Estimating Long Run Welfare Outcome in Rotating Panel with Grouped Fixed Effects: Application to Poverty Dynamics in Peru*

Hongdi Zhao† Seungmin Lee‡

Job Market Paper

September 3, 2026

Abstract

Household welfare dynamics are often difficult to investigate due to a lack of long-term panel data. Existing methods, such as pseudo-panel and synthetic panel, are widely used approaches based on repeated cross-sectional designs, but they impose ex ante assumptions that often do not hold. This paper introduces an empirical method to estimate welfare dynamics using a rotating panel design. The paper employs the grouped fixed effects (GFE) estimator to exploit within-household variation from short-term observations and to estimate longer-term welfare and its dynamics using time-varying group-level heterogeneity, thereby relaxing a few assumptions in the previous literature. The paper applied this method to Peru's National Household Survey on Living Conditions and Poverty (ENAHO) from 2007 to 2019 and examined long-term poverty and its dynamics during this period. The new method estimates household-level poverty status and its transitions reasonably well, suggesting it as a new approach to studying household welfare dynamics at the group level and providing an interpretable grouping structure that supports richer descriptions of poverty persistence and mobility.

JEL Classification: I30, I32, J60, O15

Keywords: Peru, poverty dynamics, fixed effects, development economics, poverty transition, clustering

*We thank seminar participants at Cornell University and the University of Notre Dame, and attendees at the International Conference of Empirical Economics 2026 for helpful comments. We gratefully acknowledge Chris Barrett and José Luis Montiel Olea for his valuable feedback. We also thank Cédric Schneider and Dean Jolliffe for their feedback. This material is based upon work supported in part by the National Science Foundation under Grant Number 2242507. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Comments and suggestions are welcome and may be emailed to hz399@cornell.edu and slee76@nd.edu

†Hongdi Zhao is an associate in Analysis Group.

‡Seungmin Lee is a postdoctoral research associate at Pulte Institute for Global Development in the University of Notre Dame.

1 Introduction

Individual and household well-being change over time, implying that poverty evolves as well. While some poor households are only transiently poor, quickly recovering back to non-poor status, many other poor households are chronically poor, trapped in poverty over the course of their lives. Understanding poverty dynamics - change in status and duration of status - is crucial in policy design for multiple reasons (Baulch and Hoddinott, 2000). First, it could identify a vulnerable population that remains consistently poor over time, providing an opportunity to understand why they are chronically poor. For instance, if we find that households in one region are more chronically poor than those in other regions, it could be due to a lack of public goods and infrastructure there. Second, different policies are needed to address the distinct types of poverty, since the causes of chronic and transient poverty differ. Social assistance programs, unemployment insurance, and cash transfers are effective against short-term poverty, often caused by temporary hardship, whereas productive asset accumulation, high-quality service provision, and formalization are needed to address chronic poverty, which stems from a lack of productive assets and infrastructure. Third, the duration of poverty itself could exacerbate a household's poverty status. The lack of productive assets can lead to chronic poverty, but chronically poor households may reduce their assets or investment in capital accumulation as a coping strategy, such as selling their livestock or withdrawing their children from school, trapping themselves in poverty for longer periods (Attanasio and Székely, 2001; Carter and Barrett, 2006). However, studying household-level poverty dynamics has often been hindered by the lack of reliable panel data that observe the same households over time. Household-level panel data are often unavailable or have small sample sizes due to high implementation and maintenance costs.

Several econometric and statistical approaches have been developed to estimate dynamics under such data limitations. Pseudo-panel (Deaton, 1985) is an early attempt to overcome data limitations by grouping micro-level units (individuals and households) into cohorts that share common characteristics (e.g., birth cohort, village, or district) and tracking cohort-level dynamics using repeated cross-sectional data. The synthetic panel is a more recent approach that enables estimation of micro-level welfare outcome dynamics using only two rounds of cross-sectional data (Dang et al., 2014). Based on a few assumptions on the population and its parameters, the synthetic panel estimates the boundaries (upper- and lower-bound) of poverty mobility (e.g., entry into and exit from poverty) where true mobility lies in between. The literature has shown that bounded estimates almost always capture true poverty mobility across both developed and developing countries (Dang et al., 2014; Cruces et al., 2015; Hérault and Jenkins, 2019). The synthetic panel method has been extended to generate point-level estimates with additional parametric assumptions Dang and Lanjouw (2023), and to the Least Absolute Shrinkage and Selection Operator - Predictive Mean Matching (LASSO-PMM) method, a semi-parametric method that does not require the parametric assumptions from Dang and Lanjouw (2023) (Lucchetti et al., 2025). D'Alberto et al. (2025) introduces a scenario-based approach that accounts for temporal dependencies in the data using a Bayesian framework and a matching

method.

While researchers have developed these methods for short-term dynamics using repeated cross-sectional data, there have been limitations in the current literature. First, these methods often rely on ex-ante assumptions that may not always hold. The pseudo-panel approach assumes cohort-level heterogeneity is time-invariant, which is less likely to hold in the long-run. The synthetic panel assumes that the underlying population in two different periods is the same, and its point estimates further assume that the error terms in the welfare model are bivariately normally distributed, a fundamental assumption that was quantitatively rejected in every country the authors tested (Dang and Lanjouw, 2023). Such violations may lead to welfare status and transition status being biased and/or imprecisely estimated. Furthermore, both pseudo and synthetic panel methods are sensitive to how cohorts are defined, for which there is no solid guidance.

In this paper, we introduce a new method for estimating micro-level welfare status and its dynamics, addressing these limitations. We study longer-term status by employing the grouped fixed effects (GFE) estimator (Bonhomme and Manresa, 2015) (hereafter, BM2015), in which heterogeneity is time-varying at the group level in rotating panel data, with micro-level units observed over short periods. We apply this method to study long-run poverty dynamics in Peru from 2007 to 2019 using the GFE estimator. By employing the GFE framework on Peru's National Household Survey on Living Conditions and Poverty (ENAHO), which uses a rotating panel design with a 5-year rolling sample, we leverage its structure by combining the strengths of short-panel identification with a data-driven grouping approach. We classify households into latent groups with distinct time patterns, and then leverage these estimated group-time components to recover poverty transition measures over horizons that exceed the observed panel length.

We find that our new method reasonably well tracks unobserved poverty status and its transitions. In a one-step-ahead validation that holds out each household's final observed year, the new method predicted households' unobserved poverty status with an accuracy of 83-87% and poverty transitions with error rates similar to the existing method, but with fewer assumptions. We further show that our main findings are robust to alternative sample definitions: removing the head of household's age restriction of the sample yields very similar optimal number of group selection and group structure, while restricting to the smaller subset of households observed for at least five rounds of survey leaves the optimal number of latent groups unchanged but makes the differences in group trajectories less stable with different sets of model specifications, highlighting the importance of adequate group–year support in rotating panels. We also test the model's ability to predict medium-run poverty using households that remained in the data for at least 4 years, with the last 2 years as the test set and the prior years as the training set. We show that the model can predict the two years of test data with over 80% accuracy.

This paper makes the following contributions. First, we provide a framework for studying poverty dynamics in rotating panel settings that bridges the gap between true panels and repeated cross-sections. Using observed household overlap to estimate latent groups and their

time profiles, we apply the GFE estimator, leveraging information that other conventional methods lack. Second, we show how GFE can be adapted to recover poverty transition measures over horizons longer than the observed panel lengths, offering a practical way to study medium- and longer-run mobility when long panels are unavailable. Third, to the best of our knowledge, this paper is the first to adopt GFE estimator to study poverty dynamics, providing a data-driven alternative to cohort-based synthetic panels that reduces sensitivity to arbitrary cohort definitions and yields more stable estimates of poverty dynamics in settings where only short panels are observed.

The remainder of the paper is organized as follows. Section 2 lays out the empirical framework and discussed how GFE is identified and estimated in rotating panels. Section 3 describes the ENAHO rotating panel structures and our implementation of using GFE to estimate poverty transition on ENAHO data measures over both short and longer horizons. Section 4 presents the main results on poverty persistent and mobility. Section 5 shows robustness checks to alternative sample restrictions and section 6 concludes the paper with discussion on it's policy relevance.

2 Empirical Strategy

2.1 Rotating Panels

Rotating panel design is where survey units (e.g, households, individuals) are interviewed for a limited number of consecutive years and then rotated out, while new households enter the sample each year. As a result, the dataset is neither a long true panel nor a sequence of independent cross-sectional data, but rather a hybrid design that balances sample size and panel structure at a lower cost than a true panel.[1] Due to its hybrid nature, the rotating panel design has been widely adopted internationally, including Brazil, Mexico, Peru and Vietnam.[2] Since partial units overlap over a short period, typically from two to six years, they provide within-unit information that can be very useful in estimating unit-level welfare dynamics. Since many existing poverty estimation methods are designed for purely repeated cross-sectional data, they treat each survey round as an independent draw from a stable population and recover mobility only from cohort-level information (i.e., the synthetic panel estimates dynamics via moments and assumptions about the joint distribution of unobservables). Therefore, if we apply either method to rotating panels, we necessarily discard observed within-household information and fail to utilize that overlap to discipline the correlation in unobserved shocks that drive mobility and persistence in the poverty transition measure. Overall, existing panel implementations may yield wide bounds or unstable point estimates precisely where short rotating panels provide an opportunity to do better.

Although rotating panels combine the broad coverage of repeated cross-sections with some within-household follow-up, they have limitations for studying long-term poverty dynamics.

[1] We present more detailed information about the ENAHO data and its summary statistics in Section 3.1

[2] Examples of rotating panel surveys are shown in Appendix Table A1.

First, the sample size of overlapping units is often small due to administrative costs, resulting in large standard errors in estimation (Van den Brakel and Krieg, 2015). Second, estimates can be biased if the units newly surveyed in a given period have systematically different characteristics from those surveyed in the same period, a phenomenon known as "rotating panel bias" (Bailar, 1975). Third, most rotating panel data track units for only a short period, primarily due to administrative costs, which can limit the estimation of dynamics. For example, one way to estimate dynamics is to use a unit-level fixed-effects (FE) model to capture micro-level unobservables, decompose welfare transitions into observables and unobservables, and determine which one mainly drives welfare transitions. However, estimating unit-level FE can be imprecise with modest T, and such effects lack economically interpretable implications. Even if we precisely estimate time-invariant fixed effects, they may not remain time-invariant over a longer period (Hill et al., 2020; Millimet and Bellemare, 2025)

The partial overlap in rotating panel data is central to our empirical strategy. Even though each household is followed only for a few years, repeated observations for a subset of households provide direct information on persistence in expenditures and on time-varying shocks affecting similar households. At the same time, the continuing arrival of new households preserves the representativeness of each annual cross-section and provides rich variation in covariates and outcomes over time. We exploit this rotating panel structure in the next subsection by modeling households as belonging to a finite number of latent groups that share common time patterns in unobserved heterogeneity, and we estimate these group-time profiles using all available household-year observations.

2.2 Grouped Fixed Effect (GFE) Estimator

BM2015 estimates unobserved heterogeneity at the group level: households are assigned to a predetermined number of groups through multi-stage unsupervised clustering, and the heterogeneity is time-varying. The GFE estimator provides a flexible way to model unobserved heterogeneity when full panel structures are not available. The key idea is that households can be classified into a finite number of latent "types," each of which follows its own time pattern in the unobserved components of welfare. Rather than assuming a single household fixed effect or imposing a parametric distribution of unobservables, the GFE estimator approximates continuous heterogeneity with a discrete set of groups and estimates both group membership and group-specific time profile directly from the data.

Compared with other estimators that model unobserved heterogeneity beyond classic two-way fixed effects (TWFE), such as correlated random effects (Mundlak, 1978) and interacted fixed effects (IFE) (Bai, 2009), BM2015 has the following advantage in our study setting. First, BM2015 allows heterogeneity to be time-varying, which is essential for investigating long-term welfare, where micro-level heterogeneity is likely to vary over time. Second, BM2015 models heterogeneity at the group level, which can be precisely estimated and provides meaningful implications compared to micro-level FE, which is often imprecisely estimated and has little implication. Third, it preserves intra-unit variation in welfare and covariates, allowing us to

estimate the association between welfare and covariates with limited within-unit variation (e.g., in educational attainment). Individual units are assigned to a group through clustering; units with similar trends in heterogeneity (i.e., the outcome net of covariates) are grouped together. Therefore, each group exhibits a distinct trend in heterogeneity over time. Group assignments and model parameter estimation are conducted through either an iterative algorithm or a variable neighborhood search (VNS) algorithm. That is, the GFE estimator uses within-unit information by assigning units to groups in a data-driven way, addressing issues of sensitivity to the cohort definition and the loss of within-unit information from pseudo- and synthetic-panel methods.[3]

BM2015 has been widely adopted to study causal effects on prediction. BM2015 applied the GFE method to examine the effects of income on democracy and found that, on average, income has little effect on democracy, though some countries democratized rapidly at different times. Oberlander et al. (2017) examined the effects of globalization on diet quality by grouping countries with similar trends in unobserved heterogeneity path into six groups. Janys and Siflinger (2024) estimated the causal effects of abortion on mental health by capturing the unobserved decision-making process driving risky behavior. Aparicio-Pérez and Ripollés (2025) investigated the impact of natural resources on economic growth by capturing the unobserved institutional quality that worked either positively or negatively on different groups of countries. In terms of predictive performance, the GFE estimator outperformed the standard fixed-effects estimator for predicting a rare event, such as a banking crisis (Pigini et al., 2025). In fact, GFE is one of many examples in which data-driven clustering outperforms predetermined grouping for forecasting future outcomes (Oh and Patton, 2023). Since pseudo-panel and synthetic panel group units are based on predetermined characteristics, we hypothesize that the GFE estimator can better estimate poverty status than either of these.

The GFE estimator is particularly useful in rotating panels, where households are observed only for a limited number of years, so estimating a separate household fixed effect for each unit is noisy and uninformative and does not provide a basis for extrapolating beyond the observed years. The GFE estimator instead pools information across households that exhibit similar time patterns, using the overlap in the rotating panel to learn both the latent group and the evolution of group-level unobserved components. In this way, rotating panel overlap directly contributes to identifying latent types and their dynamics. In the remainder of this section, we follow the GFE algorithm in BM2015 and adapt it to the ENAHO rotating panel structure data, providing a detailed description of our estimation procedure. We then explain how the estimated group-time paths are used to construct poverty transition measures over longer horizons.

In this study, we model the inverse Hyperbolic Sine (IHS)- transformed total household expenditure per capita as a function of observed characteristics and a group-time component

[3]Bonhomme et al. (2022) extended BM2015 by relaxing certain assumptions, such as the discreteness of heterogeneity. In this paper, we use the framework from BM2015 because it provides a simple, transparent method for recovering distinct welfare trajectories from short panel overlap, which is the central objective of this paper. Applying Bonhomme et al. (2022) to study poverty dynamics with the ENAHO data could be a potential extension to this paper.

that captures time-varying unobserved heterogeneity common to households of the same latent type.[4] Households are assigned to one of the G groups, and each group has its own sequence of year-specific intercepts. This structure captures persistent differences across households (through group membership) while allowing the unobserved component to evolve over time (through group-time effects), which is particularly useful in contexts where macro shocks and structural change can shift the welfare distribution.

2.2.1 Model setup

We model household welfare using a GFE estimator that allows for time-varying unobserved heterogeneity shared by households of the same latent type. Specifically, we estimate an expenditure equation of the form:

$$y_{it} = x'_{it}\theta + \alpha_{g_i t} + \mu_{p_i} + \varepsilon_{it} \tag{1}$$

where y_{it} is a household welfare measure, such as the (log or IHS-transformed) per capita expenditure of household i at time t; $x_{it} \in R^K$ is a vector of observed time-varying and time-invariant characteristics; $\theta \in R^K$ is a common vector of coefficients across all households; $\alpha_{g_i t}$ is a group-time fixed effect, capturing time-varying unobserved heterogeneity for group g_i that household i belongs to; μ_{p_i} is the time-invariant location fixed effects for individual i in location p, such as province fixed effect, that nets out persistent location difference; and ε_{it} is an idiosyncratic error term.

We observe units $i = 1, \cdots, N$ living in $p = 1, \cdots, L$ over periods $t = 1, \cdots, T$ with outcome y_{it} and regressors $x_{it} \in R^K$. Given Eq. (1), let $d_{it} \in \{0, 1\}$ indicates observation (ENAHO is a rotating panel at the household and year level, so $d_{it} = 0$ if a household is not surveyed in year t). It is assumed that x_{it} and ε_{it} are contemporaneously uncorrelated given $d_{it} = 1$.

Each unit belongs to one latent group $g_i \in \{1, \cdots, G\}$. $\theta \in \Theta$ (subset of R^K parameter space) are common slopes across units, and units in the same group share the same time profile $\alpha_{gt} \in \mathcal{A}$ (e.g., all i such that $g_i = 1$ share the profile α_{1t}. $\mathcal{A}$ is subset of $R^{G \times T}$ parameter space), and units in the same location p share the same location-specific heterogeneity $\mu_p \in \Lambda$ (e.g., all i such that $p_i = 1$ share the heterogeneity μ_1. Λ is subset of R^L parameter space) . The covariates vector x_{it} may include strictly exogenous regressors and lagged outcomes. x_{it} and $\alpha_{g_i t}$ are allowed to be arbitrarily correlated. We denote α as the set of all α_{gt}'s, γ as the set of all g_i's, and μ as the set of all μ_{p_i}. $\gamma \in \Gamma_g$ denotes a particular grouping of the N units, where Γ_g is the set of all groupings of $\{1, \cdots, N\}$ into at most G groups.

All members of group g share the same unobserved trajectory. Neither the group membership g_i nor the $\alpha_{g_i t}$ terms are observed prior. They are estimated jointly from the data.

In our setting, local conditions such as geography, local prices, access to markets, and infrastructure can persistently shift expenditure levels and may otherwise be picked up by latent groups. To separate these time-invariant location differences from the group-specific

[4] We use IHS-transformed expenditure to handle extremely high expenditure values, smoothing distribution.

time profile, we include a province fixed effect μ_{p_i} for each household's province of residence. This inclusion of time-invariant heterogeneity is the first extension of the GFE framework described in BM2015, in which we use location-specific heterogeneity rather than unit-specific heterogeneity. Conceptually, the model decomposes expenditures into (i) observed covariates, (ii) a time-invariant location component, and (iii) a group-by-time component that captures shared time-varying unobserved heterogeneity among households of the same latent type. We discuss the option of adding province fixed effects instead of department (region) or districts fixed effects in more details in Section 3.2.

2.2.2 Estimation procedure

We estimate $(\theta, \alpha, \gamma, \mu)$ in Eq. (1) using the GFE Algorithm 2 (Variable Neighborhood Search-VNS + Local Search) of BM2015, adapted to the unbalanced rotating panel structure of ENAHO.[5]

The GFE estimator chooses $(\theta, \alpha, \gamma, \mu)$ to minimize the sum of squared residuals over observed household year observations:

$$(\hat{\theta}, \hat{\alpha}, \hat{\gamma}, \hat{\mu}) = \underset{(\theta,\alpha,\gamma,\mu)\in\Theta\times\mathcal{A}\times\Gamma_G\times\Lambda}{\text{argmin}} \sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}(y_{it} - x'_{it}\theta - \alpha_{g_i t} - \mu_{p_i})^2 \tag{2}$$

where the minimum is taken over all possible groupings $\gamma = \{g_1, \cdots, g_N\}$ of N households into G groups, common parameters θ, group-specific time effects $\alpha = \{\alpha_{gt}\}$, and location-specific effect $\mu = \{\mu_p\}$.

Directly minimizing this objective over all partitions γ is computationally infeasible when N is large. Following BM2015, we therefore use an iterative algorithm that alternates between (i) updating the group-time effects, common slopes, and location effects given group assignments, and (ii) updating group assignments given the current parameter estimates. In our setting, all updates are computed using only the observed household-year cells (those with $d_{it} = 1$).

In each iteration, given a current grouping γ, we estimate θ, α and μ by least squares on the pooled household-year sample with group-time indicators and location indicators. Given (θ, α, μ), each household is assigned to the group that yields the smallest sum of squared residuals over the years in which the household is observed. Because the objective function is non-convex, the VNS + local search algorithm applies a multi-start strategy to reduce sensitivity to local minima.[6] We repeat this procedure until the objective function no longer decreases or the maximum number of iterations is reached.

[5] To better adapt the algorithm to our setting, we do not use the replication code from BM2015. Instead, we implement the GFE algorithm in Python and adapt key steps to handle the partial household overlap and missing household-year cells implied by the rotating panel. Our code also allows us to add the time-invariant location fixed effects into the GFE estimator. Our implementation is written to be flexible and can be applied to other household surveys with similar rotating-panel designs with or without locational fixed effects.

[6] The objective function in (2) is non-convex because it is a joint optimization over discrete group assignments and continuous parameters. Conditional on any given γ, the objective function becomes an ordinary least-squares problem in (θ, α, μ) with a unique, closed-form minimizer. However, since group assignment is discrete, joint optimization over the G^N possible group assignments is combinatorial, leading to multiple local minima and rendering the overall objective non-convex. For the multi-start, each start with a different seed.

In practice, we implement the OLS update step by estimating θ, group-by-time effects α_{gt}, and the province-fixed effects μ_p using the observed household-year observations while absorbing province fixed effects. This implementation nets out time-invariant location differences so that group assignment is driven by within-province variation and by differences in households' time profiles.

The step-by-step detailed explanation of the algorithm is described below.

- **Step 1:** Let $(\theta^{(0)}, \alpha^{(0)}, \mu^{(0)}) \in \Theta \times \mathcal{A} \times \Lambda$ be an initial value, G be the number of groups pre-determined by the researcher, and set the tuning parameters for later steps.
  - **Initializing $\theta^{(0)}$ and $\mu^{(0)}$**: We guess the initial $\theta^{(0)}$ and $\mu^{(0)}$ by regressing y_{it} on x_{it} with pooled data with province indicators using only rows with $d_{it} = 1$.[7]
  - **Initializing $\alpha^{(0)}$** (group-time fixed effects): Since group membership g_i is not yet known, $\alpha^{(0)}$ must be initialized without groups. We initialize $\alpha^{(0)}$ as common time fixed effects estimated from the pooled observed household-year data after partialling out observed covariates and province fixed effects (using only observations with $d_{it} = 1$. We then set the initial time effect as the mean residual in each year:

$$\alpha_t^{(0)} = \frac{\sum_i d_{it} \left(y_{it} - x_{it}'\theta^{(0)} - \mu_{p_i}^{(0)} \right)}{\sum_i d_{it}}, \tag{3}$$

    where $\mu_{p(i)}^{(0)}$ denotes the estimated province fixed effect for household $i's$ province. Equivalently, $\alpha_t^{(0)}$ is the year-t average of the residualized outcome $y_{it} - x_{it}'\theta^{(0)} - \mu_{p_i}^{(0)}$ over observed units. Then:

$$\alpha_{g,t}^{(0)} = \alpha_t^{(0)} \quad \forall g \in \{1, \dots, G\} \tag{4}$$

    Such that all the groups have identical time effects.
  - **Initial grouping γ_{init}**: Once we have $(\theta^{(0)}, \alpha^{(0)}, \mu^{(0)})$, we can obtain the initial group assignment γ_{init} that solves the following minimization problem:

$$g_i^{(0)} = \underset{g \in \{1,\dots,G\}}{\operatorname{argmin}} \sum_{t=1}^{T} d_{it} \left(y_{it} - x_{it}'\theta^{(0)} - \alpha_{gt}^{(0)} - \mu_{p_i}^{(0)} \right)^2 \tag{5}$$

  - Setting up tuning parameters:
    - `n_starts`: the number of starting values we initialize. Since we follow the entire steps for each starting value, the entire steps will be repeated `n_starts` times.
    - `neighmax`: the maximum neighborhood size controlling how many neighborhood jumps we allow in the following Step 3 before stopping the algorithm.[8]

[7] Using OLS estimate as the initial guess is more stable than a random vector because we do not need to do a wild random guess. As random small-noise initializations are also allowed in BM2015 algorithm, We could use a random small noise vector from a normal distribution, or start with a zero vector.

[8] An neighborhood jump means that if we pick n units, we move them to random new groups and re-run one

* `max_local_iters`: the maximum number of local iterations that controls how many times we want to repeat the iterations to update $(\hat{\theta}, \hat{\alpha}, \hat{\gamma}, \hat{\mu})$ in Step 4.
* `itermax`: the maximum number of VNS iterations that controls how many times we want to repeat the entire VNS sequence[9]. The VNS iteration counter is initialized at $j = 0$, and increased after each VNS cycle until j reaches `itermax`

- **Step 2:** Initialize neighborhood size.

  Set the neighborhood size $n = 1$. $n = 1$ means that we start by exploring solutions that differ from the current best grouping by reassigning only one randomly chosen unit.

- **Step 3:** Neighborhood jump ("shake").

  Given the current grouping γ^*, randomly select n units and relocate each selected unit to a (uniformly) randomly selected group, producing a new γ'. In our rotating-panel setting, our objective is always computed using only observed (i, t) pairs with $d_{it} = 1$, consistent with the unbalanced-panel formulation from BM2015.

  After obtaining γ', re-estimate (θ, α, μ) by OLS with group-by-time indicators, location indicators, and covariates conditional on γ', using only observations with $d_{it} = 1$ to obtain new parameter values (θ', α', μ').

$$(\theta', \alpha', \mu') = \text{ argmin }_{(\theta,\alpha,\mu)\in\Theta\times\mathcal{A}\times\Lambda} \sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}(y_{it} - x'_{it}\theta - \alpha_{g_i t} - \mu_{p_i})^2 \tag{6}$$

- **Step 4:** Capped refinement based on Algorithm 1.

  Following the logic of Step 4 in BM2015, we apply a limited refinement procedure after the neighborhood jump rather than iterating Algorithm 1 to full convergence. In the implementation, this refinement consists of updating (θ, α, μ) once conditional on the shaken grouping γ', followed by a capped sequence of greedy reassignment passes and a final OLS update. The cap is controlled by `max_local_iters`.

