A Random-effects construction of EMV models - a solution to the Identification problem?

Peter E Clarke, Deva Statistical Consulting

Credit Scoring and Credit Control Conference XV
Edinburgh, August 2017

Introduction

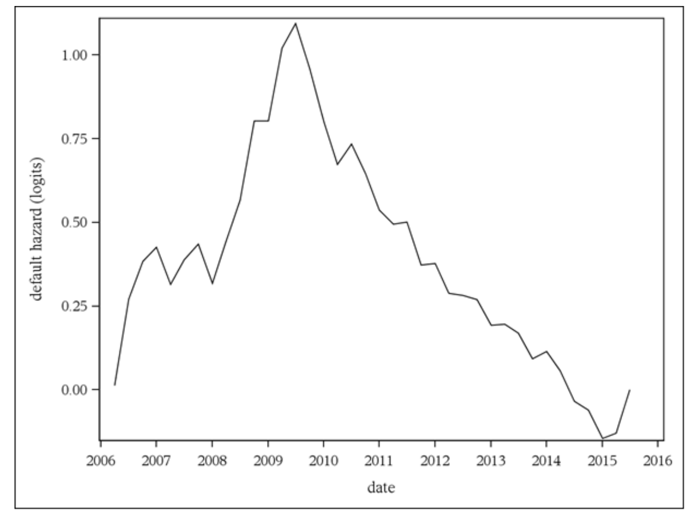
Purpose of presentation:

- Briefly describe EMV models and their attendant identification problem
- Describe the motivation for using random instead of fixed effects in mixed effects/multilevel modelling.
- Compare the results from mixed models to those of constrained fixed effects and intrinsic estimator alternatives in terms of recovering correct EMV profiles
- Show what ME models can do that these others can't

EMV

An EMV model is a class of discrete-time survival analysis that seeks to decompose a panel of time-series into three components:

- *Exogenous* – the external factors that affect the outcome, namely the economic cycle. The exogenous effect is indexed by calendar date.
- *Maturity* – the effect of time on book on the outcome. Commonly indexed by number of quarters on book.
- *Vintage* – the inherent quality of a cohort of accounts that originate at the same time. The vintage effect is indexed by origination date.



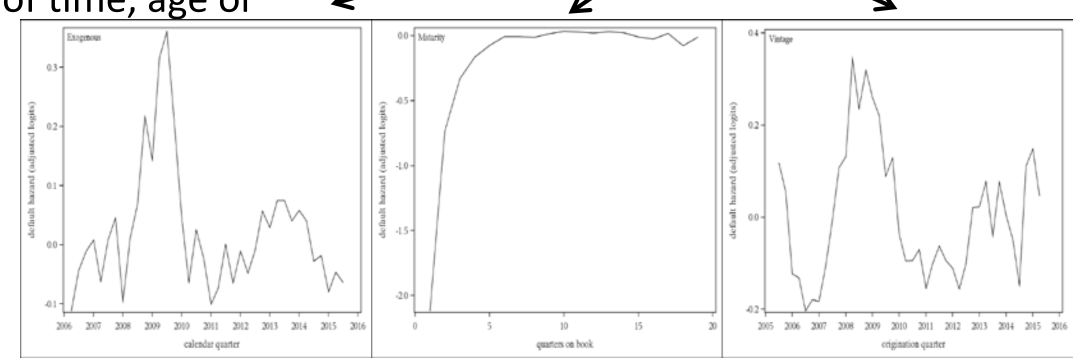
Unstructured model:

$$g(Y_{tav}) = \beta_0 + \beta_t^E + \beta_a^M + \beta_v^V$$

‘intercept’ parameter for each distinct value of time, age or vintage

Need estimators to do derive betas:

- Fixed Effect model
- Intrinsic Estimator
- Mixed/Multilevel model



Identification problem

From the origination date and maturity then calendar date is known:

