Check for updates

Received: 16 September 2022 | Revised: 20 March 2023 | Accepted: 2 May 2023

DOI: 10.1002/jae.2997

WILEY Journal of APPLIED ECONOMETRICS

REPLICATION

Approximating grouped fixed effects estimation via fuzzy clustering regression

Daniel J. Lewis[1] | **Davide Melcangi**[2] | **Laura Pilossoph**[3] | **Aidan Toner-Rodgers**[2]

[1]Department of Economics, University College London, London, UK

[2]Federal Reserve Bank of New York, New York, New York, United States

[3]Duke University, Durham, North Carolina, USA

Correspondence
Daniel J. Lewis, Department of Economics, University College London, 202 Drayton House, 30 Gordon St, London, WC1H 0AX, UK.
Email: daniel.j.lewis@ucl.ac.uk

Funding information
The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. This research was supported in part through computational resources provided by the BigTex High Performance Computing Group at the Federal Reserve Bank of Dallas.

Summary

We propose a new, computationally efficient way to approximate the "grouped fixed effects" (GFE) estimator of Bonhomme and Manresa (2015), which estimates grouped patterns of unobserved heterogeneity. To do so, we generalize the fuzzy C-means objective to regression settings. As the clustering exponent m approaches 1, the fuzzy clustering objective converges to the GFE objective, which we recast as a standard generalized method of moments problem. We replicate the empirical results of Bonhomme and Manresa (2015) and show that our estimator delivers almost identical estimates. In simulations, we show that our approach offers improvements in terms of bias, classification accuracy, and computational speed.

KEYWORDS

clustering, democracy, discrete heterogeneity, fixedeffects, panel data, unobserved heterogeneity

1 | INTRODUCTION

Bonhomme and Manresa (2015) (henceforth BM) propose the "grouped fixed effects estimator," a form of K-means regression, to study grouped patterns of heterogeneity in panel data settings. In particular, the authors study the linear model:

$$y_i = \sum_{g=1}^{G} \gamma_{ig}\theta_g x_i + v_i, \tag{1}$$

where $y_i \in \mathbb{R}^T$, $x_i \in \mathbb{R}^K$, (x_i, y_i) are independently distributed across i (and identically distributed conditional on group membership, $\tilde{g}_i$), $\gamma_{ig} = \mathbf{1}\left[\tilde{g}_i = g\right]$ and $E\left[v_i | x_i, \tilde{g}_i = g\right] = 0$. θ_g is a $T \times K$ matrix of group-specific coefficients on the covariates x, for $g = 1, \ldots, G$. The indicators γ_{ig} are equal to 1 if observation i belongs to group g, and zero otherwise. Equation (1) postulates that the outcomes y are generated linearly from x, with parameters depending on observation i's group membership, which is unobservable to the econometrician. BM consider estimating θ_g as well as group membership indicators $\gamma_{ig} = \mathbf{1}\left[i \in g\right]$ for G groups, where x_i may include a constant term, hence the "grouped fixed effects" (GFE) estimator.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2023 Federal Reserve Bank of New York and The Authors. *Journal of Applied Econometrics* published by John Wiley & Sons Ltd.

In this paper, we provide a method to approximate the GFE objective function that incorporates a degree of smoothing of the membership function, allowing for computationally more efficient estimation and delivering more accurate estimates in simulations. Our approach is particularly valuable with large data sets, which are increasingly the norm in applied economics, as well as when the specified number of groups G becomes large. BM consider the population GFE least-squares criterion

$$\left(\tilde{\theta}, \tilde{\gamma}\right) = \arg\min_{\theta,\gamma} E\left[\sum_{g=1}^{G} \gamma_{ig}\left\|y_i - \theta_g x_i\right\|^2\right], \tag{2}$$

where the minimum is taken over all possible group membership assignments, $\tilde{g} = \{\tilde{g}_1, \ldots, \tilde{g}_N\}$ of the N observed entities, and the group-specific effects $\left(\theta_g\right)$, which collectively form θ. Instead, we study

$$\left(\tilde{\theta}_m^{FCR}, \tilde{\mu}\right) = \arg\min_{\theta,\mu} E\left[\sum_{g=1}^{G} \left(\mu_{ig}\right)^m\left\|y_i - \theta_g x_i\right\|^2\right], \tag{3}$$

where the minimum is taken over the group-specific effects θ_g and the weights μ_{ig}, raised to the user-specified clustering parameter $m > 1$. Mean clustering problems following (3) are known as "fuzzy C-means" algorithms, since $m > 1$ induces continuous weights and thus "fuzzy" group assignments, rather than the binary assignments of "hard" K-means in Equation (2). We thus refer to the model (3) as "fuzzy C-regression," or FCR.

