NETWORK LISER

Discussion Paper Series

IZA DP No. 18952

September 2026

Poverty and Income Mobility in a Data-Scarce World: Concepts, Measures, and Data Challenges

Hai-Anh H. Dang
World Bank, IZA@LISER, GLO,
Indiana University, LSE and UEH

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18952 September 2026

Poverty and Income Mobility in a Data-Scarce World: Concepts, Measures, and Data Challenges*

Abstract

Poverty rates reveal how many people are poor, but not whether the same people remain poor. We review poverty and income mobility through a framework that links policy questions, time horizons, reference frames and data requirements, and that yields a data-tier-contingent reporting menu and a decision tree. An inventory of 218 economies identifies a qualifying household welfare panel in 88, with recent documented longitudinal fieldwork in 57. Three empirical illustrations then show the opportunities and the limits of the available methods. Published synthetic-panel validations yield small average deviations in selected joint transition shares, but do not establish a general success rate. Among 1,235 sampled subnational regions, about one in five crosses into a higher fixed global welfare class, which describes places rather than people. Adding noise to a Vietnamese panel changes measured entry and exit rates substantially while leaving the share poor in both years comparatively stable. Together the three illustrations distinguish what can be measured directly, what requires modeling assumptions, and what the available evidence cannot establish.

JEL classification

I32, D31, O10, O15, C81

Keywords

income mobility, poverty dynamics, synthetic panels, measurement error, subnational data, household surveys

Corresponding author

Hai-Anh H. Dang
hdang@worldbank.org

* We would like to thank Dean Jolliffe, Nayoung Lee, participants at the 39th IARIW general conference (Brussels), and various colleagues at the World Bank for helpful discussion. We would like to thank Minh Do for helpful support with data analysis. We are grateful to the UK Foreign Commonwealth and Development Office (FCDO) for funding assistance through the Data and Evidence for Tackling Extreme Poverty (DEEP) Research Program.

IZA DP No. 18952 September 2026

Poverty and Income Mobility in a Data-Scarce World: Concepts, Measures, and Data Challenges*

Abstract

Poverty rates reveal how many people are poor, but not whether the same people remain poor. We review poverty and income mobility through a framework that links policy questions, time horizons, reference frames and data requirements, and that yields a data-tier-contingent reporting menu and a decision tree. An inventory of 218 economies identifies a qualifying household welfare panel in 88, with recent documented longitudinal fieldwork in 57. Three empirical illustrations then show the opportunities and the limits of the available methods. Published synthetic-panel validations yield small average deviations in selected joint transition shares, but do not establish a general success rate. Among 1,235 sampled subnational regions, about one in five crosses into a higher fixed global welfare class, which describes places rather than people. Adding noise to a Vietnamese panel changes measured entry and exit rates substantially while leaving the share poor in both years comparatively stable. Together the three illustrations distinguish what can be measured directly, what requires modeling assumptions, and what the available evidence cannot establish.

JEL classification

I32, D31, O10, O15, C81

Keywords

income mobility, poverty dynamics, synthetic panels, measurement error, subnational data, household surveys

Corresponding author

Hai-Anh H. Dang
hdang@worldbank.org

* We would like to thank Dean Jolliffe, Nayoung Lee, participants at the 39th IARIW general conference (Brussels), and various colleagues at the World Bank for helpful discussion. We would like to thank Minh Do for helpful support with data analysis. We are grateful to the UK Foreign Commonwealth and Development Office (FCDO) for funding assistance through the Data and Evidence for Tackling Extreme Poverty (DEEP) Research Program.

1. Introduction

Most countries publish poverty headcounts, but far fewer report who escapes poverty and who falls into it, and the difference matters more than the headline suggests. The same decline in poverty is consistent with two very different societies: one in which the same households remain poor year after year while a few climb out, and another in which many households cross the line in both directions, so that the number who experience poverty over an extended period can far exceed the number poor at any one date. The two call for different policy responses.

Persistent poverty and frequent movement around the poverty line raise different questions. The first directs attention to enduring constraints, where the case is for long-term investment in human capital, infrastructure and productive assets; the second directs attention to exposure to shocks and incomplete protection, where insurance and social protection that responds to short-run shocks are the usual instruments (Baulch and Hoddinott, 2000; Carter and Barrett, 2006; Jalan and Ravallion, 2000). Mobility measures help to tell the two patterns apart; they do not by themselves identify their causes or establish which response will work. Directly measuring these transitions requires following the same households over time, and that is precisely what most countries cannot do: only about two-fifths of the world's 218 economies have ever fielded a qualifying household panel since 2000 (Section 3.1). What most of them do have is repeated cross-sections, surveys of different households at different dates, from which some transitions can be estimated under additional assumptions. This review examines what can be learned from such data, and when those assumptions are considered credible.

Understanding mobility bears on several of the central debates in development. Whether poverty traps exist is a question about dynamics rather than levels, about whether households that start poor can escape (Balboni *et al.*, 2022; Kraay and McKenzie, 2014). Equality of opportunity turns on how strongly children's prospects depend on their parents' circumstances (Corak, 2013; Narayan *et al.*, 2018). In a review of the effects of temperature change, Dang *et al.* (2024) find that the burden falls more heavily on chronic than on transient poverty. In each case, the evidence depends on credible information about change over time or links across generations, and such information is scarce: panels are often short, lose households to attrition and carry measurement error, a problem that does not spare even long-running surveys such as the Russia Longitudinal Monitoring Survey (Dang *et al.*, 2020), while the nonpanel routes reviewed below substitute assumptions for the missing links. Major policy reports have nonetheless moved beyond headcounts to mobility and vulnerability: Ferreira *et al.* (2013) on the Latin American middle class, OECD (2018) on social mobility, UNDP (2016) on multidimensional progress, and the World Development Report on the middle-income trap (World Bank, 2024).

Three developments make a review timely. The first is data. New global databases now permit comparisons that were impossible a decade ago. The Global Database on Intergenerational Mobility documents links between parents' and children's education for most of the world (Narayan *et al.*, 2018; van der Weide *et al.*, 2024); a complementary database assembles inequality-of-opportunity estimates from 196 surveys in 72 economies, with inter-generational mobility indicators announced as an extension (Ferreira *et al.*, 2026). The Subnational Poverty and Inequality Database (SPID) follows roughly 2,200 subnational units in up to 172 economies (Dang *et al.*, 2026), complementing the subnational human development indices of Smits and Permanyer (2019).[1] The second is method. Common theoretical principles now organize indices that had grown up separately (Cowell and Flachaire, 2025), and new measures distinguish upward from downward movement without requiring a panel at all (Genicot *et al.*, 2024). The third is a sobering

[1] The SPID and its cross-sectional companion, the Global Subnational Atlas of Poverty, are updated annually with the Global Monitoring Database and released through the World Bank's Development Data Hub.

body of evidence that measurement error can account for a considerable share of what is measured as mobility (Glewwe, 2012; Lee, 2022; Lee *et al.*, 2017; Millimet and Parmeter, 2025), which means that some of the movement the literature has documented may be noise.

These advances span literatures that have developed largely apart from one another, and we bring them together around a single practical question: which mobility questions can the available data answer? Our first contribution is organizational. The time frame distinguishes change within a person's lifetime, intra-generational mobility, from differences between parents and children, inter-generational mobility. The reference frame specifies whether status is assessed within a country or against a common international benchmark. In either frame, mobility can concern changes in welfare levels or changes in relative position, and our empirical global illustration follows regional means across fixed purchasing-power thresholds. We then organize estimation by the data each measure requires, so that Figure 1 presents four policy questions alongside four levels of data availability, and Section 6.2 translates the framework into a reporting menu and a decision tree that help analysts choose which measures to publish.

Our second contribution is the evidence base that makes the framework necessary. We compile an inventory of household panel availability under an explicit inclusion rule, covering all 218 economies and audited against official documentation. It shows that qualifying panels are less available in poorer countries and limited in continuous follow-up at every income level, and that access to a survey file is not the same as access to a linkable panel. The framework matters because of this scarcity: for most countries the question is not which panel to use but what can be credibly said without one.

Three empirical illustrations then address different links in the same chain: what can be estimated without household links, what changes when common international thresholds are applied to places, and how reporting error alters the resulting mobility statistics. The first assembles the validation record for synthetic panels, the method that estimates poverty transitions from repeated cross-sections and is widely used where comparable repeated cross-sections are available. In eleven tabulated settings from four publications, synthetic-panel joint transition shares differ from the corresponding reported-panel shares by 1.09 percentage points on average. This selected record is encouraging, but it neither establishes a general success rate nor shows that individual households are classified accurately. Section 5.1 distinguishes the estimators, benchmarks and uncertainty screens behind that summary.

The second illustration provides an early, descriptive assessment of how subnational regions move across fixed global welfare thresholds. Using the SPID, we follow 1,235 regions in 91 economies from the survey closest to 2010 to the latest available survey, a median span of 12 years. About one in five regions moved up a global welfare class, and more than half of those that started in the poorest class left it, while downward movement was rare. These are movements of places rather than of people, a distinction the paper maintains throughout.

The third illustration shows how selected mobility measures respond to specified additional measurement error, using simulations calibrated to the Viet Nam Household Living Standards Survey (VHLSS) panel. We add classical error, mean-zero noise independent of reported welfare, so that the reported panel accounts for four-fifths of the variance in the resulting series. The measured poverty exit rate then rises by about a fifth, while the entry rate nearly doubles (Table 4). Entry, the rarer transition, is far more sensitive in proportional terms. The exercise measures how added noise changes results relative to a reported panel that already contains error of its own, and it does not identify the true household transitions. Its lesson is not that mobility estimates are wrong but that their sensitivity to error differs by measure, and can be reported.

Our review complements existing surveys. Jenkins (2011) and the measurement-theory syntheses set out the conceptual apparatus; Cholli and Durlauf (2022) and Deutscher and Mazumder (2023) review inter-generational mobility, including the absolute and relative distinctions that recur below; Genicot *et al.* (2024) develop a particular measurement method. Each connects measures to data and to policy interpretation. What we add is the explicit integration of mobility concepts with data scarcity, particularly in poorer countries. We further document the availability and usability of household welfare panels, and use three empirical illustrations to show how validation evidence, survey comparability and measurement error shape what can responsibly be reported. Our focus is monetary welfare: we study households and individuals for mobility at the micro level, and subnational regions for the global analysis.[2]

Section 2 sets out the measures, concepts and data requirements. Section 3 reviews data availability in poorer countries, and Section 4 explains how measurement error shapes what mobility estimates can show. Section 5 presents the empirical analysis. Section 6 draws out the policy implications and concludes.

2. Measures, Concepts, and Data Requirements

Mobility is best treated as a family of estimands, quantities to be measured, rather than as a single statistic, because each measure answers a question shaped by three choices: the period over which status is followed, the reference against which movement is judged, and the data that are actually available. A report that mobility has risen in a country can therefore mean that households re-ranked faster, that more of them crossed a poverty line, or that children's outcomes

[2] Multidimensional and asset-based mobility raise parallel questions of concept and data; the framework of Section 2 applies with the welfare aggregate replaced accordingly.

depended less on their parents', and these are different claims with different data demands and different policy content. We first review the existing measures, then take the three choices in turn.

2.1 Existing measures

Mobility measures can be grouped into nine established families according to the question each answers and the data it requires (Appendix Table A1). Five of them describe movement within a distribution over time. The transition matrix is the workhorse: it records movement between classes, and the immobility ratio and the trace index of Shorrocks (1978a) compress it into a scalar, with inference methods developed by Formby *et al.* (2004). Projecting a matrix forward by applying it repeatedly demands two conditions, that current position captures all relevant history (the Markov assumption) and that transition probabilities are constant over time (time homogeneity); persistent differences between households can violate the first (Shorrocks, 1976). Movement indices measure how far income or rank changes (Fields and Ok, 1996).

Correlation-type indices measure persistence, so that a stronger association between initial and later status means less mobility. Equalization measures ask whether mobility reduces inequality in multi-period average incomes (Fields, 2010; Shorrocks, 1978b), and welfarist indices evaluate mobility through a social welfare function (Chakravarty *et al.*, 1985). Cowell and Flachaire (2018, 2025) place many of these measures in a common axiomatic framework, and Gottschalk and Spolaore (2002) draw a distinction that matters for policy, between reversals of fortune and independence of starting position.

A sixth family concerns poverty dynamics proper: entry and exit rates, spell duration, and the split between chronic and transient poverty (Duclos *et al.*, 2010; Jalan and Ravallion, 1998), with Hulme and Shepherd (2003) and McKay and Lawson (2003) reviewing the evidence from poorer countries. Vulnerability extends this family forward in time, to the risk of future poverty

(Gallardo, 2018), and related methods address misclassification around the poverty line (Porter and Yalonetzky, 2019) or date poverty escapes retrospectively where no panel exists (Krishna, 2004).

The seventh family is the panel-free measure of upward mobility of Ray and Genicot (2023). The eighth comprises anonymous and non-anonymous growth incidence curves (Grimm, 2007; Jenkins and Van Kerm, 2006, 2016; Ravallion and Chen, 2003). The ninth is inter-generational mobility, closely related to the normative literature on inequality of opportunity (Roemer and Trannoy, 2016). We also distinguish the reference frame in which these measures are applied, including common global welfare thresholds: the regional exercise of Section 5.2 applies established class-transition methods to thresholds that are the same for every unit and year, rather than introducing a further family of measures.

Four distinctions cut across these families, and keeping them apart dissolves much apparent disagreement about whether mobility is high or low. Absolute mobility concerns welfare levels and relative mobility position within a distribution, so that doubling every income produces absolute movement and no relative movement at all. Directional measures separate upward from downward movement; non-directional measures summarize movement regardless of sign (Fields and Ok, 1996; Genicot *et al.*, 2024). The difference matters because, for a fixed population, upward rank movement must be offset by downward rank movement elsewhere, whereas policy may value welfare gains among poor people even when others fall in rank without becoming worse off. Exchange mobility is reshuffling within a distribution; structural mobility results from changes in the distribution itself (Fields and Ok, 1999a). Finally, anonymous growth incidence curves compare quantiles at two dates without following the same people, while non-anonymous curves follow people grouped by their initial position (Bourguignon, 2011; Ravallion and Chen, 2003). Measures disagree, in short, because they answer different questions.

Measure choice should therefore follow the policy question and the available data, and several measures are usually needed, because their interpretation and their sensitivity to error differ. Movement and welfarist indices have well-developed axiomatic foundations. For the class-transition shares that are most commonly reported, the chosen cutoffs are central to interpretation. Section 5.2 illustrates the role of class boundaries with regional data. Section 4 examines how sensitivity to measurement error differs across families.

Measurement choices are especially consequential for inter-generational mobility, where cross-country rankings depend on both the measure and the data (Black and Devereux, 2011; Blanden, 2013). A single year of parental income is a noisy proxy for lifetime resources and can understate persistence; using multi-year averages roughly doubled many early estimates (Solon, 1992). Rank-rank slopes can be less sensitive to these choices and to lifecycle timing, the ages at which income is observed (Chetty *et al.*, 2014; Deutscher and Mazumder, 2023). Section 3.2 considers which of these measures are feasible with the data available in poorer countries.

Table 1 sets the main classification schemes beside our framework. Most of them make time explicit, and the schemes designed for data-scarce settings emphasize data requirements, but almost none separately identifies the reference frame, national or global. We combine time and reference frame and then organize estimation by data availability, while normative and axiomatic distinctions continue to cut across the resulting categories. Figure 1 and Table A2 show the framework, and the next three subsections explain its components.

2.2 Mobility over time

Intra-generational mobility follows the same household or individual over months or years, and covers income movement, rank changes, and poverty entry and exit (Fields and Ok, 1999a; Jäntti and Jenkins, 2015). Poverty dynamics is its directional, lower-tail branch, concerned with who crosses a poverty threshold, in which direction, and for how long (Bane and Ellwood, 1986; Jalan and Ravallion, 2000). Vulnerability turns the same question toward the future. One widely used definition, and the one emphasized here, is the probability of being poor later, conditional on current information (Dang and Lanjouw, 2017); the wider literature also defines vulnerability through the welfare loss associated with poverty and risk (Ligon and Schechter, 2003; López-Calva and Ortiz-Juarez, 2014).

Inter-generational mobility relates children's economic status to their parents'. Its measures include the intergenerational elasticity of income, rank-rank slopes, and absolute mobility, the share of children better off than their parents (Chetty *et al.*, 2014, 2017; Solon, 1992), and Cholli and Durlauf (2022) survey the mechanisms behind them. The distinction between generations depends on whose status is linked, not simply on elapsed time: a twenty-year within-person panel need not link parents and children, and widely separated cross-sections capture cohort replacement rather than lineage, approximating lineage mobility only under strong assumptions about household composition (Foster and Rothbaum, 2015).

Two further choices matter. One is the time horizon. Short horizons tend to reflect transitory variation and noise; longer horizons can reveal more durable change, at the price of greater attrition and weaker comparability. The spells approach studies the timing and duration of poverty episodes (Bane and Ellwood, 1986), whereas the components approach separates chronic poverty, associated with low average welfare, from transient poverty due to fluctuations around it (Jalan and Ravallion, 2000). The two can classify the same household differently, so the policy question rather than convenience should guide the choice. The other choice is the welfare aggregate. Because households smooth consumption, consumption mobility is generally lower than income mobility over short periods, and the two aggregates carry different measurement errors (Section 4). Comparisons that mix horizons or welfare aggregates therefore confound differences in concept with differences in measurement.

2.3 National and global reference frames

Two choices are often run together and are better kept apart. The first is the reference frame: whether status is judged within a country or against a benchmark common to the whole world. The second is whether the measure is relative, concerning position in a distribution, or absolute, concerning welfare levels against explicit thresholds. Both choices are available in either frame: thresholds may be national or global, and ranks may be national or global. A global frame is worth having because country of residence explains much of the worldwide variance in individual incomes (Milanovic, 2015), and because global growth patterns extend beyond any single national distribution (Lakner and Milanovic, 2016). Anand and Segal (2008) and Ravallion (2020) review how global distributions are constructed and how sensitive they are to purchasing-power conversions.

Subnational income accounts cover 1,528 regions in 83 countries (Gennaioli *et al.*, 2014), but they do not provide within-region distributions from harmonized household surveys. Absolute global-class mobility asks whether a unit crosses welfare thresholds held constant in purchasing-power terms across units and years, building on absolute welfare classifications in common purchasing-power units (Ravallion, 2010). Positional global mobility asks instead whether the unit moves in the world ranking. This paper estimates the first type of mobility, which is one illustration of the global frame rather than the whole of it. The two can move in opposite directions, since welfare can rise while rank falls if others improve faster.

A simple example shows what the frames and the measures each contribute. Consider a region whose mean welfare rises from $6 to $9 a day in comparable purchasing-power terms over a decade, while its national distribution rises proportionately. The region's national rank is unchanged despite a 50% rise in mean welfare, and in the global absolute frame it has crossed the common $8.30 threshold into the third global welfare class. What the common threshold adds is comparability. The same cutoff provides a common purchasing-power benchmark, so regions in different countries can be placed on one scale, subject to the comparability of the welfare concepts and the price adjustments behind it. A fixed national threshold in real terms would register the same improvement, but only against its own country's yardstick.

Figure 1 combines the two axes and shows what each data situation can reveal, directly or under the modeling assumptions set out in Appendix B. The upper-left cell, intra-generational mobility in the national frame, is where most research is concentrated. The upper-right cell, movement across fixed global welfare thresholds, has received little attention for tracked subnational units, and the SPID now permits it at scale (Section 5.2). The lower-left cell is inter-generational mobility in the national frame. The lower-right, parent-child transitions across fixed global welfare classes, is the thinnest of the four. Once comparable parent and child observations are linked, the ordinary weighted transition shares of the upper-left cell apply; what is missing is comparable linked data across generations and countries rather than an estimator.

Let y denote consumption or income per capita. Status may be its level, its logarithm when positive, or its position in the relevant national or global distribution, and status at two dates or across two generations has a joint distribution. A mobility measure is a functional, a rule that summarizes that distribution. Some measures depend on the association between statuses; others, such as anonymous growth incidence curves, use only the marginal distributions at each date, which is exactly why they need no panel. Appendix Table A1 compares these concepts and their data requirements, and Appendix B provides the formal statements.

2.4 Four data tiers

Everything in Section 2.1 depends on the available data. Two marginal distributions, one for each date, reveal distributional or structural change but never show who moved; separating that change from exchange mobility requires linked observations of the same units, or a model that supplies the missing links. We classify data into four tiers, and Appendix Table A1 records which measures each tier supports.

Tier D1 is a single cross-section. On its own it cannot identify transitions over time, although recall modules, coresident parent-child pairs or cohort variation can support education-based inter-generational measures, and vulnerability estimates can be obtained under further assumptions. Tier D2 is repeated cross-sections, the modal data situation in poorer countries. These directly support anonymous growth incidence analysis and panel-free upward mobility measures. Pseudo-panels follow cohorts through them, and synthetic panels estimate household (or individual) transitions under assumptions that link welfare across dates. The central identification problem at this tier is combining samples with no shared units (Ridder and Moffitt, 2007). Matching subnational units across repeated cross-sections is a different matter: the regions become a panel even though the households do not, which is what makes Section 5.2 possible without household links.

Tier D3 is an actual panel survey, observing the same units over time; repeat measurement on a subsample lies between a pure cross-section and a panel (Gibson, 2001). Tier D4 is linked administrative data, widely used in rich-country inter-generational research and rarely available for research in low-income settings. The tiers are a shorthand for the information a design carries, not a substitute for naming the variables, links and observation horizon a measure actually needs: a long household survey with parent-child links can support a rank-rank slope without a register, and a register without the relevant population or lineage links cannot. Nor do the tiers order estimate quality, since a short panel with substantial attrition and measurement error need not outperform a well-designed D2 method, a point that recurs throughout the paper. The tiers guide the inventory in Section 3, the error analysis in Section 4 and the reporting rule in Section 6.2.

3. Data Landscape in Poorer Countries

3.1 Actual panels: scarce, short, and imperfect

Table 2 lists the panel designs most used in mobility research on poorer and transition economies, and behind it stands an inventory of our own. It covers all 218 economies in the World Bank classification and records whether each reaches tier D3. To qualify, a survey must follow the same households (or individuals) over at least two waves, be nationally or near-nationally representative, collect a household welfare aggregate at two or more interviews, and have been last fielded in 2000 or later. Appendix C explains the exclusions and coding decisions and shows how alternative rules move the counts and the income gradients.

Three features stand out, and each bears on what a country can measure. First, qualifying panels are uncommon at every income level. Our search identified a qualifying national or near-national household welfare panel in 88 of the 218 economies, fielded in 2000 or later. The share rises with income, from 23% of low-income economies to 32%, 42% and 49% of lower-middle-, upper-middle- and high-income economies (Figure 3, panel a), an increase of 7.2 percentage points per unit of log GNI per capita (robust standard error 2.4).[3] These figures describe a historical stock rather than current capacity: a documented longitudinal wave in 2023 or later is recorded for 57 of the 88 programs, and two further programs are coded ongoing despite uncertain recent status.[4] The counts describe qualifying survey programs, not all longitudinal or administrative data.

Current capacity is more unequal than the stock. For ongoing programs the income gradient steepens to 12.5 percentage points per unit of log GNI per capita (2.0), or about nine points per doubling of income. Among the 76 low- and lower-middle-income economies, only six have a wave that recent and seven are coded ongoing. Scarcity is not confined to poorer countries, since about half of high-income economies also lack a qualifying panel, but richer countries can often fall back on registers and poorer ones cannot. For the remaining 130 economies our search found no program meeting the stated rule, so nationally representative household mobility estimates there rest on tier D2 or D1 unless a specialized panel or an administrative source outside our rule can be brought to bear. Population coverage tempers the economy counts: economies with a qualifying design hold 80% of the population of the 218 classified economies, which is not the share of people represented in a panel (Appendix C).

Second, continuous follow-up is often short, in richer countries too. Counting survey rounds conceals the problem, because a round means different things in annual, quarterly and irregular programs. Alongside survey frequency, the inventory therefore records four measures of depth: released rounds R, maximum linkable interviews L, interviews carrying a welfare aggregate W, and elapsed years H for the longest cohort (Appendix C gives the definitions). Because R and L depend on survey frequency, neither is compared across frequencies; L counts linked

[3] GNI per capita (Atlas method) and population are 2019 World Development Indicators values; the income grouping is the World Bank FY2026 classification. Sixteen economies without a published GNI per capita drop out of the regressions but not the counts.
[4] Israel and Morocco qualify under Appendix C's exception protocol.

observations and H measures the time they span. Among economies that do have panels, depth varies widely, and a few unusually long programs pull the average relationship.

Among the 82 programs fielded annually or less often, L rises by 2.5 interviews per unit of log GNI per capita (robust standard error 0.6). Elapsed span does not follow: median H is 5, 6, 3 and 3 years across the four income groups and 3.75 overall (Figure 3, panel c), so median spans do not rise monotonically across income groups even though the number of linked interviews does. The mean relationship is positive, 1.7 years per unit of log GNI per capita (0.8). But it rests on a few unusually long programs and is flat in a median regression, which is why we report the medians. Appendix C gives the alternative specifications, the standardized coefficients and the median regressions behind both statements. Availability has the clearest income association of the three dimensions, and coefficients for different outcomes on different scales cannot establish a formal decomposition.

Among panel economies, 61% can follow a household for five years or less, and 27 of the 43 high-income economies are among them. H is the maximum span the design permits rather than a typical household trajectory, and Table A3 flags the affected rows.[5] Access adds a further constraint, and a subtler one. Availability, recent activity and usable access are three different things. Only 43 of the 88 programs release survey files under open or registration access. Of the remaining 45, forty are restricted, which can still mean usable for an approved researcher, and five are documented through metadata alone. Twenty-nine programs are discontinued or dormant. Even an accessible file may lack what is needed to link households across rounds: of those 43 programs, 20 document the release of a persistent household key, 19 are unverified, two have no recorded access status for the key, and two explicitly release none. The strongest statement the inventory

[5] In forty-three of the 88 economies H rests on a subsample, on one cohort among several, or on a design that changed during the span; Table A3 daggers those rows and records the cohort behind H.

supports is therefore that a positive historical availability gradient does not translate into accessible, linkable data; Appendix C reports the access tiers, the key field and population coverage.

Third, panel quality varies, so tier D3 is not a uniform standard. Long-running panels face attrition, compromises in tracking respondents, and changes in questionnaires, and interview design matters in ways that can masquerade as mobility. Changes in recall periods can look like movement, and when consumption is seasonal, interviews held in different months can record seasonal fluctuation as longer-term change.

Attrition is cumulative and can be selective. In three poorer-country panels, substantial attrition was only weakly related to baseline characteristics (Alderman *et al.*, 2001), and intensive tracking in the Indonesia Family Life Survey achieves recontact rates above 90% (Thomas *et al.*, 2001). The RLMS, by contrast, has substantial cumulative attrition and periodic sample refreshment, which makes reweighting and bounds important for welfare-dynamics analysis (Dang *et al.*, 2020). Movers are more likely than others to leave a panel, so estimates based only on those who remain can be biased, in a direction that depends on the setting. Standard tools include selection-on-observables tests and inverse-probability attrition weights (Fitzgerald *et al.*, 1998), which address observed characteristics but cannot rule out selection on unobserved ones (Ghanem *et al.*, 2026).

