A Monte Carlo Study of EC-estimation in Panel Data Models with Limited Dependent Variables and Heterogeneity

Mahmoud A. El-Gamal and David M. Grether *

March 1997.

Abstract

The EC (Estimation-Classification) estimator, and its companion EC-algorithm, were introduced in El-Gamal and Grether (1995), and their properties further analyzed in El-Gamal and Grether (1996). The purpose of EC estimation is to uncover heterogeneity in panel data models in a manner which is more parsimonious and computationally less costly than some of the standard methods (e.g. fixed effects). The latter concern is particularly evident in limited dependent variable models where no simple method of estimating fixed effects is available (e.g. probits). One of the applications of El-Gamal and Grether (1996) employed the EC algorithm to estimate multiple probits in a population of 257 individuals, each being observed for 14 to 19 periods, providing very satisfactory results. Since the asymptotic theory behind EC estimation relies on “Large T” approximation (i.e. we require $T \uparrow \infty$, then $n \uparrow \infty$ in proving consistency of the estimator), we provided a diagnostic statistic (called the Average Normalized Entropy (ANE) for diagnosing the adequacy of the large $T$ approximation. In this paper, we provide a Monte Carlo analysis of the EC estimator in comparison to pooling, and fixed effects (pooled slopes) estimators. The results show that for $T$ as small as 3, the EC-estimator does significantly better than fixed effects estimators in unpoolable slope environments, and almost as well in the poolable slopes case. As $T$ gets larger the EC estimator’s performance becomes progressively superior, with $T = 20$ providing virtually perfect estimation of the number of types, what the types are, and the classification of individuals to types. The ANE statistic is found to provide a very useful indicator of the proportion of possible misclassifications.

Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706-1393, and DHSS, 228-77, Caltech, Pasadena, CA 91125. We acknowledge financial support from NSF grant #SBR-9320497 to the California Institute of Technology and The University of Wisconsin at Madison.
A Monte Carlo Study of EC-estimation in Panel Data Models with Limited Dependent Variables and Heterogeneity

Mahmoud A. El-Gamal and David M. Grether

1 Introduction

In El-Gamal and Grether (1995), El-Gamal and Grether (1996), we introduced an Estimation-Classification (EC) estimator for panel data models. Accompanying the estimator, we provided an EC-algorithm which ensured the linearity of the number of required computations in the number of types in the population (k), the number of individuals in the sample (n), and the number of time series observations per individual (T). The EC-estimator was shown to be consistent as T, goes to infinity, and then as n goes to infinity. Moreover, it was shown in El-Gamal and Grether (1996) that if we have a $\sqrt{nT}$ consistent and asymptotically normal (CAN) estimator in the homogenous case (k = 1), then the EC estimator will consistently estimate k, and obtain (CAN) estimators of the k sets of parameters, with a block-diagonal variance covariance matrix.

The general framework of the estimator is a panel model with heterogeneity of unknown type, whereby we observe data $(y_{it}, \bar{x}_{it})$ for n individuals over $T_i$ time periods each. It is assumed that there are k (unknown) types of individuals parametrized by $(\theta_1, \ldots, \theta_k) \in \Theta^k$ (unknown); where $\Theta$ is the parameter space indexing the individuals’ likelihood function. Thus, individual i’s data, is assumed to be generated by

$$(y_{it}, \bar{x}_{it})_{t=1}^{T_i} \sim \prod_{j=1}^{k} f\left(\{y_{it}, \bar{x}_{it}\}_{t=1}^{T_i}; \theta_j\right)^{\delta_{ij}},$$

where $\delta_{ij} \in \{0, 1\}$ and $\sum_{j=1}^{k} \delta_{ij} = 1$.

If, for example, each individual’s data are generated by one of k regression functions, then the types would be defined by a vector of regression parameters. Note that some parameters could be constant across types, though for notational convenience we do not specifically indicate common values.