  Our implementation follows the structure of Algorithm 2 in BM2015, but adapts the refinement step to the rotating-panel setting by using a capped practical variant rather than running Algorithm 1 to full convergence after each neighborhood jump.

- **Step 5:** Greedy local search over single-unit reassignments.

  Starting from the grouping $\gamma' = \{g'_1, \ldots, g'_N\}$, perform a *local search* that systematically checks single unit moves $i : g'_i \mapsto g \in \{1, \ldots, G\}$ for all $g \neq g'_i$, and accepts a move whenever it decreases the objective function, holding (θ', α', μ') fixed. This local search

update steps of Algorithm 1 in BM2015 to see if the objective improves. As n increases from 1 to `neighmax`, we explore "larger" perturbations of the grouping. Larger values of the `neighmax` would slower the program but provide more global exploration. We set the `neighmax` to 5 to choose the best group G.

[9] We set the `itermax` = 10, the same order of magnitude used as the BM2015.

is run for at most `max_local_iters` passes and and stop when no further single-unit reassignment reduces the objective function. Denote the resulting locally optimal grouping as $\gamma'' = \{g''_1, \ldots, g''_N\}$. Given the updated grouping γ'', re-estimate

$$(\theta'', \alpha'', \mu'') = \underset{(\theta,\alpha,\mu)\in\Theta\times\mathcal{A}\times\Lambda}{\text{argmin}} \sum_{i=1}^{N}\sum_{t=1}^{T} d_{it}(y_{it} - x'_{it}\theta - \alpha_{g''_i t} - \mu_{p_i})^2 \tag{7}$$

As noted by BM2015, "local search allows to get around local minima that are close to each other, whereas random jumps aim at efficiently exploring the objective function while avoiding to get trapped in a valley."

In our rotating panel set up, each candidate's move evaluates the following objective function:

$$\text{loss}_i(g) = \sum_{t=1}^{T} d_{it}\left(y_{it} - x'_{it}\theta - \alpha_{gt} - \mu_{p_i}\right)^2 \tag{8}$$

so missing households' information in a given year would not affect the reassignment decision.

- **Step 6:** Acceptance rule.

  If the objective function under γ'' is lower than the objective function under the current base γ^*, set $\gamma^* = \gamma''$ and return to **Step 2** (resetting $n = 1$ to restart the neighborhood exploration around the improved solution). Otherwise, set $n = n + 1$ and continue to Step 7.

- **Step 7:** Increase neighborhood size up to `neighmax`.

  If $n \leq$ `neighmax`, go back to **Step 3** and perform another neighborhood jump of size n. If $n >$ `neighmax`, proceed to Step 8.

- **Step 8:** Iterate VNS cycles up to `itermax` and stop.

  For $j = 1, \ldots,$ `itermax`, repeat Steps 2–7. After completing the `itermax`-th VNS cycle, stop and return the best grouping γ^* and associated parameters estimated.

Multiple starting values: Following the supplementary appendix in BM2015, Algorithm 2 typically runs for `n_starts` independent random initializations[10]. `n_starts` is set to reduce sensitivity to initialization. We repeat Steps 1–8 for `n_starts` different initial values and also different seeds. We keep the solution with the lowest objective function value (i.e., which minimized squared errors.)

Therefore, the choice of running parameters for the GFE Algorithm 2 in our case includes (`n_starts`; `itermax`; `neighmax`; `max_local_iters`). Setting up higher `n_starts` is computationally intensive. We start with `n_starts` = 3 to find the better number of groups G among all the G we tried, and then we increase the `n_starts` to 10 for a few chosen better G to check with the different start value to see if the algorithm yield the same results. See the Figure 1 for

[10] BM2015 calls this parameter as N_s.

the flowchart the steps of the algorithm described above, and the pseudocode (1) in Appendix A.

2.3 Estimation in an Unbalanced/Rotating Panels

A key practical feature of ENAHO data is that it is an unbalanced rotating panel: most households in this panel are observed for only a short consecutive period and are missing in other years. We therefore implement the GFE in a way that (i) uses all available household-year observations, and (ii) ensure that missing household-year observations do not mechanically affect either parameter updates or group assignments.

Let our observed set to be $\Omega \equiv \{(i,t) : d_{it} = 1\}$ where $d_{it} \in \{0,1\}$ indicates whether household i is surveyed (and has non-missing outcome and covariates) in year t. All least-squares updates and assignment steps are computed on Ω only, consistent with the masked objective in Eq. (2). More precisely, conditional on a candidate grouping $\gamma = \{g_1, \dots, g_N\}$, the updated step estimates (θ, α, μ) by regression y_{it} on x_{it}, group-by-time indicators, and province indicators using the pooled household-year sample restricted to Ω. This provides us the estimates of common slopes θ, the group-time effects $\{\alpha_{gt}\}$, and the province fixed effects $\{\mu_p\}$ based on the data observed cells only.

Given (θ, α, μ), the assignment step also respect the rotating-panel structure, for each household i, we assign the group label by minimizing the sum of squared residuals over the years in which the household is observed:

$$\hat{g}_i = \underset{g \in \{1,\dots,G\}}{\operatorname{argmin}} \sum_{t=1}^{T} d_{it} \left(y_{it} - x_{it}'\theta - \alpha_{gt} - \mu_{p_i} \right)^2 \tag{9}$$

Thus, years in which household i is not interviewed ($d_{it} = 0$) contribute zero to the assignment criterion. This is important in practice: households with shorter observed windows are not penalized for missing years, and their group membership is inferred from the information they provide. Intuitively, the rotating panel still provides identifying variations because some households are observed in consecutive years. These overlaps link adjacent years and help us pin down both (i) household group membership and (ii) how each group's unobserved component evolves over time. At the same time, the continual entry of new households each year maintain the cross-sectional sample size, ensuring the precision with which we estimate the group-specific time effects in each year.

After group assignments g_i, time effects α_{gt}, covariates coefficients θ, and province fixed effect coefficient μ_p are estimated, we can predict the outcome measured by IHS-transformed total expenditures per capita for any observed households:

$$\hat{y}_{it}^{GFE} = x_{it}'\hat{\theta} + \hat{\alpha}_{\hat{g}_i t} + \hat{\mu}_{p_i} \qquad (i,t) \in \Omega \tag{10}$$

These fitted values provide us a cleaned representation of household welfare that preserves (i) systematic variation explained by observables, (ii) persistent location differences capture by

province fixed effects, and (iii) group-specific time patterns in unobserved heterogeneity. We can then use the estimated group-time paths to construct poverty transition measured over horizons that exceed a household's observed panel length, which we talk in detail in the next Section 2.4.

2.4 From Estimated Groups to Poverty Dynamics

The GFE provides us two useful measures that we can use directly to studying poverty dynamics in a rotating panel: (i) an estimated latent group assignment for each household, $\hat{g}_i$, and (ii) an estimated group-specific time path, $\hat{\alpha}_{gt}$ (net of observed covariates we used, and also province fixed effects). The $\hat{\alpha}_{gt}$ summarizes how the expenditure trajectory evolves for households of latent type g, while $\hat{g}_i$ tells us which trajectory is most consistent with a given household's observed outcomes.

Using these estimates, we predict a model-implied expenditure series that is similar to Eq. (10), but with missing years covariates x_{it} filled by the researcher.

$$\hat{y}_{it}^{\text{comp}} = x_{it}'\hat{\theta} + \hat{\alpha}_{\hat{g}_i t} + \hat{\mu}_{p_i} \tag{11}$$

Essentially, Eq. (11) defines a model-implied complete expenditure path $\hat{y}_{it}^{\text{comp}}$ over the full time horizon, including years in which a household is not surveyed while Eq. (10) defines the in-sample fitted value for household-year cells that are actually observed in rotating panel, i.e., $(i,t) \in \Omega$ where $d_{it} = 1$. The predicted expenditure $\hat{y}_{it}^{\text{comp}}$ replaces the missing household-year information with a group-consistent trajectory rather than treating the missing years information as attrition or requiring balanced panels.

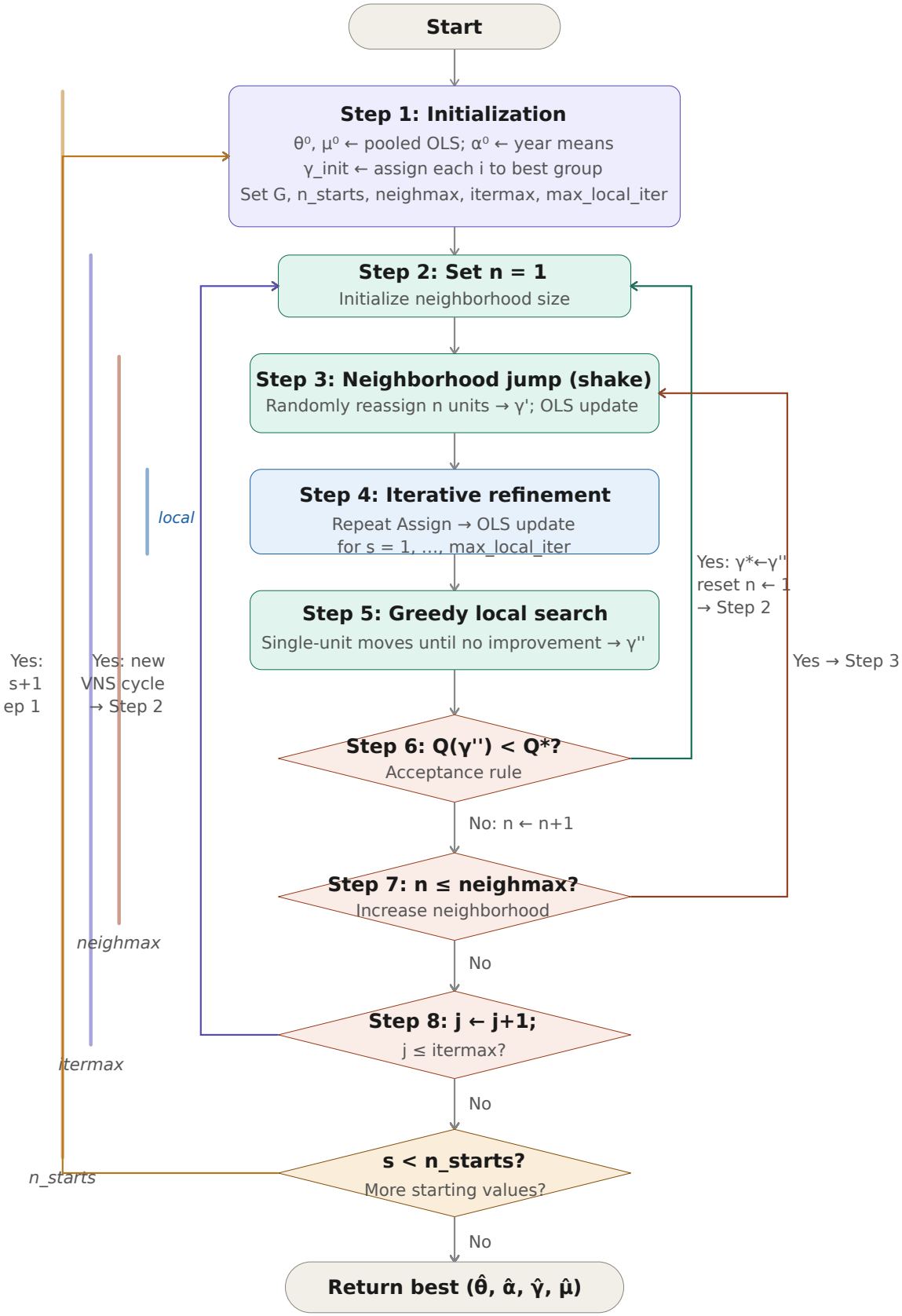
Constructing Eq. (11) therefore requires the value of x_{it} for every year t, even when the household is not observed. When the covariates are time-invariant, this extension is straightforward because the same x_{it} can be carried across years. However, if some covariates vary over time, we could specify how x_{it} is filled in for missing household-year cells. We could fill the missing by carrying forward the last observed value, anchoring to the first observed value, interpolating between observed waves, predicting x_{it} from auxiliary models or external information, or other methods deemed appropriate by the researcher. Importantly, this covariate-completion step is an additional researcher input that affects the level and timing of the predicted series in Eq. (11), while the group-time component and province fixed effect provide the structured dynamics and location adjustment learned from the observed rotating-panel overlap.

Once we have the fitted expenditures $\hat{y}_{it}^{\text{comp}}$, we can translate them into poverty status using the year-specific poverty thresholds available in ENAHO data. We can compare $\hat{y}_{it}^{\text{comp}}$ against the poverty line thresholds z_t (total poverty line or food poverty line in year t), and define a poverty indicator as

$$\hat{P}_{it} = 1\{\hat{y}_{it}^{\text{comp}} < z_t\} \tag{12}$$

and compute poverty dynamics over different horizons h by comparing $\hat{P}_{i,t}$ and $\hat{P}_{i,t+h}$. This yields transition probabilities – entry, exit, and persistence rates – as well as transition matrices

Figure 1: GFE Estimation Flowchart



(poor, non-poor). Because the fitted series is available for all the year, we can also summarize long-run measures such as the share of time spent in poverty, and incidence of chronic poverty (e.g., poor in at least certain number of years within a window), and the distribution of poverty lengths (e.g., two years, three years, or more) from the predicted outcome.

2.5 Latent Group G Selection

A key practical choice in the GFE framework is the number of latent groups, G. Larger G increases flexibility by allowing more distinct time profiles of unobserved heterogeneity, but it also raises the risk of overfitting, especially in short, unbalanced panels where many households are observed for only a few years. We therefore select G using a transparent model selection criterion, and we complement this with out-of-sample validation and qualitative checks on the stability and interpretability of the estimated group paths.

To select the number of latent group G, we combine Bayesian Information Criterion (BIC) used in BM2015 with Root Mean Square Error (RMSE). For a given G, the GFE estimator produces $(\hat{\theta}, \hat{\alpha}, \hat{\mu})$ and a minimized sum of squared residuals over observed household-year cells:

$$\text{SSE}(G) = \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} \left(y_{it} - x_{it}' \hat{\theta} - \hat{\alpha}_{\hat{g}_i, t} - \hat{\mu}_{p_i} \right)^2 \tag{13}$$

where $d_{it} = 1$ indicates an observed household cell and p_i denotes the household's time-invariant location (province).

Let $NT_{obs} = \sum_{it} d_{it}$ be the number of observed household-year cell used in estimation. We compute the estimated error variance as

$$\hat{\sigma}^2(G) = \frac{\text{SSE}(G)}{NT_{\text{obs}} - g(G)} \tag{14}$$

where $g(G) = GT + N + K + L$ is the effective parameter count. GT counts the group-by-time effects α_{gt}, N counts the unit-level group assignments (one label per household), K counts the common slope coefficients θ, and L counts the location fixed effects μ_p (one per province).

Our BIC is then computed as:

$$\text{BIC}(G) = \frac{\text{SSE}(G)}{NT_{\text{obs}}} + \hat{\sigma}^2(G) \frac{p(G)}{NT_{\text{obs}}} \ln\left(NT_{\text{obs}}\right) \tag{15}$$

The key difference between the BIC used in our application versus the one in BM2015 is that we additionally absorbed time-invariant location heterogeneity using province fixed effects. This adds L additional parameters to the penalty term and slightly reduces the effective degree of freedom in $\hat{\sigma}^2(G)$. Aside from this adjustment, the criterion is computed in the same way, using only the observed household-year cells in the unbalanced rotating panel.

Because the GFE objective is non-convex in group assignments, different starting values can lead to different local optima. To reduce sensitivity to initialization, for each candidate G, we run the estimation algorithm multiple times with different random starts and retain the

solution with the lowest objective function value, $SSE(G)$. We then compute $BIC(G)$ from the best solution. We choose the best G by looking at the both RMSE and BIC values (for both RMSE and BIC, the lower value the better).

3 Empirical Implementation

3.1 Data: ENAHO 2007 - 2019

We use the National Household Survey on Living Conditions and Poverty (ENAHO, an acronym in Spanish) from 2007 to 2019 to investigate poverty dynamics in Peru.[11] ENAHO is a nationally-representative household survey that collects data on housing, household expenditure, education, health, employment, and income (INEI, 2017). The main indicator for our study in the ENAHO is the poverty status (non-poor, poor) which is determined by whether household per capita expenditure is above the total poverty line (non-poor), below the total poverty line but above food poverty line (non-extremely poor) or below food poverty line (extremely poor).[12] Every year, 30% of the sample households are categorized as panel households which are repeatedly surveyed for up to six years. In our study period, 22.7% of the panel households were surveyed no more than three times, and only 6% of them were surveyed over 6 years. This rotating panel structure allows us to apply the GFE methods to study poverty dynamics overtime. For households who were in the data for at least three years, we can study their long term poverty dynamics using data from other households in the survey once they are grouped under the same group. The full dataset includes around 471,072 household-year observations from 375,897 unique households.

We further restrict our sample as follows. First, we restrict our study sample to households observed for at least three survey rounds. A minimum of two rounds is required for GFE estimation, and we use the final observed round for out-of-sample evaluation. Second, we restrict to households with a household head aged 25-55 in any given survey year. Because our analysis relies on identifying latent household groups with distinct time patterns, it is important that the set of households used for estimation reflects a relatively stable household structure. In rotating panel data, very young household heads are more likely to be entering newly formed households, while older heads are more likely to experience household dissolution, retirement-related changes, or mortality. All of which could generate non-random attrition and discrete shifts in household composition. These demographic transitions can mechanically alter observed expenditures and poverty status in ways that are difficult to distinguish from genuine welfare dynamics, and they can also complicate the interpretation of GFE "types" (groups).

To reduce these concerns and improve comparability of households across waves, we restricted our sample to households whose head is between 25 and 55 years old in any given survey year, an age range in which household formation and dissolution are less frequent and

[11] We processed our data based on the programs from Murillo and Sardon (2024) with further adjustments.

[12] We could not find the official reference how this poverty status in the data is determined, but we manually constructed the indicators comparing household per capita expenditure and the different poverty lines which matched the official indicators with 99.9% accuracy.

labor market attachment is more stable. After applying these sample restrictions (households surveyed for at least 3 years, and the head of household aged 25-55 in each survey year), the data sample we use for the GFE estimation includes 56,390 observations from 14,886 households. As a robustness check, we report results using the same three-wave requirement but without imposing the household-head age restriction with a sample of 104,304 observations from 27,034 households.

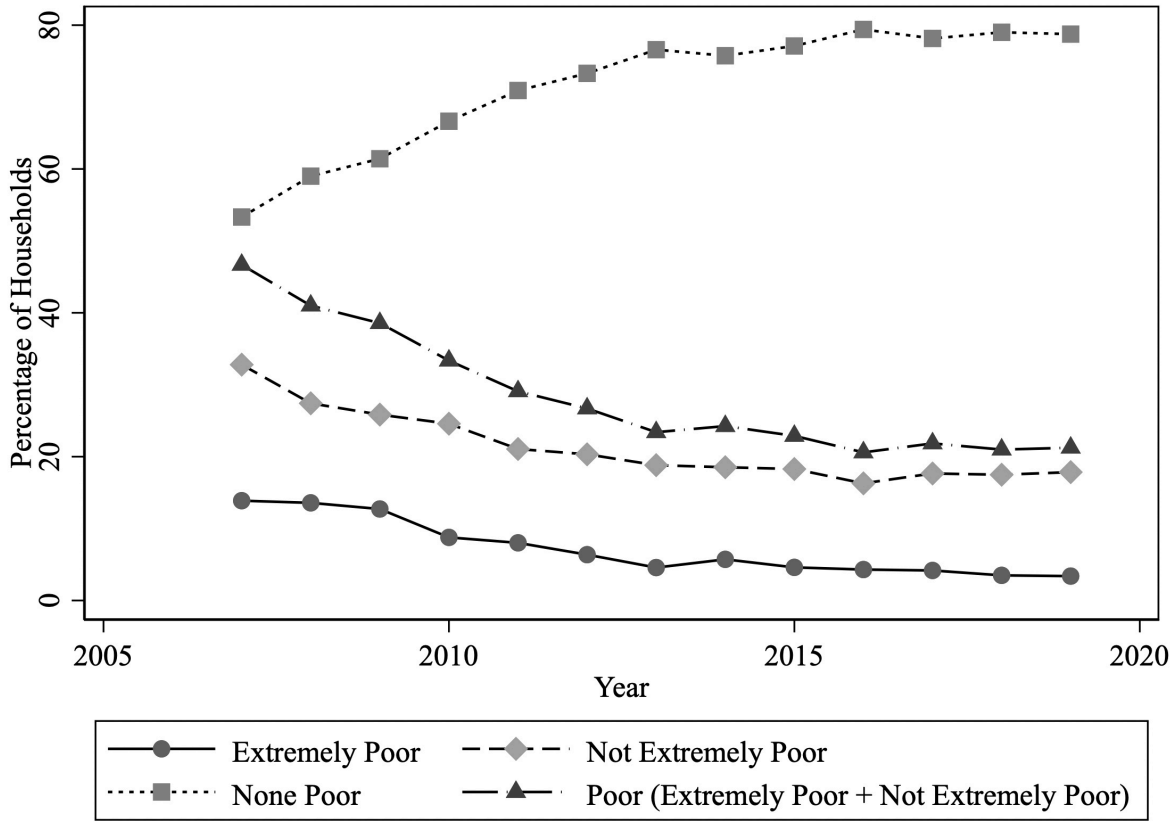
Figure 2 shows the trend in poverty status in our selected sample. Consistent with Peru's broad poverty reduction over the period, the share of households that are poor (extremely poor plus not extremely poor) declines steadily from 2007 to 2019, while the share of non-poor households rises correspondingly. Both components of poverty fall over time: extremely poverty drops sharply and becomes a relatively small share of the sample by the end of the period, and non-extreme poverty also trends downward. We present the non-restricted (full) sample poverty status in Appendix Figure A1. The overall patterns are very similar, suggesting that our main sample restriction do not change the narrative. However, the levels are slightly different, with the restricted sample exhibiting a higher poverty rate in the earlier years. Requiring repeated observation selects households that can be re-contacted and followed over time, which may differ from households observed only once (e.g., in mobility, location, or demographics). This selection can shift poverty levels relative to the full sample. Importantly, the time trends are very similar across sample, and we further assess sensitivity by reporting robustness checks that relax the age restriction and use the broader sample. In both figures, poverty categories are mutually exclusive within each year, so the shares of poor and non-poor sum to 100 percent.

Table 1 reports weighted summary statistics for our restricted ENAHO sample. Panel A summarizes household-year outcomes used in the poverty dynamic analysis. Average monthly per-capita total expenditures are around 445 (in 2007 dollars), and average total and food poverty lines are 235 and 125 (also in 2007 dollars), respectively. We also report the corresponding Inverse Hyperbolic Sine (IHS) transformed measures. Panel B describes baseline household and head characteristics measured in each household's first observed survey year to avoid over-weighting households that appear in multiple waves. In this sample, around 21 percent of the household heads are female, and the average head age is around 41 years old. 30 percent of household heads have primary education, 38 percent of household heads have secondary education, 18 percent of household heads have college education, and the rest of household heads has no primary education. Most households are Spanish-speaking (75 percent) and married or cohabiting (77 percent). Household living conditions are relatively well in our sample: 76 percent live in urban areas, 74 percent have access to water, and 89 percent have electricity.

3.2 Model Specification

Our main welfare measure y is the IHS-transformed monthly household total expenditure per capita (in 2007 dollar). We consider two specifications that differ in the set of covariates x_{it}. The first specification is a conservative one which we include characteristics that are

Figure 2: The trend of poverty status by year



Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years and to household heads aged 25–55 in all observed years.
Note: We applied our sample restriction: keep households that stayed in the survey for at three two years, and restricted the sample to household heads between 25 and 55 years old in any given survey years.

Table 1: Summary Statistics (Weighted)

	Mean	SD
Panel A: Household-year level (N=56,390)		
Total expenditures per capita (monthly) - 2007 dollars	444.92	418.24
Total poverty line - 2007 dollars	235.42	48.44
Food poverty line - 2007 dollars	124.73	18.81
IHS (Total expenditures per capita (monthly) - 2007 dollars)	6.53	0.70
IHS (Total Poverty Line - 2007 dollars)	6.13	0.20
IHS (Food Poverty Line - 2007 dollars)	5.51	0.15
Panel B: Household-level (first observed year) (N=14,924)		
Head of household		
Female	0.21	0.41
Age	41.22	7.60
Primary school	0.30	0.46
Secondary school	0.38	0.48
College	0.18	0.38
Language - Spanish	0.75	0.43
Married or cohabiting	0.77	0.42
Household		
Urban	0.76	0.43
Has water	0.74	0.44
Has electricity	0.89	0.31

Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years and to household heads aged 25–55 in all observed years.
Note: The table reports weighted means, standard deviations, and counts. The sample includes households observed at least three times over 2007-2019, with head age 25–55 in each survey year. Panel A uses household-year observations; Panel B uses the household's first observed year. All the summary statistics are weighted by the Annual Expansion Factor CPV-2007 projections from the ENAHO data.

more deterministic or unlikely to vary over time. These characteristics include age, gender, educational attainment, and the Spanish spoken language of the head of the household. The second specification adds additional characteristics that are more likely to vary over time, but important in capturing variation in expenditure. These characteristics include marital status, urban/rural residence, and whether the household has water and electricity.

To examine how much additional characteristics in the second specification capture household expenditure, we regress household expenditure on both specifications. Table 2 reports estimates of pooled household-year regression of IHS expenditure per capita on the two covariate sets. All specification include province and year fixed effect to net out time-invariant differences across local areas (e.g., geography, prices, and local labor market conditions etc.), apply the ENAHO survey expansion weight, and cluster standard errors clustered at the household level.[13] The richer specification increases the explanatory power of the model, raising the R^2 from 0.49 to 0.54, and we use these two sets of covariates as alternative inputs in our subsequent GFE analysis. In the Appendix, we show that the main qualitative patterns in Table 2 are not sensitive to using coarser (department) (Table A2) or more granular (Table A3) (district/UBIGEO) location fixed effects. We choose province fixed effects to control for a meaningful amount of local price/labor market/geography differences without fully saturating location. This specification leaves enough within-province variation for our GFE model when grouping the households into different latent type G.

