DISCUSSION PAPER SERIES

IZA DP No. 14631

Data Scarcity and Poverty Measurement

Hai-Anh H. Dang
Peter F. Lanjouw

AUGUST 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14631

Data Scarcity and Poverty Measurement

Hai-Anh H. Dang
World Bank, IZA, Indiana University, Vietnam National University and Vietnam Academy of Social Sciences

Peter F. Lanjouw
Vrije Unversiteit

AUGUST 2021

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Data Scarcity and Poverty Measurement*

Measuring poverty trends and dynamics is an important undertaking for poverty reduction policies, which is further highlighted by the SDG goal 1 on eradicating poverty by 2030. We provide a broad overview of the pros and cons of poverty imputation in data-scarce environments, update recent review papers, and point to the latest research on the topics. We briefly review two common uses of poverty imputation methods that aim at tracking poverty over time and estimating poverty dynamics. We also discuss new areas for imputation.

JEL Classification: C15, I32, O15

Keywords: poverty, imputation, consumption, wealth index, synthetic panels, household survey

Corresponding author:
Hai-Anh H. Dang
Data Production & Methods Unit Development Data Group
World Bank
1818 H St. N.W.
Washington, D.C. 20433
USA
E-mail: hdang@worldbank.org

* This is a forthcoming chapter in Jacques Silber (Eds.) "*Handbook of Research on Measuring Poverty and Deprivation*" with Edward Elgar Press. We would like to thank the UK Foreign Commonwealth and Development Office (FCDO) for funding assistance through a Knowledge for Change (KCP) Research Program.

1. Introduction

The design and formulation of poverty reduction policies is contingent on credible poverty measurement. Accurate tracking of poverty trends allows policy makers to monitor progress and to establish whether the fruits of economic growth are widely shared. Similarly, poverty measurement can assist in the identification of pockets of poverty amongst particular population groups, potentially informing the design of targeting strategies. At the global level, monitoring progress in poverty reduction across countries has been underscored as a key task in support of the first Sustainable Development Goal aimed at global poverty eradication.

The ongoing Covid-19 pandemic has further highlighted the need for accurate and timely poverty data. The pandemic has resulted in unprecedented increases in global poverty and has disproportionately impacted on the poor – in both high-income and low-income countries (Sumner *et al*., 2020; Dang *et al*., 2020). The development of appropriate and effective policy responses depends crucially on real-time, reliable, evidence on poverty outcomes.

Yet poverty measurement in many settings around the world is hamstrung by a lack of high quality data. Consumption (or income) survey data generally underpin poverty measurement efforts. Yet in many countries such data are only infrequently collected – and occasionally not available altogether. In addition, where multiple rounds of survey data over time are available and are scrutinized for evidence on poverty trends, it is often found that the data are not strictly comparable. There is growing awareness that even minor departures from strict comparability of underlying data—due to changes in questionnaire or survey design, organization of fieldwork, application of data entry and cleaning protocols—can seriously compromise the comparability of resultant poverty estimates (Lanjouw and Lanjouw, 2001; Beegle *et al*., 2018) Instead of contributing to a better understanding and assessment of distributional outcomes, flawed and problematic data can end up seriously – and dangerously - misinforming.

Recent years have seen poverty economists increasingly resorting to the application of imputation methods as a means to probe and potentially address the challenge of missing and/or problematic data. The methods used vary with the specific data challenge that is being confronted, but the general approach is to impute consumption (and/or income), or a specific poverty measure, from one data source into another based on an estimated relationship involving a set of predictors available in both data sources. The methods have been employed in a wide range of applications aimed at generating comparable poverty estimates across data sets that are *prime facie* non-comparable, or overcoming constraints on comparability posed by a lack of reliable price deflators. They have been used to generate poverty estimates at the subnational level where survey data are not suited to estimation at such levels of disaggregation. They have also seen increased use in recent years for the construction of synthetic panels, where the absence of reliable panel data have prevented the analysis of poverty dynamics.

While these imputation methods appear to offer a means to overcome a range of fundamental challenges in conventional poverty measurement, they are themselves accompanied by important caveats. Notably, depending on the specific application, the methods are predicated on underlying assumptions. These can be quite strong, and are not always readily testable with available data. There is always a risk of unsound inferences if the methods are not judiciously employed, with the necessary care and validation work

In this chapter we provide an overview of the advantages and disadvantages associated with employment of poverty imputation in a variety of data-scarce settings. The chapter updates the discussion offered in two recent review papers by Dang *et al*. (2019) and Dang (2020). We briefly review two common uses of poverty imputation methods: tracking poverty over time, and analyzing poverty dynamics. We highlight the underlying assumptions on which these two applications are predicated and point to emerging experience with the methods from recent studies. We end with a brief discussion of possible new directions that might benefit from imputation methods.