In the simplest framework considered in this paper, each individual’s data are assumed independent of each other individual, thus producing an overall log-likelihood function of the data $\{(y_{it}, \bar{x}_{it})\}_{i=1, t=1}^{n, T_i}$:

$$\ell(k; \theta_1, \ldots, \theta_k; \{\delta_{ij}\}) = \sum_{i=1}^{n} \left(\delta_{ij} \log f\left(\{y_{it}, \bar{x}_{it}\}_{t=1}^{T_i}; \theta_{h}\right)\right).$$

It is useful to consider the special case of a fixed effects model. In that framework, $k = n$, and the $\theta_j$’s differ only on one component (the intercept term). In that special case, it is
obvious that for \( n \uparrow \infty \) and a fixed \( T \), the \( \theta_j \)'s cannot be estimated consistently. It is also clear in that framework that if \( T \uparrow \infty \) for a fixed \( n \), the \( \theta_j \)'s can be consistently estimated. Whether \( T \) is large or small relative to \( n \) in a given sample is an issue that cannot be determined by asymptotic theory, it is simply a property of a data set.

In an influential paper, Heckman and MaCurdy (1980) provided the following argument for estimating fixed effects for \( n = 425 \) with only \( T = 8 \) observations per household (p.59):

"If the number of panel observations per person becomes large, and if the number of persons in the sample becomes large, maximum likelihood estimators ... are consistent and asymptotically normally distributed.

"Since the number of observations per household is fixed, it is not possible to consistently estimate fixed effects ...

"Whether or not this is a serious problem cannot be settled a priori. All samples are finite, and many cannot be imagined to become infinite ...."

As is explained in detail in Section 2 the approach we take treats the \( \delta_{ij} \)'s as parameters to be estimated. Thus, given \( k \), the individuals are classified (or clustered as there is no training set available) to one of the \( k \) types. The other parameters (\( \theta_j \)'s) are then estimated conditional upon the classification in an iterative procedure. For our approach to be useful it does not matter whether \( T \) is large of small relative to \( n \), but whether \( T \) is large enough so that the assignments to types can be made with sufficient accuracy.

In the more general heterogeneity model, the statistics literature prefers the use of the EM algorithm (Dempster et al. (1977), Redner and Walker (1984), Little and Rubin (1987)), thus treating the \( \delta_{ij} \) as missing data. In the case of \( k = n \) (as in the fixed effects model), the issue of whether the \( \delta_{ij} \)'s are parameters, nuisance parameters, or missing data, is irrelevant. However, in the case of \( k < n \) (which is preferable from the point of view of parsimony), treating the \( \delta_{ij} \)'s as parameters (as in the EM algorithm) implies the need for their estimation (via penalized maximum likelihood in EC), whereas treating them as missing data or nuisance parameters implies the need for integrating them out (as in EM). Little and Rubin (1983) criticize the estimation of the \( \delta_{ij} \) via maximum likelihood on the basis of inconsistency for a fixed \( T \) as \( n \uparrow \infty \). Our response in El-Gamal and Grether (1995), El-Gamal and Grether (1996) was similar in spirit to the large \( T \) argument of Heckman and MaCurdy (1980) discussed above for the special case of fixed effects. In El-Gamal and Grether (1996), we proved the consistency and asymptotic normality of our estimators as \( T \uparrow \infty \) and then as \( n \uparrow \infty \). Moreover, we devise a large \( T \) approximation to the EM algorithm, by means of which we can produce a diagnostic statistic (called Average Normalized Entropy, or ANE), to determine the goodness of the large \( T \) approximation. If \( T \) is found not to be sufficiently large to trust the EC estimates, the latter become consistent initial conditions for the EM algorithm as suggested by Little and Rubin (1987), which is very useful given the slowness of convergence of that algorithm.

A more recent appeal to using a large \( T \) approximation to address issues of heterogeneity in panels is discussed in Pesaran et al. (1996) (and the references therein) in the context
of dynamic panels. In this paper, as in other recent research in Econometrics (see e.g. Maddala (1991), Maddala (1996)), the question of the adequacy of fixed effects in treating panel heterogeneity is revived. The so called “pooling problem” (whether it is appropriate to assume homogeneity of all parameters other than the intercept) becomes more acute in the dynamic case with lagged dependent variables and possible persistence (e.g. unit roots) in the exogenous variables. Pesaran et al. (1996) show that the estimated parameters can exhibit perverse behavior in this environment, where the bias increases with the degree of persistence in the exogenous variables. Their recipe is to allow for heterogeneity in parameters other than the intercept (as a number of studies referenced therein suggest that “slope heterogeneity” is pervasive in economic panels), and utilize a large $T$ approximation to justify the procedure (p. 146):

“Therefore, in general, slope heterogeneity cannot be ignored and assuming it away on the grounds that one is dealing with relatively small $T$ may not be justifiable. In such circumstances it is perhaps more prudent to admit the limitations of one’s data and be more cautious about the conclusions...”