3.3 Estimation Steps

Because our intended application is to use information observed for a household in earlier survey rounds to predict its welfare in a later (unobserved) round in an unbalanced rotating panel, we therefore evaluate the GFE model using a validation design that mimics this forecasting task: for each household, we hold out its last observed survey year (test data) and estimate the model on all remaining household–year observations (training data). We prefer this test set design to other alternatives: a random subset of households for the entire period or all observations from a specific calendar year. We do not prefer leaving a random subset of households for an entire period, since our primary objective is not to predict outcomes for households that have never been in the survey. Instead, we aim to predict a household's future welfare using its earlier observed history, which is exactly the application of GFE in rotating panels. In addition, holding out entire households does not change the set of years used for estimation but tests a different notion of generalization (across households rather than forward in time within a household). A further alternative—holding out later calendar years entirely–is also not appropriate in our setting because it would remove the information needed to estimate α_{gt} for those later holdout years, and GFE does not extrapolate unknown future α_{gt} without additional structure.

[13] Peru's administrative divisions are hierarchical: the country is divided into departments (also referred to as regions), which are subdivided into provinces, which in turn are subdivided into districts. In the ENAHO data, the geographic identifier UBIGEO corresponds to the district level. In our selected sample, we have 25 departments, 195 province, and 1121 districts.

Table 2: Household Total Expenditure Per Capita Estimates

	(1)	(2)
Head of Household - Female	0.1248***	-0.1075***
	0.0120	0.0164
Head of Household - Age	-0.0241***	-0.0168***
	0.0066	0.0060
Head of Household - Age squared	0.0004***	0.0003***
	0.0001	0.0001
Head of Household - Primary school	0.1547***	0.1143***
	0.0114	0.0108
Head of Household - Secondary school	0.3849***	0.2907***
	0.0134	0.0130
Head of Household - College	0.8049***	0.6792***
	0.0178	0.0173
Language - Spanish	0.1582***	0.1258***
	0.0127	0.0121
Head of Household - Married or Cohabitant		-0.3029***
		0.0160
Urban		0.2340***
		0.0121
Has water		0.1325***
		0.0102
Has electricity		0.1656***
		0.0150
R-squared	0.492	0.538
N	56390	56390

Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years and to household heads aged 25–55 in all observed years.
Note: ***, **, * indicate significance at the 1%, 5%, and 10% levels. The dependent variable is monthly household total expenditures per capita (2007 dollars), inverse hyperbolic sine transformed. The sample includes households observed at least three times and with head age 25–55 in each survey year, 2007–2019. All regressions include province and year fixed effects and are weighted by the annual expansion factor (CPV-2007 projections). Standard errors (in parentheses) are clustered at the household level.

Step 1: Construct a rotating-panel estimation sample and mask. Let $i = 1, \dots, N$ index households and $t = 1, \dots, T$ index calendar years in our analysis window. We stack the data into household–year observations and define an observation indicator $d_{it} \in \{0, 1\}$ equal to one if household i is observed in year t with non-missing outcome and covariates, and zero otherwise. All estimation and objective function evaluations use only observations with $d_{it} = 1$, consistent with the unbalanced-panel formulation of the GFE objective.

Step 2: Define training and test data by last-observed-year holdout. For each household i, let t_i^{last} denote the last year in which the household is observed (i.e., the largest t such that $d_{it} = 1$). We construct a training mask

$$d_{it}^{\text{train}} = d_{it} \cdot \mathbf{1}\{t \neq t_i^{\text{last}}\}, \tag{16}$$

so that each household contributes all observed years except its last one to the estimation, and this data is our training data. The held out data consists of the observations $\{(i, t_i^{\text{last}}) : d_{i,t_i^{\text{last}}} = 1\}$, and this is our test data. This split preserves the rotating-panel structure: although a given household's last year is excluded from estimation, other households observed in the same calendar year remain in the training data, so the year-specific group effects α_{gt} are still identified from the cross-sectional sample in that year.

Step 3: Estimate GFE for a candidate number of groups G. For each candidate G, we estimate the GFE model in Eq. (1) by minimizing the masked sum of squared residuals over the training data:

$$\min_{\theta, \alpha, \gamma} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it}^{\text{train}} \left(y_{it} - x_{it}'\theta - \alpha_{g_i t} - \mu_{p_i} \right)^2 \tag{17}$$

where $\gamma = \{g_i\}_{i=1}^N$ denotes the household-to-group assignments. We implement Algorithm 2 of BM2015 (VNS with local search) adapted to the unbalanced panel by evaluating objectives and assignment losses only over years with $d_{it}^{\text{train}} = 1$. To reduce sensitivity to local minima, we use a multi-start strategy with `n_starts` different initializations (different random seeds) and retain the solution with the lowest objective value for each G.

Step 4: Model selection over G and predictive validation. For each G we compute (i) the BIC on the training data (based on the minimized masked SSE and the corresponding parameter count) and (ii) out-of-sample predictive performance measured using RMSE on the test data. For each held-out test observation (i, t_i^{last}), we form the prediction:

$$\hat{y}_{i,t_i^{\text{last}}} = x_{i,t_i^{\text{last}}}'\hat{\theta} + \hat{\alpha}_{\hat{g}_i,\, t_i^{\text{last}}} + \hat{\mu}_{p_i} \tag{18}$$

where $\hat{g}_i$ is the estimated group assignment for the household i obtained from the training fit. We summarize the prediction accuracy using RMSE computed over the test data.

Step 5: Final estimation for the chosen G and stability checks. After selecting G, we re-estimate the GFE model for the six best-performing candidate values of G using a larger number of random starts (higher `n_starts`) to verify that the optimal selected latent group G is stable to initialization. We then select a single final G taken into account of both the BIC and RMSE values, and take its estimated group assignments $\hat{g}$, coefficient vector $\hat{\theta}$, location fixed effects $\hat{\mu}$ (when included), and group–time effects $\hat{\alpha}_{gt}$ as the parameters used in the subsequent analysis of poverty dynamics.

To construct poverty dynamics, we distinguish between two uses of the estimated model. First, for all predictive validation exercises (e.g., one-step-ahead transitions), we use the parameter estimates obtained from the training sample that excludes each household's last observed year. This ensures that the end-year poverty status used in validation is out-of-sample at the household-year level. Importantly, because other households remain observed in each calendar year, the group–year effects $\hat{\alpha}_{gt}$ are still identified for all years in the analysis window except for 2019 because our training data does not have 2019 observations and therefore we cannot estimate $\alpha_{g,2019}$.

Second, when we move from validation to describing poverty dynamics and constructing completed welfare paths (i.e., imputing non-survey years), we re-estimate $\hat{y}_{it} = x'_{it}\hat{\theta} + \hat{\alpha}_{\hat{g}_i t} + \hat{\mu}_{p_i}$ without leaving the last year observed as a test set, whenever the covariates x_{it} are available (or completed as described in Section 2.4), which allows us to trace model-implied welfare and poverty status over the full set of years covered by the survey.

4 Results

4.1 Selecting Best G using GFE

We run the GFE model with two sets of different covariates. The first sets of covariates are the ones from Table 2 column (1), including head of household gender, age, education level, and if primary language is Spanish. The second sets of covariates are the ones from Table 2 column (2) which add additional covariates including marriage status, household living in urban area, has water, and has electricity. We call the first sets of covariates as model specification 1, and the second sets of covariates as model specification 2.

Running the GFE model with `n_starts` $= 3$, `itermax` $= 10$, `neighmax` $= 5$, and set the `max_local_iters` $= 2$ for latent groups from 1 to 11, we can calculate the BIC for each of the latent groups using the training data, and calculate the RMSE using the test data.

Figure 3 shows the change in the BIC and RMSE values between different latent groups G using specification 1 in two different periods (2009-2018, 2019). While some observations in the former period were included in the test data (e.g., those last observed years fall within this period), none of the 2019 surveys were used as test data (since 2019 is always the last year observed), making the 2019 prediction closer to pure forecasting.[14] BIC declines steadily

[14] We let $\alpha_{g,2019}{=}\alpha_{g,2018} \forall G$

as G increases and attains its minimum at the upper end of the grid (e.g., the maximal G we tried, in this case for $G = 11$ for both specifications), indicating that adding groups keeps improving in-sample fit. This monotone pattern is expected in our setting because increasing G increases the number of group-time intercepts, α_{gt}, making the model more flexible and better able to fit heterogeneity in an unbalanced rotating panel. As G becomes large, however, the model begins to fragment the sample into many small groups. In an unbalanced rotating panel, this can leave some group-year (g, t) supported by only a very small number of observations, so the corresponding α_{gt} is estimated with substantial sampling noise. This "thin support" makes group-time path less stable and can also reduce the reliability of prediction, even if the in-sample fit (measured by the BIC) continues to improve.

In contrast, predictive accuracy shows a clear interior optimum: RMSE drops sharply as G moves from very small to moderate values, then flattens and eventually increases as G becomes large. In both periods, the minimum RMSE on the held-out last-observed household-year test data occurs at $G = 4$, suggesting that additional groups beyond this point primarily capture idiosyncratic variation rather than improve forecasting performance. Since our primary goal is prediction, we emphasize the RMSE-based criterion, using BIC as a complementary diagnostic for in-sample fit and model complexity.

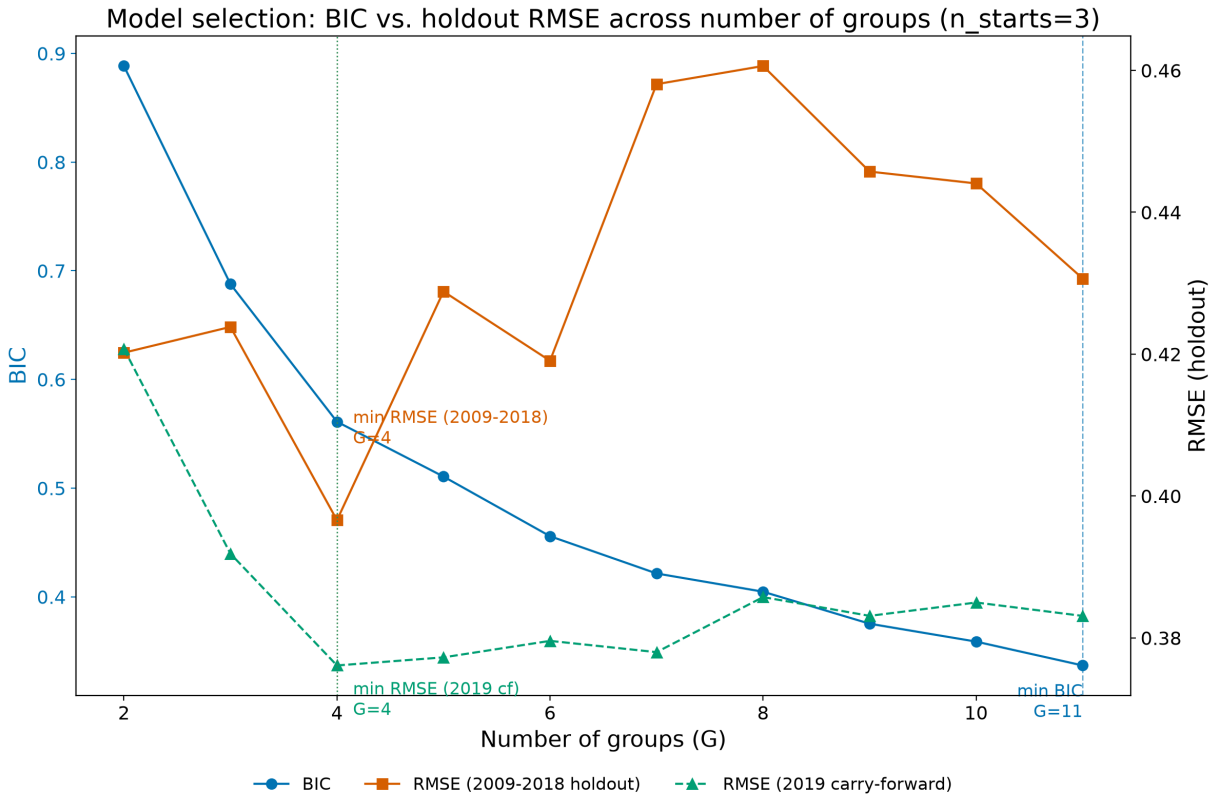
To ensure the low-RSME results are not just due to a lucky (or unlucky) random starting point that captures specific local minima rather than a truly better grouping structure, we conduct an additional stability check by increasing the number of random starts, as described in Step 5 in Section 3.3. Specifically, we take the six values of Gs with the lowest holdout RMSE in the initial grid search as the six best-performing candidate values, $G \in \{2, 3, 4, 5, 6, 11\}$, and re-estimate the model with the same algorithm setting but a larger number of initialization (we increase `n_starts` from 3 to 10). Using more random starts makes it less likely that the algorithm reaches local minima strictly greater than global minimum for a given G, so the comparison of prediction performance across different G value is more reliable.[15]

Figure 4 summarizes the resulting BIC and RMSE for these re-estimated candidate values of G with `n_starts` $= 10$. We find that the values of G that performed well in the initial run remain mostly stable. For the 2009-2018 prediction, $G = 3$ now performs slightly better than $G = 4$, while $G = 4$ yields a substantially lower RMSE than $G = 3$ for the 2019 prediction. The RMSE gap between these two G values is larger in 2019, whereas the gap in the 2009-2018 prediction is very small. Setting $G > 4$ increases RMSE, particularly for 2009-2018 prediction, reinforcing the interpretation that additional flexibility (larger G) primarily increases variance (through thinner support for some group-year cells) rather than improving generalization.

Based on these results, we select $G = 4$ as our preferred baseline choice for overall prediction accuracy for the 2009-2018 and the 2019 prediction. This value achieves the lowest (or near-lowest) out-of-sample RMSE while avoiding the instability concerns that arise when the model is forced to estimate many group-time effects with very few observations. We replicated these

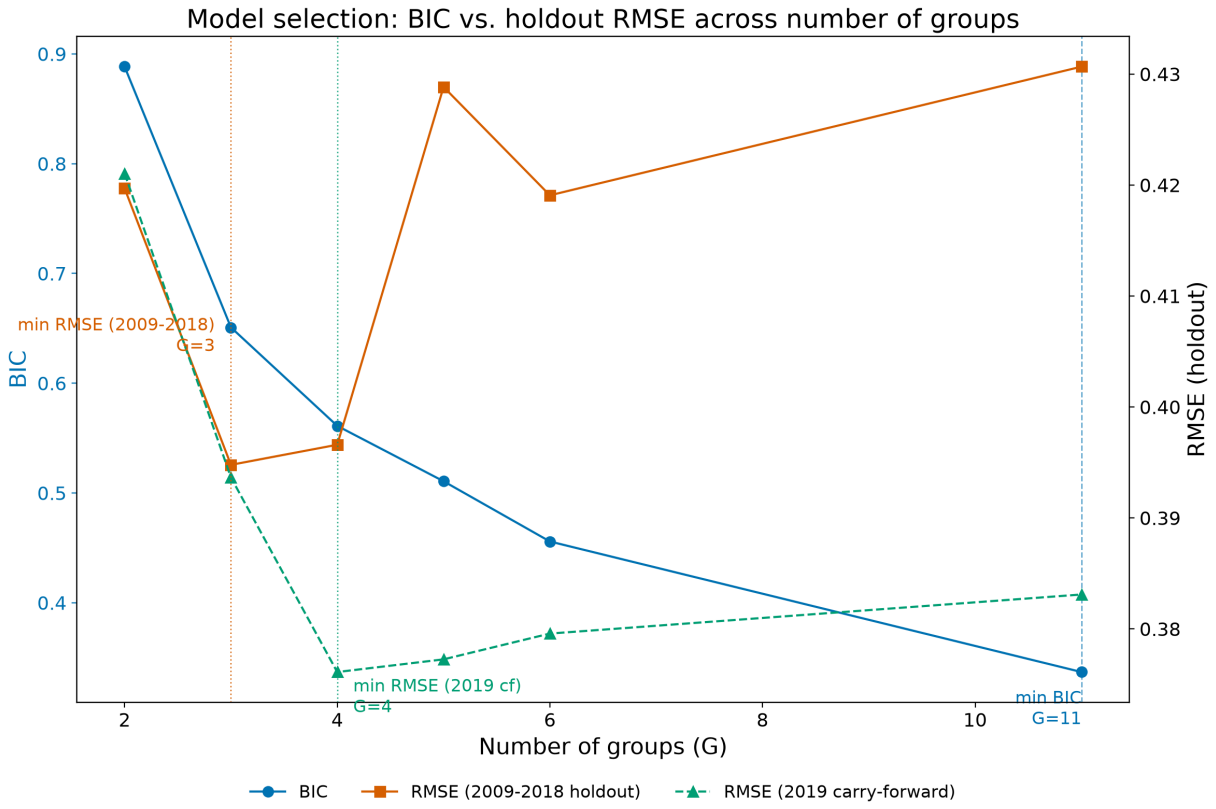
[15] We only increase to `n_starts` $= 10$ after the initial $G \in \{1, \ldots, 11\}$ grid search because computation time grows quickly with the number of random starts, so it is more efficient to reserve the higher `n_starts` run for a small set of promising G values identified in the initial run.

Figure 3: Model fit (BIC and RMSE) under specification 1 (`n_starts` = 3)



Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years and to household heads aged 25–55 in all observed years.
Note: The figure plots model selection diagnostics for the GFE estimator across initial numbers of groups $G \in \{1, \cdots, 11\}$ with `n_starts` = 3, `itermax` = 10, `neighmax` = 5, and `max_local_iters` = 2, predicted on two different periods (2009-2018, 2019). BIC is computed on the training data; RMSE is computed out-of-sample using the last observed household-year held out for each household (years 2009–2018 or 2019) test data. Lower values indicate better fit/prediction.

Figure 4: Model fit (BIC and RMSE) under specification 1 (`n_starts` = 10)



Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years and to household heads aged 25–55 in all observed years.
Note: The figure plots model selection diagnostics for the GFE estimator across candidate numbers of groups G predicted on two different periods (2009-2018, 2019), with `n_starts` = 10, `itermax` = 10, `neighmax` = 5, and `max_local_iters` = 2. BIC is computed on the training data; RMSE is computed out-of-sample using the last observed household-year held out for each household (years 2009–2018) test data. Lower values indicate better fit/prediction.

diagnoses under Specification 2 in Figure A2 and A3, where the results remain robust.

Our results on model fit across different values of G and specifications point to two practical takeaways. First, moving from specification 1 to specification 2 by adding the extra household-level controls tends to improve fit and forecasting performance at the same G: BIC is lower and the test RMSE is generally smaller in specification 2. This suggests that additional covariates capture meaningful variation that would otherwise be absorbed by the latent groups, although the reduction of RMSE by adding additional covariates in our specification is quite small.

Second, re-running the best-performing set of G values with a larger number of random starts produces almost identical RMSE across most cases, and only modest improvements in a few of them. This pattern indicates that the estimation is generally not being driven by unlucky initializations: with `n_starts` $= 3$, the algorithm already tends to find solutions with very similar out-of-sample performance, and increasing `n_starts` mainly serves as a robustness check against local minima rather than changing the substantive ranking of candidate G values. Taken together, these results support our approach of using a low `n_starts` grid search to narrow down promising values of G, followed by a higher `n_starts` rerun on the shortlist to confirm stability before selecting our optimal G.[16]

In the remainder of the analysis, unless otherwise noted, we report results using the model estimated at $G = 4$ with `n_starts` $= 10$, and we use the corresponding group assignments, group-time paths, and location effects as the basis for constructing predicted welfare trajectories and poverty transitions.

4.2 Welfare Trajectories at Selected G

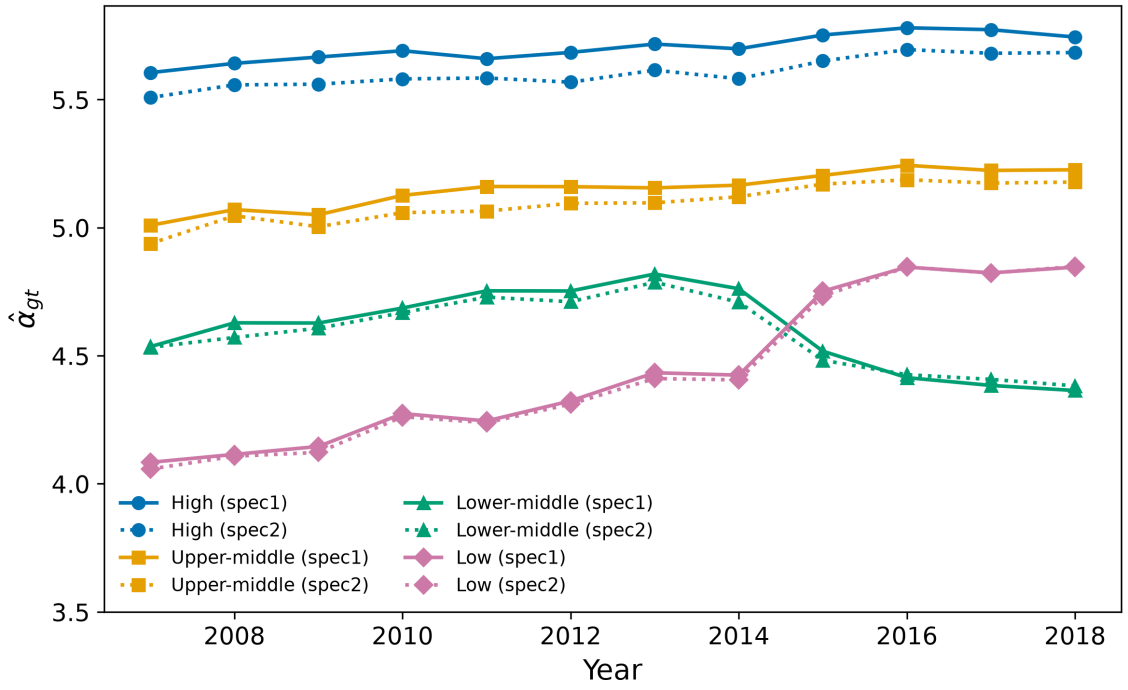
Figure 5 shows the estimated group-year intercepts $\hat{\alpha}_{gt}$ for our selected model with $G = 4$, overlaying specification 1 (solid lines) and specification 2 (dotted lines). These intercept paths summarize the evolution of latent welfare components after netting out the contribution of observed covariates and the province fixed effects, so difference in $\hat{\alpha}_{gt}$ reflect time-varying unobserved heterogeneity that is common to households assigned to the same latent group.

The figure shows four distinct welfare trajectories. "High" and "Upper-middle" groups are consistently at the top of the distribution throughout the sample with only modest movements over time, suggesting a persistently better-off latent type. "Lower-middle" group experiences an increase in the early years, followed by a gradual decline after 2013, while "Low" group is comparatively stable and keeps increasing in the later years. The similarities in welfare trajectories between the two specifications indicate that adding the richer set of controls does not significantly alter the overall pattern of latent welfare dynamics captured by the GFE.

To complement Figure 5, Figure 6 translates the latent welfare trajectories into intuitive welfare metrics: the share of households living below the poverty line. For each year t and group g, we compute the survey-weighted poverty rate among household-year observations assigned to group g under the selected $G = 4$, overlaying specification 1 (solid lines) and specification 2

[16] We report the full table for BIC and RMSE by specification and the number of random starts in Appendix Table A4. See Column "(8)-(6)" for the first point, and column "Diff (ns10-ns3)" for the second point.

Figure 5: Estimated group–year intercept paths $\hat{\alpha}_{gt}$ for $G = 4$

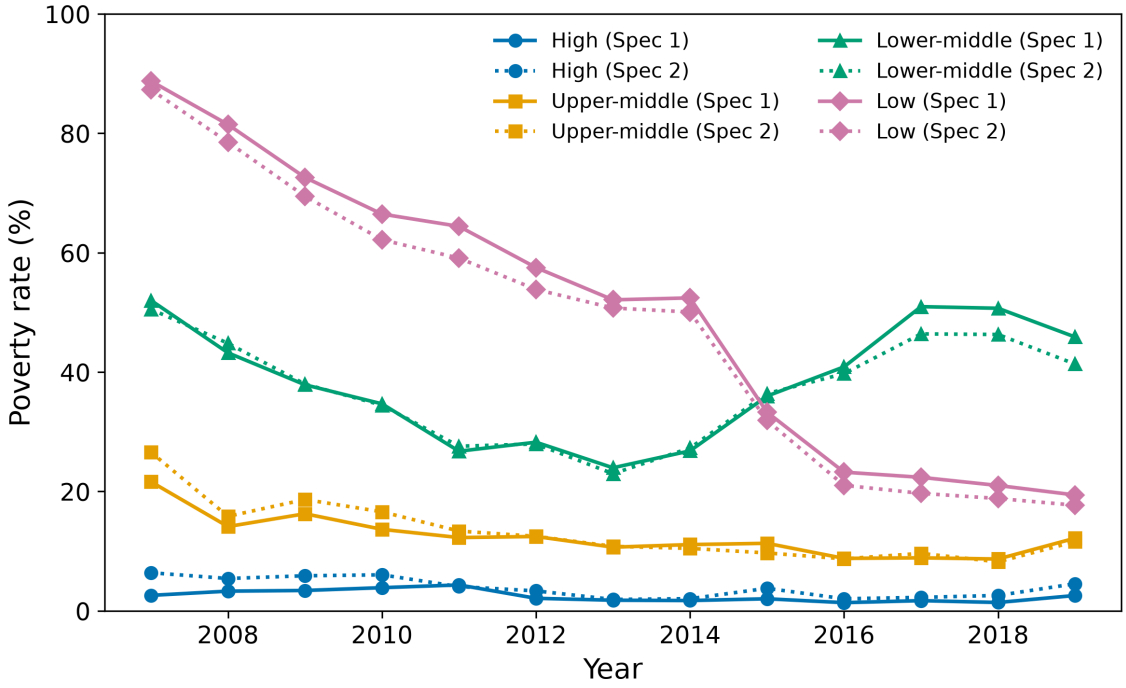


Data: Estimated group-year intercept paths for G=4 using training data.
Note: Colors indicate groups. Solid lines are specification 1 and dotted lines are specification 2. $\hat{\alpha}_{gt}$ is the estimated group×year intercept from the GFE model (net of covariates). Interpretation should focus on relative differences and dynamics across groups rather than the value.

