ECINEQ

Society for the Study of Economic Inequality

Working Paper Series

Regression-based Imputation for Poverty Measurement in Data Scarce Settings

Hai-Anh Dang

Peter Lanjouw

ECINEQ 2022 611

ECINEQ 2022 611
2022 April
www.ecineq.org

Regression-based Imputation for Poverty Measurement in Data Scarce Settings

Hai-Anh Dang
World Bank
Peter Lanjouw
Vrije Unversiteit, Amsterdam

Abstract

Measuring poverty trends and dynamics are important inputs in the formulation and design of poverty reduction policies. The empirical underpinnings of such exercises are often constrained by the absence of suitable data. We provide a broad, generalist, overview of regression-based imputation methods that have seen widespread application to estimate poverty outcomes in data-scarce environments. In particular, we review two imputation methods employed in tracking poverty over time and estimating poverty dynamics. We also discuss new areas that promise of further research.

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

JEL Cassification: C15, I32, O15

Regression-based Imputation for Poverty Measurement in Data Scarce Settings

Hai-Anh H. Dang and Peter F. Lanjouw*

January 2022

Abstract

Measuring poverty trends and dynamics are important inputs in the formulation and design of poverty reduction policies. The empirical underpinnings of such exercises are often constrained by the absence of suitable data. We provide a broad, generalist, overview of regression-based imputation methods that have seen widespread application to estimate poverty outcomes in data-scarce environments. In particular, we review two imputation methods employed in tracking poverty over time and estimating poverty dynamics. We also discuss new areas that promise of further research.

Key words: poverty, imputation, consumption, wealth index, synthetic panels, household survey

JEL: C15, I32, O15

* This is a forthcoming chapter in Jacques Silber (Eds.) "*Handbook of Research on Measuring Poverty and Deprivation*" with Edward Elgar Press. Dang (hdang@worldbank.org; corresponding author) is senior economist with the Data Production and Methods Unit, Development Data Group, World Bank and is also affiliated with GLO, IZA, Indiana University, International School, Vietnam National University, Hanoi, and Vietnam Academy of Social Sciences. Lanjouw (p.f.lanjouw@vu.nl) is professor of economics at Vrije Unversiteit, Amsterdam. We are grateful to Jacques Silber and an anonymous reviewer for helpful feedback on earlier drafts. 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 (SDG) aimed at global poverty eradication.

Yet, poverty measurement in many settings around the world, especially in poorer countries, is hamstrung by a lack of suitable consumption (or income) survey data, which generally underpin poverty measurement efforts (Beegle *et al*., 2016). A recent review of the United Nation's SDG database suggests that less than one tenth of the data points on poverty needed to properly monitor poverty trends for all the countries over the period 2000-2018 are available (Dang and Serajuddin, 2020). Furthermore, problems with data quality are also widespread. Where repeated rounds of survey data over time are scrutinized for evidence on poverty trends, the underlying data are often found not to be comparable. 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—have long been recognized to seriously compromise the comparability of resultant poverty estimates (Lanjouw and Lanjouw, 2001; Deaton and Kozel, 2005).

These issues are particularly pressing in the case of panel data. Collecting longitudinal (panel) consumption data usually involves significant resources and demands a high level of technical survey-implementation capacity. As such, panel data are more commonly available for richer countries. These data, too, are not exempt from data quality issues. For 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 first fielded is around 60 percent; this figure further decreases by half to 29 percent after another 10 years (Kozyreva *et al.*, 2016). Instead of contributing to a better understanding and assessment of distributional outcomes, flawed and problematic data can end up seriously – and dangerously - misinforming.

The ongoing Covid-19 pandemic has further added to calls for more accurate and timely poverty data. The pandemic has dramatically impacted on the poor – in both high-income and low-income countries - and has resulted in dramatic increases in global poverty (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.

Experimentation with regression-based imputation methods to assist with the estimation of poverty in the developing country setting surged in the early 2000s. Deaton and Dreze (2002), Deaton (2003), Kijima and Lanjouw (2003) and Tarozzi (2007) provide poverty estimates based on such imputation methods for India, in a context where serious concerns had been raised regarding the comparability of two contiguous rounds of the National Sample Survey Organization's household survey. Recent years have seen poverty economists increasingly resorting to the application of these imputation methods as a means to probe and potentially address the challenge of missing and/or problematic, non-comparable data. Such imputation methods can offer a cost-effective solution to a variety of data-related challenges and constraints, and are consequently seeing increasing application in practice. International agencies including World Bank, in collaboration with national statistical offices, have routinely employed and refined these methods to fill data gaps in poorer countries (Jolliffe *et al.*, 2015; World Bank, 2017). Researchers at agencies such as Asian Development Bank and UNDP have recently adopted similar methods (UNDP, 2016; Jha *et al.*, 2018).[1] Most recently, these methods have been further extended to impute poverty for population groups, such as refugees, that are typically not captured in most standard household consumption surveys (Beltramo *et al.*, 2020; Dang and Verme, 2021).

