Regularization and Classification of Linear Mixed Models via the Elastic Net Penalty with Application to the Good Judgment Project*

Jonathan Sidi[1], Ya'acov Ritov[1] and Lyle Ungar[2]

[1]Department of Statistics, Hebrew University in Jerusalem
[2]Computer and Information Statistics, University of Pennsylvania

August 13, 2017

Corresponding Author: Jonathan Sidi, Email: yoni.sidi@mail.huji.ac.il

Keywords: Linear mixed model; Elastic Net penalty; Variable selection; Probabilistic forecast

*This research was supported by a research contract to the University of Pennsylvania and the University of California from the Intelligence Advanced Research Projects Activity (IARPA) via the Department of Interior National Business Center contract number D11PC20061. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions expressed herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

Abstract

Advances in the field of model selection and prediction via regularization has forged the ability of a variety of disciplines to classify and model large-scale data. Widely used methods which apply penalties in classification are the Least Absolute Shrinkage and Selection Operator (LASSO), the Adaptive LASSO and the Elastic Net. These methods have predominately been used to classify problems of Generalized Linear Models (GLM) in which the dependency of the covariance structure is assumed to be independent. This assumption is not commonly met in practical data and the ability to model such dependencies is integral in fitting the data correctly, such data is modeled using Linear Mixed Models (LMM). Recent research applying LASSO and Adaptive LASSO to LMM's has produced promising initial results of identifying both the random and fixed effects found in data, proving both consistency and an oracle optimality. However, an inherent drawback to those variable selection methods is their performance under high correlation between covariates. To overcome this we introduce the Elastic Net penalty to LMM selection. This penalty has been found to reduce the prediction error in data with high correlation between variables; such a characteristic can be utilized in more complex data designs while optimizing the LMM problem. Findings are tested through simulations and a case study using data accumulated in an longitudinal study where probabilistic forecasts are derived from crowd sentiment. The data structure consists of repeated measures and a large number of fixed and random covariates.

1 Introduction

Generalized Linear Mixed models (GLMM) Breslow and Clayton (1993) have been applied in a variety of fields to study data designs with between-subject variation. Such designs include longitudinal, repeated measures and clustered data and have been studied thoroughly in the low dimension setting, e.g., Bates (2010) and Searle et al. (1992). In these settings the linear predictor contains in addition to fixed effects, found in Generalized Linear Models (GLM), latent random effects which capture the unique design pertaining to the data. These random effects usually are assumed to have a centered parametric distribution belonging to the exponential family.

Advances in the field of model selection and prediction via regularization, using different penalty terms, has forged the ability of a variety of disciplines to classify and model large-scale data. Widely used methods which apply penalties in classification are the Least Absolute Shrinkage and Selection Operator (LASSO), the Adaptive LASSO and the Elastic Net. These methods have predominately been used to classify problems of GLM, Friedman et al. (2010) and Van de Geer (2008), in which the dependency of the covariance structure is assumed to be independent. This assumption in practical data is not commonly met and the ability to model such dependencies is integral in fitting the data correctly, such data is modeled using Linear Mixed Models (LMM) and GLMMs.

Recent research Bondell et al. (2010) a modified Adaptive LASSO (M-ALASSO), smoothly clipped absolute deviation (SCAD) to LMMs and have produced results of identifying both the random and fixed effects found in data, proving both consistency and an oracle optimality. Model selection within the generalized linear mixed models framework has been discussed in Schelldorfer et al. (2011), Fan and Li (2012), Groll and Tutz (2014), Hui et al. (2016b) and Ibrahim et al. (2011). Schelldorfer et al. (2011) and Groll and Tutz (2014) have a drawback that only fixed effects are selected, while Ibrahim et al. (2011) apply either the SCAD or the ALASSO to each effect. Hui et al. (2016b) allow for greater flexibility for different penalty types on the fixed and random effects. It is noteworthy that Ibrahim et al. (2011) tune each penalty term to a different value through the introduction of the IC(q) criterion, a characteristic not found in the other methods.

This paper proposes a new penalty called the linear mixed model Elastic Net, LMMEN, which is better suited for regularization in highly correlated data. The LMMEN allows for regularization of both the sparsity ($\ell 1$ norm) and grouping ($\ell 2$ norm) for the fixed and random effects separately. We believe that this method will allow to better capture the design of real world data when modeling with LMMs. Through simulations and case study we find that the LMMEN out performs comparative methods in three major areas: highly correlated fixed effects data structures, high dimensionality in the fixed effects, i.e. $p >> n$, selection of random effects when the dimension of the covariance matrix is large.

In the following sections of the paper review the basic structure of the Linear Mixed Effects Model 2, the reparameterization of the LMM to allow for penalization of the random effects 3, define the LMMEN penalty 4, prove asymptotic porperties of the penalized model 6, define and analyze siumulations comparing the LMMEN to various methods 7, discuss a case study in which the LMMEN is applied to real data 8 and end with discussions 9

2 Model

The GLMM is defined as having m subjects in the sample. For the ith subject the response variable is denoted as y_{ij} for the jth observation, where $j = 1 \dots n_i$ and let $N = \sum_{i=1}^{m} n_i$. The training data $\mathbf{X}$ can be defined as two groups of covariates: the fixed effects covariates vector denoted as x_{ij} with dimensions $p \times 1$ and the random effects covariates vector denoted as z_{ij} with dimensions $q \times 1$.

y_{ij} are assumed to be conditionally independent given the subject-specific random effects, b_i, with a conditional mean $E[y_{ij}|b_i] = \mu_{ij}$ and a conditional variance $var(y_{ij}|b_i) = \phi\omega_{ij}^{-1}\nu(\mu_{ij})$. Where ϕ is a positive dispersion parameter, ω_{ij} is a pre-specified weight, and $\nu(\cdot)$ is the variance function. The relationship between μ_{ij} to $\mathbf{X}$ is defined as

$$g(\mu_{ij}) = x'_{ij}\beta + z'_{ij}b_i,$$

where $g(\cdot)$ is a strictly increasing link function, β is the fixed effects coefficient vector for x and b_i is the subject-specific random effects for z. y_{ij} are assumed to be independent and of the form $y_{ij}|b_i \sim F_y$ and b_i is assumed to be of the form $b_i \sim F_b$. The distributions F_y and F_b are predominately assumed to be normal, i.e.:

$$\begin{aligned} F_y &\sim N(\mu_{ij}, \phi\omega_{ij}^{-1}\nu(\mu_{ij})) \\ F_b &\sim N(0, D(\psi)), \end{aligned}$$

where ψ is a $c \times 1$ vector of variance components in the covariance matrix of the random effects D. Under the identity link function with normal distribution we define the LMM

$$\begin{aligned} &y_i = x'_{ij}\beta + z'_{ij}b_i + \epsilon_i, \\ &\epsilon_i \sim N(0, \sigma^2 I_{n_i}). \end{aligned} \tag{1}$$

McCulloch et al. (2011) state that distribution specification may be affected by basic characteristics of the random effects distribution, such as dependence on a covariate or the cluster sample size. For example, the mean or variance of F_b depends on a covariate. When the mean of the random effects distribution depends on a covariate, a fundamental relationship is introduced between the covariate and the distribution, potentially creating a serious bias in estimating the form of the relationship between the covariate and the outcome. Heagerty and Kurland (2001) show that the impact of highly unequal variances can lead to substantial bias. Such bias from distribution specification can cause unintended inference when testing between and within cluster covariates.

Schelldorfer et al. (2011) defined the GLMMLASSO, which solves the log likelihood of the LMM problem via a coordinate gradient descent method based on integral approximation (Laplace approximation) and submit the approximated function to numerical optimization. The advantage of integral approximation methods is to provide an actual objective function for optimization, which enables one to perform likelihood ratio tests among nested models and to compute likelihood-based fit statistics. The disadvantage of these methods is the difficulty of accommodating crossed random effects and multiple subject effects, and the inability to accommodate residual effect covariance structures, or even only residual effect over-dispersion. Moreover, the number of random effects should be small for integral approximation methods to be practically feasible. This disadvantage could potentially inhibit the estimation of random effects in a high dimensional data setting. This characteristic is inherint in all the methods that are built to solve the GLMM problem. The penalty term which is used on the approximated likelihood function is the L_1 penalty. The algorithm proposed penalizes only the fixed effects in the model, thereby estimating the parameters $\{\beta, \theta, \phi\}$ and predicting the random effects vector b using those estimates. The size of the tuning parameter is calculated in two steps: first via the AIC criterion to generate a relevant set of variables and secondly via the BIC criterion to select the final set of active fixed effects which is an unbiased estimator of degrees of freedom in linear models. Hui et al. (2016b) use a regularized penalized quasi-likelihood (rPQL) approach, which also approximates the marginal likelihood function, to simultaneously select fixed and random effects. Groll and Tutz (2014) attempt to solve a similar problem, but with the emphasis on the fixed effects selection. They deploy a gradient ascent algorithm in contrast to the coordinate gradient descent method in Schelldorfer et al. (2011).

Fan and Li (2012) introduce a class of variable selection methods for the fixed effects using a penalized profile likelihood, provided that the random effects vector has a nonsingular covariance matrix. This penalized profile likelihood is equivalent to the penalized quadratic loss function of fixed effects readily found in penalized least squared methods, such as LARS Efron et al. (2004). Random effects are selected under the constraint that the dimension of the fixed effects is smaller than the sample size. They describe an iterative solution for high dimensionality of both the fixed and random effect by which of selecting the fixed effects using the penalized least squares by ignoring all random effects to reduce the number of fixed effects to below sample size. Then in the second step, with the selected fixed effects, they select random effects and finally using the selected random effects refine the fixed effects selections.

