HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 03/11

Finite Mixture for Panels with Fixed Effects

Partha Deb
Pravin K. Trivedi

April 2011

Finite Mixture for Panels with Fixed Effects

Partha Deb
Departments of Economics
Hunter College and the Graduate Center, CUNY, and NBER

Pravin K. Trivedi
Departments of Economics
Indiana University - Bloomington

March 2011

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.

Acknowledgment: *We are grateful for comments and suggestions for improvement of an earlier version received from Anirban Basu, David Drukker, and Maarten Lindeboom.*

1. Introduction

In analyzing panel data in the context of linear models, the random effects and fixed effects models provide a central paradigm for data analysis. Finite mixture ("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 (2005) or Bago d'Uva (2005). The case of finite mixtures with fixed effects (FE) has attracted much less attention. However, the fixed effects model has a special place in the microeconometrics literature 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. In this paper we take the first steps to extend the FMM framework to panel data that includes fixed effects.

Since the finite mixture models appeared in the statistical literature in the 1960s and 1970s, they have proved to be a useful way of modeling and capturing discrete unobserved heterogeneity in the population based on the intuitive idea that different "types" may correspond to different latent classes or subpopulations. The key idea is that the unknown population distribution may be empirically approximated by a mixture of distributions with finite, but small, number of mixture components. Thus a mixture of normals has been extensively used as an approximating distribution for continuous outcomes. The monograph by McLachlan and Peel (2004) documents the enormous popularity of the mixture formulation in many areas of statistics, especially since the appearance of the path-breaking work on the expectations-maximization (EM) algorithm by Dempster, Laird and Rubin (1977) and Aitkin and Rubin (1985). The EM algorithm made the computation of the finite mixture (latent class) models accessible to applied researchers. The importance EM algorithm is demonstrated by the fact that it is used almost exclusively as the estimation method of choice in all the examples in the McLachlan and Peel monograph.

In recent years the finite mixture models has found many applications in health economics, especially for models based on cross section data; see Deb and Trivedi (1997, 2002), Conway and Deb (2005). The extension of the FMM to panel (or longitudinal data) has lagged somewhat. Bago d'Uva (2006a,b) has applied the FMM framework to British panel data; but the approach essentially uses the panel as a pooled cross section and does not explicitly introduce either fixed or random effects. This "population averaging" approach is an alternative to the RE/FE approaches. However, from the linear panel data literature it is known that if the data generating process (d.g.p.) has fixed effects then both RE and PA models yield inconsistent estimates. On the other hand, if the d.g.p. has random effects, the FE estimator is still consistent. So the fixed effects estimator has some important advantages. However, in nonlinear models some standard ways of handling fixed effects do not necessarily work, so the FE model, and by implication the FE-FMM model, is potentially problematic to estimate.

Our solution of the FE-FMM estimation problem is based on two key insights. The first, which is well-known in the panel data literature, is that the fixed effects (treated as nuisance parameters) can be eliminated either by transformation or, in fully parametric model, by concentrating them out of the likelihood and then maximizing the concentrated likelihood. This is equivalent to conditioning on a sufficient statistic for the fixed effect; hence the approach is commonly referred to as conditional likelihood (Lancaster, 2000). Then the full-data concentrated likelihood can be maximized using the EM approach. We develop this approach in the context of Normal and the Poisson regressions. Our numerical illustrations use both Monte Carlo simulation and real data within the context of 2-component finite mixtures.

Section 2 establishes the notation, the context, and the solutions for the standard one-component model. Section 3 introduces our extension to finite mixtures. Section 4 develops the EM algorithm Section 5 reports the results of a Monte Carlo study of different estimators. Section 6 provides two empirical illustrations. Section 7 concludes with discussion and remarks.

2. Handling Fixed Effects in Panel Data Models

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

Three widely used specifications are as follows:

Pooled (or population-averaged) model

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}; \quad i = 1, ..., N; \quad t = 1, ..., T \tag{1}$$

Individual-specific random effects model in which the random effect is uncorrelated with the regressors and the idiosyncratic component ε_{it} specifies

$$\begin{aligned} y_{it} &= \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, & (2) \\ \mathrm{E}[\alpha_i|\mathbf{x}'_{it}, \varepsilon_{it}] &= 0, \ t = 1, ..., T & (3) \end{aligned}$$

Individual-specific fixed effects model consists of (2)plus the assumptions $\mathrm{E}[\alpha_i \, |\mathbf{x}_{it}] \neq 0$, and $\mathrm{E}[\varepsilon_{it}|\mathbf{x}_{it}] = 0$, $i = 1, ..., N; \quad t = 1, ..., T$.

Throughout we assume that the covariates are strictly exogenous and include at least one variable that is time-varying. 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. Some relatively straightforward extensions are mentioned later. The unobserved errors ε_{it} are assumed to be i.i.d.

2.1. Incidental parameters problem

The most direct approach is to jointly estimate $\alpha_1, ..., \alpha_N$ and $\boldsymbol{\beta}$. Consistent estimation then relies on asymptotics that assume that the panel is "long", and $N \to \infty$. But in microeconometric applications a short panel is more plausible, and the asymptotic theory assume only that T is fixed while $N \to \infty$. This raises the possibility that the joint estimator of $(\alpha_1, ..., \alpha_N \ \boldsymbol{\beta})$ 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, 2004b) 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) surveys a number of other approaches for handling the incidental parameter problem.