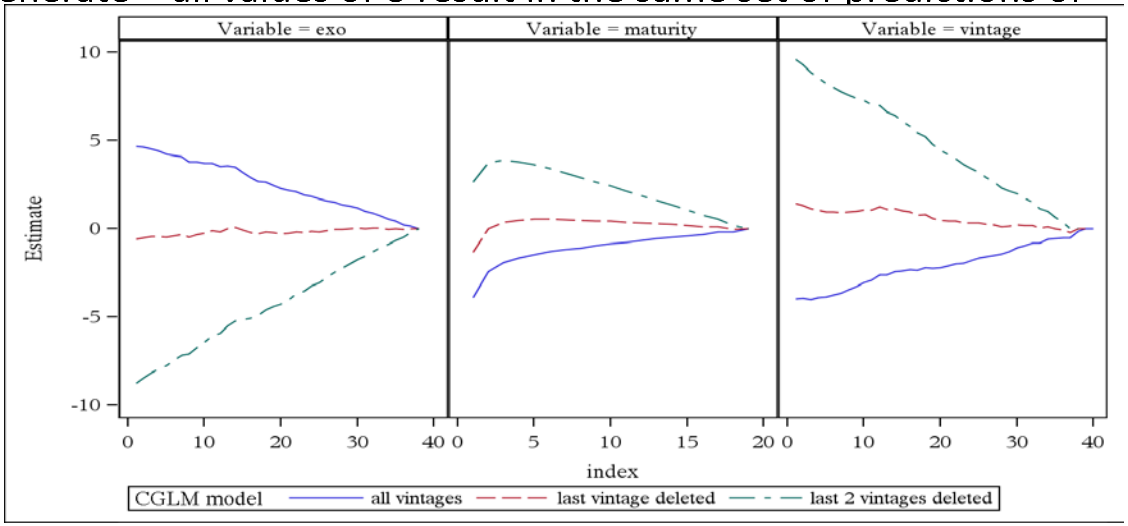
$$t = v + a$$

We can assemble a “null vector”, $\beta^{null} = (1,2.. k^E, -1,-2.. k^M, -1,-2.. k^V)$ such that $X\beta^{null} = 0$

$Y = X\beta^{EMV} = X\beta^{EMV} + \boldsymbol{\delta} X\beta^{null} = X(\beta^{EMV} + \boldsymbol{\delta}\beta^{null}) = X\beta^{*}$

Solution set for β^{EMV} is therefore degenerate – all values of $\boldsymbol{\delta}$ result in the same set of predictions of Y “Identification problem”.

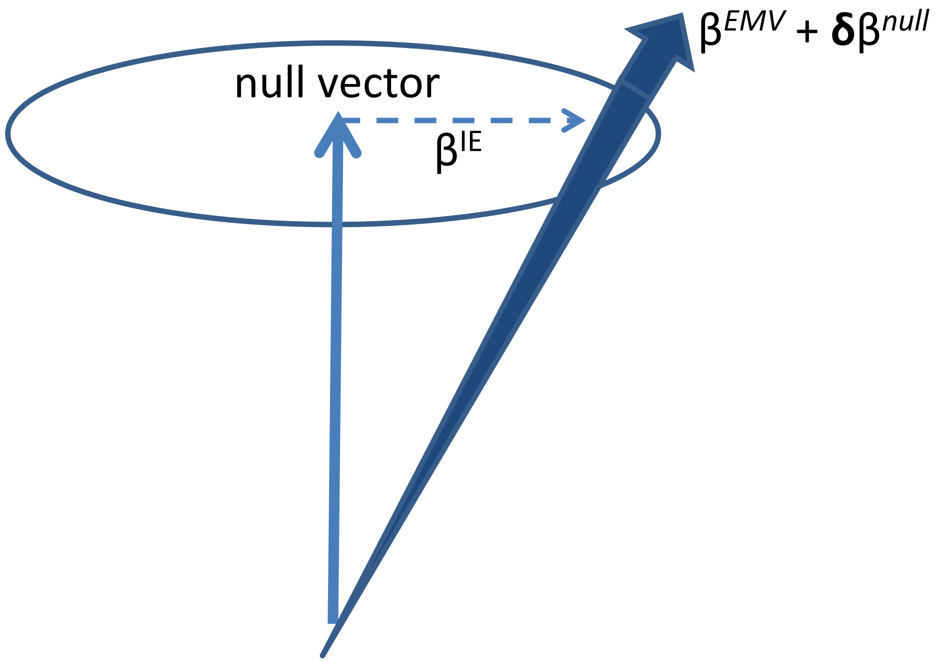
To obtain β^{*} we must apply a constraint on one of the regression parameters (“constrained GLM”: CGLM).
Typical constraint on penultimate vintage parameter: $\beta_{k-1}^{V} = 0$
Result is an artificial trend in each EMV series, +ve for exo and –ve for maturity and vintage (or vice-versa)