Bondell et al. (2010) apply linearization (Taylor expansion) to solve the LMM which is more aptly suited in models with correlated errors, a large number of random effects, crossed random effects, and multiple types of subjects. The disadvantages of this approach include the absence of a true objective function for the overall optimization process and potentially biased estimates. The likelihood function is reparameterized via a modified Cholesky decomposition of the random effects covariance structure Chen and Dunson (2003). This augmentation allows for penalties on both the fixed and random effects. The penalty used in the optimization is the Adaptive Lasso, Zou (2006), which allows for large amount of shrinkage applied to the zero-coefficients while smaller amounts are used for the non-zero ones which then results in an estimator with improved efficiency and selection properties. The level of the tuning parameter is calculated using the BIC criterion.

The regularization penalty method we propose called the linear mixed model elastic net (LMMEN) will extend the regularization characterstics of Bondell et al. (2010) and Hui et al. (2016b). A shared characteristic between the methods is the simultaneous selection fixed and random effects through penalizing the fixed and random effects seperately. Our extension allows for greater flexibility by tuning multiple regularization parameters simultaneously, instead of tuning a single penalty parameter value for both effects. We will apply the Elastic Net penalty on both the fixed and random effects estimates to allow for improved peformance when there is a high level of correlation among the fixed and random covariates.

The following table compares the different methods discussed and used in the simulations to compare the proposed method. We focus on the different type of penalties each method uses on the fixed and random effects, the criteria used to tune the regularization parameters, what type of approximations are used on the marginal likelihood function and if there is an R package that accompanies the method.

Model	Research	FE Penalty	RE Penalty	Tuning Criterion	LogLik Approx.	R Package
LMM	This Paper	Elastic Net	Elastic Net	BIC	None	lmmen
	Bondell et al. (2010)	M-ALASSO	M-ALASSO	BIC	None	None
GLMM	Groll and Tutz (2014)	LASSO	None	BIC	Laplace	glmmLasso
	Schelldorfer et al. (2011)	LASSO	None	AIC+BIC	Laplace	lmmlasso
	Hui et al. (2016b)	LASSO ALASSO SCAD	group LASSO group ALASSO group SCAD	BIC/IC(q)	PQL	rpql
	Ibrahim et al. (2011)	SCAD ALASSO	group SCAD group ALASSO	IC(q)	Laplace	None
	Fan and Li (2012)	SCAD	group SCAD	BIC	Local Linear	None

Table 1: Summary of methods to regularize linear mixed models and generalized linear mixed models. The comparative studies to this paper use the LASSO, a variant of the Adaptive LASSO (ALASSO, M-ALASSO) , or smoothly clipped absolute deviation (SCAD) as the fixed effects (FE) penalties. All but one use the grouped extension of the FE penalty as the random effects (RE) penalty, only Bondell et al. (2010) use an ungroup penalty on the RE. This paper uses the Elastic Net penalty to capture correlation characteristics between the variables.

3 Reparameterization of the Generalized Linear Mixed Model

This paper will utilize the reparameterization of the LMM model initially defined in Chen and Dunson (2003), and used in Bondell et al. (2010) and Ibrahim et al. (2011). The reparameterization offers a simple design which regularization penalties can be easily applied to the fixed and random effects simultaneously. The covariance matrix of the random effects D is factorized as follows:

$$D = \Lambda\Gamma\Gamma'\Lambda, \tag{2}$$

where $\Lambda = \text{diag}(d_1, \ldots, d_q)$ is a $q \times q$ non-negative diagonal matrix with elements proportional to standard deviations of the random effects, and Γ is a lower triangular matrix that relates to the correlations among the random effects with the (l, m) elements denoted γ_{lm}. The elements of Λ are defined as possibly equal zero, thus enabling a subset of random effects to be selected. Λ and Γ are identifiable due to the assumption that:

$$d_l \geq 0,\ \gamma_{ll} = 1, \text{ and } \gamma_{lr} = 0, \text{ for } l = 1, \ldots, q;\ r = l + 1, \ldots, q.$$

Applying the modified decomposition (2) to the LMM model (1) the reparameterized LMM is defined, where the covariance matrix of b_i is a function of Λ, Γ:

$$y_i = x_i'\beta + z_i'\Lambda\Gamma b_i + \epsilon_i.$$

4 Simultaneous Variable Selection and Estimation via Regularization Penalties

The Adaptive LASSO has been used as the penalty function on the modified LMM by Bondell et al. (2010) due to its oracle qualities. Although, there are drawbacks to its use, the primary disadvantage is that candidate covariates correlated to variables chosen in the active set are dropped from the final solution. This characteristic has been found to be a drawback in large scale data with grouped covariates. Moreover, when solving the likelihood of the LMM we can see that the fixed and random effects are dependent.

$$L(\phi|y, b) = -\frac{N + mq}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(||y - Z(I_m \otimes \Lambda)(I_m \otimes \Gamma)b - X\beta||^2 + b'b), \tag{3}$$

with $\otimes$ denoting the Kronker product, Z is a block diagonal matrix of Z_i, I_m is the identity matrix of dimension m.

To overcome these issues we apply a variation on the Elastic Net penalty to the reparameterized likelihood function, (3). The standard Elastic Net penalty denoted as P, Friedman et al. (2010), is designed to be applied on a fixed effects model where only β is penalized, as seen in (4) below. In this formulation the problem of collinearity is addressed (L_2 penalty) in conjunction with shrinkage of redundant variables (L_1 penalty).

$$\begin{aligned} \hat{\beta} &= \underset{\beta \in R^{p+1}}{\arg\min} \left[\frac{1}{2N} \sum_{i=1}^{N} (y_i - x_i'\beta)^2 + P(\beta) \right] \\ P(\beta) &= \lambda_2 \sum_{j \in P} \beta_j^2 + \lambda_1 \sum_{j \in P} |\beta_j|. \end{aligned} \tag{4}$$

We augment (4), while keeping the overall structure and characteristics of the Elastic Net, i.e., the quadratic structure in the L_2 penalty. The reparameterization of the LMM allows the penalty function, $\tilde{P}(\beta, d)$, to be dependent on both the fixed and random effects in the model.

In addition, correlated random effects can be included in the final model selection, whereas in the Adaptive LASSO settings this was not possible, thus overcoming the problematic testing of simultaneous random effects, Chen and Dunson (2003). The objective function of the LMMEN is defined as the following:

$$
\begin{aligned}
Q(\phi|y, b) &= ||y - Z\Lambda\Gamma b - X\beta||^2 + \tilde{P}(\beta, d) \\
\tilde{P}(\beta, d) &= \lambda_2^f \sum_{i\in P} \beta_i^2 + \lambda_2^r \sum_{j\in Q} d_j^2 + \lambda_1^f \sum_{i\in P} |\beta_i| + \lambda_1^r \sum_{j\in Q} |d_j|.
\end{aligned} \tag{5}
$$

Where $\tilde{P}$ and $Q(\phi)$ denote the penalty applied to the likelihood and the penalized log-likelihood. When the final model is not a mixed effects model, but either a fixed effects or random effects model then the original form of $P(\beta)$ is applied.

5 Tuning Parameters Selection

As with all penalized likelihood methods performance depends directly on being able to choose the appropriate value of the tuning parameters. As seen in equation 5 the penalty $\tilde{P}(\beta, d)$ contains four regularization parameters that need to be tuned. This is a departure from other methods such as Bondell et al. (2010), Groll (2017), Schelldorfer et al. (2011) that tune a single penalty parameter, and Hui et al. (2016a) and Ibrahim et al. (2011) that tune two penalty parameters one for each type of effect. The former papers use the BIC or an iterative method between AIC and BIC as their tuning criterions, and the latter use predominatly the IC(q) criterion to tune the two parameters simulataneously. In this paper we chose to use the BIC, It is known that under general conditions, BIC is consistent for model selection if the true model belongs to the class of models considered, while although AIC is minimax optimal, it is not consistent for selection. The BIC criterion is defined as in Bondell et al. (2010)

$$
BIC_{(\lambda_1^f, \lambda_1^r, \lambda_2^f, \lambda_2^r)} = -2L(\hat{\phi}) + \log(N) \times df_{(\lambda_1^f, \lambda_1^r, \lambda_2^f, \lambda_2^r)}. \tag{6}
$$

The penaltiy ranges are split into discrete sequences, then while holding three one is sequenced over and the golden section line search, Kiefer (1953) is applied evaluate $L(\hat{\phi})$. $L(\hat{\phi})$ is obtained from $L(\phi)$ using the estimate $\hat{\phi}$ obtaind from the value of the set $\left(\lambda_1^f, \lambda_1^r, \lambda_2^f, \lambda_2^r\right)$. The degrees of freedom are taken as the number of non-zero elements in $\hat{\phi}$. This is done until convergence to the minimal BIC. While this is computationally intensive, the flexibility in selecting specific combinations has higher priority for this research.

6 Asymptotics

Assume that the data $\{(X_i, Z_i, y_i);\ i = 1...m\}$ is a random sample of m subjects from a linear mixed-effects model with a probability density function $f(y_i|X_i, Z_i, \phi)$. Let y_i be an $n_i \times 1$ response measurements for subject i, X_i be an $n_i \times p$ design matrix of explanatory variables, and Z_i be an $n_i \times q$ design matrix of random effects.

Let $\phi = (\beta', d', \gamma')'$ be a vector of size $k \times 1$, where $\beta \in \mathbb{R}^p$, $d \in \mathbb{R}^q$ and γ is of the dimension $\frac{q(q-1)}{2}$. $p = m^\alpha$ is the number of fixed effects, and $q = m^\delta$ the number of the random effects to be estimated. Then number of free elements in the covariance matrix of the random effects, Φ, is $\frac{q(q-1)}{2}$.

