot compute $(\boldsymbol{M}_1^N)^{-1}$ from the survey, because the survey data do not contain $a_{\kappa(h)}$. Therefore, we instead use the same cross-moment from the census. We can compute this, because the census contains both $\bar{\boldsymbol{a}}_{\kappa(h)}$ and $\boldsymbol{a}_h$ for each household h in the population.

It is instructive to underscore some important differences between Hoogeveen (2005) and this study. The former takes the cluster-level averages of group indicators $\bar{\boldsymbol{a}}_{\kappa(h)}$ as if they were the group indicators $\boldsymbol{a}_h$, ignoring the fact that $\bar{\boldsymbol{a}}_{\kappa(h)}$ is essentially a proxy for $\boldsymbol{a}_h$. On the other hand, the latter clearly distinguishes $\bar{\boldsymbol{a}}_{\kappa(h)}$ from $\boldsymbol{a}_h$. In the Aggregation Approach, standard errors are computed by taking into account the aggregation involved in the estimation. Similarly, the Instrumental Variables Approach explicitly takes account of the fact that $\bar{\boldsymbol{a}}_{\kappa(h)}$ is just a proxy for $\boldsymbol{a}_h$. As a result, the cross-moment $\boldsymbol{M}_1$ between $\bar{\boldsymbol{a}}_{\kappa(h)}$ and $\boldsymbol{a}_h$, which captures how strongly they are correlated, is used in the estimation of the standard errors.

The discussion above also implies that the usefulness of the two approaches depends on the context. The Aggregation Approach would be most useful when there is a relatively large heterogeneity across clusters in the distribution of $\boldsymbol{x}$ but not within clusters. The Instrumental Variables Approach would be most useful when a good proxy for $\boldsymbol{a}_h$ is available.

2.3 Extending ELL SAE method

We now formally introduce the two alternative approaches. To this end, we first describe the data generating process and make some assumptions.

We assume that each cluster k is characterized by a couple (H_k, η_k), which is assumed to be independently and identically distributed across k. H_k is a positive integer and bounded from above and orthogonal to the cluster-specific effect η_k. To explicitly incorporate clustering, we will establish consistency as $K_C \to \infty$ instead of $N_C \to \infty$. However, we clearly have $N_C \to \infty$ as $K_C \to \infty$. As the survey selects a fixed fraction of census clusters and a fixed (and sufficiently large) fraction of the census households in the selected clusters, we also have $K_S \to \infty$ and $N_S \to \infty$ as $K_C \to \infty$.

Each household $h \in \mathcal{H}_k$ in cluster k is characterized by a sextuple $\boldsymbol{\omega}_h \equiv (\boldsymbol{x}_h^e, \boldsymbol{x}_h^d, y_h, \boldsymbol{a}_h, \tilde{\boldsymbol{a}}_h, \epsilon_h)$, where $\tilde{\boldsymbol{a}}_h$ is a G-row vector of the proxy variables for $\boldsymbol{a}_h$.

We denote all the regressors—which include the common regressors and the interaction between uncommon regressors and the group indicators—by $\boldsymbol{x}_h \equiv (\boldsymbol{x}_h^e \quad \boldsymbol{a}_h \otimes \boldsymbol{x}_h^d)$. Further, we denote the instrumental variables for $\boldsymbol{x}_h$ by $\boldsymbol{z}_h \equiv (\boldsymbol{x}_h^e \quad \tilde{\boldsymbol{a}}_h \otimes \boldsymbol{x}_h^d)$. Note here that $\boldsymbol{x}_h$ and $\boldsymbol{z}_h$ have the same dimension by construction. We assume that eq. (1) is satisfied for each household and that $\boldsymbol{\omega}_h$ is independently distributed across clusters (but not within clusters). In the census, both $\boldsymbol{x}_h$ and $\boldsymbol{z}_h$ are observed, but y_h is unobserved. In the survey, $\boldsymbol{z}_h$ and y_h are observed but not $\boldsymbol{a}_h$. Further, while the survey data do not include $\boldsymbol{a}_h$, its cluster-level average $\bar{\boldsymbol{a}}_k$ for each survey cluster $k \in \mathcal{K}_S$ can be obtained from the census. This condition is satisfied in Tanzania and many other countries. Finally, each household h has a sample weight w_h, which is normalized so that the sum of weights is equal to the number of observations, or $N_D = \sum_{h \in \mathcal{H}_D} w_h$.

We make the following assumptions in the remainder of this paper:

A1 η and ϵ follow a one-parameter mean-zero distribution with summation-reproducible property, where the parameter represents the variance.

A2 $\mathrm{var}[\epsilon_h] = \sum_g a_{hg} \sigma^2_{\epsilon,g}$.

In a typical application of the ELL method, the OLS residual $\hat{u}_h$ in eq. (1) is decomposed into the cluster-specific effect and the household-specific effect. The former is estimated as the cluster-level average of the residual $\hat{\eta}_k \equiv (N_{k,S})^{-1} \sum_{h \in \mathcal{H}_{k,S}} \hat{u}_h$, and the latter as the remainder $\hat{\epsilon}_h \equiv \hat{u}_h - \hat{\eta}_{\kappa(h)}$. Hence, the distributions of η_k and ϵ_h are estimated separately in the standard ELL method. However, as elaborated below, this decomposition is infeasible in our study. Hence, we require Assumption **A1**. For example, if both ϵ_h and η_k are normally distributed with mean zero, this assumption is satisfied.

Assumption **A2** says that ϵ_h is heteroskedastic across groups but not within each group. In other words, the variance of ϵ_h depends only on the group that the household belongs to. This specification allows for the heteroskedasticity of the household-specific random effects across groups—which may be due to the presence of local supporters for disabled people, for example.

It is useful to introduce some matrix notations here to simplify our presentation. For each $D \in \{C, S\}$, let $\boldsymbol{Y}_D$ and $\boldsymbol{U}_D$ denote N_D-vectors of y_h and u_h respectively. We let $\boldsymbol{X}_D$ and $\boldsymbol{Z}_D$ be $(N_D \times L)$-matrices of entire observations of $\boldsymbol{x}_h$ and $\boldsymbol{z}_h$, respectively. Further, we let $\boldsymbol{W}_D \equiv \mathrm{diag}(w_1, \cdots, w_{N_D})$ be a diagonal $(N_D \times N_D)$-matrix, whose (k, k) element is w_k, and $\boldsymbol{\Lambda}_D$ be a $(N_D \times K)$-matrix of cluster membership, whose (h, k) element is $\mathbf{1}(\kappa(h) = k)$. Note that eq. (1) can be written with the matrix notations as follows:

$$\boldsymbol{Y}_D = \boldsymbol{X}_D \boldsymbol{\beta}_D + \boldsymbol{U}_D, \quad D \in \{C, S\}. \tag{5}$$

We now introduce the two approaches—Aggregation and Instrumental Variables Approaches—to estimate $\boldsymbol{\beta}$.

Aggregation approach

As discussed earlier, Eq. (1) can be estimated at an aggregated level, because the cluster-level averages of $\boldsymbol{x}_h$ and y_h can be computed for each survey cluster. The

Aggregation Approach proposed here is closely related to Feige and Watts (1972), who investigate the properties of the estimator using aggregate data. Welsch and Kuh (1976) consider a related problem for a random coefficient model, and empirical application of aggregate regression models includes Polinsky (1977). However, unlike these studies, we explicitly incorporate cluster-specific random effects.

Let us now aggregate eq. (5) to the cluster level. By premultiplying $(\mathbf{\Lambda}^T \boldsymbol{W}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^T \boldsymbol{W}$ to both sides of equality, we obtain:

$$\bar{\boldsymbol{Y}} = \bar{\boldsymbol{X}}\boldsymbol{\beta} + \bar{\boldsymbol{U}}, \tag{6}$$

where the bar represents cluster-level averages. For example, $\bar{\boldsymbol{Y}} \equiv (\mathbf{\Lambda}^T \boldsymbol{W}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^T \boldsymbol{W}\boldsymbol{Y}$ is a $(K \times 1)$-matrix of the cluster-level average of y_h, where the $(k, 1)$ element of $\bar{\boldsymbol{Y}}$ is the weighted cluster-level average $\bar{y}_k (\equiv \sum_{h\in\mathcal{H}_k} w_h y_h / \sum_{h\in\mathcal{H}_k} w_h)$ for cluster k. Similarly, $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{U}}$ are, respectively, $(K \times L)$- and $(K \times 1)$-matrices of cluster-level averages $\bar{\boldsymbol{x}}_k$ and $\bar{u}_k$ of $\boldsymbol{x}_h$ and u_h, respectively. Because we can obtain $\bar{y}_k$ and $\bar{\boldsymbol{x}}_k$ for all survey clusters from the survey and census, respectively, we can run a cluster-level regression of $\bar{y}_k$ on $\bar{\boldsymbol{x}}_k$. To keep the presentation simple, we first develop the method when the cluster-level data come from one sample. We then discuss the consequences of merging survey and census data at the cluster level.

