DE GRUYTER DOI 10.1515/jem-2012-0018 — Journal of Econometric Methods 2013; 2(1): 35–51

Research Article

Partha Deb and Pravin K. Trivedi*

Finite Mixture for Panels with Fixed Effects

Abstract: This paper develops finite mixture models with fixed effects for two families of distributions for which the incidental parameter problem has a solution. Analytical results are provided for mixtures of Normals and mixtures of Poisson. We provide algorithms based on the expectations-maximization (EM) approach as well as computationally simpler equivalent estimators that can be used in the case of the mixtures of normals. We design and implement a Monte Carlo study that examines the finite sample performance of the proposed estimator and also compares it with other estimators such the Mundlak-Chamberlain conditionally correlated random effects estimator. The results of Monte Carlo experiments suggest that our proposed estimators of such models have excellent finite sample properties, even in the case of relatively small T and moderately sized N dimensions. The methods are applied to models of healthcare expenditures and counts of utilization using data from the Health and Retirement Study.

***Corresponding author: Pravin K. Trivedi,** School of Economics, University of Queensland, Level 6, Colin Clark Building, St. Lucia, QLD 4072, Australia, E-mail: p.trivedi@uq.edu.au; and Department of Economics, Wylie Hall 105, Indiana University, Bloomington, IN 47405, USA
Partha Deb: Department of Economics, Hunter College and the Graduate Center, City University of New York, and National Bureau of Economic Research

1 Introduction

In analyzing panel data in the framework of linear models, random and fixed effects models provide a central paradigm for data analysis. Finite mixture (or " latent class") models (FMM), which are appealing generally because of the additional flexibility they offer within the parametric context, have been used extensively for modeling cross-section data and also, more recently for modeling panel data. The extensions to panel data have been for either pooled or population-averaged (PA) models or the random effects (RE) models; see, for example, Skrondal and Rabe-Hesketh (2004) and Bago d'Uva (2005). There appear to have been no attempts to combine finite mixtures and fixed effects (FE). This paper takes the first step in this direction, motivated by the fact that the fixed effects model has a special place in the microeconometrics panel data literature. The fixed effect framework has considerable appeal because it makes weaker, and, perhaps more plausible, assumptions about the correlation between the unobserved individual specific effects and the observed regressors included in the model.

Important advances in analyzing finite mixture models appeared in the statistical literature in the 1960s and 1970s. Path-breaking papers by Dempster, Laird and Rubin (1977) and Aitkin and Rubin (1985) which introduced the expectations-maximization (EM) algorithm made the computation of the finite mixture (latent class) models accessible to applied researchers. The monograph by McLachlan and Peel (2004) documents the enormous popularity of the mixture formulation in many areas of statistics and highlights the importance of the EM algorithm as the estimation method of choice in numerous applications. In cross-section econometrics finite mixture models have proved to be a useful way of modeling discrete unobserved heterogeneity in the population based on the intuitive idea that different "types" may correspond to different latent classes or subpopulations (Heckman and Singer 1984; Deb and Trivedi 1997, 2002; Conway and Deb 2005). Time-series applications of finite mixtures, known as switching models, have also been popular in macro economics and are thoroughly analyzed in Frühwirth-Schnatter (2006) from a Bayesian perspective. The key idea is that the unknown population distribution may be empirically approximated by a mixture of distributions with finite, but usually small, number of mixture components. For example, a mixture of normals has been extensively used as an approximating distribution for continuous outcomes. In empirical applications of finite mixtures to panel data using a relatively small number of components, however, some latent classes may still show substantial within-class heterogeneity. Hence adding either fixed or random effects can improve the fit of the model. Adding random effects usually leads to greater computational complexity, whereas the effects of introducing fixed effects have not been previously studied.

In microeconometrics both random effect (RE) and fixed effect (FE) models are widely used to account for unobserved heterogeneity. In a fixed effects model individual-specific effect α_i, also called an incidental parameter, is an unobserved random variable that may be correlated with the regressors $\mathbf{x}_{it}$, $i=1,\ldots,N$ $t=1,\ldots,T$. In a random effect model α_i are assumed to be uncorrelated with the regressors. The common set-up of a panel model in microeconometric applications, and in this paper, is "large-N – small-T," commonly referred to as a short panel. Within this set-up consistent estimation of the incidental parameters is not possible. The solution of this incidental parameters problem is a classic problem.

In semiparametric linear models the standard way of handling fixed effects is to eliminate them by either first differencing or applying the "within or deviations-from-mean" transformations. This step sweeps out both the individual fixed effects (interpreted also as nuisance parameters) as well as all time-invariant variables. The remaining parameters can then be estimated consistently by a conditional moment type estimator. In some semiparametrically-specified nonlinear models a similar approach based on a mean-scaling transformation can be applied also. However, in nonlinear models some standard ways of handling fixed effects do not necessarily work, so the FE model, and by implication the fixed effects finite mixture model (FE-FMM), is potentially problematic to estimate.

This paper studies parametrically specified finite mixture models. In such cases another way to sweep out the incidental parameters and estimate the remaining common parameters is to use the conditional maximum likelihood (CML) approach of Andersen (1970); see also Severini (2000). The application of CML requires that there is a known sufficient statistic for the nuisance parameters; then the likelihood is maximized conditional on the data and this sufficient statistic. This is equivalent to concentrating out of the likelihood the nuisance parameters and working with the concentrated likelihood function. An alternative terminology for this is profile likelihood maximization. Unfortunately, such conditioning is known to be feasible only in a limited number of nonlinear models Lancaster (2000, 2002). How to do this in a finite mixture model appears not to have been investigated. Developing a method to do so for two leading finite mixture models is the objective of this paper. We also consider the related issue of developing computationally efficient algorithms for estimating these models.

Our solution of the FE-FMM estimation problem is based on two insights. The first is that if the incidental parameter problem can be solved for each component of the mixture, then a mixture should be formed after concentrating out the incidental parameters. Second, maximization of the resulting concentrated mixture likelihood can be accomplished in some cases by direct nonlinear optimization of the mixture likelihood, and in other cases by the application of the EM algorithm to the "full-data" variant of the concentrated mixture likelihood. We develop this approach in the context of Normal and the Poisson two-component mixture models. Numerical illustrations using both Monte Carlo simulation and real data are provided.

The remainder of this paper is organized as follows. Section 2 establishes the notation, the context, and the solutions for the standard one-component model which is used to generate the mixture. Section 3 introduces the extension to finite mixtures. Section 4 develops the EM algorithm. Section 5 reports the results of a Monte Carlo study of several estimators. Section 6 provides two empirical illustrations. Section 7 concludes with discussion and remarks.

2 Fixed Effects in Panel Data Models

For simplicity and to establish notation we begin with the familiar linear panel data model with the outcome variable denoted as y_{it}, the covariates denoted as a K-component vector of time-varying exogenous regressors $\mathbf{x}_{it}$, and individual specific effects α_i, where $i=1,\ldots,N$ and $t=1,\ldots,T$. The regression specification is

$$y_{it}=\alpha_i+\mathbf{x}_{it}^{'}\boldsymbol{\beta}+\varepsilon_{it} \tag{1}$$

Adding time-invariant regressors does not cause any additional complications and hence are omitted for simplicity. As previously mentioned the set-up assumes the large-N – small-T setting. Lagged dependent variables and endogenous regressors are excluded. For simplicity we assume a strongly balanced panel, though this restriction can be relaxed with some additional notation.

In the RE model the individual-specific effect is uncorrelated with the regressors and the unobserved i.i.d. errors ε_{it} so that $\mathrm{E}[\alpha_i|\mathbf{x}_{it}^{'},\varepsilon_{it}]=0$, $t=1,\ldots,T$. The individual-specific FE model consists of (1) plus the assumptions $\mathrm{E}[\alpha_i\,|\,\mathbf{x}_{it}]\neq 0$, and $\mathrm{E}[\varepsilon_{it}\,|\,\mathbf{x}_{it}]=0$, $i=1,\ldots,N;$ $t=1,\ldots,T.$

2.1 Incidental Parameters Problem

The most direct approach is to jointly estimate $\alpha_1,\ldots,\alpha_N$ and $\boldsymbol{\beta}$. Consistent estimation then relies on large-N – large-T asymptotics. But in microeconometric applications a short panel is more plausible, and asymptotic theory assumes only that T is fixed while $N\to\infty$. This raises the possibility that the joint estimator of $(\alpha_1,\ldots,\alpha_N,\boldsymbol{\beta})$ will be inconsistent when T is "small." We note, however, that some have argued that the approach may still "work," at least for some nonlinear models, when $T \simeq 10$. In essence this is an extension of the "least-squares-with-dummy-variables" approach for linear panel models to nonlinear panel models; see Heckman (1981), Greene (2004a,b) and Allison (2009). Drawing support from Monte Carlo evidence, Greene argues that for some nonlinear panel models the dummy variable approach produces satisfactory results even for relatively small T. Lancaster (2000) and Severini (2000) survey a number of other approaches for handling the incidental parameter problem, including especially the conditional likelihood approach. Dhaene and Jochmans (2011) provides a further analysis covering a number of nonlinear models which suffer from the incidental parameter problem. Arellano and Hahn (2007) provides an analysis of recent developments dealing with biases in estimation of nonlinear panel models.

More commonly in short panels, transformations are applied to eliminate the individual-specific fixed effect parameters α_i after which the main interest lies in estimating identifiable components of $\boldsymbol{\beta}$. In linear models with additive fixed effects, first differencing and "within" transformations are leading examples of this approach. In nonlinear models with multiplicative fixed effects, this approach has been extended to moment-based estimators for some nonlinear panels (Chamberlain 1992). Quasi-differencing analogs of within- and first-difference transformations are available for specific nonlinear panel models (Cameron and Trivedi 2005). Such semiparametric approaches are difficult to extend to finite mixture models. We, therefore, pursue solutions based on conditional maximum likelihood applied after conditioning on a sufficient statistic for α_i. In Normal linear models, this approach is also equivalent applying the within-transformation.