In Bondell et al. (2010) the hyperparameters satisfy $\alpha < 1$ and $\delta < 1$ giving a setup of $m > p$, $m > q$. The total number of unknown hyper parameters is $k = p + \frac{q(q+1)}{2} \ll m$. In this paper we are letting $\alpha > 1$, $\delta < 1$ giving a framework of $m < p$, $m > q$, i.e. a high-dimensional problem. The total number of unknown parameters that are estimated in this framework is $k = m^\alpha + \frac{m^\delta(m^\delta+1)}{2} \gg m$.

Let $L_i(\phi) = \log(f(y_i|X_i, Z_i, \phi))$ denote the contribution of observation i to the log-likelihood function, given by:

$$L_i(\phi) = -\frac{1}{2}\log|\mathbf{V}_i| - \frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\beta)'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta), \tag{7}$$

where $\mathbf{V}_i = \sigma^2(Z_i\Lambda\Gamma\Gamma'\Lambda Z_i + I_{n_i})$. Denoting the true value of ϕ as

$$\phi_0 = (\varphi_{10}, \ldots, \varphi_{k0})' = (\phi_{10}', \phi_{20}')',$$

where $\phi_{10} = (\beta_{10}', d_{10}', \gamma_{10}')'$ is an $s \times 1$ vector whose components are non-zero and ϕ_{20} are the $(k-s)$ remaining components of ϕ_0 such that $\phi_{20} = 0$. Accordingly, let $\phi = (\phi_1', \phi_2')'$. To present the theorems the following regularity conditions are imposed:

C1 The Fisher information matrix $I(\phi_{10})$ knowing $\phi_{20} = 0$ is finite and positive definite.

C2 There exists an open subset Θ of $\mathbb{R}^k$, containing the true parameter ϕ_0 such that $L_i(\phi)$ given in (7) admits all third order derivatives, which are continuous and bounded. There exists a finite mean function $M_{jlm}(y_i, X_i, Z_i)$ such that

$$\left|\frac{\partial^3}{\partial\beta\partial\varphi_l\partial\varphi_m}L_i(\phi)\right| < M(y_i, X_i, Z_i).$$

We have:

Theorem 1. *Let $\phi_0 = (\phi_{10}', 0')'$, and the observations follow the LMM model satisfying conditions C1 and C2. If $w_m m^{-1/2} \to \infty$, $(\lambda_1^f + \lambda_1^r)\sqrt{s}/mw_m \to 0$, $(\lambda_2^f + \lambda_2^r)/m \to 0$, and $(\lambda_2^f + \lambda_2^r)s/mw_m \to 0$, then there exists a local maximizer $\hat{\phi} = \begin{pmatrix}\hat{\phi}_1 \\ 0\end{pmatrix}$ of $Q\left\{\begin{pmatrix}\hat{\phi}_1 \\ 0\end{pmatrix}\right\}$ such that $\hat{\phi}_1$ is w_m consistent for ϕ_{10}.*

Theorem 2. *Let the observations follow the LMM model satisfying conditions C1 and C2. If $\lambda_m \to \infty$ then with probability tending to 1 for any given ϕ_1 satisfying $||\phi_1 - \phi_{10}||_1 \leq Mm^{-1/2}$ and some constant $M > 0$,*

$$Q\left\{\begin{pmatrix}\phi_1\\0\end{pmatrix}\right\} = \max_{||\phi_2||_1 \leq Mm^{-1/2}} Q\left\{\begin{pmatrix}\phi_1\\\phi_2\end{pmatrix}\right\}.$$

7 Simulations

Simulation testing the model selection performance were carried out on five scenarios. In each scenario 200 data sets were simulated from a multivariate normal density.

$$y_i \sim N(X_i\beta, \sigma^2(Z_i\Psi Z_i' + I_{n_i}))$$

The true values of $(\beta_1, \beta_2) = (1, 1)$, and the true variance covariance matrix

$$\Psi = \begin{pmatrix} 9 & 4.8 & 0.6 \\ 4.8 & 4 & 1 \\ 0.6 & 1 & 1 \end{pmatrix}$$

The parameterization of the five scenarios are defined in Table 2.

Scenario	Subjects m	Obs per subject n_i	Fixed Effects (real) p	Random Effects (real) q	Attribute
1	30	5	9 (2)	4 (3)	baseline
2	60	10	9 (2)	4 (3)	$+(n\uparrow)$
3	60	5	9 (2)	10 (3)	$+(n\uparrow, q\uparrow)$
4	60	10	9 (3)	4 (3)	$+(n\uparrow, \rho_\beta > 0)$
5	30	5	200 (20)	4 (3)	$+(p >> n)$

Table 2: Simulation Scenarios

LMMEN is compared to M-ALASSO Bondell et al. (2010), the R packages that solve: glmmLasso Groll (2017), lmmlasso Schelldorfer (2011) and rPQL Hui et al. (2016a) which implemets the SCAD[1] penalty PQL approximation of the GLMM marginal likelihood function. In all the methods we used the BIC criterion to select the final model in each simulation.

The first three scenarios are taken from Bondell et al. (2010), where the true model under consideration in scenarios 1 and 2 is defined as model (8a) and scenario 3 where $X = Z$ as model (8b).

$$y_{ij} = b_{i1} + \beta_1 x_{ij1} + \beta_2 x_{ij2} + b_{i2}Z_{ij1} + b_{i3}Z_{ij2} + \epsilon_{ij} \qquad \epsilon_{ij} \sim N(0,1) \quad (8a)$$
$$y_{ij} = b_{i1} + (\beta_1 + b_{i2})x_{ij1} + b_{i3}X_{ij3} + \epsilon_{ij} \qquad \epsilon_{ij} \sim N(0,1) \quad (8b)$$

We add two new scenarios to test LMMEN under situations of correlation in the fixed effects covariates and when there is a high dimension problem in the fixed effects ($p >> n$).

[1]The SCAD hyper parameter is set to a=3.7 Fan and Li (2001)

Scenario 4 tests the model performance under settings that high correlation between fixed variables exists. We extend scenario 2 by replacing X_3 with a linear combination of X_1,X_2 such that, $X_3 = wX_1 + (1 - w)X_2 + \epsilon$ where $\epsilon \sim N(0, \tau)$. This introduces high correlation in the first three fixed effects, in this setting the LASSO and Adaptive Lasso discard one of these fixed effects thus rendering the model selection inferior. This scenario was found to be beyond the limitations of the glmmlasso and the rPQL packages.

Scenario 5 tests the performance in high dimension settings. The number of fixed effects is increased to 200 and the first 20 are real parameters while the remainder 180 are nuisance, the random effects remain as in the previous scenarios. This scenario can only be run under LMMEN since the initial values are not calculated using the solution of an unpenalized mixed model, as in the other methods.

Tables 3, 4, 5 depict the summary statistics of each parameter estimated within each scenario, where the fixed effects are in Tables 3, 4 and the standard deviations of the random effects are in Table 5. The first three scenarios the real fixed effects are chosen consistently in all the method, where nusiance parameters are chosen with higher regularity. We also notice that the LMMEN and the M-ALASSO coefficient estimates are comparitavley underestimated. Scenario 4 we see that LMMEN out performs the other methods by estimating the real paramters including the β_3 which is the linear combination of the first two. As expected the other methods discards on of the three paramters due to the use of only $\ell 1$ penalty. Scenario 5 shows that the LMMEN estimates all real parameters to an weighteed average of 0.2 and sets the nuisance to zero on average. In the random effects selection we see that the LMMEN bias in the estimation of the variance components increases with higher variance levels. The glmmLasso, lmmlasso and the rPQL with the SCAD penalty in showed diminished ability to set nuisance parameters to zero.

Table 6 shows the percent of variables correctly selected for the whole model, only the fixed effects and only the random effects for each scenario. This analysis was carried out in two settings, the first summarises is if the real parameters are a subset of the final model, this is denoted as 'subset', and a stricter summary if only the real parameters were chosen in the final model, denoted as 'oracle'. In the subset analysis the real parameters are selected at high levels in all methods in the first two scenarios. In scenario 4 results seem positive for the lmmlasso, but when cross referencing the estimated values in Table 3 we see that while all varaibles are included $beta_3$ is very close to zero on average in both the lmmlasso and the M-ALASSO, thereby making it's inclusion less relevant. In the final scenario all the real parameters were in the final model 0% of the simulations, this is due to the large amount of fixed effects that were inluded, while their coefficient estimates were on average 0.2. The oracle analysis shows that correctly selecting only the real parameters is a much more difficult task and the LMMEN and the M-ALASSO outpermformed the other methods in the first three scenarios, in scenario 4 the LMMEN out performed the M-ALASSO, while the lmmen selected the correct parameters, but with near zero coefficient values.

Scenario	Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$
1	glmmLasso	1 (0.91,1.06)	1 (0.93,1.06)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	LMMEN	0.8 (0.48,0.92)	0.8 (0.5,0.93)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	lmmlasso	1 (0.92,1.07)	1 (0.93,1.06)	0 (-0.07,0.07)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	M-ALASSO	1 (0.88,1.04)	1 (0.91,1.05)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	rPQL	1 (0.9,1.13)	1 (0.89,1.09)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
2	glmmLasso	1 (0.98,1.03)	1 (0.97,1.03)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	LMMEN	1 (0.93,1)	1 (0.92,1)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	lmmlasso	1 (0.98,1.03)	1 (0.97,1.03)	0 (-0.02,0.03)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	M-ALASSO	1 (0.93,1)	1 (0.93,0.99)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	rPQL	1 (0.97,1.03)	1 (0.96,1.03)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
3	LMMEN	0.8 (0.56,0.96)	0 (0,0.01)	0.9 (0.86,1.01)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	lmmlasso	1 (0.82,1.19)	0 (-0.09,0.11)	1 (0.95,1.06)	0 (-0.05,0.06)	0 (-0.05,0.06)	0 (-0.06,0.05)	0 (-0.06,0.05)	0 (-0.07,0.05)	0 (-0.05,0.07)
	M-ALASSO	0.8 (0.75,0.92)	0 (0,0)	0.8 (0.75,0.89)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	rPQL	1 (0.74,1.17)	0 (0,0)	1 (0.9,1.09)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
4	LMMEN	0.2 (0,0.46)	0.2 (0,0.49)	0.6 (0.35,0.84)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	lmmlasso	1 (0.98,1.03)	1 (0.97,1.03)	0 (-0.01,0.02)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	M-ALASSO	0.9 (0.4,0.97)	0.9 (0.52,0.99)	0 (0,1.07)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)

Table 3: Scenarios 1-4 summary statistics of fixed effects selection. Each column represents a coefficient and rows are grouped by scenario. In each cell there is the mean value of the estimate and underneath the lower and upper quantile of the estimated coefficient. Not all methods were able to run on scenarios 3,4 due to package constraints relevant to the method.