Rotating designs offer a practical compromise. A fraction of each round's sample, half in the Vietnamese case, is re-interviewed in the next round while new households replace the rest, so that shorter links limit cumulative attrition and refreshment preserves cross-sectional representativeness. The linked subsample also provides a benchmark for validating synthetic-panel methods built from the same cross-sections (Section 5.1). The trade-off is that rotating designs identify only short-run dynamics directly. A three-wave rotation supports a chronic-transient poverty split over a few years (Appendix Table A4), but following trajectories across a lifetime, or measuring long poverty spells, requires longer links that few poorer-country surveys attempt.

Finally, a panel need not provide the best welfare measurement. India's IHDS panel collects far less detailed consumption data than the flagship NSS cross-sections, about 47 items against more than 400 (Dang *et al.*, 2019). When the panel's welfare aggregate is coarser and noisier, a synthetic panel built on the better questionnaire deserves consideration as a robustness check, and its estimates should not be presumed inferior simply because the household links are modeled rather than observed.

3.2 Inter-generational estimation without registers

Without tier-D4 linked registers, poorer-country studies usually estimate inter-generational persistence from coresident parent-child pairs or from recalled parental education and occupation. Both sources are useful, and each carries a characteristic bias. Coresidence truncates the sample, because adult children still living with their parents are a selected group. This can substantially understate some regression-based persistence measures, although rank and correlation measures need not share the same bias, and corrections are available (Emran *et al.*, 2018; Emran and Shilpi, 2021). Recalled parental schooling is coarse but comparatively robust, and censuses and Demographic and Health Surveys make education-based analysis possible at scale (Alesina *et al.*, 2021; van der Weide *et al.*, 2024), which is why education-based measures are so often the starting point in poorer countries (Asher *et al.*, 2024).

For income persistence, two-sample and copula-based methods supply the missing links through assumptions about, respectively, the transportability of a prediction relationship and the dependence between the two generations (Chetty *et al.*, 2017; Deutscher and Mazumder, 2023).

Data availability thus strongly shapes what can be estimated (Iversen *et al.*, 2019). Iversen *et al.* (2021) survey the wider field.[6]

3.3 New data methods for panels

Tier D2 supports two approaches that are often confused, and the confusion matters because they answer different questions. Pseudo-panels use repeated cross-sections to follow birth cohorts rather than individuals (Deaton, 1985). Cohort averages can yield consistent earnings-dynamics parameters under specified conditions, including cells large enough to limit noise in the cohort means (Moffitt, 1993; Verbeek and Nijman, 1992), although identifying dynamic parameters also requires exclusion restrictions that repeated cross-sections alone do not supply (Verbeek and Vella, 2005). An early application estimates individual vulnerability to poverty from Korean cohorts (Bourguignon *et al.*, 2004). Averaging within cohorts reduces idiosyncratic measurement error without removing it: in Mexico, pseudo-panel estimates of earnings mobility are far below individual-level estimates inflated by noise (Antman and McKenzie, 2007). The limitation is that averaging also removes within-cohort mobility, which is often the very thing one wants to know.

Synthetic panels instead estimate household-level joint distributions from repeated cross-sections. They use characteristics that are fixed over the interval to impute round-1 welfare for the households observed in round 2, or vice versa, and then combine imputed and observed welfare to estimate transitions. Under the partial-identification assumptions set out below, this yields bounds on transition probabilities (Dang *et al.*, 2014); under bivariate normality, with the cross-round

[6] Appendix B.7 explains these two approaches and their assumptions. Two-sample methods predict missing parental income from characteristics observed both in an auxiliary sample and in the children's recall. Copula methods combine the two marginal distributions using an assumed dependence structure to specify the joint distribution. Both use tier-D1 or D2 inputs to estimate quantities that linked tier-D4 data could measure directly.

correlation approximated from cohort-aggregated data, it yields point estimates and extends to general transition matrices and vulnerability dynamics (Dang and Lanjouw, 2023). Moreno *et al.* (2021) discuss refinements and cautions concerning the correlation parameter, and a recent alternative replaces the parametric income model with shrinkage regression and predictive mean matching (Lucchetti *et al.*, 2025). Applications now span Latin America's middle class, chronic and transient poverty in Africa and India, and female headship in the Arab region (AlAzzawi *et al.*, 2025; Dang and Dabalen, 2019; Dang and Lanjouw, 2018; Ferreira *et al.*, 2013). A separate approach works from a single cross-section, combining monetary and multidimensional indicators to identify chronic poverty (Bolch *et al.*, 2023). Section 5.1 reviews how well the synthetic panel methods validate against actual panels.

Three identification and data requirements, plus one inference requirement, are central (Appendix B.1), and each is substantive rather than technical. First, the linking covariates must be effectively time-invariant over the interval; birth cohort, sex, completed education, ethnicity and place of origin are the usual candidates. Second, the independently drawn surveys must cover comparable target cohorts or populations with adequate common support, and each period's conditional model must apply to the target-round covariates, although the coefficients need not be equal across rounds. Third, repeated cross-sections do not identify the dependence between welfare in the two rounds. Bounds avoid selecting a single dependence parameter, but still require restrictions on how welfare is associated across rounds; point estimates add a specific conditional distribution and a calibrated residual correlation, and Appendix B.1 states the assumptions for each route. Finally, uncertainty estimates must account for both sampling and model-estimation error, since treating imputed welfare as observed data understates the standard errors, and they cannot substitute for sensitivity analysis over the dependence assumption itself, which no amount of resampling identifies. The residual correlation, in particular, should be reported routinely (Moreno *et al.*, 2021; AlAzzawi *et al.*, 2025).

Tier D2 methods stand or fall on comparable cross-sections. The World Bank's global poverty monitoring standardizes welfare aggregates, adjusts for spatial price differences and flags breaks in comparability, and this infrastructure serves mobility measurement as well, since it underlies the SPID (Dang *et al.*, 2026). Figure 2 therefore gives comparability breaks a branch of their own, on which survey-to-survey imputation is the usual tool.[7]

Beyond surveys, machine-learning methods use satellite imagery and mobile-phone metadata to fill gaps in geographically detailed welfare data (Blumenstock *et al.*, 2015; Burke *et al.*, 2021; Chi *et al.*, 2022; Jean *et al.*, 2016; Merfeld *et al.*, 2023). Their value for transitions, as opposed to levels, depends on prediction bias and on how errors correlate across time, not only on fit within a year. Welfare-level estimates carry material uncertainty, especially in rural areas (Engstrom *et al.*, 2022; Gibson *et al.*, 2020), but few studies validate errors jointly over time. The evidence has not established that such maps outperform conventional small-area estimators, which have themselves been validated against population-data benchmarks for poverty levels (Corral *et al.*, 2025). Those estimators require care of their own, since standard errors can be understated when spatial correlation in area effects is ignored (Elbers *et al.*, 2003; Tarozzi and Deaton, 2009). Given the existing evidence, these maps appear a complement to survey-based mobility estimates rather than a substitute.

[7] The GMD harmonizes welfare aggregates, applies spatial and temporal deflation and attaches comparability flags; it underlies the Poverty and Inequality Platform and, at the subnational level, the SPID (Dang *et al.*, 2026). See Dang *et al.* (2019) for a recent review on survey-to-survey imputation.

4. Measurement Error and Mobility

4.1 Why noise looks like movement

Reporting error can make households appear to move between welfare classes even when the aggregate distribution barely changes, because transition measures depend on the joint distribution of welfare across periods and not only on each period's marginal. Consider measured log welfare equal to true log welfare plus classical error drawn independently in each period. The errors push households across a poverty line in both directions and weaken the correlation between periods, and with positively persistent welfare, as in the settings reviewed here, the result is lower measured persistence and inflated movement summaries. How much a particular transition statistic is biased nevertheless depends on its base rate and on the error structure (Appendix B.2). The effects can be substantial: O’Neill *et al.* (2007) find them concentrated in the tails, and Naschold and Barrett (2011) show that observed short-run income changes in Kenya and Madagascar substantially overstate structural mobility. Sensitivity also differs across index families (Cowell and Schluter, 1998).

The reliability ratio, the share of measured variance that is signal, puts a scale on the problem. Studies comparing annual earnings levels with administrative records in high-income settings place reliability at roughly 0.70 to 0.90; estimates for income changes and for household consumption are lower and not directly comparable (Bound and Krueger, 1991; Bound *et al.*, 2001; Gibson *et al.*, 2015; Table A5). Under classical error independent across waves and equal reliability in both, the measured Pearson correlation of log welfare equals the true correlation multiplied by that ratio. The arithmetic is unforgiving: treating the VHLSS reported-panel correlation of about 0.80 as a benchmark, adding such error at reliability 0.80 would lower it to about 0.64 (Section 5.3). This illustrates sensitivity to additional noise and does not identify the error already in the panel.

The structure of the error matters as much as its size. In Table 4, a plausible mean-reverting specification distorts the exit rate more than classical error does, so the classical model is a useful benchmark but not a worst case, and a correction built on it can go wrong in either direction (Bound and Krueger, 1991; Gottschalk and Huynh, 2010; Lee, 2022). Misreporting can also change over time, so that comparisons across surveys or decades partly reflect changes in measurement (Millimet and Parmeter, 2025). For consumption, recall periods, diaries versus interviews, proxy respondents and seasonal timing all shape the error process. (Appendix B.6 and Table A5 review the measurement-error evidence).

4.2 How large is the bias, and what can be done about it?

Studies that model measurement error explicitly show why these distinctions matter. In Korea, Lee *et al.* (2017) use a consumption-dynamics model with separately identified time-varying error to simulate error-free transition matrices, and the measured one-year poverty-exit rate of roughly 45% falls to between 26% and 31%. In Viet Nam, instrumenting one welfare measure with another substantially reduces measured mobility (Glewwe, 2012). The effects can be large, then, but each estimate rests on a particular setting and identification strategy, and neither supplies a general correction (Appendix B.6).

Whether a correction is defensible depends on what the data and assumptions reveal about the error process. Latent-Markov, misclassification and frontier-based models use panel structure under explicit identifying restrictions (Breen and Moisio, 2004; Lee, 2022; Lee *et al.*, 2017; Millimet and Parmeter, 2025); alternatively, two error-ridden welfare measures may instrument one another if the exclusion restriction is defensible (Glewwe, 2012). (Appendix B.6 explains these approaches). Where point identification is not supported, bounds can still show what the evidence establishes. Misclassification in a binary outcome can be modeled and corrected (Meyer and

Mittag, 2017), and partial-identification methods characterize the mobility measures consistent with a stated error rate (Li *et al.*, 2023; Millimet *et al.*, 2020).

Survey design can supply the information these assessments need. A short second consumption module for a random subsample within each wave helps to estimate reliability, under assumptions about the errors in the repeated measures. Questionnaire design matters in its own right: randomly assigned consumption modules produced substantially different poverty rates in Tanzania (Beegle *et al.*, 2012). Collecting income and consumption together offers a potential instrumental-variable strategy (Glewwe, 2012), but its exclusion restriction requires justification, since the two aggregates measure different concepts and can diverge systematically in rural low-income settings (Carletto *et al.*, 2022). Stable recall periods and documented boundaries between successive interviews, or seam design, help distinguish economic change from measurement change, and public interviewer identifiers allow the interviewer component to be assessed.

For routine reporting, we recommend publishing the observed estimate alongside sensitivity scenarios calibrated to an estimated reliability ratio (row 10 of Table 5). Reliability provides a variance scale, not a complete error model, so each scenario must also specify the error distribution, the dependence across periods and the stability assumptions; Section 5.3 illustrates how these choices affect different statistics and can guide the choice of measure. Before differences across countries or over time are read as economic differences, survey design, harmonization and error structure should be checked, which extends the caution Millimet and Parmeter (2025) urge for inequality to mobility. Measurement quality is also why an actual panel does not automatically outperform a synthetic one (Sections 2.4 and 3.1).

5. Empirical Analysis

5.1 Quantitative synthesis of synthetic-panel validation studies

Synthetic-panel validation compares estimates constructed from repeated cross-sections with an actual-panel benchmark. Of twenty-one studies identified, eighteen contribute original validation to the coded record, and only eight report numerical point comparisons. Table A6 separates point estimates, bounds, figures-only evidence and comparisons with nonmatching benchmarks, and Appendix B.4 sets out the search and coding rules.[8]

We summarize two outcomes: the absolute deviation from the reported-panel estimate, and whether the synthetic estimate lies inside the panel study's uncertainty interval. We use the term "deviation" rather than "error" because the benchmark itself contains survey noise, and the interval-agreement screen is descriptive, neither a coverage probability nor a formal equality test, since wider intervals pass more easily. Both statistics concern joint transition shares, such as the share poor in both years, under each study's own statistical unit and weighting convention; deviations in conditional entry and exit rates can be much larger.

The source-published record is the primary summary: across eleven settings in four publications, weighting settings equally, the mean absolute deviation is 1.09 percentage points, and among the eight settings in three publications with usable benchmark intervals, 30 of 32 estimates pass the agreement screen (Figure 4). Weighting and individual publications influence both figures, and Table A6 and Appendix B.4 document the aggregation and the alternative specifications. A supplementary block adds four settings that we estimated from the panels of Lucchetti *et al.* (2025); it gives 1.36 points across fifteen settings and 40 of 48 passes.[9]

[8] Rongen *et al.* (2026) is included in the bounds block because its screen places the panel estimate inside a synthetic band.

[9] Published transition cells alone do not identify the household regressions and the dependence calibration a synthetic-panel estimator requires. That block is therefore a review-side calculation that cannot yet be reproduced; it is reported apart from the primary block and enters no headline figure.

The intervals differ in level and construction, so the tally is not a common 95% screen. We use source-reported agreement counts; reconstructing the intervals from rounded inputs instead gives 28 of 32 and 38 of 48 in the source-published and supplementary blocks.[10] Results from different estimators are reported separately, because pooling them would compare unlike with unlike. Lucchetti *et al.* (2025), for example, replace the parametric income model with LASSO and predictive mean matching; their six tabulated settings give a deviation of 3.00 points and 6 of 24 passes under our screen, which is stricter than their interval-overlap criterion.[11] Table A6 also reports the alternative classifications and estimators of Dang and Lanjouw (2017) and Lucchetti (2017).

The rich-country evidence shows the limits of transferability. Hérault and Jenkins (2019) find agreement in 63 of 124 cases even with cohort-correlation choices guided by the actual panel. Colgan (2023), using the correlation directly from the panel, obtains 46 of 48 passes under a square-root transformation of income but only 26 of 48 under a log transformation. These designs rely on information that is unavailable without a panel and test different components of the procedure.[12] What they support is reporting sensitivity to both the cohort definitions and the welfare model.

For bounds, the screen asks whether the synthetic interval contains the panel estimate. The primary scope is the ten publications whose contribution we can trace to an identified published source. These publications suggest that full-range bounds that allow correlations anywhere from

[10] Gafa *et al.* (2024) is excluded because its synthetic estimates and panel benchmark refer to different household samples. Omitting Dang and Lanjouw (2023) reduces the source and extended agreement counts to 10 of 12 and 20 of 28.

[11] The matched parametric estimates are this review's re-estimation, not a comparison published by Lucchetti *et al.* (2025). The common pairs are Argentina 2013–14, Chile 2015–16, Peru 2015–16 and Nicaragua 2001–05; Table A6 reports the comparison.

[12] We use the authors' joint-transition tallies. The separate conditional-rate digitization lacks a released worksheet and comparison code.

zero to one contain the panel estimate in 200 of 212 specification-cell comparisons. Adding the Lucchetti *et al.* (2025) row gives 215 of 228 across eleven publications (which we report as a memorandum and not as the headline).[13]

Rongen *et al.* (2026) instead draw correlations from ancillary panels, adjust them for survey spacing and translate them into residual correlations. Their band reflects variation across these external correlations, so it is neither a sampling confidence interval nor an identification bound. In Tanzania, full-range bounds contain all twelve joint transition estimates, against nine under the richer model's tighter band, using rounded values, while the mean national interval width falls from 6.3 points under the sparser model to 4.8 under the richer one. The reported bands are narrower, but their informativeness depends on the external correlations being suitable for the target setting.

Benchmark quality also matters. All of the misses in the Kyrgyz validation occur in settings where attrition left the panel poorer than the cross-section (Bierbaum and Gassmann, 2012).[14] Validation, in other words, assesses the benchmark as well as the estimator.

These results support cautious use rather than a general success rate. Cells and settings are dependent, several studies reuse the same panels, and numerical reporting selects the evidence that survives into the record. Randomly splitting a panel into cross-sections, moreover, does not test the differences in sampling frames, questionnaires and nonresponse that separate genuinely

[13] We have not identified the table and page in Lucchetti *et al.* (2025). In general, tracing a row to its source is not the same as transcribing it: an exact table and page are recorded for three of the eleven rows, and the others are entered from the published tallies. That rate is not comparable to the point-estimate agreement rate, since a broad interval can pass while offering little precision, and studies reporting more specifications receive more weight. Several settings appear under multiple specifications; these are descriptive counts of dependent comparisons, without a confidence interval. Among 28 matched comparisons from three studies, full-range bounds contain the benchmark in all 28, against 25 under narrower assumed correlation ranges. Tightening buys precision at the price of additional assumptions, and the record cannot separate the two effects.

[14] We retain these cases; excluding difficult benchmarks after observing performance would bias the synthesis. An exception is Perez (2015), which uses a different Mexican survey over nonmatching intervals and is excluded from the pooled record. Our interval-by-interval scoring also differs from Perez's union rule across two tables.

independent surveys. Deviations are small in several low- and middle-income applications using absolute poverty lines, which is encouraging, but the record still remains rather limited for confident extrapolation.

5.2 Movement of sampled regional means across global welfare classes

This exercise tracks sampled regional means across fixed global welfare thresholds, giving each region equal weight. It measures the movement of places, not of people and not of positions in the world ranking, and row 9 of Table 5 accordingly requires comparable welfare data and links between units across years. The SPID supplies harmonized survey-based regional means (Dang *et al.*, 2026). We pair each region's survey closest to 2010 within 2008-2012 with its latest survey in 2017-2023, which gives 1,235 regions in 91 economies observed over a median of 12 years, and we classify their mean welfare into four bands defined by $3.00, $8.30 and $28 per person per day in 2021 purchasing-power-parity dollars.[15]

One qualification belongs beside the sample definition. Harmonization aligns the welfare aggregate and the deflation; it does not certify that a change between two particular survey rounds is free of a break in method, coverage or questionnaire, so the headline exercise measures movement in reported regional means.[16]

Table 3 shows movement that is predominantly upward. Of the regions initially below $3, 52% left that class; overall, one region in five moved up and fewer than 3% moved down. Persistence rises from 48% in the lowest class to 69%, 81% and 99% in successively higher classes,

---

[15] Regions are unweighted. All values are 2021 purchasing-power-parity dollars per person per day. The $3.00 and $8.30 lines are the World Bank's international extreme-poverty and upper-middle-income poverty lines as of the June 2025 update (World Bank, 2025); $28 is the prosperity standard behind the prosperity gap reported in the World Development Indicators (World Bank, 2026).

[16] Restricting the data on the pairs whose two surveys carry the same welfare type and the same comparability flag results in a much smaller sample of 565 regions in 50 economies, where upward movement is lower at 13.6%, and only 17.0% of the initially poorest regions leave the lowest class (Table A7, panel E). But this subsample is heavily biased toward richer countries; 94% of high-income regions survive the filter against about a third of the low-income ones.

but these rates depend on where the class boundaries sit and on how far regions start from them. In Table A7's equal-log-width specification, bottom-class persistence rises from 48% to 70% while overall upward movement changes little; because that specification changes both boundaries and widths, moving the lowest cutoff from $3.00 to $4.03, it cannot isolate the effect of class width.

Figure 5 groups regions by their country's 2010 income classification, and upward movement ranges from 27% in low-income countries to 4% in high-income countries, where most regions already occupy the top class.[17] Tracking the same regions makes this a non-anonymous growth incidence relation (Bourguignon, 2011). Persistence need not mean stagnation. Wide absolute classes permit substantial growth without crossing a boundary, and the top class has no higher category to move into. Nor does bottom-class persistence identify a poverty trap: two dates and conventional bands on regional means cannot reveal the underlying dynamics and constraints that the term describes, and the fraction of regions leaving the lowest class neither establishes nor rules one out.

Table A9 connects these regional movements to national reclassification. Twenty-three sample economies entered a higher World Bank income group between 2010 and 2024, and averaging across countries, 34% of their regions moved up a global welfare class, compared with 17% in countries whose income group was unchanged. This is an association among country averages, not evidence about residents: it does not establish broad-based gains, and what it shows is subnational variation that national reclassification alone conceals.

[17] Appendix Figure A1 relates initial welfare to subsequent annualized log growth, 100 times the log ratio of regional means divided by elapsed years. With standard errors clustered by country (91 countries), the pooled slope is indistinguishable from zero, so we do not claim convergence. Country fixed effects instead compare regions within countries, where there is limited variation in initial welfare (Table A8, panel C). Appendix Table A8, panel C, states the assumptions each specification requires and Figure A1 reports the two slopes. The clustered interval is model-based, since the SPID is a compilation of surveys rather than a probability sample of the world's regions.

For lagging-region policy, the global frame supplies thresholds that are comparable across countries, which the usual national yardsticks are not, and that domestic redistribution does not mechanically move, a property a fixed national threshold in real terms would share. Regional means nevertheless cannot reveal how gains are distributed among the people living there.[18] Dang and Lanjouw (2026) analyze poverty mobility with the SPID data in more detail.

5.3 How sensitive are reported mobility estimates to added measurement error?

We illustrate Section 4 by adding error to the VHLSS 2010–2012 panel of 3,927 households. The benchmark already contains survey error, so the experiment measures sensitivity to additional noise rather than how much existing error explains Vietnamese mobility. The poverty line reproduces the weighted 2010 cross-sectional poverty rate of 17.8%; the panel baseline is 19.1%.[19] Over 400 replications we add error to log expenditure in both years, calibrating λ, the variance of reported log welfare divided by that of the injected series, to 0.95, 0.90, 0.80 or 0.70, and we compare classical error with two mean-reverting variants in the spirit of Lee (2022). Throughout, λ refers to reported welfare and not to a known error-free series. Appendix B.3 details the procedure, and Table A5 documents the limited evidence available for calibrating these scenarios.

[18] Tables A7–A8 qualify the findings. Alternative boundaries and country-equal weighting change overall upward movement by at most three percentage points. Movement also varies with observation length: 6.7% rise over nine years or less, against 25.8% over thirteen years or more, which is why Table A7 reports endpoint-based annualized equivalents. Unequal gaps therefore complicate comparisons. Within-country differences account for 44% of observed regional growth variance, compared with just 7% for initial welfare. Unknown survey and comparability error prevents interpreting either share as economic variation alone. Regions are unweighted and class boundaries are conventions; comparability is taken up with the sample definition in Section 5.2 and in Table A7, panel E. Table 3's parenthetical figures measure country-resampling dispersion, not design-valid sampling error (Appendix B.5).

[19] The panel component links 2010 and 2012 households; expenditure is real per capita household expenditure in January 2010 prices. Household expansion weights are used throughout except for the Pearson and Spearman correlations, which are unweighted. The panel sample is slightly poorer than the full cross-section, hence the 19.1% baseline rate against the full-sample household incidence of 17.8%.

Table 4 and Figure 6 show that, in this calibration, classical error raises the entry and exit rates and the Shorrocks index while lowering the rank correlation. At $\lambda = 0.80$, exits rise from 46.1% to 55.8% and entries from 4.9% to 9.3%, increases of 21% and 87% on unrounded values; the rank correlation falls by about a fifth, yet the share poor in both years stays within half a point of 10.3%. That one stable cell is deceptive, because the other transition shares and the households recorded as moving change substantially: only about 40% of injected-data exits and 28% of entries match the reported panel, and conversely only 58% of reported-panel exits and 50% of entries survive the injection. Holding the amount of error fixed, mean-reverting error changes the poor-poor share more than classical error does, which is why the structure of the error matters as well as its size.[20]

These patterns are specific to the calibration and are not general laws. Equation (B12) perturbs only the second round, whereas the experiment adds continuous error to both. Across the classical scenarios, exits exceed reported-panel exits by 8%-27% and entries by 19%-136%, which complements the error-removal exercise of Lee *et* al. (2017). The share poor in both years is comparatively stable in this calibration, even though the other transition shares and the households classified as moving change substantially, so the stability of one aggregate is not evidence of accurate household trajectories; ranks, for their part, resist common strictly increasing transformations within each wave but remain sensitive to noise. The practical lesson is to report sensitivity under several plausible error structures alongside the raw estimates, as Table 5 proposes; these are scenarios, not corrections, and not a universal ranking of measures.[21]

[20] Holding the amount of error at the classical level, an error share of 0.25, but arranging it to revert toward the mean moves the exit rate to 59.5% against 55.8% and the joint probability of poverty in both years to 8.4% against 10.0%, so blanket corrections calibrated to classical error can misfire.

[21] Table A10 reproduces the pattern in 2012–2014, with 45% household overlap: a partial replication. Across three poverty lines, exits rise about 20%, while proportional entry increases decline as the line rises. Jointly observed income and expenditure could support instrumental-variable validation under a credible exclusion restriction; they measure different constructs and may share reporting error.

6. Policy Discussion and Conclusion

6.1 What the framework delivers

This paper organizes the mobility literature around a practical question: what can countries credibly learn with the data they have? The inventory documents qualifying panels that are less available in poorer countries and limited in continuous follow-up across all income groups, and Table 5 and Figure 2 help analysts choose what to report and make the assumptions explicit. Three empirical illustrations show the possibilities and the limits. In the studies that report comparable point estimates, synthetic-panel joint transition shares are close to reported-panel shares on average, over a sample defined by which studies report numbers at all.

In the SPID, about one in five sampled regions moved up a global welfare class, and more than half of those starting in the poorest class left it, which describes places rather than people. The VHLSS experiment shows that a stable share poor in both years can conceal quite different household transitions once noise is added. The practical lesson is to choose measures that answer the policy question and to show how the conclusions depend on data quality and assumptions.

6.2 What a statistical office could publish, and how to choose

Table 5 offers an illustrative reporting menu. Selection from the fuller catalog in Appendix Table A1 should reflect data availability, sensitivity to measurement error, policy relevance and comparability across countries, and a statistical office should start with the question it needs to answer rather than publish every feasible measure. A single cross-section (D1) can support model-based vulnerability and education-based inter-generational measures. Repeated cross-sections (D2) add distributional comparisons and, under explicit assumptions, synthetic-panel transitions (Dang *et al.*, 2014; Dang and Lanjouw, 2023). Actual panels (D3) provide observed links for rank persistence, which a modeled joint distribution can also deliver at D2 under its assumptions, and three or more waves are our recommended minimum for the split between persistent and temporary poverty rather than a mathematical prerequisite for computing it; two observations cannot establish long spells or long-run chronicity. The tiers describe the information available, not a guarantee of estimate quality, so survey strengths and limitations also shape the reporting set.

For social protection, the evidence answers two different questions and both should be reported. Entry and exit population shares (row 4) describe the volume of turnover. But the conditional rates describe two different populations, since the exit rate conditions on initial poverty and the entry rate on initial nonpoverty, and future poverty risk (row 1) extends the same question forward. Movement across ranks (row 5) adds information about relative position. Modeled links require comparable populations, stable predictors and assumptions about dependence across periods. The validation record supports cautious aggregate reporting, and it does not establish accurate classification of individual households.