The remainder of this paper will proceed as follows. In Section 2, we state the EC-estimator, EC-algorithm, and the large $T$ approximation to the EM-algorithm and ANE formally. In Section 3, we discuss the general design of our Monte Carlo study of the unpoolable-slopes probit and poolable-slopes probit. In Section 4, we show the results of our Monte Carlo analysis, and in Section 5, we provide some concluding remarks.

## 2 The EC-EM and Large $T$ Approximation Setup

Under the outlined model of heterogeneity in the introduction, our log likelihood function is:

$$\mathcal{L}(k; \theta_1, \ldots, \theta_k; \{\delta_{ij}\}) = \sum_{i=1}^{n} \left( \delta_{ij} \log f\left( \{y_{it}, x_{it}\}_{t=1}^{T_i}; \theta_h \right) \right).$$

Our EC-estimator maximizes this likelihood function less a non-stochastic penalty function which grows in $n$ and $k$:

$$\left(\hat{k}, \hat{\theta}_1, \ldots, \hat{\theta}_k\right) = \operatorname*{argmax}_{k', \theta_1, \ldots, \theta_{k'}} \left\{ \sum_{i=1}^{n} \left( \max_{h \in \{1, \ldots, k'\}} \log f\left( \{y_{it}, x_{it}\}_{t=1}^{T_i}; \theta_h \right) \right) - \text{penalty}(n, k') \right\}.$$ 

When the penalty function ($-\text{penalty}(n, k')$) is log of a prior on the parameter space, this estimator is similar to minimum description length (Rissanen (1978)) and minimum message length (Wallace and Boulton (1968)). In this paper, we shall take our prior on $(\theta_1, \ldots, \theta_k)$ to be an improper flat prior, but in any given application (as in El-Gamal and Grether (1995), El-Gamal and Grether (1996)) there may be theoretical considerations which produce a more informed prior. In this paper, we shall impose a prior $1/2^k$ for $k = 1, 2, \ldots$, and a prior $k!/k^n$ (the reciprocal of the leading term of the sum of Stirling numbers of the second type
determining the number of possible \( \{ \delta_{ij} \} \) configurations). This yields a penalty function 
\[ \text{penalty}(n, k') = k' \log(2) + n \log(k') - \log(k'!) \] .

In El-Gamal and Grether (1996), we proved that if the estimator \( \hat{\theta}_1 \) is \( \sqrt{nT} \)-CAN when 
\( k = 1 \), then our EC estimator \( (\hat{\theta}_1, \ldots, \hat{\theta}_k) \) is \( \sqrt{nT} \)-CAN when we take limits as \( T \uparrow \infty \), and then as \( n \uparrow \infty \). The implementation of the EC-estimation, however, can be very difficult if we need to compare the log-likelihood for all possible \( k^n/k! \) possible classifications. However, the EC estimator as written above implicitly maximizes over all possible \( \{ \delta_{ij} \} \) configurations by employing the EC-algorithm. All that is needed to implement this EC-estimation using 
the EC-algorithm is to perform a maximization over \( \Theta^k \), where \( \Theta \) is the parameter space for 
each type. The steps used by the EC-algorithm are as follows:

**Algorithm EC:**

For any given \( \phi = (k'; \theta_1, \ldots, \theta_{k'}) \):

- For each individual \( i \):
  - Calculate \( l f_i(h; \phi) = \log \left( f\left( \{ y_{it}, x_{it} \}_{t=1}^{T_i}; \theta_h \right) \right) \), for \( h \in \{1, \ldots, k'\} \).
  - Choose \( h \in \{1, \ldots, k'\} \) to maximize \( l f_i(h; \phi) \) over \( h \in \{1, \ldots, k'\} \). Call 
    the maximal value \( l f_i(\phi) \). (This corresponds to maximizing over the 
    \( \delta_{ij} \)'s for individual \( i \), conditional on the \( k' \) proposed parameter vectors.)
- Sum the obtained log-likelihoods \( l f_i(\phi) \) over individuals \( i \in \{1, \ldots, n\} \). Call 
  the outcome \( l f(\phi) \).