2.1.1 Normal Regression

Given (1), assume $\varepsilon_{it} \sim \mathbf{N}[0,\sigma^2]$. Denote the sample means $T^{-1}\Sigma y_{it}$ and $T^{-1}\Sigma \boldsymbol{x}_{it}$ by $(\bar{y}_i, \bar{\boldsymbol{x}}_i)$ respectively. Then the within transformation eliminates the fixed effect α_i and yields

$$y_{it}-\bar{y}_i=(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)'\boldsymbol{\beta}+(\varepsilon_{it}-\bar{\varepsilon}_i).$$

If, instead, a first-differencing transformation is applied, we obtain

$$y_{it}-y_{i,t-1}=(\mathbf{x}_{it}-\mathbf{x}_{i,t-1})'\boldsymbol{\beta}+(\varepsilon_{it}-\varepsilon_{i,t-1}).$$

where again the fixed effects do not appear in the transformed model.

Next consider conditional likelihood. The sufficient statistic for α_i is $\Sigma_{i,t} y_{it}$, or $T\bar{y}_i$. Then, under normality assumption for ε_{it}, MLE of $\boldsymbol{\beta}$ (ignoring σ^2) is based on the conditional likelihood:

$$\begin{aligned} L_{\mathrm{COND}}(\boldsymbol{\beta},\sigma^2,\alpha) &= \prod_{i=1}^{N} f(y_{i1},\dots,y_{iT}\,|\,\bar{y}_i) \\ &= \prod_{i=1}^{N} \frac{f(y_{i1},\dots,y_{iT},\bar{y}_i)}{f(\bar{y}_i)} \\ &= \prod_{i=1}^{N} \frac{(2\pi\sigma^2)^{-T/2}}{(2\pi\sigma^2/T)^{-1/2}} \exp\left\{ \sum_{t=1}^{T} -\frac{1}{2\sigma^2} \begin{bmatrix} (y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}))^2 \\ +(\bar{y}_i-\bar{\mathbf{x}}_i'\boldsymbol{\beta})^2 \end{bmatrix} \right\}. \end{aligned} \tag{2}$$

The resulting conditional ML estimator $\hat{\boldsymbol{\beta}}_{\mathrm{CML}}$ eliminates the fixed effects and solves the first-order conditions

$$\sum_{t=1}^{T}\sum_{i=1}^{N}[(y_{it}-\bar{y}_i)-(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)'\boldsymbol{\beta})](\mathbf{x}_{it}-\bar{\mathbf{x}}_i)=\mathbf{0}; \tag{3}$$

the solution coincides with the first-order conditions from least squares regression of $(y_{it}-\bar{y}_i)$ on $(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)$. Hence in this case, $\hat{\boldsymbol{\beta}}_{\mathrm{CML}}$ equals the within estimator. Here the individual-specific fixed effects $\alpha_1,\dots,\alpha_N$ are eliminated by conditioning on $\bar{y}_i$, so we may maximize the conditional log-likelihood function with respect to the common parameters only. The estimator $\hat{\boldsymbol{\beta}}_{\mathrm{CML}}$ is consistent, but the estimator $\hat{\sigma}^2_{\scriptscriptstyle CML}=\sum \hat{\varepsilon}^2_t\,/NT$ is biased to $O(T^{-1})$. (Bias adjustment is achieved by instead using the estimator $\sum \hat{\varepsilon}^2_t\,/(N(T-1)-K)$.

2.1.2 Poisson Regression

We next consider the Poisson panel model under the strong assumption that the regressors are strictly exogenous. In the standard formulation, the mean parameter is λ_{it} and the individual specific effect α_i impacts λ_{it} multiplicatively, i.e.,

$$y_{it} \sim \mathrm{P}(\lambda_{it}\alpha_i) \tag{4}$$

$$\lambda_{it}=\exp(\mathbf{x}_{it}'\boldsymbol{\beta}). \tag{5}$$

In this case also the sufficient statistic for α_i is $\sum_t y_{it}$, or $T\bar{y}_i$. So one can apply conditional maximum likelihood. This is equivalent to concentrating out of the likelihood the parameters α_i. The full log-likelihood is

$$\begin{aligned} \ln L(\boldsymbol{\beta},\boldsymbol{\alpha}) &= \ln\left[\prod_i \prod_t \{\exp(-\alpha_i\lambda_{it})(\alpha_i\lambda_{it})^{y_{it}}\,/y_{it}!\}\right] \\ &= \sum_i\left[-\alpha_i\sum_t\lambda_{it}+\ln\alpha_i\sum_t y_{it}+\sum_t y_{it}\ln\lambda_{it}-\sum_t \ln y_{it}!\right]. \end{aligned} \tag{6}$$

The first order conditions with respect to $\alpha, \partial \ln L(\beta,\alpha)/\partial\alpha = \mathbf{0},$ yield

$$\widehat{\alpha}_i = \sum_t y_{it} / \sum_t \lambda_{it} = \overline{y}_i / \overline{\lambda}_i. \tag{7}$$

Substituting (7) back into (6) yields the following concentrated likelihood function, ignoring terms not involving β:

$$\ln L_{\text{conc}}(\beta) \sum_i \sum_t \left[y_{it} \ln \lambda_{it} - y_{it} \ln\left(\sum_s \lambda_{is} \right) \right]. \tag{8}$$

Consistent estimates of β for fixed T and $N \to \infty$ can be obtained by maximization of ln $L_{\text{conc}}(\beta)$.Specifically $\partial \ln L_{\text{conc}}(\beta)/\partial\beta = \mathbf{0}$ gives the first-order conditions

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it} \left(y_{it} - \frac{\lambda_{it}}{\overline{\lambda}_i} \overline{y}_i \right) = \mathbf{0}, \tag{9}$$

where $\overline{\lambda}_i = \sum_t \lambda_{it}/T$. Time-invariant regressors drop out of the ratio $\lambda_{it}/\overline{\lambda}_i$ and their coefficients cannot be identified.

3 Finite Mixture Models with Fixed Effects

The standard definition of a C-component mixture (or C latent classes) of an arbitrary density with $f(y_{it}|\mathbf{x}_{it},\theta_j)$, $j=1,2,\ldots,C$ is

$$\sum_{j=1}^{C} \pi_j f(y_{it} | \mathbf{x}_{it}, \boldsymbol{\theta}_j) \tag{10}$$

where $0 < \pi_j < 1 \; \forall j = 1,2,\ldots,C, \; \sum_j \pi_j = 1;$ i.e.

$$y_{it} \sim f(y_{it}|\mathbf{x}_{it},\theta_j) \text{ with probability } \pi_j. \tag{11}$$

Throughout the paper we will only consider the case of fixed mixing proportions. Likelihood function based on (10–11) is referred to as mixture likelihood. As specified, this specification has no individual-specific effects, indicating that this specification is essentially a pooled data mixture model.

We wish to extend the CML approach to Normal and Poisson finite mixture models. Before considering details note that the above the conditioning approach "works" for the Normal and the Poisson regression but it will not work for the mixture of Normals or mixture of Poissons because in these cases a sufficient statistic for the α_i is not available. However, the approach can work if the mixture components are first purged of fixed effects and we then form a mixture of concentrated marginals. In the following section such an approach will be used.

Let s_i be a sufficient statistic for α_i. Then the mixture definition (10) is expressed with conditioning on s_i as follows:

$$\sum_{j=1}^{C} \pi_j f(y_{it} | \mathbf{x}_{it}, \boldsymbol{\theta}_j, s_i) \tag{12}$$

The estimation objective is to obtain consistent estimates of (π_j,θ_j), $j=1,\ldots,C$.

Computing algorithms for estimation based on expectations-maximization (EM) are discussed in McLachlan and Peel (2004) and Frühwirth-Schnatter (2006). Direct estimation based on the mixture likelihood using variants of Newton's method is also feasible in some cases.

3.1 Normal Mixture

As stated previously, under this formulation we cannot use the concentrated likelihood approach. However, as a sufficient statistic does exist for each component separately, we propose to form the mixture model using the component-wise conditional density. That is, we construct the mixture using the conditional likelihood for each component Normal regression. Thus, for $i=1,\ldots,N$,

$$f_j(\beta_j, \sigma_j^2 | \alpha) = \frac{(2\pi\sigma_j^2)^{-T/2}}{(2\pi\sigma_j^2/T)^{-1/2}} \exp\left\{ \sum_{t=1}^{T} -\frac{1}{2\sigma_j^2} [(y_{it} - \overline{y}_i) - (\mathbf{x}_{it}' - \overline{\mathbf{x}}_i')\beta_j]^2 \right\}, \tag{13}$$

and the mixture distribution is

$$\sum_{j=1}^{C} \pi_j f_j(\beta_j, \sigma_j^2 | \overline{\mathbf{y}}_i, \mathbf{x}_i) \tag{14}$$

and $0 < \pi_j < 1, \forall j = 1,2,\ldots,C, \; \sum_j \pi_j = 1$.

The expression (14) would be convenient to use if the latent class assignment of each observation is given.

As this is not the case, for the purposes of estimation it is more convenient to work with the *full-data likelihood*. This is obtained by first defining d_{it} to be an indicator variable that identifies individual *i's* latent class at time t and introducing the indicator function $1(d_{ji}=1)$. Assume observations are permanently assigned to just one latent class during the panel period, which implies that $d_{it}=d_i$. More precisely,

$$d_{ji} = \begin{cases} 1 \text{ if } i \text{ belongs to the component } j \\ 0 \text{ otherwise} \end{cases}; \tag{15}$$

Of course d_{ji} is not directly observed as classes are latent.

Then the full-data likelihood for this model, under the assumption that the observations are conditionally independent across individuals and over time, based on the concentrated density is

$$L_{conc}(\boldsymbol{\beta},\sigma^2|\alpha)=\prod_{i=1}^{N}\prod_{t=1}^{T}\sum_{j=1}^{C}\left(\pi_j f(\boldsymbol{\beta}_j,\sigma_j^2|\bar{y}_{it},\mathbf{x}_{it})\right)^{1(d_{ji}=1)}. \quad (16)$$

The maximization of this likelihood may be based on the expectations-maximization algorithm. Details are given in section (4) below. It will be shown that, in this specific case of linear panel model, it is possible to base estimation on the mixture likelihood and avoid using the EM algorithm altogether. This computational convenience is a consequence of the fact that the sufficient statistic does not depend upon latent class assignment or unknown parameters. Hence estimation can be based on the concentrated likelihood expression derived using (14). Specifically, one can apply the within transformation to the data and then use a standard program for normal mixtures. If this convenient feature is absent, then an EM algorithm based on full-data likelihood will be required as shown below for the case of Poisson mixtures with fixed effects.

