Econometrica, Vol. 90, No. 2 (March, 2022), 625–643

DISCRETIZING UNOBSERVED HETEROGENEITY

STÉPHANE BONHOMME
Kenneth C. Griffin Department of Economics, University of Chicago

THIBAUT LAMADON
Kenneth C. Griffin Department of Economics, University of Chicago

ELENA MANRESA
Department of Economics, New York University

We study discrete panel data methods where unobserved heterogeneity is revealed in a first step, in environments where population heterogeneity is not discrete. We focus on *two-step grouped fixed-effects* (GFE) estimators, where individuals are first classified into groups using *kmeans* clustering, and the model is then estimated allowing for group-specific heterogeneity. Our framework relies on two key properties: heterogeneity is a function—possibly nonlinear and time-varying—of a low-dimensional continuous latent type, and informative moments are available for classification. We illustrate the method in a model of wages and labor market participation, and in a probit model with time-varying heterogeneity. We derive asymptotic expansions of two-step GFE estimators as the number of groups grows with the two dimensions of the panel. We propose a data-driven rule for the number of groups, and discuss bias reduction and inference.

KEYWORDS: Unobserved heterogeneity, panel data, kmeans clustering, dimension reduction.

1. INTRODUCTION

IN BOTH REDUCED-FORM AND STRUCTURAL WORK IN ECONOMICS, it is common to model unobserved heterogeneity as a small number of discrete types. Various estimation strategies are available, including discrete-type random-effects (as in Keane and Wolpin (1997) and many other applications) and grouped fixed-effects (as recently studied by Hahn and Moon (2010) and Bonhomme and Manresa (2015)). These methods require the researcher to jointly estimate individual heterogeneity and model parameters.[1] In addition, little is known about their properties when individual heterogeneity is not discrete in the population. In this paper, we study two-step discrete estimators for panel data, and provide conditions for their validity when heterogeneity is continuous.

We focus on *two-step grouped fixed-effects* (GFE) estimators. In a first step, we classify individuals based on a set of individual-specific moments, using the *kmeans* clustering algorithm. The aim of the kmeans classification is to group together individuals whose latent

Stéphane Bonhomme: sbonhomme@uchicago.edu
Thibaut Lamadon: lamadon@uchicago.edu
Elena Manresa: elena.manresa@nyu.edu

We thank four anonymous referees, Manuel Arellano, Neele Balke, Jesus Carro, Gary Chamberlain, Tim Christensen, Liran Einav, Alfred Galichon, Chris Hansen, Joe Hotz, Grégory Jolivet, Arthur Lewbel, Anna Mikusheva, Roger Moon, Whitney Newey, Juan Pantano, Philippe Rigollet, Anna Simoni, Martin Weidner, and seminar audiences at various places for comments. The authors acknowledge support from the NSF grant number SES-1658920.

[1]Also related, nonparametric maximum likelihood methods (e.g., Heckman and Singer (1984)) rely on joint estimation of the distribution of heterogeneity and the parameters.

- Panel data model where unobserved heterogeneity is continuous in the population

- Panel data model where unobserved heterogeneity is continuous in the population
- But approximated with a discrete distribution (Group Fixed Effects, GFE)

- Panel data model where unobserved heterogeneity is continuous in the population
- But approximated with a discrete distribution (Group Fixed Effects, GFE)
- Propose a 2-step estimation algorithm:

- Panel data model where unobserved heterogeneity is continuous in the population
- But approximated with a discrete distribution (Group Fixed Effects, GFE)
- Propose a 2-step estimation algorithm:
  1. Classify units into groups using k-means clustering

- Panel data model where unobserved heterogeneity is continuous in the population
- But approximated with a discrete distribution (Group Fixed Effects, GFE)
- Propose a 2-step estimation algorithm:
    1. Classify units into groups using k-means clustering
    2. Estimate the model using the groups from step 1

- Panel data model where unobserved heterogeneity is continuous in the population
- But approximated with a discrete distribution (Group Fixed Effects, GFE)
- Propose a 2-step estimation algorithm:
    1. Classify units into groups using k-means clustering
    2. Estimate the model using the groups from step 1
- This is different from finite mixture models: no joint estimation required!