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 β. In linear models with additive fixed effects, first differencing and "within" transformations are leading examples of this approach. It is well known that these transformations also sweep out the parameters of time invariant factors. 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-differencing transformations are available for specific nonlinear panel models (Cameron and Trivedi, 2005). While such semiparametric approaches are attractive, they are difficult to extend to finite mixture models (Deb, Ming, Trivedi et al., 1997). We, therefore, pursue solutions that can be implemented in a fully parametric framework.

In this paper we will take Andersen's (1970) conditional maximum likelihood approach in which the fixed effects are eliminated by conditioning on a sufficient statistic, provided such a sufficient statistic is available. In this approach maximum likelihood is applied after conditioning on a sufficient statistic for α_i. Computationally this is equivalent to concentrated (profile) maximum likelihood after concentrating out the nuisance parameters. In Normal linear models, this approach is also equivalent applying the within-transformation. Whereas this approach has limited applicability because the conditioning step is not always feasible, no difficulty arises in the case of the two leading cases considered in this paper.

2.1.1. Normal regression

Given (1), assume $\varepsilon_{it} \sim N[0, \sigma^2]$.Denote the sample means $T^{-1}\sum y_{it}$ and $T^{-1}\sum x_{it}$ by $(\bar{y}_i, \bar{x}_i)$ respectively. Then the within transformation yields

$$\begin{aligned} y_{it} - \bar{y}_i &= (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i), \\ &= (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + v_{it} \end{aligned}$$

where $v_{it} = (\varepsilon_{it} - \bar{\varepsilon}_i)$ and the fixed effect α_i is eliminated, along with time-invariant regressors since $\mathbf{x}_{it} - \bar{\mathbf{x}}_i = \mathbf{0}$ if $\mathbf{x}_{it} = \mathbf{x}_i$ for all t. At this step the time-invariant

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 time-invariant regressors do not appear in the transformed model.

The sufficient statistic for α_i is $\sum_{i,t} y_{it}$, or $T\bar{y}_i$. Then, under normality assumption MLE of $\boldsymbol{\beta}$ alone is based on the conditional likelihood:

$$\begin{aligned} L_{\text{COND}}(\boldsymbol{\beta},\sigma^2,\boldsymbol{\alpha}) &= \prod_{i=1}^{N} f(y_{i1},...,y_{iT}|\bar{y}_i) \\ &= \prod_{i=1}^{N} \frac{f(y_{i1},...,y_{iT},\bar{y}_i)}{f(\bar{y}_i)} \\ &= \prod_{i=1}^{N} \frac{(2\pi\sigma^2)^{-T/2}}{(2\pi\sigma^2/T)^{-1/2}} \exp\left\{ \sum_{t=1}^{T} \begin{array}{r} -\frac{1}{2\sigma^2}[(y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}))^2 \\ +(\bar{y}_i-\bar{\mathbf{x}}_i'\boldsymbol{\beta})^2] \end{array} \right\}. \end{aligned} \tag{4}$$

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

$$\sum_{t=1}^{T}\sum_{i=1}^{N}[(y_{it}-\bar{y}_i)-(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)'\boldsymbol{\beta})](\mathbf{x}_{it}-\bar{\mathbf{x}}_i)=\mathbf{0}, \tag{5}$$

which coincide 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, $\widehat{\boldsymbol{\beta}}_{\text{CML}}$ equals the within estimator. Again the fixed effects $\boldsymbol{\alpha}$ are eliminated by conditioning on $\bar{y}_i$, so we may maximize the conditional log-likelihood function with respect to $\boldsymbol{\beta}$ and σ^2 only.

2.1.2. Poisson regression

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

$$y_{it} \sim P(\lambda_{it}\alpha_i) \tag{6}$$

$$\lambda_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta}). \tag{7}$$

In this case also the sufficient statistic for α_i is $\sum_{i,t} y_{it}$, 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

$$\begin{aligned}\ln L(\boldsymbol{\beta},\boldsymbol{\alpha}) &= \ln\left[\prod\nolimits_i \prod\nolimits_t \{\exp(-\alpha_i\lambda_{it})\,(\alpha_i\lambda_{it})^{y_{it}}\,/y_{it}!\}\right] \\ &= \sum\nolimits_i \left[-\alpha_i \sum\nolimits_t \lambda_{it} + \ln\alpha_i \sum\nolimits_t y_{it} + \sum\nolimits_t y_{it}\ln\lambda_{it} - \sum\nolimits_t \ln y_{it}!\right]. \end{aligned} \quad (8)$$

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

$$\widehat{\alpha}_i = \sum_t y_{it} / \sum_t \lambda_{it}. \quad (9)$$

Substituting (9) back into (8) yields the following concentrated likelihood function after dropping terms not involving $\boldsymbol{\beta}$:

$$\ln L_{\text{conc}}(\boldsymbol{\beta}) \propto \sum\nolimits_i \sum\nolimits_t \left[y_{it}\ln\lambda_{it} - y_{it}\ln\left(\sum\nolimits_s \lambda_{is}\right)\right]. \quad (10)$$

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

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbf{x}_{it}\left(y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right) = \mathbf{0}, \quad (11)$$

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

3. Mixture Models with Fixed Effects

We wish to extend the conditional likelihood approach to finite mixture models based on, respectively, the Normal and the Poisson distributions. Before considering details we 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 we first apply it individually to the mixture components and then form a mixture of concentrated marginals. In the following section we apply such an approach.