2. Pros and Cons of Imputation

Poverty imputation methods can offer a cost-effective solution to a variety of data-related challenges and constraints, and are consequently seeing widespread employment by national statistical offices and international agencies including World Bank.[1] These advantages relate to a variety of applications that can be listed below, in roughly decreasing order of common use.

a. fill in missing data gaps in the immediate term, especially for poorer countries
b. provide an alternative solution for expensive survey costs and/or survey implementation logistics
c. overcome issues of incomparable survey designs or non-available price deflators
d. back-cast consumption from a more recent survey for better comparison with older surveys

Some remarks on these cases are in order. Cases (a) and (b) are closely related and are the main driving factor behind poverty imputation. This is particularly relevant for poorer countries, since in almost all these countries household consumption surveys are fielded only very occasionally due to financial and logistical challenges. A recent survey by Beegle *et al.* (2016) indicates that just slightly more than half (i.e., 27) of the 48 countries in Sub-Saharan Africa had two or more comparable household surveys for the period between 1990 and 2012. Poverty imputation methods can help fill the data gaps in these contexts.

---

[1] It is related to the established statistical literature on small area estimation, which have become an integral component in the toolbox of agencies such as U.S. Census Bureau.

Case (c) achieved notoriety as a result of a prominent debate on poverty in India where changes to the recall periods for household durables and food items in the 1990s were suggested to result in the overestimation of poverty decline between 1999/00 and 2004/5 (Deaton and Kozel 2005). The debate was renewed, albeit much less heated, in the 2000s.[2] In such situations, poverty imputation has been employed as a means to provide alternative poverty estimates that can be scrutinized to assess the impact of changes in questionnaire design (Dang and Lanjouw, 2018). Another useful application of poverty imputation is allow analysts to avoid reliance on, possibly dubious, externally-sourced (intertemporal and intraregional) price deflators. Such deflators are widely applied to track poverty over time, in the face of inflation, as well as for cross-country comparisons that require different currencies to be converted to the same base.

Case (d), although less common, represents the scenario where poverty imputation is the only route to providing comparable poverty estimates for surveys fielded in the past. Besides challenges with survey design changes, various other changes such as data collection modes (e.g., the switch from paper-based interviews to computer-based interviews) or seasonality (i.e., surveys in agrarian societies being collected at different points during the crop cycle) may be encountered, and be addressed via imputation.

Despite their promise, the application of poverty imputation methods is conditional on important caveats. A key caveat is that these are model-based approaches. Consequently, the underlying modelling assumptions should be carefully assessed, and ideally validated, before the approach is employed. One example is that due to rapid technological advances, high-tech products such as cell phones are no longer the luxury good that they were decades—or even just a few

[2] Survey design issues that compromise the comparability of poverty estimates are common than one might think. They are also found in various countries such as China (Gibson *et al.*, 2003), Tanzania (Beegle *et al.*, 2012), and Vietnam (World Bank, 2012).

years—ago. Consequently, the relationship between ownership of these products and poverty is likely to have changed significantly. This argues against the casual employment of such variables in the imputation model. Another caveat is that applying poverty imputation methods is predicated on certain levels of statistical and data-analysis training and experience of local staff.

3. Typology of Data-Scarce Situations

Table 1 classifies poverty imputation situations according to the degree of missing consumption data, in a roughly decreasing order of severity: completely missing (Category A), partially missing (Category B), and available cross-sectional data but missing panel data (Category C). Table 1 also lists the typical poverty imputation (or corresponding data) situation, examples of surveys, as well as some examples of recent studies that correspond to each of the missing data categories.

Notably, Category A can be further broken down into two data scenarios, where the available survey: (i) produces no consumption data; or (ii) is designed for project targeting purposes. Examples that correspond to these scenarios are the Demographic Health Surveys (DHSs) and most small-scale (or sub-national) surveys. Category B can also be further disaggregated into three different but related data scenarios, where consumption (or income) data are: (i) non-comparable across survey rounds; (ii) unavailable in the current survey but available in some other related surveys; or (iii) unavailable at more disaggregated administrative levels than those offered in the current survey. Finally, Category C addresses the widespread situation that most surveys in developing countries do not provide (nationally representative) household panel data.[3]

[3] Notably, the examples are for presentation purposes and can overlap. For example, depending on the discussed poverty imputation methods, a DHS can appear in either Category A (i.e., generate a wealth index when consumption data are completely missing) or Category B (i.e., implement imputation when consumption data are partially missing). Category B (iii) is also known as small-area estimation (or poverty mapping) and is discussed in more details in another chapter of this handbook.