3.2 Poisson Mixture

For a fixed effect Poisson mixture we again exploit the result that for each component of the mixture distribution a sufficient statistic is available. Therefore, we can derive a finite mixture-based likelihood function based on the conditional (on the sufficient statistic) component distributions.

Specifically, given

$$y_{it}|\mathbf{x}_{it},\lambda_{it}^{(j)},\alpha_i\sim \mathrm{P}[\alpha_i\lambda_{it}^{(j)}],\ j=1,2,...,C \quad (17)$$

where $\lambda_{it}^{(j)}=\exp(\mathbf{x}_{it}'\boldsymbol{\beta}_j)$, using (7) for mixture component j gives

$$y_{it}|\mathbf{x}_{it},\lambda_{it}^{(j)},\hat{\alpha}_i\sim \mathrm{P}\left[\left(\sum_t y_{it}/\sum_t\lambda_{it}^{(j)}\right)\lambda_{it}^{(j)}\right],$$

and the C-component mixture distribution is

$$y_{it}|\mathbf{x}_{it},\lambda_{it},\hat{\alpha}_i\sim\sum_{j=1}^{C}\pi_j \mathrm{P}\left[\left(\sum_t y_{it}/\sum_t\lambda_{it}^{(j)}\right)\lambda_{it}^{(j)}\right]. \quad (18)$$

Then the full-data concentrated likelihood is derived by combining (18) and (15) which yields

$$L_{conc}(\cdot)=\prod_{i=1}^{N}\prod_{t=1}^{T}\sum_{j=1}^{C}\left(\pi_j \mathrm{P}\left(\left(\sum_t y_{it}/\sum_t\lambda_{it}^{(j)}\right)\lambda_{it}^{(j)}\right)\right)^{1(d_{ji}=1)}. \quad (19)$$

Again this likelihood can be maximized using the EM algorithm. Note that in this case the sufficient statistic depends on model parameters thus the EM algorithm is required.

4 EM Algorithm

Consider a panel with units $i=1,2,...,N$ observed at times $t=1,2,...,T$. Suppose that an observation y_{it} can be drawn from one of C latent classes, each of which has a density $f(y_{it};\boldsymbol{\theta}_j)$. Here $f(y_{it};\boldsymbol{\theta}_j)$ is shorthand for the jth component of concentrated likelihood and $\boldsymbol{\theta}_j$ is generic notation for common parameters in component j. In the Poisson case $\boldsymbol{\theta}_j=\boldsymbol{\beta}_j$. Let $y_i=(y_{i1},y_{i2},...,y_{iT})$ be the vector of observed values for unit i. Let $\mathbf{d}_i=(d_{1i},d_{2i},...,d_{Ci})$ define a set of indicator variables such that $d_{ji}=1$ if the unit i was drawn from the latent class j; $d_{ji}=0$ otherwise and $\sum_j d_{ij}=1$. Then, the *panel* finite mixture model specifies that $(\mathbf{y}_i|\mathbf{d}_i,\boldsymbol{\theta},\boldsymbol{\pi})$ are independently distributed with densities

$$\prod_{j=1}^{C}\left(f(y_{i1};\boldsymbol{\theta}_j)\times f(y_{i2};\boldsymbol{\theta}_j)\times\ldots\times f(y_{iT};\boldsymbol{\theta}_j)\right)^{d_{ji}}$$
$$=\prod_{j=1}^{C}\left(\prod_{t=1}^{T}f_j(y_{it};\boldsymbol{\theta}_j)\right)^{d_{ji}}$$

and $(d_{ji}|\boldsymbol{\theta},\boldsymbol{\pi})$ are i.i.d. with multinomial distribution

$$\prod_{j=1}^{C}\pi_j^{d_{ji}},\quad 0<\pi_j<1,\sum_{j=1}^{C}\pi_j=1.$$

Thus

$$(y_{i1},y_{i2},\ldots,y_{iT}|\boldsymbol{\theta},\boldsymbol{\pi})\sim\sum_{j=1}^{C}\left(\pi_j\prod_{t=1}^{T}f_j(y_{it};\boldsymbol{\theta}_j)\right)^{d_{ji}},$$

where $\boldsymbol{\theta}=(\boldsymbol{\theta}_1,...,\boldsymbol{\theta}_C)$ The likelihood function is then

$$L(\boldsymbol{\theta},\boldsymbol{\pi}|\mathbf{y})=\prod_{i=1}^{N}\sum_{j=1}^{C}\left(\pi_j\prod_{t=1}^{T}f_j(y_{it};\boldsymbol{\theta}_j)\right)^{d_{ji}}. \quad (20)$$

and the log likelihood function is

$$\ln L(\boldsymbol{\theta},\boldsymbol{\pi}|\mathbf{y})=\sum_{i=1}^{N}\sum_{j=1}^{C}d_{ji}\left(\ln(\pi_j)+\sum_{t=1}^{T}\ln(f_j(y_{it};\boldsymbol{\theta}_j))\right) \quad (21)$$

Replacing d_{ji} by its expected value, $\mathrm{E}\left[d_{ji}\right]=\hat{z}_{ji}$, yields the expected log-likelihood (EL),

$$\mathrm{EL}(\boldsymbol{\theta}|\mathbf{y},\boldsymbol{\pi})=\sum_{i=1}^{N}\sum_{j=1}^{C}\hat{z}_{ji}\left[\ln f_j(\mathbf{y}_i;\boldsymbol{\theta}_j)+\ln\pi_j\right]. \quad (22)$$

The M-step of the EM procedure maximizes (22) by solving the first order conditions

$$\hat{\pi}_j - \frac{\sum_{i=1}^{N} \hat{z}_{ji}}{N} = 0, \quad j = 1,..,C \tag{23}$$

$$\sum_{i=1}^{N} \sum_{j=1}^{C} \hat{z}_{ji} \frac{\partial \ln f_j(\mathbf{y}_i; \boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} = \mathbf{0}. \tag{24}$$

The marginal probability that an observation comes from the class j is the average of all individual observation probabilities of coming from the jth population. The E-step of the EM procedure obtains new values of $\mathrm{E}[d_{ji}]$ using the (23).

The posterior probability that unit i belongs to population j, $j=1,2,\ldots C$, denoted z_{ji} is defined as

$$z_{ji} = \frac{\pi_j \prod_{t=1}^{T} f_j(y_{it}; \boldsymbol{\theta}_j)}{\sum_{j=1}^{C} \pi_j \prod_{t=1}^{T} f_j(y_{it}; \boldsymbol{\theta}_j)}. \tag{25}$$

and is recomputed as new estimates of $(\pi_j, \boldsymbol{\theta}_j)$ become available.

For a given set of parameters, $\{\boldsymbol{\theta}_j, \pi_j\}_{j=1,2,\ldots,C}$, and the data, the E-step consists of calculating z_{ji} which defines the set of posterior probabilities of classification for each unit, and also the values of $\{\pi_j\}_{j=1,2,\ldots,C}$ for the next M-step. Given $\{\pi_j\}_{j=1,2,\ldots,C}$ the M-step consists of maximizing the (22). The E- and M-steps are repeated in alternating fashion until the expected log likelihood fails to increase.

4.1 Variance Estimation

Unlike Newton-Raphson type gradient based algorithm applied to the mixture likelihood, the EM algorithm does not automatically generate an estimate of the information matrix required for estimating the asymptotic variance of the maximum likelihood estimator Louis (1982), Oakes (1999). When the analytical expressions for the information matrix components are available then these can be evaluated at the converged EM estimates and used to generate the variance estimates. However, these expressions are complicated for mixture models and alternative methods need to be considered.

A convenient method of Louis (1982) and Oakes (1999) uses the observed information matrix $\mathbf{I}(\hat{\boldsymbol{\theta}}_{EM})$, which is conditioned on the $\hat{z}_{ji}$, and the likelihood gradient $\mathbf{G}(\hat{\boldsymbol{\theta}}_{EM})$, both from the full-data log-likelihood evaluated at the converged value. The Hessian is assumed to be nonsingular positive definite. Then, as shown in Louis (1982),

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_{MLE}) = \mathbf{I}(\hat{\boldsymbol{\theta}}_{EM}) - \mathbf{G}(\hat{\boldsymbol{\theta}}_{EM}) \mathbf{G}'(\hat{\boldsymbol{\theta}}_{EM}) \tag{26}$$

Once the EM estimates of $\boldsymbol{\theta}$ are available, the gradients and Hessian can be obtained, either analytically or numerically, and then used to evaluate the expression in (26).

4.2 Simplified Computation for the Normal Mixture

Some computational simplifications arise in the case of the Normal mixture because the conditioning statistic does not depend on unknown parameters. Hence, replacing $(y_{it}, \mathbf{x}_{it})$ by $(\tilde{y}_{it}, \tilde{\mathbf{x}}_{it})$, where $\tilde{}$ denotes the within transformation, and then maximizing the mixture likelihood is numerically exactly equivalent to applying the EM algorithm to the full-data likelihood. Estimation of a mixture of Normals with fixed effects can proceed therefore in essentially the same way as for the standard FM model for cross-section data provided that one first applies the within transformation to the data. Moreover, fixed effect models with more than two components can be handled using the same software as models without fixed effects.

A finite mixture likelihood based on first-differenced observations also does not require the EM algorithm. However, these results will not be numerically identical to those from the EM algorithm. First-differencing decreases the number of available time series observations from T to T–1, and it induces residual serial correlation, both of which reduce the efficiency of the standard estimator.