Trends are ***unstable*** – change when new vintages are added or removed.

Intrinsic Estimator

- The intrinsic estimator also applies a constraint to the parameter space in order to obtain estimates.
- IE identifies the part of the parameter vector obtained from CGLM which is aligned with the null vector, and strips it out. The process can be thought of as identifying where the vector of degenerate parameter solutions (β^{EMV} + $\boldsymbol{\delta}\beta^{null}$) intersects the plane at right-angles to the null vector



How does the ME estimator work

- Random effects estimators sometimes called "shrinkage" estimators, as they shrink fixed effects ("no pooling") estimates towards the common mean ("pooled" effect). Hence "partial pooling".
- Degree of shrinkage depends on the ratio of the group effect variance (between groups) to the data-level variance (within groups). A large ratio means the between group effect dominates ("no pooling").
- Statistically: $\hat{a}_j = \dfrac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\mu_\alpha}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$
- In words, when we have little evidence for an effect (i.e. the between group variance, σ_α^2, is small and volumes within group "j"(n_j) are low), random estimation of the group effect ($\hat{a}_j$) will tend to substitute the overall mean (μ_α), whereas a fixed effect model will assign a best estimate from the data ($\bar{y}_j$), even though this is likely to be unstable. When we have many observations within each group, random and fixed effect estimates converge on the same value.
- Note: its not about considerations of whether an effect is controlled or uncontrolled, although these considerations impact prediction.

ME model

Generally express vintage and exogenous terms as random effects, and maturity as fixed, hence a "mixed effect" model. For a single observation of unit "i" with age "a", the (generalized linear) model is:

$g(y_i) = \beta_0 + \beta_a^M + \beta_{i[t]} + \beta_{i[v]} + \varepsilon_i$

Where:

$\varepsilon_i \sim N(0, \sigma_\varepsilon)$

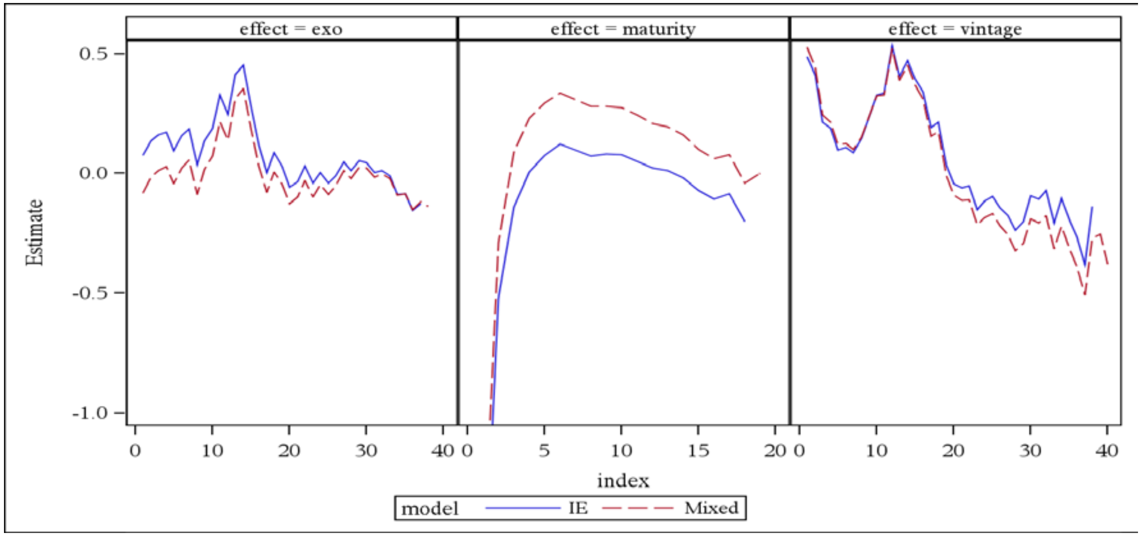
$\beta_{i[t]} \sim N(0,\sigma_E)$

$\beta_{i[v]} \sim N(0, \sigma_V)$

g(..) is the link function

Soft constraint (estimates are from a normal distribution)