(dotted lines). As we can tell from these two figures, the resulting patterns of Figure 6 closely track the estimated $\hat{\alpha}_{gt}$ paths in the opposite direction, consistent with the interpretation of $\hat{\alpha}_{gt}$ as a latent component of welfare. The "High" group, which is the consistently high $\hat{\alpha}_{gt}$ group, exhibits the lowest poverty rates throughout the sample and remains largely stable over time. "upper-middle" group also shows a similar stable pattern over the period, with a modest poverty rate. The "Lower-middle" group shows an early improvement in $\hat{\alpha}_{gt}$ followed by post-2013 declines, beginning with very high poverty rates that fall in the early years and then slowly increase in the later period. The "Low" group shows the most significant improvement over time: poverty declines substantially as $\hat{\alpha}_{gt}$ rises. As the intercept paths, the poverty trajectories are highly similar across the two specifications, indicating that the added household controls in specification 2 do not materially change the qualitative welfare ordering or the timing of improvements across latent groups.

This close similarity between the dynamics of $\hat{\alpha}_{gt}$ and poverty rates helps us validate that the selected $G = 4$ grouping captures economically meaningful differences in welfare trajectories rather than purely statistical variation. This is because in our current specifications, if we treat our covariates as time-invariant, and both the coefficient vector $\hat{\theta}$ and the location effect $\hat{\mu}$ are time-constant, then the only flexible component that can capture systematic changes in welfare over time within each latent type is the group–year intercept $\hat{\alpha}_{gt}$, making its dynamics a direct summary of the estimated welfare trajectory for group g.

Figure 6: Poverty rate trends by GFE group (G=4): Specification 1 vs. Specification 2.



Data: All available selected sample (training + test).
Note: Values are survey-weighted poverty rates by group-year (ENAHO expansion weights). Solid lines are for specification 1 and dotted lines are for specification 2. Colors denote the same group across specifications. Groups use household-level GFE assignments from the selected G=4 model.

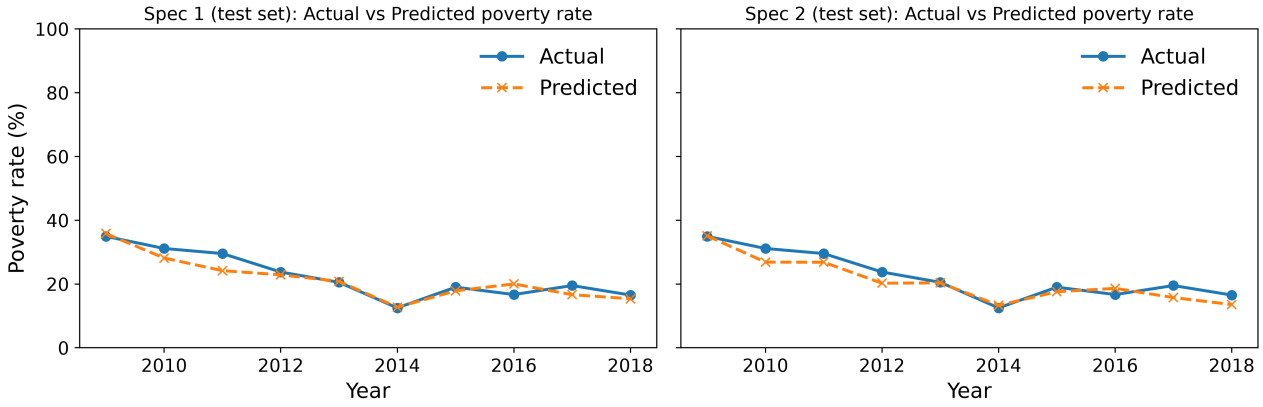
4.3 Model Validation at the Aggregate Level

Before turning to poverty transitions, we assess whether the GFE model reproduces the overall time pattern of poverty at the aggregate level when evaluated out of sample. Figure 7 compares the survey-weighted poverty rate computed from observed expenditures to the poverty rate implied by GFE-predicted welfare in the test data, where each household contributes its last observed survey year.[17] In both specifications, the predicted poverty rates are very close to the observed poverty rate across years, capturing the broad decline in poverty over the study period and remaining near the realized poverty rate in most years. This result indicates that the estimated group–year components, together with observables and location fixed effects, capture the main time variation in poverty at the population level even when predictions are formed on held-out household-years (test data).

As a reference, Appendix Figure A4 repeats the same comparison using all available observations (training plus test) and therefore shows an almost perfect overlap between actual and predicted poverty rates, as expected for an in-sample fit.

[17]Predicted poverty is defined as $\hat{P}_{i,t_i^{\text{last}}} = 1\{\hat{y}_{i,t_i^{\text{last}}} < z_{t_i^{\text{last}}}\}$, with $\hat{y}_{it} = x'_{it}\hat{\theta} + \hat{\alpha}_{\hat{g}_i,t} + \hat{\mu}_{p_i}$, and all year-level rates are aggregated using ENAHO expansion weights.

Figure 7: Actual vs. predicted poverty rate over time in the test data (G=4)



Data: Selected sample, test data only.
Note: Poverty rates are survey-weighted (ENAHO expansion weights). The test data consists of each household's last observed survey year. Predicted poverty is based on model-predicted welfare from the GFE model and the IHS poverty line (both in IHS scale).

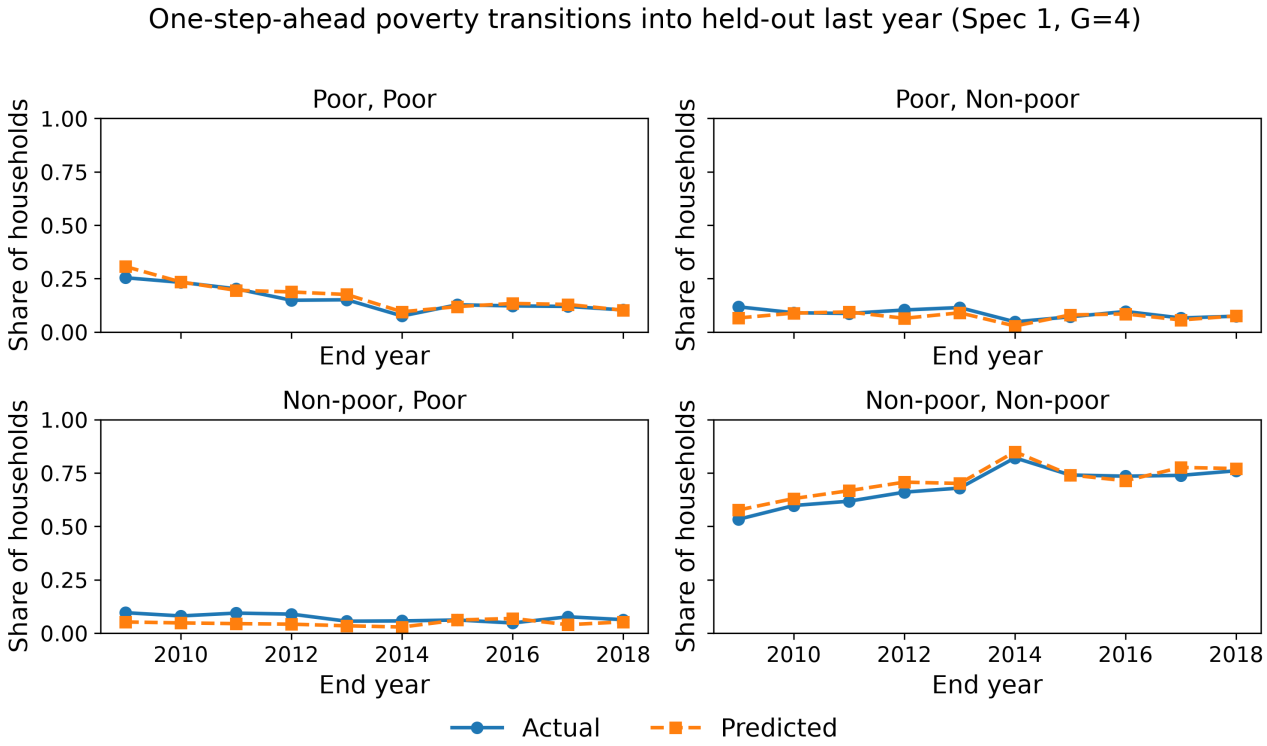
4.4 Poverty Transitions: One-step-ahead Validation

We next examine whether the GFE model can reproduce observed patterns of poverty transitions (poor-poor, poor-non-poor, non-poor-poor, and non-poor-non-poor). To provide an out-of-sample check that is feasible in our setting, we implement a one-step-ahead validation that mirrors how the model would be used in practice. For each household, we take its final observed survey year t as a held-out outcome and use the observed poverty status in the preceding year $t-1$ together with the GFE-predicted welfare in year t to construct predicted transitions. Specifically, we compare the realized transition based on $(P_{i,t-1}, P_{it})$ to the predicted transition $(P_{i,t-1}, \hat{P}_{it})$, where $\hat{P}_{it} = 1\{\hat{y}_{it} < z_t\}$. This design avoids using the household's realized welfare in the target year when evaluating transition patterns, while still allowing us to assess whether the model accurately captures changes in poverty status from one year to the next.[18]

Figure 8 reports the results of our one-step-ahead exercise for Specification 1 (For Specification 2, see Appendix Figure A7). Each point corresponds to an end year t and summarizes the distribution of two-year transitions among households whose final observed survey year is t and whose previous observed year is $t-1$ (restricting to consecutive pairs). The solid line in each panel shows the observed transition share computed from $(P_{i,t-1}, P_{it})$, while the dashed line replaces the end-year status with the model-implied classification, $(P_{i,t-1}, \hat{P}_{it})$. Across the four panels, the predicted series closely tracks the observed series, indicating that the GFE model reproduces not only the overall level of persistence in poverty (poor–poor and non-poor–non-poor), but also the relative frequency of entries into poverty (non-poor–poor) and exits from poverty (poor–non-poor) in held-out end years. This provides an out-of-sample check on the

[18] We also validated our results using all available household-year observations for which the model-implied welfare $\hat{y}_{it}$ can be constructed. Figures A5 and A6 show the results. Because these figures draw on the same observations used to estimate the model (and therefore include substantial training-sample information), they should be interpreted as showing that the estimated GFE structure delivers a transition pattern that is consistent with the data in sample, rather than as evidence of forecasting performance.

Figure 8: One-step-ahead poverty transition shares into held-out final observations: actual vs. GFE-predicted ($G = 4$), for specification 1.



Data: Last year of training data + all test data.
Note: The figure reports survey-weighted shares of households transitioning between poverty states from $t-1$ to t, where t is each household's held-out last observed year and $t-1$ is its preceding observed year (restricted to consecutive years; one transition per household). "Actual" transitions are computed using observed (IHS-transformed) per-capita expenditure relative to the (IHS-transformed) poverty line in both years. "GFE-predicted" transitions use actual poverty at $t-1$ and predicted poverty at t, where predicted poverty is $1\{\hat{y}_{it} < z_{it}\}$ based on model-predicted welfare $\hat{y}_{it} = x'_{it}\hat{\theta} + \hat{\alpha}_{\hat{g}_i,t} + \hat{\mu}_{p_i}$. Transition shares are weighted using ENAHO expansion weights in year t. In this one-step-ahead sample, the expansion-weighted misclassification rate of $\widehat{\text{poor}}_{it}$ in year t is 16.8%.

model's ability to capture short-run poverty dynamics, using information available in real time (poverty status in $t-1$ and predicted welfare in t). Overall, this one-step-ahead validation model obtains an expansion-weighted poverty-status prediction accuracy of 83.3% / 83.4% in the held-out year for Specifications 1 and 2, respectively.

4.5 Comparing GFE to Synthetic Panels

The most well-known approach to estimate poverty dynamics is the synthetic panel method proposed by Dang et al. (2014) and Dang and Lanjouw (2023), which uses repeated cross-sections to recover poverty transition rates. To benchmark our results against their approach, we replicate the synthetic panel point estimates using the same ENAHO data and plot the four transition shares over time: poor–poor, poor–non-poor, non-poor–poor, and non-poor–non-poor. Figure 9 compares the synthetic panel point estimates to the corresponding transition shares computed directly from the data for Specification 1 (see Appendix Figure A8 for Specification 2 comparison). The synthetic panel method tracks the broad time patterns in the transition shares reasonably well, but it also exhibits noticeable deviations in several years, especially for the non-poor–non-poor share, where small level differences in predicted transitions can cumulate into visible gaps.

We then evaluate how closely each method reproduces observed transition shares using a simple error metric. For each end year t, we treat the four transition shares as a probability distribution and compute the total variation distance between predicted and observed shares per each share s, $TV_t = \frac{1}{2}\sum_s |\hat{p}_{s,t} - p_{s,t}|$, along with Mean Absolute Error (MAE) and RMSE averaged across the four transition states. Table 3 summarizes these metrics for (i) our GFE-based approach where we predict in poverty status in both $t-1$ and t (ii) the synthetic panel approach which aggregates poverty transition from predicted household-level *continuous* probability in $t-1$ and t and (iii) the synthetic panel (binary) approach, which we modified by estimating poverty transition from predicted household-level *binary* poverty status. GFE-based approach yields larger errors ($RMSE \approx 0.054$ for Specification 1 and $RMSE \approx 0.055$ for Specification 2) than the original synthetic panel estimates ($RMSE \approx 0.026$ for Specification 1 and $RMSE \approx 0.023$ for Specification 2), but smaller than the synthetic panel (binary) method ($RMSE \approx 0.075$ for Specification 1 and $RMSE \approx 0.071$ for Specification 2), although the difference is modest. While the original synthetic panel estimates poverty transitions with smaller RMSE, this method assumes that the error terms in the welfare model between the two periods follow a bivariate normal distribution, allowing it to estimate continuous household-level transitions and to achieve lower RMSE. However, as we mentioned earlier, this assumption often does not hold. When we relax this assumption and estimate the binary poverty status from the synthetic panel model and then aggregate it, as in the GFE method, this modified "binary" method yields larger prediction errors. This apple-to-apple comparison between the GFE-based method and the synthetic panel (binary) method implies that GFE-based method can achieve higher prediction accuracy. Figure 10 plots the end-year RMSE in the four transition shares for the GFE one-step approach and the synthetic panel point estimates under Specification 1

Table 3: Fit of predicted vs. observed poverty transition shares (lower values indicate closer match).

dataset	spec	MAE_avg	RMSE_avg	TV_avg	TV_max
GFE	1	0.054	0.057	0.109	0.182
GFE	2	0.055	0.058	0.110	0.174
Synthetic panel	1	0.026	0.031	0.052	0.093
Synthetic panel	2	0.023	0.028	0.047	0.090
Synthetic panel (binary)	1	0.075	0.088	0.150	0.188
Synthetic panel (binary)	2	0.071	0.084	0.142	0.176

Notes: The table compares predicted and observed two-year poverty transition shares across the four transition states (poor–poor, poor–non-poor, non-poor–poor, and non-poor–non-poor). MAE and RMSE are averaged across the four states within each end year. TV denotes the total variation distance, $TV_t = \frac{1}{2}\sum_s |\hat{p}_{s,t} - p_{s,t}|$, computed for each end year t and then averaged (TV_avg) or maximized (TV_max) over years. "GFE one-step" uses observed poverty status in t–1 and predicts poverty in t, while "Synthetic panel" uses synthetic panel point estimates (Dang et al., 2014). Lower values indicate closer fit.

(Figure A9 for Specification 2). The relative performance is not driven by a single year: across end years, the three methods exhibit errors of comparable magnitude, with the GFE's RMSE generally smaller than that of the synthetic panel and larger than that of the synthetic panel (binary) approach in several years.

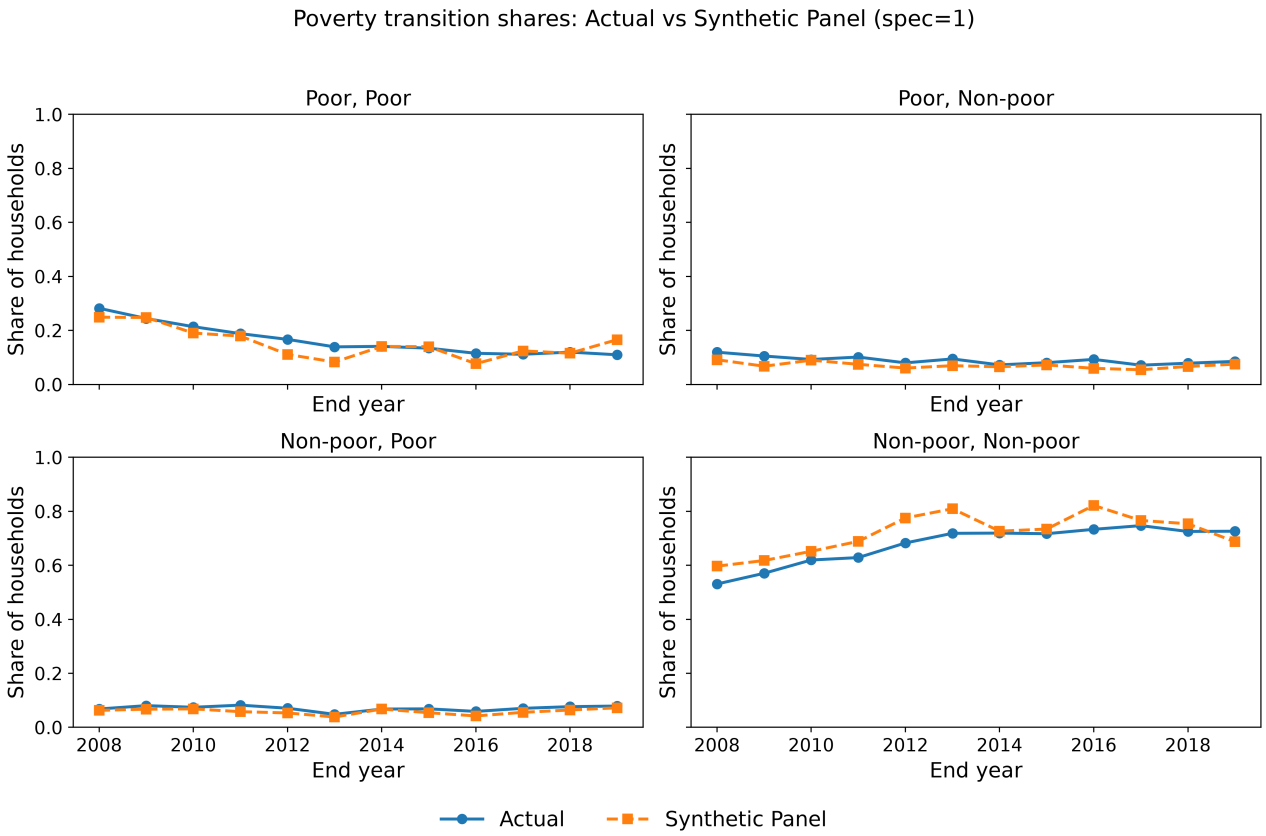
Beyond these fit comparisons, the two approaches are different in what they can deliver for more analysis. The synthetic panel method is designed to recover transition shares from repeated cross-sections, but it does not assign households to latent types and does not produce a full set of predicted welfare outcomes at the household-year level. By contrast, GFE estimates group-specific welfare components $\hat{\alpha}_{g,t}$ and assigns each household to a latent group, which allows us to summarize heterogeneity in welfare dynamics and to construct completed welfare paths for households even when some survey years are missing (under explicit assumptions about how covariates evolve between survey rounds). For researchers that plan and would benefit from describing heterogeneous welfare trajectories and using the estimated latent structure to fill missing outcomes, GFE provides additional structure that is not available in the synthetic panel framework, while achieving comparable (even better) accuracy in reproducing aggregate poverty transitions. In the next Section 4.6, we show the results of using GFE to fill all missing years' information in the rotating panel data.

4.6 GFE for Filling All Missing Years Information

One advantage for the GFE is that we can predict all the missing year's results using the group-year α_{gt} as long as we find ways to fill the missing covariates information. We discussed about the ways that researchers can fill the missing covariates information in Section 2.4.

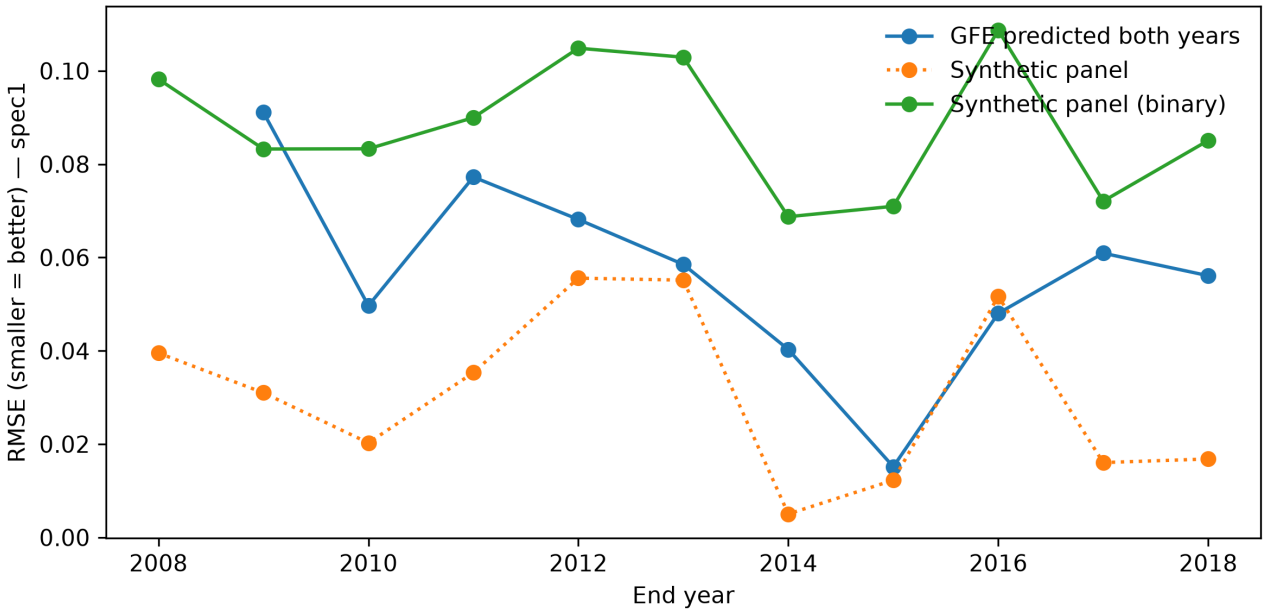
To study welfare dynamics over time, we construct a completed outcome that fills in years when a household is not observed in the survey. For each household i and year t from 2007–2018, we set y_{it}^{comp} equal to the observed welfare measure y_{it} whenever it is available (i.e., $d_{it} = 1$). When y_{it} is missing (i.e., $d_{it} = 0$), we replace it with the GFE-predicted value $\hat{y}_{it} =$

Figure 9: Poverty transition shares: actual vs. synthetic panel predictions



Note: The figures plot the root mean squared error (RMSE) between predicted and observed two-year poverty transition shares for each end year t. For a given end year, the RMSE is computed across the four transition states (poor–poor, poor–non-poor, non-poor–poor, non-poor–non-poor), comparing the predicted transition-share vector to the observed transition-share vector; smaller values indicate a closer fit. "GFE-predicted both years" line shows the RMSE from the GFE-based approach, which uses predicted poverty status in $t-1$ and in t. The other two lines show the RMSE from the synthetic panel methods, which produce transition-share predictions from repeated cross-sections, and a modified binary-version synthetic panel method that is comparable with the GFE-based method. Because the underlying samples used to compute the transition shares may differ across approaches, the figures are intended as a descriptive comparison of fit within each method. Additionally, the GFE RMSE begins in 2009 rather than 2008 because the one-step-ahead validation requires observing each household at both $t-1$ and the last observed year t. Under our analysis sample restriction (at least three observed survey years per household), no household has $t = 2008$ as its last observed year. The sample only includes household heads aged 25 to 55, the same as our sample.

Figure 10: Prediction error in poverty transition shares: GFE versus synthetic panel



Note: The figure plots the root mean squared error (RMSE) between predicted and observed two-year poverty transition shares for each end year t. For a given end year, the RMSE is computed across the four transition states (poor–poor, poor–non-poor, non-poor–poor, non-poor–non-poor), comparing the predicted transition-share vector to the observed transition-share vector; smaller values indicate a closer fit. The solid line shows the GFE approach, which uses predicted poverty status in $t-1$ and t. The dotted line shows the synthetic panel approach, which produces transition-share predictions from repeated cross-sections. Because the underlying samples used to compute the transition shares may differ across approaches, the figure is intended as a descriptive comparison of fit within each method. Additionally, the GFE one-step RMSE begins in 2009 rather than 2008 because the one-step-ahead validation requires observing each household at both $t-1$ and the last observed year t. Under our analysis sample restriction (at least three observed survey years per household), no household has $t = 2008$ as its last observed year.

$x'_{it}\hat{\theta} + \hat{\alpha}_{\hat{g}_i,t} + \hat{\mu}_{p_i}$, which combines (i) the contribution of observed household characteristics, (ii) the estimated group–year component for the household's assigned group, and (iii) province fixed effect estimates.