Scenario	Method										
5	LMMEN	$\hat{\beta}_1$ 0.2 (0,0.34)	$\hat{\beta}_2$ 0.2 (0,0.25)	$\hat{\beta}_3$ 0.2 (0,0.38)	$\hat{\beta}_4$ 0.2 (0,0.27)	$\hat{\beta}_5$ 0.2 (0,0.41)	$\hat{\beta}_6$ 0.2 (0,0.39)	$\hat{\beta}_7$ 0.2 (0,0.33)	$\hat{\beta}_8$ 0.2 (0,0.3)	$\hat{\beta}_9$ 0.2 (0,0.32)	$\hat{\beta}_{10}$ 0.2 (0,0.36)
		$\hat{\beta}_{11}$ 0.2 (0,0.41)	$\hat{\beta}_{12}$ 0.2 (0,0.25)	$\hat{\beta}_{13}$ 0.2 (0,0.34)	$\hat{\beta}_{14}$ 0.2 (0,0.4)	$\hat{\beta}_{15}$ 0.2 (0,0.33)	$\hat{\beta}_{16}$ 0.2 (0,0.26)	$\hat{\beta}_{17}$ 0.2 (0,0.34)	$\hat{\beta}_{18}$ 0.2 (0,0.37)	$\hat{\beta}_{19}$ 0.2 (0,0.33)	$\hat{\beta}_{20}$ 0.2 (0,0.39)
		$\hat{\beta}_{21-200}$ 0 (0,0)									

Table 4: Scenario 5 summary statistics of fixed effects selection. Due to the large dimension of real fixed effects estimated the values are wraped in a ribbon and the nuisance parameters are grouped together to save space. In each cell there is the mean value of the estimate and underneath the lower and upper quantile of the estimated coefficient. Not all methods were able to run on scenarios 5 due to package constraints relevant to the method.

8 Case Study

The LMMEN penalty was tested on high dimensional panel data accumulated as part the Good Judgment Project within the Aggregative Contingent Estima-

Scenario	Method	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_3$	$\hat{d}_4$	$\hat{d}_5$	$\hat{d}_6$	$\hat{d}_7$	$\hat{d}_8$	$\hat{d}_9$	$\hat{d}_{10}$
1	glmmLasso	3 (2.7,3.19)	2 (1.78,2.18)	1 (0.86,1.08)	0.3 (0.27,0.37)						
	LMMEN	2.2 (1.98,2.36)	1.6 (1.43,1.74)	0.8 (0.66,0.93)	0 (0,0)						
	lmmlasso	3 (2.69,3.15)	2 (1.82,2.21)	1 (0.87,1.09)	0.2 (0.08,0.24)						
	M-ALASSO	2.8 (2.59,3.01)	1.9 (1.67,2.07)	0.9 (0.79,1.02)	0 (0,0)						
	rPQL	2.4 (2.16,2.68)	2 (1.66,2.14)	0.5 (0.25,0.75)	0.2 (0.11,0.36)						
2	glmmLasso	3 (2.82,3.17)	2 (1.88,2.14)	1 (0.93,1.06)	0.2 (0.18,0.21)						
	LMMEN	1.8 (1.69,1.87)	1.4 (1.31,1.49)	0.9 (0.79,0.93)	0 (0,0.09)						
	lmmlasso	3 (2.81,3.17)	2 (1.88,2.14)	1 (0.93,1.06)	0.1 (0.03,0.09)						
	M-ALASSO	2.1 (2,2.21)	1.6 (1.47,1.67)	0.9 (0.81,0.94)	0 (0,0)						
	rPQL	2.9 (2.75,3.08)	2 (1.85,2.13)	0.9 (0.8,0.97)	0.1 (0.09,0.19)						
3	LMMEN	1.9 (1.79,2.12)	1.4 (1.29,1.5)	0.5 (0,0.71)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	lmmlasso	2.9 (2.75,3.19)	2 (1.88,2.14)	1 (0.93,1.09)	0.2 (0.07,0.25)	0.2 (0.09,0.24)	0.2 (0.11,0.28)	0.2 (0.12,0.27)	0.2 (0.12,0.24)	0.2 (0.12,0.28)	0.2 (0.12,0.27)
	M-ALASSO	2.2 (2.06,2.33)	1.6 (1.52,1.74)	0.7 (0.64,0.79)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	rPQL	2.2 (1.94,2.45)	1.8 (1.64,2.02)	0.4 (0.25,0.6)	0.2 (0.09,0.3)	0.2 (0.08,0.31)	0.2 (0.09,0.31)	0.2 (0.08,0.29)	0.2 (0.08,0.29)	0.2 (0.09,0.29)	0.2 (0.07,0.27)
4	LMMEN	2.2 (1.98,2.39)	1.6 (1.41,1.76)	0.8 (0.63,0.93)	0 (0,0)						
	lmmlasso	3 (2.81,3.17)	2 (1.88,2.14)	1 (0.93,1.06)	0.1 (0.02,0.09)						
	M-ALASSO	2.5 (2.27,2.67)	1.7 (1.55,1.83)	0.8 (0.72,0.98)	0 (0,0)						
5	LMMEN	2.3 (1.99,2.47)	1.7 (1.43,1.99)	0 (0,0.69)	0 (0,0)						

Table 5: Scenarios 1-5 summary statistics of random effects selection. Each column represents a coefficient and rows are grouped by scenario. In each cell there is the mean value of the estimate and underneath the lower and upper quantile of the estimated coefficient. Not all methods were able to run on scenarios 3,4,5 due to package constraints relevant to the method.

tion (ACE) Program [2]. The aim of this program is *"to dramatically enhance the accuracy, precision, and timeliness of forecasts for a broad range of event types, through the development of advanced techniques that elicit, weight, and combine the judgments of many intelligence analysts."*. The study is characterized as a longitudinal study where probabilistic forecasts are derived from crowd sentiment.

The Good Judgment team recruited approximately 3,000 users in the first year, which varied in 'expertise' and randomly assign them to 3 training groups. These training groups are classified as:

- **A: Control** No training
- **B: Probability** Trained to use probablitic techniques to compare classes and base rates of the occurence of events, average across expert opinions and assume that long term trends are consistent.
- **C: Scenario** Trained to break down initial assumption to it's causal drivers, build confidence intervals and worst case scenarios and combine assumptions through their causal drivers.

[2]Sponsored by the U.S. Intelligence Advanced Research Projects Activity (IARPA).

Scenario	Method	C (subset)	CF (subset)	CR (subset)	C (oracle)	CF (oracle)	CR (oracle)
1	glmmLasso	1.00	1.00	1.00	0.00	0.79	0.00
	LMMEN	0.86	0.94	0.90	0.49	0.49	0.78
	lmmlasso	1.00	1.00	1.00	0.00	0.00	0.07
	M-ALASSO	0.98	1.00	0.98	0.49	0.49	0.98
	rPQL	0.97	1.00	0.97	0.01	0.78	0.01
2	glmmLasso	1.00	1.00	1.00	0.00	0.87	0.00
	LMMEN	1.00	1.00	1.00	0.29	0.29	0.45
	lmmlasso	1.00	1.00	1.00	0.00	0.00	0.07
	M-ALASSO	1.00	1.00	1.00	0.64	0.64	0.93
	rPQL	1.00	1.00	1.00	0.00	0.85	0.00
3	LMMEN	0.65	0.96	0.68	0.19	0.19	0.67
	lmmlasso	0.97	0.97	0.97	0.00	0.00	0.00
	M-ALASSO	0.54	0.54	0.98	0.47	0.47	0.77
	rPQL	0.79	0.80	0.97	0.00	0.62	0.00
4	LMMEN	0.38	0.41	0.88	0.26	0.26	0.86
	lmmlasso	1.00	1.00	1.00	0.07	1.00	0.07
	M-ALASSO	0.07	0.07	0.98	0.04	0.04	0.94
5	LMMEN	0.00	0.00	0.29	0.00	0.00	0.23

Table 6: Scenarios 1-5 percentage of datasets with correctly selected parameters. Two types of selection criteria are summarized. The left hand side (denoted as subset) checks if at least the real coefficients were selected and the right hand side (denoted as oracle) checks if only the real variables were chosen and no nuisance parameters. 'C' denotes correct for the overall model, 'CF' denotes correct fixed effects and 'CR' denotes correct random effects. The cells contain the ratio of iterations that returned the correct values per column head and method. Not all methods were able to run on scenarios 3,4,5 due to package constraints relevant to the method.

Within each training group there are 4 types of opinion polls in which a user can be assigned. These groups are classified as:

- **1: Independent** Requires forecasters to work independently.
- **2: Crowd Beliefs** Forecasters see the distribution of the groupâĂŹs forecasts.
- **3: Prediction Markets**[3] System prices the bet by offering a contract to the participant that will pay a fixed amount if and only if s/he is correct.
- **4: Teamwork** Groups of 20-25 forecasters who explain why they make their forecasts, view the explanations of others, comment on them, coordinate a division of labor and enforce group beliefs.