For assessing opportunity across generations, educational mobility (row 2) uses parent-child information aligned with the Global Database on Intergenerational Mobility, with selection assessed and adjusted where justified when only co-resident pairs are observed (Emran *et al.*, 2018). Two upward-mobility statistics are in circulation and they are different estimands: the expected schooling rank of children whose parents fall in the bottom half (Asher *et al.*, 2024), and the probability that such a child reaches the top schooling quartile (Narayan *et al.*, 2018). Report whichever is chosen with its own definition rather than treating the two as interchangeable. For growth at lower ranks, row 3 requires comparable surveys but no household links, and the quantile-growth comparison is the reported measure; the Ray-Genicot index is an alternative that additionally requires a stated curvature parameter.

To ask who benefits from growth, row 6 follows groups from their initial position and should be read alongside an anonymous growth incidence curve, which compares ranks without following the same people. Rank correlations (row 7) summarize persistence in relative position; they are unchanged by a common strictly increasing transformation within each wave, including a common positive price deflator, although household- or region-specific deflators can alter ranks, as can reporting error.

Persistent and temporary poverty (row 8) separates poverty that endures over the observed period from short episodes, which bears on the balance between short-term protection and longer-term interventions; duration alone identifies neither the causes of poverty nor the appropriate response, and two observations cannot establish long spells. Global welfare classes (row 9) provide common benchmarks: comparable welfare in purchasing-power units supports class shares, transitions additionally need linked units or an explicit model, and the SPID supplies linked regions (Section 5.2). A single country can apply the same fixed thresholds to its own units without a multicountry database, which is required for the cross-country comparison rather than for the measure. For regional policy, class crossings should be read together with welfare growth, since remaining within a broad class need not mean stagnation and a rising regional mean need not mean gains for every resident.

The inventory points to practical investments in survey usability. Releasing files is not enough if analysts cannot link observations across rounds, so statistical offices should provide documented linkage variables through appropriate access arrangements, together with comparable welfare aggregates and information on attrition. A rotating panel component can deliver short-run transitions and a local benchmark for synthetic panels, although longer poverty spells require longer follow-up, and rotation carries tracking, refreshment, questionnaire, legal, access and capacity costs that this paper does not estimate. A panel alone cannot identify the measurement-error calibration; repeated welfare measures within a wave, or external validation, can help under assumptions about their errors. Common definitions and documented survey changes would also aid comparison. Appendix Table A4 illustrates the reporting set for Viet Nam using the 2010, 2012 and 2014 VHLSS rounds and the SPID regional panel, and Appendix B.8 gives the specifications.

Figure 2 provides a transparent sequence for these choices. First define what moves and against which benchmark, using Figure 1 and Table A2; then assess whether the data support the intended measure and whether their defects call for diagnostics, sensitivity analysis or an identified correction. A census alone, for example, cannot reveal income transitions without additional information or assumptions. Row 10 makes sensitivity analysis part of reporting for rows 4 and 5: publish the observed or model-based estimate alongside scenarios that specify the size, form and cross-period dependence of reporting error. The VHLSS results show why entry and exit should be examined separately. These scenarios display dependence on assumptions; they neither recover true mobility nor provide identification bounds, and other measures require a suitable error model. A report should say whether its substantive conclusion survives the scenarios examined.

A first mobility release can be specified compactly. Name the unit and the population it represents; state the period and the welfare concept; report both joint shares and conditional rates wherever the data allow; say which links are observed and which are modeled; and state whether the substantive conclusion survives the sensitivity scenarios of row 10. Documentation of this kind does more for the usability of a mobility statistic than an additional index.

6.3 Research and data agenda

The paper's limitations show where further evidence would help most. Our global-mobility estimates track unweighted sampled regions, which can be improved with appropriate population weights. More refined analysis of household transitions would additionally require linked households or an explicit joint-distribution model. The synthesis covers eighteen studies with original validation, but only eight publish numerical point comparisons, and it describes selected settings rather than offering a pooled inferential meta-analysis. The error experiment adds stylized noise to an imperfect panel, and survey-specific estimates of the error process would make those scenarios more informative.

Extending synthetic panels to generational horizons requires validation suited to longer intervals and to the outcome studied. Alternative-data estimates need validation of their errors across time, since good prediction of welfare levels in one year is not enough for mobility. Work on fixed global welfare classes should clarify how boundaries and the units observed shape interpretation. Comparable evidence on parent-child transitions across fixed global welfare classes remains limited. The main research gap is assembling and validating comparable data across generations and countries. These advances require continued investment in survey links, repeated welfare measures, harmonization and usable access. Administrative registers (D4) are still the exception than the norms in poorer countries. But they can provide useful, detailed on mobility patterns, particularly if they can be linked up with household surveys.

Better data and the transparent use of existing methods must therefore advance together. For the next mobility report, the advice is short: select policy-relevant indicators, distinguish observed from modeled links, identify the population and period covered, and place the essential assumptions and sensitivity results beside the headline estimates. That would make mobility evidence more useful for policy while keeping its limitation in view.

References

Ahmed, N., Brzozowski, M., & Crossley, T. F. (2006). Measurement errors in recall food consumption data. IFS Working Paper W06/21. London: Institute for Fiscal Studies.

AlAzzawi, S., Dang, H.-A. H., Hlasny, V., Abanokova, K., & Behrman, J. (2025). Female headship and poverty in the Arab region: Analysis of trends and dynamics based on a new taxonomy. *Social Indicators Research*, 180, 1067–1202.

Alderman, H., Behrman, J. R., Kohler, H.-P., Maluccio, J. A., & Watkins, S. C. (2001). Attrition in longitudinal household survey data: Some tests for three developing-country samples. *Demographic Research*, 5, 79–124.

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2021). Intergenerational mobility in Africa. *Econometrica*, 89(1), 1–35.

Anand, S., & Segal, P. (2008). What do we know about global income inequality? *Journal of Economic Literature*, 46(1), 57–94.

Antman, F., & McKenzie, D. (2007). Earnings mobility and measurement error: A pseudo-panel approach. *Economic Development and Cultural Change*, 56(1), 125–161.

Arseneault, L., Bryan, B. T., Canning, T., Okundi, M., Stephens, A., Triantafillopoulou, E., Yu, D., Bolivar, M., Mansoori, P., Milligan, L., Evans, E., & Gatera, G. (2025). Exploring the world for longitudinal datasets: Data resources for transformative mental health research. International Journal of Epidemiology, 54(4), dyaf128.

Asher, S., Novosad, P., & Rafkin, C. (2024). Intergenerational mobility in India: New measures and estimates across time and social groups. *American Economic Journal: Applied Economics*, 16(2), 66–98.

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., & Heil, A. (2022). Why do people stay poor? *Quarterly Journal of Economics*, 137(2), 785–844.

Bane, M. J., & Ellwood, D. T. (1986). Slipping into and out of poverty: The dynamics of spells. *Journal of Human Resources*, 21(1), 1–23.

Baulch, B., & Hoddinott, J. (2000). Economic mobility and poverty dynamics in developing countries. *Journal of Development Studies*, 36(6), 1–24.

Beegle, K., De Weerdt, J., Friedman, J., & Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*, 98(1), 3–18.

Bierbaum, M., & Gassmann, F. (2012). Chronic and transitory poverty in the Kyrgyz Republic: What can synthetic panels tell us? (UNU-MERIT Working Paper 2012-064; MGSoG Working Paper 2012-035). Maastricht: Maastricht Graduate School of Governance.

Black, S. E., & Devereux, P. J. (2011). Recent developments in intergenerational mobility. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 4B, pp. 1487–1541). Amsterdam: Elsevier.

Blanden, J. (2013). Cross-country rankings in intergenerational mobility: A comparison of approaches from economics and sociology. *Journal of Economic Surveys*, 27(1), 38–73.

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264), 1073–1076.

Bolch, K., López-Calva, L. F., & Ortiz-Juárez, E. (2023). “When life gives you lemons”: Using cross-sectional surveys to identify chronic poverty in the absence of panel data. *Review of Income and Wealth*, 69(2), 474–503.

Bound, J., & Krueger, A. B. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics*, 9(1), 1–24.

Bound, J., Brown, C., Duncan, G. J., & Rodgers, W. L. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, 12(3), 345–368.
Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J. J. Heckman & E. Leamer (Eds.), *Handbook of Econometrics* (Vol. 5, pp. 3705–3843). Amsterdam: Elsevier.
Bourguignon, F. (2011). Non-anonymous growth incidence curves, income mobility and social welfare dominance. *Journal of Economic Inequality*, 9(4), 605–627.
Bourguignon, F., Goh, C.-C., & Kim, D. I. (2004). Estimating individual vulnerability to poverty with pseudo-panel data. World Bank Policy Research Working Paper 3375. Washington, DC: World Bank.
Breen, R., & Moisio, P. (2004). Poverty dynamics corrected for measurement error. *Journal of Economic Inequality*, 2(3), 171–191.
Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (2021). Using satellite imagery to understand and promote sustainable development. *Science*, 371(6535), eabe8628.
Carletto, G., Tiberti, M., & Zezza, A. (2022). Measure for measure: Comparing survey based estimates of income and consumption for rural households. *World Bank Research Observer*, 37(1), 1–38.
Carter, M. R., & Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: An asset-based approach. *Journal of Development Studies*, *42*(2), 178-199.
Chakravarty, S. R., Dutta, B., & Weymark, J. A. (1985). Ethical indices of income mobility. *Social Choice and Welfare*, 2(1), 1–21.
Chaudhuri, S., Jalan, J., & Suryahadi, A. (2002). Assessing household vulnerability to poverty from cross-sectional data: A methodology and estimates from Indonesia. Discussion Paper No. 0102-52. New York: Department of Economics, Columbia University.
Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R., & Narang, J. (2017). The fading American dream: Trends in absolute income mobility since 1940. *Science*, 356(6336), 398–406.
Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Quarterly Journal of Economics*, 129(4), 1553–1623.
Chi, G., Fang, H., Chatterjee, S., & Blumenstock, J. E. (2022). Microestimates of wealth for all low- and middle-income countries. *Proceedings of the National Academy of Sciences*, 119(3), e2113658119.
Cholli, N. A., & Durlauf, S. N. (2022). Intergenerational mobility. NBER Working Paper No. 29760. Cambridge, MA: National Bureau of Economic Research.
Colgan, B. (2023). EU-SILC and the potential for synthetic panel estimates. *Empirical Economics*, 64(3), 1247–1280.
Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, 27(3), 79–102.
Corral, P., Henderson, H., & Segovia, S. (2025). Poverty mapping in the age of machine learning. *Journal of Development Economics*, 172, 103377.
Cowell, F. A., & Flachaire, E. (2018). Measuring mobility. *Quantitative Economics*, 9(2), 865–901.
Cowell, F. A., & Flachaire, E. (2025). Mobility and mobility measures. Working paper. London School of Economics and Aix-Marseille University.

Cowell, F. A., & Schluter, C. (1998). Measuring income mobility with dirty data. CASEpaper 16. London: Centre for Analysis of Social Exclusion, London School of Economics.
Cruces, G., Lanjouw, P., Lucchetti, L., Perova, E., Vakis, R., & Viollaz, M. (2015). Estimating poverty transitions using repeated cross-sections: A three-country validation exercise. *Journal of Economic Inequality*, 13(2), 161–179.
Dang, H.-A. H., & Dabalen, A. L. (2019). Is poverty in Africa mostly chronic or transient? Evidence from synthetic panel data. *Journal of Development Studies*, 55(7), 1527–1547.
Dang, H.-A. H., & Lanjouw, P. (2013). Measuring poverty dynamics with synthetic panels based on cross-sections. Policy Research Working Paper 6504. World Bank, Washington, DC.
Dang, H.-A. H., & Lanjouw, P. F. (2017). Welfare dynamics measurement: Two definitions of a vulnerability line and their empirical application. *Review of Income and Wealth*, 63(4), 633–660.
Dang, H.-A. H., & Lanjouw, P. F. (2018). Poverty dynamics in India between 2004 and 2012: Insights from longitudinal analysis using synthetic panel data. *Economic Development and Cultural Change*, 67(1), 131–170.
Dang, H.-A. H., & Lanjouw, P. F. (2023). Measuring poverty dynamics with synthetic panels based on repeated cross sections. *Oxford Bulletin of Economics and Statistics*, 85(3), 599–622.
Dang, H.-A. H., & Lanjouw, P. F. (2026). Poverty Mobility in Developing Countries: Stylized Facts, Data Constraints, and Subnational Illustrations. Unpublished manuscript, World Bank, Washington, DC.
Dang, H.-A. H., Hallegatte, S., & Trinh, T.-A. (2024). Does global warming worsen poverty and inequality? An updated review. *Journal of Economic Surveys*, 38(5), 1873–1905.
Dang, H.-A. H., 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(3), 757–797.
Dang, H.-A. H., Lanjouw, P. F., 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. H., Lokshin, M. M., Abanokova, K., & Bussolo, M. (2020). Inequality and welfare dynamics in the Russian Federation during 1994–2015. *European Journal of Development Research*, 32(4), 812–846.
Dang, H.-A. H., Nguyen, M. C., Brunckhorst, B. J., & Yang, J. (2026). A new global subnational poverty and inequality database based on harmonized household surveys. IZA Discussion Paper No. 18672. Bonn: IZA Institute of Labor Economics.
Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30(1–2), 109–126.
DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177–188.
Deutscher, N., & Mazumder, B. (2023). Measuring intergenerational income mobility: A synthesis of approaches. *Journal of Economic Literature*, 61(3), 988–1036.
Duclos, J.-Y., Araar, A., & Giles, J. (2010). Chronic and transient poverty: Measurement and estimation, with evidence from China. *Journal of Development Economics*, 91(2), 266–277.
Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, 71(1), 355–364.

Emran, M. S., & Shilpi, F. (2021). Economic approach to intergenerational mobility: Measures, methods, and challenges in developing countries. In V. Iversen, A. Krishna, & K. Sen (Eds.), *Social Mobility in Developing Countries: Concepts, Methods, and Determinants* (pp. 197–220). Oxford: Oxford University Press.

Emran, M. S., Greene, W., & Shilpi, F. (2018). When measure matters: Coresidency, truncation bias, and intergenerational mobility in developing countries. *Journal of Human Resources*, 53(3), 589–607.

Engstrom, R., Hersh, J., & Newhouse, D. (2022). Poverty from space: Using high resolution satellite imagery for estimating economic well-being. *World Bank Economic Review*, 36(2), 382–412.

Ferreira, F. H. G., Messina, J., Rigolini, J., López-Calva, L.-F., Lugo, M. A., & Vakis, R. (2013). *Economic Mobility and the Rise of the Latin American Middle Class*. Washington, DC: World Bank.

Ferreira, F. H. G., Peragine, V., Brunori, P., Salas-Rojo, P., Moramarco, D., Barajas, L., Barbieri, T., Daza-Báez, N., Datt, G., de Sandi, V., Farella, F., Martinez, A., Jr., Nguyen, J., Park, A., Simeone, E., Sirugue, L., Torres López, P., & Zotti, G. (2026). Global estimates of opportunity and mobility: A database. III Working Paper No. 158. London: International Inequalities Institute, London School of Economics and Political Science.

Fields, G. S. (2010). Does income mobility equalize longer-term incomes? New measures of an old concept. *Journal of Economic Inequality*, 8(4), 409–427.

Fields, G. S., & Ok, E. A. (1996). The meaning and measurement of income mobility. *Journal of Economic Theory*, 71(2), 349–377.

Fields, G. S., & Ok, E. A. (1999a). The measurement of income mobility: An introduction to the literature. In J. Silber (Ed.), *Handbook of Income Inequality Measurement* (pp. 557–598). Boston: Kluwer Academic Publishers.

Fields, G. S., & Ok, E. A. (1999b). Measuring movement of incomes. *Economica*, 66(264), 455–471.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources*, 33(2), 251–299.

Formby, J. P., Smith, W. J., & Zheng, B. (2004). Mobility measurement, transition matrices and statistical inference. *Journal of Econometrics*, 120(1), 181–205.

Foster, J. E., & Rothbaum, J. (2015). Using synthetic panels to estimate intergenerational mobility. Working Paper No. 013/2015. Mexico City: Espinosa Yglesias Research Centre (CEEY).

Gafa, D. W., Hodey, L. S., & Senadza, B. (2024). Examining poverty dynamics in Ghana: Evidence from longitudinal and repeated cross-sectional data. *Journal of African Economies*, 33(Suppl. 1), i136–i157.

Gallardo, M. (2018). Identifying vulnerability to poverty: A critical survey. *Journal of Economic Surveys*, 32(4), 1074–1105.

Garbero, A. (2014). Estimating poverty dynamics using synthetic panels for IFAD-supported projects: A case study from Vietnam. *Journal of Development Effectiveness*, 6(4), 490–510.

Garcés-Urzainqui, D., Lanjouw, P., & Rongen, G. (2021). Constructing synthetic panels for the purpose of studying poverty dynamics: A primer. *Review of Development Economics*, 25(4), 1803–1815.

Genicot, G., Ray, D., & Concha-Arriagada, C. (2024). Upward mobility in developing countries. NBER Working Paper No. 32391. Cambridge, MA: National Bureau of Economic Research.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2014). Growth in regions. *Journal of Economic Growth*, 19(3), 259–309.

Ghanem, D., Hirshleifer, S., & Ortiz-Becerra, K. (2026). Testing attrition bias in field experiments. *Journal of Human Resources*, 61(2), 690–731.

Gibson, J. (2001). Measuring chronic poverty without a panel. *Journal of Development Economics*, 65(2), 243–266.

Gibson, J., & Kim, B. (2013). How reliable are household expenditures as a proxy for permanent income? Implications for the income–nutrition relationship. *Economics Letters*, 118(1), 23–25.

Gibson, J., Beegle, K., De Weerdt, J., & Friedman, J. (2015). What does variation in survey design reveal about the nature of measurement errors in household consumption? *Oxford Bulletin of Economics and Statistics*, 77(3), 466–474.

Gibson, J., Olivia, S., & Boe-Gibson, G. (2020). Night lights in economics: Sources and uses. *Journal of Economic Surveys*, 34(5), 955–980.

Glewwe, P. (2012). How much of observed economic mobility is measurement error? IV methods to reduce measurement error bias, with an application to Vietnam. *World Bank Economic Review*, 26(2), 236–264.

Gottschalk, P., & Huynh, M. (2010). Are earnings inequality and mobility overstated? The impact of nonclassical measurement error. *Review of Economics and Statistics*, 92(2), 302–315.

Gottschalk, P., & Spolaore, E. (2002). On the evaluation of economic mobility. *Review of Economic Studies*, 69(1), 191–208.

Grimm, M. (2007). Removing the anonymity axiom in assessing pro-poor growth. *Journal of Economic Inequality*, 5(2), 179–197.

Hérault, N., & Jenkins, S. P. (2019). How valid are synthetic panel estimates of poverty dynamics? *Journal of Economic Inequality*, 17(1), 51–76.

Hulme, D., & Shepherd, A. (2003). Conceptualizing chronic poverty. *World Development*, 31(3), 403–423.

Iversen, V., Krishna, A., & Sen, K. (2019). Beyond poverty escapes—Social mobility in developing countries: A review article. *World Bank Research Observer*, 34(2), 239–273.

Iversen, V., Krishna, A., & Sen, K. (Eds.) (2021). *Social Mobility in Developing Countries: Concepts, Methods, and Determinants*. Oxford: Oxford University Press.

Jalan, J., & Ravallion, M. (1998). Transient poverty in postreform rural China. *Journal of Comparative Economics*, 26(2), 338–357.

Jalan, J., & Ravallion, M. (2000). Is transient poverty different? Evidence for rural China. *Journal of Development Studies*, 36(6), 82–99.

Jäntti, M., & Jenkins, S. P. (2015). Income mobility. In A. B. Atkinson & F. Bourguignon (Eds.), *Handbook of Income Distribution* (Vol. 2A, pp. 807–935). Amsterdam: Elsevier.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790–794.

Jenkins, S. P. (2011). *Changing Fortunes: Income Mobility and Poverty Dynamics in Britain*. Oxford: Oxford University Press.

Jenkins, S. P., & Van Kerm, P. (2006). Trends in income inequality, pro-poor income growth, and income mobility. *Oxford Economic Papers*, 58(3), 531–548.

Jenkins, S. P., & Van Kerm, P. (2016). Assessing individual income growth. *Economica*, 83(332), 679–703.
Kraay, A., & McKenzie, D. (2014). Do poverty traps exist? Assessing the evidence. *Journal of Economic Perspectives*, 28(3), 127–148.
Krishna, A. (2004). Escaping poverty and becoming poor: Who gains, who loses, and why? *World Development*, 32(1), 121–136.
Lakner, C., & Milanovic, B. (2016). Global income distribution: From the fall of the Berlin Wall to the Great Recession. *World Bank Economic Review*, 30(2), 203–232.
Lee, N. (2022). Measurement error and its impact on estimates of income dynamics. *Empirical Economics*, 63(5), 2539–2550.
Lee, N., Ridder, G., & Strauss, J. (2017). Estimation of poverty transition matrices with noisy data. *Journal of Applied Econometrics*, 32(1), 37–55.
Li, H., Millimet, D. L., & Roychowdhury, P. (2023). Measuring economic mobility in India using noisy data: A partial identification approach. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 186(1), 84–109.
Ligon, E., & Schechter, L. (2003). Measuring vulnerability. *Economic Journal*, 113(486), C95–C102.
Limanli, Ö. (2021). Estimating poverty transitions in Turkey using repeated cross-sectional data. *Ege Academic Review*, 21(2), 81–91.
López-Calva, L. F., & Ortiz-Juarez, E. (2014). A vulnerability approach to the definition of the middle class. *Journal of Economic Inequality*, *12*(1), 23-47.
Lucchetti, L. (2017). Who escaped poverty and who was left behind? A non-parametric approach to explore welfare dynamics using cross-sections (Policy Research Working Paper 8220). Washington, DC: World Bank.
Lucchetti, L., Corral, P., Ham, A., & Garriga, S. (2025). An application of LASSO and multiple imputation techniques to income dynamics with cross-sectional data. *Review of Income and Wealth*, 71(1), e12693.
Luttmer, E. F. P. (2001). Measuring poverty dynamics and inequality in transition economies: Disentangling real events from noisy data. World Bank Policy Research Working Paper 2549. Washington, DC: World Bank.
Martinez, A., Jr., Western, M., Haynes, M., & Tomaszewski, W. (2013). Measuring income mobility using pseudo-panel data. *The Philippine Statistician*, 62(2), 71–99.
McKay, A., & Lawson, D. (2003). Assessing the extent and nature of chronic poverty in low income countries: Issues and evidence. *World Development*, 31(3), 425–439.
Mekasha, T. J., & Tarp, F. (2021). Understanding poverty dynamics in Ethiopia: Implications for the likely impact of COVID-19. *Review of Development Economics*, 25(4), 1838–1868.
Merfeld, J. D., Dang, H. A., & Newhouse, D. (2023). *Improving Estimates of Mean Welfare and Uncertainty in Developing Countries*. Policy Research Working Paper 10348. World Bank.
Meyer, B. D., & Mittag, N. (2017). Misclassification in binary choice models. *Journal of Econometrics*, 200(2), 295–311.
Milanovic, B. (2015). Global inequality of opportunity: How much of our income is determined by where we live? *Review of Economics and Statistics*, 97(2), 452–460.
Millimet, D. L., Li, H., & Roychowdhury, P. (2020). Partial identification of economic mobility: With an application to the United States. *Journal of Business & Economic Statistics*, 38(4), 732–753.

Millimet, D. L., & Parmeter, C. F. (2025). The impact of measurement error on trends in earnings inequality in the USA. *Empirical Economics*, 69(5), 2727–2753.
Moffitt, R. (1993). Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics*, 59(1–2), 99–123.
Moreno, H., Bourguignon, F., & Dang, H.-A. H. (2021). On synthetic income panels. IZA Discussion Paper No. 14236. Bonn: IZA Institute of Labor Economics.
Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Mahler, D. G., Ramasubbaiah, R. G. N., & Thewissen, S. (2018). *Fair Progress? Economic Mobility Across Generations Around the World*. Washington, DC: World Bank.
Naschold, F., & Barrett, C. B. (2011). Do short-term observed income changes overstate structural economic mobility? *Oxford Bulletin of Economics and Statistics*, 73(5), 705–717.
OECD (2018). *A Broken Social Elevator? How to Promote Social Mobility*. Paris: OECD Publishing.
O'Neill, D., Sweetman, O., & Van de gaer, D. (2007). The effects of measurement error and omitted variables when using transition matrices to measure intergenerational mobility. *Journal of Economic Inequality*, 5(2), 159–178.
Perez, V. (2015). Moving in and out of poverty in Mexico: What can we learn from pseudo-panel methods? ISER Working Paper Series No. 2015-16. Colchester: Institute for Social and Economic Research, University of Essex.
Porter, C., & Yalonetzky, G. (2019). Fuzzy chronic poverty: A proposed response to measurement error for intertemporal poverty measurement. *Review of Income and Wealth*, 65(1), 119–143.
Prais, S. J. (1955). Measuring social mobility. *Journal of the Royal Statistical Society, Series A*, 118(1), 56–66.
Ravallion, M. (2010). The developing world's bulging (but vulnerable) middle class. *World Development*, 38(4), 445–454.
Ravallion, M. (2020). On measuring global poverty. *Annual Review of Economics*, 12, 167–188.
Ravallion, M., & Chen, S. (2003). Measuring pro-poor growth. *Economics Letters*, 78(1), 93–99.
Ray, D., & Genicot, G. (2023). Measuring upward mobility. *American Economic Review*, 113(11), 3044–3089.
Ridder, G., & Moffitt, R. (2007). The econometrics of data combination. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics* (Vol. 6B, pp. 5469–5547). Amsterdam: Elsevier.
Roemer, J. E., & Trannoy, A. (2016). Equality of opportunity: Theory and measurement. *Journal of Economic Literature*, 54(4), 1288–1332.
Rongen, G., Elbers, C., & Lanjouw, P. (2026). A Bayesian-inspired approach to improving the precision of synthetic panel estimates of poverty dynamics. Tinbergen Institute Discussion Paper TI 2026-066/V. Amsterdam: Tinbergen Institute.
Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434), 473–489.
Salvucci, V., & Tarp, F. (2021). Estimating poverty transitions in Mozambique using synthetic panels: A validation exercise and an application to cross-sectional survey data. WIDER Working Paper 2021/26. Helsinki: UNU-WIDER.
Shorrocks, A. F. (1976). Income mobility and the Markov assumption. *Economic Journal*, 86(343), 566–578.
Shorrocks, A. F. (1978a). The measurement of mobility. *Econometrica*, 46(5), 1013–1024.

Shorrocks, A. F. (1978b). Income inequality and income mobility. *Journal of Economic Theory*, 19(2), 376–393.
Smits, J., & Permanyer, I. (2019). The subnational human development database. *Scientific Data*, 6, Article 190038.
Solon, G. (1992). Intergenerational income mobility in the United States. *American Economic Review*, 82(3), 393–408.
Stanley, T. D., & Doucouliagos, H. (2012). *Meta-Regression Analysis in Economics and Business*. London: Routledge.
Tarozzi, A., & Deaton, A. (2009). Using census and survey data to estimate poverty and inequality for small areas. *Review of Economics and Statistics*, 91(4), 773–792.
Thomas, D., Frankenberg, E., & Smith, J. P. (2001). Lost but not forgotten: Attrition and follow-up in the Indonesia Family Life Survey. *Journal of Human Resources*, 36(3), 556–592.
Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, 40(6), 604–634.
UNDP (2016). *Multidimensional Progress: Human Resilience beyond Income*. Caribbean Human Development Report. New York: United Nations Development Programme.
Van der Weide, R., Lakner, C., Mahler, D. G., Narayan, A., & Ramasubbaiah, R. (2024). Intergenerational mobility around the world: A new database. *Journal of Development Economics*, 166, 103167.
Verbeek, M., & Nijman, T. (1992). Can cohort data be treated as genuine panel data? *Empirical Economics*, 17(1), 9–23.
Verbeek, M., & Vella, F. (2005). Estimating dynamic models from repeated cross-sections. *Journal of Econometrics*, 127(1), 83–102.
World Bank (2024). *World Development Report 2024: The Middle-Income Trap*. Washington, DC: World Bank.
World Bank (2025). June 2025 update to global poverty lines. Washington, DC: World Bank.
World Bank (2026). Prosperity gap. World Development Indicators, indicator SI.SPR.PGAP. Washington, DC: World Bank. Accessed 18 September 2026.