The standard definition of a C-component mixture (or C latent classes) of an arbitrary density with fixed effects α_{ji}, $f(y_{it}|\mathbf{x}_{it},\boldsymbol{\theta}_j,\alpha_{ji})$, $j = 1, 2, ..., C$ is

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

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

$$y_{it} \sim f(y_{it}|\mathbf{x}_{it}, \boldsymbol{\theta}_j, \alpha_{ji}) \text{ with probability } \pi_j \tag{13}$$

Throughout the paper we consider the case of fixed mixing proportions, leaving to future work the task of extending this to the case where π_j is parametrized as a function of observable variables.

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 $i = 1, ..., N$,

$$f(\boldsymbol{\beta}_j, \sigma_j^2|\boldsymbol{\alpha}) = \frac{(2\pi\sigma_j^2)^{-T/2}}{(2\pi\sigma_j^2/T)^{-1/2}} \exp\left\{\sum_{t=1}^{T} -\frac{1}{2\sigma_j^2}[(y_{it} - \bar{y}_i) - (\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i)\boldsymbol{\beta}_j]^2\right\} \tag{14}$$

and the mixture distribution is

$$\sum_{j=1}^{C} \pi_j f_j(\boldsymbol{\beta}_j, \sigma_j^2|\overline{\mathbf{y}}_i, \mathbf{x}_i) \tag{15}$$

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

The expression (15) 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. Assume observations are permanently assigned to just one latent class during the panel period, which implies that $d_{it} = d_i$. More precisely,

$$d_{ji} = \begin{cases} 1 \text{ if } i \text{ belongs to the component } j \\ 0 \text{ otherwise} \end{cases}; \tag{16}$$

d_{ji} is not observed.

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

$$L_{conc}(\boldsymbol{\beta},\sigma^2|\boldsymbol{\alpha}) = \prod_{i=1}^{N}\prod_{t=1}^{T}\sum_{j=1}^{C}\left(\pi_j f(\boldsymbol{\beta}_j,\sigma_j^2|\overline{\mathbf{y}}_{it},\mathbf{x}_{it})\right)^{1(d_{ji}=1)}. \tag{17}$$

The maximization of this likelihood is based on the expectations-maximization algorithm. Details are given later in the paper. Later we also show that in this specific case of linear panel model, it is possible to avoid using the EM algorithm altogether and to instead use the marginal likelihood expressed in terms of transformed data $(\widetilde{\mathbf{y}},\widetilde{\mathbf{x}})$ where tilde denotes either within- or first-difference transformation.

3.2. Poisson mixture

Again our approach uses the fact that for each component of the distribution a sufficient statistic is available. Therefore, we can derive the conditional (on the sufficient statistic) component distributions and use them to define a finite mixture based likelihood function.

Specifically we have

$$y_{it}|\mathbf{x}_{it},\boldsymbol{\lambda}_{it}^{(j)},\alpha_{ji} \sim P[\alpha_{ji}\lambda_{it}^{(j)}],\ j=1,2,...,C \tag{18}$$

where $\lambda_{it}^{(j)} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta}_j)$.

Then using (9) we obtain

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

and the C-component mixture distribution is

$$y_{it}|\mathbf{x}_{it},\boldsymbol{\lambda}_{it},\widehat{\boldsymbol{\alpha}}_i \sim \sum_{j=1}^{C}\pi_j P\left[\left(\sum_t y_{it}/\sum_t \lambda_{it}^{(j)}\right)\lambda_{it}^{(j)}\right]. \tag{19}$$

Then the full-data concentrated likelihood can be formed combining (19) and (16) which yields

$$L_{conc}(\cdot) = \prod_{i=1}^{N}\prod_{t=1}^{T}\sum_{j=1}^{C}\left(\pi_j P(\lambda_{it}^{(j)})\right)^{1(d_{ji}=1)}. \tag{20}$$

Again this likelihood can be maximized using the EM algorithm.

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 $f(y_{it}; \theta_j)$. Let $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})$ 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 was drawn from the latent class j; $d_{ji} = 0$ otherwise and $\sum_j d_{ji} = 1$. Then, the *panel* finite mixture model specifies that $(\mathbf{y}_i | \mathbf{d}_i, \boldsymbol{\theta}, \boldsymbol{\pi})$ are independently distributed with densities

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

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

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

Thus

$$(y_{i1}, y_{i2}, ..., y_{iT} | \boldsymbol{\theta}, \boldsymbol{\pi}) \;\tilde{}\; \sum_{j=1}^{C} \left(\pi_j \prod_{t=1}^{T} f(y_{it}; \theta_j) \right)^{d_{ji}}.$$

The likelihood function is then

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

and the log likelihood function is

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

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

$$\mathrm{EL}(\boldsymbol{\theta} | \mathbf{y}, \boldsymbol{\pi}) = \sum_{i=1}^{N} \sum_{j=1}^{C} \widehat{z}_{ji} \left[\ln f_j(y_i; \boldsymbol{\theta}_j) + \ln \pi_j \right]. \tag{21}$$

The M-step of the EM procedure maximizes (21) by solving the first order conditions

$$\widehat{\pi}_j - \frac{\sum_{i=1}^{N} \widehat{z}_{ji}}{N} = 0, \quad j = 1, .., C \tag{22}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{C}\widehat{z}_{ij}\frac{\partial \ln f_j(y_i;\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}=0. \tag{23}$$

The marginal probability that an observation comes from the class j is the average of all individual observation probabilities of coming from the j^{th} population. The E-step of the EM procedure obtains new values of $\mathrm{E}[d_{ji}]$ using the (22).