The decreasing order of the severity of missing consumption data in Table 1 roughly corresponds to an increasing order of country incomes. In particular, Category A and Category B (i) mostly pertain to low-income or lower-middle-income countries, where resources and technical constraints either hinder the collection of consumption data or render such processes not fully effective. Category B (ii) and (iii) can apply to lower-middle-income or upper-middle-income countries. In fact, Category B (iii) also applies to a high-income country context such as the U.S. where there is an interest to provide poverty (and income) estimates at more disaggregated levels. Finally, Category C is also relevant for middle-income and high-income countries where there are no panel data, or the panel data suffer from quality issues. As an example, due to attrition, the percentage of households that remain in the Russia Longitudinal Monitoring Survey (RLMS) panel in the first 10 years after it was fielded is around 60 percent; this figure further decreases by half to 29 percent after another 10 years (Kozyreva *et al.*, 2016).

Other classifications may also be used besides that offered in Table 1. For example, using the time dimension poverty imputation methods can be classified into across-year imputation and within-year imputation. While across-year imputation typically offers estimates that monitor poverty trend over time, within-year imputation usually offers estimates that are relevant to small-scale projects (including project targeting). Another classification is based on survey types and includes across-survey imputation and within-survey imputation. Clearly, imputation from one survey into another often requires more careful data checks and invokes stronger modelling assumptions than imputation from one round to another of the same survey type. Furthermore, appropriate adjustments (e.g., standardizing the variables) may need be conducted for imputation using surveys of different designs (Dang *et al.*, 2017).

4. Workhorse Equations and Key Imputation Challenges

Household consumption is typically estimated using the following reduced-form linear model

$$y_{ij} = \beta' x_{ij} + \mu_{ij} \qquad (1)$$

for household *i* in survey *j*, for *i*= 1,…, N (see, e.g., Elbers *et al.* (2003), Ravallion (2016)). x_{ij} can include household variables such as the household head's age, sex, education, ethnicity, religion, language (i.e., which can represent household tastes), occupation, and household assets or incomes. We will examine the extensions of Equation (1) in the following two situations, tracking poverty trends and measuring poverty dynamics (which respectively correspond to Categories B and C in Table 1).

4.1. Tracking poverty trends

We extend Equation (1) as follows

$$y_j = \beta_j' x_j + \upsilon_{cj} + \varepsilon_j \qquad (2)$$

where the error term μ_{ij} is broken down into two components, one (υ_{cj}) a cluster random effects and the other (ε_j) the idiosyncratic error term. We suppress the subscript that indexes households to make the notation less cluttered in this sub-section. Conditional on household characteristics, the cluster random effects and the error terms are usually assumed uncorrelated with each other and to follow a normal distribution such that $\upsilon_{cj}|x_j \sim N(0, \sigma_{\upsilon_j}^2)$ and $\varepsilon_j|x_j \sim N(0, \sigma_{\varepsilon_j}^2)$. While the normal distribution assumption results in the standard linear random effects model that is more convenient for mathematical manipulations and computation, an alternative modelling option is to draw from the empirical distribution of the error terms.

Assume that the explanatory variables x_j are comparable for both surveys (Assumption 1), and that the changes in x_j between the two periods can capture the change in poverty rate in the next period (Assumption 2). Dang *et al.* (2017) define the imputed consumption y_2^1 as

$$y_2^1 = \beta_1' x_2 + \upsilon_1 + \varepsilon_1 \quad (3)$$

and estimate it as

$$\hat{y}_{2,s}^1 = \hat{\beta}_1' x_2 + \tilde{\hat{\upsilon}}_{1,s} + \tilde{\hat{\varepsilon}}_{1,s} \quad (4)$$

where β_1' and the distributions of the error terms υ_1 and ε_1 are estimated using Equation (2). $\tilde{\hat{\upsilon}}_{1,s}$ and $\tilde{\hat{\varepsilon}}_{1,s}$ represent the s^{th} random draw (simulation) from their estimated distributions, for *s*= 1,…, S. We suggest using 1,000 simulations or more. The poverty rate in period 2 and its variance can then be estimated as

i) $\hat{P}_2 = \frac{1}{S}\sum_{s=1}^{S} P(\hat{y}_{2,s}^1 \le z_1)$ (5)

ii) $V(\hat{P}_2) = \frac{1}{S}\sum_{s=1}^{S} V(\hat{P}_{2,s}|x_2) + V(\frac{1}{S}\sum_{s=1}^{S} \hat{P}_{2,s}|x_2)$ (6)

This imputation method is related to but improves on the widely-used proxy means testing approach. The predicted household consumption generated using survey-to-survey imputation methods is composed of *both* the two terms on the right-hand side of Equation (1), that is $\hat{\beta}' x_{ij}$ and $\hat{\mu}_{ij}$. In contrast, the predicted household consumption (or wealth) with the traditional proxy means testing approach only uses the term $\hat{\beta}' x_{ij}$. Theoretical and empirical evidence suggest that including the predicted error term $\hat{\mu}_{ij}$ could help to significantly improve prediction accuracy (Dang *et al.*, 2017; Dang *et al.*, 2019). For consistency, the poverty line in the base survey—rather than that in the target survey—should be used together with the predicted consumption to obtain poverty estimates.