We show that in the limit as $m \to 1$, the "fuzzy" objective in Equation (3) is equivalent to the K-means objective in Equation (2). However, by writing the optimal weights $\mu_{ig}(y, x; \theta, m)$ as a function of m and θ, FCR can be reframed as a Generalized Method of Moments (GMM) problem. For a fixed m, a sample estimator $\hat{\theta}_m^{FCR}$ is then consistent for $\tilde{\theta}_m^{FCR}$, with an asymptotically normal limiting distribution.

Why is this a useful representation of the original problem in Equation (2)? First, as m approaches 1, the "fuzzy" objective function converges to the original GFE objective function, providing a convenient approximation. We find in both simulations and BM's empirical application that for $m = 1.001$, say, the quality of the approximation is excellent, as evidenced by the performance of $\hat{\theta}_m^{FCR}$ relative to either the true parameters and group assignment, or BM's empirical estimates. Second, because the objective in Equation (3) is continuous for $m > 1$ (but potentially close to 1), it is natural to estimate $\tilde{\theta}_m^{FCR}$ not through an iterative procedure, alternating between choosing a set of discrete group membership functions and conditionally optimizing the model parameters, as is conventional for clustering problems, but rather through direct minimization of the objective function, as in any other GMM problem. This means the problem can be solved in a single step using any built-in minimization routine, without the need to explicitly code the specialized algorithms described in BM. Indeed, the preferred solution method of BM involves searching over essentially all possible group assignments of the N entities, which is a very computationally intensive task as N and G grow. Perhaps due to resulting numerical challenges of searching over the binary group assignments, we find that the ability of our approximating estimator to recover the truth in simulations (in terms of both parameter estimates and group assignment) is superior to the original GFE estimator, despite the lesser computation time noted above.

In summary, we find in simulations that our approximating estimator yields more reliable parameter estimates, delivers lower misclassification rates, and requires far shorter computation time, particularly as N, T, and G grow. We thus think of the fuzzy clustering approach as a fast and easy way to estimate grouped fixed effects.

Our paper relates to several recent papers that provide different ways of estimating grouped patterns of heterogeneity in a panel data setting. Kim and Wang (2019) take a Bayesian approach to the problem and demonstrate that their method leads to similar estimates in the BM empirical application, much like ours. We instead retain a frequentist framework and replace the binary membership function in BM with continuous weights. In another recent paper, Mugnier (2022) proposes a "pairwise distance" estimator based on the adjacency matrix of the units of observation, which also offers computational improvements over the GFE estimator and achieves consistency for $N, T \to \infty$. However, he does not extend his method to the case of heterogeneous coefficients on regressors, x. While simulation results illustrate consistency, no performance comparison is conducted between the proposed estimator and GFE. Chetverikov and Manresa (2022) propose a "post-spectral" estimator that likewise is computationally simpler than the GFE estimator, where an initial parameter estimate is obtained from an auxiliary convex function that can be consistently estimated and is also minimized at $\tilde{\theta}$. However, they require a particular factor structure for the covariates, x. While for a different data generating process, their simulation results also point to significant improvements over GFE, in particular, in terms of misclassification. In a further strand of the literature, Su et al. (2016) and Su and Ju (2018), among others, consider approaches based on the Lasso procedure to achieve classification, as an alternative to more traditional clustering methods.

10991255, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jae.2997 by University College London UCL Library Services, Wiley Online Library on [25/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | FUZZY CLUSTERING REGRESSION

In this section, we describe the fuzzy clustering regression (FCR) methodology. We first show that it converges to the "hard" K-means objective of BM, justifying the use of m close to 1 to approximate GFE estimators. Next, we argue that the FCR problem can be rewritten as a standard GMM problem, immediately yielding an asymptotic distribution.