The posterior probability that unit i belongs to population j, $j = 1, 2, ..., C$, denoted z_{ji} is given by

$$z_{ji}=\frac{\pi_j\prod_{t=1}^{T}f_j(y_{it};\theta_k)}{\sum_{j=1}^{C}\pi_j\prod_{t=1}^{T}f_j(y_{it};\theta_j)}.$$

The expected log likelihood function then is given by

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

For a given set of parameters, $\{\theta_j,\pi_j\}_{j=1,2,...,C}$ the E-step consists of calculating z_{ki} which gives the set of posterior probabilities of classification for each unit and also the values of $\{\pi_j\}_{j=1,2,...,C}$ for the next M-step. Given $\{\pi_j\}_{j=1,2,...,C}$, the M-step consists of maximizing the expected log likelihood function. The E- and M-steps are repeated in alternating fashion until the expected log likelihood fails to increase. At that point, we conduct a final M-step in which the $\{\pi_j\}_{j=1,2,...,C}$ are also estimated. The last step does not change any of the values of $\{\theta_j,\pi_j\}_{j=1,2,...,C}$ but is appealing because it provides standard errors for the estimates of $\{\pi_j\}_{j=1,2,...,C}$ which the preceding E-step estimates do not provide.

4.1. Simplified computation for the Normal mixture

The conditioning statistic in the case of the Normal mixture does not depend on unknown parameters. This implies that replacing $(y_i\ , x_{it})$ by $(\widetilde{\mathbf{y}}_i,\widetilde{\mathbf{x}}_{it})$, where $\widetilde{}$ denotes the within transformation, and then maximizing the mixture likelihood is numerically exactly equivalent to applying the full EM algorithm. 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. This implies also that fixed effect models with more than 2-components do not pose additional difficulties; they can be handled using the same software as for models without fixed effects. Inference procedures for determining the number of mixture components also extend to fixed effect panels.

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.

While the analysis has focussed on individual-specific effects, the approach extends straight-forwardly to cluster- or group-specific effects if clusters and groups are directly identified; within-individual transformation is replaced by within-group transformation.

5. A Monte Carlo Study

The main objective of the set of Monte Carlo experiments is to examine the finite sample properties of the proposed estimators. But we also discuss (a) what parameter a standard fixed effects model (which may be thought of as a one-component fixed effects model) identifies and (b) the biases of standard finite mixture models that do not account for the random or fixed effects of the data generating process (d.g.p.).

5.1. Normal mixtures

5.1.1. The d.g.p.

In the case of Normal mixtures, we specify the data generating process as follows.

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

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 experiments are based on the following parameter configurations, $T = (4, 8)$, $NT = 10000$, $\sigma_1 = \sigma_2 = 1$, $\beta_{11} = 1$, $\beta_{12} = 2$, 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, the individual effects are assumed to be uncorrelated with x_{it}, i.e., α_{i1} and α_{i2} are drawn from i.i.d. $N(0,1)$. In the second set of experiments, the individual effects are correlated with x_{it}. More precisely, they are correlated with $\overline{x}_{i.}$ which are the within-group means of x_{it} with the correlation generated by specifying $\alpha_{ij} = u_{ij} + \sqrt{T}\overline{x}_{i.}$ where u_{ij} are drawn from i.i.d. $N(0,1)$. Note that $\mathrm{Var}(\sqrt{T}\overline{x}_{i.}) = 1$. Finally, α_{ij} are rescaled to have unit standard deviation.

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 z_i such as

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

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_{ij} has unit standard deviation. Thus conditional on including z_i as regressors, the FE panel may be treated as a RE panel. Adding z_i as additional regressors then implements the "Mundlak correction". Given our d.g.p., the $\overline{x}_{i.}$ corresponds to z_i; therefore, adding $\overline{x}_{i.}$ as an additional regressor and then maximizing the marginal likelihood (without the EM algorithm) will 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

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

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_{ij} has unit standard deviation.

5.1.2. Estimator comparison

We estimate and compare the following estimators:

1. FE1: a standard fixed-effects linear regression which ignores the mixture aspect of the d.g.p.;

2. FM2: a standard 2-component finite mixture of Normal densities which ignores the mechanism generating individual specific effects;

3. FM2-M: a standard 2-component finite mixture of Normal densities with Mundlak correction;

4. FM2-FE: a 2-component fixed-effects finite mixture of Normal densities which accommodates the individual specific effects, using the EM algorithm.

Note that 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

$$\begin{aligned} \mu &= \pi\mu_1 + (1-\pi)\mu_2 \\ \sigma^2 &= \pi\sigma_1^2 + (1-\pi)\sigma_2^2 + \pi(1-\pi)(\mu_1 - \mu_2)^2. \end{aligned}$$

This particular $N(\mu, \sigma^2)$ is regarded as a projection of the two-component model on the one-component model. FE1 identifies (μ, σ^2).

The misspecified one-component model therefore identifies μ, σ^2 which are weighted functions of the underlying parameters. When the two components are well separated, so that $\mu_1 >> \mu_2$, variance estimator $\widehat{\sigma}^2$ 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 $\widehat{\mu}$. 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 the Normal mixtures are shown in Tables 1-3. Both RE and FE specifications are considered. The summary statistics are based on 1000 replications. Table 1 is for the (orthogonal) RE d.g.p., for $T = 4$ and 8, respectively, and 2 for the corresponding (correlated) FE specification.

Under the RE d.g.p., and given that the regression has only one time-varying regressor, FE1 yields a consistent estimate of $\beta = \pi\beta_1 + (1-\pi)\beta_2 = 1.5$. The result for both $T = 4$

and $T = 8$ are consistent with the theoretical expectation. Under the RE formulation, each component of the two-component mixture is overdispersed, a feature neglected by the standard finite mixture of Normals. Yet the estimator retains its consistency property. The Monte Carlo results are again in line with the theoretical expectations; the results for $T = 8$ are more precise, but the estimator is not efficient.