This analysis for individual-specific effects extends to panel data with cluster- or group-specific effects if clusters and groups are directly observable and not latent. Suppose individuals $i=1,\ldots,N$ are uniquely assigned to G groups, and the data are denoted as $(y_{it}^{(g)}, \mathbf{x}_{it}^{(g)})$, $g=1,\ldots,G$, then $\left(\tilde{y}_{it}^{(g)}, \tilde{x}_{it}^{(g)}\right)$ denotes the data after a within-group transformation; specifically, $\tilde{y}_{it}^{(g)} = y_{it} - \bar{y}^{(g)}$ denotes deviation of y_{it} from group-specific mean $\bar{y}^{(g)}$. In some instances panel data may have a third dimension in addition to (i, t); for example, in trade data that dimension could be country pair. The approach of this paper could be extended to this case also.

5 A Monte Carlo Study

We next report a Monte Carlo study of Normal and Poisson mixtures with fixed effects whose simulation design is motivated by the following objectives. Our first objective is to examine the finite sample properties of the proposed estimators and the robustness and computational efficiency of the proposed algorithms. Our second objective is to compare the proposed models with other selected models that some practitioners would regard as suitable alternatives to the FE-FMM. These include mixture models that incorporate a "Mundlak correction" factor, mixture models without fixed effects, and a fixed effect model without mixture specification (Mundlak 1978; Chamberlain 1984). These alternative specifications do not necessarily identify the same parameters as the FE-FMM, but they are potentially interesting.

5.1 Normal Mixtures

5.1.1 The Simulation Design

In the case of a two-component Normal mixture, the data generating process is as follows:

$$y_{it} \sim \begin{cases} \mathrm{N}(\mu_{1it}, \sigma_1^2) & \text{with probability } \pi \\ \mathrm{N}(\mu_{2it}, \sigma_2^2) & \text{with probability} (1-\pi) \end{cases}$$

where μ and σ denote the mean and standard deviation of the Normal distribution respectively and

$$\mu_{jit} = \beta_{1j} x_{it} + \alpha_{ij}.$$

The simulations are based on the following design configurations, T=(4, 8), N=2500, $\sigma_1=\sigma_2=1$, $\beta_{11}=1$, $\beta_{12}=2$, and π =0.5. The covariate x_{it} is drawn from a uniform distribution scaled and translated to have a mean of 1 and a standard deviation of 1.

In one set of experiments the individual-specific effects are assumed to be uncorrelated with x_{it}, i.e., α_{i1} and α_{i2} are drawn from i.i.d. N(0, 1). This experiment simulates the behavior of the FE estimator when the d.g.p. is a RE model. In general the FE estimator is consistent in this case and the simulation experiments checks this property.

In the second set of experiments, the individual-specific effects are correlated with x_{it}. More precisely, they are correlated with $\bar{x}_{i.}$, the within-group mean of x_{it}. Correlation is generated by specifying $\alpha_{ij} = u_{ij} + \sqrt{T}\,\bar{x}_{i.}$ where u_{ij} are drawn from i.i.d. N(0, 1). Note that $\mathrm{Var}(\sqrt{T}\,\bar{x}_{i.}) = 1$. Finally, α_{ij} are rescaled to have unit standard deviation.

It was suggested that the results may be sensitive to the choice of π in the Monte Carlo design. Specifically, it could be suggested that small values of π may be computationally problematic. Heuristically, $N\pi$ and $N(1-\pi)$ are the effective numbers of observations for the two components when C=2; therefore, one might expect that with a small value of π a component may be harder to identify (differentiate). This difficulty is mitigated as N increases. We considered the case of π =0.80, 1–π =0.2, C=2, and N=2500, both for the Poisson and Normal cases. The pattern of the results generated under this design were very similar to those for cases reported above.

We also considered simulations for samples with N=2500 but with the baseline parameter configurations. The results were as expected: precision of parameter estimates improved.

The Monte Carlo design we have used implies a direct connection of our FE model with the Mundlak-Chamberlain type conditionally correlated random effects (CCRE) model, which posits a relationship between α_i and observable variables $\boldsymbol{z}_i$ such as

$$\alpha_i = \kappa_1 (\mathbf{z}_i' \gamma + \eta_i),$$

where η_i is an i.i.d. random effect and $\mathrm{E}[\varepsilon_{it}\eta_i]=0$. The constant κ_1 is chosen so that α_i has unit standard deviation. Thus conditional on including $\boldsymbol{z}_i$ as regressors, the FE panel may be treated as a RE panel. Adding $\boldsymbol{z}_i$ as additional regressors then implements the "Mundlak correction." Given our d.g.p., the $\bar{\boldsymbol{x}}_{i.}$ corresponds to $\boldsymbol{z}_i$; therefore, adding $\bar{\boldsymbol{x}}_{i.}$ as an additional regressor and then maximizing the mixture likelihood (without the EM algorithm) should yield the same results as the EM algorithm.

In empirical analysis, there is no a priori reason to believe that the individual effects will be of the Mundlak-Chamberlain type. Therefore, to evaluate the performance of alternative specifications in the case of individual effects which are not of the Mundlak-Chamberlain type, we specify

$$\alpha_i = \kappa_1 \exp(\mathbf{z}_i' \gamma + \kappa_2 \eta_i) + \eta_i,$$

where η_i is an i.i.d. random effect and $\mathrm{E}[\varepsilon_{it}\eta_i]=0$. The constants κ_1 and κ_2 are chosen so that α_i has unit standard deviation.

5.1.2 Estimator Comparison

We estimate and compare the following estimators:

1. N-FE1: a standard fixed-effects linear regression which ignores the mixture aspect of the d.g.p.;
2. N-FM2: a standard two-component finite mixture of Normal densities which ignores the mechanism generating individual specific effects;
3. N-FM2-M: a standard two-component finite mixture of Normal densities with Mundlak correction;
4. N-FM2-FE: a two-component fixed-effects finite mixture of Normal densities which accommodates the individual specific effects, using the EM algorithm.

The first two models are misspecified relative to the correctly specified third and fourth models. Each model identifies parameters that may differ from the target parameter. For example, for two component distributions $N(\mu_1,\sigma_1^2)$ and $N(\mu_2,\sigma_2^2)$, and the mixture $\pi N(\mu_1,\sigma_1^2)+(1-\pi)N(\mu_2,\sigma_2^2)$, the closest 1-component distribution (in the Kullback-Leibler metric) that minimizes the expected distance measure is $N(\mu,\sigma^2)$ where

$$\mu=\pi\mu_1+(1-\pi)\mu_2$$
$$\sigma^2=\pi\sigma_1^2+(1-\pi)\sigma_2^2+\pi(1-\pi)(\mu_1-\mu_2)^2.$$

This particular $N(\mu,\sigma^2)$ is regarded as a projection of the two-component model on the one-component model. Hence, N-FE1 identifies (μ,σ^2). That is, the misspecified one-component model identifies weighted functions of the underlying parameters. When the two components are well separated, so that $\mu_1 >> \mu_2$, variance estimator $\hat{\sigma}^2$ is not a weighted sum of the two variances but instead is contaminated by a multiple of the "cross-parameter" factor $(\mu_1-\mu_2)^2$ which will also inflate the estimated variance of $\hat{\mu}$. In the special case of a scale mixture, where $\mu_1=\mu_2$, but $\sigma_1^2\neq\sigma_2^2$, the one-component model provides a consistent estimate of the mean and of the weighted variance parameter $\pi\sigma_1^2+(1-\pi)\sigma_2^2$.

5.1.3 Results

The Monte Carlo results for Normal mixtures, shown in Tables 1–3, are based on 1000 replications.

Table 1 reports results for the (orthogonal) RE d.g.p., for T=4 and 8. Under the RE d.g.p., and given that the regression has a single time-varying regressor, N-FE1 yields a consistent estimate of the linear combination $\pi\beta_1+(1-\pi)\beta_2=1.5$. The results for both T=4 and T=8 are in line with the theoretical expectation. Variability in the T=8 case is substantially smaller than that in the under the RE formulation, each component of the two-component mixture has extra variation added to the error, a feature neglected by the N-FM2 model. Yet the estimator of the slope parameter retains its consistency property. However, the residual variance now identifies the sum of variances of the idiosyncratic component and the random effect. The N-FM2-M specification incorporates a Mundlak correction where none is needed, and this also contributes to greater variability of the estimated slopes. The N-FM2-FE estimator shows smaller variability. In the T=4 case, there appear to be small biases in the estimates of β_1 and β_2 but these disappear when T=8.

Table 2 gives results for the first variant of the correlated FE specification. For the FE d.g.p. specified using the Mundlak-Chamberlain formulation, N-FE1 is a consistent estimator of the linear combination $\pi\beta_1+(1-\pi)\beta_2$. The variance of the distribution falls by a small amount as T goes from 4 to 8. N-FM2, which ignores correlation between the fixed effect and regressors, is not a consistent estimator of β_1 and β_2. The Monte Carlo results indicate a larger bias for both coefficients of around 0.3 for the T=4 case and a smaller bias of around 0.25 for the T=8 case. Perhaps surprisingly, the mixing fraction π shows almost no bias. Note that we currently do not have any theoretical results to support this simulation result. N-FM2-FE should be consistent and the results once again satisfy this expectation. Because the d.g.p. used to generate correlation is exactly of the type that fully validates the "Mundlak correction," the N-FM2-M

Table 1 Two-Component Mixture of Normals with Orthogonal Random Effects.

Parameter	True value	T=4		T=8	
		Mean	Std. Dev.	Mean	Std. Dev.
N-FE: Fixed Effects Linear Regression					
β	0.0	1.500	0.016	1.501	0.014
N-FM2: 2-Component Mixture					
β_1	1.0	1.000	0.078	0.999	0.055
β_2	2.0	1.998	0.076	2.003	0.054
π_1	0.5	0.499	0.066	0.500	0.047
N-FM2-M: Mixture with Mundlak Correction					
β_1	1.0	0.999	0.079	1.000	0.055
β_2	2.0	1.997	0.078	2.003	0.055
π_1	0.5	0.498	0.066	0.500	0.047
N-FM2-FE: Mixture with Fixed Effects					
β_1	1.0	0.945	0.035	0.992	0.015
β_2	2.0	2.053	0.035	2.010	0.015
π_1	0.5	0.499	0.031	0.500	0.015

Note: N=2500; number of replications=1000.