2.1 | FCR approximates GFE

Consider the linear model introduced in Equation (1). The t dimension of y_i allows for a panel structure, with repeated observations of each entity i over time, or simply multiple observed outcomes for each entity. The "hard" K-means (HKM), or in the terminology of BM, GFE, objective function is

$$J^{HKM}(\theta) = E\left[\min_g \left\|y - \theta_g x\right\|^2\right]. \tag{4}$$

Alternatively, the objective above can be rewritten as a weighted sum replacing minimization with the resulting binary group membership function,

$$L^{HKM}(\theta) = E\left[\sum_{g=1}^{G} \gamma_g^* (y, x; \theta) \left\|y - \theta_g x\right\|^2\right], \tag{5}$$

where $\gamma_g^* (y, x; \theta) = \mathbf{1}\left[\left\|y - \theta_g x\right\|^2 \leq \|y - \theta_h x\|^2 \forall h \neq g\right]$.

The FCR objective function imposes smoothness on the estimated group membership function, $\gamma_g^*(\cdot)$, so that it is no longer binary. In particular, the objective function is instead

$$L_m^{FCR}(\theta, \mu) = E\left[\sum_{g=1}^{G} \left(\mu_g\right)^m \left\|y - \theta_g x\right\|^2\right] \tag{6}$$

where $m > 1$ is the clustering parameter and μ_g represent group weights. Bezdek (1981) derives the MSE-optimal weights μ_g given m and θ, $\mu_g(y, x; \theta, m)$, in the pure fixed effects case (i.e., a single constant regressor in x). Analogously, the optimal weights in the regression case, Equation (6), can be derived as

$$\mu_g (y, x; \theta, m) = \left(\sum_{h=1}^{G} \frac{\left\|y - \theta_g x\right\|^{2/(m-1)}}{\left\|y - \theta_h x\right\|^{2/(m-1)}}\right)^{-1}, g = 1, \dots, G, \tag{7}$$

Yang and Yu (1992) show that in the pure fixed effects case the objective in Equation (6) can be rewritten subsuming the optimal parameter-dependent weights in Equation (7) into the objective function itself. For fixed m, define $\mu (y, x; \theta, m) = (\mu_1 (y, x; \theta, m), \dots, \mu_G (y, x; \theta, m))$. Then, Equation (6) can similarly be rewritten as

$$\begin{aligned} L_m^{FCR} (\theta, \mu) &= L_m^{FCR} (\theta, \mu (y, x; \theta, m)) \\ &= E\left[\sum_{g=1}^{G} \left(\sum_{h=1}^{G} \frac{\left\|y - \theta_g x\right\|^{2/(m-1)}}{\left\|y - \theta_h x\right\|^{2/(m-1)}}\right)^{-m} \left\|y - \theta_g x\right\|^2\right] \\ &= E\left[\left(\sum_{g=1}^{G} \left\|y - \theta_g x\right\|^{-2/(m-1)}\right)^{1-m}\right] \\ &\equiv J_m^{FCR}(\theta) \end{aligned} \tag{8}$$

Significantly, $J_m^{FCR}(\theta)$ has replaced an objective function with weights linked to the parameters with a nonlinear function in (the norm of) the group-specific errors $\left\|y - \theta_g x\right\|$, since the weights themselves are simply a function of those errors.

Assumption 1 provides standard conditions such that $J_m^{FCR}(\theta)$ approximates $J^{HKM}(\theta)$.

10991255, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jae.2997 by University College London UCL Library Services, Wiley Online Library on [25/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Assumption 1. $E\left[\|y\|^2\right] < \infty, E\left[\|x\|^2\right] < \infty$, and $\theta \in \Theta$, which is compact.