To compute $\hat{y}_{it}$ in non-survey years, we also need values of the covariates x_{it} in those years. There is no single "correct" way to do this, and researchers should choose an imputation rule that matches their setting and the variables they use. For example, characteristics that are truly fixed (such as gender or language) can be carried forward and backward within household, while other variables may reasonably evolve over time (such as age) and can be updated mechanically. In richer specifications, some covariates may change in ways that are not mechanical (such as education, marital status, or assets), and different assumptions—e.g., holding them constant, carrying forward the last observed value, interpolating between observed values, or using external information—will change the predicted component $x'_{it}\hat{\theta}$ and therefore can shift the level and shape of the group-average paths.

To make this possible in non-survey years in our setting, we fill in missing household characteristics using simple rules: characteristics that are effectively time-invariant in both specifications are carried forward/backward within the same household, and age is updated when missing so that it moves with the calendar year.

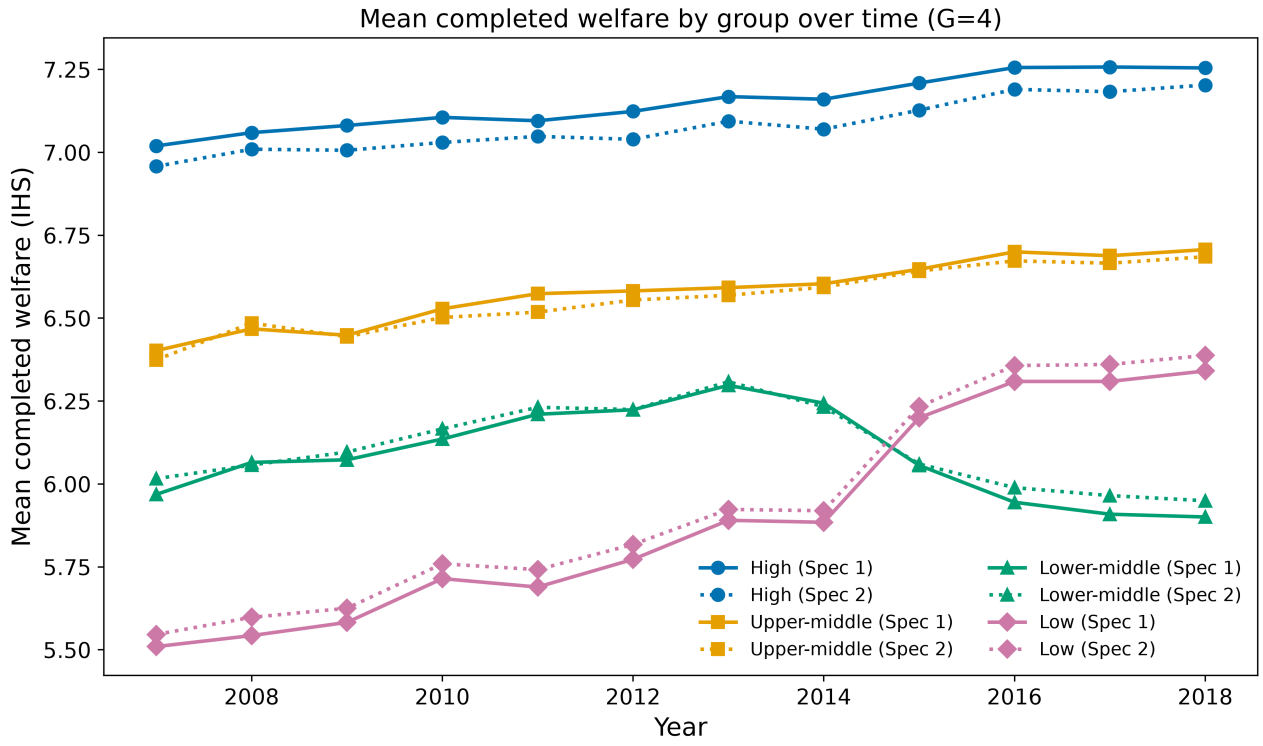
Figure 11 shows the average y_{it}^{comp} by group and year, overlaying Specification 1 and 2. The completed series shows us very clear and stable differences across groups: Group 1 remains the highest throughout the period, Group 2 keep declining in the later years, Group 3 shows an increase in the early years followed by a gradual decline after year 2013, and Group 4 exhibits the largest improvement. The patterns are very similar to our estimated group-year α_{gt} path as shown in Figure 5.

4.7 Group Composition by Baseline Characteristics

To understand the baseline profile by each latent group, we prepared two summary Tables 4 and 5 to summarize group size and baseline characteristics for the selected $G = 4$ model under Specifications 1 and 2, respectively. Baseline characteristics are measured in each household's first observed survey year, and all statistics are weighted using the ENAHO expansion factors so that the reported shares and means are population-representative. To make the group labels interpretable, we order groups by the mean estimated group effect $\bar{\alpha}_g$ and refer to them as *Top*, *Upper-middle*, *Lower-middle*, and *Low*.

The two tables show clear differences in baseline welfare levels across groups. In Specification 1 (Table 4), the Top group accounts for about 12% of the weighted population and has the highest baseline welfare (mean IHS expenditure of 7.24), while the Low group accounts for about 21% and has the lowest baseline welfare (5.56). The middle groups together comprise roughly two-thirds of the population, with baseline welfare levels that fall between these extremes (6.08 for the Lower-middle group and 6.55 for the Upper-middle group). The education profile also varies widely across groups: the share with a college education is much higher in the Top group (36.4%) than in the Low group (9.3%), and the Top group is more likely to report Spanish as

Figure 11: Mean completed welfare by GFE group over time ($G = 4$), overlaying specifications.



Data: All available selected sample (training + test).
Note: The figure plots the mean of the completed welfare measure y_{it}^{comp} by latent group g and year t for 2007–2018. The completed outcome equals observed welfare when available and is imputed using the GFE prediction when missing: $y_{it}^{\text{comp}} = y_{it}$ if observed and $y_{it}^{\text{comp}} = \hat{y}_{it}$ otherwise, where $\hat{y}_{it} = x_{it}'\hat{\theta} + \hat{\alpha}_{\hat{g}_i,t} + \hat{\mu}_{p_i}$. Solid lines correspond to Specification 1 and dotted lines correspond to Specification 2; colors identify groups. Imputation for non-survey years relies on within-household filling rules for covariates (time-invariant characteristics are carried forward/backward within household; age is updated by one year when missing. The welfare measure that we use is the IHS-transformed household expenditure per capita.

the household language (81.7% versus 70.1%), which is important in Peru labor market, where many jobs require Spanish fluency. In contrast, the average age is very similar across groups (around 41 years), suggesting that the group ordering primarily reflects persistent welfare and human capital differences rather than age composition.

Table 5 extends this comparison by adding additional household characteristics. The same broad patterns remain: the Top group is smaller (13% of the population) and has higher baseline welfare (7.15), while the Low group is larger (19%) and has lower baseline welfare (5.62). The expanded set of covariates also reveals meaningful differences in living conditions. In particular, access to basic services is substantially higher in the Top group than in the Low group (water access: 74.6% vs. 55.3%; electricity: 84.4% vs. 78.2%). Language differences are also pronounced (Spanish: 84.6% in the Top group vs. 65.5% in the Bottom group).

Overall, the baseline profiles in both specifications help clarify which kinds of households are classified into each latent group. The ordering by $\bar{\alpha}_g$ aligns closely with a drop in baseline living standards: households in the Top group enter the sample with higher per-capita expenditure, substantially higher schooling attainment, and a higher likelihood of speaking Spanish, while households in the Bottom group start with lower expenditure, much lower rates of college education, and a lower prevalence of Spanish. In Specification 2, this decline extends to housing and local amenities—the Top group is more likely to have access to water and electricity than the Bottom group. These systematic baseline differences suggest that the GFE classification is not arbitrary: it sorts households into groups that are already meaningfully distinct in observable socioeconomic conditions at baseline. At the same time, the groups are not determined by any single observed characteristic alone. For example, average age is very similar across groups, and the middle groups have comparable age profiles but differ in education and baseline welfare. This reinforces the interpretation that the latent grouping captures broader "types" of households defined by a bundle of correlated living-standard indicators, which then map into different latent welfare trajectories over time through the estimated group–year components.

The baseline characteristics by GFE grouping provide researchers very useful information to understand poverty dynamics. It moves the analysis beyond a single snapshot of who is poor in a given year and instead organizes households into a small number of economically meaningful "types" that evolve differently over time. When groups reflect a bundle of correlated living-standard indicators (education, language, and basic services), they provide a transparent way to summarize heterogeneity in both the level of welfare and the path of welfare. In practice, this helps researchers distinguish households that are persistently better-off from those that remain structurally disadvantaged, and it also helps identify intermediate groups that may experience upward mobility or vulnerability to downturns. Because the group–year components trace how each type's latent welfare changes over time, researchers can connect poverty transitions to macroeconomic shocks and policy periods in a more structured way, and can report dynamic patterns (e.g., catch-up, stagnation, or decline) that are difficult to see from year-by-year poverty rates alone. Finally, once households are classified into these types, the framework can be used to describe who is most likely to enter or exit poverty and to target sub-

sequent analysis (and policy discussion) toward the groups whose trajectories suggest persistent deprivation or heightened vulnerability.

Table 4: Group size and baseline characteristics (Specification 1, $G = 4$).

Group	Type (by $\bar{\alpha}_g$)	Households N	Pop. share (%)	IHS exp. pc (baseline)	Female head (%)	Age (years)	Primary edu (%)	Secondary edu (%)	College+ (%)	Spanish (%)
1	Top	2,457	11.89	7.24	19.4	41.33	19.6	31.9	36.4	81.7
2	Upper-middle	5,233	31.88	6.55	15.4	40.80	31.6	37.6	16.2	75.2
3	Lower-middle	3,675	35.61	6.08	19.7	40.77	35.1	34.2	12.0	70.3
4	Low	3,559	20.62	5.56	16.1	40.31	33.7	35.5	9.3	70.1

Notes: Baseline covariates are measured in each household's first observed survey year. All entries are weighted using ENAHO expansion factors. Binary variables are reported as percentages. Groups are ordered by mean estimated group effect ($\bar{\alpha}_g$), from Top to Bottom. Population share is calculated by summing the sample weight for each G and dividing by the sum of weights for the entire sample.

Table 5: Group size and baseline characteristics (Specification 2, $G = 4$).

Group	Type (by $\bar{\alpha}_g$)	Households N	Pop. share (%)	IHS exp. pc (baseline)	Female head (%)	Age (years)	Primary edu (%)	Secondary edu (%)	College+ (%)	Spanish (%)	Married (%)	Urban (%)	Water (%)	Ele (%
1	Top	2,467	12.87	7.15	18.0	41.33	21.5	30.9	34.2	84.6	72.7	71.3	74.6	84.
2	Upper-middle	5,217	32.36	6.45	14.9	40.87	32.4	35.4	15.8	73.7	82.0	69.1	67.7	80.
3	Lower-middle	3,861	35.93	6.09	19.0	40.43	35.0	36.1	11.6	72.7	79.6	73.4	64.1	80.
4	Low	3,379	18.84	5.62	19.1	40.76	31.9	36.6	10.7	65.5	77.2	71.1	55.3	78.

Notes: Baseline covariates are measured in each household's first observed survey year. All entries are weighted using ENAHO expansion factors. Binary variables are reported as percentages. Groups are ordered by mean estimated group effect ($\bar{\alpha}_g$), from Top to Bottom. Population share is calculated by summing the sample weight for each G and dividing by the sum of weights for the entire sample.

5 Robustness Check

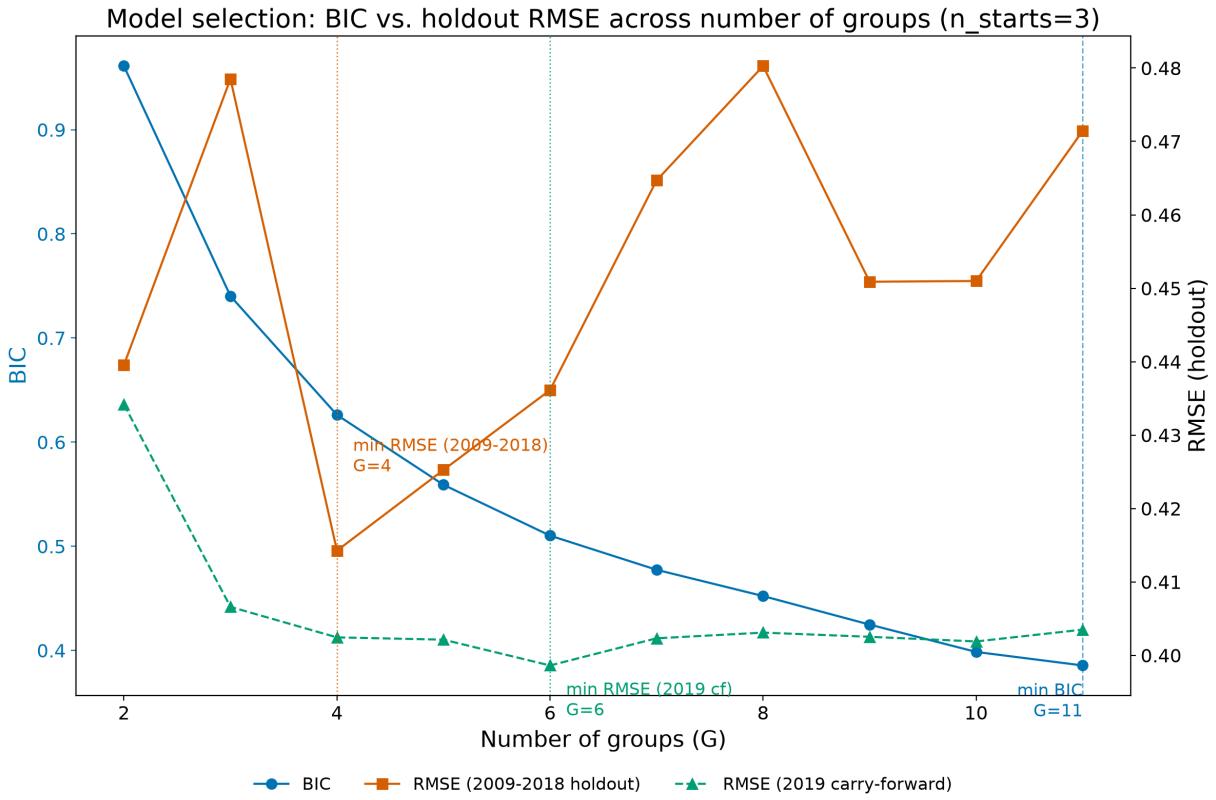
5.1 No Age Restriction

As the first robustness check, we re-estimate the GFE model on a broader sample that does not impose any age restriction on the household head for the ENAHO data. We keep the same rotating-panel requirement used in the main analysis (households observed in at least three survey years) and the same estimation and evaluation set up. We end up having a sample of 104,304 household-year observation from 27,034 unique households. The average age of head of households for the sample is around 52, with a standard deviation of 15 based on the household head's first survey year age. Approximately 31% of household heads are aged 60 or older.

We first run the GFE model with a grid over $G = 1, \ldots, 11$ with `n_starts` $= 3$ random initialization for each G. Similarly to Figure 3 that shows the change in the BIC and RMSE values between different latent groups G using under specification 1, our result with the sample of no age restriction (Figure 12) continue to favor a small number of latent groups: the holdout RMSE is minimized at $G = 4$ for the 2009-2018 prediction and at $G = 6$ for the 2019 prediction, while the BIC decreases mechanically with increase of G and thus prefer much larger G. After rerunning the model with a higher random initialization (`n_starts` $= 10$) for the six Gs that have the lowest RMSE, the model still shows that $G = 4$ yields the lowest RMSE for 2009-2018 prediction, and $G = 6$ for the 2019 prediction (Figure 13). Overall, dropping the age restriction does not materially affect the model-selection conclusion: the preferred number of latent groups remains as $G = 4$. This suggests that the main structural patterns identified by the GFE model are not driven by the head-age restriction. Expanding the sample changes household composition but does not alter the degree of latent heterogeneity favored by the out-of-sample fit.

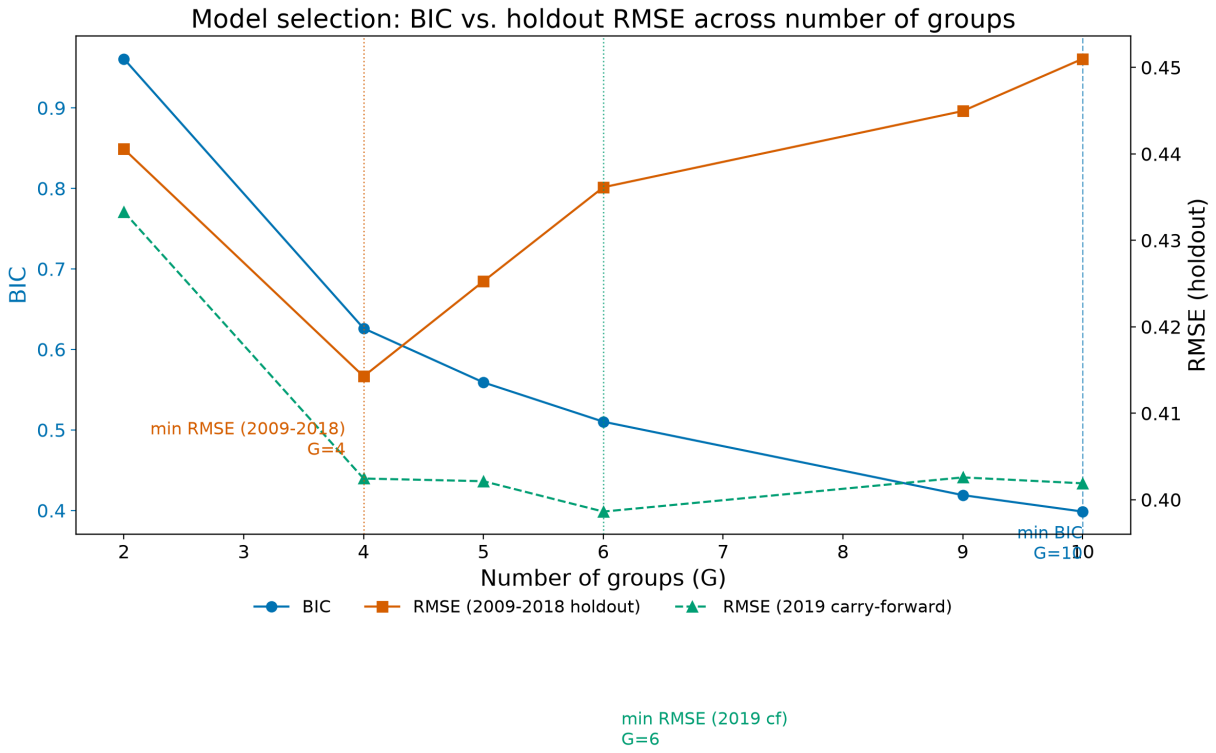
Figure 14 reproduces the Figure 5 exercise using the sample without any head-age restriction, plotting the estimated group-year intercepts $\hat{\alpha}_{gt}$ for $G = 4$ under Specification 1 and 2. The qualitative welfare dynamics remain very similar to the sample used for the main results. We can see a clear and persistent separation between the top and bottom groups throughout the period. Importantly, the estimated $\hat{\alpha}_{gt}$ for $G = 4$ paths are closer to each other with Specifications 1 and 2 (solid vs. dotted lines) than that of the results in the main results (Figure 5). A natural explanation is that removing the age restriction increases the effective sample size and improves support within each group-year cell, thereby stabilizing the estimation of group-year intercepts $\hat{\alpha}_{gt}$. However, one major difference is that while different specifications yield almost identical dynamics across all groups, removing age restrictions yields different poverty patterns between the "Lower-middle" and "Low" groups after 2013. A possible interpretation is that households whose head is either very young (25 or younger) or very old (55 or older) experience rapid changes in household characteristics that are present only in Specification 2. For instance, younger household heads during the period may be more likely to marry or migrate to an urban area, which deviates their poverty patterns from those of the model with more conservative covariates. These findings suggest that estimating poverty status may be more

Figure 12: BIC vs. RMSE over G - specification 1 (no age restriction), `n_starts` = 3



Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years but does not apply age restrictions for household heads.
Note: The figure plots model selection diagnostics for the GFE estimator across initial numbers of groups $G \in \{1, \cdots, 11\}$ with `n_starts` = 3, `itermax` = 10, `neighmax` = 5, and `max_local_iters` = 2 under specification 1. BIC is computed on the training data; RMSE is computed out-of-sample using the last observed household-year held out for each household (years 2009–2018 and year 2019) test data. Lower values indicate better fit/prediction.

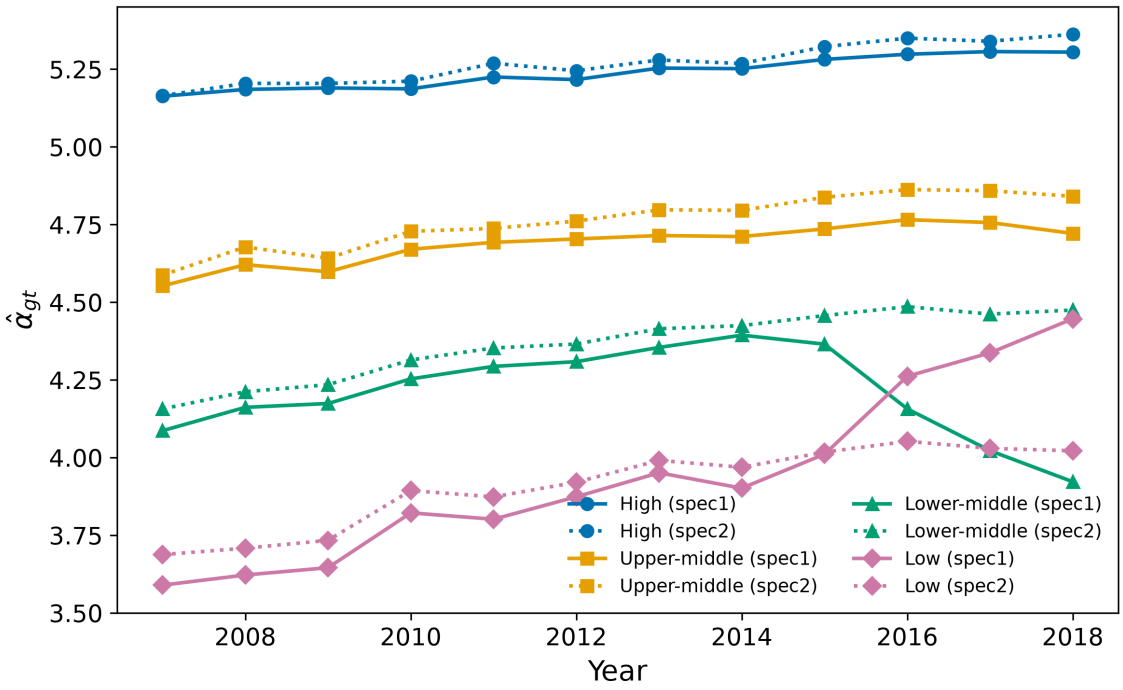
Figure 13: BIC vs. RMSE over G - specification 1 (no age restriction), `n_starts` = 10



Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years but does not apply age restrictions for household heads.
Note: The figure plots model selection diagnostics for the GFE estimator across candidate numbers of groups G under two covariate specifications (spec1 and spec2) with `n_starts` = 10, `itermax` = 10, `neighmax` = 5, and `max_local_iters` = 2. BIC is computed on the training data; RMSE is computed out-of-sample using the last observed household-year held out for each household (years 2009–2018) test data. Lower values indicate better fit/prediction.

difficult for younger or older generations, whose characteristics are more likely to change. More broadly, the results highlight a practical trade-off faced by researchers working with rotating panels: including households that are more likely to experience covariate changes increases within-household information but can make the estimated trajectories more sensitive to model specification choices.

Figure 14: Estimated group–year intercept paths $\hat{\alpha}_{gt}$ for $G = 4$ (no age restriction)



Data: Estimated group-year intercept paths for G=4 using training data of sample without age restriction.
Note: Colors indicate groups. Solid lines are specification 1 and dotted lines are specification 2. $\hat{\alpha}_{gt}$ is the estimated group×year intercept from the GFE model (net of covariates). Interpretation should focuses on relative differences and dynamics across groups rather than the value.

Table 6 replicates Table 3, summarizing the fit between the GFE-based and synthetic panel methods. Consistent with the main results, the GFE-based method yields higher RMSE than that of the original synthetic panel method, but lower RMSE than that of the binary-predicted synthetic panel method, which is equivalent to the GFE-based method. Appendix Figure A10 plot the end-year RMSE in the four transition shares for the GFE one-step approach and the synthetic panel point estimates under Specification 1 and 2. Similarly to our conclusion in the main results (Figure 10 and Figure A9), the relative performance is not driven by a single year and the GFE one-step RMSEs are generally smaller than the synthetic panel approach.

Tables 7 and 8 replicate the baseline group-profile exercise using the sample without age restriction for Specification 1 and 2, respectively. The resulting group composition and baseline characteristics are consistent with Tables 4 and 5 in the main results. The Top group has the highest baseline IHS expenditure per capita and the largest share with college education, while the Low group has the lowest baseline welfare and substantially lower educational attainment. The Top group is more likely to have access to water and electricity than the Bottom group in

Table 6: Fit of predicted vs. observed poverty transition shares using data without age restriction (lower values indicate closer match).

dataset	spec	MAE_avg	$RMSE\_avg$	TV_avg	TV_max
GFE predicted both years	1	0.052	0.055	0.104	0.163
GFE predicted both years	2	0.060	0.063	0.120	0.161
Synthetic panel	1	0.026	0.031	0.052	0.093
Synthetic panel	2	0.023	0.028	0.047	0.090
Synthetic panel (binary)	1	0.075	0.088	0.150	0.188
Synthetic panel (binary)	2	0.071	0.084	0.142	0.176

Notes: The table compares predicted and observed two-year poverty transition shares across the four transition states (poor–poor, poor–non-poor, non-poor–poor, and non-poor–non-poor). MAE and RMSE are averaged across the four states within each end year. TV denotes the total variation distance, $TV_t = \frac{1}{2}\sum_s |\hat{p}_{s,t} - p_{s,t}|$, computed for each end year t and then averaged (TV_avg) or maximized (TV_max) over years. "GFE one-step" uses observed poverty status in t–1 and predicts poverty in t, while "Synthetic panel" uses synthetic panel point estimates (Dang et al., 2014). Lower values indicate closer fit.