Table 2 Two-Component Mixture of Normals with Mundlak Type Correlated Random Effects.

Parameter	True value	T=4		T=8	
		Mean	Std. Dev.	Mean	Std. Dev.
N-FE: Fixed Effects Linear Regression					
β	0.0	1.500	0.015	1.500	0.014
N-FM2: 2-Component Mixture					
β_1	1.0	1.348	0.068	1.245	0.053
β_2	2.0	2.346	0.069	2.262	0.051
π_1	0.5	0.499	0.058	0.498	0.045
N-FM2-M: Mixture with Mundlak Correction					
β_1	1.0	1.000	0.055	0.999	0.039
β_2	2.0	2.001	0.056	1.999	0.037
π_1	0.5	0.500	0.046	0.499	0.033
N-FM2-FE: Mixture with Fixed Effects					
β_1	1.0	0.947	0.057	0.991	0.016
β_2	2.0	2.056	0.054	2.009	0.016
π_1	0.5	0.502	0.051	0.500	0.017

Note: N=2500; number of replications=1000.

estimator is equivalent to N-FM2-FE based on maximizing the full-data likelihood. The standard deviations of the parameter estimates from the N-FM2-FE estimator, however, are much smaller than those of N-FM2-M when T=8.

In Table 3, we report the results of experiments in which the fixed effects are generated using a non-Mundlak-Chamberlain formulation, N-FE1 is still a consistent estimator of $\pi\beta_1 + (1-\pi)\beta_2$. Once again, N-FM2 is not a consistent estimator of β_1 and β_2. Unlike the results in Table 2, in this case, N-FM2 is not a consistent estimator for π either. The Monte Carlo results indicate a large bias for both slope coefficients and the mixing probability regardless of whether T=4 or T=8. As the simulation scheme used for generating dependence between regressor and fixed effect is not exactly consistent with the "Mundlak correction," N-FM2-M is also not a consistent estimator. The estimates of β_1 are upward biased by about 0.2 while estimates for β_2 are downward biased by about 0.12. The estimates for π are also substantially upward biased. There is no evidence of the bias decreasing when T goes from 4 to 8. Only N-FM2-FE is a consistent estimator for this d.g.p. Although there is evidence of small bias of 2–4% in the T=4 case, the bias disappears when T=8.

5.2 Poisson Mixtures

5.2.1 The d.g.p.

In the case of Poisson mixtures, we specify the d.g.p. as follows:

$$y_{it} \sim \begin{cases} \mathrm{P}(\mu_{1it}) & \text{with probability } \pi \\ \mathrm{P}(\mu_{2it}) & \text{with probability } (1-\pi) \end{cases}$$

$$\begin{aligned} \mu_{jit} &= \alpha_i \exp(\beta_{1j} x_{it}) \\ &\equiv \exp(\beta_{1j} x_{it} + \tau_i). \end{aligned}$$

where $\alpha_i \equiv \exp(\tau_i)$. The experiments are based on the following parameter configurations, T=(4,8), N=2500, β_1=0.2, β_2=1.0, and x_{it} is drawn from a uniform distribution scaled and translated to have a mean of 1 and a standard deviation of 1.

In one set of experiments, reported in Table 4, the individual effects are random, i.e., τ_i are drawn from i.i.d. N(0,0.4). This component generates component-wise overdispersed counts. Ignoring overdispersion in the standard one-component model does not affect the consistency property of the Poisson MLE. However, this property does not extend to a mixture of overdispersed Poissons. Hence, even in the case of random effects, the MLE of Poisson FM2 model will not be consistent.

In a second set of experiments, reported in Table 5, the individual-specific effects are correlated. More precisely, they are correlated with $\bar{x}_{i.}$ which are the within-group means of x_{it} with the correlation generated by specifying $\tau_i = u_i + \sqrt{T}\bar{x}_{i.}$ where u_i are drawn from i.i.d. N(0,1). Note that $\mathrm{Var}(\sqrt{T}\bar{x}_{i.}) = 1$. Finally, τ_i are rescaled to have unit standard deviation. Like random effects, fixed effects also generate component-wise overdispersion leading to inconsistency of Poisson FM2. Given the nonlinear conditional mean, the "Mundlak correction" in this context is ad hoc. In fact, neither the existence nor the form of the

Table 3 Two-Component Mixture of Normals with non-Mundlak Type Correlated Random Effects.

Parameter	True value	T=4		T=8	
		Mean	Std. Dev.	Mean	Std. Dev.
N-FE: Fixed Effects Linear Regression					
β	0.0	1.500	0.016	1.501	0.014
N-FM2: 2-Component Mixture					
β_1	1.0	1.309	0.090	1.359	0.117
β_2	2.0	2.286	0.159	2.270	0.171
π_1	0.5	0.569	0.094	0.645	0.133
N-FM2-M: Mixture with Mundlak Correction					
β_1	1.0	1.188	0.067	1.219	0.073
β_2	2.0	1.884	0.047	1.881	0.039
π_1	0.5	0.563	0.078	0.594	0.089
N-FM2-FE: Mixture with Fixed Effects					
β_1	1.0	0.979	0.068	0.994	0.029
β_2	2.0	2.084	0.064	2.012	0.026
π_1	0.5	0.531	0.063	0.503	0.028

Note: N=2500; number of replications =1000.

Table 4 Two-Component Mixture of Poissons with Orthogonal Random Effects.

Parameter	True value	T=4		T=8	
		Mean	Std. Dev.	Mean	Std. Dev.
P-FE: Fixed Effects Poisson Regression					
β		0.770	0.011	0.775	0.009
P-FM2: 2-Component Mixture					
β_1	0.2	0.402	0.024	0.401	0.018
β_2	1.0	0.904	0.017	0.903	0.012
π_1	0.5	0.601	0.013	0.600	0.012
P-FM2-M: Mixture with Mundlak Correction					
β_1	0.2	0.402	0.025	0.401	0.018
β_2	1.0	0.904	0.018	0.903	0.012
π_1	0.5	0.601	0.013	0.600	0.012
P-FM2-FE: Mixture with Fixed Effects					
β_1	0.2	0.236	0.022	0.241	0.013
β_2	1.0	0.983	0.013	0.987	0.008
π_1	0.5	0.514	0.015	0.522	0.012

Note: N=2500; number of replications =1000.

Table 5 Two-Component Mixture of Poissons with Correlated Random Effects.

Parameter	True value	T=4		T=8	
		Mean	Std. Dev.	Mean	Std. Dev.
P-FE: Fixed Effects Poisson Regression					
β		0.775	0.010	0.778	0.011
P-FM2: 2-Component Mixture					
β_1	0.2	0.466	0.022	0.445	0.020
β_2	1.0	0.990	0.015	0.956	0.014
π_1	0.5	0.585	0.012	0.586	0.015
P-FM2-M: Mixture with Mundlak Correction					
β_1	0.2	0.349	0.022	0.348	0.020
β_2	1.0	0.919	0.015	0.915	0.014
π_1	0.5	0.571	0.012	0.570	0.015
P-FM2-FE: Mixture with Fixed Effects					
β_1	0.2	0.192	0.034	0.277	0.019
β_2	1.0	0.960	0.015	0.994	0.009
π_1	0.5	0.459	0.038	0.544	0.021

Note: N=2500; number of replications=1000.

appropriate correction is known to us. So including the same additional regressor as in the Normal case is speculative, and, in theory, it does not remove the bias due to the correlated random effect. Hence, for this d.g.p., only the Poisson FM2-FE estimator is consistent.

5.2.2 Estimator Comparison

Given these data generating processes, we estimate and compare

– P-FE1: a standard fixed-effects one-component Poisson regression without fixed effect adjustment;
– P-FM2: a standard two-component mixture of Poisson ignoring individual specific effects;
– P-FM2-M: a standard two-component mixture of Poisson with "Mundlak-type" correction;
– P-FM2-FE: a two-component fixed-effects finite mixture of Poisson individual specific effects, using the EM algorithm.

As previously noted, the first three are misspecified models and only the fourth is correctly specified.

In the case of P-FE1 and P-FM2-FE, if every outcome y_{it} for a particular group i is zero, then that group of observations must be dropped from the sample prior to estimation. This occurs with a reasonably substantial frequency in the case where T=4 but is less probable when T=8. Note that asymptotically, one does not expect this situation to occur.

The simulation results reported in Tables 4 and 5 are based on 1000 replications.

5.2.3 Results

The results in Table 4 are for the experiments with orthogonal random effects. Although we have reported results for P-FE1, unlike the case of the normal mixtures, one does not expect the estimated coefficient to be a simple mixture-probability weighted combination of the component parameters. Thus is it not possible to comment on the quality of that estimator, except to note that the variance of the estimator is smaller in the T=8 case as compared to the T=4 case. As expected, the simulation results confirm that P-FM2 and P-FM2-M are inconsistent estimators. The bias of each of the three model parameters is large and does not decrease as the panel dimension T increases. P-FM2-FE is theoretically consistent under the RE d.g.p. but the simulation results show that there are small biases in the estimates of β_2 and π and somewhat larger biases in the estimates of β_1. These biases are considerably smaller, however, than those obtained with P-FM2 or P-FM2-M.

The results of the specifications in Table 5, in which the individual effects are correlated with the regressors, are qualitatively identical to those shown in Table 4. Once again, P-FM2 and P-FM2-M have large biases. P-FM2-FE performs substantially better, but there remains evidence of finite sample biases for all three model parameters.

6 Empirical Applications

The empirical question of interest is the effect of turning 65 and becoming Medicare eligible on medical utilization among those who have been uninsured prior to Medicare eligibility, relative to the previously-insured group. Specifically, we model the logarithm of total medical expenditures using the Normal family of models and the number of doctor visits using the Poisson family of models.