Table 1. How earlier studies inform the framework of Figure 1

Study	Main distinction	Use in this review
1. Shorrocks (1978a, 1978b)	Changes in position versus reductions in longer-term inequality.	Separates movement from its welfare benefits.
2. Chakravarty *et al.* (1985)	Describing mobility versus judging whether it improves well-being.	Makes value judgments explicit.
3. Bane and Ellwood (1986); Jalan and Ravallion (2000)	Poverty spells versus persistent and temporary components of poverty.	Distinguishes how long poverty lasts from how much persists.
4. Fields and Ok (1999a)	Persistence, rank or income-share changes, size and direction of income changes, and longer-term inequality.	Organizes measures of mobility within a lifetime.
5. Baulch and Hoddinott (2000)	Persistent and temporary poverty under limited data.	Connects the policy question to the available data.
6. Ferreira *et al.* (2013)	Movement among poor, vulnerable and middle-class groups.	Connects class transitions to fixed welfare thresholds.
7. Jäntti and Jenkins (2015)	Persistence, rank changes, income growth and longer-term inequality.	Clarifies what different mobility measures reveal.
8. Dang *et al.* (2019)	Missing surveys, missing welfare measures and missing panel links.	Organizes methods by the information available.
9. Cholli and Durlauf (2022)	Parent-child relationships measured with regressions or transition tables.	Shows how the statistical model shapes the findings.
10. Deutscher and Mazumder (2023)	Mobility across generations depends on the outcome, age and measurement period.	Makes results easier to compare across studies.
11. Genicot *et al.* (2024)	Size versus direction of change; absolute versus relative measures; options without panels.	Extends measurement to settings without household links.
12. Cowell and Flachaire (2025)	Common principles for measures based on income levels, ranks and changes.	Clarifies the properties and value judgments of each index.

Note: The table summarizes selected contributions rather than reproducing their full classifications. This review brings them together around three choices: mobility within or across generations, status judged within a country or against common global welfare thresholds, and the data available. Absolute and relative measurement is a further choice available in either frame. Figure 1 sets out the framework, and Table A1 lists the individual measures.

Table 2. Selected panel surveys used in mobility research on poorer and transition economies

Survey (country)	Period and rounds	How participants are followed	Main strength or limitation
1. CFPS (China)	2010 onward; every two years	National family panel	Broad coverage across provinces.
2. ELMPS (Egypt)	1998-2023; 5 rounds	Labor market panel; new households added	No comprehensive household consumption or income measure.
3. GSPS (Ghana)	2009/10-2018/19; 3 released rounds; fourth fielded from 2022	National household panel	Supports national and regional analysis.
4. IFLS (Indonesia)	1993-2014/15; 5 rounds	Follows movers and newly formed households	Baseline covered 83% of the population.
5. IHDS (India)	2004/05 and 2011/12; third round fielded in 2022-24	National household panel	Short consumption questionnaire limits comparison with NSS surveys.
6. KLIPS (Korea)	1998 onward; annual	Labor and income panel	Urban baseline; coverage depends on the sample used.
7. LSMS-ISA (Ethiopia, Malawi, Niger, Nigeria, Tanzania, Uganda)	2008/09 onward; 2-7 rounds per country	Household panels; coverage varies	Sample replacement and respondent losses can limit continuous follow-up.
8. MxFLS (Mexico)	2002-2012; 3 rounds	Follows households and migrants abroad	Tracks migrants to the United States; long gaps between rounds.
9. NIDS (South Africa)	2008-2017; 5 rounds	National household panel	Respondent losses are greater among richer and urban households.
10. RLMS-HSE (Russia)	1994-2024; 29 released rounds	Household panel; new households added	Long follow-up; substantial loss of original households requires careful weighting.
11. VHLSS (Viet Nam)	2002 onward; full welfare rounds every two years	Rotating panel; about half re-interviewed	Short linked sequences; the 2010-2012 link is used in Section 5.3.
12. Young Lives (Ethiopia, India, Peru, Viet Nam)	2002 onward; 7 rounds overall	Two groups followed from childhood	Not nationally representative; the seventh round excludes Viet Nam.

Notes: The list is illustrative and includes designs outside Table A3's eligibility rule: the IFLS covered 13 of Indonesia's then 27 provinces, the ELMPS lacks the comprehensive welfare measure the rule requires and is excluded from Table A3. Young Lives follows selected child cohorts, its seventh round covering Ethiopia, India and Peru only. Table A3 instead lists Susenas for Indonesia and the Korea Welfare Panel Study for Korea. Released rounds need not equal the observations available for one household, and rounds fielded but not released are identified separately. Viet Nam's 2010-2012 link is used in Section 5.3; Table A4 and Appendix B.8 use 2010, 2012 and 2014. Appendix C states the inventory rules.

Sources: GSPS, Gafa *et al.* (2024); IFLS tracking, Thomas *et al.* (2001); IHDS consumption measurement, Dang *et al.* (2019); KLIPS measurement error, Lee *et al.* (2017); MxFLS validation, Moreno *et al.* (2021); RLMS sample loss and weighting, Dang *et al.* (2020). Coverage and round counts follow the producing agencies.

Table 3. Regional transitions across global welfare classes, circa 2010 to circa 2022 (SPID)

Initial class (circa 2010)	To: below $3	To: $3-8.30	To: $8.30-28	To: at least $28	Regions
Panel A. Number of regions					
Below $3.00	66	72	0	0	138
$3.00-8.30	24	243	88	0	355
$8.30-28	0	9	436	92	537
At least $28	0	0	2	203	205
Panel B. Row percentages					
Below $3.00	47.8	52.2	0.0	0.0	100
	(11.8)	(11.8)	(0.0)	(0.0)	
$3.00-8.30	6.8	68.5	24.8	0.0	100
	(2.3)	(4.7)	(5.1)	(0.0)	
$8.30-28	0.0	1.7	81.2	17.1	100
	(0.0)	(0.8)	(6.8)	(6.9)	
At least $28	0.0	0.0	1.0	99.0	100
	(0.0)	(0.0)	(1.2)	(1.2)	

Note: The sample is 1,235 subnational regions in 91 economies observed in both the 2008-2012 and 2017-2023 windows, median span 12 years, and regions are unweighted. Classes are absolute and fixed at $3.00, $8.30 and $28 per person per day in 2021 PPP, each class running from its lower cutoff inclusive to the next cutoff exclusive, so a regional mean exactly at a cutoff falls in the higher class, so the table measures movement across fixed global thresholds and not movement in the global ranking. Parentheses are resampling dispersion rather than design-based standard errors: the 91 countries are drawn with replacement 2,000 times, all of a drawn country's regions together. Binomial standard errors, 1.7 to 4.0 times smaller, are in Table A8, panel B; Appendix B.5 states what the quantity is and is not. Source: our calculations from the SPID.

Table 4. Simulated effects of measurement error on mobility estimates, Viet Nam 2010-2012

Target variance ratio λ and error design	Realized first-wave error-variance ratio	Poverty rate, 2010 (%)	Exit rate (%)	Entry rate (%)	Poor both years (%)	Shorrocks M	Rank corr.
1.00 (reported panel) †	0.00	19.1	46.1	4.9	10.3	0.641	0.780
		(0.6)	(1.9)	(0.4)	(0.5)	(0.011)	(0.008)
0.95	0.05	20.3	49.6	5.9	10.2	0.694	0.735
		(0.3)	(1.1)	(0.3)	(0.2)	(0.008)	(0.004)
0.90	0.11	21.0	52.1	7.0	10.1	0.729	0.692
		(0.4)	(1.3)	(0.4)	(0.3)	(0.009)	(0.005)
0.80	0.25	22.6	55.8	9.3	10.0	0.782	0.608
		(0.5)	(1.3)	(0.4)	(0.3)	(0.009)	(0.008)
0.70	0.43	24.4	58.5	11.7	10.1	0.824	0.527
		(0.6)	(1.4)	(0.5)	(0.4)	(0.009)	(0.010)
0.80, mean-reverting (d = 0.148)	0.55	23.0	64.7	12.1	8.1	0.864	0.435
		(0.6)	(1.3)	(0.6)	(0.3)	(0.009)	(0.012)
0.95 variance ratio, mean-reverting (d = 0.100)	0.25	20.6	59.5	8.5	8.4	0.796	0.584
		(0.5)	(1.4)	(0.5)	(0.3)	(0.009)	(0.008)

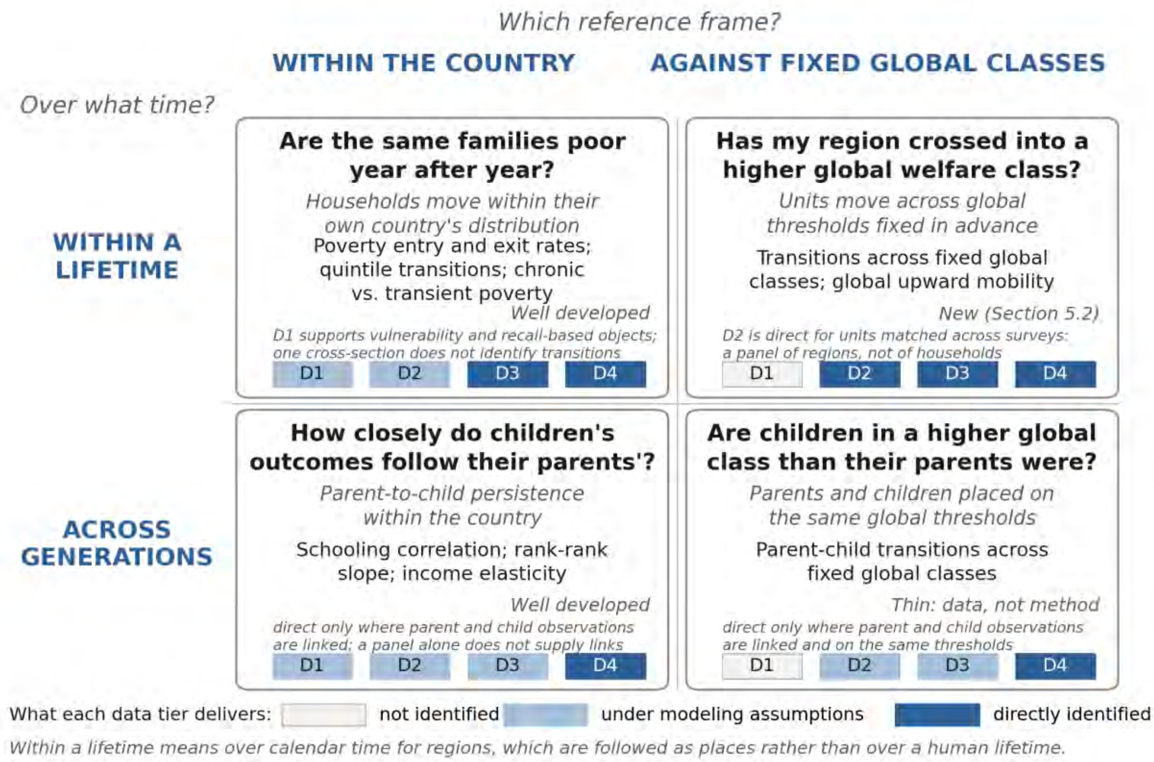
Note: The laboratory is the VHLSS 2010-2012 panel component, 3,927 linked households, at the released 2010 household weight renormalized to the panel sample. The link retains 41.8% of 2010 households and is poorer and more rural than the full sample, whose household poverty incidence is 17.8%, so every entry is a linked-sample estimate. † Parentheses in this row are household-bootstrap dispersion over 2,000 resamples, which is not a design-valid standard error; elsewhere they are Monte Carlo standard deviations across 400 replications, and the two are not comparable. λ is a reliability ratio in the classical rows and a variance ratio in the mean-reverting ones, which both set Corr(u, y) = -0.20 and differ in what they hold fixed. M is the Shorrocks trace index and Rank corr. the unweighted Spearman rank correlation. Appendix B.3 gives the error model, equations (B13) to (B16), and a diagnostic with correlated wave errors. Source: our calculations from the VHLSS 2010-2012 panel.

Table 5. A practical menu for reporting mobility: nine measure families and one reporting practice, ordered by minimum data requirement

Measure or practice	Policy question	Data and essential checks
1. Risk of future poverty	How many households face a high risk of poverty?	One survey (D1) gives the shares once a vulnerability line exists; estimating that line needs linked observations (D3) or a line borrowed from elsewhere. State the risk cutoff, the horizon it refers to, and where the line comes from.
2. Educational mobility across generations	How closely is children's schooling related to their parents' schooling?	Parent-child information in one survey or census (D1). Report the schooling correlation and one stated upward-mobility statistic. Check selection from observing only co-resident pairs. A household survey carrying the same links serves equally; a register without lineage links does not.
3. Growth at lower ranks	How does welfare change at lower ranks?	Two comparable surveys (D2). Report the quantile-growth comparison, which compares ranks across surveys rather than following the same households.
4. Poverty entry and exit	How many households enter, leave or remain in poverty?	Modeled links (D2) or an actual panel (D3). Report both population shares and rates conditional on initial poverty status.
5. Movement up and down the ranking	How much do people move up or down the welfare ranking?	Modeled links (D2) or an actual panel (D3). Report a quintile transition table and the Shorrocks trace index.
6. Growth by initial position	Which initially poor or better-off groups gain from growth?	Modeled links (D2) or an actual panel (D3). Compare with growth at each rank without tracking the same people.
7. Persistence in ranks	How closely are earlier and later welfare ranks related?	Two linked waves (D3) recommended; a modeled joint distribution (D2) supports it under its assumptions. Report the rank correlation; ranks remain sensitive to reporting error.
8. Persistent and temporary poverty	How much poverty persists over the observed period?	Three or more linked waves (D3) recommended; two observations cannot establish long spells. State how persistent poverty is defined and check attrition.
9. Global welfare classes	How many people or places cross common welfare thresholds?	Comparable welfare in common purchasing-power units. Shares need cross-sections; transitions also need linked units or a model. A single country can apply the same fixed thresholds to its own units; multicountry data are needed for the cross-country comparison, not for the measure.
10. Sensitivity to reporting error	How much could reporting error change rows 4 and 5?	Use the underlying data and estimator. Specify the error size, form and dependence across periods; report alternative scenarios.

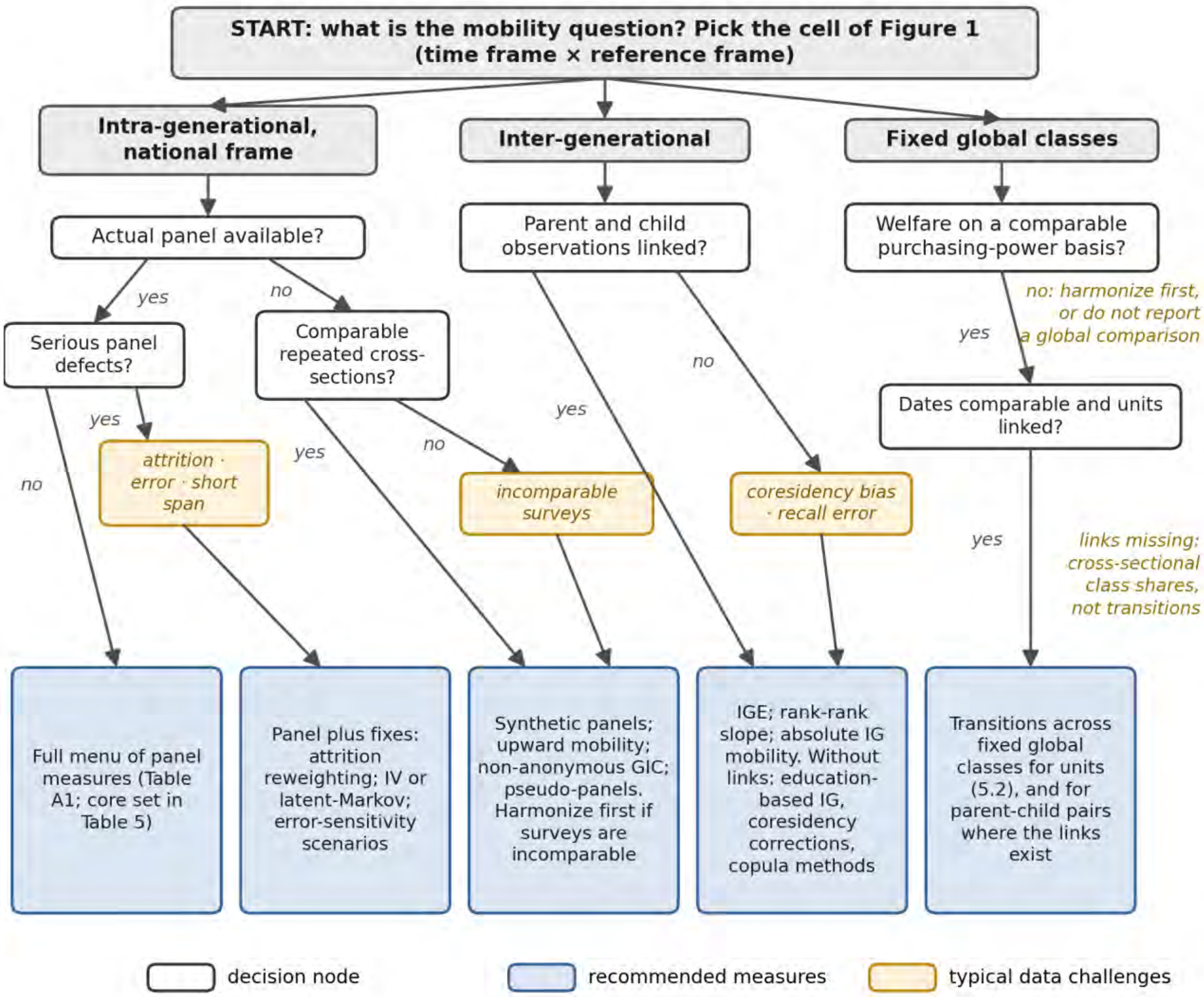
Note: This is an illustrative menu of nine measure families and one reporting practice; report selected measures whenever new comparable observations become available, choosing them as Section 6.2 sets out. D1 is one cross-section, D2 repeated cross-sections and D3 an actual panel, and a tier supports its measures only under the stated assumptions. In row 2, the two upward-mobility statistics in circulation are different estimands: the expected schooling rank of children whose parents fall in the bottom half (Asher *et al.*, 2024), and the probability that such a child reaches the top schooling quartile (Narayan *et al.*, 2018). State which one is reported, whose rank distribution defines the cutoffs, the parent and child cohorts, and how multiple parents are handled. In row 3, the reported measure is the quantile-growth comparison; the Ray-Genicot index is an alternative that also requires a stated curvature parameter. In rows 4 to 6, modeled links are synthetic panels: they need comparable populations, credible stable predictors and a defensible assumption about welfare persistence, and sensitivity to that assumption should be shown under at least two cohort definitions. Row 9 uses fixed PPP thresholds rather than world ranks, and the SPID unit is an area, not a household. Row 10 is a sensitivity exercise, not an error correction and not an identification bound. Table A1 gives definitions and references, and Table A4 illustrates the menu for Viet Nam.

Figure 1. A framework for deciding what mobility a country can measure



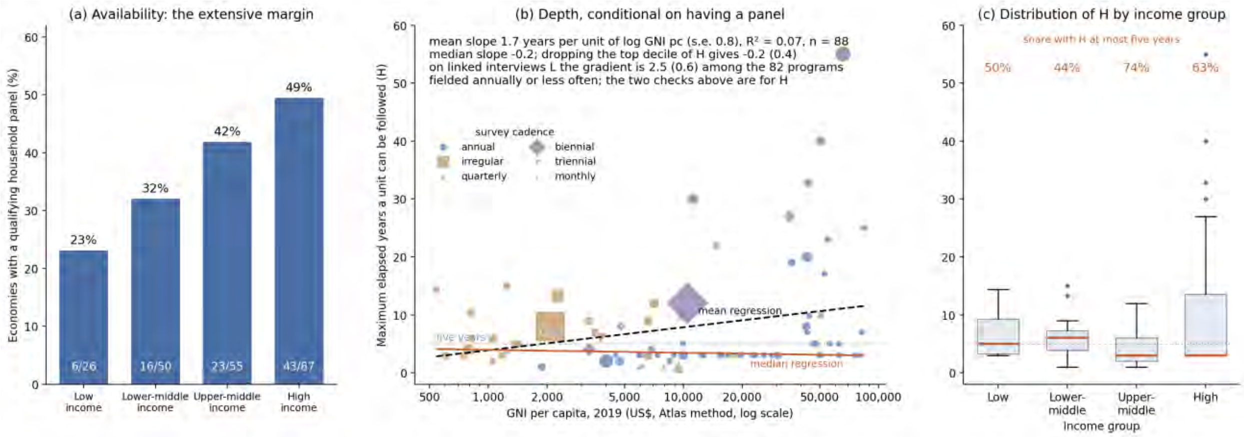
Note: D1 to D4 are the four data tiers of Section 2.4, and the chips record what each tier delivers at the cell's own unit. The right-hand column applies the same dollar cutoffs to every unit and year. Repeated cross-sections identify household transitions only under synthetic-panel assumptions, so D2 is model-based on the left; for an aggregate unit matched across surveys without tracking anyone, such as a subnational region, the same data identify the transition directly, so D2 is direct in the upper-right cell, where the object followed is a panel of regions rather than of households. A chip describes the cell's measures as a group, and the italic line inside each cell states the qualification that applies to it: in the two inter-generational cells, D3 and D4 are direct only where parent and child observations are linked, since a panel or a register without those links does not identify a parent-child transition; in the upper-left cell, a single cross-section supports vulnerability and recall-based objects but does not identify household transitions. Table A2 restates the taxonomy in question-first form, with references by cell.

Figure 2. A practitioner's decision tree: from question and data to measures and methods



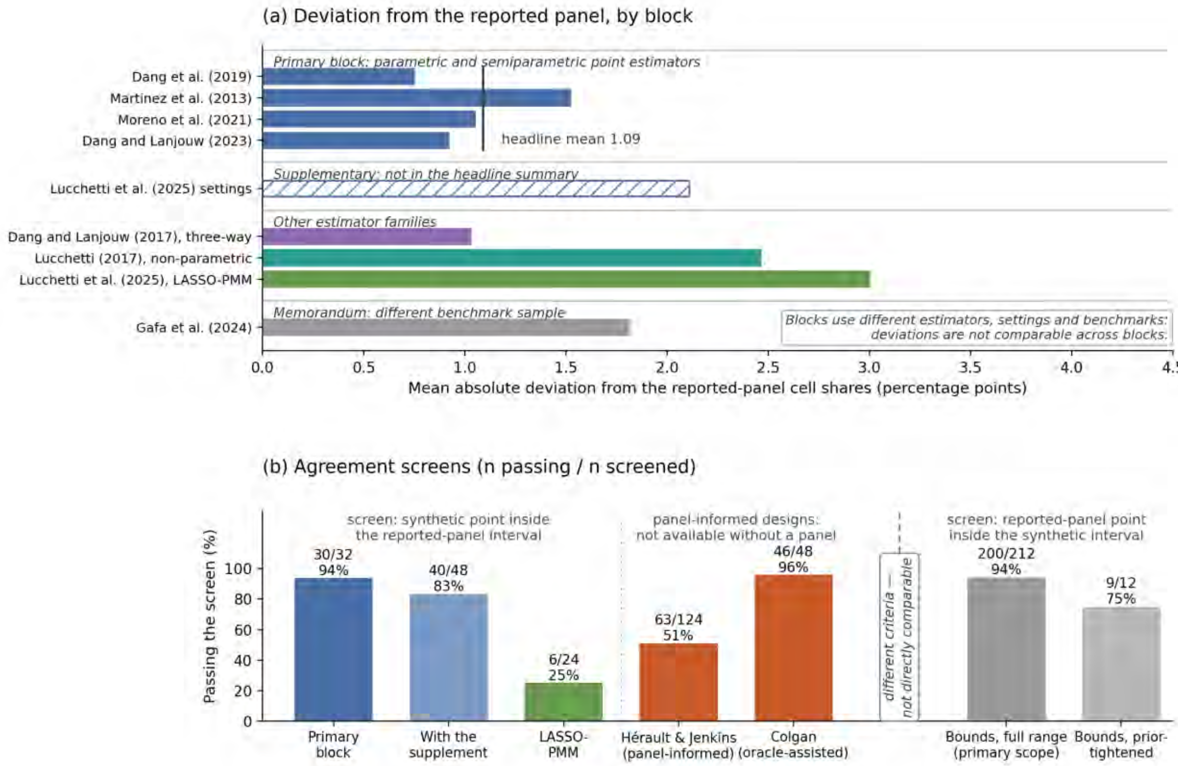
Note: Instrumental-variable and latent-Markov routes are assumption-dependent diagnostics or corrections rather than automatic panel fixes. The inter-generational cell of the fixed-global-class branch carries a linked-data route: standard transition estimators apply once parent and child observations are linked and placed on the same thresholds, but the data such an implementation needs are rarely assembled. The fixed-global-class branch asks two questions in sequence: first whether welfare can be expressed on a comparable purchasing-power basis, and only then whether the dates are comparable and the units linked. Failing the first calls for harmonization, or for not reporting a global comparison at all, since levels read against global lines need the same purchasing-power basis; failing linkage alone still permits cross-sectional class shares. The branch illustrates fixed global classes, and global ranks are a further option not shown there. Blue boxes carry pointers to Table A1, Table 5 and Sections 3 to 5. GIC = growth incidence curve; IG = inter-generational; IV = instrumental variables.

Figure 3. Household panel data are scarce everywhere, and shallow almost everywhere



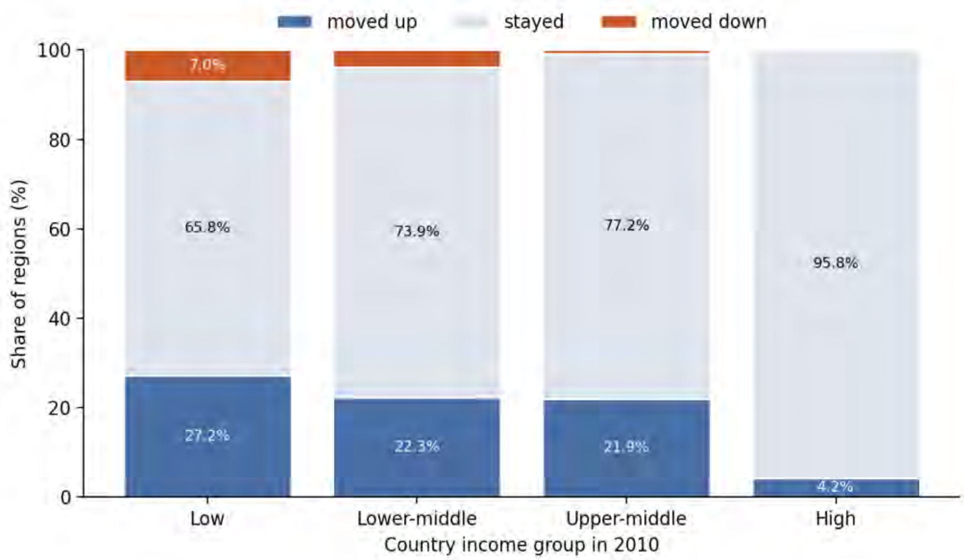
Note: The unit is an economy in the World Bank classification, all 218 of them, and the inclusion rule is that of Table A3, stated in Appendix C. Panel (a) counts economies with a qualifying survey fielded since 2000, so it is a stock rather than current capacity; Section 3.1 gives the running-component counts. Panels (b) and (c) cover the 88 that qualify, and bubble area is population. The released runner generates the slopes shown. The median and top-decile checks annotated in panel (b) are for elapsed span H; the separate gradient reported there for linked interviews L covers the programs fielded annually or less often. Both checks are author-supplied. Source: our compilation, September 2026.