Since the number of households and the group composition may vary from cluster to cluster in the survey, $\bar{u}_k$ is in general heteroskedastic. It is straightforward to show that the covariance matrix $\mathbf{\Omega}_A \equiv E[\bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^T]$ of $\bar{\boldsymbol{U}}$ is a diagonal matrix, whose (k, k) element is $\sigma^2_{u,k} = \sigma^2_\eta + \sum_{g=1}^G \bar{A}_{k,g}\sigma^2_{\epsilon,g}$ for $\bar{A}_{k,g} \equiv (\sum_{h\in\mathcal{H}_k} w_h)^{-2} \sum_{h\in\mathcal{H}_k} w_h^2 a_{hg}$. This suggests the following GLS estimation procedure. First, we take the OLS residual $\hat{u}_k$ from eq. (6). In the second step, we regress $\hat{u}_k^2$ (instead of $\sigma^2_{u,k}$) on a constant and $\bar{A}_{k,1}, \cdots, \bar{A}_{k,G}$. This gives consistent estimates $\hat{\sigma}^2_\eta$ and $\hat{\sigma}^2_{\epsilon,g}$ of σ^2_η and $\sigma^2_{\epsilon,g}$, which in turn give a consistent estimate $\hat{\mathbf{\Omega}}_A \equiv \mathrm{diag}(\hat{\sigma}^2_{u,1}, \ldots, \hat{\sigma}^2_{u,K})$ of $\mathbf{\Omega}_A$, where $\hat{\sigma}^2_{u,k} \equiv \hat{\sigma}^2_\eta + \sum_{g=1}^G \bar{A}_{k,g}\hat{\sigma}^2_{\epsilon,g}$.[7] We can also check if ϵ_h is heteroskedastic under the following null hypothesis: $\sigma^2_{\epsilon,1} = \cdots = \sigma^2_{\epsilon,G}$.

Let us now premultiply $\mathbf{\Omega}_A^{-1/2}$ to eq. (6), so that we have $\mathbf{\Omega}_A^{-1/2}\bar{\boldsymbol{Y}} = \mathbf{\Omega}_A^{-1/2}\bar{\boldsymbol{X}}\boldsymbol{\beta} + \mathbf{\Omega}_A^{-1/2}\bar{\boldsymbol{U}}$ and denote the diagonal matrix of cluster-level averages of the sample weights by $\bar{\boldsymbol{W}} \equiv \mathrm{diag}(\bar{w}_1, \cdots, \bar{w}_K)$ for $\bar{w}_k \equiv H_k^{-1} \sum_{h\in\mathcal{H}_k} w_h$. With this, we can run a weighted least squares regression of $\mathbf{\Omega}_A^{-1/2}\bar{\boldsymbol{Y}}$ on $\mathbf{\Omega}_A^{-1/2}\bar{\boldsymbol{X}}$.

We now add the data subscript to demonstrate how the aggregation estimator can be constructed in our context. We take the regressors from the census and the dependent variable from the survey, such that $\bar{X}_C$ and $\bar{Y}_S$. Since the heteroskedasticity of $\bar{u}_k$ and sample weights originate from the survey, these are computed from the survey. Therefore, our aggregation estimator $\hat{\boldsymbol{\beta}}_{AGG}$ is given as follows:

$$\hat{\boldsymbol{\beta}}_{AGG} \equiv (\bar{\boldsymbol{X}}_C^T \hat{\mathbf{\Omega}}_A^{-T/2} \bar{\boldsymbol{W}}_S \hat{\mathbf{\Omega}}_A^{-1/2} \bar{\boldsymbol{X}}_C)^{-1} \bar{\boldsymbol{X}}_C^T \hat{\mathbf{\Omega}}_A^{-T/2} \bar{\boldsymbol{W}}_S \hat{\mathbf{\Omega}}_A^{-1/2} \bar{\boldsymbol{Y}}_S \tag{7}$$

[7] Note that the regression in the second step is unweighted, since the accuracy of $\hat{u}_k^2$ as an estimate of $\sigma^2_{u,k}$ does not systematically improve as the number of observations in each cluster k grows. This is because the error term η does not go away by aggregation. Weighting can actually make the estimate of σ^2_η less accurate.

To establish the consistency and asymptotic normality of this estimator, we define $\boldsymbol{J}^0 \equiv I_{K_S}$ and $\boldsymbol{J}^1 \equiv \boldsymbol{\Omega}_A^{-T/2}\bar{\boldsymbol{W}}_S\boldsymbol{\Omega}_A^{-1/2}$ and make the following three additional assumptions in addition to Assumptions **A1** and **A2**:

A3 $K_S^{-1}\bar{\boldsymbol{X}}_C^T\boldsymbol{J}^j\bar{\boldsymbol{X}}_C \xrightarrow{p} \Xi_1$ for $j \in \{0, 1\}$ as $K_C \to \infty$, where Ξ_1^j is a non-singular symmetric matrix.

A4 $K_S^{-1/2}\bar{\boldsymbol{X}}_C^T\boldsymbol{J}^j\bar{\boldsymbol{U}}_S \xrightarrow{d} \mathcal{N}(0, \Xi_2^j)$ for $j \in \{0, 1\}$ as $K_C \to \infty$, where Ξ_2^j is a positive-definite and symmetric matrix.

A5 $K_S^{-1}\bar{\boldsymbol{X}}_C^T\boldsymbol{J}^j(\bar{\boldsymbol{X}}_S - \bar{\boldsymbol{X}}_C)\boldsymbol{\beta} \xrightarrow{p} 0_L$ for $j \in \{0, 1\}$ as $K_C \to \infty$.

Assumption **A3** essentially requires a convergence in the second moment of cluster averages. Assumption **A4** is somewhat different from a standard assumption because the cluster averages for the regressors come from the census but the cluster averages for the error terms comes from the survey. Nevertheless, since the error terms are orthogonal to the regressors, this assumption is reasonable. Assumption **A5** essentially requires that (possibly weighted) correlation between the cluster average of the regressors and the sampling error in the cluster averages, $\bar{\boldsymbol{X}}_S - \bar{\boldsymbol{X}}_C$, are orthogonal. Since survey sampling within each cluster is random, this is also a reasonable assumption to maintain. In principle, the sampling error affects the asymptotic variance of $\hat{\boldsymbol{\beta}}_{AGG}$. However, since a sufficiently high proportion of households are selected in each cluster, variance of $\hat{\boldsymbol{\beta}}_{AGG}$ due to the sampling error is small relative to those due to $\bar{U}_S$. With these, we have the following proposition (all proofs are in Appendix 1):

Proposition 1 *Under assumptions* **A1**-**A5** *and some regularity conditions,* $\hat{\boldsymbol{\beta}}_{AGG}$ *is a consistent estimate of* $\boldsymbol{\beta}$ *and asymptotically normally distributed with the asymptotic variance approximately equal to* $(\Xi_1^1)^{-1}\Xi_2^1(\Xi_1^1)^{-1}$ *as* $K_C \to \infty$.

Instrumental variables approach

While $\hat{\boldsymbol{\beta}}_{AGG}$ is consistent, it throws away household-level information and thus may not be efficient. That is, even though the common regressors $\boldsymbol{x}_h^e$ are observed for each household, the Aggregation Approach uses only their cluster-level averages for estimation. Therefore, we consider an alternative approach that uses a household-level regression model with instrumental variables z_h. It should be noted that we use instrumental variables not to address endogeneity issue but to combine two samples, as it becomes apparent shortly.

We are clearly unable to use the standard instrumental variables estimator here, because we do not observe y_h, $\boldsymbol{x}_h$, and z_h in the same sample. We instead need to use instrumental variables in a two-sample setting. To understand how this works, note first that the instrumental variables estimator and two-stage least-squares estimator can be written as a product of sample moments. As we discuss below, different moments can be computed from the two samples. Therefore, by combining moments from the two samples, we can run instrumental variables regressions. This approach was first proposed and used by Angrist and Krueger (1992) to investigate the relationship between the age at the school entry and ultimate educational attainment. It was subsequently adopted in various studies such as Lusardi (1996); Bjorklund and Jantti (1997); Currie and Yelowitz (2000), and Dee and Evans (2003).

It is noteworthy that, unlike the standard one-sample case, instrumental variables and two-stage least-squares estimators are not identical even when the estimation equation is exactly identified. We choose to use the two-sample two-stage least squares (TS2SLS) estimator instead of the two-stage instrumental variables estimator, because the former is asymptotically more efficient than the latter (Inoue and Solon 2010). However, the version of TS2SLS we use differs from Inoue and Solon (2010), as we allow for clustering.