These opinion polls allow for different levels of interactions between the users in each group. Thus making users randomly assigned to 12 groups. 75 active questions were opened over the first year. These questions had various themes such as the future outcome of: financial turbulence, election results, economic stability and diplomatic security, a full list of the questions can be found in the

[3] This group has been omitted due to technical problems that arose during the first year of the competition.

Appendix B Table 7. Each user could answer an active question at any time until the question was closed and resolved. This design is a natural one for a repeated measures model with random effects, in which the questions are designated as subjects with random intercepts and for each group a random effect is estimated. In addition there are 40 fixed variables that contain demographic, psychological and past performance information, a full list of the variables can be found in the Appendix B Table 8. The data tested was 100 random samples of 50 answers from 20 randomly sampled questions, giving a block structure of 1,000 observations.

A model defined to predict weights to assign to each user per question and aggregating user outcome predictions to a group outcome prediction.

$$\begin{gathered}\frac{y_i}{1-y_i} = x'_{ij}\beta + z'_{ij}b_i + \epsilon_i, \\ \epsilon_i \sim N(0, \sigma^2 I_{n_i}).\end{gathered} \tag{9}$$

where y_i is the probabliity of each true outcome, β is the fixed effects coefficient vector for x and b_i is the subject-specific random effects for z, which are the (training::opinion poll) group designations for each forecaster. We define $\hat{w}_i$ as the predicted weight. The estimated weight is transformed to better separate predictions that are in the middle the $[0, 1]$, which indicates indecision and forecasts that are closer to extremes of the range. To acheive this the weights are tranformed by exponenation $\tilde{w}_i = \exp(\hat{w}_i)$ and truncated to the 20^{th} and 80^{th} percentile of the estimated vector.

The LMMEN with specifications for the design structure will be compared to the rPQL method with the SCAD penalty on both the fixed and random effects. Both of the methods are suitable to the data structure to select relevant fixed effects and determine which training/opinion poll groups have non-zero variance parameters. Additionally both methods share the regularization characteristic of penalizing the fixed and random effects seperately, but the rPQL uses the same penalty value for both, while the LMMEN allows for greater flexibility. Two levels of algorithm performance will be investigated, first is the model selection and second is the accuracy of the aggregated predictions. The statistic which will be used to test performance of the aggregated predictions is the Brier score. In this case study only binary events are taken under consideration and 6 questions are omitted under this constraint. The Brier score equation is defined as

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o)^2,$$

in which f_t is the prediction at time t, o is the question outcome, and N is the number of prediction instances. First we compare the fixed effect covariate selection between the two algorithms as seen in panel (b) of Figure 1. It can be seen that the LMMEN produces a higher level of sparsity than the rPQL with SCAD penalty and the variables chosen are persistent in the simulation. The groups of users are assumed to be distributed normally with a variation parameter. The results of random effects selection can be found in panel (c) of Figure 1. In the

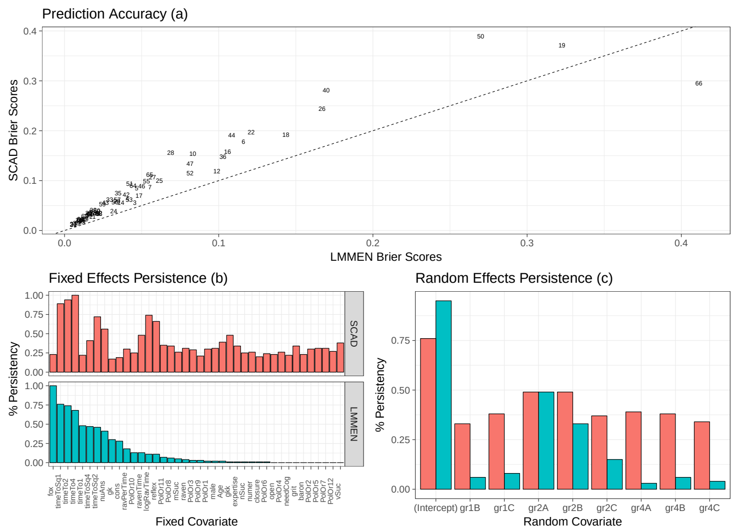

Figure 1: Model performance of LMMEN and rPQL with SCAD penalty tested on 100 random samples of 1000 observations from the Good Judgment study. Panel (a) compares the probabilistic forecast accuracy of the two methods using the Brier Score as the loss function. Panel (b) compares the distribution of fixed effects selection persistency between the two methods. Panel (c) compares the distribution of random effects selection persistency between the two methods.

LMMEN solution we see that the variance estimate of groups {gr1B: Independent::Probability, gr1C: Independent::Scenario, gr4A: Teamwork::Control, gr4B: Teamwork::Probability, gr4C: Teamwork::Scenario} is equal to zero in a large percent of the simulations, thus concluding that there is no difference between the user responses in those groups, whereas in the regularized PQL solution there is a some random effects selection but it does not have persistency among any of the groups.

The second level of performance investigated is the prediction accuracy. The estimated non-zero covariates after selection are used to aggregate out of sample user predictions of active questions. The results of the two selection methods can be found in panel (a) of Figure 1. We see that the LMMEN out performs the rPQL in nearly every question, the average brier scores for LMMEN and rqpl are .055 and .085 respectively.

9 Discussion

In the paper we have shown that fixed and random effects in high dimensional linear mixed models can be simultaneously selected. This selection method introduces the ability to select variables under conditions of multicollinearity both in the fixed and random effects. This method, LMMEN, furthers current variable selection of these models with the introduction of a ridge penalty into the optimization.

It was found through simulations that this method correctly selects fixed and random effects under sparse data designs. Simulations were carried out under the Gaussian assumption for both the conditional distribution and the distribution of the random effects. Further simulations will be carried out which relax the assumption of the conditional distribution. When testing the LMMEN in the case study the variable selection was more apparent both in the fixed and random effects. The LMMEN gave further insight into the characteristics of groups of users, where a subset of them were found not have prediction difference within the groups. Finally the prediction accuracy of the LMMEN model compared to the rPQL with a SCAD penalty.

This paper applies the Brier Score (L2 loss) as the loss function to tune the penalty parameters in the case study. One could calibrate the penalty parameters is the intraclass correlation (ICC) levels. The ICC is intrinsic to random effects models, and is regularly used for evaluating the level of correlation between different groups as defined by the model. Applying the LMMEN while calibrating to minimize the ICC could be a vital tool for correctly selecting candidate random effects to model the data design and will be assessed in future work.

A Appendix: Proofs

For the penalized log-likelihood in (5), let $\phi = (\phi_1', 0')'$ and let

$$L^1(\phi_1) \equiv L\left\{\begin{pmatrix}\phi_1\\0\end{pmatrix}\right\} \text{ and } Q^1(\phi_1) \equiv Q\left\{\begin{pmatrix}\phi_1\\0\end{pmatrix}\right\}$$

denote the log-likelihood and the penalized log-likelihood of the first s components of ϕ.

Proof Theorem 1. Consider the penalized log-likelihood $Q(\phi)$ given in (5) in the neighborhood of the true value ϕ_{10}. Let $u \neq 0$, and $\phi_1 = \phi_{10} + w_m u$. Setting $\phi_2 = 0$, we show that for a small enough $\epsilon > 0$, there exists a large constant C such that for a sufficiently large m,

$$P\Big(\sup_{\|u\|=C} Q^1(\phi_{10} + w_m u) < Q(\phi_{10})\Big) \geq 1 - \epsilon.$$

Thus, with probability $1-\epsilon$ the maximum is within the ball of radios Cw_m.

Note that

$$\begin{aligned} mD_m(u) &\equiv Q^1(\phi_1) - Q^1(\phi_{10}) \\ &= -\left[L^1(\phi_{10} + w_m u) - L^1(\phi_{10})\right] \\ &\quad + \lambda_1^f\big(\|\beta_0 + w_m u_\beta\|_1 - \|\beta_0\|_1\big) + \lambda_1^r\big(\|d_0 + w_m u_d\|_1 - \|d_0\|_1\big) \\ &\quad + \lambda_2^f\big(\|\beta_0 + w_m u_\beta\|_2^2 - \|\beta_0\|_2^2\big) + \lambda_2^r\big(\|d_0 + w_m u_d\|_2^2 - \|d_0\|_2^2\big), \end{aligned}$$

where we divided u to its natural components $u_\beta \in R^p$ and $u_d \in R^q$. Using the Taylor series expansion we have

$$\begin{aligned}
&D_m(u)\\
&= -w_m(m^{-1}\nabla L(\phi_{10}))'u - \frac{w_m^2}{2m}u'[\nabla^2 L(\phi_{10})]u + R_m\\
&\quad + m^{-1}\lambda_1^f\big(\|\beta_0 + w_m u_\beta\|_1 - \|\beta_0\|_1\big) + m^{-1}\lambda_1^r\big(\|d_0 + w_m u_d\|_1 - \|d_0\|_1\big)\\
&\quad + m^{-1}\lambda_2^f\big(\|\beta_0 + w_m u_\beta\|_2^2 - \|\beta_0\|_2^2\big) + m^{-1}\lambda_2^r\big(\|d_0 + w_m u_d\|_2^2 - \|d_0\|_2^2\big),
\end{aligned}$$

where $\nabla L(\phi_{10}), \nabla^2 L(\phi_{10})$ denote the vector and matrix of the first and second order partial derivatives of $L(\phi_1)$ at ϕ_{10} respectively. $\nabla P(\beta, d), \nabla^2 P(\beta, d)$ denote the first and second derivatives of the penalty term at (β_0, d_0). The remainder R_n tends to zero as m $\to \infty$ since, by C2, $|R_m|$ can be bounded by

$$\left(\frac{w_m^3\|u\|_2^3}{6m}\right)\sum_{i=1}^m M(y_i, X_i, Z_i) = O_P(w_m^3).$$

The jth partial derivative for each corresponding β_1, d_1, γ_1 the $\nabla L(\phi_{10})$ satisfies $E\left\{\frac{\partial}{\partial\beta_j}L(\phi_1)\right\} = E\left\{\frac{\partial}{\partial d_j}L(\phi_1)\right\} = E\left\{\frac{\partial}{\partial\gamma_j}L(\phi_1)\right\} = 0$ and thus the corresponding empirical means are $O_p(m^{-1/2})$.