Two key modelling challenges are associated with Equation (1). First, the coefficients β estimated from the previous consumption survey can be combined with the variables in the more recent survey to obtain poverty estimates. This is often referred to as the constant parameter assumption.[4] Second, good model selection is crucial for obtaining accurate estimates. Meta-analysis of estimates using data from various countries suggests that imputation models that include household assets and housing characteristics or utilities expenditure appear to perform best (Christiaensen *et al.*, 2012; Carletto *et al.*, 2021). We provide a brief review of selected studies on poverty imputation with validation in the past 20 years in Appendix A, Table A.1.

4.2. Measuring poverty dynamics

To express the equations for measuring poverty dynamics, we return to the notations with Equation (1). Let x_{ij} be a vector of *time-invariant* household characteristics that are observed in both survey rounds. Subject to data availability, these characteristics could include such variables as sex, ethnicity, religion, language, place of birth, and parental education as well as variables that can be converted into time-invariant versions based, for example, on information about household heads' age and education. The vector x_{ij} can also include time-varying household characteristics if retrospective questions about the round-1 values of such characteristics are asked in the second round survey.

Let z_j be the poverty line in period j. We are interested in knowing the unconditional measures of poverty mobility such as

$$P(y_{i1} < z_1 \text{ } and \text{ } y_{i2} > z_2) \quad (7)$$

which represents the percentage of households that are poor in the first survey round (year) but nonpoor in the second survey round, or the conditional measures such as

[4] Notably, this assumption is also needed for consistency for the $\hat{\beta}' x_{ij}$ part in the case of proxy means tests. While concerns exist that this assumption is likely to be valid only under normal circumstances, rather than during periods of fast (economic growth and) poverty reduction, it has been found to hold during a period of dramatic economic growth in China and Vietnam where poverty incidence was cut by around half (Christiaensen *et al.*, 2012). Furthermore, a weaker version of this assumption has been proposed and validated for data from various countries such as India, Jordan, and Vietnam (Dang *et al.*, 2017; Dang and Lanjouw, 2018; Dang *et al.*, 2019).

$$P(y_{i2} > z_2 | y_{i1} < z_1) \quad (8)$$

which represents the percentage of poor households in the first round that escape poverty in the second round.

If panel data are available, we can estimate the quantities in (7) and (8); but in the absence of such data, we can use synthetic panels to study mobility. To operationalize the framework, we make two standard assumptions. First, we assume that the underlying populations being sampled in survey rounds 1 and 2 are identical, such that their characteristics remain time-invariant (Assumption 3). More specifically, coupled with Equation (1), this implies the conditional distribution of expenditure in a given period is identical whether it is conditional on the given household characteristics in period 1 or period 2 (i.e., $x_{i1} \equiv x_{i2}$ implies $y_{i1}|x_{i1}$ and $y_{i1}|x_{i2}$ have identical distributions). Second, we assume that μ_{i1} and μ_{i2} have a bivariate normal distribution with positive correlation coefficient ρ and standard deviations σ_{μ_1} and σ_{μ_2} respectively (Assumption 4). (Note that we refer to these assumptions as Assumptions 3 and 4 for presentation purposes only, they are not related to Assumptions 1 and 2 discussed earlier). Given these assumptions, Quantity (7) can be estimated by

$$P(y_{i1} < z_1 \; and \; y_{i2} > z_2) = \Phi_2(\frac{z_1 - \beta_1' x_{i2}}{\sigma_{\mu_1}}, -\frac{z_2 - \beta_2' x_{i2}}{\sigma_{\mu_2}}, -\rho) \quad (9)$$

where $\Phi_2(.)$ stands for the bivariate normal cumulative distribution function) (and $\phi_2(.)$ stands for the bivariate normal probability density function) (Dang *et al.*, 2014; Dang and Lanjouw, 2013). Note that in Equation (9), the estimated parameters obtained from data in both survey rounds are applied to data from the second survey round (x_2) (or the base year) for prediction, but we can use data from the first survey round as the base year as well. It is then straightforward to estimate quantity (8) by dividing quantity (7) by $\Phi(\frac{z_1-\beta_1' x_{i2}}{\sigma_{\mu_1}})$, where $\Phi(.)$ stands for the univariate normal cumulative distribution function (cdf).