- Panel data model where unobserved heterogeneity is continuous in the population
- But approximated with a discrete distribution (Group Fixed Effects, GFE)
- Propose a 2-step estimation algorithm:
    1. Classify units into groups using k-means clustering
    2. Estimate the model using the groups from step 1
- This is different from finite mixture models: no joint estimation required!
- Handles both time-invariant and time-varying unobserved heterogeneity

GFE vs. Random Effects (EM) and Fixed Effects

GFE vs. Random Effects (EM) and Fixed Effects

- Random Effects (EM/MM): Efficient but requires distributional assumptions

GFE vs. Random Effects (EM) and Fixed Effects

- Random Effects (EM/MM): Efficient but requires distributional assumptions
  - Must specify type distribution (e.g., 3 discrete types, normal mixture)

GFE vs. Random Effects (EM) and Fixed Effects

- Random Effects (EM/MM): Efficient but requires distributional assumptions
  - Must specify type distribution (e.g., 3 discrete types, normal mixture)
  - Misspecification can bias all estimates

GFE vs. Random Effects (EM) and Fixed Effects

- Random Effects (EM/MM): Efficient but requires distributional assumptions
  - Must specify type distribution (e.g., 3 discrete types, normal mixture)
  - Misspecification can bias all estimates
- Fixed Effects: Flexible but suffers from incidental parameters

GFE vs. Random Effects (EM) and Fixed Effects

- Random Effects (EM/MM): Efficient but requires distributional assumptions
  - Must specify type distribution (e.g., 3 discrete types, normal mixture)
  - Misspecification can bias all estimates
- Fixed Effects: Flexible but suffers from incidental parameters
  - In nonlinear models, can't eliminate α_i via transformation

GFE vs. Random Effects (EM) and Fixed Effects

- Random Effects (EM/MM): Efficient but requires distributional assumptions
  - Must specify type distribution (e.g., 3 discrete types, normal mixture)
  - Misspecification can bias all estimates
- Fixed Effects: Flexible but suffers from incidental parameters
  - In nonlinear models, can't eliminate α_i via transformation
  - Estimating N parameters with T observations each leads to bias in θ

GFE vs. Random Effects (EM) and Fixed Effects

- Random Effects (EM/MM): Efficient but requires distributional assumptions
  - Must specify type distribution (e.g., 3 discrete types, normal mixture)
  - Misspecification can bias all estimates
- Fixed Effects: Flexible but suffers from incidental parameters
  - In nonlinear models, can't eliminate α_i via transformation
  - Estimating N parameters with T observations each leads to bias in θ
- GFE: Fixed effects logic with dimension reduction

GFE vs. Random Effects (EM) and Fixed Effects

- Random Effects (EM/MM): Efficient but requires distributional assumptions
  - Must specify type distribution (e.g., 3 discrete types, normal mixture)
  - Misspecification can bias all estimates
- Fixed Effects: Flexible but suffers from incidental parameters
  - In nonlinear models, can't eliminate α_i via transformation
  - Estimating N parameters with T observations each leads to bias in θ
- GFE: Fixed effects logic with dimension reduction
  - Group similar individuals, estimate K group parameters (where $K \ll N$)

GFE vs. Random Effects (EM) and Fixed Effects

- Random Effects (EM/MM): Efficient but requires distributional assumptions
    - Must specify type distribution (e.g., 3 discrete types, normal mixture)
    - Misspecification can bias all estimates
- Fixed Effects: Flexible but suffers from incidental parameters
    - In nonlinear models, can't eliminate α_i via transformation
    - Estimating N parameters with T observations each leads to bias in θ
- GFE: Fixed effects logic with dimension reduction
    - Group similar individuals, estimate K group parameters (where $K \ll N$)
    - No distributional assumptions, but reduces incidental parameters problem