To derive the TS2SLS estimator in our setup, consider the following first-stage regression: $\boldsymbol{X} = \boldsymbol{Z\Gamma} + \boldsymbol{\Delta}$, where $\boldsymbol{\Gamma}$ and $\boldsymbol{\Delta}$ are $(L \times L)$- and $(N \times L)$-matrices, respectively. The first-stage estimate of $\boldsymbol{\Gamma}$ can be obtained by $\hat{\boldsymbol{\Gamma}}_C = (\boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{Z}_C)^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{X}_C$ using the census sample. We then replace (unobservable) $\boldsymbol{X}_S$ with its predicted value $\boldsymbol{Z}_S \hat{\boldsymbol{\Gamma}}_C$ in the weighted least squares estimator $(\boldsymbol{X}_S^T \boldsymbol{W}_S \boldsymbol{X}_S)^{-1}(\boldsymbol{X}_S^T \boldsymbol{W}_S \boldsymbol{Y}_S)$ to arrive at the following TS2SLS estimator:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{TS2SLS} \equiv\ & (N_C^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{X}_C)^{-1} (N_C^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{Z}_C)(N_S^{-1} \boldsymbol{Z}_S \boldsymbol{W}_S \boldsymbol{Z}_S)^{-1} \\ & (N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{Y}_S),\end{aligned} \tag{8}$$

To derive the asymptotic characteristics of $\hat{\boldsymbol{\beta}}_{TS2SLS}$, we let $\boldsymbol{\Omega}^0 \equiv \mathrm{diag}(A \cdot (\sigma_{\epsilon,1}^2, \cdots, \sigma_{\epsilon,G}^2)^T) + \sigma_\eta^2 \boldsymbol{\Gamma}^T \boldsymbol{\Gamma}$ and make the following assumptions:

A6 $N_D^{-1} \boldsymbol{Z}_D^T \boldsymbol{W}_D \boldsymbol{X}_D \xrightarrow{p} \boldsymbol{Q}_0$ as $K_C \to \infty$ for each $D \in \{C, S\}$, where $\boldsymbol{Q}_0$ is a non-singular matrix.

A7 $N_D^{-1} \boldsymbol{Z}_D^T \boldsymbol{W}_D \boldsymbol{Z}_D \xrightarrow{p} \boldsymbol{Q}_1$ as $K_C \to \infty$ for each $D \in \{C, S\}$, where $\boldsymbol{Q}_1$ is a non-singular symmetric matrix.

A8 $N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{\Omega}_S^0 \boldsymbol{W}_S \boldsymbol{Z}_S \xrightarrow{p} \boldsymbol{Q}_2$ as $K_C \to \infty$, where $\boldsymbol{Q}_2$ is a positive-definite and symmetric matrix.

A9 $\sqrt{K_S}(N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{U}_S) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{Q}_2)$ as $K_C \to \infty$.

With these assumptions, we have the following result:

Proposition 2 *Under Assumptions* **A1**, **A2**, **A6**-**A9**, *and some regularity conditions,* $\hat{\boldsymbol{\beta}}_{TS2SLS}$ *is a consistent estimate of* $\boldsymbol{\beta}$ *and asymptotically normally distributed with the asymptotic variance of* $\boldsymbol{Q}_1^{-1} \boldsymbol{Q}_2 \boldsymbol{Q}_1^{-1}$ *as* $K_C \to \infty$.

It should be noted that it is straightforward to obtain the sample analogue of $\boldsymbol{Q}_1$. However, the sample analogue of $\boldsymbol{Q}_2$ cannot be computed readily, because $\boldsymbol{\Omega}_S^0$ is in general unknown. To estimate the variance of $\hat{\boldsymbol{\beta}}_{TS2SLS}$, one may resort to bootstrapping by independently bootstrapping the two samples and repeatedly estimate $\hat{\boldsymbol{\beta}}$. While this is potentially a useful way to find an estimate of var$[\hat{\boldsymbol{\beta}}]$, it does not work well for the ELL method because we still need estimates of $\sigma_{\epsilon,g}^2$ and σ_η^2 for imputing y_h. Therefore, we choose to use the estimates $\hat{\sigma}_{\epsilon,g}^2$ and $\hat{\sigma}_\eta^2$ taken from the Aggregation Approach. This in turn allows us to get a consistent estimate $\hat{\boldsymbol{\Omega}}^0 \equiv \mathrm{diag}(A \cdot (\hat{\sigma}_{\epsilon,1}^2, \cdots, \hat{\sigma}_{\epsilon,G}^2)^T) + \hat{\sigma}_\eta^2 \boldsymbol{\Gamma}^T \boldsymbol{\Gamma}$ of $\boldsymbol{\Omega}^0$ using the census data. Replacing $\boldsymbol{\Omega}^0$ with $\hat{\boldsymbol{\Omega}}^0$ in the definition of $\boldsymbol{Q}_2$, we can obtain an estimate of $\boldsymbol{Q}_2$. With these sample analogues of $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, we can estimate the variance of $\hat{\boldsymbol{\beta}}_{TS2SLS}$.

The choice we make for the estimate of $\boldsymbol{Q}_2$ has an added advantage. Because both the Aggregation and Instrumental Variables Approaches use the same estimates of $\sigma_{\epsilon,g}$ and σ_η, the differences in poverty estimates only result from the differences in the estimates of $\hat{\boldsymbol{\beta}}$ and $\widehat{\text{var}}[\hat{\boldsymbol{\beta}}]$. This allows us to make a fair comparison between these two approaches.

So far, we have been agnostic about the choice of the proxy variable $\tilde{a}_h$. In our empirical application, we consider two alternatives. The first alternative ("TS2SLS-A") is the conditional probability of being in each group. That is, we assume that the probability of household being in group g conditional on v_h is $Prob_g(v_h;\theta)$, where v_h is a vector of household characteristics observed both in the survey and census and θ is the parameter of the model. The parameter estimate $\hat{\theta}$ is obtained from the census, and the estimated probability is imputed into each survey household by $\hat{a}_{hg} \equiv Prob_g(v_h;\hat{\theta})$.

The second alternative ("TS2SLS-B") is the average of the group indicators in the cluster, or $\hat{\boldsymbol{a}}_h = \bar{\boldsymbol{a}}_{\kappa(h)}$. The second alternative can be considered as a special case of first alternative, where the former uses only the cluster membership to predict the probability of household being in group g. In our empirical application, this approach is straightforward to implement as the cluster-level average of disability status can be easily computed from the census. The second alternative has an advantage that the results are directly comparable with the results from the Aggregation Approach, because exactly the same set of variables is used for estimation. In contrast, variables not used in the Aggregation Approach can be included in v_h under the first alternative.

Whether we use Aggregation Approach or Instrumental Variables Approach, we carry out Monte-Carlo simulations in the same way as the ELL method once we obtain the relevant parameter estimates. That is, we draw $\tilde{\boldsymbol{\beta}}^{(r)}$ for the rth round of simulation from a normal distribution with mean $\hat{\boldsymbol{\beta}}$ and variance $\widehat{\text{var}}[\hat{\boldsymbol{\beta}}]$. The variances of the error terms, $\tilde{\sigma}_\eta^{2,(r)}$ and $\tilde{\sigma}_\epsilon^{2,(r)}$, are jointly drawn from a normal distribution using the point estimate and variance-covariance matrix from the residual regression estimates.[8] Cluster-specific random effect $\tilde{\eta}_k^{(r)}$ and household-specific random effect $\tilde{\epsilon}_{hg}^{(r)}$ are drawn from the empirical distribution of $\hat{u}_k$ standardized to have mean zero and a unit standard deviation, and augmented by $\tilde{\sigma}_\eta^{(r)}$ and $\tilde{\sigma}_\epsilon^{(r)}$, respectively. Once we have drawn these parameters, we calculate $\tilde{y}_h^{(r)}$ for each census record. The remaining steps are the same as the standard ELL method.

The extension of the ELL method discussed above is fairly general. While the empirical focus of this study is on disabled households, both the Aggregation and Instrumental Variables Approaches are potentially applicable to many other situations where the survey does not contain the group membership indicators but their proxy variables can be constructed from the survey and census.

[8] If either $\tilde{\sigma}_\eta^{2,(r)}$ or $\tilde{\sigma}_\epsilon^{2,(r)}$ is negative, we redraw until both are positive.

3 Data and measurement

We apply the extension of the ELL SAE developed in the previous section to a survey and a census in Tanzania to derive poverty statistics for disabled and non-disabled households. In this section, we first describe the data and then discuss the measurement of poverty and disability.

3.1 Data

As noted above, we use a census and a survey for this study. For the census, we use the long-form questionnaire of the Population and Housing Census of Tanzania for 2002, which includes questions on the age, sex, relation to the household head, marital status, education and economic activity of each household member as well as the housing conditions and asset holdings of the household. In addition, the long-form questionnaire asks about the disability status of each household member. The long-form questionnaire was used for about 1.2 million households out of about 6.8 million households in mainland Tanzania. The sample for the long-form questionnaire was drawn by Systematic Simple Random Sampling, where each enumeration area (EA) was selected with equal probability and all households in sample enumeration areas are included in the sample. We apply the census weights that incorporate this sampling procedure. We excluded Zanzibar from the analysis, because it is not covered in the survey. We also excluded less than 0.01 percent of households, whose size exceed the observed maximum household size in the survey. Further details on the census can be found in National Bureau of Statistics (2003a, b).