For $\nabla^2 L(\phi_{10})$ we have

$$m^{-1}\nabla^2 L(\phi_{10}) \to_p -I(\phi_{10}),$$

where $I(\phi_{10})$ is the Fisher information evaluated at ϕ_{10}, which is positive definite by (C1). By choosing a sufficiently large C, the second term dominates the first term uniformly in $\|u\| = C$.

For the penalty term if $w_m P(\beta, d) \to 0$ as $m \to \infty$ it follows that $P(\beta, d) \to_p 0$, and thus also dominated by the second term. The absolute value of the penalty component of $D_m(u)$ is bounded by

$$\begin{aligned}
m^{-1}&w_m\lambda_1^f\|u_\beta\|_1 + m^{-1}w_m\lambda_1^r\|u_d\|_1 + m^{-1}\lambda_2^f\big(2w_m\|\beta_0\|_2\|u_\beta\|_2 + w_m^2\|u_0\|_2^2\big)\\
&+ m^{-1}\lambda_2^r\big(2w_m\|d_0\|_2\|u_d\|_2 + w_m^2\|u_d\|_2^2\big)\\
&\le m^{-1}w_m C\big(\lambda_1^f\sqrt{s} + \lambda_1^r\sqrt{s} + \lambda_2^f(2\|\beta_0\|_2 + w_m C) + \lambda_2^r(2\|d_0\|_2 + w_m C)\big).
\end{aligned}$$

which is dominated by the second term of $D_m(u)$. Therefore, by choosing a sufficiently large C there exists a local maximum inside $\{\phi_{10} + w_m u : \|u\| < C\}$ with probability $1 - \epsilon$, thus there exists a local maximizer $\hat{\phi} = (\hat{\phi_1}', 0')'$ of $\phi_0 = (\phi_1', 0')'$ such that $\|\hat{\phi_1} - \phi_{10}\| = O_p(w_m)$. $\square$

For the following proof we define $\phi = (\beta', d', \gamma')$ as a $k \times 1$ vector of unknown parameters of size $k = k_\beta + k_d + k_\gamma$. Let $\phi_2 = (\beta_2', d_2', \gamma_2')$ be a vector of size $k_2 = k - s$ corresponding to the true zero parameters, given $k_2 = k_{\beta_2} + k_{d_2} + k_{\gamma_2}$. Reminding that we defined earlier that the likelihood and the penalized log likelihood as

$$L(\phi) = L\left\{\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}\right\} \text{ and } Q(\phi) = Q\left\{\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}\right\}.$$

Proof Theorem 2. For $m \to \infty$ and any $\phi_1 : ||\phi_1 - \phi_{10}||_1 \leq Mm^{-1/2}$ and for $\epsilon_m = Mm^{-1/2}$ and for each $j = (s+1), \dots, (k_{\beta_2} + k_{d_2})$ we have with probability tending to 1 that

$$\begin{aligned} &\frac{\partial}{\partial \varphi_j} Q(\phi) < 0 \text{ for } 0 < \varphi_j < \epsilon_m \\ &\frac{\partial}{\partial \varphi_j} Q(\phi) > 0 \text{ for } -\epsilon_m < \varphi_j < 0 \end{aligned} \tag{10}$$

The partial derivative of $Q(\phi)$ with respect to φ_j is given by:

$$\frac{\partial}{\partial \varphi_j} Q(\phi) = \frac{\partial}{\partial \varphi_j} L(\phi) - (\lambda_1 \mathrm{sgn}(\varphi_j) + 2\lambda_2 \varphi_j)\,,$$

noting that the penalty is dependent on whether φ_j is β or d.

One can verify (10) through the Taylor Series expansion of $\frac{\partial}{\partial \varphi_j} L(\phi) = \frac{\partial}{\partial \varphi_j} L(\phi)$ around ϕ_0:

$$\begin{aligned} \frac{\partial}{\partial \varphi_j} Q(\phi) =& \frac{\partial}{\partial \varphi_j} L(\phi_0) - \sum_{l=1}^{k} \frac{\partial}{\partial \varphi_l} \left(\frac{\partial}{\partial \varphi_j} L(\phi_0) \right) (\varphi_l - \varphi_{l0}) \\ &+ \frac{1}{2} \sum_{i=1}^{m} \sum_{l=1}^{k} \sum_{g=1}^{k} \frac{\partial^2}{\partial \varphi_l \partial \varphi_g} \left(\frac{\partial}{\partial \varphi_j} L_i(\phi_*) \right) (\varphi_l - \varphi_{l0})(\varphi_g - \varphi_{g0}) \\ &- (\lambda_1 \mathrm{sgn}(\varphi_j) + 2\lambda_2 \varphi_j)\,, \end{aligned} \tag{11}$$

where ϕ_* is on the interval connecting ϕ and ϕ_0. Next we define the first order derivatives needed to numerically solve (11):

$$\begin{aligned} L_\beta =& \frac{\partial}{\partial \beta_j} L(\phi_0) = X_j' V^{-1} (y - X\beta) = O_p(m^{-1/2}) \\ L_d =& \frac{\partial}{\partial d_j} L(\phi_0) = \frac{1}{2} \left[\mathrm{Tr}(V^{-1} S^j) + (y - X\beta)' (V^{-1} S^j V^{-1})(y - X\beta) \right] = O_p(m^{-1/2}), \end{aligned}$$

where $S^j = Z(\frac{\partial}{\partial d_j} D\Gamma\Gamma' D)Z'$ and $\mathrm{Tr}(A)$ is the trace operator on a given matrix A. We now define the second order derivatives which follow $\frac{1}{m}\nabla^2 L(\phi)|_{\phi=\phi_0} \to E_{\phi_1=\phi_{10}}[\nabla^2 L(\phi)]$, where

$$E[\nabla^2 L(\phi)] = E \begin{bmatrix} L_{\beta\beta} & L_{\beta d} & L_{\beta\gamma} \\ L'_{\beta d} & L_{dd} & L_{d\gamma} \\ L'_{\beta\gamma} & L'_{d\gamma} & L_{\gamma\gamma} \end{bmatrix},$$

$$
\begin{aligned}
E[L_{\beta\beta}]_j =& - XV^{-1}X \\
E[L_{\beta d}]_j =& - E\left[X_j'(V^{-1}S^jV^{-1})(y - X\beta)\right]|_{\phi=\phi_0} = 0 \\
E[L_{\beta\gamma}]_j =& - E\left[X_j'(V^{-1}T^jV^{-1})(y - X\beta)\right]|_{\phi=\phi_0} = 0 \\
E[L_{dd}]_{jl} =& - \mathrm{Tr}(V^{-1}S^jV^{-1}S^l)|_{\{j\geq(s+1),\phi_j=0\}} = 0 \\
E[L_{\gamma\gamma}]_{jl} =& - \mathrm{Tr}(V^{-1}T^jV^{-1}T^l)|_{\{j\geq(s+1),\phi_j=0\}} = 0 \\
E[L_{d\gamma}]_{jl} =& - \mathrm{Tr}(V^{-1}S^jV^{-1}T^l)|_{\{j\geq(s+1),\phi_j=0\}} = 0,
\end{aligned}
$$

where $T^j = ZD(\frac{\partial}{\partial\gamma_j}\Gamma\Gamma')DZ'$.
Using these partial derivatives we solve (11) first for $\phi_j = \beta_j$ and then for $\phi_j = d_j$.

$$
\begin{aligned}
&\frac{1}{\sqrt{m}}\left(\frac{\partial}{\partial\beta_j}Q(\phi)\right) \\
=&\frac{1}{\sqrt{m}}\Bigg[L_\beta - m\left(\sum_{l=1}^{k_\beta}L_{\beta\beta}(\beta_l-\beta_{l0}) + \sum_{l=k_{\beta+1}}^{k_d}L_{\beta d}(d_l - d_{l0}) + \sum_{l=k_{d+1}}^{k_\gamma}L_{\beta\gamma}(\gamma_l-\gamma_{l0})\right) \\
&+ \sum_{i=1}^{m}\sum_{l=1}^{k_\beta}\sum_{g=k_\beta+1}^{k_d}\frac{\partial}{\partial\beta_g}L_{\beta d}(\beta_l-\beta_{l0})(d_g-d_{g0}) \\
&+ \sum_{i=1}^{m}\sum_{l=1}^{k_\beta}\sum_{g=k_d+1}^{k_\gamma}\frac{\partial}{\partial\beta_g}L_{\beta\gamma}(\beta_l-\beta_{l0})(\gamma_g-\gamma_{g0}) \\
&+ \sum_{i=1}^{m}\sum_{l=k_\beta+1}^{k_d}\sum_{g=k_{d+1}}^{k_\gamma}\frac{\partial}{\partial\gamma_g}L_{\beta d}(d_l-d_{l0})(\gamma_g-\gamma_{g0}) \\
&+ \frac{1}{2}\Bigg(\sum_{i=1}^{m}\sum_{l=k_\beta+1}^{k_d}\sum_{g=k_\beta+1}^{k_d}\frac{\partial}{\partial d_g}L_{\beta d}(d_l-d_{l0})(d_g-d_{g0}) \\
&+ \sum_{i=1}^{m}\sum_{l=k_d+1}^{k_\gamma}\sum_{g=k_{d+1}}^{k_\gamma}\frac{\partial}{\partial\gamma_g}L_{\beta\gamma}(\gamma_l-\gamma_{l0})(\gamma_g-\gamma_{g0})\Bigg) - \left(\lambda_1^f\mathrm{sgn}(\beta_j) + 2\lambda_2^f(\beta_j)\right)\Bigg],
\end{aligned}
$$

given $||\phi - \phi_0||_1 \leq Mm^{-1/2}$ then we have

$$
\frac{1}{\sqrt{m}}\left(\frac{\partial}{\partial\beta_j}Q(\phi)\right) = -\left(\lambda_1^f\mathrm{sgn}(\beta_j) + 2\lambda_2^f(\beta_j)\right) + O_p(1). \tag{12}
$$