The motivation for the investigation is as follows. Medicare eligibility at age 65 results in a large and abrupt decline in the probability of being uninsured in the US. The large decrease in the chance of being uninsured at age 65 should help to reduce disparities in the use of health services after age 65 compared to before. We investigate the effect of universal health insurance coverage on health outcomes and the use of health services by exploiting the natural experiment that exogenously changes the insurance status of most Americans at age 65 – that is, eligibility for the Medicare program.

Since almost all individuals turning 65 become automatically eligible for Medicare, a difference-in-difference or regression discontinuity design could provide estimates of the effect of insurance on the use of health services free of confounding, since the choice of insurance is not endogenous and hence selection effects are absent. However, there remains the issue of unobserved individual-specific heterogeneity. In difference-in-difference studies, the data are often cross-sectional, and hence it is not possible to identify individuals after age 65 who were uninsured before age 65; that is, good controls are harder to find. Instead, a synthetic control group is used consisting of similar individuals before age 65 (see Card, Dobkin, and Maestas 2008). Because the HRS is a panel of individuals, many of whom turn 65 during the surveys, panel data automatically generates good controls. Moreover, panel data allow us to control for individual-specific unobserved heterogeneity much better than is possible with cross-sectional data.

Though a FM2-FE is the main model of interest, to facilitate comparisons we also estimated models under RE assumptions using pooled cross-sectional methods, including the standard FM2 model.

Table 6 Summary Statistics of HRS Data.

Variable	Ever uninsured=0		Ever uninsured=1	
	Mean	SD	Mean	SD
Total expenditures (given >0)	7342	24,949	7420	29,380
Logarithm of total expenditures	7.630	1.534	7.461	1.603
Number of visits to the doctor	6.145	8.025	5.479	8.831
Age	59.70	4.504	60.34	4.342
Medicare eligible (M)	0.111	0.315	0.121	0.326
Married	0.799	0.400	0.684	0.465
Income quartile 1	0.109	0.312	0.364	0.481
Income quartile 2	0.243	0.429	0.298	0.457
Income quartile 3	0.305	0.460	0.188	0.391
Female	0.525	0.499	0.566	0.496
Black	0.110	0.313	0.169	0.375
Hispanic	0.0452	0.208	0.188	0.391
Other race	0.0153	0.123	0.0279	0.165
High school dropout	0.145	0.352	0.421	0.494
High school degree	0.404	0.491	0.331	0.471
College attendee	0.219	0.414	0.160	0.366
Lives in midwest	0.273	0.445	0.177	0.381
Lives in south	0.376	0.484	0.510	0.500
Lives in west	0.163	0.369	0.182	0.386

Note: N=25,282 for the sample of ever uninsured=0 and N=5011 for the sample of ever uninsured=1. Sample sizes for positive expenditures (and its logarithm) are 24,324 and 4403 for the two samples, respectively.

6.1 Data

We apply the models described above to an analysis of medical expenditures and the number of doctor visits using panel data from six waves of the Health and Retirement Survey (HRS): 1992, 1994, 1996, 1998, 2000 and 2002. The HRS is a population-based sample of US community dwellers aged 51–61 in 1992 (and their spouses) with follow-up interviews every 2 years. During the past few years, the HRS has become one of the most widely used data sets for analyzing health and the use of health services in the US e.g, Smith and Kington 1997; Smith 1998; Johnson and Crystal 2000).

We analyze data for age eligible respondents and their spouses, if the spouse is within the 51–61 year age range in 1992 and if they are in the data in at least four waves. We eliminate any individual who is on Medicaid or Medicare at any time before turning 65. Our final sample size across six waves totals 30,293 person-wave observations, with N=5860 and $6 \geq T \geq 4$. There are no missing observations for 33.19% of the sample, and an additional 49.76% of the sample has at most one missing observation. It is assumed that observations are missing at random and there is no attrition bias.

We next consider some a priori arguments that support the finite mixture fixed effects specification.

Approximately 17% of our sample has been uninsured before the age of 65. Table 6 contains descriptive statistics for dependent variables measuring the use of health services and health status both for those who have been uninsured before the age of 65 and for those who were consistently insured. Those who have been uninsured have significantly lower log expenditures (about 17%) and number of doctor visits (12%). These differences are raw differences that do not control for observed and unobserved heterogeneity. The kernel density plot of log-expenditures by insurance status in Figure 1 shows that the distributions have more than one mode. Such a feature is consistent with a finite mixture type unobserved heterogeneity in the two groups. Figure 1 also shows the density plots of OLS fitted values and residuals after a pooled regression of log(expenditures) on the full set of regressors as in column 1 of Table 7. These plots are only intended to be suggestive and descriptive and not formal tests of multimodality, but they hint at the possibility of mixture components that are not well separated.

The summary statistics show that the uninsured are less likely to be married, and more likely to have lower education and income. Older Americans who have experienced uninsurance are, statistically, more likely to be black, Hispanic and female. With so many differences in observed characteristics, one may reasonably conclude

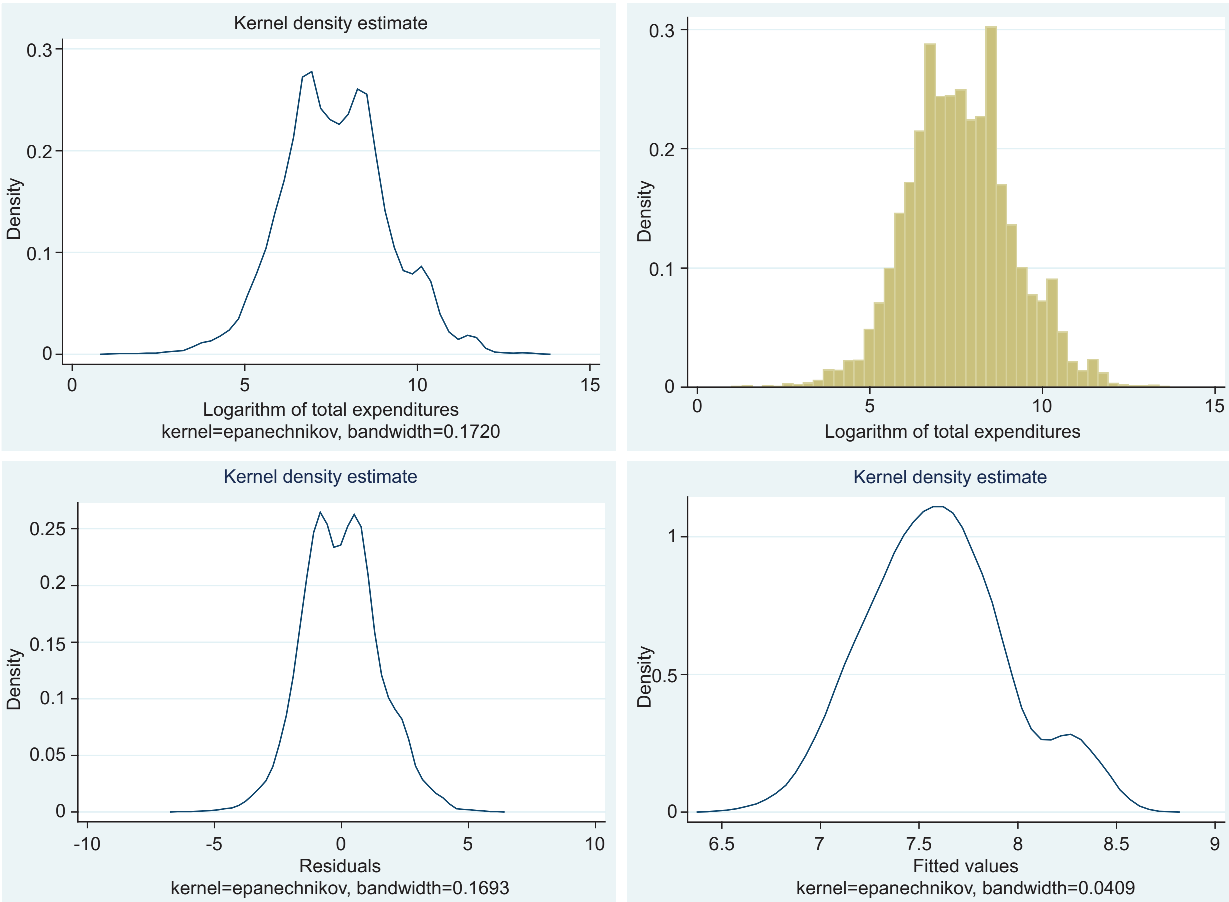

Figure 1 K-Density Plots of Log(Expenditures), their Fitted Values and Residuals.

that there are substantive differences in unobserved characteristics as well. There are two facts that support this view. First the HRS sample is deficient in measures of health status, which potentially could be worse on average for those previously without insurance. Second, before-and-after Medicare comparison, using those without and those with Medicare, to estimate the treatment effect of Medicare is potentially flawed because those with Medicare may also purchase supplementary insurance such as Medigap or an employer-sponsored supplementary plan to cover expenses that are not covered by Medicare. That is, those with Medicare are likely to be heterogeneous in terms of their insurance coverage. Medicare beneficiaries data from the Medicare Beneficiaries Surveys of 2003–05 show that more than 60% have some type of supplementary coverage. Our sample does not control for heterogeneity of coverage under Medicare, and this has motivated a model with individual-specific effects. The purchase of supplementary insurance is likely to be correlated with regressors like income, education, and ethnicity – a consideration that justifies the fixed effects specification.

6.2 Results

In the empirical analysis, we estimate two models that do not explicitly handle individual-specific effects and four that incorporate individual-specific effects in different ways. Four of the six models allow for discrete finite-mixture type heterogeneity while and two do not. The key variable of interest in the analysis is the interaction variable (U × M) whose coefficient measures the impact of having Medicare on those with no insurance coverage prior to becoming Medicare eligible. Inference is based on the cluster-robust formulation of the covariance matrix given in equation (26) without further finite sample adjustments.

6.2.1 Expenditure Models

Six estimated models of log expenditure are shown in Tables 7 and 8. These include the linear standard linear FE model, the standard FM model, and the FE-FM model. The tables

Table 7 Models of (log) Expenditures.