Specification 2. Overall, dropping the age restriction does not change the interpretation of the latent groups as distinct welfare types.

Table 7: Group size and baseline characteristics using sample without age restriction (Specification 1, $G = 4$).

Group	Type (by $\bar{\alpha}_g$)	Households N	Pop. share (%)	IHS exp. pc (baseline)	Female head (%)	Age (years)	Primary edu (%)	Secondary edu (%)	College+ (%)	Spanish (%)
1	Top	4,523	13.32	7.32	27.7	53.65	24.5	25.5	28.9	84.0
2	Upper-middle	9,518	32.40	6.62	21.9	51.03	29.8	29.8	15.8	75.4
3	Lower-middle	7,223	35.07	6.07	21.7	49.05	32.4	27.6	9.1	66.2
4	Low	5,791	19.20	5.52	22.7	49.78	29.9	25.3	8.3	61.8

Notes: Baseline covariates are measured in each household's first observed survey year. All entries are weighted using ENAHO expansion factors. Binary variables are reported as percentages. Groups are ordered by mean estimated group effect ($\bar{\alpha}_g$), from Top to Bottom. Population share is calculated by summing the sample weight for each G and dividing by the sum of weights for the entire sample.

Table 8: Group size and baseline characteristics using sample without age restriction (Specification 2, $G = 4$).

Group	Type (by $\bar{\alpha}_g$)	Households N	Pop. share (%)	IHS exp. pc (baseline)	Female head (%)	Age (years)	Primary edu (%)	Secondary edu (%)	College+ (%)	Spanish (%)	Married (%)	Urban (%)	Water (%)	Ele (%
1	Top	4,041	13.24	7.28	23.1	53.12	21.0	25.8	31.6	84.4	68.7	72.2	78.7	84.
2	Upper-middle	8,947	31.93	6.57	22.0	50.69	31.8	29.0	15.1	74.0	72.2	70.9	73.3	82.
3	Lower-middle	9,389	35.84	6.08	21.7	49.13	31.6	26.7	8.8	67.9	74.4	70.0	64.8	78.
4	Low	4,678	18.99	5.64	25.8	50.66	30.5	28.2	8.4	60.9	67.7	71.4	57.6	75.

Notes: Baseline covariates are measured in each household's first observed survey year. All entries are weighted using ENAHO expansion factors. Binary variables are reported as percentages. Groups are ordered by mean estimated group effect ($\bar{\alpha}_g$), from Top to Bottom. Population share is calculated by summing the sample weight for each G and dividing by the sum of weights for the entire sample.

5.2 Household Surveyed At Least Five Years

We also re-estimate the GFE model that apply the age restriction but keep only households surveyed at least five years in the ENAHO data. This restriction gives us a sample of 15,529 household-year observation from 2,946 unique households, which only account for around 7.8 % (2,946 / 375,897) of unique households in the sample that we use for our main analysis, a significant sample decreases as compared to the sample used in the main analysis.

After running the GFE model with a grid over $G = 1, \ldots, 11$ with `n_starts` $= 3$ random initialization for each G, and then select the best six Gs that with the lowest RMSEs to re-run the model with `n_starts` $= 10$, Figure 15 shows that the best G with the overall lowest RMSE is still 4. This result is consistent with our main result.

Figure 16 reproduces the Figure 5 exercise using the sample with households surveyed for at least five years, plotting the estimated group-year intercepts $\hat{\alpha}_{gt}$ for $G = 4$ under Specifications 1 and 2. Relative to the sample we used in our main results, the two specifications produce slightly different poverty trajectories between specifications (solid vs. dotted lines), but their patterns are consistent over the period. This consistency implies that while households surveyed over a longer period may differ in overall poverty rates, their patterns are not systematically different from those of households surveyed over a shorter period, maintaining sample representativeness.

5.3 Medium-run Poverty Prediction

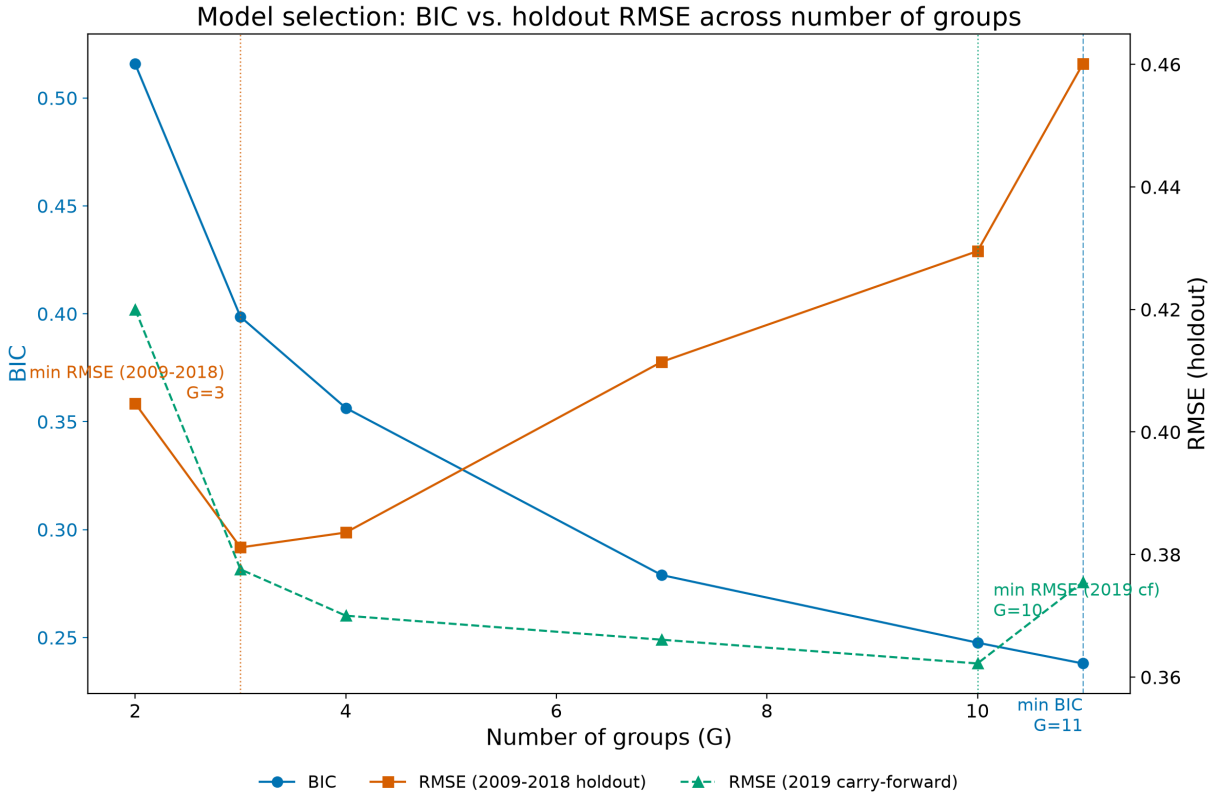
As a robustness check, we restrict the sample to households surveyed for at least four years and reserve the final two observed years for a validation. We therefore estimate the GFE model using the earlier observed years. This provides us an opportunity to demonstrate how well the model is in predicting medium-run outcomes based on the alpha trajectories since we emphasize earlier that one of the advantages of the GFE model is its ability to fill all the missing years information. This restriction gives us a sample of 35,693 household-year observation from 7,987 unique households.

Table 9 reports weighted poverty classification results using households observed for at least four years. The final two observed years are reserved for testing, and the GFE model is estimated using only earlier observed years. We can tell that poverty classification accuracy remains high throughout the held-out period for both specifications. Averaging across all eligible held-out observations, the weighted classification accuracy is 79.9% for Specification 1 and 83.29% for Specification 2. Year-specific accuracy generally lies between about 75 and 88 percent. These prediction accuracies suggest that the model captures medium-run poverty reasonably well.

5.4 First year as Holdout

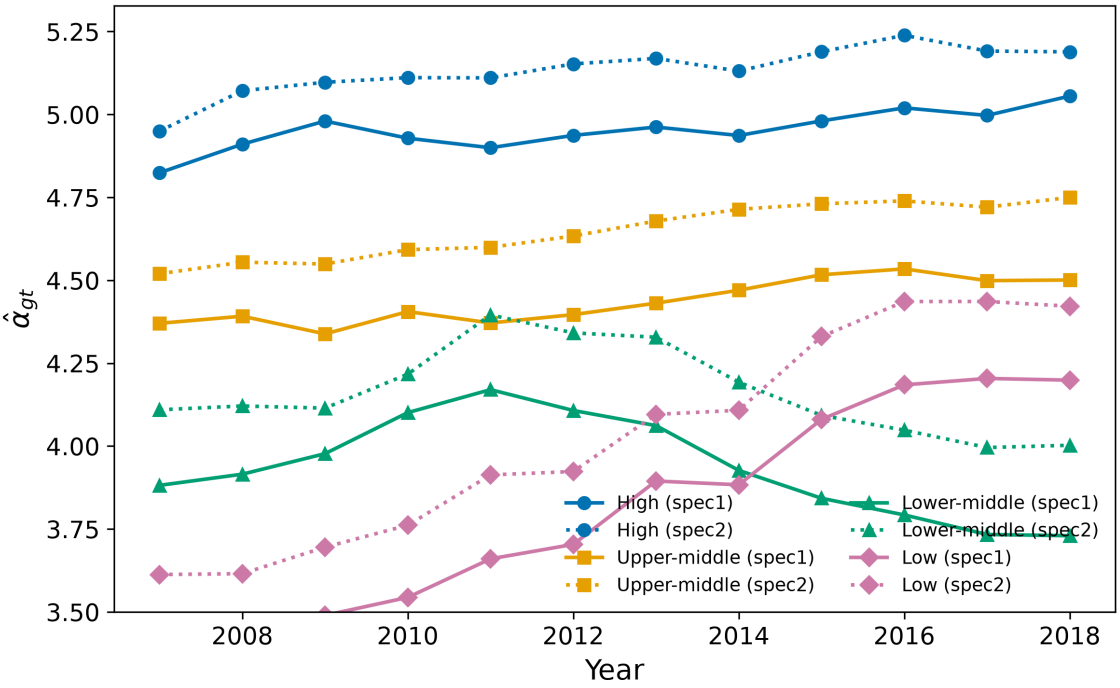
We replicate our analysis by holding out the first year surveyed as the test set, rather than the last year surveyed in our main analysis. For example, for a household surveyed from 2007 to 2009, we use 2008-2009 as the training set and 2007 as the test set. Therefore, all

Figure 15: BIC vs. RMSE by different sets of covariates for selected G (households surveyed at least five years)



Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least five survey years and to household heads aged 25-55 in all observed years.
Note: The figure plots model selection diagnostics for the GFE estimator across candidate numbers of groups G under two covariate specifications (spec1 and spec2) with `n_starts` = 10, `itermax` = 10, `neighmax` = 5, and `max_local_iters` = 2. BIC is computed on the training data; RMSE is computed out-of-sample using the last observed household-year held out for each household (years 2009–2018) test data. Lower values indicate better fit/prediction.

Figure 16: Estimated group–year intercept paths $\hat{\alpha}_{gt}$ for $G = 4$ (households surveyed at least five years)



Data: Estimated group-year intercept paths for G=4 using training data of households stayed in the survey for at least five years.
Note: Colors indicate groups. Solid lines are specification 1 and dotted lines are specification 2. $\hat{\alpha}_{gt}$ is the estimated group×year intercept from the GFE model (net of covariates). Interpretation should focuses on relative differences and dynamics across groups rather than the value.

Table 9: Poverty classification accuracy in held-out years

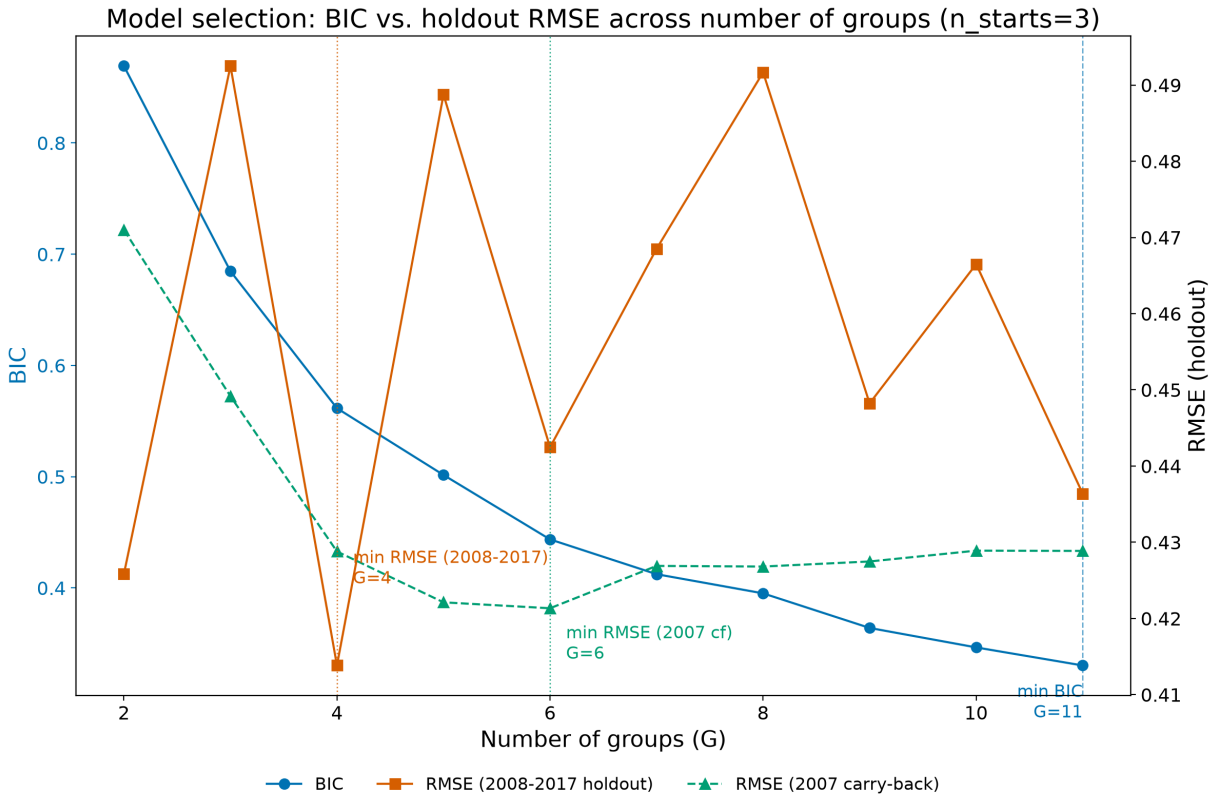
Specification	Year	Households	Obs.	Actual poverty rate (%)	Predicted poverty rate (%)	Classification accuracy (%)
Spec 1	2009	856	856	33.99	32.40	81.63
	2010	1,391	1,391	30.66	25.34	80.26
	2011	674	674	29.34	24.04	78.12
	2012	1,054	1,054	29.10	25.11	80.81
	2013	1,110	1,110	20.03	17.05	83.97
	2014	1,246	1,246	18.00	13.63	86.07
	2015	2,220	2,220	21.37	13.34	75.41
	2016	1,239	1,239	16.17	13.53	78.11
	2017	1,101	1,101	17.39	15.60	78.38
	Overall	5,928	10,891	23.48	19.00	79.92
Spec 2	2009	856	856	33.99	29.80	83.36
	2010	1,391	1,391	30.66	22.92	79.22
	2011	674	674	29.34	20.49	79.65
	2012	1,054	1,054	29.10	25.49	80.94
	2013	1,110	1,110	20.03	17.80	82.95
	2014	1,246	1,246	18.00	16.11	87.51
	2015	2,220	2,220	21.37	16.10	84.28
	2016	1,239	1,239	16.17	14.45	85.57
	2017	1,101	1,101	17.39	16.43	83.64
	Overall	5,928	10,891	23.48	19.37	83.22

Notes: For each held-out household-year observation within the supported period of the estimated group-time effects, poverty status is predicted and compared with realized poverty status. Classification accuracy is the weighted share of correctly classified poor/non-poor observations, using ENAHO household survey weights.

households and their survey observations were identical to those in the main analysis, except for the training set and test set. This check investigates whether group assignment and poverty trends are robust to the hold-out year.

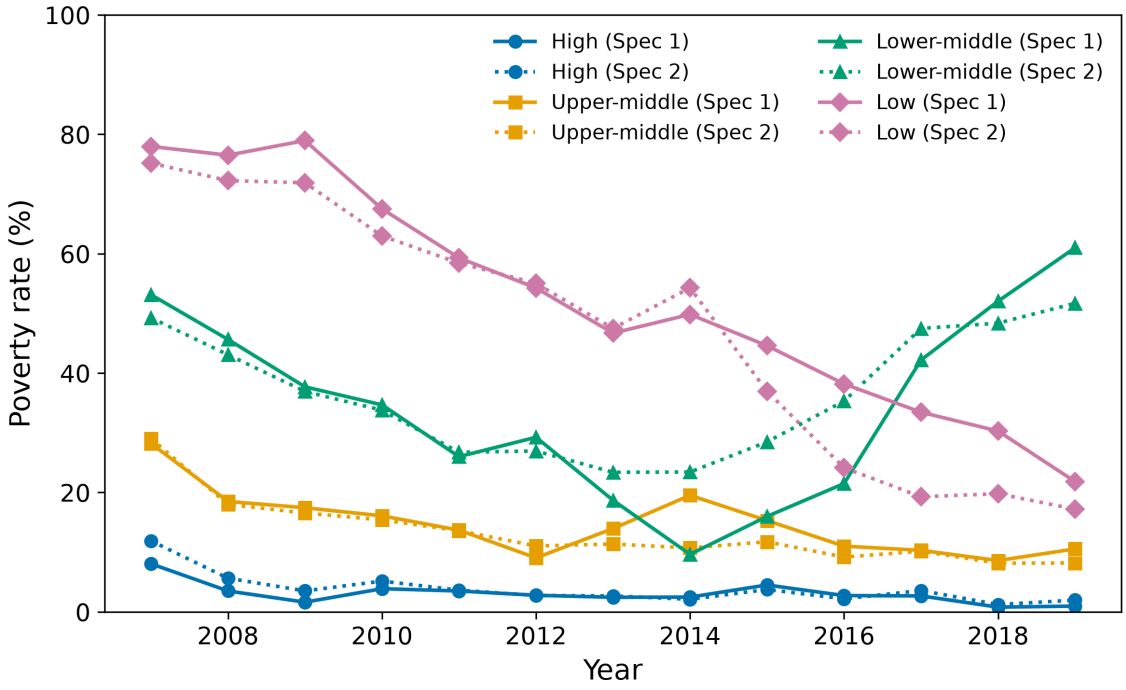
Figure 17 replicates Figure 3. RMSE is the lowest when G=4 (2008-2017) and G=6 (2007). While G=6 is marginally better in the 2007 prediction, G=4 is significantly better in the 2008-2017 prediction. Overall, G=4 remains the best performance. Figure 18 replicates Figure 6. The Top and Upper-middle groups remain stable, while the lower-middle and low groups show some differences in patterns across specifications. Specification 2 shows nearly identical trajectories to those in the main analysis, but specification 1 shows slight divergence after 2013. Regarding group assignment stability, 61% (68%) of households were assigned to the same group when the 1st year was held out using specification 1 (specification 2). On the one hand, group assignments were relatively robust for those in the Top and Upper-middle groups; 80% (62%) of households in the Top (Upper-middle) group in the main analysis remained in the same group in the 1st-year held-out setting. On the other hand, group assignments were less robust for the other two groups; 58% (49%) of the households in the Low (Lower-middle) group in the main analysis remained the same group in the 1st-year held-out setting. These results imply that while the selection of train and test data does not make a major difference in national-level projection, unit-level group assignment is more sensitive to the selection.

Figure 17: Model fit (BIC and RMSE) under specification 1 - 1st year hold-out (`n_starts` = 3)



Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years and to household heads aged 25–55 in all observed years.
Note: The figure plots model selection diagnostics for the GFE estimator across initial numbers of groups $G \in \{1, \cdots, 11\}$ with `n_starts` = 3, `itermax` = 10, `neighmax` = 5, and `max_local_iters` = 2, predicted on two different periods (2008-2017, 2007). BIC is computed on the training data; RMSE is computed out-of-sample using the last observed household-year held out for each household (years 2008–2017 or 2007) test data. Lower values indicate better fit/prediction.

Figure 18: Poverty rate trends by GFE group (G=4): Specification 1 vs. Specification 2. 1st year holdout



Data: All available selected sample (training + test).
Note: Values are survey-weighted poverty rates by group-year (ENAHO expansion weights). Solid lines are for specification 1 and dotted lines are for specification 2. Colors denote the same group across specifications. Groups use household-level GFE assignments from the selected G=4 model.

6 Discussion and Implications

This paper implements a GFE approach to measuring poverty dynamics in rotating panel surveys, where households are observed for only a few years but there is partial overlap across waves. Using Peru's ENAHO rotating panel as an application, we show that the GFE framework can leverage short-panel overlap to recover interpretable latent welfare trajectories and to construct poverty transition measures over a long period of time. The model classifies households into a number of latent types and estimates group–year welfare trajectories that summarize how these types evolve over time, after accounting for observed covariates and province fixed effects.

We show that empirically, the estimated groups corresponding to meaningful baseline differences in living conditions: higher-ranked groups have higher baseline expenditure, higher education level, and better access to water and electricity, while lower-ranked groups are disadvantaged in multiple margins. The estimated group–year trajectories show distinct welfare paths over the periods of 2007–2018, and these patterns are robust in two different covariate specifications we have. In addition, the model-implied poverty rates are closely aligned with the true poverty rates in the test data, and our one-step-ahead validation shows that GFE can reproduce observed two-year transition shares in out-of-sample end years very nicely.

Our findings show the advantage of applying GFE to rotating panels which contains more information about welfare and poverty dynamics than repeated cross-section data. An important advantage of GFE is that it provides us two outputs that are both useful for applied work: (i) a summary of heterogeneous welfare dynamics through a number of group trajectories, and (ii) a very practical group-year parameter for constructing completed welfare paths in years when households are not observed, conditional on researchers makes transparent assumptions about how covariates evolve overtime. This can be very helpful when the researchers want to understand and describe medium to long run poverty persistence and mobility and to relate these patterns to policy changes.

One caveat is important for the interpretation. The completed welfare paths rely on assumptions about missing covariates in non-survey years. In our application, we use a simple completion rule (forward/backward filling for time-invariant characteristics and deterministic updating for age), but other reasonable choices are possible, and they can change the level and slope of the completed welfare trajectories. We therefore recommend that researchers treat covariate completion as part of the empirical design and assess robustness to alternative imputation rules whenever completed panels are used for substantive conclusions.

A second practical lesson concerns sample design in rotating panels. Our robustness checks show that removing the household head's age restriction leaves the optimal number of latent group selection and preferred group structure conclusion unchanged, consistent with the idea that richer group-year support stabilizes $\hat{\alpha}_{gt}$ and reduces sensitivity to the covariate set. By contrast, restricting samples to households observed for at least 5 years sharply reduces the effective sample size and yields greater divergence in estimated trajectories across specifications, even though the RMSE-based choice of optimal G remains the same. This divergence may not

be an issue for most rotating panel designs, where the same unit is tracked for up to 2 years (e.g., the Current Population Survey (CPS) and the Vietnam Household Living Standards Survey (VHLSS)). However, for a rotating panel design where the same units are tracked more than two rounds, our robustness check suggests that maintaining coverage with shorter household participation can be more valuable for having stable trajectory inference than for focusing on the small subset of long-duration households.

While our implementation follows the discrete GFE framework of BM2015, extensions that allow for richer forms of heterogeneity may be valuable in some settings, but the discrete grouping approach has the practical benefit of transparency and interpretability, which is central for describing poverty dynamics in applied work.

Looking forward, our findings suggest that GFE can serve as a practical bridge between short rotating panels and the long-run welfare dynamics that researchers and policymakers often cares about. By combining interpretable latent types with a transparent completion strategy to fill all the missing years dynamics, the approach can produce transition predictions that align closely with observed mobility patterns. Our results motivate the use of GFE model for studying poverty dynamics when true long panels are unavailable, and they point the value of developing diagnostics and model-selection criteria that explicitly account for cell support, imputation assumptions, and out-of-sample performance.

References

Aparicio-Pérez, D. and Ripollés, J. (2025). Disentangling the heterogeneous effect of natural resources on economic growth. *Economic Modelling*, 142:106927.

Attanasio, O. and Székely, M. (2001). Going beyond income: Redefining poverty in latin america. In *Portrait of the Poor: An Assets-Based Approach*, pages 1–43. Inter-American Development Bank.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bailar, B. A. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*, 70(349):23–30.

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

Bonhomme, S., Lamadon, T., and Manresa, E. (2022). Discretizing unobserved heterogeneity. *Econometrica*, 90(2):625–643.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Carter, M. R. and Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: An asset-based approach. *Journal of Development Studies*, 42(2):178–199.

Cruces, G., Lanjouw, P., Lucchetti, L., Perova, E., Vakis, R., and Viollaz, M. (2015). Estimating poverty transitions using repeated cross-sections: a three-country validation exercise. *The Journal of Economic Inequality*, 13(2):161–179.

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

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30(1):109–126.