For $\beta_{j0} = 0$ and $\{\lambda_1^f, \lambda_2^f\} \to \infty$ the sign of the derivative is completely determined by β_j, more specifically:

$$
\begin{array}{llll}
\text{if} & M > \beta_j > 0 & \text{then} & \frac{\partial}{\partial\beta_j}Q(\phi) < 0 \\
\text{if} & -M < \beta_j < 0 & \text{then} & \frac{\partial}{\partial\beta_j}Q(\phi) > 0
\end{array}.
$$

Similarly,

$$
\begin{aligned}
&\frac{1}{\sqrt{m}}\left(\frac{\partial}{\partial d_j}Q(\phi)\right)\\
&=\frac{1}{\sqrt{m}}\Bigg[L_d - m\left(\sum_{l=1}^{k_\beta} L_{\beta\beta}(\beta_l-\beta_{l0}) + \sum_{l=k_{\beta+1}}^{k_d} L_{\beta d}(d_l-d_{l0}) + \sum_{l=k_{d+1}}^{k_\gamma} L_{\beta\gamma}(\gamma_l-\gamma_{l0})\right)\\
&\quad+\sum_{i=1}^{m}\sum_{l=1}^{k_\beta}\sum_{g=1}^{k_\beta}\frac{\partial}{\partial\beta_g}L_{d\beta}(\beta_l-\beta_{l0})(\beta_g-\beta_{g0})\\
&\quad+\sum_{i=1}^{m}\sum_{l=k_\beta+1}^{k_d}\sum_{g=k_d+1}^{k_\gamma}\frac{\partial}{\partial\gamma_g}L_{dd}(d_l-d_{l0})(\gamma_g-\gamma_{g0})\\
&\quad+\sum_{i=1}^{m}\sum_{l=k_\beta+1}^{k_d}\sum_{g=k_{d+1}}^{k_\gamma}\frac{\partial}{\partial d_g}L_{d\beta}(\beta_l-\beta_{l0})(d_g-d_{g0})\\
&\quad+\frac{1}{2}\Bigg(\sum_{i=1}^{m}\sum_{l=k_\beta+1}^{k_d}\sum_{g=k_\beta+1}^{k_d}\frac{\partial}{\partial d_g}L_{dd}(d_l-d_{l0})(d_g-d_{g0})\\
&\quad+\sum_{i=1}^{m}\sum_{l=k_d+1}^{k_\gamma}\sum_{g=k_{d+1}}^{k_\gamma}\frac{\partial}{\partial\gamma_g}L_{d\gamma}(\gamma_l-\gamma_{l0})(\gamma_g-\gamma_{g0})\Bigg) - (\lambda_1^r\mathrm{sgn}(d_j) + 2\lambda_2^r(d_j))\Bigg],
\end{aligned}
$$

given $||\phi-\phi_0||_1 \le Mm^{-1/2}$ then we have

$$\frac{1}{\sqrt{m}}\left(\frac{\partial}{\partial d_j}Q(\phi)\right) = -\left(\lambda_1^r\mathrm{sgn}(d_j) + 2\lambda_2^r(d_j)\right) + O_p(1).$$

For $d_{j0}=0$ and $(\lambda_1^r,\lambda_2^r)\to\infty$ the sign of the derivative is completely determined by d_j, more specifically:

$$
\begin{array}{llll}
\text{if} & M > d_j > 0 & \text{then} & \frac{\partial}{\partial d_j}Q(\phi) < 0\\
\text{if} & -M < d_j < 0 & \text{then} & \frac{\partial}{\partial d_j}Q(\phi) > 0
\end{array}.
$$

□

B Appendix: Case Study Tables

Id	Question Text	Correct Answer	Date Activated	Date Closed	Duration (days)
1	Will the Six-Party talks (among the US, North Korea, South Korea, Russia, China, and Japan) formally resume in 2011?	No	2011-08-31	2012-01-03	125
2	Will Serbia be officially granted EU candidacy by 31 December 2011?	No	2011-08-31	2012-01-03	125
3	Will the United Nations General Assembly recognize a Palestinian state by 30 September 2011?	No	2011-08-30	2011-10-03	34
4	Will Daniel Ortega win another term as President of Nicaragua during the late 2011 elections?	Yes	2011-08-31	2011-11-09	70
5	Will Italy restructure or default on its debt by 31 December 2011?	No	2011-08-31	2012-01-03	125
6	By 31 December 2011, will the World Trade Organization General Council or Ministerial Conference approve the "accession package" for WTO membership for Russia?	Yes	2011-08-31	2011-12-16	107
7	Will the 30 Sept 2011 'last' PPB for Nov 2011 Brent Crude oil futures exceed $115?	No	2011-09-06	2011-10-03	27
8	Will the Nikkei 225 index finish trading at or above 9,500 on 30 September 2011?	No	2011-09-06	2011-10-03	27
9	Will Italy's Silvio Berlusconi resign, lose re-election/confidence vote, OR otherwise vacate office before 1 October 2011?	No	2011-09-06	2011-10-03	27
10	Will the London Gold Market Fixing price of gold (USD per ounce) exceed $1850 on 30 September 2011 (10am ET)?	No	2011-09-06	2011-10-03	27
11	Will Israel's ambassador be formally invited to return to Turkey by 30 September 2011?	No	2011-09-06	2011-10-03	27
12	Will PM Donald Tusk's Civic Platform Party win more seats than any other party in the October 2011 Polish parliamentary elections?	Yes	2011-09-06	2011-10-11	35
13	Will Robert Mugabe cease to be President of Zimbabwe by 30 September 2011?	No	2011-09-06	2011-10-03	27
14	Will Muqtada al-Sadr formally withdraw support for the current Iraqi government of Nouri al-Maliki by 30 September 2011?	No	2011-09-06	2011-10-03	27
15	Will peace talks between Israel and Palestine formally resume at some point between 3 October 2011 and 1 November 2011?	No	2011-10-03	2011-11-02	30
16	Will the expansion of the European bailout fund be ratified by all 17 Eurozone nations before 1 November 2011?	Yes	2011-10-03	2011-10-17	14
17	Will the South African government grant the Dalai Lama a visa before 7 October 2011?	No	2011-10-03	2011-10-11	8
18	Will former Ukrainian Prime Minister Yulia Tymoshenko be found guilty on any charges in a Ukrainian court before 1 November 2011?	Yes	2011-10-03	2011-10-11	8
19	Will Abdoulaye Wade win re-election as President of Senegal?	No	2011-10-03	2012-03-26	175
20	Will the Freedom and Justice Party win at least 20 percent of the seats in the first People's Assembly (Majlis al-Sha'b) election in post-Mubarak Egypt?	Yes	2012-01-07	2012-01-24	17
21	Will Joseph Kabila remain president of the Democratic Republic of the Congo through 31 January 2012?	Yes	2011-10-03	2012-02-01	121
22	Will Moody's issue a new downgrade of the sovereign debt rating of the Government of Greece between 3 October 2011 and 30 November 2011?	No	2011-10-03	2011-12-01	59
23	Will the UN Security Council pass a measure/resolution concerning Syria in October 2011?	No	2011-10-03	2011-11-01	29
24	Will the U.S. Congress pass a joint resolution of disapproval in October 2011 concerning the proposed $5+ billion F-16 fleet upgrade deal with Taiwan?	No	2011-10-03	2011-10-26	23
25	Will the Japanese government formally announce the decision to buy at least 40 new jet fighters by 30 November 2011?	No	2011-10-03	2011-12-01	59
26	Will the Tunisian Ennahda party officially announce the formation of an interim coalition government by 15 November 2011?	No	2011-11-07	2011-11-19	12
27	Will Japan officially become a member of the Trans-Pacific Partnership before 1 March 2012?	No	2011-11-07	2012-03-01	115
28	Will the United Nations Security Council pass a new resolution concerning Iran by 1 April 2012?	No	2012-03-21	2012-04-02	12
29	Will Hamad bin Isa al-Khalifa remain King of Bahrain through 31 January 2012?	Yes	2011-11-07	2012-02-01	86
30	Will Bashar al-Assad remain President of Syria through 31 January 2012?	Yes	2011-11-07	2012-02-01	86
31	Will Italy's Silvio Berlusconi resign, lose re-election/confidence vote, OR otherwise vacate office before 1 January 2012?	Yes	2011-11-13	2011-11-15	2
32	Will Lucas Papademos be the next Prime Minister of Greece?	Yes	2011-11-11	2011-11-11	0
33	Will Lucas Papademos resign, lose re-election/confidence vote, or vacate the office of Prime Minister of Greece before 1 March 2012?	No	2011-12-12	2012-03-01	80
34	Will the United Kingdom's Tehran embassy officially reopen by 29 February 2012?	No	2011-12-12	2012-03-01	80
35	Will a trial for Saif al-Islam Gaddafi begin in any venue by 31 March 2012?	No	2011-12-12	2012-04-02	112
36	Will S&P downgrade the AAA long-term credit rating of the European Financial Stability Facility (EFSF) by 30 March 2012?	Yes	2011-12-14	2012-01-17	34
37	Will North Korea successfully detonate a nuclear weapon, either atmospherically, underground, or underwater, between 9 January 2012 and 1 April 2012?	No	2012-01-09	2012-04-02	84
38	By 1 April 2012, will Egypt officially announce its withdrawal from its 1979 peace treaty with Israel?	No	2012-01-09	2012-04-02	84
39	Will Kim Jong-un attend an official, in-person meeting with any G8 head of government before 1 April 2012?	No	2012-01-09	2012-04-02	84
40	Will Christian Wulff resign or vacate the office of President of Germany before 1 April 2012?	Yes	2012-01-09	2012-02-17	39
41	Will the daily Europe Brent Crude FOB spot price per barrel be greater than or equal to $150 before 3 April 2012?	No	2012-01-09	2012-04-03	85
42	Will the Taliban begin official in-person negotiations with either the US or Afghan government by 1 April 2012?	No	2012-02-22	2012-04-02	40
43	Will Yousaf Raza Gillani resign, lose confidence vote, or vacate the office of Prime Minister of Pakistan before 1 April 2012?	No	2012-01-23	2012-04-02	70
44	Will Yemen's next presidential election commence before 1 April 2012?	Yes	2012-01-23	2012-02-21	29
45	Will Traian Basescu resign, lose referendum vote, or vacate the office of President of Romania before 1 April 2012?	No	2012-01-23	2012-04-02	70
46	Will the UN Security Council pass a new measure/resolution directly concerning Syria between 23 January 2012 and 31 March 2012?	No	2012-03-21	2012-04-02	12
47	Before 1 April 2012, will South Korea officially announce a policy of reducing Iranian oil imports in 2012?	No	2012-01-23	2012-04-02	70
48	Will Israel release Palestinian politician Aziz Duwaik from prison before 1 March 2012?	No	2012-01-23	2012-03-01	38
49	Will Iran and the U.S. commence official nuclear program talks before 1 April 2012?	No	2012-01-30	2012-04-02	63
50	Will Serbia be officially granted EU candidacy before 1 April 2012?	Yes	2012-01-30	2012-03-02	32
51	Will the IMF officially announce before 1 April 2012 that an agreement has been reached to lend Hungary an additional 15+ Billion Euros?	No	2012-01-30	2012-04-02	63
52	Will Libyan government forces regain control of the city of Bani Walid before 6 February 2012?	No	2012-01-30	2012-02-08	9
53	Will a run-off be required in the 2012 Russian presidential election?	No	2012-01-30	2012-03-05	35
54	Will the Iraqi government officially announce before 1 April 2012 that it has dropped all criminal charges against its VP Tareq al-Hashemi?	No	2012-01-30	2012-04-02	63
55	Will Egypt officially announce by 15 February 2012 that it is lifting its travel ban on Americans currently in Egypt?	No	2012-01-30	2012-02-16	17
56	Will a Japanese whaling ship enter Australia's territorial waters between 7 February 2012 and 10 April 2012?	No	2012-02-07	2012-04-11	64
57	Will William Ruto cease to be a candidate for President of Kenya before 10 April 2012?	No	2012-02-07	2012-04-10	63
58	Will Marine LePen cease to be a candidate for President of France before 10 April 2012?	No	2012-02-07	2012-04-10	63
59	Between 21 February 2012 and 1 April 2012, will the UN Security Council announce any reduction of its peacekeeping force in Haiti?	No	2012-02-21	2012-04-02	41
60	Will Mohamed Waheed Hussain Manik resign or otherwise vacate the office of President of Maldives before 10 April 2012?	No	2012-02-21	2012-04-10	49
61	Will Japan commence parliamentary elections before 1 April 2012?	No	2012-02-21	2012-04-02	41
62	Before 13 April 2012, will the Turkish government officially announce that the Turkish ambassador to France has been recalled?	No	2012-02-21	2012-04-13	52
63	Will Standard and Poor's downgrade Japan's Foreign Long Term credit rating at any point between 21 February 2012 and 1 April 2012?	No	2012-02-21	2012-04-02	41
64	Will Myanmar release at least 100 more political prisoners between 21 February 2012 and 1 April 2012?	No	2012-02-21	2012-04-02	41
65	Will a civil war break out in Syria between 21 February 2012 and 1 April 2012?	No	2012-02-21	2012-04-02	41
66	Will Tunisia officially announce an extension of its current state of emergency before 1 April 2012?	Yes	2012-03-05	2012-04-03	29
67	Before 1 April 2012, will Al-Saadi Gaddafi be extradited to Libya?	No	2012-03-05	2012-04-02	28
68	Before 1 April 2012, will the Sudan and South Sudan governments officially announce an agreement on oil transit fees?	No	2012-03-05	2012-04-02	28
69	Will Yemeni government forces regain control of the towns of Jaar and Zinjibar from Al-Qaida in the Arabian Peninsula (AQAP) before 1 April 2012?	No	2012-03-05	2012-04-02	28