Estimates look similar to IE – but which is better?



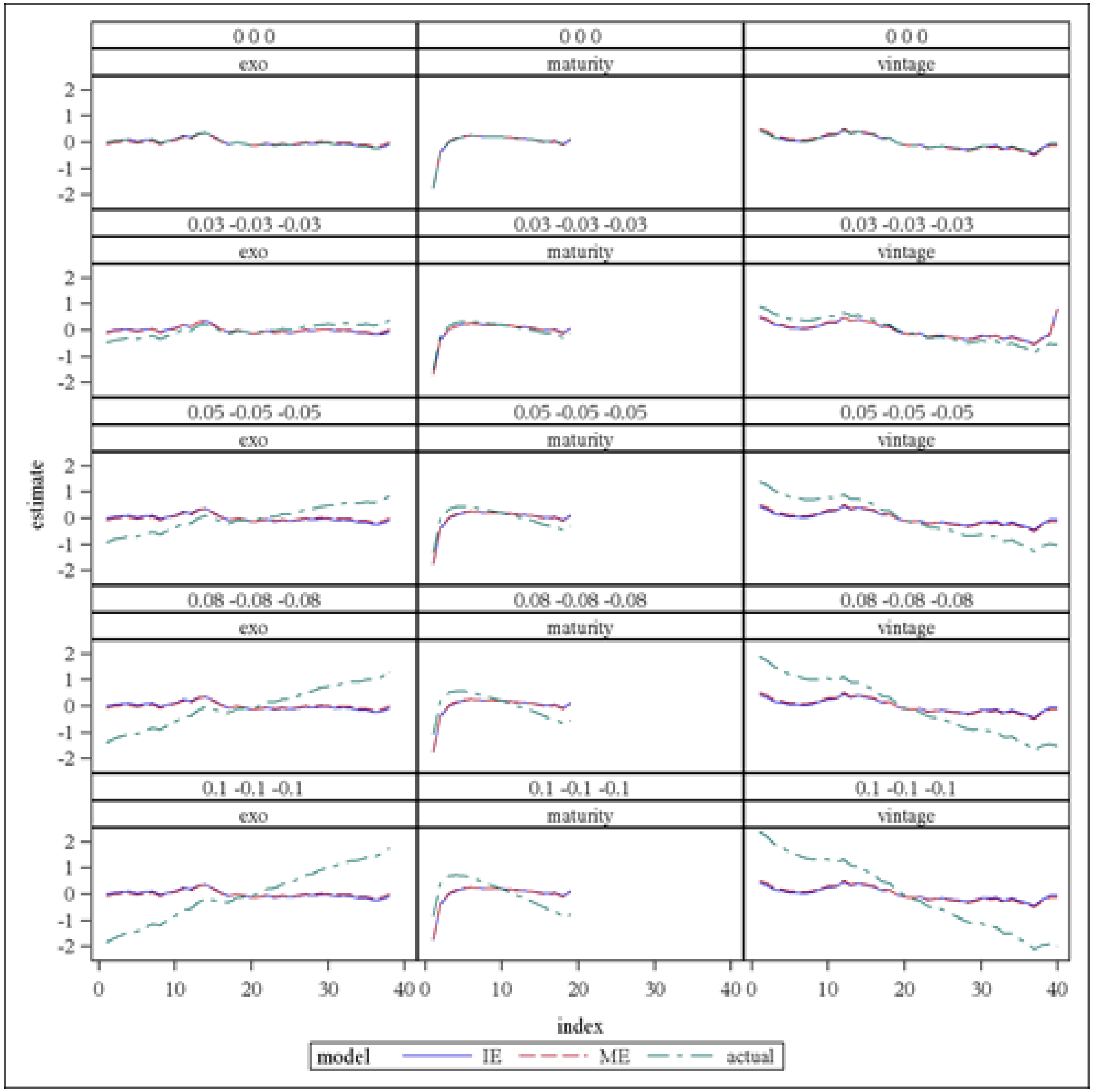
Which is better: ME or IE?