Proposition 1. *Under Assumption 1,*

$$\lim_{m \to 1^+} J_m^{FCR}(\theta) = J^{HKM}(\theta).$$

Proposition 1 shows that the FCR objective function converges to the HKM objective function as $m \to 1$ from above, so setting the clustering parameter close to 1 approximates the GFE clustering problem using the FCR objective function. However, for $m > 1$, it retains the continuity of the FCR membership function, and the benefits we describe below. How does the FCR solution behave as $m \to 1^+$? In the limit,

$$\tilde{\theta}_1^{FCR} = \arg\min \lim_{m \to 1^+} J_m^{FCR}(\theta) = \tilde{\theta}. \tag{9}$$

The minimizer of the limiting objective function coincides with the GFE minimizer, $\tilde{\theta}$. However, $\tilde{\theta}_m^{FCR}$ does not converge to $\tilde{\theta}$, since the convergence of J_m^{FCR} is non-uniform due to the exponential structure of the weights. If the true model is (1), then for fixed $m > 1$, $\tilde{\theta}_m^{FCR}$ is a pseudo-true parameter. This is because it minimizes an objective function that does not align with the true model, since the weights are not indicator functions; $\tilde{\theta}_m^{FCR}$ converges to the limit of these pseudo-true parameters. However, unless $T \to \infty$, $\tilde{\theta}$, the K-means solution, is also a pseudo-true parameter, due to uncertainty in the membership function. Thus, in short panels, whether the pseudo-true parameters associated with FCR or K-means will be closer to the parameters is entirely application-specific; any deviation of the FCR estimator from K-means may actually bring it closer to the true value. Indeed, the simulation study in Section 3.2 shows that, at least in finite samples, the bias of FCR for all parameters is in fact lower than that of K-means in all specifications.

2.2 | FCR as GMM

The FCR objective in Equation (8) is differentiable for fixed m, and the first-order conditions, stated in Proposition 2, constitute a set of just-identifying moment conditions:

Proposition 2. *The solution $\tilde{\theta}_m^{FCR}$ satisfies the moment conditions*

$$E\left[\left(\sum_{h=1}^{G} \frac{\|y_i - \theta_g x_i\|^{2/(m-1)}}{\|y_i - \theta_h x_i\|^{2/(m-1)}}\right)^{-m} \left(y_{it} - \theta_{g(t)} x_i\right) x_i\right] = 0 \;\; for \;\; g = 1, \ldots, G \;\; and \;\; t = 1, \ldots, T, \tag{10}$$

where t indexes dimensions of y_i and (t) rows of θ_g; FCR is a GMM problem.

This result has two main implications. First, $\tilde{\theta}_m^{FCR}$ can be estimated via direct minimization of the FCR objective function. This is a standard nonlinear minimization problem. Typically, clustering problems are solved iteratively, alternating between assignment (or computing group weights), then re-estimating model parameters. The econometrician is required to search over many possible group assignments, particularly in the case of binary weights, as in the GFE estimator. Such algorithms, where individual observations or subsets of observations are systematically reallocated between groups from one iteration to the next, are susceptible to local minima. Moreover, they are not well suited to parallelization, since the search over possible groupings is usually conducted sequentially. While the direct minimization required by FCR may also be prone to local minima since it is non-convex, it is straightforward to parallelize across start values to mitigate such concerns, and the computation is fast enough to facilitate many start values.

Second, the asymptotic properties of both HKM (or GFE) and FCM problems have previously proven difficult to establish, requiring extensive technical arguments (e.g., Bonhomme & Manresa, 2015; Pollard, 1981; 1982; Yang, 1994; Yang & Yu, 1992). However, Proposition 2 shows that the FCR clustering problem is simply a GMM problem, and the asymptotic properties of the estimator (for fixed m) follow by standard arguments. Define $\eta(\theta, y_i, x_i)$ as the $(G \times T \times K) \times 1$ vector-valued moment function that stacks (10) for $g = 1, \ldots, G$ and $t = 1, \ldots, T$, and $\hat{\theta}_m^{FCR}$ as the value that minimizes the associated sample objective function. Under standard assumptions, it follows that

$$\sqrt{N}\left(\hat{\theta}_m^{FCR} - \tilde{\theta}_m^{FCR}\right) \xrightarrow{d} \mathcal{N}(0, H^{-1} V H^{-1}),$$

10991255, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jae.2997 by University College London UCL Library Services, Wiley Online Library on [25/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where H and V are the Jacobian and moment covariance, respectively. Appendix S1 contains a formal discussion. It is straightforward to impose restrictions on various elements of θ, either that certain parameters are homogeneous across groups or time-invariant.