For the survey, we use the 2000/01 Household Budget Survey (HBS). It is representative at the level of 20 regions in mainland Tanzania.[9] The survey data cover a wide range of household and individual characteristics, including many of the variables included in the census, and detailed information on consumption expenditure. The survey does not, however, include a question on the disability status for each household member. Therefore, we cannot identify the individuals with disability in the survey. However, the survey data allow us to identify the individuals who are not economically active due to disability.

The survey data contain 22,178 households from 1,158 enumeration areas. It comes with a sample weight for each observation. After eliminating observations with missing values for the variables used in this study, we are left with 21,608 observations from 1,148 enumeration areas. National Bureau of Statistics (2002) offers further information on the HBS data, including a range of summary statistics.

Since the administrative identifiers are not fully harmonized at the level of enumeration areas between the survey and census, we choose to regard a ward as a cluster in this paper and merge the survey and census datasets at the this level to implement the Aggregation Approach. There are 2,457 and 801 clusters (wards) in the census and

[9] Some regions were split after the survey was conducted. In 2002, the Manyara region broke out of the Arusha region and the census data allow us to distinguishes between Manyara and Arusha regions, but the survey data do not.

survey, respectively. It is worth noting here that the census average of the disability indicator at the cluster level can be merged into the survey.

3.2 Measurement of poverty and disability

To measure poverty, we adopt the official definition of poverty given by National Bureau of Statistics (2002). National Bureau of Statistics (2002) first calculates the real household consumption expenditure per adult equivalent for 28 days, which accounts for regional price differences and accommodates different needs for different age and sex groups.[10] The household is deemed poor when the real household consumption expenditure per adult equivalent for 28 days is below the official poverty line, which is set at 7,253 Tanzanian schillings, or 16.09 US dollars using the purchasing power parity conversion factor reported in the World Development Indicators for 2002. The poverty line covers the cost for satisfying the minimum adult caloric requirement and some essential non-food consumption expenditure. According to this definition, 35.7 percent of the people in mainland Tanzania are poor.

As for disability, we simply use the self-reported disability status in the census. That is, a person who reported any disability is classified as disabled in this study. While self-reported disability measure has some potential issues such as self-reporting bias, it is likely to capture severe forms of disability.[11]

Table 1 provides some summary statistics of the disability status in mainland Tanzania. As column (8) of the table shows, 1.8 percent of the population has some disability in Tanzania. Columns (1), (2), and (3) break down the summary statistics by different age groups. They show that elderly people (aged over 65) have substantially higher prevalence of disability than children (under 15) and working-age adults (aged between 15 and 64). The gender difference in the prevalence of disability is not large, as columns (4) and (5) show.

In columns (6) and (7) of Table 1, we report the summary statistics by whether the individual is the household head. The difference between head and other household members in the distribution of disability reflects the fact that household heads are on average older than other household members, but there are also differences in the type of disability between the household heads and other household members. Hence, disabled household heads do not necessarily represent disabled people in general.[12]

Nevertheless, as with Hoogeveen (2005), we classify a household headed by a disabled person as a disabled household without distinguishing different types of disability. We make this choice for three reasons. First, we prefer to focus on the disability status of adults, because children's disability status is likely to be less reliable than adults' disability status. For example, UNICEF (1999) argues that some disabilities may go unrecognized until children go to school when learning difficulties as well

[10] See National Bureau of Statistics (2002) for the weights to derived the adult equivalence scale, which is similar to those cited in Collier et al. (1986). See World Bank (2015) for additional discussions on the adult equivalence scale used in Tanzania.

[11] Evidence from Malawi and Zambia indicates that self-reported disabled people, on average, experience much more activity limitations than self-reported non-disabled people (Loeb and Eide 2004, 2006).

[12] In Online Appendix Table A1, we report the distribution of types of disability.

Table 1 Prevalence of disability and population and disability shares by age groups, gender, and relationship to the household head in mainland Tanzania

	Age Group			Gender		Rel. to Head		Total
	-14 (1)	15-64 (2)	65+ (3)	M (4)	F (5)	Head (6)	Others (7)	(8)
Prevalence of disability (%)	1.0	2.2	6.8	2.1	1.6	2.7	1.6	1.8
Population share (%)	44.7	51.1	4.1	48.6	51.4	21.6	78.4	100.0
Disability share (%)	23.5	61.2	15.2	55.0	45.0	31.5	68.5	100.0

Source: Author's calculations based on the Population and Housing Census for 2002

as visual and hearing impairment are brought to notice. Since the household heads are adults, we can largely avoid the issue of unrecognized (and thus unreported) disabilities. Second, we focus on the disability status of the household head, because the household head's status is likely to be important for the households' poverty status. Third, we choose to use a binary disability indicator instead of disaggregating different types of disability, because the number of observations for each *specific type* of disability is extremely low in each cluster. As a result, we are unable to reliably estimate the coefficients for different types of disability. Hence, in our empirical application, we let $G = 2$ with $g = 1$ and $g = 2$ representing non-disabled and disabled households.

While neither the survey nor census data alone allows us to find the poverty rates disaggregated by the households' disability status, we are able to get some insights into the difference in poverty by the disability status. In the survey, we can disaggregate the poverty rate by whether the household is headed by an "economically disabled" member, or a member who is economically inactive due to disability as reported in Online Appendix Table A2. In the census, we observe the housing conditions and asset holdings of the households, which can be expected to be correlated to the household welfare. Hence, we can disaggregate these variables by the household disability status as reported in Online Appendix Table A3. These tables indicate that disabled households are poorer than non-disabled households.[13]

4 Results

As with the ELL method, our first step is to estimate the model parameters. Unlike most other empirical studies, estimated regression parameters are *not* the quantities of intrinsic interest in this study. Nevertheless, we briefly explain the regression results reported in Table 2.

As noted above, our dependent variable y_h is logarithmic household consumption per adult equivalent for 28 days. In the Aggregation Approach, we need to merge the cluster-level average of y_h for the survey with the cluster-level averages of census variables, including the indicator variable for the household head's disability status.

[13] Online Appendix Table A2 also shows that the head's economic disability status is more strongly correlated with poverty than the presence of an economically disabled member, showing the importance of household head's disability status relative to other member's disability status. Online Appendix also provides further discussions.

We take a ward to be a cluster, because we cannot merge the two data sources at the level of enumeration areas. As a result, the standard errors reported in this study may be slightly upward biased, because wards are larger than enumeration areas. Nevertheless, the bias is likely to be small because most of the survey wards have only one enumeration area.

In principle, different consumption models can be used for different zones or regions. However, we choose to have a single consumption model for mainland Tanzania to ensure that there is a sufficiently large number of clusters (wards) in the estimation sample. Further, using a single consumption model also helps us to keep the presentation simple.

As detailed in Sect. 2, we first run an OLS regression based on the Aggregation Approach. The OLS regression results are reported in column (1) ("Aggregate OLS") of Table 2. We then run a regression of the squared OLS residuals $\hat{u}_k^2$ on a constant, $\bar{A}_{k,\text{non-disabled}}$, and $\bar{A}_{k,\text{disabled}}$. The coefficient on these terms provide estimates of σ_η^2 and $\sigma_{\epsilon,g}^2$, respectively, and are reported in Table 3.[14] While these quantities are not of intrinsic interest, there are two points to note here. First, the variance σ_η^2 of the cluster-specific random effects is statistically significant, even though it is much smaller than the magnitude of the variance $\sigma_{\epsilon,g}^2$ of the household-specific random effects. Because the errors due to the cluster-specific effects do not decrease as fast as the errors due to the household-specific effects as we aggregate, it is important to allow for cluster-specific effects. Second, while the difference between $\sigma_{\epsilon,\text{non-disabled}}^2$ and $\sigma_{\epsilon,\text{disabled}}^2$ is not statistically significant at a conventional level, the latter is much larger than the former. This is consistent with our earlier conjecture that there may exist household-specific random effects that are only relevant to disabled households.

With these estimates, we are able to obtain the GLS estimates of $\boldsymbol{\beta}$. In column (2) ("Aggregate GLS") of Table 2, we report the regression coefficients based on the Aggregation Approach using eq. (7). This column shows that people live in larger households tend to be poorer in general. Column (2) also indicates that grass leaves and bamboo roofs are significantly negatively associated with logarithmic consumption per adult equivalent, whereas households with concrete roofs, telephones, and bikes tend to have significantly higher levels of consumption. These results are to be expected. One coefficient of particular interest is the disability status of household. It is interesting to note that the point estimate is negative even though it is statistically insignificant in column (2).

In columns (3) ("TS2SLS-A") and (4) ("TS2SLS-B") of Table 2, we report the results of TS2SLS regressions. In column (3), we use the predicted probability of disability status as an instrument. Specifically, we estimate a logit regression of the disability indicator variable using the census data. The results of the logit regression are shown in Online Appendix Table A5. The covariates used in the logit regression are commonly observed in both the census and survey such that we can compute $\hat{a}_{h1}$, $\hat{a}_{h2}$, and z_h for both samples. With these, we can run the TS2SLS regression using eq. (8) and derive its variance based on the results in Proposition 2.