- Results from several portfolios show that IE and ME are both much better than CGLM at recovering stable EMV components.
- For loans the EMV components look similar under IE and ME – don’t know which is better because we do not know the true profiles.
- However, if we simulate a dataset using our own profiles, then we do know what the correct profile looks like.
- Don’t need to vary the short-term non-linear features of each component, as all methods should recover these well.
- Only need to vary the long-term linear component. We do this by rotation (“tilting”) the original IE components : adding a small positive or negative slope to each component.
- Vary the amount of rotation systematically. Each set of rotations (one for each EMV component) then used to generate a set of data.
- Each dataset analysed with fixed (CGLM and IE) and ME models. The slope of the linear, long term element of each component summarizes the performance of the model. We compare this to the slope of the true component.
- Obviously impossible for any estimator to obtain true estimates, as solution is degenerate

Special case: null rotations

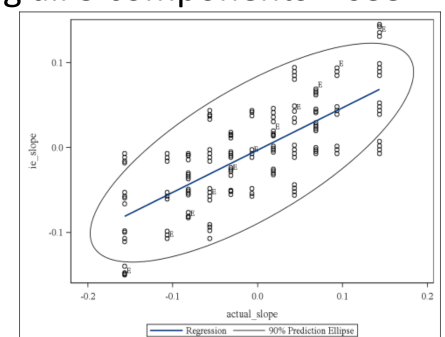
One particular subset of rotations of interest are “null rotations”: because these rotations are coincident with the null vector they result in the same simulated data as the original. A selection of these is shown below.

It goes without saying that no estimation procedure can recover the exact profiles when two profiles generate the same data.