The regression-based imputation methods aimed at poverty estimation have been employed in a wide range of applications and vary with the specific data challenge that is being confronted. But the general point of departure is to predict consumption (and/or income), or a specific poverty measure, in a target data source, based on an estimated relationship involving a set of predictors that are available in both the target data source and another base data source. This basic method has also seen increased use in recent years for the construction of synthetic panels, where the absence of reliable panel data has constrained the analysis of poverty dynamics. While these imputation methods appear to offer a means to overcome, or at least mitigate, a range of fundamental challenges in conventional poverty measurement, they are themselves 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 updated overview of the advantages and disadvantages associated with employment, in a variety of data-scarce settings, of the particular approach to imputation described above. We classify these settings before briefly reviewing two recent imputation approaches: aimed at, respectively, tracking poverty over time and analyzing poverty dynamics. While these approaches may not account for all the various poverty measurement-

[1] A related area is small area estimation, which provides poverty (income) estimates for lower administrative areas than what are available in the household survey. Small area estimation methods have become an integral component in the toolbox of agencies in richer countries such as U.S. Census Bureau and are discussed in a separate chapter in this book. This is related to an established statistical literature on multiple imputation (e.g., Carpenter and Kenward, 2013; Little and Rubin, 2020). Imputation methods have also applied to study related topics such as labor transitions and welfare mobility in richer countries (OECD, 2015 and 2018).

oriented imputation applications, they have seen increased used in a variety of different settings, ranging from low-income to upper-middle income countries and over different geographical regions. We highlight the key underlying assumptions on which these two approaches are predicated and point to emerging experience with several methods proposed in recent studies. We end with a brief discussion of possible new research directions for imputation methods. The chapter serves as a generalist introduction to the topic of imputation in poverty measurement and refers readers to in-depth treatments elsewhere.[2]

2. Where imputation can be potentially useful

The appeal of resorting to regression-based imputation methods depends on a variety of circumstances. These can be listed below, in roughly decreasing order of common use:

a. To fill in missing data gaps (most commonly in poorer countries)
b. To provide an alternative to conducting new surveys that are prohibitively expensive or for which technical and administrative capacity is unavailable.
c. To overcome issues of non-comparability in existing surveys or to side-step the non-availability of reliable price deflators
d. To back-cast consumption from a more recent to an older 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 most common reason for adopting imputation efforts in poverty measurement. 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

[2] We refer interested readers to more technical details and literature reviews provided in two recent review papers by Dang *et al*. (2019) and Dang (2021).

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. Imputation methods can help fill the data gaps in these contexts.

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 argued to result in the overstatement of poverty decline between 1999/00 and 2004/5 (Deaton and Kozel 2005). The debate was renewed, albeit much less heated, in the 2000s.[3] In such situations, regression-based imputation methods were 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 such imputation approaches is to 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 regression-based imputation provides the only route to producing 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 can potentially be addressed via imputation.

[3] Survey design issues that compromise the comparability of poverty estimates are more common than one might think and are found in various countries including China (Gibson *et al.*, 2003), Tanzania (Beegle *et al.*, 2012), and Vietnam (World Bank, 2012). See Dang and Lanjouw (2021) for more discussion on a typology of poverty imputation situations and recent sample studies.

Despite their promise, the application of these 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, wherever the approach is employed. One example of the issues that can arise can be seen in the context of rapid technological change, where high-tech products such as cell phones are no longer the luxury good that they were decades—or even just a few years—ago. In such settings, the relationship between ownership of these products and economic wellbeing is likely to have changed significantly. This argues against the casual employment of such covariates in the imputation model. Another consideration relates to the fact that the appropriate application of imputation methods is predicated on familiarity with the requisite statistical methods, and involves a degree of data-analysis training and experience of local staff. We discuss some of these key challenges in more detail in the next section.[4]

3. Workhorse equations and key imputation challenges

A household maximizes utility subject to an income budget constraint that includes choice variables such as quantities of goods, durables, and leisure (or labor supply) (Deaton and Muellbauer, 1980). These in turn are determined by different factors, such as household tastes. It follows that a model of (log) household consumption (y_{ij}) is typically estimated using the following reduced-form linear model

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

[4] A variety of statistical issues arise with respect to the specification and estimation of the underlying prediction models. One needs, for example, to consider the criteria for selection of variables into the imputation models and different types of errors (e.g., measurement error in covariates, sampling error, and out-of-sample prediction error). These issues must be attended to with care and transparency. See Elbers, Lanjouw and Lanjouw (2003) and Dang *et al*. (2019) for detailed discussion.