Compared to tracking poverty trends over time, estimating poverty dynamics does not require the assumption of constant parameters β. But the additional assumptions that the time-invariant characteristics remain the same over time, and that a good estimate of the correlation coefficient ρ is available (Assumptions 3 and 4 above) are critical. Additional technical details, alternative methods, and also limitations, are discussed in various studies (Dang and Lanjouw, 2013; Bourguignon and Dang, 2019; Moreno *et al.*, 2021, and Herault and Jenkins, 2018). Recent validations and applications of synthetic panel methods by various researchers for different country contexts ranging from India to Africa, Latin America, and Europe have yielded encouraging results (Ferreira *et al.*, 2012; Beegle *et al.*, 2016; UNDP, 2016; OECD, 2018; Dang *et al.*, 2019).

5. Conclusion

We end this paper by pointing to a few promising directions for future research both in terms of topics and methods. Regarding topics, poverty imputation methods can be quite useful in helping to estimate and track poverty for marginalized groups such as forcefully displaced populations and refugees. As the number of refugees is increasing worldwide, estimating poverty for these disadvantaged populations is receiving growing attention. However, this task is challenged by the fact that these populations are scattered in hard-to-reach places and are not commonly included in the sampling frame of the host countries. Recent studies suggest that a judicious combination of household consumption surveys, administrative data and imputation methods can provide poverty estimates for refugees (Beltramo *et al.*, 2020; Dang and Verme, 2021).

Poverty imputation can potentially be further improved following recent statistical advances. For example, recent evidence suggests that machine learning techniques such as lasso could result in good targeting estimates (Altindag *et al.*, 2021). Another study by Steele *et al.* (2017) applies machine learning techniques and big data (i.e., cell phone and satellite data) to evaluate poverty mapping. Early adopters of multiple imputation (MI) techniques, which are well developed in the statistical literature, suggest MI may offer an alternative method (Douidich *et al.*, 2016; Dang *et al.*, 2017).

Finally, a novel idea recently put forward proposes to employ rapid assessment surveys (e.g., 60-minute style surveys) in combination with imputation methods to make better use of the advantages of both (Pape, 2021). But more validations would certainly be needed for further progress in this direction.

References

Beegle, K., Christiaensen, L., Dabalen, A., & Gaddis, I. (2016). *Poverty in a rising Africa*. World Bank: Washington, DC.

Beegle, K., de Weerdt, J., Friedman, J., and Gibson, J. (2012) Methods of Consumption Measurement Through Surveys: Experimental Evidence from Tanzania, *Journal of Development Economics*, 98(1), 3-18.

Beltramo, T., Dang, H.-A., Sarr, I., & Verme, P. (2020). *Is imputing poverty efficient? An example from refugee data in Chad*. World Bank Policy Research Paper # 9222.

Bourguignon, F. & Dang, H.-A. (2019). "Investigating Welfare Dynamics with Repeated Cross Sections: A Copula Approach". *Paper presented at the WB-IARIW conference*, Washington DC.

Brown, C., Ravallion, M., & van de Walle, D. (2018). A poor means test? Econometric targeting in Africa. *Journal of Development Economics, 134*, 109–124.

Carletto, C., Dang, H.-A., & Kilic, T. (2021). "Poverty Imputation in Contexts without Consumption Data: A Revisit with Further Refinements". Working paper. World Bank: Washington, DC.

Christiaensen, L., Lanjouw, P., Luoto, J., & Stifel, D. (2012). Small area estimation-based prediction methods to track poverty: Validation and applications. *Journal of Economic Inequality, 10*, 267–297.

Coady, D., Grosh, M., & Hoddinott, J. (2004). Targeting outcomes redux. *World Bank Research Observer, 19*, 61–85.

Dang, H.-A. (2020). "To Impute or Not to Impute, and How? A Review of Alternative Poverty Estimation Methods in the Context of Unavailable Consumption Data". *Development Policy Review*. doi: https://onlinelibrary.wiley.com/doi/abs/10.1111/dpr.12495

Dang, H.-A., & Lanjouw, P. (2013). *Measuring poverty dynamics with synthetic panels based on cross-sections* (World Bank Policy Research Working Paper No. 6504).

---. (2016). Toward a new definition of shared prosperity: a dynamic perspective from three countries. In *Inequality and growth: Patterns and policy* (pp. 151–171). London: Palgrave Macmillan.