For the FE d.g.p. specified using the Mundlak-Chamberlain formulation, FE1 is a consistent estimator of $\beta = \pi\beta_1 + (1-\pi)\beta_2$. FM2 is not a consistent estimator of β_1 and β_2. The Monte Carlo results indicate a larger bias for both coefficients of around 35% for the $T = 4$ case and a smaller bias of around 25% for the $T = 8$ case. On the other hand the mixing fraction π shows almost zero bias (though we currently do not have any theoretical results to support this simulation result). FM2-FE should be consistent and the results are once again satisfy this expectation. As expected, applying FM2-M leads to results very close to those from the EM algorithm for maximizing the full-data likelihood.

Finally, for the FE d.g.p. specified using a non-Mundlak-Chamberlain formulation, FE1 is still a consistent estimator of $\beta = \pi\beta_1 + (1-\pi)\beta_2$. As with the other FE case, FM2 is not a consistent estimator of β_1 and β_2. The Monte Carlo results indicate a large bias for both coefficients for both $T = 4$ and $T = 8$ cases. Unlike the results of the previous case, the results reported in Table 3 show that applying FM2-M leads to biased coefficients for β_1 and β_2 as well as for π. Only the results from FM2-FE are consistent.

5.2. Poisson mixtures

5.2.1. The d.g.p.

In the case of Poisson mixtures, we specify the data generating process as follows:

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

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

where $\alpha_{ij} \equiv \exp(\tau_{ij})$. The experiments are based on the following parameter configurations, $T = (4, 8)$, $NT = 10000$, $\beta_{11} = 0.2$, $\beta_{12} = 1.0$, 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, the individual effects are random, i.e., τ_{i1} and τ_{i2} are drawn from i.i.d. $N(0, 0.4)$. In the second set of experiments, the individual effects are correlated with x_{it}. More precisely, they are correlated with $\overline{x}_{i.}$ which are the within-group means of x_{it} with the correlation generated by specifying $\tau_{ij} = u_{ij} + \sqrt{T}\overline{x}_{i.}$ where u_{ij} are drawn from i.i.d. $N(0, 1)$. Note that $\text{Var}(\sqrt{T}\overline{x}_{i.}) = 1$. Finally, τ_{ij} are rescaled to have unit standard deviation.

5.2.2. Estimator comparison

Given these data generating processes, we estimate and compare

1. FE1: a standard fixed-effects one-component Poisson regression without fixed effect adjustment;
2. FM2: a standard 2-component mixture of Poisson ignoring individual specific effects;
3. FM2-M: a standard 2-component mixture of Poisson with "Mundlak-type" correction;
4. FM2-FE: a 2-component fixed-effects finite mixture of Poisson individual specific effects, using the EM algorithm.

The first three are misspecified models and the fourth is correctly specified.

In the case of FE1 and 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 $T = 8$.

5.2.3. Results

The results for the Poisson simulations are given in Table 4-5. Once again, the summary statistics reported are based on 1000 simulations. The results for the (orthogonal) RE d.g.p. are in Table 4, and those for the FE d.g.p. are in Table 5.

A RE Poisson mixture is an overdispersed model, so under the RE specification the d.g.p. is a mixture of two such overdispersed models. The simulation results confirm that FE1 and FM2 are inconsistent estimators. This result is significantly different from that for the standard (one-component) Poisson model in which the Poisson MLE remains consistent under overdispersion. On the other hand, FM2-FE remains consistent under the RE d.g.p.; and there is very little difference between $T = 4$ and $T = 8$ outcomes.

In the case where the d.g.p. is a 2-component Poisson mixture with fixed effects, the maximum likelihood estimators of the two misspecified models are again both inconsistent. By contrast, our proposed fixed effects MLE for the mixture model has a very good performance both when $T = 4$ and $T = 8$, with bias not significantly different from zero.

Given the nonlinearity of the conditional mean, the "Mundlak correction" that we apply no longer has a formal mathematical basis. In fact, neither the existence nor the form of the appropriate correction is known to us. So including the same additional regressor as in the Normal case is speculative, and it does not remove the bias due to the correlated random effect.

6. Empirical Applications

We further apply the models described in the Monte Carlo study above using panel data from six waves of the HRS: 1992, 1994, 1996, 1998, 2000 and 2002, The empirical question of interest is the effect of turning 65 and becoming Medicare eligible on medical utilization among those who have been uninsured compared to others conditional on a broad set of measures of the use of health services and health status available in the HRS. We study the expenditures on healthcare excluding zeros and the number of office-based doctor visits. Specifically, we model the logarithm of expenditures using the normal family of models and 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 U.S. 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 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 design arguably provides unbiased estimates of the effect of insurance on the use of health services, since individuals are not selecting Medicare coverage based on current health status or other variables. In most difference-in-difference studies, the data are a cross-section, thus it is not possible to identify individuals after age 65 who were uninsured before age 65. Instead, a synthetic control group is used consisting of similar individuals before age 65. (Card, et al. 2008) Because the HRS is a panel of individuals, many of whom turn 65 during the surveys, panel data methods should be able to control for individual-specific unobserved heterogeneity much better than is possible with cross-sectional data. Thus we estimated models under RE assumptions using pooled cross-sectional methods including standard FM2 and FM2-FE panel data models.