Figure 4. Synthetic-panel validation: deviations from reported-panel transition shares and interval-agreement screens



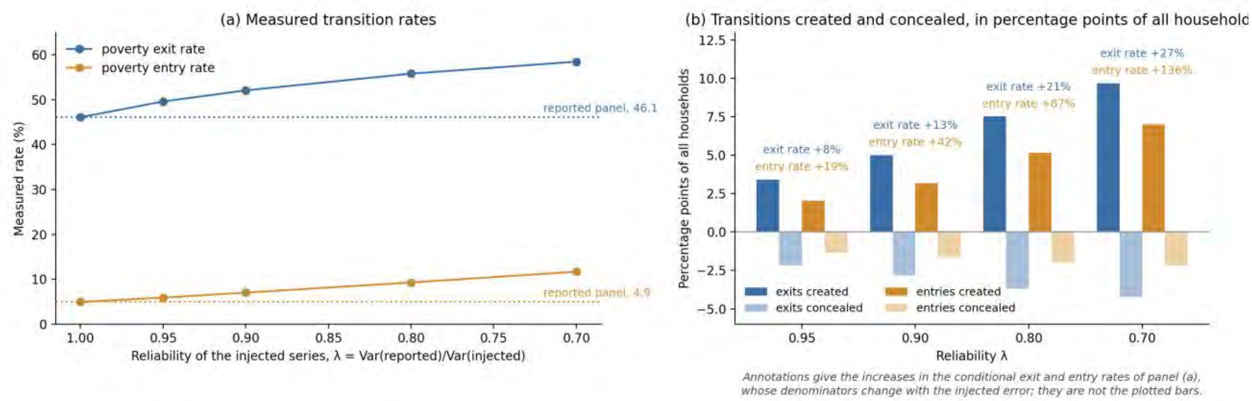
Note: Each bar in panel (a) is one study-estimator row of Table A6, the unweighted mean of its setting-level mean absolute deviations from the reported-panel cell shares. The supplementary block is our own estimation on the Lucchetti *et al.* (2025) panels; the unreleased inputs and code behind it are ours rather than the source's, and it is shown apart from the primary block and excluded from the headline summary; deviations are not comparable across blocks, which use different estimators, settings and benchmarks; Dang and Lanjouw's (2017) three-way classification provides cells that are not comparable per cell; Gafa *et al.*'s (2024) memorandum is excluded for a different benchmark sample. In panel (b) the denominators are transition cells, four per binary matrix, and not year-pairs; all tallies take the source-reported convention, and reconstructing the rounded cells gives 28 of 32 and 38 of 48. The bounds bar in panel (b) is the primary ten-publication tally; the eleven-publication memorandum that adds the unresolved row is in Table A6. The panel-informed designs are separated because they use panel information that is unavailable without a panel. LASSO-PMM is not one of them: it is estimated from cross-sectional inputs, and only its validation uses a panel benchmark. Appendix B.4 defines every statistic. Source: our calculations from the studies in Appendix Table A6.

Figure 5. Regional movement across global welfare classes, by the country's 2010 income group



Note: Sample and class definitions are those of Table 3; the unit is the sampled regional mean, not the household. Income groups are the World Bank classification for 2010, and a region moves when it crosses at least one class boundary. Source: our calculations from the SPID and the World Bank income-classification panel.

Figure 6. Sensitivity of measured mobility to injected measurement error: simulation evidence from the VHLSS panel



Note: Simulation design as in Table 4 and Appendix B.3; each point is the mean over 400 replications. Panel (a) plots measured poverty exit and entry rates as classical error is injected, indexed by λ; dotted lines mark the reported-panel rates. Panel (b) separates the gross flows the net change conceals: a created transition is one the measured data record and the reported panel does not, a concealed transition the reverse, both in percentage points of all households. The annotations are not read off the bars: they give the increase in the conditional exit and entry rates of panel (a) relative to the reported panel, computed on unrounded values, and those rates have denominators that change with the injected error, whereas the bars are percentage points of all households. None of these quantities is a statement about latent true transitions: they measure disagreement with the reported-panel classification, which is itself error-ridden. Source: our calculations from the VHLSS 2010-2012 panel.

Appendix A. Supplementary tables and figures

Table A1. A catalog of mobility measures: concepts and data requirements

Measure (key references)	Concept captured	Minimum data requirement	Notes for use in poorer countries
1. Quintile/decile transition matrix; immobility ratio (Prais, 1955) *	Positional (relative) movement between classes	D3, or D2 via synthetic panel	The workhorse; cutoff-sensitive; measurement error (ME) can distort transitions
2. Shorrocks trace index M (Shorrocks, 1978a) *	Scalar summary of transition-matrix mobility	D3 or D2 (synthetic)	Easy to report; inherits the matrix's ME sensitivity
3. Average rank/quantile jump, Bartholomew-type (see Fields and Ok, 1999a)	Distance of positional movement	D3 or D2 (synthetic)	Rank-based; unchanged by a common strictly increasing transformation within a wave, including a common deflator
4. Income correlation; rank correlation (Hart-type; Jäntti and Jenkins, 2015) *	Temporal dependence (immobility)	D3 recommended; D2 (synthetic) under the modeled joint distribution	Classical ME attenuates Pearson correlation; with positive persistence this overstates correlation-based mobility
5. Shorrocks rigidity index R (Shorrocks, 1978b)	Whether mobility equalizes multi-period incomes	D3 (3+ waves ideal)	Rarely computed in poorer countries for lack of long panels
6. Fields-Ok absolute income movement (Fields and Ok, 1996)	Total (non-directional) income movement	D3	Decomposes into growth and transfer components
7. Fields-Ok per capita log-income movement (Fields and Ok, 1999b)	Symmetric percentage movement; directional version available	D3	Sensitive to noise in small incomes
8. Ethical/welfarist mobility indices (Chakravarty *et al.*, 1985)	Welfare value of mobility relative to an immobile benchmark	D3	Axiomatically grounded; little application in poorer countries
9. Class-1/Class-2 status-movement measures (Cowell and Flachaire, 2025)	Axiomatic aggregation of individual status changes (income or rank)	D3 or D2 (synthetic)	Unifying framework; nests many familiar indices
10. Panel-free upward mobility measure (Ray and Genicot, 2023; Genicot *et al.*, 2024) *	Directional income growth of poorer ranks (pro-poor growth)	D2 (two cross-sections)	Designed for data-scarce settings; estimated for 122 countries
11. Growth incidence curve, anonymous (Ravallion and Chen, 2003) *	Quantile-by-quantile growth (no tracking)	D2	Benchmark against which non-anonymous mobility is contrasted
12. Non-anonymous GIC / mobility profiles (Bourguignon, 2011) *	Growth conditional on initial position	D3 or D2 (synthetic)	Reconciles growth and mobility perspectives

Measure (key references)	Concept captured	Minimum data requirement	Notes for use in poorer countries
13. Poverty entry/exit rates; spell durations (Bane and Ellwood, 1986) *	Poverty dynamics (directional, lower tail)	D2 for two-period entry and exit; D3 for spell durations	D2 via a synthetic panel identifies two-period entry and exit, not spell duration, which needs three or more linked waves; long spell analysis is rare in poorer countries
14. Chronic vs. transient poverty decomposition (Jalan and Ravallion, 2000) *	Persistence vs. variability of poverty	D3, three or more waves recommended	Policy triage: insurance vs. structural interventions; two observations cannot establish long-run chronicity
15. Joint/conditional poverty-transition probabilities from synthetic panels: bounds and point estimates (Dang *et al.*, 2014; Dang and Lanjouw, 2023) *	Poverty dynamics without panel data	D2 (repeated cross-sections)	Validated against actual panels (Table A6); accuracy is setting-specific and the residual correlation must be reported
16. Vulnerability to poverty; vulnerability lines (Ligon and Schechter, 2003; Dang and Lanjouw, 2017) *	Forward-looking mobility risk	D1 with assumptions; better with D3/D2	Complements backward-looking transition measures; a vulnerability line needs linked observations or a borrowed line, after which one cross-section gives the shares
17. Inter-generational income elasticity, IGE (Solon, 1992; Deutscher and Mazumder, 2023)	Relative persistence of income across generations	D4 or long D3 with parent-child links for direct identification; D1/D2 two-sample and copula routes are assumption-based	Lifecycle and attenuation biases; rarely feasible in poorer countries
18. Inter-generational rank-rank slope (Chetty *et al.*, 2014)	Positional persistence across generations	Linked parent-child observations for direct identification, from registers (D4) or from a survey carrying the links; D1/D2 copula routes are assumption-based	More robust to lifecycle/ME issues than the IGE
19. Absolute inter-generational mobility: share of children out-earning parents (Chetty *et al.*, 2017)	Absolute progress across generations	Linked parent-child observations, or marginal distributions plus a copula assumption	Copula methods relax the linked-data requirement; sensitivity to the copula depends on how far the two marginal distributions overlap

Measure (key references)	Concept captured	Minimum data requirement	Notes for use in poorer countries
20. Educational inter-generational mobility: schooling correlation; bottom-half-to-top-quartile upward mobility (Narayan *et al.*, 2018; van der Weide *et al.*, 2024; Alesina *et al.*, 2021; Asher *et al.*, 2024) *	Human-capital persistence and upward movement	D1 (coresidence or recall); censuses/DHS	The standard in poorer countries; coresidency/truncation bias (Emran *et al.*, 2018); with discrete schooling, assigning each level the mean rank of its band does not break ties, since everyone in a level keeps the same rank, and a level straddling the parental median or the child 75th percentile leaves groups that are not exactly one half or one quarter of the population, so the cutoffs are schooling categories rather than exact quantiles unless an explicit allocation rule is stated
21a. Absolute global-class mobility: movement of households or regions across globally fixed welfare thresholds (Dang *et al.*, 2026) *	Crossing dollar cutoffs that are the same for every unit and year	Regions: harmonized multi-country surveys with a subnational panel (SPID), which identify regional transitions directly. Households: population weights and within-region distributions give class shares; transitions additionally need linked households or a joint-distribution model	New frontier: Section 5.2 estimates it for 1,235 regions; sensitive to where the cutoffs are placed (Table A7, panel A)
21b. Global positional mobility: movement in the world ranking (Milanovic, 2015; Lakner and Milanovic, 2016)	Overtaking other units in the global distribution	Harmonized multi-country surveys with linked or modelled units	Distinct from 21a and can move in the opposite direction when the whole distribution shifts; harmonized ranks alone do not identify it, since a unit's movement through the world ranking requires linking or modeling the unit across rounds

Note: D1 = single cross-section; D2 = repeated cross-sections (pseudo/synthetic panels); D3 = actual panel; D4 = linked administrative registers. Measures marked * feed the core reporting set of Table 5. The minimum-data column names the data a measure needs to be identified; where a tier is given as a recommendation rather than a requirement, the row says so. Many D3 measures become feasible at D2 via synthetic panels (Section 3.3). Appendix B.1 states the synthetic-panel estimator formally.

Table A2. An alternative, question-first statement of the framework in Figure 1

Cell of Figure 1	The question a policymaker asks	What is measured	Data needed (feasible tier)	State of the evidence
Within a lifetime, within the country	Are the same families poor year after year, or do many escape while others fall in?	Poverty entry and exit rates; quintile transitions; chronic versus transient poverty	Repeated surveys (D2) via synthetic panels; an actual panel (D3) identifies them directly	Dense; the bulk of the literature
Within a lifetime, against the world	Has my region crossed into a higher global welfare class, or stayed where it was while others crossed?	Movement across globally fixed welfare classes (absolute, not positional); global upward mobility	Harmonized multi-country surveys (SPID, GMD)	Thin; the estimates in Section 5.2
Across generations, within the country	How strongly are children's outcomes related to their parents' circumstances?	Parent-child schooling correlation; rank-rank slope; income elasticity	One survey with recall or coresidence (D1); registers (D4) where they exist	Dense for education; thin for income in poorer countries
Across generations, against the world	Are today's children better off than their parents were, evaluated against the same fixed purchasing-power thresholds in both generations?	Parent-child transitions across fixed global welfare classes: the weighted count of linked parent-child pairs in a cell divided by the weighted count in the parental origin class, which is not the same as a child out-earning a parent without crossing a class boundary	Harmonized surveys plus recall or linked data; the estimator is the ordinary transition share once the links exist	Thin: comparable linked evidence is scarce, not the method

Note: The four rows are the four cells of Figure 1, restated so that each begins from a policy question rather than from a concept, with the data tiers of Section 2.4 applying within every cell. Key references by cell: (i) intra-generational, national frame: Fields and Ok (1999a); Jäntti and Jenkins (2015); Dang *et al.* (2014); (ii) intra-generational, fixed global classes: Dang *et al.* (2026) for the operational definition used here, with Milanovic (2015) and Lakner and Milanovic (2016) as background on global distributions rather than as validations of class transitions; (iii) inter-generational, national frame: Solon (1992); Chetty *et al.* (2014); Narayan *et al.* (2018); (iv) inter-generational, fixed global classes: van der Weide *et al.* (2024) as background on inter-generational mobility, principally in education, rather than as direct evidence on parent-child transitions across fixed dollar classes; a cell for which comparable linked evidence remains scarce rather than one for which no estimator exists. Readers who prefer a tabular statement of the framework may substitute this table for Figure 1.

Table A3. One qualifying national or near-national household-panel program per economy, all 218 economies, August 2026

Economy	Survey	Span	Cadence	R	L	W	H, yrs	Universe	Status	Rel.	Access	Official source
Panel A. Economies with a nationally or near-nationally representative household panel												
Albania *	Survey on Income and Living Conditions (SILC)	2017-2025	Annual	8	4	4	3.0	National	Ongoing	2024	C	instat.gov.al
Australia	Household, Income and Labour Dynamics in Australia (HILDA)	2001-2024	Annual	24 (25)	24	24	23.0 †	National	Ongoing	2024	C	melbourneinstitute.unimelb.edu.au
Austria	EU-SILC Austria	2004-2025	Annual	22	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Belgium	EU-SILC Belgium	2004-2025	Annual	22	6	6	5.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Benin *	Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM) panel	2018-2022	Irregular	2 s	2	2	3.0 †	National	Dormant	2022	B	microdata.worldbank.org
Bosnia and Herz.	Living Standards Measurement Study panel (LSMS/LiBiH)	2001-2004	Annual	4	4	2	3.0	National	Ended	2004	B	microdata.worldbank.org
Brazil	Pesquisa Nacional por Amostra de Domicilios Continua (PNAD Continua)	2012-2026	Quarterly	58 q	5	2	1.0 †	National	Ongoing	2026	A	ibge.gov.br
Bulgaria	EU-SILC Bulgaria	2006-2025	Annual	20	6	6	5.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Burkina Faso	Enquête Harmonisée sur les Conditions de Vie des Ménages, panel (EHCVM-P)	2018-2022	Irregular	2 s	2	2	3.0 †	National	Dormant	2022	B	microdata.worldbank.org
Canada	Survey of Labour and Income Dynamics (SLID)	1993-2011	Annual	19	6	6	5.0	Near-national	Ended	2011	C	www23.statcan.gc.ca
Chile *	Encuesta de Proteccion Social (EPS)	2002-2024	Irregular	8 s	8	8	22.0 †	Near-national	Ongoing	2024	B	previsionsocial.gob.cl
China	China Family Panel Studies (CFPS)	2010-2022	Biennial	7 (8) s	7	7	12.0 †	National	Ongoing	2022	B	isss.pku.edu.cn
Colombia	Encuesta Longitudinal Colombiana (ELCO)	2019-2022	Irregular	2 s	2	2	3.0	Near-national	Dormant	2022	A	microdatos.dane.gov.co
Costa Rica *	Encuesta Nacional de Hogares (ENAHO) rotating panel	2010-2025	Annual	16	4	4	3.0 †	National	Ongoing	2025	D	inec.cr
Croatia	EU-SILC Croatia	2010-2025	Annual	16	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Cyprus	EU-SILC Cyprus	2005-2025	Annual	21	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Czechia	EU-SILC Czechia (Zivotni podminky)	2005-2025	Annual	21	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Denmark	EU-SILC Denmark	2004-2025	Annual	22	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Dominican Rep.	Encuesta Nacional Continua de Fuerza de Trabajo (ENCFT)	2014-2026	Quarterly	47 q	5	5	1.0 †	National	Ongoing	2026	A	bancentral.gov.do
Ecuador	Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU)	2021-2026	Quarterly	21 q	4	4	1.3 †	National	Ongoing	2026	A	ecuadorencifras.gob.ec

Economy	Survey	Span	Cadence	R	L	W	H, yrs	Universe	Status	Rel.	Access	Official source
Estonia	Estonian Social Survey (EU-SILC)	2004-2025	Annual	22	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Ethiopia	Ethiopia Socioeconomic Survey (ESS/ESPS, LSMS-ISA)	2011-2022	Irregular	5 s	3	3	4.0 †	Near-national	Dormant	2022	B	microdata.worldbank.org
Finland	EU-SILC Finland (Income Distribution Statistics)	2004-2025	Annual	22	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
France	EU-SILC France (SRCV)	2004-2025	Annual	22	9	9	8.0 †	National	Ongoing	2025	C	insee.fr
Georgia	Welfare Monitoring Survey (WMS)	2009-2017	Biennial	5	5	5	8.0 †	Near-national	Dormant	2017	D	unicef.org
Germany	German Socio-Economic Panel (SOEP)	1984-2024	Annual	41	41	41	40.0 †	National	Ongoing	2024	C	diw.de
Ghana *	Ghana Socioeconomic Panel Survey (GSPS)	2009-2019	Irregular	3 (4) s	3	3	9.0	National	Dormant	2019	B	microdata.worldbank.org
Greece	EU-SILC Greece	2003-2025	Annual	23	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Hong Kong SAR *	Hong Kong Panel Study of Social Dynamics (HKPSSD)	2011-2021	Irregular	5 s	5	5	10.0 †	National	Dormant	2021	C	caeser.hkust.edu.hk
Hungary	EU-SILC Hungary	2005-2025	Annual	21	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Iceland	EU-SILC Iceland	2004-2024	Annual	17 (21)	4	4	3.0	National	Ongoing	2020	C	hagstofa.is
India	India Human Development Survey (IHDS)	2004-2012	Irregular	2 (3) s	2	2	8.0	National	Ongoing	2012	B	ihds.umd.edu
Indonesia *	Susenas panel (National Socio-Economic Survey panel component)	2003-2010	Annual	8	3	3	2.0	National	Ended	2010	B	catalog.ihsn.org
Iran, Islamic Rep. *	Household Expenditure and Income Survey (HEIS) rotating address panel	2010-2024	Annual	15	3	3	2.0 †	National	Ongoing	2024	A	amar.org.ir
Ireland	EU-SILC Ireland (CSO)	2003-2025	Annual	22	6	6	5.0	National	Ongoing	2024	C	ec.europa.eu/eurostat
Israel	Longitudinal Survey	2012-2023	Annual	9	9	9	9.8 †	National	Ongoing	2022	C	cbs.gov.il
Italy *	Survey on Household Income and Wealth (SHIW), Banca d'Italia	1987-2022	Biennial	17	14	14	27.0 †	National	Ongoing	2022	A	bancaditalia.it
Jamaica *	Jamaica Survey of Living Conditions (JSLC) rotating panel	1990-2010	Annual	21	4	4	3.0 †	National	Dormant	2010	C	statinja.gov.jm
Japan	Japan/Keio Household Panel Survey (JHPS/KHPS)	2004-2024	Annual	21	21	21	20.0 †	National	Ongoing	2024	C	pdrc.keio.ac.jp
Kazakhstan *	Household Budget Survey (HBS)	2001-2025	Annual	25	3	3	2.0 †	National	Ongoing	2025	D	stat.gov.kz
Kenya	Kenya Continuous Household Survey (KCHS) panel component	2019-2020	Annual	2	2	2	1.0 †	National	Ended	2020	B	statistics.knbs.or.ke
Korea, Rep.	Korea Welfare Panel Study (KOWEPS)	2006-2025	Annual	20	20	20	19.0 †	National	Ongoing	2025	B	koweps.re.kr

Economy	Survey	Span	Cadence	R	L	W	H, yrs	Universe	Status	Rel.	Access	Official source
Kyrgyz Rep.	Listening to the Kyrgyz Republic (L2KGZ) monthly household panel	2021-2025	Monthly	42 (43)	42	42	3.5	National	Ongoing	2025	B	microdata.worldbank.org
Latvia	EU-SILC Latvia	2005-2025	Annual	21	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Lithuania	EU-SILC Lithuania	2005-2025	Annual	21	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Luxembourg	EU-SILC Luxembourg (fielded as PSELL)	2004-2025	Annual	22	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Malawi	Integrated Household Panel Survey (IHPS)	2010-2024	Irregular	5 s	5	5	14.4 †	National	Ongoing	2024	C	microdata.worldbank.org
Malta	EU-SILC Malta	2005-2025	Annual	21	4	4	3.0	National	Ongoing	2025	C	nso.gov.mt
Mexico	Mexican Family Life Survey (MxFLS/ENNViH)	2002-2012	Irregular	3 s	3	3	10.0	National	Dormant	2012	B	ennvih-mxfls.org
Moldova	Household Budget Survey panel (CBGC)	2006-2018	Annual	13	4	4	3.0	Near-national	Ended	2018	B	statistica.gov.md
Montenegro	Survey on Income and Living Conditions (EU-SILC)	2013-2025	Annual	13	4	4	3.0	National	Ongoing	2025	C	monstat.org
Morocco *	Enquete Panel de Menages (ONDH)	2012-2023	Irregular	5 s	5	5	7.0 †	National	Ongoing	2019	C	ondh.ma
Myanmar	Integrated Household Living Conditions Assessment (IHLCA I and II)	2004-2010	Irregular	2 s	2	2	5.0 †	Near-national	Ended	2010	D	catalog.ihsn.org
Nepal	Nepal Living Standards Survey (NLSS) panel component	1995-2010	Irregular	3 s	3	3	15.0 †	National	Ended	2010	B	microdata.worldbank.org
Netherlands	LISS panel (Longitudinal Internet Studies for the Social Sciences)	2008-2025	Annual	18	18	18	17.1 †	National	Ongoing	2025	B	lissdata.nl
New Zealand	Survey of Family, Income and Employment (SoFIE)	2002-2010	Annual	8	8	8	7.0	National	Ended	2010	C	stats.govt.nz
Nicaragua	Encuesta de Medicion de Nivel de Vida (EMNV)	1998-2005	Irregular	3 s	3	3	7.0 †	National	Ended	2005	A	inide.gob.ni
Niger	Enquête Nationale sur les Conditions de Vie des Ménages et l'Agriculture (ECVM/A, LSMS-ISA)	2011-2014	Irregular	2 s	2	2	3.0	National	Ended	2014	B	microdata.worldbank.org
Nigeria	General Household Survey-Panel (GHS-Panel, LSMS-ISA)	2010-2024	Irregular	5 s	5	5	13.3 †	National	Ongoing	2024	B	microdata.worldbank.org
North Macedonia	Survey on Income and Living Conditions (SILC)	2010-2024	Annual	15	4	4	3.0	National	Ongoing	2024	C	stat.gov.mk
Norway	EU-SILC Norway (Survey of Living Conditions)	2003-2025	Annual	23	8	8	7.0 †	National	Ongoing	2025	C	ssb.no
Paraguay	Encuesta Permanente de Hogares Continua (EPHC) semi-panel	2017-2026	Quarterly	38 q	2	2	1.0 †	Near-national	Ongoing	2026	A	ine.gov.py
Peru	Encuesta Nacional de Hogares (ENAHO) panel	2007-2025	Annual	19	5	5	4.0 †	National	Ongoing	2025	A	proyectos.inei.gob.pe

Economy	Survey	Span	Cadence	R	L	W	H, yrs	Universe	Status	Rel.	Access	Official source
Philippines *	Family Income and Expenditure Survey (FIES) master-sample rotation	2003-2009	Triennial	3	3	3	6.0	National	Ended	2009	B	psa.gov.ph
Poland	EU-SILC Poland (GUS)	2005-2025	Annual	21	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Portugal	EU-SILC Portugal (ICOR)	2004-2025	Annual	22	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Romania	EU-SILC Romania	2007-2025	Annual	19	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
Russia	Russia Longitudinal Monitoring Survey (RLMS-HSE)	1994-2024	Annual	29	29	29	30.0 †	National	Ongoing	2024	A	rlms-hse.cpc.unc.edu
Rwanda	Integrated Household Living Conditions Survey poverty panel (EICV3-EICV5)	2010-2017	Irregular	3 s	3	3	6.0 †	National	Dormant	2017	B	statistics.gov.rw
Senegal *	Enquête Pauvreté et Structure Familiale (EPSF)	2006/07-2010/11	Irregular	2 s	2	2	5.0	National	Ended	2011	B	demostaf.web.ined.fr
Serbia	Survey on Income and Living Conditions (SILC)	2013-2025	Annual	13	4	4	3.0	National	Ongoing	2025	C	stat.gov.rs
Slovakia	EU-SILC Slovakia	2005-2025	Annual	20	4	4	3.0	National	Ongoing	2024	C	ec.europa.eu/eurostat
Slovenia	EU-SILC Slovenia (SURS)	2005-2025	Annual	21	4	4	3.0	National	Ongoing	2025	C	ec.europa.eu/eurostat
South Africa	National Income Dynamics Study (NIDS)	2008-2017	Irregular	5 s	5	5	9.0	National	Dormant	2017	B	datafirst.uct.ac.za
Spain	EU-SILC Spain (ECV)	2004-2025	Annual	22	4	4	3.0	National	Ongoing	2025	A	ec.europa.eu/eurostat
Sweden	EU-SILC Sweden (SCB)	2004-2025	Annual	22	6	6	5.0 †	National	Ongoing	2025	C	ec.europa.eu/eurostat
Switzerland	Swiss Household Panel (SHP)	1999-2025	Annual	27	27	26	25.0 †	National	Ongoing	2025	B	forscenter.ch
Tajikistan *	Tajikistan Living Standards Survey (TLSS)	2007-2009	Irregular	2 s	2	2	2.0	National	Ended	2009	B	microdata.worldbank.org
Tanzania	National Panel Survey (NPS, LSMS-ISA)	2008-2021	Irregular	5 s	4	4	6.0 †	National	Dormant	2021	B	microdata.worldbank.org
Thailand *	Panel Socio-Economic Survey (Panel SES, NSO)	2005-2017	Irregular	6	6	6	12.0 †	National	Dormant	2017	D	nso.go.th
Türkiye	Survey on Income and Living Conditions (SILC/GYKA) panel	2006-2025	Annual	20	4	4	3.0	National	Ongoing	2025	C	evam.tuik.gov.tr
Uganda	Uganda National Panel Survey (UNPS, LSMS-ISA)	2009-2020	Irregular	7 s	7	7	10.4 †	National	Dormant	2020	B	microdata.worldbank.org
Ukraine *	Ukrainian Longitudinal Monitoring Survey (ULMS)	2003-2012	Irregular	4	4	4	9.0 †	National	Dormant	2012	B	iza.org
United Kingdom	Understanding Society: UK Household Longitudinal Study (UKHLS) with BHPS	1991-2024	Annual	33	32	32	32.8 †	National	Ongoing	2024	B	understandingsociety.ac.uk

Economy	Survey	Span	Cadence	R	L	W	H, yrs	Universe	Status	Rel.	Access	Official source
United States	Panel Study of Income Dynamics (PSID)	1968-2025	Annual to 1997; biennial since	44	44	43	55.0 †	National	Ongoing	2025	B	psidonline.isr.umich.edu
Uruguay	Encuesta Longitudinal de Proteccion Social (ELPS, BPS)	2012-2016	Irregular	2 s	2	2	3.0	National	Dormant	2016	B	bps.gub.uy
Uzbekistan *	Listening to the Citizens of Uzbekistan (L2CU) monthly panel	2018-2025	Monthly	82	82	82	6.8	National	Ongoing	2025	B	microdata.worldbank.org
Viet Nam *	Viet Nam Household Living Standards Survey (VHLSS) rotating panel	2002-2024	Biennial panel component	9 (12)	3	3	4.0 †	National	Ongoing	2024	B	nso.gov.vn

Panel B. Economies with no nationally or near-nationally representative household panel

Low-income (20): Afghanistan; Burundi; Central African Republic; Chad; Congo, Dem. Rep.; Eritrea; Gambia, The; Guinea-Bissau; Korea, Dem. Rep.; Liberia; Madagascar; Mali; Mozambique; Sierra Leone; Somalia; South Sudan; Sudan; Syrian Arab Republic; Togo; Yemen.