[14] In Online Appendix Table A4, we also report the summary statistics and correlations of $\hat{u}_k^2$, $\bar{A}_{k,\text{non-disabled}}$, and $\bar{A}_{k,\text{disabled}}$.

Table 2 Regression results for the aggregation and instrumental variables approaches

Variable	(1) Aggregate OLS			(2) Aggregate GLS			(3) TS2SLS-A			(4) TS2SLS-B		
	Coef		(S.E.)	Coef		(S.E.)	Coef		(S.E.)	Coef		(S.E.)
Constant	9.750	***	(0.129)	9.761	***	(0.169)	9.539	***	(0.157)	9.524	***	(0.165)
Head's age	−0.004	**	(0.002)	−0.003		(0.003)	0.003		(0.002)	−0.001		(0.002)
Household size	−155.8	***	(18.2)	−130.1	***	(24.2)	−136.5	***	(13.7)	−131.8	***	(14.9)
Household size squared	0.005	***	(0.001)	0.003	***	(0.001)	0.004	***	(0.001)	0.004	***	(0.001)
Electricity available	0.154	**	(0.069)	0.226	**	(0.101)	0.087		(0.081)	0.111		(0.083)
Protected well water	−0.008		(0.050)	0.021		(0.068)	0.043		(0.105)	0.042		(0.106)
Piped water	0.026		(0.038)	−0.003		(0.050)	0.094		(0.096)	0.104		(0.097)
River/dam/lake water	0.097	**	(0.049)	−0.040		(0.065)	0.117		(0.113)	0.112		(0.113)
Grass leaves & bamboo roofs	−0.241	***	(0.039)	−0.260	***	(0.052)	−0.199	***	(0.067)	−0.220	***	(0.065)
Concrete roofs	0.347	**	(0.145)	0.300	*	(0.178)	0.226		(0.284)	0.224		(0.285)
Total number of rooms	0.083	***	(0.020)	0.045	*	(0.027)	0.041		(0.038)	0.046		(0.037)
Have telephone	0.548	***	(0.202)	0.448		(0.280)	0.389		(0.479)	0.397		(0.479)
Have bike	0.202	***	(0.045)	0.219	***	(0.061)	0.112	**	(0.047)	0.136	***	(0.051)
Head's education	0.024	***	(0.008)	0.010		(0.010)	0.023	***	(0.005)	0.027	***	(0.006)
Tanga Region	0.111	**	(0.048)	0.121	*	(0.067)	0.074		(0.216)	0.124		(0.216)

Table 2 continued

Variable	(1) Aggregate OLS			(2) Aggregate GLS			(3) TS2SLS-A			(4) TS2SLS-B		
	Coef		(S.E.)	Coef		(S.E.)	Coef		(S.E.)	Coef		(S.E.)
Morogoro Region	0.094	**	(0.046)	0.140	**	(0.061)	0.063		(0.249)	0.049		(0.249)
Mtwara Region	−0.192	***	(0.054)	−0.207	***	(0.066)	−0.060		(0.265)	−0.069		(0.266)
Mbeya Region	0.115	**	(0.049)	0.093		(0.064)	0.162		(0.198)	0.223		(0.202)
Singida Region	−0.174	***	(0.055)	−0.227	***	(0.075)	−0.289		(0.233)	−0.325		(0.232)
Tabora Region	0.116	**	(0.050)	0.083		(0.076)	0.136		(0.270)	0.178		(0.269)
Rukwa Region	0.127	**	(0.060)	0.112		(0.090)	0.077		(0.373)	0.129		(0.380)
Disabled household	−1.049	*	(0.618)	−0.644		(0.792)	−4.087	***	(1.261)	−0.008		(2.538)

Source: Author's calculations based on the Population and Housing Census for 2002 and 2000/01 Household Budget Survey

Dependent variable is logarithmic consumption per adult equivalent for 28 days. *, **, and *** denote statistical significance at 10, 5, and 1 percent levels, respectively. TS2SLS-A [TS2SLS-B] uses the predicted probability of disability [cluster-level prevalence of disability] as the instrumental variable for the disabled household indicator. The standard errors reported in parentheses in column (1) are heteroskedasticity robust. Those in column (2) are based on the empirical analogue of the approximate asymptotic variance in Proposition 1, whereas those in columns (3) and (4) are empirical analogues of the asymptotic variance in Proposition 2

Table 3 Residual regression results

Variable	Coef		(S.E.)
σ^2_{η}	0.086	***	(0.008)
$\sigma^2_{\epsilon,\text{non-disabled}}$	0.039		(1.493)
$\sigma^2_{\epsilon,\text{disabled}}$	2.606		(27.315)

Source: Author's calculations based on the Population and Housing Census for 2002 and 2000/01 Household Budget Survey

The dependent variable is squared residuals $\hat{u}^2_k$ from the cluster-level OLS regression. The standard errors reported in parentheses are heteroskedasticity robust. *, **, and *** denote statistical significance at 10, 5, and 1 percent levels, respectively

Table 4 Poverty estimates based on different choices of approaches and instruments

	(1) Aggregate		(2) TS2SLS-A		(3) TS2SLS-B	
	Est	(S.E.)	Est	(S.E.)	Est	(S.E.)
Non-Disabled	35.4	(7.0)	32.0	(9.0)	36.8	(8.5)
Disabled	51.8	(9.1)	66.7	(14.7)	48.3	(15.7)
Mainland Tanzania	35.8	(6.8)	32.9	(8.9)	37.1	(8.2)

Source: Author's calculations based on the Population and Housing Census for 2002 and 2000/01 Household Budget Survey

All figures are in percentage

In column (4), we use the cluster-level prevalence of disability as an instrument for the household-level disability status. This instrumental variable uses the same set of information as the Aggregation Approach, except that the Instrumental Variables Approach uses observations of covariates at the household level, rather than the cluster level. Therefore, the results using this instrument are directly comparable to those based on the Aggregation Approach.

Overall, point estimates have expected signs, when they are significant. Further, the differences in the point estimates across different estimation methods are small relative to the standard errors for the point estimates. Therefore, the regression results are reasonable and broadly consistent with each other.

After all relevant model parameters are estimated, we conduct a Monte-Carlo simulation. As discussed in Sect. 2, we first draw the parameters $(\tilde{\boldsymbol{\beta}}^{(r)}, \tilde{\sigma}^{2,(r)}_{\eta}, \tilde{\sigma}^{2,(r)}_{\epsilon})$ in the rth round of the simulation. Then, we also draw the cluster-specific effect $\eta^{(r)}_k$ and household-specific effect $\epsilon^{(r)}_h$ for each cluster k and household h, respectively. These draws allow us to compute the imputed consumption per adult equivalent $\tilde{y}^{(r)}_h$ in the rth round for household h, which in turn can be used to compute the poverty statistics $\tilde{P}^{(r)}_J$ for each set J of households.

Table 4 reports the estimates of poverty rates for disabled and non-disabled households based on different choices of approaches and instruments. Regardless of the choice, the estimated poverty rates for the non-disabled households are around 35 percent, whereas those for disabled households are around 50 percent or above. Because the point estimates for non-disabled and disabled households are subject to the same model errors (i.e., errors associated with the estimation of $\boldsymbol{\beta}$), their errors are positively correlated across simulation rounds. As a result, we can conclude that the poverty rate for disabled households is significantly higher than that for non-disabled households at a five percent level of significance in all columns, even though the standard error for each estimate is not small. This result also corroborates the descriptive statistics derived from only the census or survey, which are respectively reported in Online Appendix Tables A2 and A3. In particular, the poverty rates for mainland Tanzania reported in Table 4 are close to the survey-only estimate of 35.7 percent, regardless of the approach used. Therefore, the SAE results are all consistent with the survey-only estimates.

Our preferred results are the one based on the Aggregation Approach, because the Aggregation Approach yields a smallest standard error and a point estimate that is closest to the survey-only estimate among the three estimates reported in Table 4. However, it is not always the case that Aggregation Approach is most preferable. When we are able to get very strong instrumental variables, it is likely that the Instrumental Variables Approach yields more favorable results than the Aggregation Approach. We will briefly revisit this point in Sect. 5.

Let us now turn to the regional-level estimates of poverty disaggregated by the disability status. For the ease of presentation, we plot the regional estimates of poverty rates for non-disabled and disabled households on maps in Fig. 1a and b, respectively.[15] As the comparison of these figures show, the poverty rates for disabled households are higher than those for non-disabled households for all regions. In Appendix 1, we also make a regional-level comparison between the survey-only and SAE estimates of poverty rate. It shows that the SAE poverty estimates are consistent with the survey-only estimates and have acceptable levels of standard errors. Therefore, we are indeed able to obtain meaningful spatial disaggregation of poverty and disability by extending the ELL SAE method.

It should be underscored that we have used consumption per adult equivalent as a measure of the standards of living that is comparable between disabled and non-disabled households. However, if there is a cost of disability, or additional consumption that needs to be given to disabled households to achieve the same level of standards of living as non-disabled households for a given level of consumption per adult equivalent, then we would underestimate the poverty for disabled households. Therefore, if we could take the cost of disability into account, our finding that disabled households are poorer than non-disabled households would be strengthened.