---. (2018). Poverty dynamics in India between 2004 and 2012: Insights from longitudinal analysis using synthetic panel data. *Economic Development and Cultural Change, 67*, 131–170.

Dang, Hai-Anh and Paolo Verme. (2021). "Estimating Poverty for Refugee Populations Can Cross-Survey Imputation Methods Substitute for Data Scarcity?" *ECINEQ Working Paper 2021-578*.

Dang, H.-A., Jolliffe, D., & Carletto, C. (2019). Data gaps, data incomparability, and data imputation: A review of poverty measurement methods for data-scarce environments. *Journal of Economic Surveys, 33*, 757–797.

Dang, H.-A., Lanjouw, P., Luoto, J., & McKenzie, D. (2014). Using repeated cross-sections to explore movements into and out of poverty. *Journal of Development Economics, 107*, 112–128.

Dang, H.-A., Lanjouw, P. F., & Serajuddin, U. (2017). Updating poverty estimates in the absence of regular and comparable consumption data: Methods and illustration with reference to a middle-income country. *Oxford Economic Papers, 69*, 939–962.

Dang, H.-A., Huynh, T. L. D., & Nguyen, M. H. (2020). Does the COVID-19 Pandemic Disproportionately Affect the Poor? Evidence from a Six-Country Survey. *IZA Discussion Paper No. 13352.*

Deaton, A., & Kozel, V. (2005). *The great Indian poverty debate*. New Delhi: Macmillan.

Douidich, M., Ezzrari, A., Van der Weide, R., & Verme, P. (2016). Estimating quarterly poverty rates using labor force surveys: a primer. *World Bank Economic Review, 30*, 475–500.

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

Ferreira, F. H., Messina, J., Rigolini, J., López-Calva, L. F., Lugo, M. A., & Vakis, R. (2012). *Economic mobility and the rise of the Latin American middle class*.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography, 38*, 115–132.

Filmer, D., & Scott, K. (2012). Assessing asset indices. *Demography, 49*, 359–392.

Gibson, J., Huang, J., & Rozelle, S. (2003). Improving estimates of inequality and poverty from urban China's household income and expenditure survey. *Review of Income and Wealth, 49*, 53–68.

Grosh, M. E., del Ninno, C., Tesliuc, E., & Ouerghi, A. (2008). *For protection and promotion: The design and implementation of effective safety nets*.

Herault, N., & Jenkins, S. (2019). How Valid are Synthetic Panel Estimates of Poverty Dynamics? *Journal of Economic Inequality*, 17, 51-76.

Kozyreva, P., Kosolapov, M., & Popkin, B. M. (2016). Data resource profile: The Russia Longitudinal Monitoring Survey—Higher school of economics (RLMS-HSE) phase II: Monitoring the economic and health situation in Russia, 1994–2013. *International Journal of Epidemiology, 45*, 395–401.

Lanjouw, J.O., & Lanjouw, P. (2001). How to Compare Apples and Oranges: Poverty Measurement Based on Different Definitions of Consumption, *Review of Income and Wealth,* 47(1), 25-42.

Mathiassen, A. (2009). A model based approach for predicting annual poverty rates without expenditure data. *Journal of Economic Inequality, 7*, 117–135.

Mathiassen, A. (2013). Testing prediction performance of poverty models: Empirical evidence from Uganda. *Review of Income and Wealth, 59*, 91–112.

Moreno, H., Bourguignon, F., & Dang, H.-A. (2021). *On the construction of synthetic panels.* IZA Discussion Paper # 14236.

OECD. (2018). *A broken social elevator? How to promote social mobility.* Paris: OECD.

Pape, U. J. (2021). Measuring Poverty Rapidly Using Within-Survey Imputations. *World Bank Policy Research Paper* No. 9530.

Ravallion, M. (2016). *The economics of poverty: History, measurement, and policy*. New York, NY: Oxford University Press.

Sahn, D. E., & Stifel, D. C. (2000). Poverty comparisons over time and across countries in Africa. *World Development, 28*, 2123–2155.

Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., ... & Hadiuzzaman, K. N. (2017). Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface, 14*, 20160690.

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

Sumner, Andy, Chris Hoy, and Eduardo Ortiz-Juarez. (2020). "Estimates of the Impact of Covid-19 on Global Poverty". *WIDER Working Paper 2020/43*. Helsinki: UNU-WIDER.

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

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

UNDP. (2016). *Multidimensional progress: Well-being beyond income*.