Table 7: Case study questions asked to participants.

Id	Type	Variable	Label
1	General Knowledge	gk	General knowledge Score
2		gkk	Adjusted General Knowledge Score
3	Current Question	expertise	User given expertise level in question subject matter (scale 1-5)
4		nuAns	Number of new answers user submitted to question
5		timeTo1	Time that passed from activation of question to answer Opinion Poll 1
6		timeTo2	Time that passed from activation of question to answer Opinion Poll 2
7		timeTo4	Time that passed from activation of question to answer Opinion Poll 4
8		timeToSq1	power(timeTo1,2)
9		timeToSq2	power(timeTo2,2)
10		timeToSq4	power(timeTo4,2)
11	Past Performance	nSuc	Number of total correct answers
12		mSuc	Mean number of total correct answers
13		vSuc	Variance of total correct answers
14	User Deomgraphic	Age	Age of user
15		male	gender of user (boolean)
16	User Psychological	baron	Cognitive Reflection Test and Extended Cognitive Reflection Test (by Jon Baron)
17		closure	Need for Closure
18		cons	Political Philosophies
19		fox	Fox-Hedgehog test
20		grit	Grit test
21		needCog	Need for Cognition
22		numer	Berlin Numeracy
23		open	Actively Open-Minded Thinking
24		raven	Number of correct items Raven's Progressive Matrices
25		ravenTime	Truncated time (in seconds) of submit for Raven's item (3 times the user median)
26		ravPerTime	raven/ravenTime
27		logRavTime	log(ravenTime)
28		reflex	Cognitive Reflection Test (CRT)
30		PolOr1	World politics remains a jungle in which (to quote Thucydides) the strong do what they will and the weak accept what they must
30		PolOr2	International institutions increasingly constrain the conduct of nation-states
31		PolOr3	Economic and population growth are stretching nature to its breaking point.
32		PolOr4	Just when humanity seems to be stretching resources to their limits humans are ingenious at inventing cost-effective technological fixes that permit economic growth to continue.
33		PolOr5	The rise of China to superpower status will inevitably entail sharp conflicts with the United States.
34		PolOr6	The rise of radical Islam will be short lived, and pragmatic forces will prevail in contested areas.
35		PolOr7	I doubt that global climate change modelers know as much about climate trends as they claim.
36		PolOr8	European monetary integration should be scaled back sharply.
37		PolOr9	On political and economic issues, I am more liberal than conservative.
38		PolOr10	Government should routinely intervene in the economy to achieve fairer outcomes.
39		PolOr11	Free markets function well with minimal government intervention.
40		PolOr12	I would rather be wrong in an interesting way than right in an uninteresting way.

Table 8: Case study candidate fixed effects variables.

References

D. M. Bates. lme4: Mixed-effects modeling with r. *URL http://lme4. r-forge. r-project. org/book*, 2010.

H. D. Bondell, A. Krishna, and S. K. Ghosh. Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*, 66(4):1069–1077, 2010.

N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.

Z. Chen and D. B. Dunson. Random effects selection in linear mixed models. *Biometrics*, 59(4):762–769, 2003.

B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456): 1348–1360, 2001.

Y. Fan and R. Li. Variable selection in linear mixed effects models. *Annals of statistics*, 40(4):2043, 2012.

J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

A. Groll. *glmmLasso: Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation*, 2017. URL `https://CRAN.R-project.org/package=glmmLasso`. R package version 1.5.1.

A. Groll and G. Tutz. Variable selection for generalized linear mixed models by l 1-penalized estimation. *Statistics and Computing*, pages 1–18, 2014.

P. J. Heagerty and B. F. Kurland. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, 88(4):973–985, 2001.

F. K. Hui, S. Mueller, and A. Welsh. *rpql: Regularized PQL for Joint Selection in GLMMs*, 2016a. URL `https://CRAN.R-project.org/package=rpql`. R package version 0.5.

F. K. Hui, S. Müller, and A. Welsh. Joint selection in mixed models using regularized pql. *Journal of the American Statistical Association*, (just-accepted), 2016b.

J. G. Ibrahim, H. Zhu, R. I. Garcia, and R. Guo. Fixed and random effects selection in mixed effects models. *Biometrics*, 67(2):495–503, 2011.

J. Kiefer. Sequential minimax search for a maximum. *Proceedings of the American mathematical society*, 4(3):502–506, 1953.

C. E. McCulloch, J. M. Neuhaus, et al. Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical Science*, 26(3):388–402, 2011.

J. Schelldorfer. *lmmlasso: Linear mixed-effects models with Lasso*, 2011. URL `https://CRAN.R-project.org/package=lmmlasso`. R package version 0.1-2.

J. Schelldorfer, P. Bühlmann, and S. Van de Geer. Estimation for high-dimensional linear mixed-effects models using l1-penalization. *Scandinavian Journal of Statistics*, 38(2):197–214, 2011.

S. R. Searle, G. Casella, and C. E. McCulloch. *Variance Components*, volume 242. John Wiley and Sons, New York, 1992.

S. A. Van de Geer. High-dimensional generalized linear models and the lasso. *The Annals of Statistics*, pages 614–645, 2008.

H. Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.