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. μ_{ij} is the error term. We will examine the extensions of Equation (1) in the following two situations, tracking poverty trends and measuring poverty dynamics.

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

Assume that the explanatory variables x_j are comparable for both surveys (Assumption 1), in the sense that they are defined and calculated in the same way, and that the changes in x_j between the two periods can reflect the change in poverty rate in the next period (Assumption 2).[5] Dang *et al.* (2017) define the imputed consumption y_2^1 as

[5] A typical example where the explanatory variables x_j are comparable for both surveys is that these two surveys represent two rounds of the same household consumption survey and there is no change in survey design across these

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

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 a large number of simulations, so as to minimize computation error. The poverty rate in period 2 and its variance can then be estimated as

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

Recent theoretical and empirical evidence suggest that this imputation method can improve on the prediction accuracy offered by the "proxy means testing" approach that is widely applied in practice (Dang *et al.*, 2017; Dang *et al.*, 2019). Specifically, the predicted household consumption generated using the regression-based imputation method described above 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 $\hat{\beta}' x_{ij}$. For consistency, the poverty line in the base survey—rather than that in the target survey—should be used in combination 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 applied to the variables in the more recent

survey rounds. Also note that x_j are likely more comparable when the survey rounds are implemented closer to each other (i.e., the time difference between the survey rounds is shorter).

survey to obtain poverty estimates. This is often referred to as the constant parameter assumption.[6] 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 in validation exercises (Christiaensen *et al.*, 2012; Dang *et al.*, 2021).

3.2. Measuring poverty dynamics

Poverty dynamics involves the study of household-level poverty outcomes at multiple time periods. Let z_j be the poverty line in period j. We are interested in knowing the unconditional measures of poverty mobility such as

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

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

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

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 directly estimate the quantities in (7) and (8); but in the absence of such data, we can employ imputation to construct "synthetic panels" to study mobility. We return to the notations with Equation (1). Let x_{ij} be a vector of *time-invariant* household characteristics that are observed in two cross-section survey rounds fielded in the same country at

[6] 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 observed 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).

two time periods. 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.

To operationalize the synthetic panel framework, two standard assumptions are typically made. First, the underlying populations being sampled in survey rounds 1 and 2 are assumed to be identical, such that their characteristics remain time-invariant (Assumption 3). More specifically, coupled with Equation (1), this implies the conditional distribution of consumption/income in a given period is identical whether it is conditional on the given household characteristics in period 1 or period 2 (i.e., $\mathrm{x_{i1}} \equiv \mathrm{x_{i2}}$ implies $\mathrm{y_{i1}}|\mathrm{x_{i1}}$ and $\mathrm{y_{i1}}|\mathrm{x_{i2}}$ have identical distributions). Second, μ_{i1} and μ_{i2} are assumed to 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

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

where $\Phi_2(.)$ stands for the bivariate normal cumulative distribution 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 survey round 2 (x_2) (the base year) for prediction, but we can use data from survey round 1 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 now the additional assumption that a good estimate of the correlation coefficient ρ is available (Assumptions 3 and 4 above), is critical. Dang *et al.* (2014) describe how in the absence of an independent estimate of ρ, bounds on poverty mobility can be estimated by postulating a range of values for ρ (in the extreme case one can utilize the theoretical bound values of 0 and 1). The resulting bounds on the mobility estimates will be narrower the better the prediction power of the underlying regression model – pointing to the importance of assembling as rich a set of time-invariant covariates as possible.

Dang and Lanjouw (2013) outline an approach to estimating ρ using cross section survey data from several different countries. While this is contingent on stronger assumptions that should be thoroughly examined before employment of this approach, validation exercises and applications of these synthetic panel methods by various researchers for different country contexts ranging from India to Africa, Asia, the Middle East, Latin America, and Europe have yielded encouraging results (Ferreira *et al.*, 2012; Cruces et al, 2015, Beegle *et al.*, 2016; UNDP, 2016; OECD, 2018; Dang *et al.*, 2019). Additional technical details, alternative imputation methods, and limitations, are discussed in various studies (Dang and Lanjouw, 2013; Herault and Jenkins, 2019; Bourguignon and Dang, 2019; Moreno *et al.*, 2021).

It is important to underscore that the assumptions and caveats involved in the construction of synthetic panels are onerous. As a result, caution is warranted and one would not recommend generating synthetic panel data where actual high-quality panel data already exist or may be collected in a cost-efficient and timely manner. In other contexts, where actual panel data do not exist or are too expensive to collect, there may be scope to implement the synthetic panel approach described here. Taken with the necessary precautions, such data can offer plausible estimates for poverty dynamics, information that is sorely needed, particularly in the developing country context.