US Census Bureau. (n.d.). Center for Statistical Research and Methodology (CSRM): Small area estimation. Retrieved from the United States Census Bureau website : https://www.census.gov/srd/csrm/SmallArea.html

Table 1. Categories of Missing Household Consumption Data and Recent Sample Studies

Type	Extent of Missing Consumption Data (and Associated Methods)	Typical Situation	Example	Recent Sample Studies
A	Completely missing (e.g., wealth index)	a) Non-consumption surveys	DHSs and most small-scale (or sub-national) surveys	Sahn and Stifel (2000); Filmer and Pritchett (2001); Filmer and Scott (2012)
		b) Proxy means test/project targeting	Most small-scale surveys	Coady, Grosh, and Hoddinott (2004); Grosh *et al.* (2008); Brown, Ravallion, and van de Walle (2018)
B	Partially missing (e.g., imputed consumption)	(a) Consumption data not comparable across survey rounds	Some rounds of India's National Sample Surveys (NSSs)	Tarozzi (2007); Christiaensen *et al.* (2012); Mathiassen (2013)
		(b) Consumption data unavailable in current survey but available in another related survey	The annual LFS does not have consumption data, but the household consumption survey is implemented every few years	Mathiassen (2009); Douidich *et al.* (2016); Dang, Lanjouw, and Serajuddin (2017)
		(c) Consumption data unavailable at more disaggregated administrative levels than those in current survey	Population census data are representative at lower administrative level than a household consumption survey, but do not collect consumption data	Elbers, Lanjouw, and Lanjouw (2003); Elbers *et al.* (2007); Tarozzi and Deaton (2007)
C	Cross sections available, but missing panel data (e.g., synthetic panels)	(a) Most surveys in developing countries do not offer panel data	India's National Sample Surveys, China's Household Income Project (CHIPs)	Dang *et al.* (2014); Dang and Lanjouw (2013); Moreno, Bourguignon, and Dang (2021)
		(b) Some surveys offer short-term panel data only	Vietnam Household Living Standards Surveys (VHLSS)	
		(c) Some surveys offer long-run panels but with high attrition rates or panels that are not nationally representative	Russia's Longitudinal Monitoring Surveys (RLMSs) or Indonesia's Family and Life Surveys (IFLSs)	

Note: DHS and LFS respectively stand for Demographic and Health Surveys and Labor Force Surveys. This table is a modified and expanded version of similar tables in Dang, Jolliffe, Carletto (2019) and Dang (2020). The number of sample studies are restricted to three or fewer to save space. More references are provided in Tables A.1 and A.2 in Appendix A and discussed in the main text.

Appendix A: Additional Tables

Table A.1. Overview of Some Key Poverty Imputation Studies (with Validation) since the 2000s