To summarize, we have argued that for suitably chosen m close to 1, the FCR objective function approximates the GFE objective function, with equality in the limit. The FCR problem facilitates a GMM implementation with direct minimization, unlike the iterative procedures used to implement most clustering algorithms. For fixed m, the corresponding finite sample estimator is consistent for its population counterpart, which, in the limit, is equal to the GFE solution. The estimator is also asymptotically normal, and analytical standard errors are available for the parameter estimates.

We focus on FCR as a tool to approximate GFE, but the results above hold for any $m > 1$. FCR with larger m may be valuable in its own right in many economic applications. Indeed, "fuzzy C-means" algorithms were introduced as a form of "possibilistic" clustering to better accommodate uncertain group membership in realistic datasets, where noise implies cluster membership cannot be discerned with certainty. Setting a higher m flattens the membership function, increasing the uncertainty of assignment. Thus, FCR could be a non-parametric alternative to finite mixture models, such as the Gaussian mixture model considered in Lewis et al. (2022).

3 | APPLICATION AND SIMULATIONS

We first replicate the empirical results of BM using their estimator and replication code. Then, to illustrate our approach, we implement the FCR estimator using the same specifications and extend BM's simulations calibrated to that application to include our proposed estimator. Doing so allows us to directly compare the performance of the two methods in terms of accuracy, coverage, and computational speed. In particular, we consider the panel of countries from Acemoglu et al. (2008), who study the coevolution of income and democracy from 1970 to 2000.[1] Our specification takes the following form, where we regress an index of democracy (the Freedom House indicator) on its lagged value, lagged log GDP per capita, and the group-time effect α_{gt}:

$$democracy_{it} = \beta_1 democracy_{it-1} + \beta_2 logGDPpc_{it-1} + \alpha_{\tilde{g}_i t} + v_{it}, \tag{11}$$

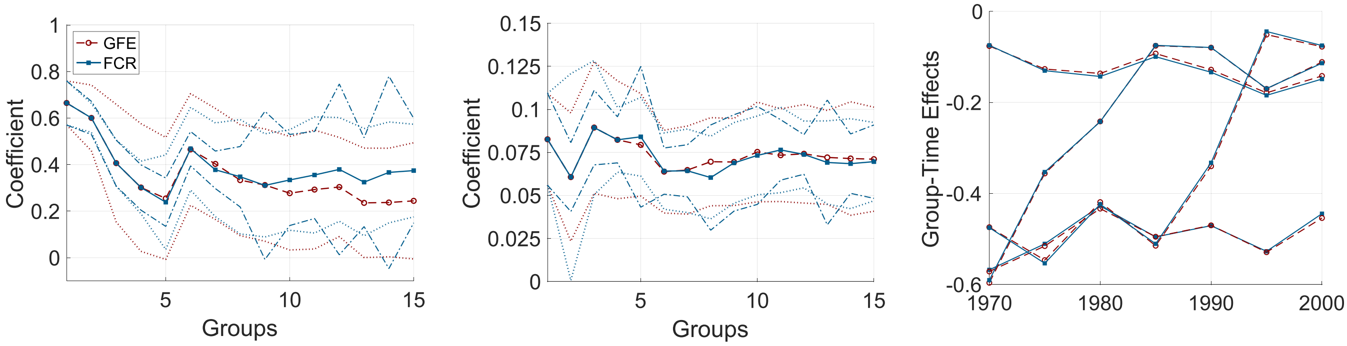
where $\tilde{g}_i$ denotes the group membership of country i. The "fuzziness" of the FCR membership function is governed by the clustering parameter m, with group assignment becoming binary as $m \to 1^+$. Following experimentation, we set $m = 1.001$ to replicate the BM estimates. In practice, this generates group weights equal to either 0 or 1 to 6 decimal places. In our main results, we consider 1000 starting values for minimization, which we find offers the best mix of performance and computational speed; Table S1 shows the effect of varying the number of start values. Our algorithm can be parallelized across starting values, and our main estimation is run with 250 parallel cores.