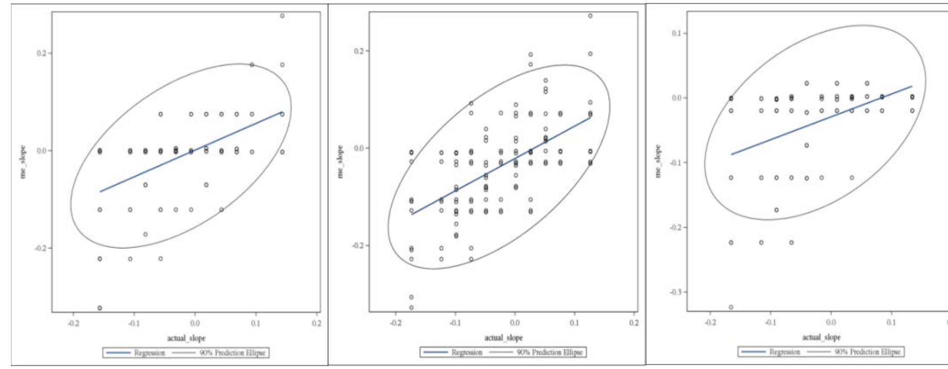


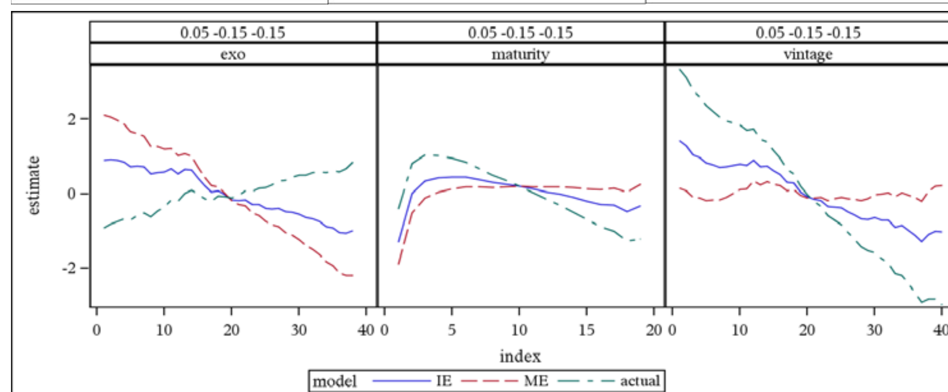
Findings

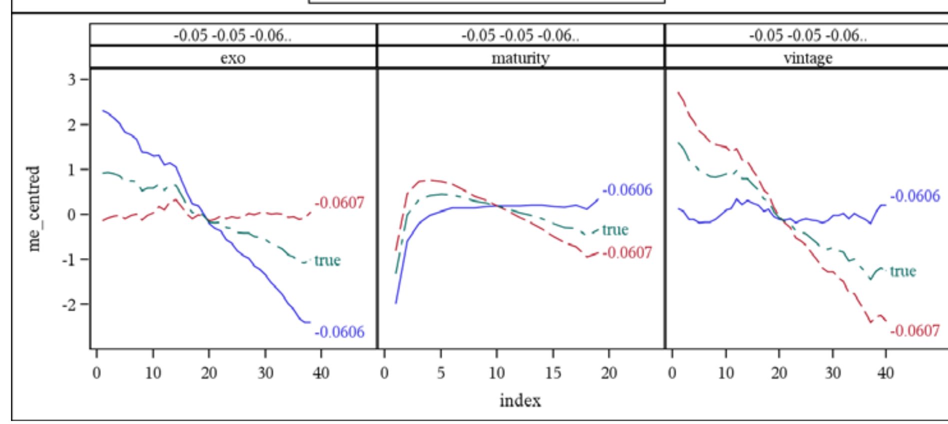
- The non-linear patterns are always recovered successfully, but the linear trends not so successfully – the greater the “net” trend, the worse the error in the slope. Top row shows the trend in each EMV component from the ME model vs actual trend. Agreement is much better than CGLM, but worse than IE (IE performs better for 72% of simulations).
- There are major differences in between IE and ME in terms of recovery of the linear trend:
- IE divides net slope among all 3 components – see blue curves in the middle row of charts. Hence a rotation set with all true slopes equal is correctly modelled by IE (indicated by ‘E’ symbol in plot of IE slope vs actual for exo).



- ME concentrates slope in one component – see red curves in the middle row. Maturity and vintage are quite flat, exo has extreme slope. Whether it is exo or vintage component that acquires the extreme slope is fluid: with a rotation of the vintage component of -0.606, it is exo that is highly sloped. Increase the rotation to -0.607 and now it is the vintage component (bottom row).