GFE vs. Random Effects (EM) and Fixed Effects

- Random Effects (EM/MM): Efficient but requires distributional assumptions
  - Must specify type distribution (e.g., 3 discrete types, normal mixture)
  - Misspecification can bias all estimates
- Fixed Effects: Flexible but suffers from incidental parameters
  - In nonlinear models, can't eliminate α_i via transformation
  - Estimating N parameters with T observations each leads to bias in θ
- GFE: Fixed effects logic with dimension reduction
  - Group similar individuals, estimate K group parameters (where $K \ll N$)
  - No distributional assumptions, but reduces incidental parameters problem
  - Treats continuous heterogeneity as approximately discrete

There are two main assumptions:

There are two main assumptions:

1. Low-dimensional types: All individual heterogeneity is driven by a small set of latent characteristics

There are two main assumptions:

1. Low-dimensional types: All individual heterogeneity is driven by a small set of latent characteristics
   - e.g. worker productivity and firm quality both depend on a single match type

There are two main assumptions:

1. Low-dimensional types: All individual heterogeneity is driven by a small set of latent characteristics

    - e.g. worker productivity and firm quality both depend on a single match type
    - Can handle time-varying effects through nonlinear factor structures

There are two main assumptions:

1. Low-dimensional types: All individual heterogeneity is driven by a small set of latent characteristics
   - e.g. worker productivity and firm quality both depend on a single match type
   - Can handle time-varying effects through nonlinear factor structures

There are two main assumptions:

1. Low-dimensional types: All individual heterogeneity is driven by a small set of latent characteristics
    - e.g. worker productivity and firm quality both depend on a single match type
    - Can handle time-varying effects through nonlinear factor structures
2. Types are learnable: We can distinguish individuals based on observable patterns

There are two main assumptions:

1. Low-dimensional types: All individual heterogeneity is driven by a small set of latent characteristics
    - e.g. worker productivity and firm quality both depend on a single match type
    - Can handle time-varying effects through nonlinear factor structures
2. Types are learnable: We can distinguish individuals based on observable patterns
    - e.g. wage histories reveal underlying productivity

There are two main assumptions:

1. Low-dimensional types: All individual heterogeneity is driven by a small set of latent characteristics
    - e.g. worker productivity and firm quality both depend on a single match type
    - Can handle time-varying effects through nonlinear factor structures
2. Types are learnable: We can distinguish individuals based on observable patterns
    - e.g. wage histories reveal underlying productivity
    - e.g. choice patterns reveal preferences

Consider the log likelihood of a dynamic discrete choice model:

Consider the log likelihood of a dynamic discrete choice model:

$$\ell_i(\alpha_i, \theta; d_{it}, X_{it}, Y_{it}) = \sum_t \underbrace{\log f(d_{it}|X_{it}, \alpha_i, \theta)}_{\text{choices}} + \underbrace{\log f(X_{it}|d_{it-1}, X_{it-1}, \alpha_i, \theta)}_{\text{state transitions}} + \underbrace{\log f(Y_{it}|d_{it}, X_{it}, \alpha_i, \theta)}_{\text{outcomes}}$$

Consider the log likelihood of a dynamic discrete choice model:

$$\ell_i\left(\alpha_i, \theta; d_{it}, X_{it}, Y_{it}\right) = \sum_t \underbrace{\log f\left(d_{it} | X_{it}, \alpha_i, \theta\right)}_{\text{choices}} + \underbrace{\log f\left(X_{it} | d_{it-1}, X_{it-1}, \alpha_i, \theta\right)}_{\text{state transitions}} + \underbrace{\log f\left(Y_{it} | d_{it}, X_{it}, \alpha_i, \theta\right)}_{\text{outcomes}}$$