3.1 | Replication of BM results

First, we replicate the empirical results reported in BM using their estimator and replication code. Estimates for β_1 and β_2 are plotted for $G = 1$ to $G = 15$ by the dashed line with circles in Figure 1a,b; they are identical to those reported in BM. Our approach is able to closely replicate these empirical results: FCR estimates are plotted by the solid line with squares. The point estimates are nearly identical, especially for small G; the average absolute difference is 0.040 for β_1 and 0.002 for β_2. Dashed lines represent 95% bootstrap confidence intervals, with dash-dotted lines plotting analytical errors for comparison. Our analytical standard errors tend to be somewhat narrower than both bootstrap intervals, since while both take $T = 7$ as finite, the bootstrap additionally accounts for finite N. Bootstrapping introduces substantial additional computational burden, as the model is re-estimated for each bootstrapped sample, making the computational speed of FCR, discussed below, very appealing. Figure 1c plots the fixed effects for $G = 4$; again, the estimates closely match BM, identical to at least 3 decimal places with average absolute difference 0.0004.

[1] We use the balanced panel from Acemoglu et al. (2008), which includes 90 countries observed over seven five-year periods. All data files are available for download at https://economics.mit.edu/files/5000.

10991255, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jae.2997 by University College London UCL Library Services, Wiley Online Library on [25/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) Democracy (β_1) (b) Income (β_2) (c) Fixed-effects, $G = 4$

FIGURE 1 Empirical coefficient estimates. *Note*: (a) and (b) plot FCR (solid blue lines with squares) and GFE estimates (red dashed lines with circles) for the lagged democracy and lagged income coefficients as the number of groups varies. Dotted lines show 95% bootstrapped confidence intervals and dash-dot 95% analytical confidence intervals for FCR. (c) plots group fixed effects over time for $G = 4$.

TABLE 1 Simulation performance.

	$G = 3$		$G = 5$		$G = 10$	
	FCR	**GFE**	**FCR**	**GFE**	**FCR**	**GFE**
Bias						
Average lagged democracy bias	0.035	0.084	0.042	0.056	0.051	0.054
Average lagged democracy RMSE	0.043	0.094	0.056	0.070	0.067	0.075
Average lagged income bias	0.013	0.032	0.010	0.007	0.009	0.013
Average lagged income RMSE	0.016	0.035	0.012	0.017	0.012	0.015
Group Misclassification						
Average misclassification rate	9.37%	9.50%	7.69%	9.68%	16.11%	44.73%
Inference with analytical standard errors						
Median lagged democracy standard errors	0.051	0.051	0.050	0.068	0.056	0.048
Coverage rate for lagged democracy	0.894	0.790	0.911	0.840	0.937	0.940
Median lagged income standard errors	0.011	0.013	0.010	0.014	0.014	0.010
Coverage rate for lagged income	0.885	0.840	0.938	0.960	0.939	0.930
Inference with bootstrapped standard errors						
Median lagged democracy standard errors	0.070	0.068	0.094	0.097	0.093	0.091
Coverage rate for lagged democracy	0.982	0.970	0.975	0.973	0.994	0.992
Median lagged income standard errors	0.014	0.016	0.015	0.016	0.016	0.016
Coverage rate for lagged income	0.925	0.914	0.935	0.943	0.995	0.996
Computation time						
Total time (seconds)	17.5	24.8	27.6	26.5	127.4	78.2

Note: This table compares the performance of the FCR and GFE estimators on a simulated panel with $N = 90$ and $T = 7$. GFE results come from Table S3 and Table S4 in their paper. While BM's preferred implementation of their Algorithm 2 uses 10 starting values, 10 neighbors, and maximum steps of 10, they use only 5 starting values and 5 maximum steps in simulations, since their baseline specification "resulted in prohibitive computation times." For consistency within this table, we calculate computation times ourselves under this faster specification. GFE analytical standard errors and coverage use the Pollard (1982) fixed-T formula, reported in columns (2) of BM Table S7. The GFE bootstrapped standard errors are reported in columns (3) of BM Table S7. We compute the FCR bootstrapped standard errors following the same procedure as BM. Bias, misclassification, computation time, and non-rejection probability are means across 1000 simulations, while the standard errors are medians (to match the reported estimates in BM). Coverage rates are for nominal 95% confidence intervals. For FCR, we use 1000 starting values and 250 parallel cores; the reported computation time is the total across these starting values, not the time per starting value.