Variable	OLS	FM Normal		FE Regression
		Component1	Component2	
Age	−0.102	−0.248*	0.256	−0.067
	(0.068)	(0.122)	(0.215)	(0.069)
$Age^2/100$	0.152**	0.286**	−0.176	0.166**
	(0.058)	(0.103)	(0.182)	(0.059)
Medicare eligible (M)	−0.158*	−0.277**	0.133	−0.175**
	(0.063)	(0.106)	(0.174)	(0.063)
Been uninsured (U)	−0.168**	−0.138*	−0.231*	−0.185**
	(0.034)	(0.060)	(0.094)	(0.071)
$U \times M$	0.329**	0.452**	0.066	0.475**
	(0.082)	(0.124)	(0.159)	(0.090)
Married	−0.059*	−0.072	−0.023	0.078
	(0.029)	(0.048)	(0.080)	(0.053)
Income quartile 1	−0.252**	−0.143	−0.517**	−0.102*
	(0.037)	(0.074)	(0.117)	(0.043)
Income quartile 2	−0.135**	−0.034	−0.368**	−0.059
	(0.029)	(0.055)	(0.081)	(0.035)
Income quartile 3	−0.058*	0.027	−0.253**	−0.003
	(0.025)	(0.048)	(0.067)	(0.029)
Female	0.023	−0.121*	0.367**	
	(0.023)	(0.049)	(0.065)	
Black	−0.047	−0.152*	0.195	
	(0.035)	(0.066)	(0.111)	
Hispanic	−0.153**	−0.122	−0.210	
	(0.050)	(0.091)	(0.142)	
Other race	−0.115	−0.063	−0.285	
	(0.098)	(0.171)	(0.319)	
High school dropout	−0.076	0.035	−0.336**	
	(0.040)	(0.070)	(0.128)	
High school degree	−0.109**	−0.007	−0.348**	
	(0.031)	(0.054)	(0.085)	
College attendee	−0.035	0.096	−0.353**	
	(0.035)	(0.060)	(0.097)	
Lives in midwest	0.019	0.009	0.038	
	(0.035)	(0.058)	(0.091)	
Lives in south	−0.011	0.007	−0.052	
	(0.033)	(0.055)	(0.082)	
Lives in west	−0.020	−0.036	0.024	
	(0.040)	(0.068)	(0.103)	
π_1		0.692		
		(0.052)		
lnL	−52435.733	−52249.655		−46405.764
Observations	28,727	28,727		28,727

Cluster-robust standard errors in parentheses.
$^{**}p < 0.01$, $^{*}p < 0.05$.

provide coefficient estimates and log-likelihood values. The latter value is generated using maximum likelihood estimates and the expression for the marginal log-likelihood - (14) in the Normal case and (18) in the Poisson case. Note, however, that the fixed effect models only include time-varying variables, and the estimated log-likelihood value for these models is log-concentrated likelihood which is not directly comparable with specifications that permit time-invariant variables. One may, however, compare the fit of the standard FE and FE-FM2 models using log-concentrated likelihood because they both exclude time-invariant regressors. Specifically, that value is −46323.793 for FE-FM2 and −46405.764 for the FE model. The standard likelihood ratio test statistic is 163.94 (=−2((46405.764+46323.793), which is asymptotically χ^2 distributed with 9 degrees of freedom – a result that confirms statistically significant improvement in the fit of the model.

A detailed analysis of the FM2 specification indicated two components. Fitted values were generated for each component; the larger of which shows smaller (mean, standard deviation) – (7.7, 0.95) compared with (9.59,11.0) for the second, confirming that the components are not well separated when we do not allow for individual-specific effects. The estimated mixing proportion for the FM2-FE model is around 0.64, but it is higher for three other specifications based on mixtures. The estimated coefficient on the interaction term in the OLS estimation is positive and significant (0.329), indicating 32.8% higher expenditure. All the mixture specifications indicate that there is a substantial difference in this coefficient between the two components, and the relative magnitude is between two to four times higher for the larger component. Thus, Medicare eligibility for the previously-uninsured group substantially increases expenditures compared to those who have been insured. The size effects are considerably different between the two classes of individuals. For the majority of individuals ($\hat{\pi}$=69%), the effect is large and significant (0.452) while for a substantial (31%) minority, the result is small and insignificant. The estimates in the first two columns of Table 8 show that the results do not change much when all the time-invariant covariates are dropped from the model. The introduction of individual fixed effects does, however, make a substantial difference. First, a comparison of OLS with a linear fixed effects model in Table 8 shows that the coefficient on the interaction term increases from 0.329 to 0.475. The results in Table 8 also show that the effects for each of the two components in the FM model with FE are larger than those in the FM model without FE. Indeed, now the coefficient for the smaller group is statistically significant and almost three times larger. This result may seem counter-intuitive because in many cases allowing for fixed (correlated) effects would lower this coefficient, not increase it. However, the reader is reminded that the higher estimate could also reflect the effect of supplementary private insurance.

6.2.2 Count Models

As in the case of the standard (one-component) fixed effect Poisson model, the marginal effect of variation in covariate

Table 8 Models of (log) Expenditures.

Variable	FM Normal		Mundlak FM Normal		Fixed Efffects FM Normal	
	Component1	Component2	Component1	Component2	Component1	Component2
Age	−0.210	0.190	−0.200	0.308	0.087	−0.413**
	(0.120)	(0.214)	(0.121)	(0.232)	(0.083)	(0.059)
$Age^2/100$	0.255*	−0.124	0.299**	−0.205	0.080	0.369**
	(0.102)	(0.182)	(0.103)	(0.197)	(0.071)	(0.050)
Medicare eligible (M)	−0.248*	0.065	−0.315**	0.198	−0.117	−0.312**
	(0.106)	(0.178)	(0.106)	(0.195)	(0.077)	(0.055)
Been uninsured (U)	−0.133*	−0.349**	−0.247	0.008	−0.083	−0.450**
	(0.057)	(0.088)	(0.128)	(0.246)	(0.079)	(0.063)
U×M	0.443**	0.105	0.571**	0.262	0.550**	0.251**
	(0.121)	(0.160)	(0.136)	(0.206)	(0.098)	(0.077)
Married	−0.056	−0.127	0.088	0.038	0.088	0.049
	(0.050)	(0.081)	(0.087)	(0.146)	(0.063)	(0.046)
Income quartile 1	−0.216**	−0.520**	−0.082	−0.182	0.022	−0.361**
	(0.074)	(0.116)	(0.077)	(0.146)	(0.052)	(0.037)
Income quartile 2	−0.081	−0.402**	−0.072	−0.030	0.040	−0.254**
	(0.056)	(0.083)	(0.062)	(0.114)	(0.042)	(0.030)
Income quartile 3	0.006	−0.298**	0.025	−0.090	0.054	−0.130**
	(0.049)	(0.070)	(0.054)	(0.095)	(0.035)	(0.025)
π_1	0.703		0.718		0.670	
	(0.054)		(0.050)		(0.004)	
lnL	−52307.121		−46300.367		−46323.793	
Observations	28,727		28,727		28,727	

Robust standard errors in parentheses.
$^{**}p<0.01$, $^{*}p<0.05$.

is not identified. This is because the component-wise conditional mean is also not identified; as $\partial E[y]/\partial x_k = \beta_k \exp(\boldsymbol{x}'\boldsymbol{\beta})$, only the effect of a time-varying regressor x_k is identified as time-invariant regressors are swept out by the fixed effects transformation. The semielasticity is consistently estimated by the coefficient β, and the ratio of marginal effects of x_k and x_j, defined as $(\partial E[y]/\partial x_k)/\ \partial E[y]/\partial x_j) = \beta_k/\beta_j$, is also identified when the index function $\boldsymbol{x}'\boldsymbol{\beta}$ is linear. Analogously, in the FE-FM case ratios of within-component marginal effects may be identified. The relative marginal effects can also be compared across components. If the index function is not linear in all components of $\boldsymbol{x}$, the ratio of marginal effects of some regressors may also depend upon regressors and estimated coefficients.

Results for the count models of doctor visits are shown in Tables 9 and 10. Detailed analysis of the FM-2

Table 9 Models of Number of Visits to the Doctor.

Variable	Poisson	FM Poisson		FE Poisson
		Component1	Component2	
Age	0.303**	0.297**	0.125	0.342**
	(0.058)	(0.056)	(0.142)	(0.055)
$Age^2/100$	−0.216**	−0.199**	−0.080	−0.221**
	(0.049)	(0.048)	(0.121)	(0.046)
Medicare eligible (M)	0.066	−0.037	0.012	0.080
	(0.048)	(0.053)	(0.128)	(0.047)
Been uninsured (U)	−0.157**	−0.247**	−0.046	−0.098
	(0.035)	(0.055)	(0.092)	(0.055)
U × M	0.232**	0.203**	0.123	0.166*
	(0.067)	(0.078)	(0.142)	(0.081)
Married	−0.051	−0.026	−0.092	−0.037
	(0.029)	(0.034)	(0.064)	(0.043)
Income quartile 1	−0.089**	−0.216**	−0.031	0.025
	(0.034)	(0.044)	(0.076)	(0.034)
Income quartile 2	−0.028	−0.075*	0.098	0.018
	(0.027)	(0.032)	(0.067)	(0.027)
Income quartile 3	0.010	0.029	0.148*	0.034
	(0.023)	(0.028)	(0.068)	(0.022)
Female	0.218**	0.246**	0.168**	
	(0.023)	(0.028)	(0.053)	
Black	0.074*	0.125**	−0.006	
	(0.035)	(0.038)	(0.070)	
Hispanic	0.024	0.037	0.045	
	(0.053)	(0.067)	(0.122)	
Other race	0.004	−0.117	−0.076	
	(0.098)	(0.118)	(0.156)	
High school dropout	−0.096*	−0.220**	−0.031	
	(0.041)	(0.048)	(0.090)	
High school degree	−0.094**	−0.164**	−0.128	
	(0.030)	(0.031)	(0.066)	
College attendee	−0.035	−0.093*	−0.010	
	(0.034)	(0.039)	(0.086)	
Lives in midwest	−0.097**	−0.082*	−0.109	
	(0.034)	(0.036)	(0.072)	
Lives in south	−0.129**	−0.135**	−0.112	
	(0.032)	(0.035)	(0.067)	
Lives in west	−0.066	−0.090	−0.016	
	(0.039)	(0.047)	(0.093)	
π_1		0.863		
		(0.006)		
lnL	−145697.614	−98508.989		−75477.206
Observations	30,293	30,293		29,996

Robust standard errors in parentheses.
**$p<0.01$, *$p<0.05$.