6.1. Data

The HRS is a population-based sample of U.S. community dwellers aged 51 to 61 in 1992 (and their spouses) with follow-up interviews every two 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 U.S. (for example, 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 to 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. The final data set is a mildly unbalanced panel.

Approximately 17 percent 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 and number of doctor visits. But these differences cannot be thought of as causal impacts of uninsurance because of observed and unobserved heterogeneity. The summary statistics show that they 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 that there are substantive differences in unobserved characteristics as well, motivating the importance of the fixed effect approach to eliminating individual-specific unobserved heterogeneity.

6.2. Results

In the models for log expenditure shown in Table 7, the estimate on the interaction term in OLS estimation is positive and significant (0.328). Thus, Medicare eligibility for those who have been uninsured substantially increases expenditures compared to those who have been insured. The effect sizes are considerably different between the two classes of individuals. For the majority of individuals ($\widehat{\pi} = 69\%$), the effect is large and significant (0.444) while for a substantial (31%) minority, the result is small and insignificant. The estimates in the first two columns of Table 3b 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 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 0.328 to 0.474. The results in Table 8 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.

Results for the models of counts of doctor visits are shown in Tables 9 and 10. 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 23.4% compared to those who have been insured. This estimate decreases to 17% once fixed effects are introduced; note the contrast with the direction of the change in effects with the introduction of fixed effects in the case of positive expenditures. Both 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 for both components, with the larger effect observed for the component with the higher probability of occurrence ($\widehat{\pi} = 0.862$). Once fixed effects are introduced into the FM Poisson regression, the effect size for the second component becomes 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. Conclusion and qualification

This paper has focussed on the fixed effects panel MLE for 2-component finite mixtures of Normal and Poisson distributions. We extend the concentrated (conditional) likelihood approach of Andersen (1970) for handling the fixed effects to a mixture model. 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.

8. References

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.

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

Bago d'Uva. T. 2005. Latent class models for use of primary care: evidence from a British panel. *Health Economics* 14: 873–892

Bago d'Uva, T. 2006. Latent class models for use of health care. *Health Economics.* 15: 329–343.

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), 2242–58.

Chamberlain, G. 1984. Panel Data. In *Handbook of Econometrics*, Volume II, ed. 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., X. Ming, and P.K. Trivedi. 1997. Semiparametric Estimation of Finite Mixture Models of Event Counts, With Application to Demand for Medical Care by the Elderly. Unpublished discussion paper.

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.* Series B (Methodological), 39 (1), 1-38.

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*, 179-195. Cambridge, MA: MIT Press:

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.

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.

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

Smith, James P. 1998. Socioeconomic Status and Health. *American Economic Review*, 88, 192-196.

Smith, James. P. and Raynard Kington. 1997. Demographic and Economic Correlates of Health in Old Age, *Demography*, 43, 159-70.

Winkelmann, R. 2004. Health care reform and the number of doctor visits-an econometric analysis, *Journal of Applied Econometrics*, 19, 455-472.

Wooldridge, J. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20, 39-54.

Table 1: 2-Component Mixture of Normals with Orthogonal Random Effects

		T = 4		T = 8	
Parameter	True value	Mean	Std. Dev.	Mean	Std. Dev.
FE: Fixed Effects Linear Regression					
β		1.501	0.017	1.500	0.018
FM2: 2-Component Mixture					
β_1	1.0	0.995	0.075	1.002	0.078
β_2	2.0	2.000	0.075	2.003	0.079
π_1	0.5	0.497	0.063	0.503	0.067
FM2-M: Mixture with Mundlak Correction					
β_1	1.0	0.997	0.076	1.004	0.078
β_2	2.0	2.000	0.078	2.000	0.080
π_1	0.5	0.498	0.064	0.502	0.068
FM2-FE: Mixture with Fixed Effects					
β_1	1.0	0.945	0.034	0.991	0.022
β_2	2.0	2.054	0.033	2.010	0.022
π_1	1.5	0.499	0.028	0.501	0.022

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

		T = 4		T = 8	
Parameter	True value	Mean	Std. Dev.	Mean	Std. Dev.
FE: Fixed Effects Linear Regression					
β		1.500	0.016	1.499	0.019
FM2: 2-Component Mixture					
β_1	1.0	1.349	0.072	1.233	0.077
β_2	2.0	2.345	0.072	2.251	0.074
π_1	0.5	0.500	0.060	0.494	0.064
FM2-M: Mixture with Mundlak Correction					
β_1	1.0	1.000	0.056	1.000	0.054
β_2	2.0	2.000	0.056	1.997	0.056
π_1	0.5	0.500	0.046	0.500	0.047
FM2-FE: Mixture with Fixed Effects					
β_1	1.0	0.951	0.058	0.991	0.022
β_2	2.0	2.060	0.055	2.010	0.022
π_1	1.5	0.505	0.052	0.501	0.023

Table 3: 2-Component Mixture of Normals with non-Mundlak Type Correlated Random Effects

		T = 4		T = 8	
Parameter	True value	Mean	Std. Dev.	Mean	Std. Dev.
FE: Fixed Effects Linear Regression					
β		1.501	0.017	1.500	0.018
FM2: 2-Component Mixture					
β_1	1.0	1.304	0.087	1.329	0.135
β_2	2.0	2.277	0.138	2.248	0.243
π_1	0.5	0.563	0.091	0.616	0.153
FM2-M: Mixture with Mundlak Correction					
β_1	1.0	1.184	0.067	1.211	0.094
β_2	2.0	1.885	0.043	1.873	0.054
π_1	0.5	0.559	0.071	0.586	0.119
FM2-FE: Mixture with Fixed Effects					
β_1	1.0	0.977	0.055	1.000	0.051
β_2	2.0	2.086	0.053	2.012	0.037
π_1	1.5	0.529	0.050	0.507	0.052

