Comparison of the Fixed Effects estimator and the Mundlak mixed effects estimator

Fixed effects (FE) and random effects (RE) models refer to models estimated with all observations in groups represented by dummy variables or correlated disturbances. Mixed effect (ME) models are alternative parameterizations of the same model. They decompose variation in X_{ij} into within group variability and between group variability while preserving the benefits of FE. The standard FE estimator does this either by decomposing the between group variability into dummy variables (FE_1) or by removing that variability from the residuals by subtracting the mean residual within groups (FE_2) and modeling the remaining variance with a set of $X_i's$ which are deviations from means within groups. A random effects model under the Mundlak formulation, also called a contextual model in the multilevel literature, decomposes the variability by splitting X_{ij} into two sets of variables: means $\beta\bar{X}_j$ and deviations $\beta(X_{ij} - \bar{X}_j)$.

There are thus three basic ways to reorganize between group variability: dummy variables—FE_1, decomposition into within and between versions of X_{ij}—ME, and differencing the between variance out of the data set completely—FE_2. The choice between the FE estimators and the ME estimator is based on how the researcher wants to interpret between group variability and whether the particular approach is statistically or numerically feasible.

- If the researcher wants to rank order groups then the dummy variable approach is better.
- If the researcher wants to ignore all between group variability because it is not of substantive interest then purging it from the data is better.
- If the researcher is so paralyzed by fear of that omitted variable bias at the group level would render any between group comparisons automatically suspect then simply purging it from the data set is better. This is an extreme position which was not apparent in the early literature of panel data but is now accepted practice in some fields, but omitted variable bias is a potential problem for everything in econometrics.
- If the researcher wants to test group-based hypotheses based on substantive explanatory variables then the ME version is better.

Note that so far the discussion concerns differences in modeling between group differences. That's because within-group differences and coefficients and standard errors on the FE estimators and the within-group varying variables in the ME should be very nearly identical assuming some basic conditions. All that is substantively different about the models is how to capture and represent between group variability.

In practice the coefficients and standard errors will rarely be absolutely identical because of rounding error from the estimation but the differences should be trivial. More than trivial differences between the FE estimators and within group varying variables of the mixed effects estimator are a signal that 1) the relationships among groups and variables are more complex than the current model allows; 2) there is serious measurement error 3) the model is biased by endogeneity in the estimation of either the fixed effects or the random effects 4) the estimation technique used to produce fixed or random effects produced biased estimates. Note that any or all of these problems can happen at once and result in major or trivial differences.

Measurement Error

Measurement error of $\bar{X}_j$ is always possible. When it happens with the dummy variable based FE estimator it can result in dummy coefficients of the wrong size, sign, or statistical significance. When it occurs in the ME model it can mean that the actual random effect might be biased. In either case it also means that the beta coefficients are likely to be biased for everything in the model related to the measurement error. This is important in understanding the underlying endogeneity problem.

$\bar{X}_j$ can be calculated in two general ways:

- If $\bar{X}_j$ is built by averaging over X_{ij} and there is measurement error in some observations of X_{ij} then the $\bar{X}_j$ is biased by small or large amount.
    - So is everything else based on the biased data. The random effect is biased and the within group $\beta's$ are biased because they are all based on an error in the data. In this case any model is biased, FE, RE, or ME.
- If $\bar{X}_j$ is built from a different data source and it has measurement error or the errors are uncorrelated with the errors in X_{ij} then it will possibility result in endogeneity. An example here would be using American Community Survey (census) data on state education levels and individual-level data on personal educational attainment from a typical survey. Both sources have some measurement error but as their errors are unrelated there might be some remaining endogeneity of lower level education and the random effect.
    - Note that this can be true even if the version of $\bar{X}_j$ from a different data source is more accurate! Variance and bias are separate issues.

Differences in the coefficients and standard errors from one of the FE models and the ME model could be the result of $\bar{X}_j$ being a *weak instrument* (weakly correlated with the badly measured explanatory variable) because of measurement error. If there is measurement error in the main data source (X_{ij}) then it will bias the fixed effects (because the values of X's are wrong), the random effects (because the expected values of the groups will be wrong), and $\bar{X}_j$ (because the values used to make them are wrong). The familiar endogeneity problem exists between the biased version of (X_{ij}) and the biased random effect. The biased version of $\bar{X}_j$ is needed to properly instrument the endogeneity. By using a more accurate (less biased) version of $\bar{X}_j$ the effectiveness of the instrument can be reduced. This is true if the $\bar{X}_j$ alternative source is more accurate, less accurate, or equally accurate but biased in some different pattern or direction.