No	Authors	Data	Estimation method	Main variables in the imputation model	Main findings
1	Stifel and Christiaensen (20(	Kenya's Welfare Monitoring Survey (WMS) in 1997 and Demographic and Health Survey (DHS) in 1993, 1998, 2003	Elber *et al.* 's (2003) method	Housing characteristics (quality of floor, roof, drinking water sources), house durables (ownership of radio, television, refrigerator, bike), cluster characteristics (cluster averages of households with low-quality floors and with access to piped water), and district characteristics (district averages of household with access to electricity, early onset of rainfall, malaria prevalence, household under-five height-for-age z scores).	The imputation-based poverty estimates closely track the survey-based poverty estimates in 1998.
2	Tarozzi (2007)	Indian National Sample Survey 1994/95-1999/2000	Inverse probability weighting	Demographics, education, employment characteristics, scheduled castes or tribe, land ownership, energy source for cooking and for lighting	Predicted poverty estimates are higher than the official poverty rtaes, but follows the same trend.
3	Christiaensen *et al. (* 2012)	Vietnam's VLSS in 1992/93 and 1997/98; Russia's RLMS in 1993, 1998, 2003; China's Gansu and Inner Mongolia survey in 2000/04; Kenya's WMS in 1997 and KIHBS in 2005/06	Elber *et al.* 's (2003) method	Demographics, geographics, education/ profession, location, housing quality, consumer durables, food expenditure (rice and non-rice expenditure), nonfood expenditure (30 day and annual recalls)	The poverty prediction method works well for Vietnam both with models using certain expenditure components (particularly non-rice food spending and non-food spending) and with comprehensive models of non-consumption assets. In rural Gansu and Inner Mongolia, models based on non-expenditure assets work consistently, while models using certain expenditure subcomponents sometimes work.
4	Mathiassen (2013)	Uganda Monitoring Survey (MS) 1-4, Uganda National Household Survey (UNHS) 1-3	Elber et al.'s (2003) method with refinements for estimating the variance of the error term	Demographics, education, employment characteristics, occupation, housing, consumption of food, non-durable and semi-durable expenditures, welfare indicators, and regional dummies. Model is estimated by urban/rural.	Predicted poverty trends are very similar for each survey model regardless of base survey. While in most cases predictions at rural, urban, and subregional levels are in line with the official poverty figures, predicted urban poverty trends follow more closely the actual trends than is the case for rural areas.
5	Daniels and Minot (2015)	Uganda National Household Survey in 2005/06, Demographic and Household Surveys (DHS) in 1995, 2000/2001, 2006 and 2009	Elber *et al.* 's (2003) method	Demographics, ownership of assets (ownership of motorbike, bicycle, tv or radio) and housing characteristics (type of floor, source of water, type of toilet, electricity). Model is estimated by urban/rural and regions separately.	Asset-based poverty estimates in the 2006 DHS are very close to the consumption-based poverty estimates from 2005/06 UNHS. In 2009/2010, however, the asset-based poverty rates using the DHS data are greater than those estimated directly from the UNHS in most regions of the country.
6	Douidich *et al.* (2016)	Morocco 2000/01 National Survey on Consumption and Expenditure (NSCE) and the 2006/07 National Living Standards Survey (NLSS), LFS 2000–2009	Elber *et al.* 's (2003) method	Demographics, education, employment characteristics, household assets and durables (kitchen, douche, tv, parabole), house characteristics (number of rooms, electricity, sewage, drinking water, flush toilet), interactions of urban/rural variable with employment or with house characteristics. Model is estimated by urban/rural separately.	Imputation estimates obtained with the 2001 and 2007 models are very close, but model with assets does not add improvement in poverty estimates. Adding the asset variables improves model 2001's estimate of the 2007 poverty rate but not model 2007's estimate of the 2001 poverty rate. Imputation poverty estimates in LFSs for the period 2001–2009 provide almost overlapping poverty trends using NSCE and NLSS, even when disaggregated by urban and rural areas.
7	Cuesta and Ibarra (2017)	Tunisia National Consumption Survey (ENBCV) of 2010 and the Labor Force Surveys (ENPE) of 2009, 2010 and 2012	Elber *et al.* 's (2003) and Dang *et al.* 's (2017) methods and macro-based projections of the national poverty rates.	Demographics, geographics, education, employment characteristics, access to tap water and electricity, household assets and house durables (ownership of car, motorcycle, and/or bicycle; television and/or radio; washing machine, refrigerator, freezer, dishwasher, or oven), rural/urban location and regional characteristics	1) Estimated consumption model provides a reasonable approximation to the observed poverty rates in 2010. Dang et al., (2014) method of imputation provides a closer estimate of poverty to the official rate in 2010. Random residual imputations and Dang et al., (2014) method of imputation also work well in predicting full consumption distributions. 2) Macro-projections are in line with respect to the survey-to-survey imputation.
8	Dang *et al.* (2017)	Jordan's Household Expenditure and Income Survey (HEIS) 2008 and 2010, Unemployment and Employment Survey (LFS)	Refinements to Elber *et al.* 's (2003) method for survey-to-survey imputation	Demographics (as well as household demographics such as the shares of household members in different age ranges), marital status, nationality, employment characteristics, urban/rural location, household assets, log of income per household member.	Models that include demographic, work sector, household assets, and/or income variables provide reasonable estimates using the consumption data in the HEIS 2008 survey round in combination with the household characteristics in the HEIS 2010 round. Estimates from within-year and across-year imputations from the HEIS into LFS fell within the 95 confidence interval of the true rates.
9	Dang and Lanjouw (2018)	India National Sample Surveys (NSSs) 2009/10 and 2011/12	Dang *et al.* 's (2017) method	Demographics, religion, social classes, education, employment status and work sector, assets, house durables and home ownership, urban/rural location	Imputation method underestimates poverty in 2011/12, but underestimation is not very large. The largest difference between true and imputed poverty rates in models including household assets.
10	Christiaensen *et al.* (2020)	Rwanda: Enquete Int´egrale sur les Conditions de Vie des m´enages de Rwanda (EICV1) 2001 and (EICV2) 2006. Uganda: Uganda National Household Survey (UNHS) 2005/06 and 2009/10. Tanzania: Tanzania National Panel Survey (NPS) 2008/09 and 2010/11	Demand theory, including Engel law to predict linear changes in consumption sub-aggregates	Total number of non-durable consumption items: Tanzania - 112, Rwanda - 284, Uganda - 126. Final number of consumption items to construct linear sub-aggregates: Tanzania - 17, Rwanda - 28, Uganda - 18.	Linear combination of consumption sub-aggregates does not accurately predict poverty headcount in a subsequent period. Estimated poverty headcounts are outside the 95 % CI of the poverty estimates for the full consumption aggregate.

Sources: Based on Carletto *et al.* (2021).