3.2 | Simulations and FCR performance

Having obtained very similar estimates empirically, we now compare the performance of FCR and GFE in simulations. We start from the simulation code found in BM's replication package and extend it to include our estimator. In particular, we simulate a panel of the same size as the application ($N = 90$, $T = 7$), where data are generated by first estimating the GFE model on the empirical dataset and then using the resulting group-time, common coefficient, and binary group

10991255, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jae.2997 by University College London UCL Library Services, Wiley Online Library on [25/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

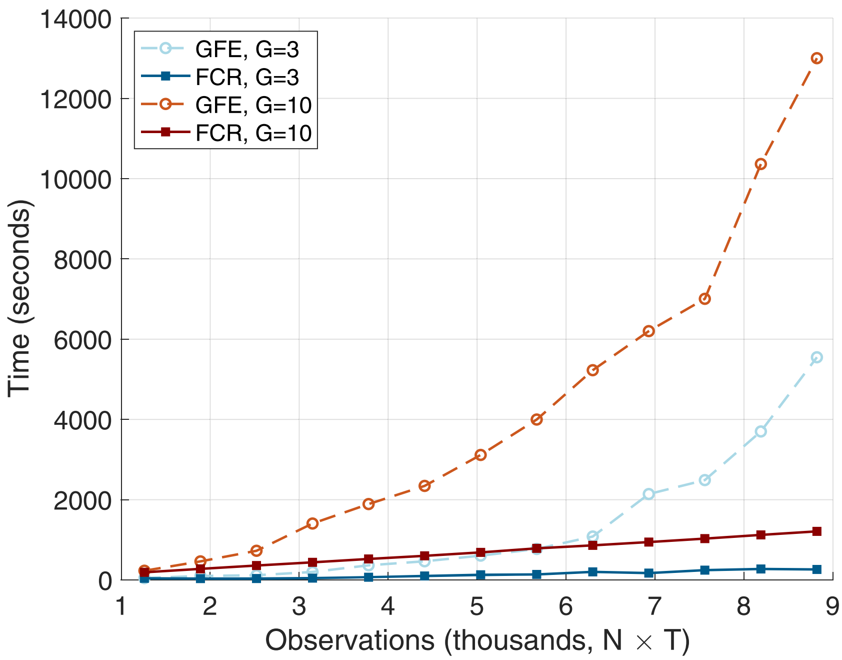


FIGURE 2 Computation time by dataset size. *Note*: Computation time by dataset size for the FCR and GFE estimators. The horizontal axis indicates the total number of observations ($N \times T$), where T is fixed at 7 and N increased. The solid blue line shows computation time for FCR with 3 groups, the solid red for 10 groups, the dashed blue GFE with 3 groups, and the dashed red GFE with 10 groups. The BM GFE estimator is run using their preferred implementation, "Algorithm 2" with 10 simulations, 10 neighbors, and maximum steps of 10. The FCR algorithm is run with 1000 starting values and 250 parallel cores.

membership estimates to create the DGP. In particular,

$$y_{it}^{s} = \hat{\alpha}_{g_i t} + \hat{\beta}_1 democracy_{it-1} + \hat{\beta}_2 logGDPpc_{it-1} + v_{it}^{s}$$

where $\hat{g}_i$ is the estimated group membership from BM, $\hat{\alpha}_{g_i t}$ is the estimated group-time effect, $\hat{\beta}_1$ and $\hat{\beta}_2$ are the estimated common coefficients, and the errors v_{it}^{s} are i.i.d. normal draws for simulation sample s, with variance equal to the mean squared residual. We consider 1000 Monte Carlo samples.

The simulation results are presented in Table 1, where we report the performance of the two estimators in terms of bias, group misclassification rate, coverage, and computation time. FCR compares favorably on all measures, particularly bias when G is small and misclassification when G is large. When $G = 10$, for example, the GFE estimator classifies 44.7% of units incorrectly, while FCR misclassifies only 16.11%. Across simulations, computation time is similar for the two estimators, with GFE slightly faster for some specifications given the particular choice of tuning parameters. However, note that the tuning parameters used in BM's simulations are not their recommended tuning parameters, compared to which FCR proves universally faster.[2] In terms of inference, the median standard errors for the common coefficients are closely comparable across the two estimators. On the other hand, the coverage rate of the 95% confidence intervals based on analytical standard errors is generally much closer to its nominal level for FCR than for BM's GFE; bootstrapped errors yield similar rates for both estimators.