5 Discussion and conclusion

This paper extended the ELL SAE method to estimate poverty rates for small groups such as disabled households at a spatially disaggregated level. This extension is useful because survey data available so far typically do not permit meaningful spatial dis-

[15] Online Appendix Table A6 reports the point estimates and standard errors by disability status in each region.

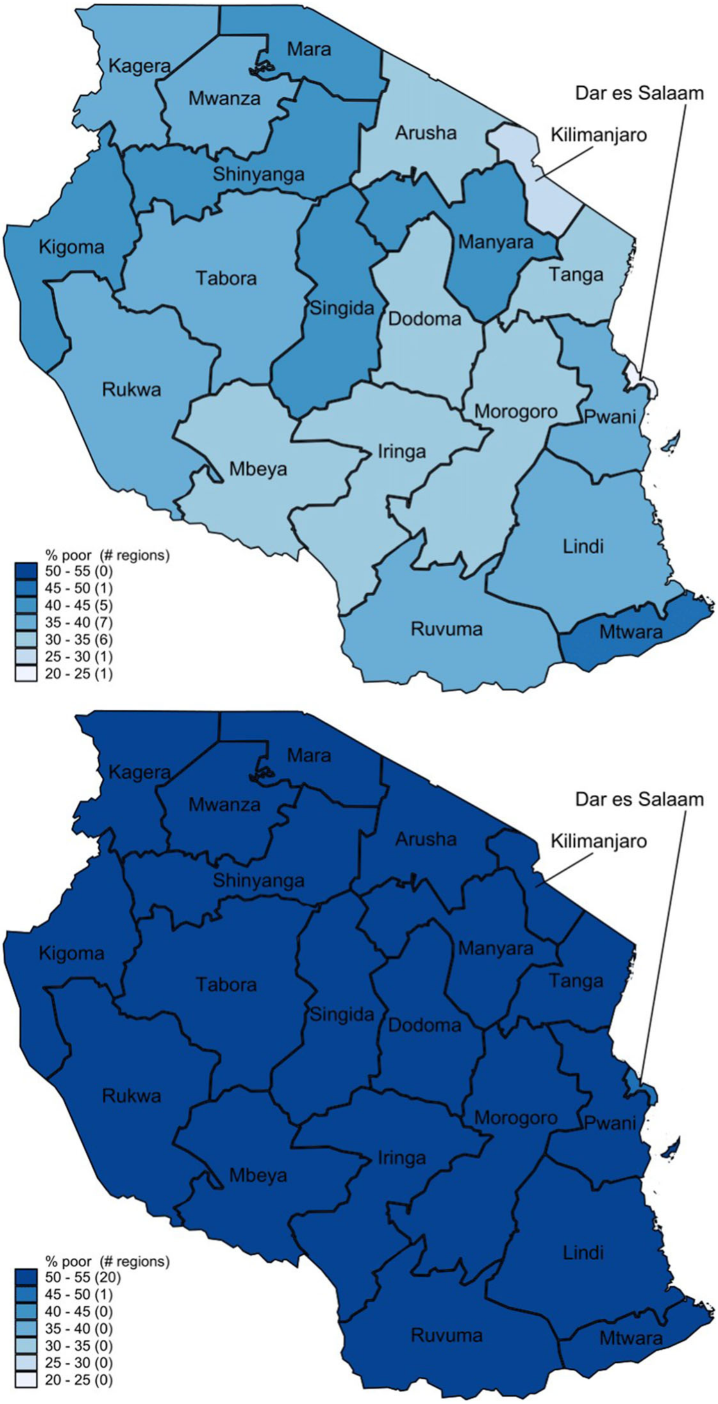

Fig. 1 Estimated poverty rates for **a** non-disabled households (top) and **b** disabled households (bottom) based on the Aggregate Approach

aggregation of poverty and disability.[16] However, unlike the ELL method, we do not require all covariates to be included both in the census and survey. This extension is particularly important when there is a systematic difference across different groups.

We enable an estimation of a consumption model that includes group indicators without observing them in the survey, while allowing for the potential heteroskedasticity between disabled and non-disabled households. While the difference between $\sigma_{\epsilon,\text{disabled}}$ and $\sigma_{\epsilon,\text{non-disabled}}$ is statistically insignificant in our application, allowing for this heteroskedasticity seems to be still important. We created a version of SAE estimates under the assumption that idiosyncratic error terms are homoskedastic. The point estimates (standard errors) of P_0 for non-disabled households, disabled households, and both combined (i.e., mainland Tanzania) are 0.365 (0.065), 0.413 (0.044), and 0.366 (0.064), respectively, using the Aggregation Approach. While the point estimate and standard error for mainland Tanzania are similar to the corresponding estimates in column (1) of Table 4, the standard errors under homoskedasticity appears to be too optimistic for disabled households. This is because the point estimate of $\sigma_{\epsilon,\text{disabled}}$ is much larger than $\sigma_{\epsilon,\text{non-disabled}}$ and because the number of disabled households is much smaller than that of non-disabled households This conclusion remains the same when the Instrumental Variables Approach is used.[17]

Both Aggregation and Instrumental Variables Approaches discussed in this paper produce poverty estimates that are consistent with the survey-only estimates in mainland Tanzania. The standard errors of the regional SAE estimates are comparable to those of the corresponding survey-only estimates at the regional level. Hence, our method successfully disaggregated regional poverty estimates by the disability status of the household head, which is unobserved in the survey.

In our empirical application, the Aggregation Approach provides an estimate that is closer to the survey-only estimate and has the smallest standard error among all alternatives we considered. Therefore, our preferred results are based on the Aggregation Approach. However, there are situations where Instrumental Variables Approach may work better than the Aggregation Approach. To highlight this point, first note that the Aggregation Approach does not work when there is no variation in the cluster-level prevalence of disability, because the dependent variable is a constant in this case. However, the Instrumental Variables Approach may work even in this case, provided that there are variations in the proxy variable $\tilde{a}_h$ within each cluster.

[16] Note, however, that there is an ongoing effort to improve the data collection standards and data availability (Abualghaib et al. 2019; Groce and Mont 2017). The methodology proposed in this paper will remain relevant because it takes time to improve data availability and because it is still useful to see the change from the past situation.

[17] In addition, we also implemented a version of Hoogeveen (2005), where we use the same set of covariates as those in Table 2, except that we replace the individual disability status with the cluster-level prevalence of disability. The point estimate (standard error) of P_0 for non-disabled households, disabled, and mainland Tanzania are 0.490 (0.009), 0.554 (0.010), and 0.491 (0.008), respectively. Hence, poverty appears to be overestimated when a comparable model is used. While this apparent overestimation may be driven by a particular specification we use, the standard errors also appear to be too low, particularly for non-disabled households. This is also the case when the Instrumental Variables Approach is used. This results is expected, since Hoogeveen (2005) does not take into consideration the fact that the cluster-level prevalence of disability is just a proxy for the individual-level disability status.

The extension of the ELL method considered in this study can be easily applied to a study of poverty and disability in other countries, if census and survey data that are similar to the ones used in this study are available. It can also be applied to a variety of other contexts in which poverty comparisons across groups are important (e.g., poverty comparisons across ethnic or religious groups). Further, the proposed extension has an advantage that it enables spatial disaggregation of poverty estimates, even in cases where the survey does not permit such disaggregation.

Finally, this study also makes an empirical contribution to the growing body of literature on disability and poverty. We show that disabled households are significantly poorer than non-disabled households in every region of mainland Tanzania. This conclusion would be further strengthened when we take into account the consumption cost of disability and possibly social exclusion.[18] Hence, this study underscores the importance of considering disability in the formulation of anti-poverty policies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00181-023-02478-8.

Funding An earlier version of this study was supported by SMU Research Grant (07-C208-SMU-007).

Declarations

Conflict of interest Fujii declares that he has no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the Singapore Management University and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Appendix

Appendix A: Comparison of regional poverty estimates

Table 5 provides the survey-only estimates of poverty rates at this level in column (1) and their corresponding SAE estimates in column (2) at the level of 20 survey regions. The survey-only and SAE estimates are generally close and have a strong positive correlation of 0.71. The null hypothesis for a two-sided z-test of equality of the survey-only and SAE estimates is not rejected at a five percent level of significance in all regions. The null hypothesis for the joint χ^2-test of the equality of the two estimates in all regions was also not rejected. Therefore, on balance, even though the difference between survey-only and SAE estimates can be large in absolute value for a few regions, there is no evidence that survey-only and SAE estimates are systematically different.

In column (3), we report the share of people living in disabled households. As the comparisons of column (3) with columns (1) and (2) show, there is a strong positive correlation between the poverty rate and share of people in disable households. The

[18] In Tanzania, social exclusion of disabled people may not be among the most important issues facing those who are disabled (Kisanji 1995a, b; UNICEF 1999).