Lower-middle-income (34): Angola; Bangladesh; Bhutan; Bolivia; Cambodia; Cameroon; Comoros; Congo, Rep; Côte d'Ivoire; Djibouti; Egypt; Eswatini; Guinea; Haiti; Honduras; Jordan; Kiribati; Lao PDR; Lebanon; Lesotho; Mauritania; Micronesia; Namibia; Pakistan; Papua New Guinea; Solomon Islands; Sri Lanka; São Tomé and Príncipe; Timor-Leste; Tunisia; Vanuatu; West Bank and Gaza; Zambia; Zimbabwe.

Upper-middle-income (32): Algeria; Argentina; Armenia; Azerbaijan; Belarus; Belize; Botswana; Cabo Verde; Cuba; Dominica; El Salvador; Equatorial Guinea; Fiji; Gabon; Grenada; Guatemala; Iraq; Kosovo; Libya; Malaysia; Maldives; Marshall Islands; Mauritius; Mongolia; Samoa; St. Lucia; St. Vincent and the Grenadines; Suriname; Tonga; Turkmenistan; Tuvalu; Venezuela.

High-income (44): American Samoa; Andorra; Antigua and Barbuda; Aruba; Bahamas, The; Bahrain; Barbados; Bermuda; British Virgin Islands; Brunei Darussalam; Cayman Islands; Channel Islands; Curaçao; Faeroe Islands; French Polynesia; Gibraltar; Greenland; Guam; Guyana; Isle of Man; Kuwait; Liechtenstein; Macao SAR, China; Monaco; Nauru; New Caledonia; Northern Mariana Islands; Oman; Palau; Panama; Puerto Rico (U.S.); Qatar; San Marino; Saudi Arabia; Seychelles; Singapore; Sint Maarten (Dutch part); St. Kitts and Nevis; St. Martin (French part); Taiwan, China; Trinidad and Tobago; Turks and Caicos Islands; United Arab Emirates; Virgin Islands (U.S.).

Note: Panel A lists every economy for which we identified a nationally or near-nationally representative household panel or rotating-panel survey; 80 of the 88 are strictly national and eight near-national, each flagged in its row. Panel B names the 130 with none. Appendix C states the inclusion rule, the protocol, the exclusions and what the audit could not establish. Column definitions. Cadence and R describe the selected longitudinal instrument rather than the survey program that carries it. R is released rounds of that instrument in its own unit, annual unless marked q for quarterly, m for monthly or s for irregular; the column reads R (F), where F is a documented lower bound on the longitudinal rounds fielded and appears only where it exceeds R. R is not compared across cadences. L is the maximum number of consecutive interviews linkable for one household or person, and W how many of those carry a welfare aggregate. H is the elapsed years between the first and last welfare observation of the longest single cohort; a dagger marks every row whose H is reached only by a subsample or one cohort. Status is ongoing where the longitudinal component was fielded in 2023 or later, ended if discontinued, and dormant otherwise, under the exception protocol of Appendix C. Rel. is the survey or reference year of the latest released operation, not the year its file was published. Access tiers record the availability of the survey file: A open download, B free registration or license, C approved researcher or secure facility, and D metadata only; an asterisk marks medium-confidence coding. Sources: our compilation from official survey documentation, audited 19 August and re-audited 1 and 2 September 2026; where a national office and a secondary catalog disagree, the national office is cited.

Table A4. The reporting menu of Table 5, computed for one country: Viet Nam, VHLSS 2010-2014 and the SPID

Row of Table 5	Tier used	Result
1. Vulnerability line and the shares it defines	D1 (given the existing line estimated from the 2010-2012 panel)	At a vulnerability index of 10%, the line is 14,449 thousand dong per person per year in January-2010 prices, 1.84 times the poverty line, with a bootstrap standard error of 915. It divides the 2012 cross-section into 15.0% poor, 32.6% vulnerable and 52.4% secure. Carried forward to the 2012-2014 panel, which the estimate does not use, the same line delivers a realized index of 10.2%.
2. Educational inter-generational mobility	D1, census or DHS	Not computable from this extract, a household file without an individual roster, so parents and children cannot be linked. The only row that requires a different instrument rather than a different tier.
3. Quantile growth below and above the median	D2 (2010 and 2012 cross-sections)	Quantile welfare grew 4.3% a year below the median against 2.5% above it, so growth over the period was pro-poor in the anonymous sense.
4a. Poverty entry and exit rates; joint and conditional transition probabilities, from the panel	D3 (2012-2014 panel)	Poor in both years 10.7%, poor then nonpoor 6.8%, nonpoor then poor 4.6%, nonpoor in both 77.9%; conditional exit 38.9% and entry 5.6%.
4b. The same quantities, from the two cross-sections alone	D2 (synthetic panel, 2012 and 2014)	Poor in both years 10.8%, poor then nonpoor 5.9%, nonpoor then poor 4.0%, nonpoor in both 79.3%; conditional exit 35.1% and entry 4.8%. Every cell falls inside the 95% interval of the panel estimate.
5. Quintile transition matrix and its Shorrocks summary	D3 (2010-2012 panel)	Shorrocks mobility index 0.641 against a maximum of 1.25. Diagonal, poorest to richest quintile: 65%, 38%, 37%, 37%, 66%.
6. Non-anonymous growth incidence curve against its anonymous benchmark	D3 and D2	Both curves are in annualized log growth, 100 times the log ratio divided by the elapsed years. The anonymous curve, over the decile cutpoints of the two cross-sections, runs from 4.4 at the first decile to 0.9 at the ninth. The non-anonymous curve, for the panel households that started in each decile, runs from +19.5 at the first decile to -11.4 at the tenth. The gap between them is not a clean measure of reranking: the two differ in sample and in aggregation as well as in anonymity.
7. Rank correlation of welfare across waves	D3 (two linked waves)	Spearman rank correlation 0.780 over two years.

Row of Table 5	Tier used	Result
8. Chronic versus transient poverty	D3 (three linked waves)	Of the 1,765 households linked across 2010, 2012 and 2014, 7.4% are poor in all three years, 19.3% in one or two, and 73.3% in none. The all-period chronic share of the squared poverty gap is 0.70 on the standard construction, a ratio of weighted aggregates whose numerator evaluates the gap at each household's arithmetic mean consumption and whose denominator is the mean over waves of the contemporaneous gaps. Substituting the geometric mean for the arithmetic mean in the same level-based expression gives 0.76, reported as a sensitivity (Appendix B.8). A second decomposition answers a different question and is labeled separately: the permanent component is 75% of the variance of log consumption in a single wave. That decomposes welfare rather than poverty, and it is not a reliability, since persistent reporting error also loads into the permanent component.
9. Absolute global-class shares and transitions	Harmonized multi-country data (SPID)	The country's six SPID regions had mean welfare of $7.1 to $15.9 a day in 2010 and $11.3 to $21.4 in 2022, growing 3.2% a year on average. One region crossed from the $3.00-8.30 class into the $8.30-28 class; none moved down.
10. Measurement-error sensitivity scenarios for rows 4 and 5	Same tier as the underlying row-4/5 estimator (D2 or D3), plus a measure-specific error model and calibration	These are sensitivity scenarios for the entry and exit measures of Table 4, on the 2010-2012 panel, and not a recalculation of the published 2012-2014 comparison of rows 4a and 4b. At a variance ratio of 0.80: measured exit rate 55.8% against that panel's 46.1%, entry rate 9.3% against 4.9%. Row 5, at the same ratio: Shorrocks M of 0.782 against the panel's 0.641. The scenarios are computed at D3 here, but they need only an error model and a stated calibration and can accompany a D2 estimate on the same terms.

Note: Viet Nam Household Living Standards Surveys 2010, 2012 and 2014, with the SPID regional panel for row 9; row numbering follows Table 5, and row 10 reproduces the classical row of Table 4. Appendix B.8 states the poverty line, the weighting and every specification. For Row 1, the line follows the second definition of Dang and Lanjouw (2017): the welfare level at which 10% of the households lying between it and the poverty line are poor two years later. Appendix B.8 states the estimation and compares it with two probability-based alternatives. For Row 3, this is the anonymous quantile-growth diagnostic. The Ray-Genicot index for panel-free upward mobility can offer an alternative. Rows 4a and 4b reproduce the actual-panel and synthetic-panel poverty transitions of Table 6 in Dang *et al.* (2019) for the 2012-2014 VHLSS pair. The other VHLSS rows use the 2010-2012 pair or all three rounds, as stated in each row and in Appendix B.8. In row 6 both curves are in annualized log growth. Row 9 comes from Table 3. Source for rows 4a and 4b: Dang *et al.* (2019), Table 6, whose synthetic panel is built from the cross-sectional component with household heads aged 25 to 55 in the first round.

Table A5. Published estimates of the reliability of survey welfare measures, and of the share of measured mobility attributable to measurement error

Study	Economy	Design and benchmark	Measure	Estimate
Panel A. Reliability of the welfare measure: share of measured variance that is signal				
Bound and Krueger (1991)	United States	March CPS matched to Social Security payroll records	Annual earnings, level	0.82 men, 0.92 women
Bound and Krueger (1991)	United States	as above	Annual earnings, one-year change	0.65 men, 0.81 women
Bound *et al.* (1994)	United States	PSID Validation Study, employer payroll records	Annual earnings, level	0.85 (1982), 0.70 (1986)
Bound *et al.* (1994)	United States	as above	Annual earnings, four-year change	0.71
Gottschalk and Huynh (2010)	United States	SIPP matched to Social Security earnings records	Annual earnings, level	0.67; 0.73 excluding imputed
Ahmed *et al.* (2006)	Canada	Four-week recall module against a two-week household diary, same households	Food expenditure, level	0.22 to 0.36
Gibson and Kim (2013)	Papua New Guinea	Same households remeasured six months apart	Per capita expenditure, log, against permanent income	0.51
Luttmer (2001)	Russia, Poland	RLMS and the Polish household budget survey, instrumental variables	Income and expenditure, log, level	0.32 to 0.68
Luttmer (2001)	Russia, Poland	as above	Year-to-year change	0.20 to 0.45
Gibson *et al.* (2015)	Tanzania	Eight randomly assigned questionnaires; benchmark a supervised personal diary	Consumption, log; attenuation factor, not a reliability ratio	0.42 to 0.66
Panel B. Share of measured mobility attributed to measurement error				
Glewwe (2012)	Viet Nam	VLSS panel; the income aggregate instruments the expenditure aggregate	Measured mobility	15% to 42% is upward bias
Lee *et al.* (2017)	Korea	KLIPS; error removed by a model-based simulation	One-year poverty exit rate	45% measured against 26% to 31% error-free
Memorandum, not an estimate of the share attributable to error: the response of mobility statistics to added noise of known structure				
This paper, Table 4	Viet Nam	VHLSS 2010-2012 panel; error of known structure INJECTED into the reported panel	Response of exit and entry rates to added noise at a variance ratio of 0.80	Measured exits 21% above and entries 87% above the reported panel

Note: The reliability ratio is the share of the measured variance of a welfare measure that is true signal. It is the quantity denoted λ in Table 4, and is a reliability only under the classical additive-error model of Appendix B.3. The studies with an administrative benchmark all measure earnings in high-income countries, whereas this review is mostly about consumption in poorer countries, where estimates are lower. Bound *et al.* (1994) and several others report the complementary noise share, so the entries above are one minus that share. Gibson and Kim (2013) take permanent income as the target, so their 0.51 counts real transitory variation as well as reporting error, and the estimate of Gibson *et al.* (2015) is an attenuation factor rather than a reliability ratio. Panel B reports shares of measured movement rather than of variance. The table gives the variance ratios of Table 4 an empirical scale; it does not supply a correction factor. Sources: as cited.

Table A6. A structured synthesis of the synthetic-panel validation record

Study	Setting (survey years)	Estimator	Cells	Mean absolute deviation (pp)	Interval-agreement tally (criterion in block note)	Format
Panel A. Studies publishing cell values: deviation from the reported panel and containment, by estimator						
Binary joint-probability point estimators with a parametric or semiparametric dependence model (poor and nonpoor)						
Dang *et al.* (2019)	Viet Nam (2012-14)	Point, cohort rho	4	0.75	4/4	Table
Martinez *et al.* (2013)	Philippines, 3 round-pairs (2003-09)	Point, semiparametric DL(B), cohort-derived correlation, 100 replications (study's primary specification)	12	1.52	n.a.	Table
Moreno *et al.* (2021)	Mexico, 2 lines (2002-05)	Point, Model 1 Regime 1, fixed rho = 0.25, two-normal mixture, 500 draws, two lines	8	1.05	6/8	Table
Dang and Lanjouw (2023)	5 economies (2001-09)	Point, cohort rho	20	0.92	20/20 source-reported; 18/20 reconstructed from printed precision	Table
Subtotal, source-published families	11 settings in 4 publications	-	44	1.09	30/32 source-reported; 28/32 reconstructed (94%; 88%)	-
Supplementary block, not in the headline summary: Lucchetti *et al.* (2025) settings	4 economies (2001-16)	Point, cohort rho; our own estimation on these panels. The unreleased inputs and code are ours rather than the source's, so this row cannot yet be reproduced	16	2.11	10/16	Table
Subtotal, extended record (adds the re-estimation)	15 settings in 5 publications	-	60	1.36	40/48 source-reported; 38/48 reconstructed (83%; 79%)	-
Memorandum: Gafa *et al.* (2024), benchmark on a different sample	Ghana, 2 round-pairs (2010-14)	Point, cohort rho; excluded from both subtotals	8	1.81	6/8	Table
Parametric point estimates, poor, vulnerable and middle class						
Dang and Lanjouw (2017)	Viet Nam (2006-08)	Point, cohort rho	9	1.03	6/9	Table

Study	Setting (survey years)	Estimator	Cells	Mean absolute deviation (pp)	Interval-agreement tally (criterion in block note)	Format
Non-parametric point estimates with a fixed equal weight on the two bound draws, gamma = 0.5 (Lucchetti, 2017)						
Lucchetti (2017)	Peru, 2 periods (2007-11)	Non-param., equal weights	8	2.46	7/8	Table
LASSO with predictive mean matching (Lucchetti *et al.*, 2025)						
Lucchetti *et al.* (2025)	6 tabulated settings of 36 validation panels, 4 economies (2001-16)	LASSO-PMM	24	3.00	6/24 (25%)	Table
Memorandum: the four settings in which both our parametric re-estimation and the published LASSO-PMM estimates exist						
Lucchetti *et al.* (2025), matched	4 matched settings: Argentina, Chile, Nicaragua, Peru	Our parametric re-estimation vs published LASSO-PMM	16 vs 16	2.11 vs 2.94	10/16 vs 3/16	Table
Panel B. Studies publishing figures only: containment tallies (joint transition probabilities)						
Hérault and Jenkins (2019)	Australia, HILDA and Britain, BHPS (31 year-pairs)	Point, cohort rho; panel-informed leading-case specification	124	not published	63/124 (51%)	Figures
Colgan (2023)	Greece (12 year-pairs), sqrt transform; panel-based rho (oracle-assisted)	Point, panel-based rho	48	not published	46/48 (96%); 26/48 under a log transform	Figures
No pooled subtotal	Panel-informed (Hérault and Jenkins) and oracle-assisted (Colgan) designs are not pooled	-	172	-	see rows	-
Panel C. Studies contributing bounds, by how the residual correlation is restricted						
Full range (non-parametric, or parametric at rho = 0 and 1)						
Bierbaum and Gassmann (2012)	Kyrgyz Republic	-	36	-	24/36	Table
Cruces *et al.* (2015)	Chile; Nicaragua; Peru	-	40	-	40/40	Table
Dang and Lanjouw (2023)	5 economies	-	20	-	20/20	Table
Dang *et al.* (2014)	Indonesia; Viet Nam	-	8	-	8/8	Table
Memorandum: Gafa *et al.* (2024), benchmark on a different sample	Ghana	excluded from the subtotal; 4/8 joint cells and 8/8 conditional cells	16	-	12/16	Table
Limanli (2021)	Turkey	-	24	-	24/24	Table
Lucchetti *et al.* (2025)	Argentina; Chile; Nicaragua; Peru	-	16	-	15/16	Table

Study	Setting (survey years)	Estimator	Cells	Mean absolute deviation (pp)	Interval-agreement tally (criterion in block note)	Format
Martinez *et al.* (2013)	Philippines	DLLM(A)-DLLM(B), the semiparametric endpoints at rho = 0 and rho = 1; 100 replications	12	-	12/12	Table
Mekasha and Tarp (2021)	Ethiopia	-	4	-	4/4	Table
Moreno *et al.* (2021)	Mexico, 2 lines (2002-05)	Parametric endpoints at rho = 0 and rho = 1, labeled bounds by the source	8	-	8/8 (100%)	Table
Rongen *et al.* (2026)	Tanzania, 2014-2020, split-panel design	Zero-one bounds, Specification 2	12	-	12/12	Figures
Salvucci and Tarp (2021)	Mozambique	-	48	-	48/48	Table
Tightened by an assumed rho range						
Dang *et al.* (2014)	Indonesia; Viet Nam	-	8	-	6/8	Table
Limanli (2021)	Turkey	-	16	-	16/16	Table
Mekasha and Tarp (2021)	Ethiopia	-	4	-	3/4	Table
Tightened by a rho range estimated from the data						
Martinez *et al.* (2013)	Philippines	DL(A)-DL(C), the narrower interval around DL(B) from cohort-derived rho formulas; 100 replications	12	-	12/12	Table
Tightened by a bootstrap prior over correlations observed in ancillary panels						
Rongen *et al.* (2026)	Tanzania, 2014-2020, split-panel design	Four-standard-deviation band, Specification 2; Specification 1 gives 7/12	12	-	9/12 (75%)	Figures
Subtotal, full range	10 publications whose source is identified; all coded full-range comparisons; specification cells, not distinct transitions	-	212	-	200/212 (94.3%)	-
Memorandum: full range, adding the row whose provenance is unresolved	11 publications; adds the 16 Lucchetti *et al.* (2025) comparisons	-	228	-	215/228 (94.3%)	-

Study	Setting (survey years)	Estimator	Cells	Mean absolute deviation (pp)	Interval-agreement tally (criterion in block note)	Format
Memorandum: full range including Gafa *et al.* (2024)	12 publications; adds Gafa *et al.* (2024) to the memorandum above	-	244	-	227/244 (93.0%)	-
Subtotal, tightened by an assumed rho range	3 publications	-	28	-	25/28 (89.3%)	-
Subtotal, tightened by a rho range estimated from the data	1 publication	-	12	-	12/12 (100.0%)	-
Subtotal, tightened by a bootstrap prior over ancillary correlations	1 publication	-	12	-	9/12 (75.0%)	-
Memorandum: all bounds pooled (repeated specifications and Gafa included)	12 publications	-	296	-	273/296 (92.2%); not a headline	-
Memorandum, not included in any subtotal: benchmark drawn from a different survey (Perez, 2015)						
Perez (2015)	Mexico, ENIGH against the MxFLS	Bounds, three specifications	48	-	22/48	Table
Perez (2015)	as above, point estimates at rho = 0.442	Point, panel-based rho	16	6.04	4/16	Table
Panel D. Summary (deviation and agreement statistics; definitions in Appendix B.4)						
Absolute deviation, extended record	60 cells, 15 settings in 5 publications	MAD (setting-equal) / publication-equal / leave-one-out range	-	1.36 / 1.27 / 1.09–1.59	-	-
Memorandum: all coded binary settings, noncomparable samples pooled	68 cells, 17 settings in 6 publications; includes Gafa	MAD from the released rounded rows (stored unrounded); RMSD and max cell stored	-	1.42 from the released rounded rows; 1.41 stored unrounded	-	-
Absolute deviation, source-published families only (excludes the re-estimation)	44 cells, 11 settings in 4 publications	MAD (setting-equal) / publication-equal / leave-one-out range	-	1.09 / 1.06 / 0.93-1.24	30/32 (94%)	-

Study	Setting (survey years)	Estimator	Cells	Mean absolute deviation (pp)	Interval-agreement tally (criterion in block note)	Format
Absolute deviation, three-way; non-parametric; LASSO-PMM	9; 8; 24 cells	MAD reported; RMSD and max cell stored	-	MAD / RMSD / maximum deviation: 1.03/1.59/3.70; 2.46/2.62/4.00; 3.00/3.32/5.70. The LASSO-PMM MAD is 2.98 from stored unrounded values.	-	-
Total-variation distance	all 17 coded settings; equals 2 × MAD for every 2×2 block and is retained for interpretation only	median / max	-	2.40 / 9.90	-	-
Agreement, extended record	48 cells with a printed benchmark interval	-	-	-	40/48 (83%)	-
Agreement, all coded binary settings	56 cells	-	-	-	46/56 (82%); noncomparable samples pooled	-
Agreement, three-way; non-parametric	9; 8 cells	-	-	-	6/9; 7/8	-
Agreement, LASSO-PMM (this review's screen)	24 cells, 6 tabulated settings	-	-	-	6/24 (25%)	-
Agreement, figures only, reported separately	124 and 48 transition-cell comparisons (31 and 12 year-pairs)	-	-	-	63/124 (51%); 46/48 (96%)	-
Agreement, conditional rates	62 transition-cell comparisons (Hérault and Jenkins)	-	-	-	22/62 (35%) under our digitization of the published figures; the source prose elsewhere implies 23/62	-
Bounds, full range	228 coded full-range specification cells, Gafa excluded, Moreno and Rongen included	-	-	-	215/228 (94.3%)	-
Bounds, tightened (nested specifications and the prior-tightened band)	52 specification-cell comparisons	-	-	-	46/52 (88.5%)	-
Agreement, prior-tightened band	12 cells (Rongen *et al.*, 2026), Specification 2	-	-	-	9/12 (75%); 7/12 under Specification 1	-

Study	Setting (survey years)	Estimator	Cells	Mean absolute deviation (pp)	Interval-agreement tally (criterion in block note)	Format
Bounds, matched contrast within study	28 paired comparisons from three studies, reported under both a free and an assumed correlation	-	-	-	28/28 vs 25/28	-
Cross-survey benchmark, bounds	48 cells (Perez, 2015)	-	-	-	22/48 (45.8%)	-
Cross-survey benchmark, point estimates	16 cells (Perez, 2015)	MAE	-	6.04	4/16	-

Note: pp = percentage points. The table covers the eighteen studies with original validation that have been extracted. Two contribute no comparison and are not listed: Garbero (2014), which applies the bounds method without a panel benchmark, and Ferreira *et al.* (2013), which reports the validation of Cruces *et al.* (2015) qualitatively; a further direct validation, for Thailand, is not yet extracted. Rongen *et al.* (2026) screen a reported-panel estimate against a synthetic interval, so they enter Panel C. Each study contributes one row per estimator block, with its total in the cells column and the unweighted mean over its settings in the deviation column; Appendix B.4 defines every statistic and explains why Dang and Lanjouw (2023) is reported both at 20 of 20 and at 18 of 20. The source-published point subtotal excludes the Gafa *et al.* (2024) memorandum, whose benchmark uses a different household sample, and the four Lucchetti *et al.* (2025) settings, which are our estimation on those panels. Panel C's primary subtotal is the ten publications whose source is identified; the eleven-publication total that adds the row whose provenance is unresolved is reported as a memorandum. Identifying a row's source is not the same as transcribing it: an exact table and page are recorded for three of the eleven contributing bounds rows, and the others are entered from the published tallies. Panel B counts transition cells, four per binary matrix and not year-pairs, coded as the sources report them; Panel C counts specification cells, because several studies report several models for one benchmark. The conditional tally of 22 of 62 in Panel D, for Hérault and Jenkins, is our own digitization of published figures.

Table A7. Sensitivity of the global mobility estimates to the definition of the classes and of the sample

Specification	Regions moving up a class (%)	Regions unchanged (%)	Regions moving down (%)	Persistence, bottom class (%)	Persistence, top class (%)
Panel A. Class boundaries and class widths (2021 PPP dollars per person per day)					
Baseline: \$3.00 / \$8.30 / \$28	20.4	76.8	2.8	47.8	99.0
Lower cutoffs: \$2.15 / \$6.85 / \$24	17.6	80.4	2.0	42.9	99.6
Higher cutoffs: \$3.65 / \$10.00 / \$32	19.7	77.4	2.9	61.5	97.9
Median-based: \$2.50 / \$7.00 / \$25	20.3	77.5	2.2	33.8	99.1
Equal log widths: \$4.03 / \$13.08 / \$42.46	20.6	76.6	2.8	70.3	98.2
Panel B. Distance to the boundary the region would have to cross					
Quartile of the gap to the boundary the region must cross	*Regions*	*Mean gap (log points)*	*Crossing that boundary (%)*		
Upward, gap to the next higher class (n = 1,030)					
Nearest quartile	258	0.14	62.8		
Second quartile	257	0.41	29.2		
Third quartile	257	0.70	5.1		
Farthest quartile	258	0.99	0.8		
Downward, gap to the next lower class (n = 1,097)					
Nearest quartile	275	0.15	11.6		
Second quartile	274	0.43	1.1		
Third quartile	274	0.68	0.0		
Farthest quartile	274	0.93	0.0		
Panel C. Weighting					
Weighting scheme	*Moving up (%)*	*Moving down (%)*	*Mean growth (% per year)*		
Unweighted: each region counts once (baseline)	20.4	2.8	1.75		
Each country weighted equally	20.8	3.4	1.75		
Panel D. Length of the observation window, with annualized equivalent rates					
Span between anchor surveys	*Regions*	*Mean span (years)*	*Moving up (%)*	Annualized equivalent rate (% per year)	*Implied 10-year probability (%)*
9 years or less	163	7.5	6.7	0.93	8.9

Specification	**Regions moving up a class (%)**	**Regions unchanged (%)**	**Regions moving down (%)**	**Persistence, bottom class (%)**	**Persistence, top class (%)**
10 to 12 years	502	11.0	18.7	1.87	17.2
13 years or more	570	13.1	25.8	2.25	20.3
Panel E. Comparability of the two surveys behind each pair					
Sample	*Regions*	*Countries*	*Moving up (%)*	Leaving the bottom class (%)	*High-income share of regions (%)*
All matched pairs (baseline)	1,235	91	20.4	52.2	15.4
Same welfare type and comparability flag at both dates	565	50	13.6	17.0	31.5

Note: The sample is the 1,235 regions in 91 economies of Table 3 unless stated otherwise. Panel A is a joint sensitivity to where the boundaries sit and how they are spaced. Each row shifts the cutoffs, and the last row also makes the interior log intervals equally wide, which moves every cutoff, the lowest from $3.00 to $4.03; the level and the spacing therefore change together and their effects cannot be separated here. Panel B groups regions by the distance in logs to the boundary they would have to cross; it establishes that distance predicts crossing, not how much classification error the estimated regional means induce. Panel C contrasts the unweighted baseline with country-equal weights. Panel D converts the share moving up into an annualized equivalent and an implied ten-year equivalent share: both are endpoint-based transformations of the observed movement, not estimated transition hazards and not forecasts, since two dates cannot reveal reversals in between. Panel E restricts the sample to pairs whose two surveys carry the same welfare type and the same comparability flag in the SPID, so that the change is not read across a documented break; the retained subsample is differently composed, as its last column shows. Source: our calculations from the SPID.