Choose \( (k'; \theta_1, \ldots, \theta_{k'}) \) to maximize \( [l f(\phi) - \text{penalty}(n, k')] \). (For each \( k' \), use a 
maximization subroutine to maximize over \( (\theta_1, \ldots, \theta_{k'}) \).)

A natural and consistent starting point for the parameter estimation can be constructed 
from individual level estimates. First, estimate a separate function for each individual, and 
then cluster the estimates into \( k \) groups taking the centroids as the initial estimates. One 
can attempt to guard against getting trapped in local maxima by restarting the procedure 
at different starting points obtained by random perturbations on the original point.\(^1\)

The EC estimator is consistent as \( T \uparrow \infty \), but when \( T \) is small, misclassifications of 
subjects will result in a small sample bias in the estimated \( \hat{\theta}_j \)'s. To determine whether this 
bias is sufficiently significant to warrant our attention, we proposed a large-\( T \) approximation 
to the EM algorithm for obtaining posterior probabilities on individual classifications. The 
EM algorithm for this mixture model would start with guesses \( \pi^0_j \) on the proportions of each 
type in the population, and guess \( \theta^0 = (\theta^0_1, \ldots, \theta^0_k) \). Then, for the E-step of the \( r \)th iteration, 
the posteriors on individual classifications would be calculated using Bayes’ rule:

\[ p^r(\{ \delta_{ij} = 1 \}) = p^r_{ij} = \frac{\pi^r_j f(\{ y_{it}, x_{it} \}_{t=1}^{T_i}; \theta^r, \{ \delta_{ij} = 1 \})}{\sum_{k=1}^k \pi^r_k f(\{ y_{it}, x_{it} \}_{t=1}^{T_i}; \theta^r, \{ \delta_{ij} = 1 \})}, \]

\(^1\) This randomized strategy was used in El-Gamal and Grether (1996), but not in the current paper.
\[ \pi_j^r = \sum_{i=1}^{k} p_{ij}^r, \]
and calculate
\[ Q(\theta, \theta^r) = \sum_{i=1}^{n} \sum_{j=1}^{k} p_{ij}^r \log \left[ f(\{y_i, \tilde{x}_{ij}^r\}_{i=1}^{T_i} | \theta^r) \right]. \]

Then, the M-step would maximize \( Q(\theta, \theta^r) \) to produce \( \theta^{r+1} \), and we iterate the E and M-steps until convergence.

A large \( T \) approximation to this EM-algorithm would start with \( \theta^0 \) initialized at the EC estimate \( \hat{\theta}, \ldots, \hat{\theta}_k \), and the M-step would be by-passed in each iteration. The iteration of the E-steps would converge to posteriors on the classifications. If those posteriors are very concentrated on the groups to which the individuals were assigned, then the expected number of possible misclassifications is small, and the resulting small sample bias is also small. The summary statistic we use to quantify the closeness of the \( p_{ij} \)'s to zeros and ones (and thus their closeness to the \( \delta_{ij} \)'s) is called the Average Normalized Entropy (ANE):

\[ \text{ANE}(k) = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} p_{ij} \log_k(p_{ij}), \]

which is zero when the \( p_{ij} \)'s are equal to the \( \delta_{ij} \)'s. ANE always lies between zero and one, with small numbers reflecting crisp classifications and negligible small sample biases, and large ANEs reflecting weak classifications and significant potential biases.

3 The Monte Carlo Probits Experimental Design

In this paper, we implement two experimental designs to study the relative performance of EC estimation and fixed effects estimation in panel models with limited dependent variables. The data generating process in both cases is a simple Probit model:

\[ y_{it}^* = \sum_{j=1}^{2} \delta_{ij}(\alpha_j + \beta_j x_{it}) + \epsilon_{it}, \]

\[ y_{it} = \begin{cases} 1 & \text{if } y_{it}^* > 0, \\ 0 & \text{otherwise}. \end{cases} \]

Where we shall generate \( \epsilon_{it} \sim \text{i.i.d. } N(0,1) \), \( x_{it} \sim \text{i.i.d. } N(0,1) \), and \( \Pr\{\delta_{i1}\} = \Pr\{\delta_{i2}\} = 0.5 \).