Table 5 Comparison of regional poverty rates between survey-only and SAE estimates

Region	(1) Survey only		(2) SAE		(3) Share in
	Est.	(S.E.)	Est.	(S.E.)	Disabled HH
Dodoma	34.0	(3.5)	31.2	(9.6)	2.9
Arusha	38.9	(3.9)	36.4	(6.6)	2.3
Kilimanjaro	30.4	(4.6)	29.8	(10.0)	1.6
Tanga	34.6	(3.9)	31.0	(9.2)	1.8
Morogoro	29.5	(3.5)	30.7	(9.2)	3.0
Pwani	46.1	(5.2)	38.6	(5.9)	3.5
Dar Es Salaam	17.4	(2.1)	24.2	(10.4)	1.3
Lindi	53.5	(5.1)	38.6	(6.3)	2.8
Mtwara	37.9	(3.8)	46.2	(4.8)	3.5
Ruvuma	41.2	(6.1)	36.3	(7.3)	3.2
Iringa	31.8	(4.1)	33.3	(8.3)	2.6
Mbeya	19.8	(2.4)	30.8	(9.6)	1.4
Singida	55.7	(4.5)	44.3	(6.1)	3.3
Tabora	27.3	(4.4)	38.2	(7.0)	1.9
Rukwa	30.8	(4.9)	36.1	(8.3)	1.3
Kigoma	37.6	(5.0)	42.6	(4.3)	3.7
Shinyanga	41.8	(4.2)	41.0	(4.8)	2.7
Kagera	28.7	(5.1)	36.4	(7.1)	2.7
Mwanza	47.9	(3.8)	39.8	(5.2)	2.7
Mara	45.5	(4.1)	41.0	(4.7)	3.2

All figures are in percentage. Column (2) is based on the Aggregation Approach

correlation coefficient between columns (1) and (3) [(2) and (3)] is 0.66 [0.68]. Because the calculation of the correlation between columns (1) and (3) does not involve any imputation, the positive correlation cannot be attributed to imputation.

Appendix B: Proofs

Proof of Proposition 1 We first establish the consistency of the cluster-level OLS estimator $\hat{\boldsymbol{\beta}}_{COLS} \equiv (\bar{\boldsymbol{X}}_C^T \bar{\boldsymbol{X}}_C)^{-1} \bar{\boldsymbol{X}}_C \bar{\boldsymbol{Y}}_S$. Letting $j = 0$ in Assumptions **A3–A5** and using $K_S \to \infty$ as $K_C \to \infty$, we have:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{COLS} - \boldsymbol{\beta} &= (K_S^{-1} \bar{\boldsymbol{X}}_C^T \bar{\boldsymbol{X}}_C)^{-1} K_S^{-1} \bar{\boldsymbol{X}}_C (\boldsymbol{X}_S \boldsymbol{\beta} + \boldsymbol{U}_S) - \boldsymbol{\beta} \\
&= (K_S^{-1} \bar{\boldsymbol{X}}_C^T \bar{\boldsymbol{X}}_C)^{-1} K_S^{-1} \bar{\boldsymbol{X}}_C (\bar{\boldsymbol{X}}_S - \bar{\boldsymbol{X}}_C) \beta \\
&\quad + (K_S^{-1} \bar{\boldsymbol{X}}_C^T \bar{\boldsymbol{X}}_C)^{-1} K_S^{-1} \bar{\boldsymbol{X}}_C \bar{\boldsymbol{U}}_S \xrightarrow{p} 0_L
\end{aligned}
$$

This means that we can obtain consistent estimates of σ_η^2 and $\sigma_{\epsilon,g}^2$ by running a regression of the squared OLS residual $\hat{u}_k^2$ on a constant and $\bar{A}_{k,1}, \cdots, \bar{A}_{k,G}$. Hence, we can obtain a consistent estimate $\hat{\boldsymbol{\Omega}}_A$ of $\boldsymbol{\Omega}_A$ by replacing σ_η^2 and $\sigma_{\epsilon,g}^2$ with their consistent estimates in the definition of $\boldsymbol{\Omega}_A$. Then, letting $\hat{\boldsymbol{J}}_1 \equiv \hat{\boldsymbol{\Omega}}_A \boldsymbol{W}_S \hat{\boldsymbol{\Omega}}_A$, we have $\hat{\boldsymbol{J}}_1 \xrightarrow{p} \boldsymbol{J}_1$ as $K_C \to \infty$. From this and assumptions **A1–A5** and using $K_S \to \infty$ as $K_C \to \infty$, we have the following results as $K_C \to \infty$:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{AGG} - \boldsymbol{\beta} &= (\bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{\Omega}}_A^{-T/2} \bar{\boldsymbol{W}}_S \\
&\quad \hat{\boldsymbol{\Omega}}_A^{-1/2} \bar{\boldsymbol{X}}_C)^{-1} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{\Omega}}_A^{-T/2} \bar{\boldsymbol{W}}_S \hat{\boldsymbol{\Omega}}_A^{-1/2} (\bar{\boldsymbol{X}}_S \boldsymbol{\beta} + \bar{\boldsymbol{U}}_S) - \boldsymbol{\beta} \\
&= (\bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 \bar{\boldsymbol{X}}_C)^{-1} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 (\bar{\boldsymbol{X}}_S \boldsymbol{\beta} + \bar{\boldsymbol{U}}_S) - \boldsymbol{\beta} \\
&= (K_S^{-1} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 \bar{\boldsymbol{X}}_C)^{-1} K_S^{-1} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 (\bar{\boldsymbol{X}}_S - \bar{\boldsymbol{X}}_C) \boldsymbol{\beta} \\
&\quad + (K_S^{-1} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 \bar{\boldsymbol{X}}_C)^{-1} K_S^{-1} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 \bar{\boldsymbol{U}}_S \\
&\xrightarrow{p} 0_L, \qquad (9)\\
\sqrt{K_S}(\hat{\boldsymbol{\beta}}_{AGG} - \boldsymbol{\beta}) &= (K_S^{-1} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 \bar{\boldsymbol{X}}_C)^{-1} K_S^{-1/2} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 (\bar{\boldsymbol{X}}_S - \bar{\boldsymbol{X}}_C) \boldsymbol{\beta} \\
&\quad + (K_S^{-1} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 \bar{\boldsymbol{X}}_C)^{-1} K_S^{-1/2} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 \bar{\boldsymbol{U}}_S \\
&\xrightarrow{d} \mathcal{N}(0, \Xi_1^{-1}(\Xi_0 + \Xi_2)\Xi_1^{-1}), \qquad (10)
\end{aligned}$$

where $\Xi_0 \equiv \lim_{K_S \to \infty} K_S^{-1/2} \bar{\boldsymbol{X}}_C^T \hat{\boldsymbol{J}}_1 (\bar{\boldsymbol{X}}_S - \bar{\boldsymbol{X}}_C) \boldsymbol{\beta}\boldsymbol{\beta}^T (\bar{\boldsymbol{X}}_S - \bar{\boldsymbol{X}}_C)^T \hat{\boldsymbol{J}}_1 \bar{\boldsymbol{X}}_C$. Notice that the between-cluster variations in the regressors are already differenced out in $(\bar{\boldsymbol{X}}_S - \bar{\boldsymbol{X}}_C)$. Therefore, $(\bar{\boldsymbol{X}}_S - \bar{\boldsymbol{X}}_C)$ is influenced by the within-cluster variation in the regressor. With a sufficiently large fraction of households sampled in each cluster, the contribution of $\Xi_1^{-1}\Xi_0\Xi_1^{-1}$ to the asymptotic variance of $\hat{\boldsymbol{\beta}}_{AGG}$ is small relative to $\Xi_1^{-1}\Xi_2\Xi_1^{-1}$. Hence, the asymptotic variance of $\hat{\beta}_{AGG}$ is approximately equal to $\Xi_1^{-1}\Xi_2\Xi_1^{-1}$. □

Proof of Proposition 2 First, note that N_D is approximately proportionate to K_D and that $N_D \to \infty$ as $K_D \to \infty$ for $D \in \{C, S\}$. The consistency and asymptotic normality of the TS2SLS estimator can be derived from a variant of the standard argument (White 1984, Chap. 5). Under assumptions **A1**-**A2** and **A6**-**A9** and using $K_S \to \infty$ as $K_C \to \infty$, we have the following results as $K_S \to \infty$:

$$\begin{aligned}
&\hat{\boldsymbol{\beta}}_{TS2SLS} - \boldsymbol{\beta} \\
&\quad = (N_C^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{X}_C)^{-1} (N_C^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{Z}_C)(N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{Z}_S)^{-1} N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{X}_S \boldsymbol{\beta} \\
&\quad\quad + (N_C^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{X}_C)^{-1} (N_C^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{Z}_C)(N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{Z}_s)(N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{U}_s) - \boldsymbol{\beta} \\
&\quad \xrightarrow{p} (\boldsymbol{Q}_0^{-1} \boldsymbol{Q}_1 \boldsymbol{Q}_1^{-1} \boldsymbol{Q}_0 - I)\boldsymbol{\beta} = 0, \\
&\sqrt{K_C}(\boldsymbol{\beta}_{TS2SLS} - \boldsymbol{\beta}) \\
&\quad = \sqrt{K_C} \left[(N_C^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{X}_C)^{-1} (N_C^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{Z}_C)(N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{Z}_S)^{-1} N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{X}_S - I \right] \boldsymbol{\beta} \\
&\quad\quad + \sqrt{K_C} \left[(N_C^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{X}_C)^{-1} (N_C^{-1} \boldsymbol{Z}_C^T \boldsymbol{W}_C \boldsymbol{Z}_C)(N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{Z}_S)(N_S^{-1} \boldsymbol{Z}_S^T \boldsymbol{W}_S \boldsymbol{U}_S) \right] \\
&\quad \xrightarrow{d} \mathcal{N}(0, \boldsymbol{Q}_0^{-1} \boldsymbol{Q}_2 \boldsymbol{Q}_0^{-1}).
\end{aligned}$$