Table A8. Decomposition of regional growth, dispersion under resampling of countries, and the growth-baseline relation of Figure A1

Component or specification					
Panel A. Between- and within-country decomposition of the variance of regional growth					
Component	*Variance*	*Share of total (%)*			
Total variance of annualized regional growth	5.26	100.0			
Between countries	2.96	56.3			
Within countries	2.30	43.7			
Memo: between-country share of the variance of initial log welfare	-	93.0			
Panel B. Dispersion of the diagonal of Table 3 under resampling of countries					
Initial global welfare class	*Persistence (%)*	*Nominal binomial SE*	*Country-resampling dispersion*	*Ratio*	
Below $3.00	47.8	4.3	11.8	2.7	
$3.00 to $8.30	68.5	2.5	4.7	1.9	
$8.30 to $28	81.2	1.7	6.8	4.0	
At least $28	99.0	0.7	1.2	1.7	
Panel C. The growth-baseline relation of Figure A1 under clustering, fixed effects, a common horizon and baseline measurement error					
Specification for the growth-baseline relation	*Regions*	*Slope*	*Dispersion*	*t*	*Units*
Initial global percentile, unclustered	1,235	-0.008	0.002	-3.6	pp/yr per percentile
Initial global percentile, clustered on 91 countries	1,235	-0.008	0.007	-1.1	pp/yr per percentile
Adding country fixed effects, initial log welfare	1,235	-2.42	0.32	-7.7	pp/yr per log point
Initial log welfare, clustered	1,235	-0.28	0.20	-1.4	pp/yr per log point
Initial log welfare, horizon of exactly 12 years	154	-0.01	0.51	-0.0	pp/yr per log point
implied slope under an assumed reliability $\lambda = 0.95$	1,235	+0.18	0.21	n.a.	pp/yr per log point
implied slope under an assumed reliability $\lambda = 0.90$	1,235	+0.70	0.22	n.a.	pp/yr per log point
implied slope under an assumed reliability $\lambda = 0.80$	1,235	+1.92	0.25	n.a.	pp/yr per log point

Note: Components may not sum to totals because of rounding. In panel C every row uses standard errors clustered by country; the fixed-effects row counts the 91 absorbed country intercepts in its degrees-of-freedom factor, as Appendix B.5 states. The sample is the 1,235 regions in 91 economies of Table 3. Panel A splits the variance of annualized regional growth into between- and within-country components and repeats the exercise for initial log mean welfare; both components include survey sampling and comparability error in unknown proportions, so the split is neither an estimate nor a bound on the within-country share of underlying economic variation. Panel B resamples the 91 countries with replacement 2,000 times, taking each drawn country's regions together, which is not design-valid sampling error. Panel C regresses annualized regional growth on initial position. Its fixed-effects row is not a robust version of the pooled slope, since only 7% of the variance of initial log welfare is within countries, and the rows labeled implied slope are a sensitivity exercise under an assumed reliability, not a correction; Appendix B.5 states both conventions and the assumptions they require. Source: our calculations from the SPID.

Table A9. Regional global mobility and national income-classification transitions, 2010-2024

Country classification change (2010-2024)	Countries	Regions	Mean within-country share of regions moving up (%)	Mean regional growth (% per year)	Within-country SD of growth (pp)
Upgraded	23	313	33.9	2.48	1.13
			(6.80)	(0.45)	(0.14)
Unchanged	67	915	16.6	1.51	1.43
			(2.96)	(0.25)	(0.14)
Downgraded	1	7	0.0	1.45	1.08

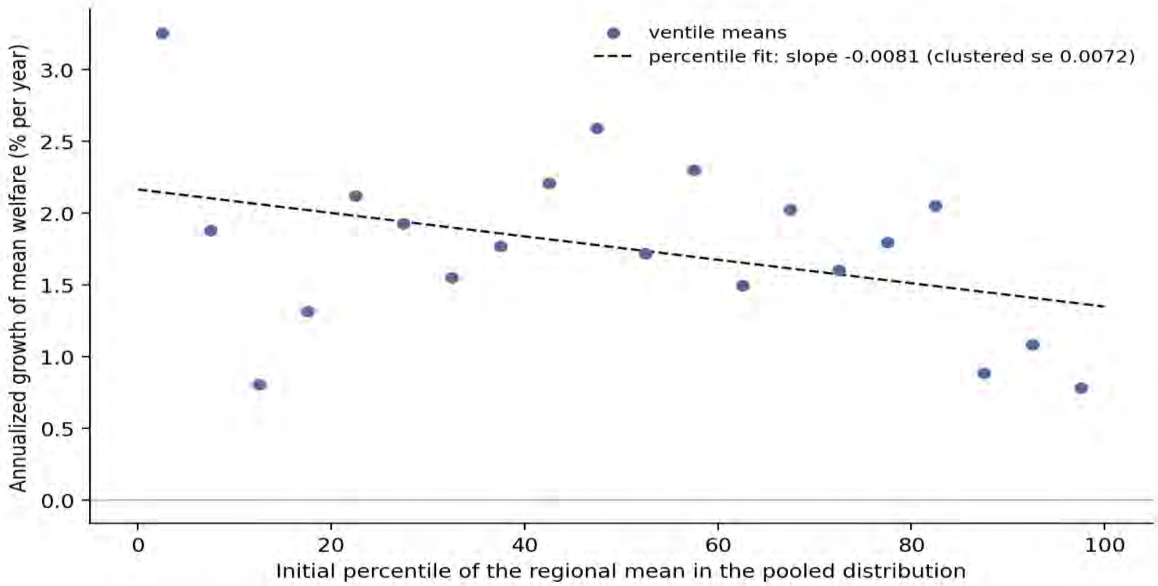
Note: Standard errors in parentheses are computed across countries within each group, as the standard deviation divided by the square root of the number of countries; they are not reported for the downgraded group, which contains one country. Countries with fewer than two sample regions are excluded, and regional classes and the estimation sample are those of Table 3. The last three columns are unweighted averages across countries. pp = percentage points. Source: our calculations from the SPID and the World Bank income-classification panel.

Table A10. Robustness of the measurement-error simulations: replication on the adjacent panel, sensitivity to the poverty line, and the gross flows behind the net change

Sample and reliability λ	Poverty rate (%)	Exit rate (%)	Entry rate (%)	Joint poor-poor (%)	Shorrocks mobility index M	Rank correlation
Panel A. Replication on the adjacent 2012-2014 panel						
2010-2012 panel (baseline), 1.00	19.1	46.1	4.9	10.3	0.641	0.780
0.90	21.0	52.1	7.0	10.1	0.729	0.692
0.80	22.6	55.8	9.3	10.0	0.782	0.608
2012-2014 panel (replication), 1.00	16.0	47.1	4.7	8.5	0.634	0.773
0.90	17.5	52.5	6.4	8.3	0.730	0.687
0.80	19.0	56.2	8.4	8.3	0.784	0.607
Panel B. Sensitivity to the poverty line (2010-2012 panel, λ = 0.80)						
Poverty line	*Poverty rate (%)*	*Exit,* reported panel *(%)*	*Exit, λ = 0.80 (%)*	*Increase on reported (%)*	*Entry,* reported panel *(%)*	*Entry, λ = 0.80 (%)*
10th percentile	11.3	52.6	63.4	20	2.1	5.5 (+158%)
national rate, 17.8% (baseline)	19.1	46.1	55.8	21	4.9	9.3 (+87%)
25th percentile	26.4	41.8	50.5	21	7.0	12.3 (+75%)
Panel C. Transitions created and concealed by the injected error, exits and entries separately (2010-2012 panel)						
Reliability λ	*Created (pp of all households)*	*Concealed (pp of all households)*	*Match rate (%)*	*Retention rate (%)*		
Poverty exits						
0.95	3.4	2.2	66	75		
0.90	5.0	2.8	54	68		
0.80	7.5	3.7	40	58		
0.70	9.7	4.2	32	52		
Poverty entries						
0.95	2.1	1.4	56	66		
0.90	3.2	1.7	42	59		
0.80	5.2	2.0	28	50		
0.70	7.0	2.2	21	45		

Note: Rows labeled no added error are the reported panel itself, which carries its own survey error. The design is that of Table 4 and Appendix B.3: classical error is injected into both waves so that λ equals the stated value, and each entry is the mean over 400 replications. Panel B reports the increase in each rate relative to the reported panel, computed on unrounded values; its percentile cutoffs are taken from the full 2010 cross-section, while the rates displayed against them are computed on the linked panel sample. Panel C compares the two classifications household by household: a created transition is one the injected data record and the reported panel does not, and a concealed transition the reverse. All of these quantities measure disagreement with the reported-panel classification, not with latent transitions. The 2012-2014 panel links 3,869 households, 1,740 of them also in the 2010-2012 panel, so it is a replication on a partly overlapping sample. Source: our calculations from the VHLSS 2010, 2012 and 2014 panels.

Figure A1. Initial position and subsequent growth of sampled regional means



Note: Each dot is the mean of one twentieth of the 1,235 sample regions, ordered by initial percentile; the binned points are descriptive. Growth is annualized log growth throughout, as defined in Section 5.2. The dashed line is the fit on the underlying regions, with a slope of -0.008 per percentile (0.007). Regressing the same growth rates on initial log mean welfare is a separate specification, not a rescaling, and gives -0.28 (0.20). Both standard errors are clustered on the 91 countries and neither slope is distinguishable from zero, so the line describes the data and does not identify convergence. Table A8, panel C, reports the baseline-error sensitivity with the assumptions it requires. Source: our calculations from the SPID.

Appendix B. Formulae and technical details

This appendix collects the formal statements behind the measures and procedures of the main text. Throughout, y denotes log welfare, so y = log(Y) for a welfare aggregate Y. All distributional assumptions are made on the log scale, and poverty lines are written in the same units. The sections follow the order of the main text, and equations are numbered (B1) to (B19).

B.1 Synthetic panels: bounds and point estimates

Our exposition follows Dang *et al.* (2014) and Dang and Lanjouw (2023). Suppressing the country index, let x(i,j) be a vector of characteristics for household i observed in survey round j, j = 1 or 2, that is also observed, or can be reconstructed, in the other round: the household head's year of birth, sex, ethnicity, religion, language, place of birth, parental education, and education completed by adulthood, together with any round-1 values recovered by retrospective questions asked in round 2. A data prerequisite is that the two rounds be comparable.

Estimation samples are usually restricted to household heads aged 25 to 55 in the first cross-section, with the range shifted by the interval length in the second, which limits artifactual change from household formation and dissolution and keeps schooling close to fixed; population weights are then applied, so that estimates represent the population the selected cohorts cover rather than the whole population automatically. Let y(i,j) denote log household consumption or income, following the convention of this appendix. The linear projection of welfare on characteristics in each round is

$$y_{ij} = \beta_j' x_{ij} + \varepsilon_{ij} \tag{B1}$$

with the coefficients allowed to differ across rounds. Let z(j) be the poverty line in round j. The unconditional, or joint, measures of poverty mobility take the form

$$P(y_{i1} < z_1 \; and \; y_{i2} > z_2) \tag{B2}$$

which is the share of households poor in the first round and nonpoor in the second, while the conditional measures take the form

$$P(y_{i2} > z_2 \mid y_{i1} < z_1) \tag{B3}$$

which is the share of the initially poor who escape poverty. With an actual panel both are estimated directly. With two independent cross-sections the marginal distributions are observed but the pair is not, so these joint probabilities are not identified without further structure. Two assumptions supply it. Assumption 1 requires that the two rounds sample comparable target cohorts over a common covariate support, and that each round's conditional model transport credibly within that support; equality of coefficients across rounds, or of the conditional distributions themselves, is not required.

Assumption 2 assumes that the error terms ε_{ij} are positive quadrant dependent conditional on the x(i,j) characteristics, implying that their correlation is non-negative (i.e., ranging between larger than 0 and 1), which can result in lower bound and upper bound estimates of poverty mobility. For point estimates, Dang *et al.* (2014) assume a stronger version of Assumption 2, which assumes joint normality of the error terms, with standard deviations s(1) and s(2) and non-negative correlation r, under which

$$P(y_{i1} < z_1 \; and \; y_{i2} > z_2 \mid x_{i2}) = \Phi_2(a_1(x_{i2}), -a_2(x_{i2}), -\rho), a_t(x) = \frac{z_t - \beta_t{}'x}{\sigma_{\varepsilon t}} \tag{B4}$$

The probability itself is the expectation of this conditional probability over the distribution of the characteristics, and its estimator is the survey-weighted sample analogue over the households of the target round:

$$\hat{P}(y_{i1} < z_1 \; and \; y_{i2} > z_2) = \frac{\sum w_i \Phi_2(a_1(x_{i2}), -a_2(x_{i2}), -\rho)}{\sum w_i} \tag{B5}$$

where the bivariate normal cumulative distribution function is evaluated at each household in the target round. The other three cells of the transition matrix follow analogously. The conditional rate is the ratio of two weighted sums and not the weighted mean of household-level ratios: the estimator is the sum over households of the weighted bivariate probability divided by the sum over the same households of the weighted univariate marginal, so that the denominator is the estimated share of the conditioning group rather than a per-household normalization.

In the sample estimator the coefficient vector, the residual standard deviation and the residual correlation are replaced by their estimates, which we write without hats in the display equations only to keep them legible; the generated-regressor uncertainty this introduces is the reason standard errors must propagate the first stage as well as the second. The estimated parameters are applied to round-2 data for prediction, though round 1 can equally serve as the base year.

What remains is the residual correlation, which two cross-sections cannot observe. The standard approach approximates it from the correlation of cohort-aggregated welfare between the two surveys, netting out the part explained by observed characteristics:

$$\rho = \frac{\rho_{y_{i1}y_{i2}}\sqrt{var(y_{i1})var(y_{i2})} - \beta_1{}'var(x_i)\beta_2}{\sigma_{\varepsilon 1}\sigma_{\varepsilon 2}} \tag{B6}$$

Dang and Lanjouw (2023) note a second route, which we recommend reporting alongside the first. Assume the residuals contain a cohort fixed effect, aggregate all the time-invariant variables to the cohort level, and estimate

$$y_{cj} = \beta_j{}' x_{cj} + \varepsilon_{cj} \tag{B7}$$

where c indexes cohorts, x(cj) is the vector of time-invariant characteristics averaged within cohort, and the error term contains a cohort fixed effect and an idiosyncratic component. The estimator is the correlation across cohorts of the two rounds' fitted residuals, weighted by cohort size. What it recovers is cohort-level persistence, and that is not automatically the household-level residual correlation that equation (B4) requires: aggregation removes idiosyncratic variation, and finite cohort cells add sampling error to the aggregates, so the two can differ in either direction. We therefore treat agreement between the two routes as sensitivity evidence rather than as validation: similar values are reassuring, and divergent ones warn that the parameter is doing too much work. That warning is not hypothetical.

The correlation is sensitive to how cohorts are defined, and no statistic available in the absence of panel data reliably identifies the better definition (Hérault and Jenkins, 2019; Garcés-Urzainqui *et al.*, 2021; Colgan, 2023). It may also be inconsistent even in large samples (Moreno *et al.*, 2021). The point estimator is in this sense an imputation of the unobserved second observation rather than a link, so its uncertainty carries a between-imputation component that a single draw does not display (Rubin, 1996). A third route replaces the cohort estimate with a bootstrap draw from correlations observed in ancillary panels, which narrows the admissible interval at the price of assuming that the external collection is informative for the country at hand (Rongen *et al.*, 2026). Section 5.1 supplies the empirical counterpart: across 28 paired comparisons from three studies, each reporting both a free and an assumed interval for the same cell, imposing the assumed range lowers containment of the benchmark value from 28 of 28 to 25 of 28, while the one study that tightens with a correlation range estimated from its own data loses no containment at all, and the two studies reporting in figures trace their weaker performance largely to this parameter. Standard errors should propagate both sampling and estimation error, for instance by bootstrapping the entire two-round procedure, since treating imputed welfare as data understates uncertainty.

B.2 Measurement error and mobility statistics

Let measured log welfare be true log welfare plus a classical error: mean zero, independent of true welfare and of the error in the other round. Define the reliability ratio in round t as

$$\lambda_t \;=\; \frac{var(y_t)}{var(y_t) + var(u_t)} \;=\; \frac{var(y_t)}{var(w_t)} \tag{B8}$$

the share of measured variance that is signal. Two consequences organize Section 5. First, correlations attenuate:

$$corr(w_1, w_2) \;=\; \sqrt{\lambda_1 \lambda_2}\, corr(y_1, y_2) \tag{B9}$$

which reduces to the reliability ratio times the true correlation when the two rounds are equally reliable. Any mobility index decreasing in the inter-temporal correlation is then biased upward, provided the process is positively persistent. Rank correlations are also affected, and they decline in the scenarios examined here, but no general proportional attenuation formula holds for them. A general direction would require stronger distributional restrictions than the classical conditions above. Second, threshold statistics are distorted through misclassification. Let a(t) be the probability that a truly nonpoor household is measured poor in round t, and b(t) the reverse, collected in the classification matrix

$$M_t \;=\; \begin{bmatrix} 1 \,-\, b_t & b_t \\ a_t & 1 \,-\, a_t \end{bmatrix} \tag{B10}$$

whose rows are true states and whose columns are observed states, poor first. If J is the true joint-state matrix, the observed joint matrix is the two-sided product

$$\tilde{J} \;=\; M'_1\, J\, M_2 \tag{B11}$$

and the observed conditional rates follow on row-normalizing it. Two conditions are needed, and neither follows from additive error. Classification must be nondifferential within a round, and errors must be independent across rounds. Independent additive errors do not generally guarantee these conditions: crossing probabilities depend on distance from the line, and those distances may be related across waves. The misclassification representation is therefore an approximation, which is why Section 5.3 simulates the additive model directly. Where only the second round is misclassified, and e and x are the true entry and exit rates, the observed rates are

$$e_{obs} \;=\; (1 \,-\, b_2)e \,+\, a_2(1 \,-\, e), x_{obs} \;=\; (1 \,-\, a_2)x \,+\, b_2(1 \,-\, x) \tag{B12}$$

Setting a(2) = b(2) = m gives e(obs) - e = m(1 - 2e) and x(obs) - x = m(1 - 2x). Three things follow. A rate rises under symmetric misclassification only if its true value lies below one half, so the direction is not general. The absolute distortion is governed by distance from one half: at the Vietnamese values the same m moves entry by 0.90m and exit by 0.08m. It does not follow that joint probabilities are generally the least distorted statistics. The change in a joint cell is the change in the corresponding conditional rate multiplied by the share of the origin group, which is a denominator identity rather than evidence about accuracy: in Table 4 the share poor in both years barely moves, while the poor-then-nonpoor cell is the exit rate times the initial poverty rate, and both of those rise with the injected error. What the exercise supports is the narrower point that a stable aggregate cell need not imply stable household trajectories. These are properties of this model at these rates, not a general law. Under mean-reverting error, correcting as though the error were classical mis-corrects, and the direction depends on the statistic: in the fixed-size example of Table 4 the mean-reverting design inflates the exit rate more than classical error does, so a classical-only inversion would under-correct that inflation (Bound and Krueger, 1991; Lee, 2022).

B.3 The measurement-error simulation design

The experiment behind Table 4 and Figure 6 injects error of known structure into an observed panel. Write y for log per capita expenditure as the VHLSS panel reports it, and w for the series after injection. y is the benchmark, not latent welfare: the VHLSS aggregate already contains error of unknown size and structure, and nothing here identifies it.

The ratio λ is the variance of the reported panel over that of the injected series. It is defined against y, and it is the reliability of the injected series only in the classical rows, not the total reliability of the VHLSS against latent welfare. The estimates in Table A5 therefore scale it; they do not calibrate it. For a target ratio the classical error is drawn independently in each round as

$$u_{it} \sim N(0, V_t \frac{1-\lambda}{\lambda}) \tag{B13}$$

which delivers an injected series whose variance ratio against the reported panel matches the target in expectation. The draws are not centered or made weighted-orthogonal to the reported series within a replication, so the realized ratio varies around the target; Table 4 therefore reports the realized error size for every row.

The mean-reverting variant requires more care. The measured value must be written in centered form, w = m + (1 - d)(y - m) + v, where m is the round's weighted mean, so that the error has mean zero and injection leaves the mean of log welfare unchanged. It does not leave the poverty rate unchanged: a fixed line cuts a distribution whose variance and shape both move, and Table 4 records measured poverty rising from 19.1% to 23.0% and 20.6%. Centering removes a mechanical location shift, not the headcount change.

The correct calibration for a target variance ratio λ is

$$var(v) = V[\frac{1}{\lambda} - (1 - d)^2] \tag{B14}$$

under which the correlation between the error and reported-panel log welfare is

$$corr(u, y) = \frac{-d}{\sqrt{\frac{1}{\lambda} - 1 + 2d}} \tag{B15}$$

so that if a particular negative correlation c is also targeted, the mean-reversion parameter solves

$$d = c^2 + |c|\sqrt{c^2 + \frac{1}{\lambda} - 1} \tag{B16}$$

At λ = 0.80 and a target correlation of -0.20 this gives d = 0.1477 and a disturbance variance of 0.5236 times V, the calibration used in the first mean-reverting row of Table 4. Two finite-sample points follow. The shrinkage leaves mean log welfare unchanged only in expectation, because the disturbance is drawn rather than centered within each replication. The realized error share reported in Table 4 is computed on the first wave, although error is injected in both, and the two waves differ only by sampling variation. The two error models give λ different meanings. Under classical injection it is a conventional reliability ratio. Under the centered mean-reverting form it is only a variance ratio: the squared correlation is (1 - d) squared times λ.

We therefore use reliability for the classical rows only, and variance ratio throughout. The second mean-reverting row fixes the error size instead, setting Var(u)/Var(y) to the classical value 0.25, which requires d = 0.100 and implies λ = 0.9524; the second mean-reverting row and the classical row at λ = 0.80 both set Var(u)/Var(y) to 0.25, so comparing them varies error-welfare dependence at a fixed marginal error variance, with the small additional change in temporal dependence noted below.

The classical errors, and the innovations in the mean-reverting design, are drawn independently across waves. The total mean-reverting error is not independent across waves: because it contains the term that pulls reported welfare toward its mean, Cov(u0, u1) = d^2 Cov(y0, y1), which on these data is a correlation of about 0.03 in both calibrations, so the fixed-size comparison varies induced temporal dependence slightly as well as error-welfare dependence. The exercise therefore illustrates transitory error and does not cover persistent reporting error; as a diagnostic, imposing a cross-wave error correlation of 0.5 at λ = 0.80 gives exit and entry rates of 49.2% and 7.3% rather than 55.8% and 9.3%. The exit rate is the share of the initially poor measured on the other side of the line in the second wave, the entry rate the same quantity for the initially nonpoor, and the joint probability the share of all households poor in both waves. For each replication we recompute the poverty rate, the entry and exit rates, the joint probability of poverty in both rounds, the rank correlation, and the Shorrocks trace index,

$$M = \frac{K - tr(T)}{K - 1}, K = 5 \tag{B17}$$

where T is the row-stochastic quintile transition matrix. M is zero under the identity matrix and one when rows are independent of the origin state, but it is not bounded above by one: it can reach K/(K - 1), or 1.25 at K = 5.

Table 4 reports means across replications, with Monte Carlo standard deviations for the injected-error rows and bootstrap standard errors for the reported-panel row. The bootstrap draws panel households uniformly with replacement and recomputes the weighted statistic with the original survey weights. It does not resample primary sampling units or respect strata, which the released file does not identify, so it is resampling dispersion rather than design-valid sampling uncertainty.

For each replication we also record whether the reported panel and the injected data place each household on opposite sides of the line. The match rate is the share of injected-data transitions that are also reported-panel transitions, and the retention rate the share of reported-panel transitions that survive injection. Both are statements about agreement with the reported panel. Because that panel is itself measured with error, neither estimates nor bounds disagreement with latent true transitions.

B.4 The quantitative synthesis of validation studies

Primary validation comparisons require the synthetic estimate and the panel benchmark to refer to the same target population and period, with materially comparable welfare concepts, transition definitions, statistical units and weighting conventions; comparisons failing one or more of these conditions are retained as memorandum evidence and excluded from the pooled descriptive summaries. Section 5.1 assembles the studies that compare a synthetic panel with an actual panel under that rule. It is a structured descriptive synthesis, not a meta-analysis. This appendix defines each statistic, states the aggregation levels, and records the search and coding rules.

Let p index publications, s validation settings, and k = 1, …, K(s) the cells of a complete joint transition matrix. A validation setting is one study family by economy by survey pair by poverty line or welfare classification by estimator specification; journal and working-paper versions of one study form a single family and are counted once. Let S(p,s,k) be the synthetic estimate and T(p,s,k) the reported-panel benchmark of the same joint transition share, both in percentage points, and define the deviation d(p,s,k) = S − T. The benchmark is an estimated comparison value rather than latent truth, so d is called a deviation rather than an error or a bias.

For each setting the mean absolute deviation is $MAD(p,s) = K(s)^{-1} \Sigma_k |d(p,s,k)|$ and the root mean squared deviation is $RMSD(p,s) = [K(s)^{-1} \Sigma_k d(p,s,k)^2]^{½}$. The headline MAD is the unweighted mean of MAD(p,s) over the settings of the source-published block; an extended figure adds our own re-estimation on the Lucchetti *et al.* (2025) panels. A publication-equal mean, which first averages within each publication, and a leave-one-publication-out range are reported alongside both. The RMSD and the maximum cell deviation in Table A6 are pooled over the cells of the block. For a complete matrix the total-variation distance is $TV(p,s) = ½ \Sigma_k |d(p,s,k)| = (K(s)/2)$ MAD(p,s), so for every two-by-two block TV equals twice the MAD. Because the cell differences are expressed in percentage points, TV is reported in percentage points as well.