Table 4: 2-Component Mixture of Poissons with Orthogonal Random Effects

		T = 4		T = 8	
Parameter	True value	Mean	Std. Dev.	Mean	Std. Dev.
FE: Fixed Effects Poisson Regression					
β		0.769	0.012	0.775	0.013
FM2: 2-Component Mixture					
β_1	1.0	0.401	0.024	0.402	0.026
β_2	2.0	0.903	0.017	0.904	0.017
π_1	0.5	0.601	0.014	0.601	0.016
FM2-M: Mixture with Mundlak Correction					
β_1	1.0	0.401	0.025	0.401	0.026
β_2	2.0	0.903	0.017	0.905	0.017
π_1	0.5	0.601	0.014	0.601	0.016
FM2-FE: Mixture with Fixed Effects					
β_1	1.0	0.235	0.021	0.243	0.020
β_2	2.0	0.982	0.013	0.988	0.011
π_1	1.5	0.513	0.015	0.523	0.016

Table 5: 2-Component of Poissons with
Correlated Random Effects

		T = 4		T = 8	
Parameter	True value	Mean	Std. Dev.	Mean	Std. Dev.
FE: Fixed Effects Poisson Regression					
β		0.775	0.010	0.778	0.011
FM2: 2-Component Mixture					
β_1	1.0	0.466	0.022	0.445	0.020
β_2	2.0	0.990	0.015	0.956	0.014
π_1	0.5	0.585	0.012	0.586	0.015
FM2-M: Mixture with Mundlak Correction					
β_1	1.0	0.349	0.022	0.348	0.020
β_2	2.0	0.919	0.015	0.915	0.014
π_1	0.5	0.571	0.012	0.570	0.015
FM2-FE: Mixture with Fixed Effects					
β_1	1.0	0.192	0.034	0.277	0.019
β_2	2.0	0.960	0.015	0.994	0.009
π_1	1.5	0.459	0.038	0.544	0.021

Table 6: Summary Statistics of HRS Data

	ever uninsured = 0		ever uninsured = 1	
Variable	mean	sd	mean	sd
total expenditures (given > 0)	7,342	24,949	7,420	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=5,011 for the sample of ever uninsured=1. Sample sizes for positive expenditures (and its logarithm) are 24,324 and 4,403 for the two samples respectively.

Table 7: Models of (log) Expenditures

Variable	OLS	FM Normal component1	FM Normal component2	FE Regression
age	0.074**	0.086**	0.048**	0.127**
	(0.003)	(0.005)	(0.009)	(0.003)
Medicare eligible	-0.021	-0.026	-0.003	-0.033
	(0.040)	(0.069)	(0.107)	(0.040)
been uninsured	-0.168**	-0.136**	-0.233**	-0.187**
	(0.028)	(0.049)	(0.082)	(0.063)
been uninsured X Medicare eligible	0.328**	0.444**	0.072	0.474**
	(0.074)	(0.120)	(0.161)	(0.083)
married	-0.060*	-0.077	-0.017	0.072
	(0.024)	(0.042)	(0.070)	(0.052)
income quartile 1	-0.255**	-0.153*	-0.501**	-0.108**
	(0.033)	(0.062)	(0.100)	(0.042)
income quartile 2	-0.135**	-0.037	-0.358**	-0.060
	(0.026)	(0.048)	(0.075)	(0.034)
income quartile 3	-0.058*	0.028	-0.252**	-0.002
	(0.024)	(0.043)	(0.066)	(0.029)
female	0.023	-0.121**	0.362**	
	(0.018)	(0.038)	(0.056)	
black	-0.047	-0.150**	0.192*	
	(0.029)	(0.054)	(0.094)	
hispanic	-0.152**	-0.118	-0.219	
	(0.039)	(0.068)	(0.112)	
other race	-0.114	-0.066	-0.272	
	(0.069)	(0.130)	(0.232)	
high school dropout	-0.075*	0.035	-0.334**	
	(0.032)	(0.058)	(0.100)	
high school degree	-0.109**	-0.009	-0.339**	
	(0.025)	(0.046)	(0.074)	
college attendee	-0.035	0.094	-0.344**	
	(0.028)	(0.050)	(0.084)	
lives in midwest	0.019	0.007	0.041	
	(0.027)	(0.049)	(0.081)	
lives in south	-0.011	0.004	-0.047	
	(0.025)	(0.045)	(0.074)	
lives in west	-0.020	-0.038	0.025	
	(0.031)	(0.055)	(0.091)	
π_1		0.691		
		(0.037)		
lnL	-52439.030	-52253.789		-46410.991
Observations	28,727	28,727		28,727