D'Alberto, R., De Nicolò, S., and Gardini, A. (2025). Estimating poverty transitions from repeated cross-sections: A statistical perspective. *Social Indicators Research*, 179(3):1143–1164.

Hill, T. D., Davis, A. P., Roos, J. M., and French, M. T. (2020). Limitations of fixed-effects models for panel data. *Sociological Perspectives*, 63(3):357–369.

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

INEI (2017). Peru final version of OECD review of official statistics and statistical system.

Janys, L. and Siflinger, B. (2024). Mental health and abortions among young women: time-varying unobserved heterogeneity, health behaviors, and risky decisions. *Journal of Econometrics*, 238(1):105580.

Lucchetti, L., Corral, P., Ham, A., and 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.

Millimet, D. L. and Bellemare, M. F. (2025). On the (mis) use of the fixed effects estimator. *Oxford Bulletin of Economics and Statistics*.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85.

Murillo, D. and Sardon, S. (2024). Commodity booms, local state capacity, and development.

Oberlander, L., Disdier, A.-C., and Etilé, F. (2017). Globalisation and national trends in nutrition and health: A grouped fixed-effects approach to intercountry heterogeneity. *Health Economics*, 26(9):1146–1161.

Oh, D. H. and Patton, A. J. (2023). Dynamic factor copula models with estimated cluster assignments. *Journal of Econometrics*, 237(2):105374.

Pigini, C., Pionati, A., and Valentini, F. (2025). Grouped fixed effects regularization for binary choice models. *arXiv*, (arXiv:2502.06446).

Van den Brakel, J. A. and Krieg, S. (2015). Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design. *Survey Methodology*, 41(2):267–297.

Appendix

Appendix A Pseudocode

Algorithm: GFE Estimation via VNS + Local Search

Input : $\{y_{it}, x_{it}, d_{it}\}$, number of groups G to explore, tuning: `n_starts`, `neighmax`, `itermax`, `max_local_iters`
Output: $(\hat{\theta}, \hat{\alpha}, \hat{\gamma}, \hat{\mu})$

Function $Q(\theta, \alpha, \gamma, \mu)$
  return $\sum_{i,t} d_{it}(y_{it} - x'_{it}\theta - \alpha_{g_i t} - \mu_{p_i})^2$

Function *Assign*(θ, α, μ)
  For each i: $g_i \leftarrow \arg\min_g \sum_t d_{it}(y_{it} - x'_{it}\theta - \alpha_{gt} - \mu_{p_i})^2$;
  return γ

Function *OLSUpdate*(γ)
  OLS of y_{it} on x_{it} and group×time dummies and province fixed effects using $d_{it} = 1$;
  return (θ, α, μ)

Function *LocalSearch*$(\gamma, \theta, \alpha, \mu,$ *max_local_iters*$)$
  `// Greedy single-unit moves holding` (θ, α, μ) `fixed`
  perform at most `max_local_iters` greedy reassignment passes, stopping earlier if no improving move exists
  return γ

$Q^* \leftarrow +\infty$
for $r = 1$ **to** *n_starts* **do**
  `// Step 1: Initialization`
  $\theta^{(0)}, \mu^{(0)} \leftarrow$ pooled OLS using $d_{it} = 1$
  $\alpha_t^{(0)} \leftarrow \overline{y_{it} - x'_{it}\theta^{(0)} - \mu_{p(i)}^{(0)}}$ (means over $d_{it} = 1$)
  $\alpha_{gt}^{(0)} \leftarrow \alpha_t^{(0)}$ for all g, t
  $\gamma \leftarrow \text{Assign}(\theta^{(0)}, \alpha^{(0)}, \mu^{(0)})$
  $(\theta, \alpha, \mu) \leftarrow \text{OLSUpdate}(\gamma)$
  $\gamma^\star \leftarrow \gamma$;
  $(\theta^\star, \alpha^\star, \mu^\star) \leftarrow (\theta, \alpha, \mu)$;
  $Q_r^\star \leftarrow Q(\theta^\star, \alpha^\star, \gamma^\star, \mu^\star)$
  `// Outer VNS cycles`
  for $j = 1$ **to** *itermax* **do**
    `// Step 2: Start from neighborhood size` $n = 1$ `each cycle`
    $n \leftarrow 1$
    while $n \le$ *neighmax* **do**
      `// Step 3: Neighborhood jump ("shake") from current best` $\gamma^\star$
      Select random set $\mathcal{J}_n$ with $|\mathcal{J}_n| = n$
      $\gamma' \leftarrow \gamma^\star$
      Reassign each $i \in \mathcal{J}_n$ to a uniformly random group in $\{1, \dots, G\}$
      `// Step 4: Refinement after the neighborhood jump`
      $(\theta', \alpha', \mu') \leftarrow \text{OLSUpdate}(\gamma')$
      `// Step 5: Greedy local search over single-unit moves`
      $\gamma'' \leftarrow \text{LocalSearch}(\gamma', \theta', \alpha', \mu',$ `max_local_iters`$)$
      $(\theta'', \alpha'', \mu'') \leftarrow \text{OLSUpdate}(\gamma'')$
      $Q'' \leftarrow Q(\theta'', \alpha'', \gamma'', \mu'')$
      `// Step 6: Acceptance rule`
      if $Q'' < Q_r^\star$ **then**
        $\gamma^\star \leftarrow \gamma''$;
        $(\theta^\star, \alpha^\star, \mu^\star) \leftarrow (\theta'', \alpha'', \mu'')$;
        $Q_r^\star \leftarrow Q''$
        $n \leftarrow 1$
        `// restart neighborhood search around improved solution`
      else
        $n \leftarrow n + 1$
        `// Step 7: increase neighborhood size`
  `// Keep best start`
  if $Q_r^\star < Q^*$ **then**
    $(\hat{\theta}, \hat{\alpha}, \hat{\gamma}, \hat{\mu}) \leftarrow (\theta^\star, \alpha^\star, \gamma^\star, \mu^\star)$;
    $Q^* \leftarrow Q_r^\star$
return $(\hat{\theta}, \hat{\alpha}, \hat{\gamma}, \hat{\mu})$

Appendix B Tables

Table A1: Rotating Panel Design Surveys

Country	Name	Rotation Design
Argentina	Permanent Household Survey (EPH)	2-2-2 rotation scheme: Households are interviewed for two consecutive quarters, dropped for the next two, and interviewed again for the following two.
European Union	EU-SILC (European Union Statistics on Income and Living Conditions)	Each year, 25% of the sample is rotated out and replaced by a new sub-sample.
Kyrgyz Republic	Kyrgyz Integrated Household Survey (KIHS)	25% of the sampled households are replaced each year.
Mexico	National Survey of Occupation and Employment (ENOE)	5-quarter rotating panel where 20% of the sample is replaced each quarter.
Peru	National Household Survey (ENAHO)	Each year, 30% of households are selected as panel households, surveyed from 2 to 5 years.
South Africa	Quarterly Labour Force Survey (QLFS)	One-quarter of the sample being refreshed each quarter.
United States	Current Population Survey (CPS)	4-8-4: Households are interviewed for 4 consecutive months, excluded for 8, and re-interviewed for 4 more before being retired. This means 50% of households are in the survey during the same month one year earlier.
Vietnam	Vietnam Household Living Standards Surveys (VHLSS)	50% of households are retained, 50% are newly surveyed

Table A2: Household Total Expenditure Per Capita Estimates

	(1)	(2)
Head of Household - Female	0.1542***	-0.1063***
	0.0123	0.0167
Head of Household - Age	-0.0244***	-0.0175***
	0.0068	0.0061
Head of Household - Age squared	0.0004***	0.0003***
	0.0001	0.0001
Head of Household - Primary school	0.1910***	0.1287***
	0.0125	0.0114
Head of Household - Secondary school	0.4579***	0.3147***
	0.0142	0.0135
Head of Household - College	0.8820***	0.6995***
	0.0182	0.0177
Language - Spanish	0.1911***	0.1368***
	0.0128	0.0119
Head of Household - Married or Cohabitant		-0.3141***
		0.0163
Urban		0.3018***
		0.0121
Has water		0.1257***
		0.0105
Has electricity		0.1776***
		0.0158
R-squared	0.443	0.508
N	56390	56390

Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years and to household heads aged 25–55 in all observed years.
Note: ***, **, * indicate significance at the 1%, 5%, and 10% levels. The dependent variable is monthly household total expenditures per capita (2007 dollars), inverse hyperbolic sine transformed. The sample includes households observed at least three times and with head age 25–55 in each survey year, 2007–2019. All regressions include department (Peru region) and year fixed effects and are weighted by the annual expansion factor (CPV-2007 projections). Standard errors (in parentheses) are clustered at the household level.

Table A3: Household Total Expenditure Per Capita Estimates

	(1)	(2)
Head of Household - Female	0.0967***	-0.1020***
	0.0116	0.0149
Head of Household - Age	-0.0179***	-0.0111*
	0.0062	0.0058
Head of Household - Age squared	0.0003***	0.0002***
	0.0001	0.0001
Head of Household - Primary school	0.1202***	0.0977***
	0.0113	0.0107
Head of Household - Secondary school	0.3106***	0.2582***
	0.0133	0.0128
Head of Household - College	0.6621***	0.5898***
	0.0167	0.0161
Language - Spanish	0.0994***	0.0870***
	0.0130	0.0125
Head of Household - Married or Cohabitant		-0.2759***
		0.0140
Urban		0.2045***
		0.0149
Has water		0.1321***
		0.0107
Has electricity		0.1638***
		0.0159
R-squared	0.567	0.595
N	56390	56390

Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years and to household heads aged 25–55 in all observed years.
Note: ***, **, * indicate significance at the 1%, 5%, and 10% levels. The dependent variable is monthly household total expenditures per capita (2007 dollars), inverse hyperbolic sine transformed. The sample includes households observed at least three times and with head age 25–55 in each survey year, 2007–2019. All regressions include UBIGEO (department–province–district level location) and year fixed effects and are weighted by the annual expansion factor (CPV-2007 projections). Standard errors (in parentheses) are clustered at the household level.

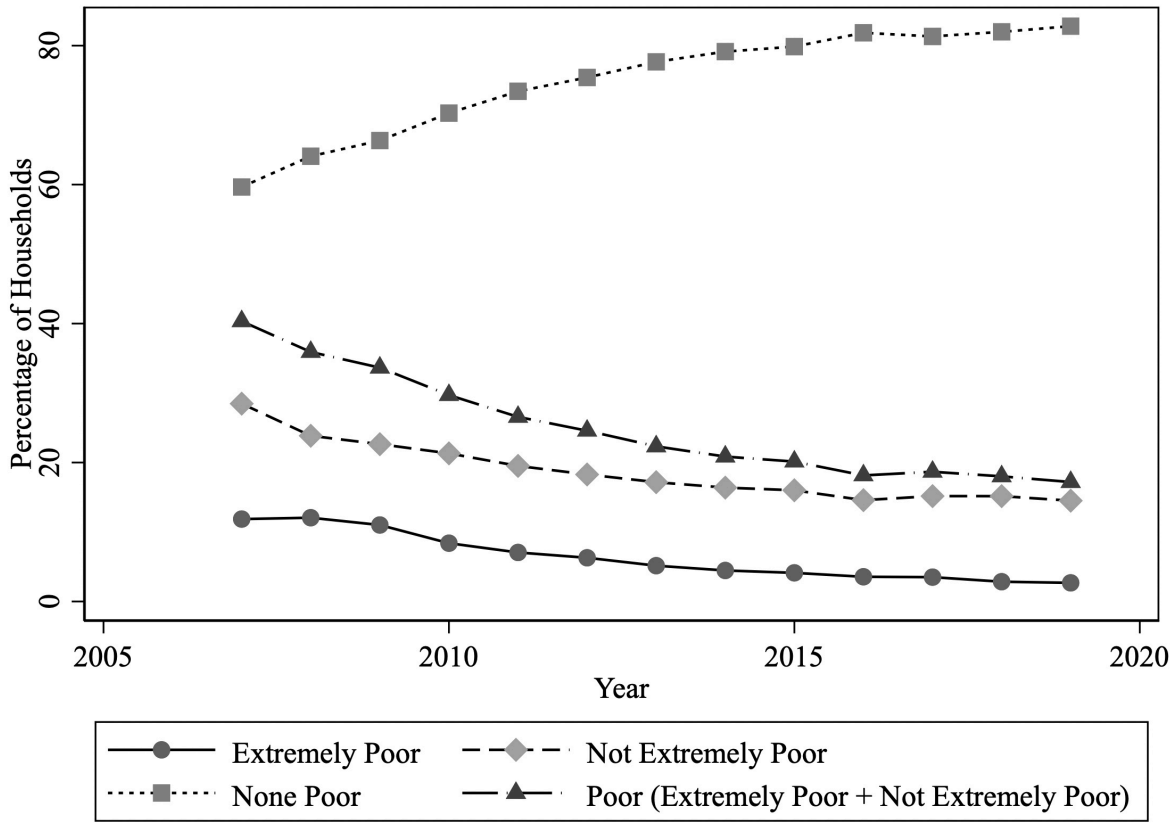
Table A4: BICs and RMSEs by specification and number of random starts

	n_start=3					n_start=10					Diff (ns10-ns3)	
	spec1		spec2		s2-s1	spec1		spec2		s2-s1	spec1	spec2
G	BIC (1)	RMSE (2)	BIC (3)	RMSE (4)	Diff (4)-(2)	BIC (5)	RMSE (6)	BIC (7)	RMSE (8)	Diff (8)-(6)	RMSE (6)-(2)	RMSE (8)-(4)
1	1.764	0.503	1.590	0.478	-0.025							
2	0.958	0.485	0.878	0.466	-0.019							
3	0.717	0.470	0.669	0.455	-0.015							
4	0.584	0.416	0.548	0.408	-0.008	0.584	0.416	0.548	0.408	-0.008	0.000	0.000
5	0.529	0.449	0.495	0.440	-0.009	0.521	0.453	0.491	0.438	-0.015	0.004	-0.002
6	0.464	0.427	0.444	0.425	-0.002	0.464	0.427	0.439	0.426	-0.001	0.000	0.001
7	0.430	0.448	0.411	0.436	-0.011	0.430	0.448	0.408	0.448	0.000	0.000	0.012
8	0.402	0.449	0.385	0.457	0.008							
9	0.380	0.453	0.357	0.434	-0.019			0.356	0.425			-0.010
10	0.363	0.441	0.339	0.438	-0.004	0.359	0.447	0.337	0.431	-0.016	0.006	-0.006
11	0.338	0.442	0.322	0.456	0.015	0.338	0.442				0.000	
12	0.328	0.464	0.304	0.442	-0.021							
13	0.316	0.472	0.297	0.481	0.009							
14	0.303	0.472	0.290	0.471	-0.001							
15	0.295	0.518	0.277	0.496	-0.022							
16	0.281	0.476	0.266	0.489	0.013							
17	0.273	0.485	0.257	0.479	-0.006							
18	0.264	0.474	0.249	0.475	0.001							
19	0.259	0.518	0.243	0.483	-0.035							
20	0.251	0.510	0.236	0.500	-0.010							
21	0.246	0.506	0.231	0.509	0.003							
22	0.237	0.520	0.225	0.499	-0.021							
23	0.230	0.496	0.218	0.485	-0.011							
24	0.226	0.505	0.215	0.520	0.015							
25	0.221	0.496	0.209	0.491	-0.005							
26	0.218	0.511	0.207	0.512	0.001							
27	0.211	0.504	0.201	0.513	0.010							
28	0.207	0.519	0.198	0.505	-0.014							
29	0.204	0.521	0.192	0.484	-0.037							
30	0.198	0.508	0.189	0.508	0.000							
31	0.198	0.503	0.186	0.512	0.009							
32	0.193	0.512	0.184	0.513	0.001							
33	0.189	0.509	0.179	0.495	-0.014							
34	0.189	0.534	0.180	0.529	-0.005							
35	0.185	0.515	0.175	0.515	0.000							
36	0.183	0.538	0.173	0.514	-0.023							
37	0.180	0.562	0.169	0.535	-0.028							
38	0.178	0.523	0.167	0.514	-0.009							
39	0.175	0.529	0.166	0.521	-0.008							
40	0.173	0.533	0.166	0.519	-0.014							

Notes: BIC is computed on the training data. RMSE is computed on the test data. For each specification, we estimate the model over the grid of G with `n_starts` $= 3$ and rerun selected G values with `n_starts` $= 10$ to assess sensitivity to initialization. "Diff" columns report spec2–spec1 and `n_starts` $= 10 -$ `n_starts` $= 3$ RMSE differences. Blank entries indicate G values not rerun with `n_starts` $= 10$.

Appendix C Figures

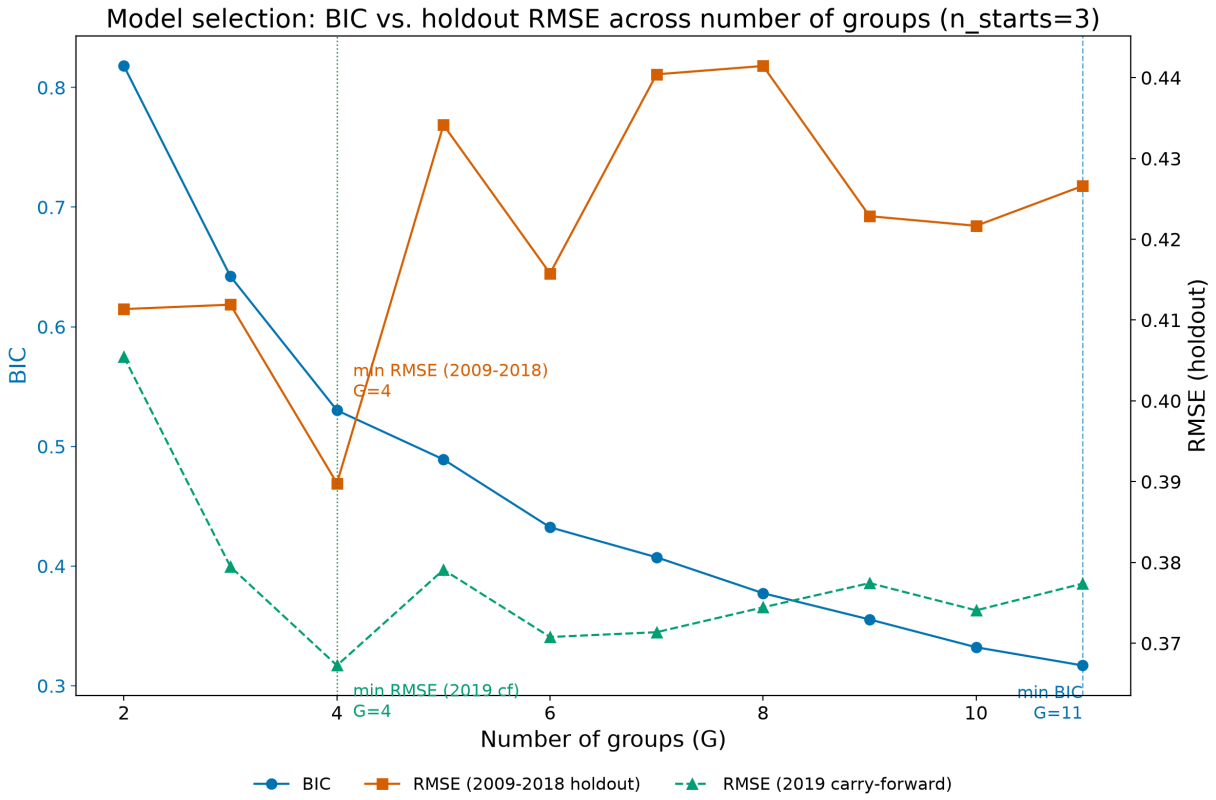
Figure A1: The trend of poverty status by year - Unrestricted full sample



Source: Author's calculation using full ENAHO dataset.

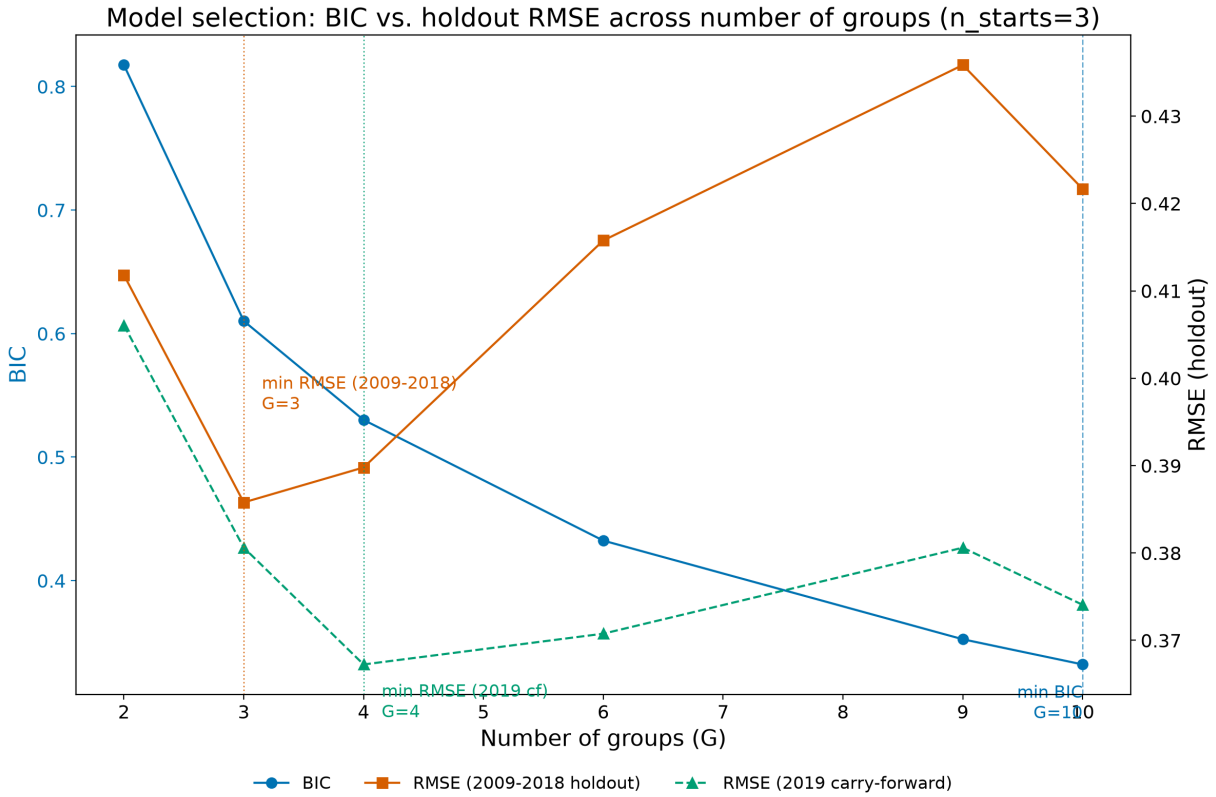
Note: This figure present the trend of poverty status using the original sample without years and age restriction.

Figure A2: Model fit (BIC and RMSE) under specification 2 (`n_starts` = 3)



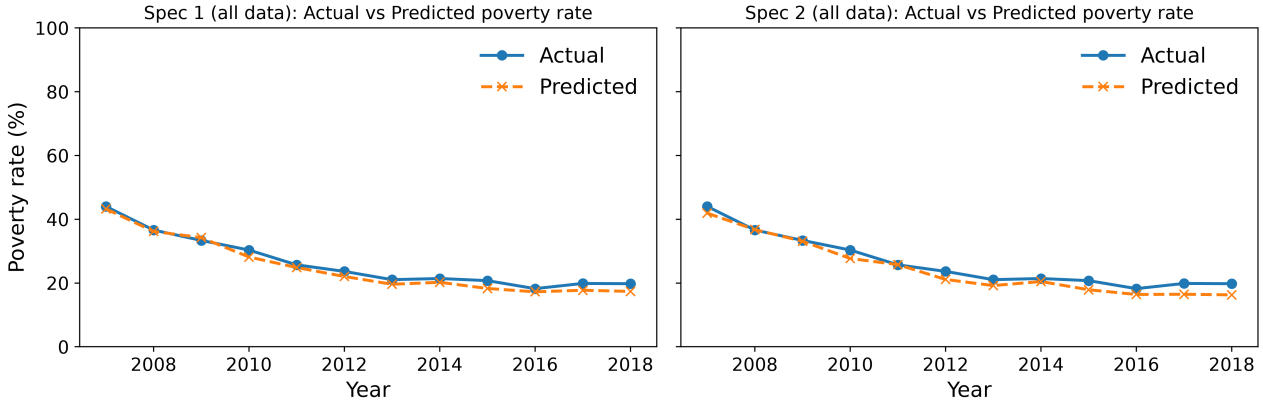
Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years and to household heads aged 25–55 in all observed years.
Note: The figure plots model selection diagnostics for the GFE estimator across initial numbers of groups $G \in \{1, \cdots, 11\}$ with `n_starts` = 3, `itermax` = 10, `neighmax` = 5, and `max_local_iters` = 2, predicted on two different periods (2009-2018, 2019). BIC is computed on the training data; RMSE is computed out-of-sample using the last observed household-year held out for each household (years 2009–2018 or 2019) test data. Lower values indicate better fit/prediction.