It is retained only because it reads as the largest discrepancy in any aggregate formed from the matrix, and it is not independent evidence. Complete, consistently normalized matrices sum to one hundred, so signed cell deviations sum to zero apart from rounding, which is why absolute deviations are summarized. No inverse-variance synthesis is fitted, because the covariance between synthetic and benchmark estimates is unavailable, cells and settings are dependent, several benchmark panels recur across rows, reporting of numeric cells is selective, and the source-published block contains four publications, the extended block five. A random-effects synthesis (DerSimonian and Laird, 1986), a meta-regression on study characteristics (Stanley and Doucouliagos, 2012) or robust variance estimation with small-sample adjustments (Tipton and Pustejovsky, 2015) would each require inputs that this record does not carry.

$$TV_{ps} = \frac{1}{2}\sum_{k=1}^{K(s)} |S_{psk} - T_{psk}| = \frac{K_s}{2} MAD_{ps} \qquad \text{(B18)}$$

For point estimators the interval-agreement indicator is C(p,s,k) = 1 if S(p,s,k) lies inside the benchmark interval [L, U] the panel study reports, with endpoints inclusive. Those intervals are not uniform in level or construction across sources, one being a bootstrap percentile band rather than a 95% interval, so the pooled measure is a source-reported benchmark-interval screen and not a common 95% screen. For synthetic bounds under specification m the direction reverses: C(p,s,k,m) = 1 if T(p,s,k) lies inside the synthetic interval. Published unrounded values, printed interval endpoints and source-reported tallies take precedence over anything reconstructed. Where a source states its own count of estimates falling inside the benchmark interval, that count governs, and any cell that a reconstruction from rounded inputs would place outside is flagged in the cell-level record with the margin and the standard error that would reconcile it. For Dang and Lanjouw (2023) the source reports 20 of 20; rebuilding each interval from the printed one-decimal standard errors places the United States and Viet Nam poor-to-nonpoor cells outside by 0.012 and 0.020 points, giving 18 of 20. Standard errors of at least about 0.306 and 0.510, consistent with the printed values, would place both inside, but those values are unpublished, so both conventions are reported. The screen needs only the endpoints L and U, so no standard error is inferred from an interval.

The reverse case does arise: where a study prints a point estimate T and an ordinary standard error but no interval, the interval is constructed as T ± 1.96 SE, and no such construction is applied where asymmetric, transformed or bootstrap endpoints are published, which are used as printed. The tallies are descriptive screens, not coverage probabilities or equality tests: they ignore the uncertainty of the synthetic estimate, the covariance between the two estimates and the dependence among cells, and a wider interval passes more easily. Numerators and denominators are therefore reported with every tally, and the containment of bounds should be read alongside their width, which the cell-level record will carry. The point-estimator primary block uses one stated specification for each defined setting. The full-range bounds subtotal instead counts all coded full-range specification–cell comparisons and therefore gives greater weight to studies reporting more variants; it is the tally reported in the main text, on the primary ten-publication scope, and should be read as an all-specifications count rather than as one comparison per estimand.

The primary bounds subtotal covers the ten publications whose contribution is traceable to an identified published source; the Lucchetti *et al.* (2025) row is reported separately, as a memorandum, because its table and page are unresolved. Identification at that level is weaker than transcription: exact table and page references are held for three of the eleven full-range rows, and the rest are entered from the published tallies. and the Thailand validation identified in the search is outside the coded record because its cells were not extracted. One-specification-per-estimand and interval-width summaries are not reported. Oracle-assisted calibrations that use a panel-based correlation, and conditional rates, are reported separately. The search covered journal articles and working papers citing Dang *et al.* (2014) or Dang and Lanjouw (2013, 2023), their reference lists and forward citations, through 31 August 2026, with working papers admitted.

B.5 Standard errors for the subnational exhibits

In Table 3 the entries of Panel B are row shares of a multinomial classification, so the standard error of a share p in a row containing n regions is

$$se(p) = \sqrt{\frac{p(1-p)}{n}} \tag{B19}$$

which is the nominal binomial reference under independent regions. It is not the inference Table 3 reports. Before a standard error is interpreted the target must be named, and two are defensible. Under the finite-set target the 1,235 regions are the population of interest, so no uncertainty from sampling that finite set attaches to the shares; the survey-estimation error in each regional mean and the classification error near a threshold remain. Under the superpopulation target the sampled countries are draws from a wider set that might have been surveyed.

The binomial formula corresponds to neither. It ignores the sampling error of the underlying regional means, and it ignores the clustering of regions within countries, which is severe because most of a country's regions move together. Table 3 therefore reports resampling dispersion: we draw the 91 countries with replacement 2,000 times, take all of a drawn country's regions together, recompute the matrix, and report the standard deviation of each row share. It is 1.7 to 4.0 times the binomial figure.

We are deliberate about what this measures: how much the estimates move when the set of countries is perturbed, under the assumption that countries are exchangeable. That is an assumption about the compilation, not a design fact. Two components sit outside it: the regional means are themselves survey estimates with sampling error, for which the extract carries no replicate weights, and class assignment is a deterministic function of an estimated mean, so error near a boundary produces classification error. Neither source of uncertainty is propagated by these resampling figures. In Table A9 the reported quantities are averages across countries, so the standard error is the between-country standard deviation divided by the square root of the number of countries.

Regarding the fixed-effects row of Table A8, panel C, its standard error is clustered by country, as the pooled rows are. Absorbing the 91 country intercepts does not free their degrees of freedom, so the residual count is 1,235 - 91 - 1 = 1,143 rather than 1,233; the reported standard error of 0.3163 carries that correction and gives t = -7.65. The rows labeled implied slope apply a classical-error calibration to the estimated baseline: with a region-specific horizon h the annualized growth rate carries 1/h of the level error, so the implied slope is the observed slope plus 100(1 - λ) times the mean reciprocal span, 0.0907, all divided by λ, and the standard error is the observed one divided by λ. That requires classical error in the baseline, no correlation with the growth innovation and a common reliability across horizons; it rescales the conditional standard error and does not add uncertainty about λ itself. The rows are a sensitivity exercise and not a correction, they carry no t statistic, and the implied slope is zero at a reliability of 0.97.

The implied-slope rows also need an assumption that common reliability across horizons does not by itself deliver, since conditional welfare variances can differ across horizons. A sufficient scenario is that baseline errors have mean zero, constant variance, and are independent of latent initial welfare, of true growth, of the follow-up error and of the observation horizon, so that the baseline error variance is unrelated to the reciprocal span and λ is a pooled reliability. The numerical scenario is unchanged under those restrictions, and, as the rows state, uncertainty about λ itself is not included.

B.6 Evidence on the magnitude of measurement-error bias

Section 4.2 quotes the two studies that estimate the noise share directly for settings of this kind. The wider evidence comes from different methods, error models and welfare aggregates. For Korea, Lee *et al.* (2017) identify time-varying measurement error separately within a consumption-dynamics model and simulate error-free transition matrices: a measured one-year poverty-exit rate of roughly 45% falls to between 26% and 31%. For Viet Nam, Glewwe (2012) instruments one error-ridden welfare measure with another, and measured mobility shrinks substantially.

Matched survey-administrative data for the United States confirm that both the size and the nonclassical structure of the error matter (Gottschalk and Huynh, 2010; Millimet and Parmeter, 2025). These studies agree that transitory classical error inflates measured movement, often substantially, but they do not share a design and the direction is not universal: time-invariant and mean-reverting components can offset or reverse it. A direction should be claimed only alongside an explicit error model.

On the structure of the error rather than its size, Lee (2022) estimates a framework nesting classical, mean-reverting and time-invariant components, finding classical error large enough to bias estimated income dynamics toward zero. Millimet and Parmeter (2025) show that the structure of misreporting in United States earnings data has itself changed across decades, by enough to alter measured inequality trends. Comparisons of dynamics across surveys, countries or decades may therefore partly reflect changes in the error process rather than in the underlying economy.

The inter-generational literature learned the same lesson earlier and in reverse. Its object is persistence rather than movement, so classical error in parental income biases the headline estimate downward: single-year proxies for lifetime status roughly halved early elasticity estimates until multi-year averages repaired the attenuation (Solon, 1992). The modern synthesis accordingly treats measurement choices as first-order determinants of reported mobility (Deutscher and Mazumder, 2023). The intra-generational literature for poorer countries has been slower to internalize the point.

B.7 Register-free inter-generational methods: mechanics and assumptions

Section 3.2 describes two families of method that construct inter-generational objects without linked parent-child records, both estimating tier-D4 quantities from tier-D1 and tier-D2 inputs. Two-sample methods project missing parental income on characteristics observed both for parents in an auxiliary sample and in the children's recall, then relate child outcomes to the projection. The condition the method needs is transportability: the relationship predicting parental income from those characteristics in the auxiliary sample must carry over to the target parents, which requires the two samples to represent the same population of parents. That is not the same as equal returns to characteristics in the parent and child generations. Where the projection is read as a two-sample instrumental-variable estimator, the characteristics must also be relevant for parental income and excluded from the child's outcome equation given the controls; accurate recall and transportability alone do not recover the intended persistence coefficient, and estimates are sensitive to the instrument set (Emran and Shilpi, 2021). Projected parental income is an estimate, so standard errors should propagate the first stage as well as the second.

Copula methods take the two marginal distributions as known and complete the joint distribution with an assumed dependence structure. The assumption supplies the missing association, not the whole of the answer: absolute mobility depends on both marginals and on their dependence, and the sensitivity of a given statistic to the copula is governed by how far the two marginal distributions overlap (Chetty *et al.*, 2017). Out-earning a parent is not intrinsically a tail event, so no general rule makes the middle robust and the tails fragile. Both families would benefit from the validation discipline synthetic panels have accumulated (Section 5.1).

B.8 The illustrative computation of Appendix Table A4

Appendix Table A4 computes the core reporting set of Table 5 for one country, from the VHLSS rounds of 2010, 2012 and 2014 and the Vietnamese regions of the SPID. The poverty line is the 2010 GSO–World Bank poverty line of 7,831.3 thousand dong per person per year, about 653 thousand per month, expressed here in January-2010 prices; the estimand is the household, so that line places 17.8% of households below it in 2010 under household expansion weights, which is full-sample household incidence, and 20.7% of people once household size is applied, which is the population incidence reported by that source. All VHLSS estimates use the household weight of the first wave of each linked pair, renormalized to the linked sample and not adjusted for attrition, except the Spearman rank correlation, which is unweighted; the SPID regions in row 9 are unweighted.

Row 1 reports a vulnerability line and the shares it defines, estimated by the second definition of Dang and Lanjouw (2017). Their vulnerability index is the share of households lying between the poverty line and a candidate line at the first date who are poor at the second, and the line is found by raising the candidate from the poverty line until that share falls to the stated index, here 10%. Estimation uses the 3,927 linked households of the 2010-2012 panel with first-wave weights, steps of a quarter of a percent of the poverty line, and a minimum of 500 households between the two lines. Nothing is assumed about the distribution of welfare and no covariate model enters. The line is then applied to the 2012 cross-section, whose welfare variable is already in January-2010 prices, so no deflation is required.

Three devices were compared before this one was chosen, and each trades assumptions against data. The variance-equation approach of Chaudhuri *et al.* (2002) needs only one cross-section, which is its advantage in D1-tier situations. But it reads the dispersion between similar households as the risk each of them faces over time. Its probability is not monotone in current welfare, so it does not define a line at all. The equivalent line it implies at a probability of 10% priced a realized risk of 2.9% when carried forward to 2012-2014. The probit of López-Calva and Ortiz-Juarez (2014) conditions on household characteristics, so it says which households are exposed as well as how many, but it needs a panel and a functional form. It also provides a higher estimate of 15.5% against the expected 10%. The line used here assumes the least and was the only one whose stated index survived the same test, at 10.2%. Its disadvantages are that it needs a panel, that it prices risk for a group rather than for a household, and that the range of index values the data support is bounded, since only 4.9% of the initially non-poor became poor between 2010 and 2012 and that figure caps the first of the two definitions.

But the estimates are consistent with earlier estimates for Viet Nam. Dang and Lanjouw (2017) report that 6% of the non-poor of 2006 were poor in 2008, and that a vulnerability index of 10% then placed the line 114% above the poverty line. Over 2010-2012 the corresponding entry rate is 4.9% and the same index places the line 85% above.

Rows 4a and 4b report the poverty transitions of Table 6 in Dang *et al.* (2019), for the 2012-2014 VHLSS pair: row 4a their actual-panel estimates on the panel component, row 4b their synthetic-panel estimates constructed from the cross-sectional component, with household heads aged 25 to 55 in the first round. That table is also the first row of Table A6, so the illustration and the validation synthesis rest on the same published comparison rather than on a separate calculation of our own. The remaining VHLSS rows of Table A4 use the 2010-2012 pair or all three rounds, as each row states.

Rows 3, 5 and 6 use the following conventions. Row 3 averages annualized quantile growth over ventiles, that is over the quantiles 0.05, 0.10, …, 0.95, separately below and above the median. Row 5 assigns quintiles with survey weights at each date. Row 6 reports both curves in the same unit, annualized log growth, so that the difference between them is one of estimand rather than of measurement: the anonymous curve is the annualized log growth of each decile cutpoint of the two cross-sections, and the non-anonymous curve is the weighted mean of household annualized log change, grouped by the household's decile in the initial round. The two curves still differ in sample and in aggregation as well as in anonymity, since one is computed on the two cross-sections at decile cut points and the other on the linked panel as a mean of household growth rates, so the gap between them is not a clean measure of reranking. Row 3 keeps compound annual quantile growth, the convention in which pro-poor growth comparisons are usually reported.

Row 8 reports the chronic share of the squared poverty gap as a ratio of weighted aggregates, not as an average of household-specific ratios. Write y(i,t) for real per capita consumption of household i in wave t, z for the common real poverty line, w(i) for the three-wave household weight and T = 3. The contemporaneous squared gap is $g(i,t) = [\max(0, 1 - y(i,t)/z)]^2$. The denominator is the weighted mean over households of the average of g(i,t) over the three waves. The numerator of the standard construction replaces the three waves by the household's arithmetic mean consumption, $[\max(0, 1 - \bar{y}(i)/z)]^2$ with $\bar{y}(i)$ the mean of y(i,t) over t, and the chronic share is the ratio of the two weighted sums, 0.70 here. Since the denominator is an aggregate, households with no poverty in any wave contribute zero to both sums and raise no division question. The variant substitutes the geometric mean, exp of the mean of log y(i,t), for $\bar{y}(i)$ in the same level-based expression and against the same line, which raises the share to 0.76; it is a sensitivity to how a household's central consumption is summarized, not a poverty measure defined on logs. Because the substitution is made inside a level-based expression, the variant does not inherit the usual properties of a chronic share and can in principle exceed one, which limits how far it should be read as an alternative decomposition.

The three-wave file holds 1,765 households, 25 more than the 1,740 common to both adjacent pair extracts, because seventeen are absent from the 2010-2012 extract and eight from the 2012-2014 extract while appearing in the corresponding cross-sections; the identifier-level sample-flow file that would explain each case is not yet released. Restricting to the common 1,740 moves the chronic share from 0.696 to 0.700, the mean-log variant from 0.763 to 0.766 and the permanent share from 0.7505 to 0.7531. That restriction is sensitivity evidence and not a substitute for the ledger, which would have to identify each of the twenty-five cases. Row 8 uses the 1,765 households observed in all three rounds.

The permanent and transitory components of log per capita expenditure are separated with the small-sample corrections a three-wave panel requires: the transitory variance is the weighted mean of the within-household sum of squared deviations divided by T minus one, and the permanent variance is the variance of household means less the transitory variance divided by T. Without those corrections the permanent share is overstated. The resulting share is 0.75. It is a random-intercept decomposition, in which the household effect is fixed over the three waves and the transitory component is stationary and serially uncorrelated, with no period effects removed; removing weighted wave means instead raises the share to about 0.77. It measures persistence in reported welfare, and it is not a reliability, because a household that consistently under-reports loads into the permanent component: persistent reporting error inflates 0.75 rather than deflating it.

Row 2 is not computed. The public VHLSS extract is a household file without an individual roster, so parents and children cannot be linked and the coresidence design of Section 3.2 cannot be implemented. We report the row as infeasible rather than substituting a different estimand.

Appendix C. The panel inventory: estimand, protocol, and sensitivity

Table A3 and Figure 3 rest on an inventory we assembled, and the count they report is a function of an inclusion rule as much as of the world. This appendix states the rule and the definitions, records what the audit could and could not establish, and reports how the headline numbers move when the rule changes. Three properties of the estimand deserve emphasis. It is a stock of capacity, not a statement about data an outside researcher can obtain today, so the running component and the access tier are reported separately. It is a property of the program rather than of any cohort within it. It turns, finally, on the welfare aggregate rather than on linkage alone.

What the inventory measures. The estimand is whether an economy has ever fielded, since 2000, a nationally or near-nationally representative survey that links the same households or individuals across at least two interviews and collects a household welfare aggregate at two or more of them. Rotating panels count; regional and urban-only panels, birth-cohort and aging panels, within-year revisit designs and phone panels without a welfare aggregate do not.

Definitions. Table A3's note defines the four depth measures the inventory records instead of a wave count: released rounds R, maximum linkable interviews L, how many of those carry a welfare aggregate W, and the elapsed years H of the longest single cohort. Cadence describes the selected longitudinal instrument rather than the program that carries it, so a biennial component inside an annual survey is coded biennial. R and L inherit the unit of the cadence and are not comparable across frequencies, so R enters no regression except among annually fielded programs and the L gradient is reported over the 82 programs fielded annually or less often. That group still mixes annual, biennial, triennial and irregular designs; restricting to the 52 annual programs gives a steeper gradient of 4.2 (1.3), so the pooled figure is the more cautious of the two.

H is a maximum design span rather than a typical trajectory: it can be attained by a small surviving cohort, and every documented case carries a dagger in Table A3. Where an economy has more than one qualifying program we list the one with the largest W, and where two differ in W only because of cadence we list the one with the larger H and record the others as alternates. Rel. records the survey or reference year of the latest released operation, which is why it can precede the end of the span.

Round counts carry a weaker evidentiary claim than the other fields. Released rounds are counted from release records and are exact where a release log exists. Fielded rounds are a lower bound implied by what has been released together with documented fielding dates, not an independently verified total; 27 rows whose two counts had been compiled on different clocks were reconciled to the invariant that a released round must have been fielded. Until a country-by-wave ledger is assembled, R should be read as a release count and F as a documented lower bound on fieldings.

Evidence and audit status. Every row carries a source and all 88 positive rows were audited, but the evidence is tiered rather than uniform. Every positive row carries an evidence URL, custodian, access date and a coded status on each of five inclusion dimensions: unit linkage, a welfare aggregate at two or more interviews, dates, coverage and access to a persistent key. Nineteen rows carry a medium rather than a high confidence flag and are asterisked in Table A3.

The status coding needs one further statement, because the rule and the count do not coincide. A program is coded ongoing when its longitudinal component has a documented fielding in 2023 or later, and 57 of the 88 rows meet that rule. Israel and Morocco are carried as ongoing although their most recent documented fieldings are 2022 and 2019, because neither producing agency documents the program as discontinued. We do not claim documented recent capacity for them: the defensible reading is 57 rows with a recent documented wave plus two whose status is uncertain. Table C1 reports the strict rule as a sensitivity, at 57 economies and a gradient of 12.2 (2.0) against 12.5 (2.0) on the coded 59, so nothing in Section 3.1 turns on the two exceptions.

Six rows name a weaker dimension rather than a weaker inclusion. Iran's rotation is documented in a sampling article rather than a producing-agency manual, and its operational key is the sampled address, so movers are not followed. Jamaica's panel is dwelling based, and the improved household matching its technical guide reports for 1995-1996 does not carry to the later spell. The Kyrgyz Republic's monthly panel repeats household income through a short phone module rather than a full consumption aggregate. Ukraine's survey represents the working-age population aged 15 to 72, the coverage question discussed below. Viet Nam's 2024 survey plan supports current fielding and welfare content, but its nine released longitudinal rounds are not reconstructed here. The Philippine row is the sixth, on an inferred rather than documented cross-round linkage. Of the 88 rows, 69 are coded high and 19 medium confidence.

Two fields could not always be established. Three distinct things are recorded separately and should not be conflated: whether the survey file can be obtained, which is the access tier; whether a released file carries a persistent household key; and how well documented that key status is. The linkable-file field records the key as present for 49 rows and absent for four, unverified for 33 and not coded for two. The broader access-key evidence field records the status as documented for 53 rows and unverified for 35, the 53 comprising the 49 present and the four absent. The access tier records whether the file can be obtained and nothing else, so a freely downloadable file with no linkage variable stays at tier A or B and its missing key is recorded in the key field, as for Kenya. Rows unverified on this dimension are counted in the stock but excluded from any statement about what a researcher can obtain. The Philippine row exercises both fields at once: the Philippine Statistics Authority documents two visits within each Family Income and Expenditure Survey round, but not re-interviewing across rounds, and the cross-round linkage used in the literature rests on a bespoke extract rather than a released file. Its access tier is nonetheless B, because the agency supplies microdata free of charge on a data-request form.

The inventory was cross-checked in August 2026 against the World Bank Microdata Library, the cited source for fourteen panel A rows, and against the Atlas of Longitudinal Datasets, whose broader rule admits birth-cohort, aging and clinical studies that ours excludes. The cross-check confirmed program identity and span for the rows those catalogs cover. Later audits accepted three additions, Iran, Jamaica and Paraguay; declined Mauritius, whose design documents only a two-quarter recontact; declined the deletion of the Philippines; declined two substitutions under the cadence tie-break, since each would have shortened H while raising W; and reinstated Uzbekistan, whose listening survey collects household income from the first monthly round. Each adjudication is documented.

Two coding decisions are worth stating because a reader may prefer the other choice. Ukraine is retained as national although its survey represents the population aged 15 to 72: the restriction is on respondent age rather than on geographic or sectoral coverage, and the welfare aggregate is collected for the whole household. Recoding it as near-national would leave 79 national panels, income-group counts of 5, 15, 18 and 41 and a slope of 7.9 (2.4), which changes no conclusion. The nationally representative labor-market panel surveys of the Economic Research Forum family remain an open class, because their earnings and enterprise-income modules may or may not satisfy the welfare-aggregate requirement.

A structural comparison with the Atlas is instructive because the two count different things. The Atlas identifies longitudinal data of some kind in 146 countries, against the 88 economies in which our rule finds a nationally representative household welfare panel (Arseneault *et al.*, 2025). The gap is not a discrepancy: a birth cohort or a biobank makes a country visible in the Atlas without giving it the capacity to measure poverty mobility, and the Atlas prioritizes samples above 8,000 at inception, ages 14 to 30, and mental health content, none of which a rotating panel attached to a household budget survey is designed to satisfy. What the inventory adds is a complete denominator: it records both the identified qualifying programs and the economies for which the search found no program meeting the stated rule.

Sensitivity of the counts and of Figure 3. Table C1 reports how the count and the extensive-margin gradient move as the rule changes, and Table C2 capacity by income group under the base rule. The scarcity result is robust: no rule we consider brings the share of economies with usable panel capacity near the share that would make direct measurement the norm. The tightest rule, requiring both a running component and an obtainable survey file, leaves twenty-one economies of 218, and requiring a documented persistent household key as well leaves fourteen.

The income gradient is not robust, and its instability is informative. On the stock it is 7.2 points per unit of log GNI per capita (2.4); on the running component 12.5 (2.0); and on accessible survey files the point estimate becomes negative, -4.1 (2.0), because high-income panels are predominantly gated behind approved-researcher or secure-facility conditions while LSMS-ISA and comparable operations in poorer countries are released under registration.

Requiring a documented linkage key as well leaves 20 programs and a slope of -1.5 (1.6), and requiring a running component too leaves fourteen. The file-access-only gradient is modestly negative under this coding, with a two-sided p-value of about 0.04, but it is statistically indistinguishable from zero once a documented persistent key is also required. What the evidence supports is therefore the disappearance of the positive historical-stock gradient once usable access is required, rather than a negative gradient robust to the definition of usable access.

The depth gradients of Figure 3, panel b, are computed over qualifying economies only and move little under the re-audit, to 2.5 (0.6) in linked interviews and 1.7 (0.8) in elapsed span. The interview gradient is a standardized effect of 0.43 and persists in a median regression; the elapsed-span gradient rests on a few unusually long programs, explains about 7% of the variance and is approximately flat at -0.2 in a median regression, which is why Section 3.1 reports the median spans rather than the mean relationship. The gradient in released rounds among the 52 annually fielded programs, 4.3 (1.0), is the statistic most exposed to release-count updates and to which programs count as annually fielded: dropping the United States, whose survey moved from an annual to a biennial cadence during its span, gives 3.8 (0.9).

Income groups are the World Bank FY2026 classification, frozen deliberately: the inventory was audited against documentation available in August 2026, and the regressions use 2019 Atlas GNI per capita as a pre-pandemic anchor. The grouping therefore enters the tabulations only, and re-basing it to FY2027 moves the shares by one to two points without changing the ordering or any conclusion.

Population coverage. Using 2019 population, economies that have ever fielded a qualifying design account for 80% of the summed population of the 218 classified economies and 66% of that of the low- and lower-middle-income ones. This is the share of people living in an economy with such a design, not the share represented in a panel, which is far smaller.

Table C1. How the count and the extensive-margin gradient of Figure 3 move under alternative inclusion rules

Inclusion rule	Economies	By income group (L; LM; UM; H)	Gradient	Robust s.e.
Base rule, as coded in Table A3	88	6; 16; 23; 43	7.2	(2.4)
Requiring a longitudinal component coded ongoing	59	1; 6; 13; 39	12.5	(2.0)
Sensitivity: a documented longitudinal wave in 2023 or later (strict rule)	57	1; 5; 13; 38	12.2	(2.0)
Requiring at least three comparable welfare waves (W >= 3; W inherits cadence, so this favors high-frequency programs)	75	4; 10; 19; 42	10.4	(2.2)
Requiring a strictly national universe	80	5; 15; 19; 41	7.7	(2.4)
Requiring open or registration access (tier A or B)	43	5; 14; 14; 10	−4.1	(2.0)
Requiring both a running component and tier A or B	21	0; 5; 7; 9	2.1	(1.4)
Requiring open or registration access and a documented persistent key	20	3; 7; 4; 6	−1.5	(1.6)
Requiring a running component, tier A or B and a documented key	14	0; 5; 3; 6	1.0	(1.2)

Note: The gradient is the coefficient on log GNI per capita in 2019 in a linear probability model over the 202 economies with a published value, in percentage points, with heteroskedasticity-robust standard errors, and is the quantity plotted in Figure 3, panel a.

Table C2. Panel capacity by income group, under the base inclusion rule

Income group	Economies	Ever qualified	Coded ongoing	Access tier A or B	Both	Median W	Median H
Low income	26	6 (23%)	1 (4%)	5 (19%)	0	3.0	5.0
Lower middle income	50	16 (32%)	6 (12%)	14 (28%)	5	3.0	6.0
Upper middle income	55	23 (42%)	13 (24%)	14 (25%)	7	4.0	3.0
High income	87	43 (49%)	39 (45%)	10 (11%)	9	4.0	3.0
All economies	218	88 (40%)	59 (27%)	43 (20%)	21	4.0	3.8

Note: W is the number of linked interviews carrying a household welfare aggregate and H the elapsed years of the longest single cohort. Ever qualified counts economies that have fielded a qualifying design at any point since 2000, and Coded ongoing those still operating under the protocol of Appendix C: 57 rows with a documented wave in 2023 or later, plus two whose status is uncertain. Both requires a row coded ongoing and access tier A or B, defined in Table A3's note. Percentages are over all economies in the row; the medians of W and H are over the qualifying economies only. Computed from the inventory behind Table A3, and therefore conditional on the rule stated above.