4. Conclusion

We end this chapter by pointing to a few promising directions for future research. Regarding imputation topics, these methods can be useful in helping to track poverty for marginalized groups that are not typically covered well in traditional surveys. For example, as the number of refugees is increasing worldwide, estimating poverty for these disadvantaged populations is receiving growing attention. Recent studies suggest that a judicious combination of household consumption surveys, administrative data and imputation methods can provide plausible poverty estimates for refugees (Beltramo *et al.*, 2020; Dang and Verme, 2021).

Poverty imputation can be further improved following recent statistical advances. For example, recent evidence suggests that various machine learning techniques can complement or strengthen current procedures (Altindag *et al.*, 2021). Multiple imputation (MI) techniques, which are well developed in the statistical literature (Little and Rubin, 2020), offer an alternative method for imputing poverty and labor outcomes (Douidich *et al.*, 2016; Dang and Carletto, 2021).

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). Alternatively, in an effort to save time and financial resources, consumption data can be collected for only a subset of households in a full-scale survey, and then imputation methods applied to project consumption into the balance of sampled households

(Gibson, 2001; Ahmed *et al*., 2014). While promising, additional validation work is needed for further progress along these lines.

References

Ahmed, F., Dorji, C, Takamatsu, S. and Yoshida, N. (2014) Hybrid Survey to Improve Reliability of Poverty Statistics in a Cost-Effective Manner. *World Bank Policy Research Working Paper No 6909.*

ltındağ, O., O'Connell, S. D., Şaşmaz, A., Balcıoğlu, Z., Cadoni, P., Jerneck, M., & Foong, A. K. (2021). Targeting humanitarian aid using administrative data: model design and validation. *Journal of Development Economics, 148,* 102564.

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, World Bank: Washington, DC.

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.

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.

Cruces, G., Lanjouw, P., Lucchetti, L., Perova, E., Vakis, R., and Viollaz, M., (2015) Intra-Generational Mobility and Repeated Cross-Sections: A Three Country Validation Exercise *Journal of Economic Inequality*, 13:161-179.

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

---. (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.

---. (2021). *Data scarcity and poverty measurement* (IZA Discussion Paper No. 14631).

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., 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. H., & Serajuddin, U. (2020). Tracking the sustainable development goals: Emerging measurement challenges and further reflections. *World Development*, *127*, 104570.

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.*

Dang, H.-A. (2021). "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*, 39(6), 1008-1030.

Dang, H.-A. & Carletto, C. (2021). “Recall Bias Revisited: Measure Farm Labor Using Mixed-Mode Surveys and Multiple Imputation”. IZA Discussion Paper No. 14997.

Dang, H.-A. & Verme, P. (2021). “Estimating Poverty for Refugee Populations Can Cross-Survey Imputation Methods Substitute for Data Scarcity?” *ECINEQ Working Paper 2021-578*.

Dang, H. A. H., Kilic, T., Carletto, C., & Abanokova, K. (2021). “Poverty Imputation in Contexts without Consumption Data: A Revisit with Further Refinements”. *Policy Research Working Paper no. 9838*. World Bank: Washington, DC.

Deaton, A. and Muellbauer, J. (1980). *Economics and Consumer Behavior*. Cambridge University Press.

Deaton, A. and Dreze, J. (2002) “Poverty and Inequality in India: A Re-Examination”. *Economic and Political Weekly,* 37, 3729-3748.

Deaton, A. (2003) ‘Adjusted Indian Poverty Estimates for 1999-2000’ *Economic and Political Weekly*, January 25, 2003.

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*.

Gibson, J. (2001). Measuring chronic poverty without a panel. *Journal of Development Economics*, *65*, 243-266.

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*(1), 53-68.

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

Kijima, Y. and Lanjouw, P. (2003) Poverty in India during the 1990s: A Regional Perspective, *Policy Research Working Paper No. 3141*, DECRG, The World Bank.

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.

Little, R. J. A. & Rubin, D. B. (2020). *Statistical Analysis with Missing Data*. 3rd Edition. New Jersey: Wiley.

Moreno, H., Bourguignon, F., & Dang, H.-A. (2021). *On the construction of synthetic panels.* IZA Discussion Paper No. 14236.

OECD. (2015). *OECD employment outlook 2015*. Paris: OECD Publishing.

---. (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.

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.

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

World Bank. (2012). "Well Begun, Not Yet Done: Vietnam's Remarkable Progress on Poverty Reduction and the Emerging Challenges". *Vietnam Poverty Assessment Report 2012*. Hanoi: World Bank.

---. (2017). *Monitoring global poverty: report of the commission on global poverty*. Washington, DC: The World Bank.