A tradeoff might exist between an exogenous within group variable based on accurate data and weakly correlated with the measurement error and a badly measured $\bar{X}_j$ which is a better instrument for testing a substantive hypothesis. One way to mitigate this problem by taking the difference of the two versions of $\bar{X}_j$ and including it in the model as a separate instrument! This problem arises only when there is a lot of measurement error.

Parametrization Error

Dummy variables and using the ME estimator are different ways to parametrize between group variability. The main substantive difference is in the interpretation of the between group information (e.g. as group intercepts or as substantive variables). However, the equality of these two parametrizations is based on the underlying assumption that $\bar{X}_j$ is a good representation of between group variability related to X_{ij}.

In a properly specified model it almost always should be. In an improperly specified model it might not be. Consider a model that predicts voting based on age in different states. Under the FE_1 estimator the model would be:

$$\text{Y(voting)} = \alpha + \beta(age)_i + \beta(state\ intercepts)_j + \varepsilon$$

Under the ME estimator the model would be:

$$\text{Y(voting)} = \alpha + \beta(age)_i + \beta(age - your\ state's\ average\ age)_j + \mu + \varepsilon$$

State intercepts (dummies) are essentially the most flexible way of parametrizing the between effects here. Using the state average may not properly capture the between group differences if age has a nonlinear effect. Since the effect of age on voting changes as a person gets older for demographic, financial, and medical reasons the effect is typically modeled as *age* and age^2.

Error can arise from attempting to parametrize the distribution of the X with a mean instead of splines, percentiles, or a median or from not taking the average of each in a set of splines or polynomials of X_{ij}. If deviances around the group mean are random and approximately symmetrical then the mean does an excellent job of capturing the central tendency. However, if the within group variation is clustered along two modes or is seriously skewed then the mean will do a poor job of capturing between group variability.

In other words, if a variable appears to have a nonlinear effect when it should have a linear effect then the traditional Mundlak specification might perform poorly for that variable. However, the estimated coefficient for that variable would be uninformative anyway because of the misspecification. That is a signal to check for linearity conditions in the model and examine the specification of variables. The effect on the FE model differs from the effect on the ME model, so the difference between the two as a diagnostic for a deeper problem.

Statistical/Estimation Error

Thus far, the problems discussed imply that when there is a difference between an FE estimator and the ME estimator the ME estimator is probably more wrong. In other words, that the relevant coefficients and standard errors from the FE estimator are probably more accurate (though that doesn't mean that they are actually totally accurate). While that is probably true in linear models it is not the case in nonlinear models. In nonlinear models the FE_1 (or dummy variable) estimator is often problematic because of the incidental parameters problem whereby additional dummy variables can induce bias in all model parameters. That still leaves it as an improvement on the FE_2 (or differencing) method since that approach is impossible in most nonlinear cases. One alternative is a higher order Taylor Series in the otherwise linear probability model which can be estimated by first regressing to obtain fitted values then estimating a polynomial in the fitted values. However, this is often not done by most applied researchers.

The incidental parameter problem happens with dummy variables in a nonlinear maximum likelihood problem. By increasing the dimensionality (every dummy variable adds another dimension) of the PDF or CDF it becomes more difficult to solve the underlying calculus problem and more likely to return biased coefficients and standard errors for some or all parameters and not just the dummies. The less linear the data (with binary problems being the worst) the more likely the incidental parameters bias is to occur.

Because random effects aren't estimated from dummies in nonlinear models they don't add unnecessary extra dimensions to the problem and thus do not suffer from incidental parameters bias. If the random effect isn't calculated properly then it could still cause the coefficients and standard errors to be biased but that is an issue that can be solved by picking the right numerical technique. It is especially important to know what Stata is doing to create random effects in nonlinear models to avoid these problems.

In this case, using the ME model and comparing it to FE_1 is actually a way to test for incidental parameters bias assuming there is no measurement error or parametrization error. Major differences in results between the two methods likely means the FE method is more wrong (though the ME method can still be biased). Therefore, any differences should be taken as a sign of bias in the FE model.

The ME method is also just a different way of calculating fixed effects than the typical FE researcher uses. It is automatically easier in a nonlinear setting than other standard approaches to calculating fixed effects and will allow the researcher to use the FE estimator to placate an economist! In that case, the researcher can just pretend the $\bar{X}_j$'s aren't there and focus on the within estimates. The expected value of each group from the random effect estimates the dummy variable coefficients from the ME model.

The point of all of this is that the differences between fixed effects from the FE and ME models are a quick diagnostic for other problems. If they are seriously different then there is a reason that is likely causing other problems with the results.