In Design I, we shall make the slopes unpoolable, with \( \alpha_1 = 1 \), \( \alpha_2 = -1 \), \( \beta_1 = -1 \), \( \beta_2 = 1 \). We shall then simulate a 1000 data sets with \( n = 20 \), and \( T = 3, 5, 10, 20 \). For each simulated data set, we shall estimate \( (\alpha_1, \beta_1, \alpha_2, \beta_2, \{\delta_{ij}\}) \) using the EC-algorithm. We shall also estimate \( (\{\alpha_i\}_{i=1}^{20}, \beta) \) for the fixed effects model, and - for comparison - \( (\alpha, \beta) \) for a
pooled model. In Design II, we implement the same steps as in Design I, except for making the slopes poolable, by setting $\beta_1 = \beta_2 = 0.5$.

Using the 1000 simulated data sets for each design and each longitudinal sample size, we can construct and compare the distributions of resulting point estimates, and — in the case of the EC-algorithm — misclassifications and ANEs. We can, thus, compare the efficiency gains of allowing $k << n$ in the poolable slopes case, as well as the gains from allowing for unpoolable slopes in Design I.

All of the simulations were conducted on a Sun Ultra2 with two 167MHz CPUs. The simulation code was written in Gauss, using the MaxLik4.0 likelihood maximization subroutine to estimate the pooled probits (2 parameters), the EC probits ($2 \times k$ parameters for $k = 1, 2, 3$), and the fixed effects probits ($n + 1$ parameters). Following the “road-map to the large $T$ approximation” which we introduced in El-Gamal and Grether (1996), we started by estimating $n$ probits (2 parameters each) for the $n$ individuals, as well as one pooled probit (2 parameters) for the entire sample ($k = 1$). For $k = 2, 3$, we then applied a $k$-means clustering algorithm to the $n$ point estimates for the $n$ individuals, using the algorithm of MacQueen (1967). The $k$-centroids so produced were used as an initial condition for the EC multiprobit estimation, and produced reasonably fast convergence. For the fixed effects estimates, we used the individual $\alpha_i$'s from the $n$ individual probits, and the $\beta$ from the pooled probit, to initialize the search for the $(n + 1)$ parameters that maximize the fixed effects probit likelihood function. The average number of seconds used per simulation (for individual estimates, $k$-clustering, the $k = 1, 2, 3$ EC-estimates and the fixed effects estimates) ranged from approximately 9 seconds for $T = 3$ to approximately 30 seconds for $T = 20$.

4 Monte Carlo Results

For each of the two designs (unpoolable, and poolable), we produce ten graphs. The first two graphs show the densities of the number of misclassifications and the ANE statistics, respectively, in the 1000 simulations each for $T = 3, 5, 10, 20$. The next eight graphs per design are in four pairs for $T = 3, 5, 10, 20$. For each value of $T$, the pair of graphs show the densities of $\alpha$ and $\beta$ estimates over the 1000 simulations each, for the pooled estimates ($k = 1$), and the EC estimates ($k = 2$), as well as the fixed effects estimates. All densities are smoothed with a Gaussian kernel, and Silverman’s rule of thumb bandwidth. In the remainder of this section, we shall bring to the reader’s attention some of the more interesting features of the 10 plots for each design. It is our view, however, that “a picture is worth a 1000 words”, and ocular examination of the smoothed densities is the best way to understand the behavior of our EC estimator relative to the fixed effects approach.

---

2 We note that for sufficiently small sample sizes, the $y_{it}$'s may be perfectly explained using a linear combination of the $x_{it}$'s for some individuals. In such cases, the parameters for such individuals are identifiable only up to scale. In our Monte Carlo analysis, we applied a uniform rescaling of all individual probit estimates in such cases. We note that this is only used for finding reasonable (consistent) starting positions for the EC maximization, which identifies the parameters by pooling the data of all individuals within each type.
4.1 Design I: Unpoolable Slopes

Figures 1 and 2 illustrate the speed with which the proportion of individuals misclassified by EC declines as T increases. Table 1 provides some summary statistics on the means and standard deviations of the number of misclassifications (out of 20 individuals), and the ANE statistic. It is clear that the proportion misclassified, and the ANEs move together as predicted; hence reinforcing our recommendation to use the ANE as a measure of the degree of misclassification (and small sample estimation bias). We note that for T’s of modest size, the proportion misclassified is below 5%, as reflected also by the ANE of 0.35. This suggests that T values similar to those in Heckman and MaCurdy (1980) may be sufficient for EC to perform very well, whereas we shall see in later discussion that fixed effects estimators perform rather poorly for this value of T. For T larger than 10, misclassifications are extremely rare, and the potential small sample estimation bias is negligible. This is in accordance with our results in El-Gamal and Grether (1995), El-Gamal and Grether (1996), where it was found with T=14 to 19 that the ANEs were reasonably small.