- Likelihoods are assumed to be additively separable conditional on the FE α_i

Consider the log likelihood of a dynamic discrete choice model:

$$\ell_i(\alpha_i, \theta; d_{it}, X_{it}, Y_{it}) = \sum_t \underbrace{\log f(d_{it}|X_{it}, \alpha_i, \theta)}_{\text{choices}} + \underbrace{\log f(X_{it}|d_{it-1}, X_{it-1}, \alpha_i, \theta)}_{\text{state transitions}} + \underbrace{\log f(Y_{it}|d_{it}, X_{it}, \alpha_i, \theta)}_{\text{outcomes}}$$

- Likelihoods are assumed to be additively separable conditional on the FE α_i
- Strict concavity of the log-likelihood in α is required

Consider the log likelihood of a dynamic discrete choice model:

$$\ell_i\left(\alpha_i, \theta; d_{it}, X_{it}, Y_{it}\right) = \sum_t \underbrace{\log f\left(d_{it} | X_{it}, \alpha_i, \theta\right)}_{\text{choices}} + \underbrace{\log f\left(X_{it} | d_{it-1}, X_{it-1}, \alpha_i, \theta\right)}_{\text{state transitions}} + \underbrace{\log f\left(Y_{it} | d_{it}, X_{it}, \alpha_i, \theta\right)}_{\text{outcomes}}$$

- Likelihoods are assumed to be additively separable conditional on the FE α_i
- Strict concavity of the log-likelihood in α is required
- This holds for many models: probit, logit, Poisson, tobit, etc.

Bias and Inference

Bias and Inference

- GFE has two sources of bias:

Bias and Inference

- GFE has two sources of bias:

  1. Incidental parameter bias (like standard FE)

Bias and Inference

- GFE has two sources of bias:

  1. Incidental parameter bias (like standard FE)

  2. Approximation error from discretizing continuous heterogeneity

Bias and Inference

- GFE has two sources of bias:
    1. Incidental parameter bias (like standard FE)
    2. Approximation error from discretizing continuous heterogeneity
- Use half-panel jackknife to correct bias

Bias and Inference

- GFE has two sources of bias:

  1. Incidental parameter bias (like standard FE)

  2. Approximation error from discretizing continuous heterogeneity

- Use half-panel jackknife to correct bias

  - Estimate on full sample, first half, and second half

Bias and Inference

- GFE has two sources of bias:
  1. Incidental parameter bias (like standard FE)
  2. Approximation error from discretizing continuous heterogeneity
- Use half-panel jackknife to correct bias
  - Estimate on full sample, first half, and second half
  - Combine: $\hat{\theta}_{BR} = 2\hat{\theta} - \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}$

Bias and Inference

- GFE has two sources of bias:
  1. Incidental parameter bias (like standard FE)
  2. Approximation error from discretizing continuous heterogeneity
- Use half-panel jackknife to correct bias
  - Estimate on full sample, first half, and second half
  - Combine: $\hat{\theta}_{BR} = 2\hat{\theta} - \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}$
- Choice of K: Data-driven rule controls approximation error

Extensions to Improve Performance

Extensions to Improve Performance

- Conditional moments: Control for covariate heterogeneity in step 1
  - Example: group workers by wage residuals after controlling for education

Extensions to Improve Performance

- Conditional moments: Control for covariate heterogeneity in step 1
    - Example: group workers by wage residuals after controlling for education
- Model-based iteration: Refine groups using full model structure
    - Reassign individuals to best-fitting groups, then re-estimate
    - Like EM algorithm but simpler (good starting point from step 1)

Extensions to Improve Performance

- Conditional moments: Control for covariate heterogeneity in step 1
  - Example: group workers by wage residuals after controlling for education
- Model-based iteration: Refine groups using full model structure
  - Reassign individuals to best-fitting groups, then re-estimate
  - Like EM algorithm but simpler (good starting point from step 1)
- Two-way grouping: Classify both individuals and time periods
  - When time effects also have low-dimensional structure
  - Reduces parameters from $K \times T$ to $K \times p$ (e.g., 3 business cycle regimes)