Notice that $\boldsymbol{Q}_1$ does not show up in the final expression because $(N_C^{-1}\boldsymbol{Z}_C^T\boldsymbol{W}_C\boldsymbol{Z}_C)(N_S^{-1}\boldsymbol{Z}_S^T\boldsymbol{W}_S\boldsymbol{Z}_S)^{-1} \xrightarrow{p} \boldsymbol{Q}_1\boldsymbol{Q}_1^{-1} = I$ as $K_C \to \infty$. □

References

Abualghaib O, Groce N, Simeu N, Carew MT, Mont D (2019) Making visible the invisible: why disability-disaggregated data is vital to “leave no-one behind”. Sustainability 11(11):3091

Angrist J, Krueger A (1992) The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. J Am Stat Assoc 87(418):328–336

Araujo MC, Ferreira FHG, Lanjouw P, Özler B (2008) Local inequality and project choice: theory and evidence from Ecuador. J Public Econ 92(5–6):1022–1046

Bjorklund A, Jantti M (1997) Intergenerational income mobility in Sweden compared to the USA. Am Econ Rev 87(5):1009–1018

Braithwaite J, Mont D (2009) Disability and poverty: a survey of World Bank Poverty Assessments and implications. ALTER Eur J Disabil Res 3:219–232

Collier P, Radwan S, Wangwe S, Wagner A (1986) Labour and poverty in rural Tanzania: Ujamaa and rural development in the United Republic of Tanzania. Oxford University Press, Oxford

Currie J, Yelowitz A (2000) Are public housing projects good for kids? J Public Econ 75:99–124

de Weerdt J (2010) Moving out of poverty in Tanzania: evidence from Kagera. J Dev Stud 46(2):331–349

Dee TS, Evans WN (2003) Teen drinking and educational attainment: evidence from two-sample instrumental variable estimates. J Law Econ 21(1):178–209

Demombynes G, Özler B (2005) Crime and local inequality in South Africa. J Dev Econ 76(2):265–292

Elbers C, Lanjouw JO, Lanjouw P (2002) ‘Micro-level estimation of welfare.’ Policy Research Department Working Paper 2911, The World Bank

Elbers C, Lanjouw JO, Lanjouw P (2003) Micro-level estimation of poverty and inequality. Econometrica 71(1):355–364

Elbers C, Lanjouw JO, Lanjouw P (2005) Imputed welfare estimates in regression analysis. J Econ Geogr 5(1):101–118

Elbers C, Lanjouw P, Mistiaen A, Özler B, Simler K (2004) On the unequal inequality of poor communities. World Bank Econ Rev 18(3):401–421

Elbers C, Fujii T, Lanjouw P, Özler B, Yin W (2007) Poverty alleviation through geographic targeting. J Dev Econ 83(1):198–213

Elwan A (1999) ‘Poverty and disability: a survey of the literature.’ SP Discussion Paper 9932, The World Bank

Feige EL, Watts HW (1972) An investigation of the consequences of partial aggregation of micro-economic data. Econometrica 40(2):343–360

Filmer D (2008) Disability, poverty and schooling in developing countries: results from 14 household surveys. World Bank Econ Rev 22(1):141–163

Foster J, Greer J, Thorbecke E (1984) A class of decomposable poverty measures. Econometrica 52(3):761–766

Fujii T (2008) How well can we target aid with rapidly collected data?: empirical results for poverty mapping from Cambodia. World Dev 36(10):1830–1842

Fujii T (2010) Micro-level estimation of child undernutrition indicators in Cambodia. World Bank Econ Rev 24(3):520–553

Fujii T (2019) Regional prevalence of health worker absenteeism in Tanzania. Health Econ 28(2):311–316

Grech S (2016) Disability and poverty: complex interactions and critical reframings. In: Grech and Soldatic (eds) Disability in the Global South: the Critical Handbook. Springer. chapter 14, pp 217–236

Groce NE, Mont D (2017) Counting disability: emerging consensus on the Washington Group questionnaire. Lancet 5:E649–E650

Haveman R, Wolfe B (2000) The economics of disability and disability policy. In: Culyer AJ, Newhouse JP (eds) Handbook of Health Economics: Volume 1B, pp 995–1051

Hoogeveen J (2005) Measuring welfare for small but vulnerable groups poverty and disability in Uganda. J Afr Econ 14(4):603–631

Inoue A, Solon G (2010) Two-sample instrumental variable estimators. Rev Econ Stud 92(3):557–561

Ivaschenko O, Lanjouw P (2010) A new approach to producing geographic profiles of HIV prevalence: an application to Malawi. World Med Health Policy 2(1):235–266
Kisanji J (1995) Interface between culture and disability in the Tanzanian context: part I. Int J Disabil Dev Educ 42(2):93–108
Kisanji J (1995) Interface between culture and disability in the Tanzanian context: part II. Int J Disabil Dev Educ 42(2):109–124
Loeb ME, Eide AH (2004) Living conditions among people with activity limitations in Malawi: a national representative study. SINTEF Report STF78 A044511, SINTEF Health Research, Oslo, Norway
Loeb ME, Eide AH (2006) Living conditions among people with activity limitations in Zambia: a national representative study. SINTEF Report SINTEF A262, SINTEF Health Research, Oslo, Norway
Loyalka P, Liu L, Chen G, Zheng X (2014) The cost of disability in China. Demography 51:97–118
Lusardi A (1996) Permanent income, current income, and consumption: evidence from two panel data sets. J Bus Econ Stat 14(1):81–90
Masset E, White H (2004) Are chronically poor people being left out of progress towards the Millennium Development Goals? a quantitative analysis of older people, disabled people and orphans. J Hum Dev 5(2):279–297
Mitra S (2004) Disability and social safety nets in developing countries. SP Discussion Paper 0509, The World Bank
Mitra S (2018) Disability, health and human development. Palgrave, New York
Mitra S, Palmer M, Kim H, Mont D, Groce N (2017) Extra costs of living with a disability: a review and agenda for research. Disabil Health J 10:475–484
Mont D, Nguyen C (2018) Spatial variation in the poverty gap between people with and without disabilities: evidence from Vietnam. Soc Indic Res 137:745–763
Mont D, Cuong NV (2011) Disability and poverty in Vietnam. World Bank Econ Rev 25(2):323–359
National Bureau of Statistics (2002) Household Budget Survey 2000/01 (Dar es Salaam, Tanzania: National Bureau of Statistics, United Republic of Tanzania)
National Bureau of Statistics (2003a) 2002 Population and Housing Census: General Report (Dar es Salaam, Tanzania: National Bureau of Statistics, United Republic of Tanzania)
National Bureau of Statistics (2003b) 2002 Population and Housing Census: Volume I Methodology Report (Dar es Salaam, Tanzania: National Bureau of Statistics, United Republic of Tanzania)
Polinsky AM (1977) The demand for housing: a study in specification and grouping. Econometrica 45(2):447–461
Sohnesen TP, Ambel AA, Fisker P, Andrews C, Khan Q (2017) Small area estimation of child undernutrition in Ethiopian woredas. PLoS ONE 12(4):e0175445
Taylor HR, Katala S, Muñoz B, Turner V (1991) Increase in mortality associated with blindness in rural Africa. Bull World Health Organ 69(3):335–338
UNICEF (1999) Children in Need of Special Protection Measures. Dar es Salaam, Tanzania
United Nations (2019) Disability and Development Reprot 2018: Realizing the Sustainable Development Goals by, for and with persons with disabilities. New York: United Nations
Welsch RE, Kuh E (1976) The variances of regression coefficient estimates using aggregate data. Econometrica 44(2):353–363
White H (1984) Asymptotic theory for econometricians. Academic Press, London
World Bank (2015) Tanzania Mainland Poverty Assessment. Washington DC, USA: World Bank. Available from http://hdl.handle.net/10986/22021
World Health Organization (2011) World Report on Disability. World Health Organization
Yeo R, Moore K (2003) Including disabled people in poverty reduction work: "nothing about us, without us". World Dev 31(3):571–590
Zaidi A, Burchardi T (2005) Comparing incomes when needs differ: Equivalization for extra costs of disability in the U.K. Rev Income Wealth 51(1):89–114