<table>
<thead>
<tr>
<th>T</th>
<th>μ(mis)</th>
<th>σ(mis)</th>
<th>μ(ANE)</th>
<th>σ(ANE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.19</td>
<td>2.39</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>5</td>
<td>0.79</td>
<td>1.4</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>10</td>
<td>0.07</td>
<td>0.43</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>20</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 2 shows the distribution of estimates of k (the number of types) as we vary T. It is clear in this table that the vast proportion of cases lead to correct estimation of k = 2, with the proportion starting above half, and growing quickly as T gets larger.

<table>
<thead>
<tr>
<th>T</th>
<th>4.8</th>
<th>63.1</th>
<th>32.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.0</td>
<td>81.9</td>
<td>18.1</td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>88.7</td>
<td>11.3</td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>92.5</td>
<td>7.5</td>
</tr>
<tr>
<td>20</td>
<td>0.0</td>
<td>95.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Figures 3 through 10 illustrate the improvement in EC and fixed effects estimation. They show the rapid improvement in EC estimation, with modest biases for T = 3, 5, which are - however - much smaller than the corresponding biases in fixed effects estimates. For T = 10, 20, the EC estimators are performing extremely well, whereas the fixed effects estimates still exhibit a strong inward bias, and much larger variance, due to their inability to pool individuals with similar behavior.
Figure 1: Densities of misclassifications for $n=20$, $T=3,5,10,20$

Figure 2: Densities of ANE for $n=20$, $T=3,5,10,20$
Figure 3: $\hat{\alpha}$'s $T=3$, $n=20$, unpoolable $\beta$'s

Figure 4: $\hat{\beta}$'s, $T=3$, $n=20$, unpoolable $\beta$'s

Figure 5: $\hat{\alpha}$'s $T=5$, $n=20$, unpoolable $\beta$'s

Figure 6: $\hat{\beta}$'s, $T=5$, $n=20$, unpoolable $\beta$'s
Figure 7: $\hat{\alpha}$'s T=10, n=20, unpoolable $\beta$'s

Figure 8: $\hat{\beta}$'s, T=10, n=20, unpoolable $\beta$'s

Figure 9: $\hat{\alpha}$'s T=20, n=20, unpoolable $\beta$'s

Figure 10: $\hat{\beta}$'s, T=20, n=20, unpoolable $\beta$'s
4.2 Design II: Poolable Slopes

Figures 11 and 12 illustrate the speed with which the proportion of individuals misclassified by EC declines as $T$ increases. Table 3 provides some summary statistics on the means and standard deviations of the number of misclassifications (out of 20 individuals), and the ANE statistic. We notice from Table 2 that the proportion of misclassifications, and the ANEs are somewhat higher in this poolable case than they were in the non-poolable case. This is mainly due to the fact that we are not imposing the condition that $\beta_1 = \beta_2$. We note, however, that for $T = 20$, the ANEs are virtually identical with those for the nonpoolable case, and the proportion of misclassified individuals is very insignificant, despite the fact that we do not impose $\beta_1 = \beta_2$.

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\mu$(mis)</th>
<th>$\sigma$(mis)</th>
<th>$\mu$(ANE)</th>
<th>$\sigma$(ANE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.93</td>
<td>2.12</td>
<td>0.20</td>
<td>0.12</td>
</tr>
<tr>
<td>5</td>
<td>1.65</td>
<td>1.94</td>
<td>0.13</td>
<td>0.09</td>
</tr>
<tr>
<td>10</td>
<td>0.36</td>
<td>1.02</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>20</td>
<td>0.04</td>
<td>0.52</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 4 is the analog of Table 2, showing the behavior of $\hat{k}$ as $T$ varies, for the case of $\beta_1 = \beta_2$. It can readily be seen by comparing Tables 2 and 4 that there is a tendency in the poolable case to estimate $\hat{k} = 2$ correctly somewhat less often. For small values of $T$, both $\hat{k} = 1$ and $\hat{k} > 2$ occur more often in the poolable case, though not significantly so. In general, the main pattern in Tables 2 and 4 is the fast convergence of almost all simulations to the correct estimation of $\hat{k} = 2$ as $T$ gets larger.