Table 10 Models of Number of Visits to the Doctor.

Variable	FM Poisson		FM-Linear Mundlak		FM-Fixed Effects	
	Comp.1	Comp.2	Comp.1	Comp.2	Comp.1	Comp.2
Age	0.308**	0.130	0.391**	0.234	0.294**	0.468**
	(0.054)	(0.136)	(0.054)	(0.136)	(0.046)	(0.124)
$Age^2/100$	−0.208**	−0.085	−0.253**	−0.147	−0.174**	−0.337**
	(0.046)	(0.116)	(0.047)	(0.116)	(0.039)	(0.106)
Medicare eligible (M)	−0.033	0.034	0.058	0.121	−0.002	0.247*
	(0.052)	(0.126)	(0.054)	(0.125)	(0.040)	(0.116)
Been uninsured (U)	−0.288**	−0.050	−0.093	0.025	−0.102*	−0.084
	(0.066)	(0.106)	(0.071)	(0.167)	(0.049)	(0.106)
U×M	0.215**	0.130	0.151	−0.216	0.261**	0.035
	(0.081)	(0.147)	(0.116)	(0.285)	(0.059)	(0.174)
Married	−0.119**	−0.134*	0.011	−0.056	−0.030	−0.051
	(0.034)	(0.067)	(0.051)	(0.125)	(0.037)	(0.091)
Income quartile 1	−0.237**	−0.018	−0.012	0.047	−0.050	0.147*
	(0.040)	(0.071)	(0.040)	(0.100)	(0.028)	(0.074)
Income quartile 2	−0.104**	0.088	0.012	0.086	−0.034	0.117
	(0.029)	(0.059)	(0.030)	(0.080)	(0.023)	(0.061)
Income quartile 3	0.003	0.112	0.049	0.147*	−0.013	0.124*
	(0.027)	(0.064)	(0.026)	(0.072)	(0.018)	(0.052)
π_1	0.864		0.863		0.845	
	(0.007)		(0.006)		(0.005)	
lnL	−99336.688		−98002.905		−74577.131	
Observations	30,293		30,293		29,996	

Robust standard errors in parentheses.
$^{**}p<0.01$, $^{*}p<0.05$.

model showed that the components are better separated. The fitted values from the two components have (mean, std. dev.) of (3.91, 1.58) and (22.5, 199), respectively. Note that the expenditure analysis pertains to those with positive expenditures, where as the count analysis includes individuals with some zero doctor visits. The estimates from a Poisson regression show that Medicare eligibility for those who have been uninsured increases the number of visits by 23.2% compared to those who have been insured. This estimate *decreases* to 17% once fixed effects are introduced. Two specifications of the finite mixture models with all covariates reported in Table 9, and that with only time varying covariates reported in Table 10, show significant effects across both components, with the larger effect observed for the component with the higher probability of occurrence ($\hat{\pi}$=0.862). Once fixed effects are introduced into the FM Poisson regression, the effect for the second component is substantially smaller and statistically insignificant. The coefficient for the first component is a little larger than the corresponding estimate in the model without FE.

7 Concluding Remarks

This paper has focussed on the fixed effects panel MLE for two-component finite mixtures of Normal and Poisson distributions. We extend the concentrated (conditional) likelihood approach of Andersen (1970) to handle fixed effects in a mixture model. Computation is simpler for the Normal mixture than for the Poisson mixture. Our Monte Carlo results confirm that under correct specification of the d.g.p. this approach works satisfactorily. We also show the connections with the correlated random effects model. The approach can be extended to finite mixtures with more than two components, as well as to other families of distributions that share some properties of the Normal and Poisson models.

We note that the issues about specification, estimation and analysis of finite mixture models remain valid topics for discussion and comment, but the scope of our paper does not extend to dealing with these issues, including: choice of the number of mixture components; identification of "small" components. A common approach in applied work is to choose the number and types of subpopulations on the basis of some sort of a priori information. For a more information criteria based approach, we refer the reader to Lindsay (1995), Lindsay and Roeder (1992) and McLachlan and Peel (2004) for further discussion. However, they are not intrinsic to this paper and the justification for the current approach does not depend upon how we handle these issues. That is, these are important modeling issues but quite separate from the main theme of our paper.

We have not considered empirically important complications like censoring and lagged or endogenous regressors. Those issues are left for future work.

Acknowledgment: We are grateful for comments and suggestions for improvement of an earlier version received from Anirban Basu, David Drukker, Maarten Lindeboom, and two anonymous reviewers and the editor. We also received helpful suggestions from participating audiences at 2011 SETA Conference and the 17th International Panel Data Conference. We alone are responsible for any errors.

Previously published online March 28, 2013

References

Aitkin, M., and D. B. Rubin. 1985. “Estimation and Hypothesis Testing in Finite Mixture Models.” *Journal of the Royal Statistical Society. Series B (Methodological)* 47(1): 67–75

Allison, P. 2009. *Fixed Effects Regression Models*. Los Angeles: Sage Publications.

Andersen, E. B. 1970. “Asymptotic Properties of Conditional Maximum Likelihood Estimators.” *Journal of the Royal Statistical Society* B 32: 283–301.

Arellano, M., and J. Hahn. 2007. “Understanding Bias in Nonlinear Panel Models: Some Recent Developments.” In *Advances in Economics and Econometrics: Theory and Applications, Volume III,* edited by R. Blundell, W. K. Newey, and T. Persson, 381–409. Cambridge: Cambridge University Press.

Bago d’Uva. T. 2005. “Latent Class Models for Use of Primary Care: Evidence From a British Panel.” *Health Economics* 14: 873–892.

Cameron, A.C., and P.K Trivedi. 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

Card, D., C. Dobkin, and N. Maestas. 2008. “The Impact of Nearly Universal Insurance Coverage on Health Care Utilization and Health: Evidence from Medicare.” *American Economic Review* 98(5): 2242–2258.

Chamberlain, G. 1984. “Panel Data.” In *Handbook of Econometrics*, *Volume II*, edited by Z. Griliches, and M. Intriligator, 1247–1318. Amsterdam: North-Holland.

Chamberlain, G. 1992. “Comment: Sequential Moment Restrictions in Panel Data.” *Journal of Business and Economic Statistics* 10: 20–26.

Conway, K., and P. Deb. 2005. “Is Prenatal Care Really Ineffective? Or, is the ’Devil’ in the Distribution?.” *Journal of Health Economics* 24: 489–513.

Deb, P., and P. K. Trivedi 1997. “Demand for Medical Care by the Elderly: a Finite Mixture Approach.” *Journal of Applied Econometrics* 12: 313–336.

Deb, P. and P. K. Trivedi. 2002. The Structure of Demand for Health Care: Latent Class versus Two-part Models.” *Journal of Health Economics* 21: 601–625.

Dempster A. P., N. M. Laird, and D. B. Rubin. 1977. “Maximum Likelihood from Incomplete Data via EM Algorithm.” *Journal of the Royal Statistical Society, B* 39(1): 1–38.

Dhaene, G., and K. Jochmans. 2011. Profile-score Adjustments for Nonlinear Fixed-effect Models. Paper presented at the 17th International Panel Data Conference, Montreal.

Frühwirth-Schnatter, S. 2006. *Finite Mixture and Markov Switching Models*. New York: Springer.

Greene, W. 2004a. “The Behavior of Fixed Effects Estimator in Nonlinear Models.” *Econometrics Journal* 7: 98–119.

Greene, W. 2004b. “Estimating Econometric Models with Fixed Effects.” *Econometric Reviews* 23: 125–147.

Heckman, J. J. 1981. “The Incidental Parameters Problem and the Problem of Initial Conditions in estimating a Discrete Time-discrete Data Stochastic Process.” In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C.F. Manski and D. McFadden 179–195. Cambridge, MA: MIT Press.

Heckman, J. J., and B. Singer. 1984. “A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data.” *Econometrica* 52: 271–320.

Johnson R. W., and S. Crystal. 2000. “Uninsured Status and Out-of-Pocket Costs at Midlife.” *Health Services Research* 35(Part 1): 911–031.

Lancaster, T. 2000. “The Incidental Parameter Problem Since 1948.” *Journal of Econometrics* 95: 391–413.

Lancaster, T. 2002. “Orthogonal Parameters and Panel Data.” *Review of Economic Studies* 69: 647–666.

Lindsay, B. G. 1995. *Mixture Models: Theory, Geometry and Applications,* NSF-CBMS Regional Conference Series in Probability and Statistics, volume 5, IMS-ASA.

Lindsay, B. G., and K. Roeder. 1992. “Residual Diagnostics in the Mixture Model.” *Journal of American Statistical Association* 87: 785–795.

Louis, T. A. 1982. “Finding the Observed Information Matrix when Using the EM Algorithm.” *Journal of the Royal Statistical Society, B* 44(2): 226–233.

McLachlan, G., and D. Peel. 2004. *Finite Mixture Models*. New York: John Wiley.

Mundlak, Y. 1978. “On the Pooling of Time Series and Cross Section Data.” *Econometrica* 46: 69–85.

Oakes, D. 1999. “Direct Calculation of the Information Matrix via the EM Algorithm.” *Journal of the Royal Statistical Society, B* 61(2): 479–482.

Severini, T. A. 2000. *Likelihood Methods in Statistics*. Oxford: Oxford University Press.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling*. New York: Chapman & Hall.

Smith, J. P. 1998. “Socioeconomic Status and Health.” *American Economic Review* 88: 192–196.

Smith, J. P., and R. Kington. 1997. “Demographic and Economic Correlates of Health in Old Age.” *Demography* 43: 159–170.