Standard errors in parentheses

** p<0.01, * p<0.05

Table 8: Models of (log) Expenditures

	FM Normal		Mundlak FM Normal		Fixed Efffects FM Normal	
Variable	comp.1	comp.2	comp.1	comp.2	comp.1	comp.2
age	0.088**	0.044**	0.150**	0.066**	0.174**	0.043**
	(0.005)	(0.009)	(0.006)	(0.012)	(0.004)	(0.003)
Medicare eligible (M)	-0.021	-0.031	-0.062	0.038	-0.044	-0.020
	(0.069)	(0.114)	(0.072)	(0.128)	(0.049)	(0.043)
been uninsured (U)	-0.132**	-0.351**	-0.252*	0.021	-0.077	-0.430**
	(0.048)	(0.082)	(0.120)	(0.221)	(0.077)	(0.069)
U X M	0.435**	0.109	0.554**	0.283	0.547**	0.289**
	(0.119)	(0.166)	(0.143)	(0.221)	(0.101)	(0.092)
married	-0.060	-0.122	0.077	0.043	0.090	0.025
	(0.042)	(0.071)	(0.092)	(0.157)	(0.065)	(0.055)
income quartile 1	-0.224**	-0.510**	-0.100	-0.156	-0.004	-0.302**
	(0.059)	(0.100)	(0.076)	(0.139)	(0.051)	(0.045)
income quartile 2	-0.084	-0.394**	-0.076	-0.019	0.031	-0.220**
	(0.047)	(0.077)	(0.061)	(0.111)	(0.042)	(0.036)
income quartile 3	0.007	-0.299**	0.023	-0.081	0.047	-0.096**
	(0.043)	(0.069)	(0.053)	(0.094)	(0.035)	(0.030)
π_1	0.703		0.721		0.637	
	(0.040)		(0.036)		(0.006)	
lnL	-52310.674		-51911.336		-48721.800	
Observations	28,727		28,727		28,727	

Standard errors in parentheses

** $p<0.01$, * $p<0.05$

Table 9: Models of Number of Visits to the Doctor

	Poisson	FM Poisson		FE Poisson
Variable		component1	component2	
age	0.050**	0.064**	0.032**	0.083**
	(0.001)	(0.001)	(0.002)	(0.001)
Medicare eligible (M)	-0.115**	-0.208**	-0.061**	-0.097**
	(0.010)	(0.016)	(0.022)	(0.011)
been uninsured (U)	-0.157**	-0.247**	-0.046*	-0.094**
	(0.008)	(0.015)	(0.018)	(0.019)
U X M	0.234**	0.201**	0.121**	0.169**
	(0.017)	(0.029)	(0.032)	(0.022)
married	-0.048**	-0.023*	-0.090**	-0.026
	(0.006)	(0.010)	(0.013)	(0.015)
income quartile 1	-0.083**	-0.208**	-0.023	0.038**
	(0.009)	(0.016)	(0.018)	(0.012)
income quartile 2	-0.028**	-0.071**	0.102**	0.020*
	(0.007)	(0.012)	(0.015)	(0.010)
income quartile 3	0.009	0.028**	0.146**	0.033**
	(0.006)	(0.010)	(0.015)	(0.008)
female	0.219**	0.247**	0.170**	
	(0.005)	(0.008)	(0.011)	
black	0.073**	0.123**	-0.010	
	(0.007)	(0.012)	(0.016)	
hispanic	0.022*	0.035	0.044	
	(0.010)	(0.019)	(0.023)	
other race	0.004	-0.119**	-0.075*	
	(0.018)	(0.036)	(0.038)	
high school dropout	-0.097**	-0.224**	-0.037*	
	(0.008)	(0.014)	(0.018)	
high school degree	-0.094**	-0.166**	-0.132**	
	(0.007)	(0.011)	(0.015)	
college attendee	-0.035**	-0.096**	-0.015	
	(0.007)	(0.012)	(0.016)	
lives in midwest	-0.097**	-0.084**	-0.113**	
	(0.007)	(0.012)	(0.016)	
lives in south	-0.129**	-0.136**	-0.117**	
	(0.007)	(0.011)	(0.014)	
lives in west	-0.066**	-0.090**	-0.019	
	(0.008)	(0.013)	(0.018)	
π_1		0.862		
		(0.002)		
lnL	-145795.089	-98542.228		-75565.675
Observations	30,293	30,293		29,996

Standard errors in parentheses
** p<0.01, * p<0.05

Table 10: Models of Number of Visits to the Doctor

	FM Poisson		FM-Linear Mundlak		FM-Fixed Effects	
Variable	comp.1	comp.2	comp.1	comp.2	comp.1	comp.2
age	0.064**	0.031**	0.095**	0.062**	0.084**	0.082**
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Medicare eligible (M)	-0.209**	-0.041	-0.154**	-0.003	-0.105**	-0.085**
	(0.016)	(0.022)	(0.018)	(0.024)	(0.014)	(0.019)
been uninsured (U)	-0.289**	-0.051**	-0.096**	0.017	-0.096**	-0.090**
	(0.015)	(0.018)	(0.031)	(0.041)	(0.026)	(0.028)
U X M	0.221**	0.137**	0.150**	-0.216**	0.266**	0.042
	(0.029)	(0.033)	(0.040)	(0.048)	(0.029)	(0.035)
married	-0.114**	-0.128**	0.019	-0.047	-0.036	-0.011
	(0.010)	(0.013)	(0.023)	(0.030)	(0.019)	(0.024)
income quartile 1	-0.231**	-0.012	0.004	0.063*	-0.050**	0.181**
	(0.015)	(0.017)	(0.020)	(0.026)	(0.016)	(0.019)
income quartile 2	-0.103**	0.088**	0.018	0.094**	-0.037**	0.123**
	(0.011)	(0.014)	(0.016)	(0.021)	(0.013)	(0.016)
income quartile 3	0.000	0.107**	0.048**	0.145**	-0.018	0.125**
	(0.010)	(0.014)	(0.013)	(0.019)	(0.010)	(0.014)
π_1	0.862		0.863		0.842	
	(0.002)		(0.002)		(0.005)	
lnL	-99373.994		-98052.544		-92428.822	
Observations	30,293		30,293		29,996	

Standard errors in parentheses
** p<0.01, * p<0.05