<table>
<thead>
<tr>
<th>$T$</th>
<th>$% \hat{k} = 1$</th>
<th>$% \hat{k} = 2$</th>
<th>$% \hat{k} &gt; 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>16.9</td>
<td>58.2</td>
<td>24.9</td>
</tr>
<tr>
<td>5</td>
<td>1.6</td>
<td>75.0</td>
<td>23.4</td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>85.4</td>
<td>14.6</td>
</tr>
<tr>
<td>20</td>
<td>0.0</td>
<td>89.7</td>
<td>10.3</td>
</tr>
</tbody>
</table>

Figures 13 through 20 show that for $T = 3, 5$, the EC estimators still do significantly better on estimating $\alpha_1$ and $\alpha_2$, both in terms of bias and variance, than the fixed effects model. Despite the fact that the fixed effects model imposes the restriction that $\beta_1 = \beta_2$, its performance in estimating $\beta$ is not significantly better than the EC estimators, even for $T$ as small as 3 and 5. With $T = 10, 20$, the EC estimators are clearly superior to the fixed effects estimators for $\alpha$, and the performance in estimating $\beta$ is virtually identical in terms of bias and variance. Hence, even with the handicap of not imposing $\beta_1 = \beta_2$, the EC estimator recovers quickly for reasonable values of $T$. Needless to say, larger values of $n$ can only improve the behavior of the EC estimate once $T$ reaches this sufficiently large range (due to pooling more data within each group, thus reducing the estimation error).
Figure 11: Densities of misclassifications for $n=20$, $T=3,5,10,20$, poolable slopes

Figure 12: Densities of ANE for $n=20$, $T=3,5,10,20$, poolable slopes
Figure 13: $\hat{\alpha}$'s $T=3$, $n=20$, poolable $\beta$'s

Figure 14: $\hat{\beta}$'s, $T=3$, $n=20$, poolable $\beta$'s

Figure 15: $\hat{\alpha}$'s $T=5$, $n=20$, poolable $\beta$'s

Figure 16: $\hat{\beta}$'s, $T=5$, $n=20$, poolable $\beta$'s
Figure 17: $\hat{\alpha}$'s, $T = 10$, $n = 20$, poolable $\beta$'s

Figure 18: $\hat{\beta}$'s, $T = 10$, $n = 20$, poolable $\beta$'s

Figure 19: $\hat{\alpha}$'s, $T = 20$, $n = 20$, poolable $\beta$'s

Figure 20: $\hat{\beta}$'s, $T = 20$, $n = 20$, poolable $\beta$'s
5 Concluding Remarks

The calculations exhibited in the previous section illustrate that the EC algorithm is not only feasible, but performs well in panels with values of \( T \) well within the range of most panel data sets. This is encouraging as there are many situations in which the basic idea underlying the method is directly applicable. That is, the observed individuals may each belong to one of a fixed but possibly unknown number of types. Examples include consumers adopting different shopping strategies, countries following different kinds of development paths, workers responding to intermittent spells of unemployment and, as in our original application, people making inference judgements.

The statistical theory of the EC estimator is based on asymptotic approximations as \( T \) goes to infinity with \( n \) going to infinity or remaining constant. As pointed out earlier in the paper it does not matter whether \( T \) is large relative to \( n \) or not. What is crucial is that \( T \) be sufficiently large for the classification of the units to types to be made with sufficient precision. If such precise classification is possible, small sample biases arising from miscategorisation can be ignored safely.

Given this asymptotic approximation, the econometrician needs a means to determine if the sample sizes available are sufficiently large. We suggest that once the individuals have been classified into types, the researcher may calculate the posterior probability that each individual unit belongs to each of the active types using our large-\( T \) approximation to the EM algorithm. If these probabilities are close to zeroes and ones, then the data are strongly in agreement with the assignment of individuals to types. If the classifications are not crisp, then one may wish to study the mixture of the types using the EM-algorithm, or consider an entirely different approach (e.g. random effects models) to the data. We have introduced a diagnostic statistic ANE designed to aid the researcher in judging the crispness of the classifications. The performance in this study suggests that the measure is indeed useful in identifying small samples for which misclassifications may result in biases in parameter estimates.
References