The fuzzy clustering approximation is particularly valuable in settings with large panels, which are increasingly common in empirical work. Specifically, the smooth weights incorporated in the FCR objective permit direct minimization in a single step, which is substantially faster than previously implemented approaches, even after parallelization across many start values. To illustrate this point, we plot computation time by dataset size in Figure 2, where the data are expanded in the N dimension. We construct these larger datasets by stacking the simulated panels from our first exercise, with independent error draws for each observation. While FCR remains quite efficient and appears to scale linearly with number of observations, computation time for the GFE estimator becomes prohibitive even on moderately sized datasets and grows nonlinearly.[3] Thus, fuzzy clustering is a fast and easy way to estimate grouped patterns of heterogeneity, with particular advantages in larger datasets.

[2]As in BM, GFE simulation results are presented with a faster Algorithm 2 (see note to the table) than their preferred specification. BM report that computation time for their preferred algorithm is 38.4 and 228.4 s for 3 and 10 groups, respectively, thus slower than FCR. We present a more thorough analysis of computation time below, revealing that FCR is considerably faster on larger datasets.

[3]Computational time for GFE appears to grow exponentially in the size of the dataset. This is likely related to the established NP-hardness of the K-means clustering problem (Aloise et al., 2009). We thank an anonymous referee for pointing this out.

10991255, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jae.2997 by University College London UCL Library Services, Wiley Online Library on [25/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4 | CONCLUSION

We generalize fuzzy C-means clustering to regression problems. As the clustering parameter approaches 1, the resulting estimator approximates the GFE estimator. The FCR estimator is a GMM estimator, with standard asymptotic distribution theory, and offers substantial improvements in computational efficiency. In simulations, it also shows reduced bias and misclassification compared to GFE and yields quantitatively similar results empirically.

DATA AVAILABILITY STATEMENT

The data supporting this paper are freely available from the journal's replication archive, https://journaldata.zbw.eu/journals/jae.

REFERENCES

Acemoglu, D., Johnson, S., Robinson, J. A., & Yared, P. (2008). Income and democracy. *American Economic Review*, *98*(3), 808–42.

Aloise, D., Deshpande, A., Hansen, P., & Popat, P. (2009). NP-hardness of Euclidean sum-of-squares clustering. *Machine learning*, *75*(2), 245–248.

Bezdek, J. (1981). *Pattern recognition with fuzzy objective function algorithms*: Plenum Press.

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

Chetverikov, D., & Manresa, E. (2022). Spectral and post-spectral estimators for grouped panel data models.

Kim, J., & Wang, L. (2019). Hidden group patterns in democracy developments: Bayesian inference for grouped heterogeneity. *Journal of Applied Econometrics*, *34*(6), 1016–1028.

Lewis, D. J., Melcangi, D., & Pilossoph, L. (2022). Latent heterogeneity in the marginal propensity to consume. (*902*). New York, NY.

Mugnier, M. (2022). A simple and computationally trivial estimator for grouped fixed effects models.

Pollard, D. (1981). Strong consistency of k-means clustering. *The Annals of Statistics*, *9*(1), 135–140.

Pollard, D. (1982). A central limit theorem for k-means clustering. *The Annals of Probability*, *10*(4), 919–926.

Su, L., & Ju, G. (2018). Identifying latent grouped patterns in panel data models with interactive fixed effects. *Journal of Econometrics*, *206*(2), 554–573.

Su, L., Shi, Z., & Phillips, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica*, *84*(6), 2215–2264.

Yang, M.-S. (1994). On asymptotic normality of a class of fuzzy c-means clustering procedures. *International Journal of General Systems*, *22*(4), 391–403.

Yang, M.-S., & Yu, K. F. (1992). On existence and strong consistency of a class of fuzzy c-means clustering procedures. *Cybernetics and Systems*, *23*(6), 583–602.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of the article.

How to cite this article: Lewis D. J., Melcangi D., Pilossoph L., & Toner-Rodgers A. (2023). Approximating grouped fixed effects estimation via fuzzy clustering regression. *Journal of Applied Econometrics*, 1–8. https://doi.org/10.1002/jae.2997

10991255, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jae.2997 by University College London UCL Library Services, Wiley Online Library on [25/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License