Figure A3: Model fit (BIC and RMSE) under specification 2 (`n_starts` = 10)



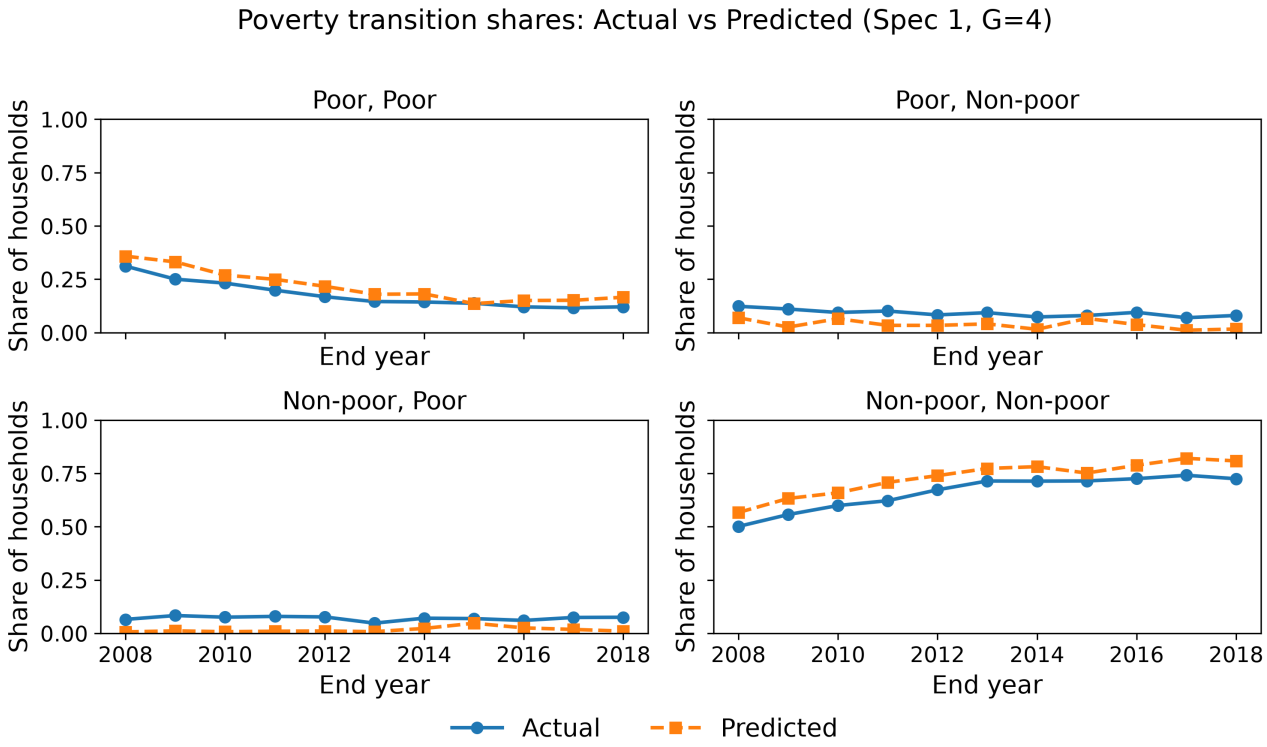
Source: Authors' calculations using ENAHO 2007–2019, restricted to households observed in at least three survey years and to household heads aged 25–55 in all observed years.
Note: The figure plots model selection diagnostics for the GFE estimator across candidate numbers of groups G predicted on two different periods (2009-2018, 2019), with `n_starts` = 10, `itermax` = 10, `neighmax` = 5, and `max_local_iters` = 2. BIC is computed on the training data; RMSE is computed out-of-sample using the last observed household-year held out for each household (years 2009–2018) test data. Lower values indicate better fit/prediction.

Figure A4: Actual vs. predicted poverty rate over time in all available data (G=4)



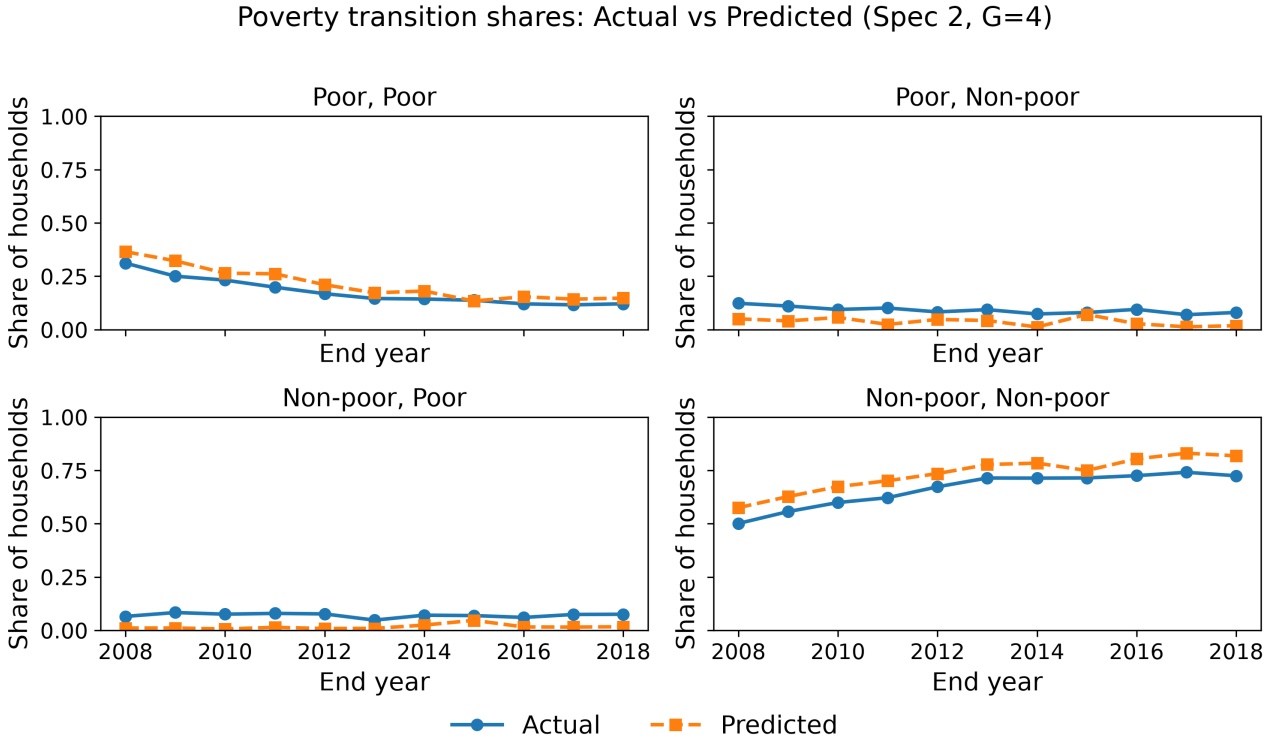
Data: All available selected sample (training + test).
Note: Poverty rates are survey-weighted (ENAHO expansion weights). This figure is prepared using all the available data (training + test). Predicted poverty is based on model-predicted welfare from the GFE model and the IHS poverty line (both in IHS scale).

Figure A5: Two-year poverty transition shares over time: actual vs. GFE-predicted (all observed household–year pairs, $G = 4$), for specification 1.



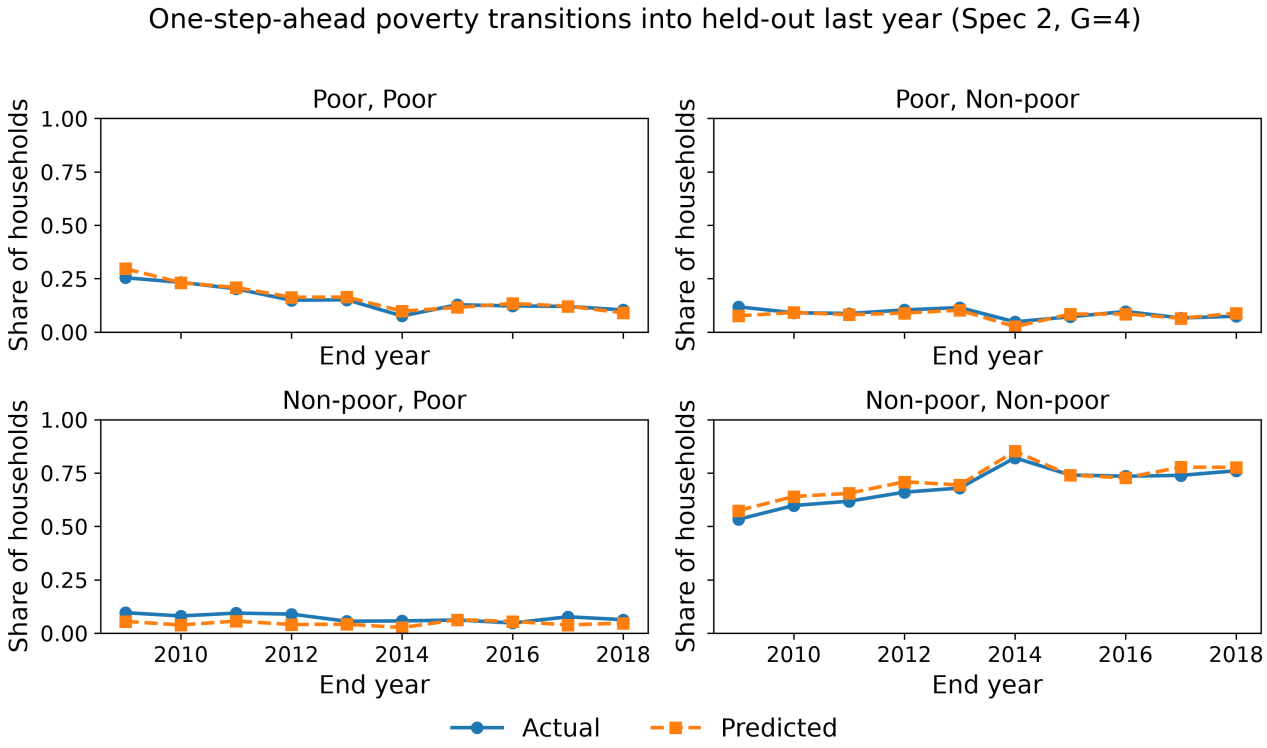
Data: All available selected sample (training + test).
Note: Each panel plots the survey-weighted share of households in one of four poverty transition states between consecutive survey years $(t-1, t)$: Poor→Poor, Poor→Non-poor, Non-poor→Poor, and Non-poor→Non-poor. "Actual" transitions are computed from observed (IHS-transformed) per-capita expenditure relative to the (IHS-transformed) total poverty line. "Predicted" transitions replace observed expenditure with GFE-predicted welfare $\hat{y}_{it} = x'_{it}\hat{\theta} + \hat{\alpha}_{\hat{g}_i,t} + \hat{\mu}_{p_i}$ and classify predicted poverty as $1\{\hat{y}_{it} < z_{it}\}$. Only households observed in both years of a given pair are included; transition shares are weighted using ENAHO expansion weights (measured in the end year t). The final year-pair is omitted when group–year intercepts are unavailable (e.g., no $\hat{\alpha}_{g,2019}$).

Figure A6: Two-year poverty transition shares over time: actual vs. GFE-predicted (all observed household–year pairs, $G = 4$), for specification 2.



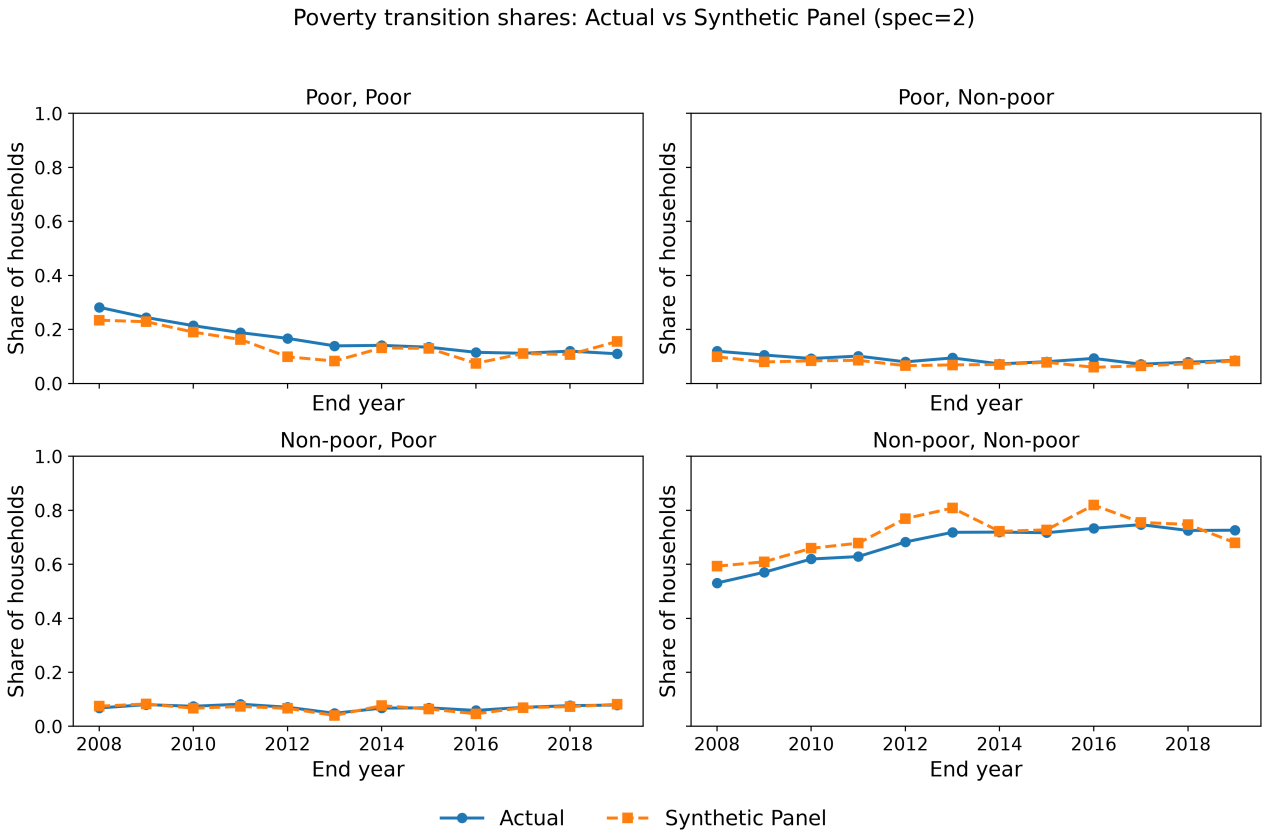
Data: All available selected sample (training + test).
Note: Each panel plots the survey-weighted share of households in one of four poverty transition states between consecutive survey years $(t-1, t)$: Poor→Poor, Poor→Non-poor, Non-poor→Poor, and Non-poor→Non-poor. "Actual" transitions are computed from observed (IHS-transformed) per-capita expenditure relative to the (IHS-transformed) total poverty line. "Predicted" transitions replace observed expenditure with GFE-predicted welfare $\hat{y}_{it} = x'_{it}\hat{\theta} + \hat{\alpha}_{\hat{g}_i,t} + \hat{\mu}_{p_i}$ and classify predicted poverty as $1\{\hat{y}_{it} < z_{it}\}$. Only households observed in both years of a given pair are included; transition shares are weighted using ENAHO expansion weights (measured in the end year t). The final year-pair is omitted when group–year intercepts are unavailable (e.g., no $\hat{\alpha}_{g,2019}$).

Figure A7: One-step-ahead poverty transition shares into held-out final observations: actual vs. GFE-predicted ($G = 4$), for specification 2.



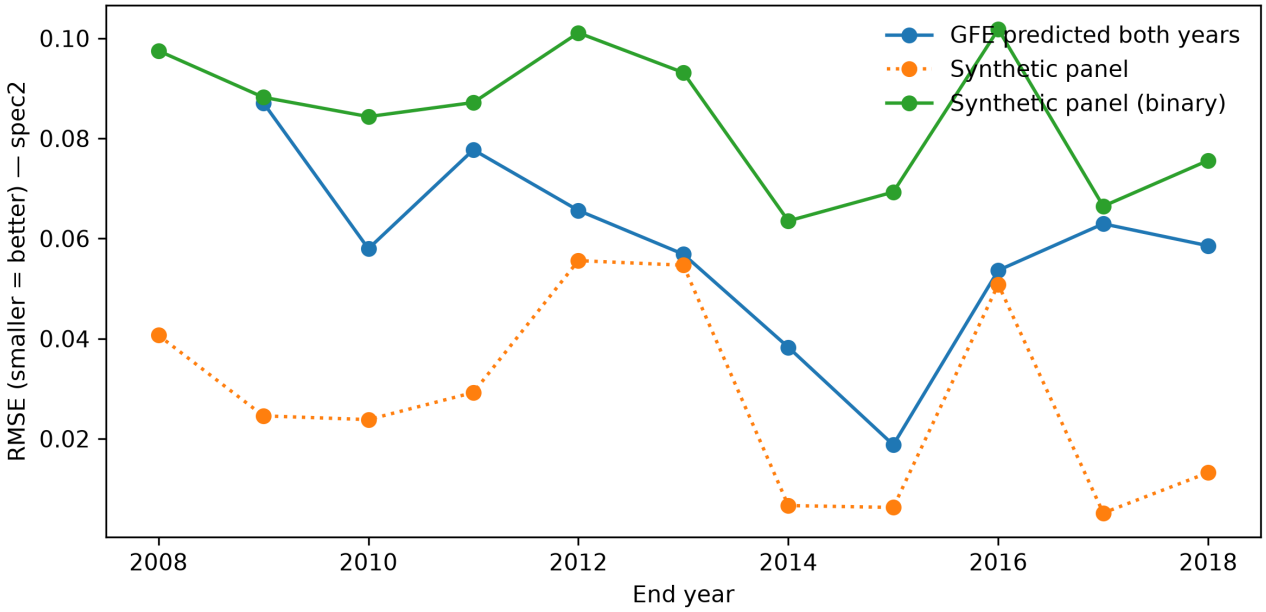
Data: Last year of training data + all test data.
Note: The figure reports survey-weighted shares of households transitioning between poverty states from $t-1$ to t, where t is each household's held-out last observed year and $t-1$ is its preceding observed year (restricted to consecutive years; one transition per household). "Actual" transitions are computed using observed (IHS-transformed) per-capita expenditure relative to the (IHS-transformed) poverty line in both years. "GFE-predicted" transitions use actual poverty at $t-1$ and predicted poverty at t, where predicted poverty is $1\{\hat{y}_{it} < z_{it}\}$ based on model-predicted welfare $\hat{y}_{it} = x'_{it}\hat{\theta} + \hat{\alpha}_{\hat{g}_i,t} + \hat{\mu}_{p_i}$. Transition shares are weighted using ENAHO expansion weights in year t. In this one-step-ahead sample, the expansion-weighted misclassification rate of $\widehat{\text{poor}}_{it}$ in year t is 17.6%.

Figure A8: Poverty transition shares: actual vs. synthetic panel predictions, for specification 2



Note: Each panel plots the share of households in one of four poverty transition states between $t-1$ and t: poor–poor, poor–non-poor, non-poor–poor, and non-poor–non-poor. The solid line ("Actual") reports transitions computed directly from observed survey data using per-capita expenditure relative to the poverty line in both years. The dashed line reports the corresponding transition shares predicted by the synthetic panel method, which is estimated using repeated cross-sections rather than true panel links. The horizontal axis uses the end year t for each two-year pair. Because of the set up and use of repeated cross-sections rather than true panel, the actual and prediction poverty transition can go up to $t = 2019$. The sample only use household head between 25-55 years old as well, the same as our sample.

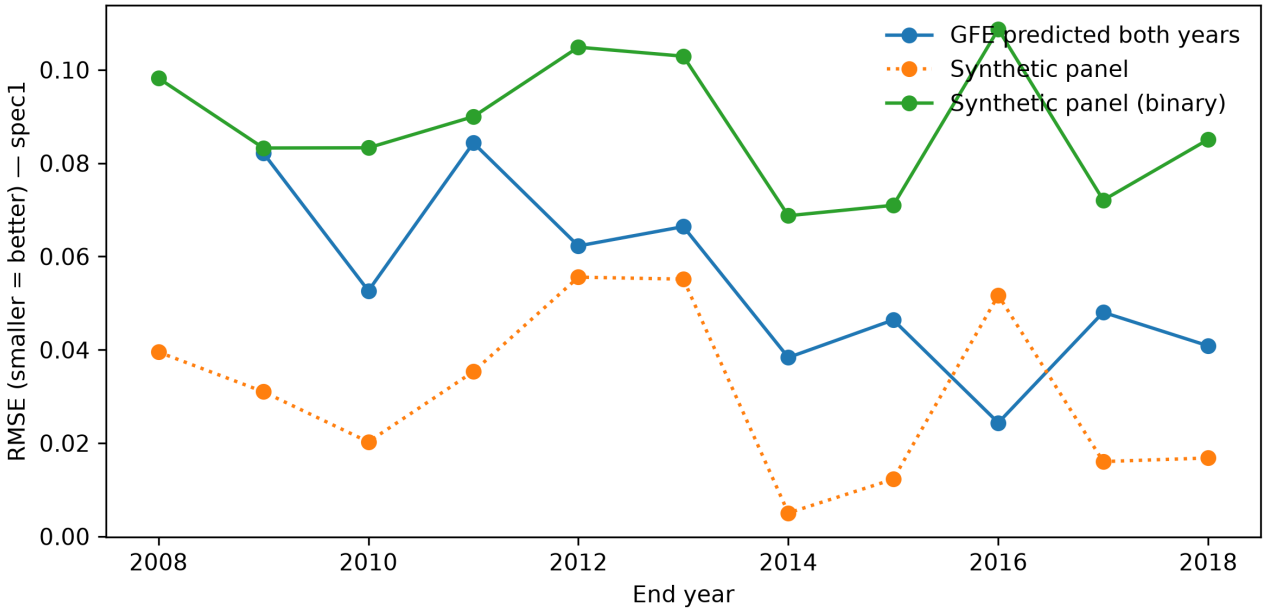
Figure A9: Prediction error in poverty transition shares: GFE one-step versus synthetic panel.



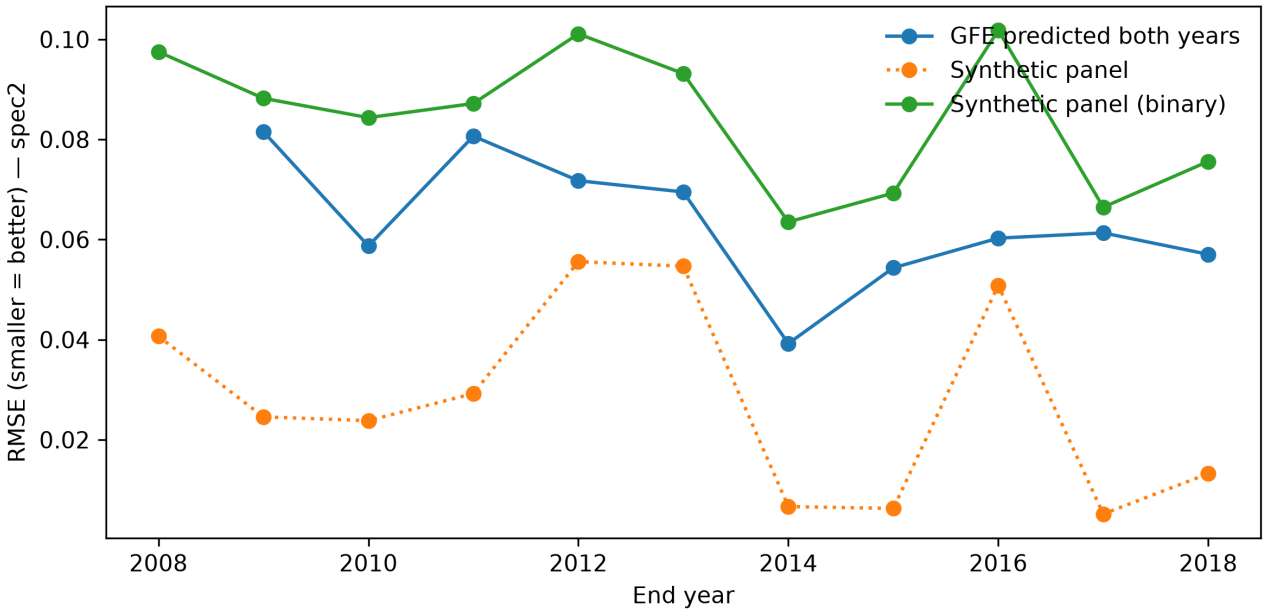
Note: The figures plot the root mean squared error (RMSE) between predicted and observed two-year poverty transition shares for each end year t. For a given end year, the RMSE is computed across the four transition states (poor–poor, poor–non-poor, non-poor–poor, non-poor–non-poor), comparing the predicted transition-share vector to the observed transition-share vector; smaller values indicate a closer fit. "GFE-predicted both years" line shows the RMSE from GFE-based approach, which uses predicted poverty status in t–1 and in t. The other two lines show the RMSE from the synthetic panel methods, which produce transition-share predictions from repeated cross-sections, and a modified binary-version synthetic panel method that is comparable with the GFE-based method. Because the underlying samples used to compute the transition shares may differ across approaches, the figures are intended as a descriptive comparison of fit within each method. Additionally, the GFE RMSE begins in 2009 rather than 2008 because the one-step-ahead validation requires observing each household at both $t-1$ and the last observed year t. Under our analysis sample restriction (at least three observed survey years per household), no household has $t = 2008$ as its last observed year.

Figure A10: Prediction error in poverty transition shares: GFE one-step versus synthetic panel using data without age restriction

(a) Specification 1



(b) Specification 2



Note: The figures plot the root mean squared error (RMSE) between predicted and observed two-year poverty transition shares for each end year t. For a given end year, the RMSE is computed across the four transition states (poor–poor, poor–non-poor, non-poor–poor, non-poor–non-poor), comparing the predicted transition-share vector to the observed transition-share vector; smaller values indicate a closer fit. "GFE-predicted both years" line shows the RMSE from GFE-based approach, which uses predicted poverty status in t–1 and in t. The other two lines show the RMSE from the synthetic panel methods, which produce transition-share predictions from repeated cross-sections, and a modified binary-version synthetic panel method that is comparable with the GFE-based method. Because the underlying samples used to compute the transition shares may differ across approaches, the figures are intended as a descriptive comparison of fit within each method. Additionally, the GFE RMSE begins in 2009 rather than 2008 because the one-step-ahead validation requires observing each household at both $t-1$ and the last observed year t. Under our analysis sample restriction (at least three observed survey years per household), no household has $t = 2008$ as its last observed year.