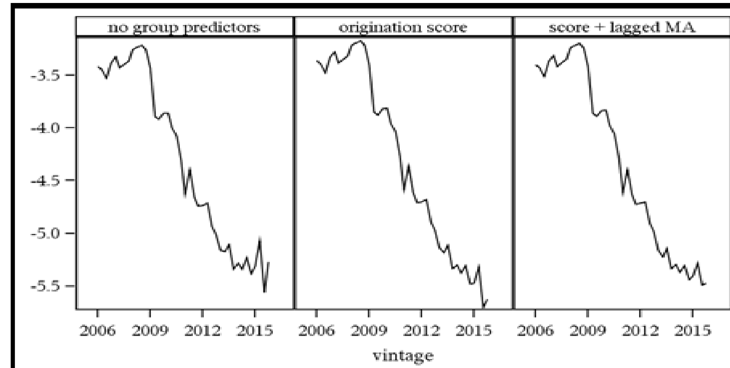


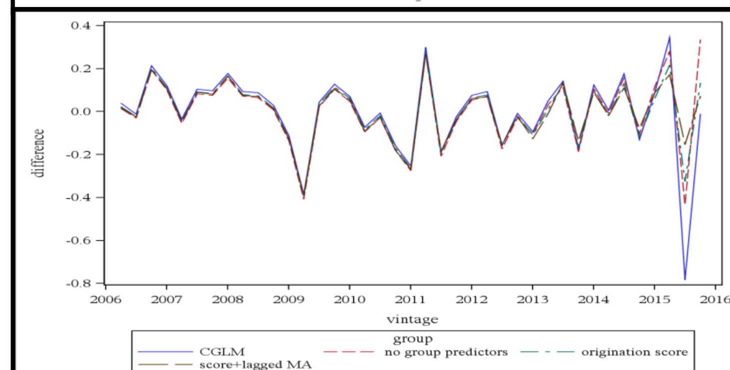
Mixed Effect Models – Part 2

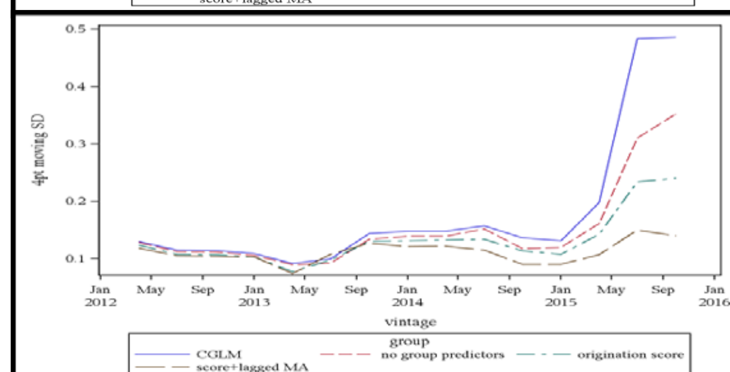
- Mixed effect models offer far richer modelling opportunities than simply extracting "true" EMV components.
- Indeed, conventional "data level" fixed effects models can be regarded as a special case of mixed effects models.
- Mixed effects models build relationships at multiple levels, from the "data level" up through hierarchies of groupings, eg customer – region – country.
- Each grouping is defined by its variance contribution (or covariance for special cases) and a model for its expectation in terms of "hyperparameters", for instance we might utilize regional unemployment rates in a hierarchical stress testing model.
- Groupings above the data level need not be hierarchical, for instance socioeconomic and geographic factors that help determine propensity. In EMV models, exogenous and vintage effects are non-hierarchical "random" effects.
- The examples here enrich the EMV model with group level predictors of exogenous and vintage components.

Improving the stability of the Vintage effect

- Getting vintage effect right is crucial for IFRS9 forecasting.
- One of the biggest challenges we have is that we know least about the most important vintages – ie those originating within the last year.
- Raw EMV model can only use outcomes to estimate vintage: estimates for recent vintages are imprecise because we have few events to work from. Multilevel EMV can use other predictors.
- This is illustrated in the top figure for a mortgage EMV, which shows the vintage effect estimated by a multilevel model without group predictors (LHS), including average origination score of each vintage as a group level predictor (middle), using origination score + autoregressive component (lagged moving average). The trends become smoother as we move from left to right.
- The middle figure overlays the <u>differenced</u> series of these three models, and adds the CGLM estimates. Differencing removes the long-term trend. We can see that for most of the vintages there is no dissimilarity between models, but as we process more recent vintages the quarter-on-quarter variation becomes lower for multilevel models in general, and for models that include group-level predictors in particular (this is the shrinkage effect)
- The bottom figure shows the moving standard deviation (assessed over a 4-point window) of vintage effect. There is clearly much less variation in the last of the three ME models.







How does ME models allow both level 1 and level 2 descriptors of the same factor?

- We can think of the group-level model as describing a prior distribution of the factor's effect.
- The likelihood of the vintage effect is the fixed effect for the group, typically the group mean or adjusted mean, distributed over a confidence interval
- Together, the prior and likelihood form the posterior probability of the random effect
- The better the group level model ("prior") the smaller the variance of the group effect (level 2 "between groups" variance) and the smoother trend in the overall (posterior) behaviour – the magic of random effects.
- Where we have little information on a specific group, the prior dominates (eg recent vintages)
- This is a major benefit of fitting group predictors to random effects – they stabilize poorly estimated group effects by substituting portfolio level averages (or estimates) instead.