Design document for Statsmodels MixedLM

The model for observation j in group i is:

$$Y_{ij} = x'_{ij}\beta + z'_{ij}\gamma_i + \sum_{\ell=1}^{d}\sum_{k=1}^{m_{i\ell}} \theta_{i\ell k} v_{ijlk} + \epsilon_{ij}$$

- $x_{ij} \in \mathcal{R}^p$ are covariates for the fixed effects
- $\beta \in \mathcal{R}^p$ is an unknown vector of fixed effects parameters
- $z_{ij} \in \mathcal{R}^q$ are covariates for the random effects
- $\gamma_i \sim N(0, \Psi)$ is a q-dimensional vector of random coefficients
- $\theta_{i\ell k} \sim N(0, \tau_\ell^2)$ is a realization of a variance component

There are d variance components. The variances of these terms are τ_1^2, ..., τ_d^2 (unknown scalar parameters). The ℓ^{th} variance component is realized $m_{i\ell}$ independent times in the i^{th} group. The v_{ijlk} values are scalar covariates.

Correspondence between notation above and Statsmodels source code:

- p above is `k_fe` in the source code
- q above is `k_re` in the source code
- d above is `k_vc` in the source code
- β above is `fe_params` in the source code
- Ψ above is `cov_re` in the source code
- $(\tau_1^2, \ldots, \tau_d^2)$ is `vcomp` in the source code
- x_{ij} above is `exog` in the source code
- z_{ij} above is `exog_re` in the source code
- v_{ijlk} above are stored in `exog_vc` in the source code

Notes

- During the optimization process, the scale parameter σ^2 and the fixed effects parameter β are always profiled out of the likelihood and score.

- The scipy optimizers require the parameters to be packed into a single vector. The packing used by default is $[\text{vec}_{\text{lt}}(\Psi^{1/2}), \text{vcomp}]$, where vec_{lt} vectorizes the lower triangle (including the diagonal) of a square matrix. These are all the parameters of the profile log-likelihoo (since β and σ^2 are profiled out). If `use_sqrt` is `False`, the optimization is performed with respect to $[\text{vec}_{\text{lt}}(\Psi), \text{vcomp}]$.

- The log-likelihood and score function can take as arguments either a MixedLMParams object or a packed vector. If a packed vector, it must contain $[\text{vec}_{\text{lt}}(\Psi^{1/2}), \text{vcomp}]$. If the argument is a MixedLMParams object it is possible to specify a value of β, and the log-likelihood or score is calculated at the specified point. If the profile_fe is set to `True`, whatever β is provided in the MixedLMParams object is replaced with its GLS estimate.

- There are two versions of the score function. One is taken with respect to $[\text{vec}_{\text{lt}}(\Psi), \text{vcomp}]$, the other is taken with respect to $[\text{vec}_{\text{lt}}(\Psi^{1/2}), \text{vcomp}]$. The latter is obtained by applying the standard change of variables technique to the former. Note that the score function of the log-likelihood profiled over β is the same as the un-profiled score, restricted to the profile subset of the parameter space. This is not true of the Hessian, therefore we do not provide a way to calculate the Hessian of the profiled log-likelihood (which is very complicated to obtain). As a result, we cannot use Newton-Raphson type methods for optimization. However we do have the Hessian with respect to the full parameter $[\beta, \text{vec}_{\text{lt}}(\Psi), \text{vcomp}]$ which is used to produce the sampling covariance matrix of the estimates.