Generating and visualizing predictor effects???

September 23, 2021

1 Introduction

In many applications, predictions of expected responses, or response probabilities, are often of interest. Predictions provide a great way to summarize what the regression model is telling us and are very useful for interpreting and visualizing models estimates using graphs. For example, in logistic regression models, the coefficient estimates are usually not easy to interpret, and we may want to explore the "effects" of covariates on predicted probabilities, or we may require predictions of responses at various levels of covariates. For example, as a way to plan, public health officials may want to know the predicted probability of improved water services in slums on the basis of the dwellers' household size, income, wealth index, etc.

Both simple and generalized linear (mixed) models (GL(M)Ms) can examine very complex relationships, including nonlinear relationships between response and predictors, interactions between predictors (and via splines), and nonlinear transformations via link functions due to their flexibility. This flexibility comes at a cost, for example, complex multivariate models may risk misinterpretation, and miscalculation of quantities of interest. Also, coefficient estimates of models involving nonlinear link functions or interactions lose their direct interpretation [Leeper, 2017], meaning that interpretation of derived quantities from these estimates requires some understanding of the specified model.

When visually presented, predictions provide a unified and intuitive way of describing relationships from a fitted model, especially complex models involving interaction terms or some kind transformations on the dependent variables whose estimates are usually, but not always, a subject to less clarity

of interpretation. However, generating predictions and associated confidence intervals for regression models has a number of challenges. In particular:

1. choice of representative values of *focal* variable(s) and appropriate 'model center' for *non-focal* variables in the case of multiple predictors

2. propagation of uncertainty ([how] can we incorporate uncertainty in nonlinear components of (G)LMMs? Should we exclude variation in non-focal parameters?)

Most common way of dealing with the first challenge is taking unique levels of the focal predictor if discrete or taking appropriately sized quantiles (or bins) if continuous, and then calculating the predictions while holding non-focal predictors at their mean values. There a number of variations on description of this quantity – *predictor effect* [Fox and Hong, 2009], *marginal predictions* [Leeper et al., 2017] and *estimated marginal means* [Lenth and Lenth, 2018]. In this article, we refer to this quantity as *variable (effect?) predictions* since this quantity should, for example, tell us what we would expect the outcome to be at a particular value or level of the variable. An alternative, is the population-based approach which involves computing the prediction over the population of the non-focal predictors and then averaging across the values of the focal predictor. We give special focus to the second challenge later in the article.

When dealing with nonlinear link functions, the correct predictions for example, are even much harder to estimate. One approach is to make predictions on the transformed scale (linear predictor scale), and then back-transform back to the original scale. However, the back-transformation may either result in biased predictions or requires some approximation. In particular, Jensen's inequality and bias in expected mean prediction induced by nonlinear transformation of the response variable can lead inaccurate predictions. The main purpose of this article is to discuss and implement various approaches for computing variable predictions and how to correctly compute the associated confidence intervals. We further explore approaches for correcting bias in predictions for GL(M)Ms involving nonlinear link functions. The outline of this article is as follows: Section 2 provides formal definition various quantities of interest, Section 3 provides statistical background of estimation of model predictions, Section 4 describes mathematical and computation implementation of the proposed method for uncertainty propa-

gation, and lastly Section 5 describes methods for bias correction in GL(M)M together with the computational implementation.

2 Quantities of interest

Several quantities of interest may be derived from regression models. The first one being the coefficient estimates. Others being predictors, model matrix, predicted values and marginal effects. *predictors* are model input variables and is associated with one or more *variables* (in polynomials, splines and models with interaction). Model matrix refers to the design matrix whose rows include all combination of variables appearing in the interaction terms, along with the "typical" values of the focal and non-focal predictors.

To illustrate some of the concepts, consider results from, hypothetical simulated example, regression of household size as function of household wealth index and age of household head, and the second model involving these variables plus the interaction between them.

Table 1: Example of simple linear model output

	Dependent variable:	
	hhsize	
	(1)	(2)
age	0.997*** (0.032)	0.996*** (0.032)
wealthindex	1.986*** (0.031)	1.993*** (0.032)
age:wealthindex		−0.023 (0.031)
Constant	1.538*** (0.033)	1.539*** (0.033)
Observations	1,000	1,000
R^2	0.841	0.841
Adjusted R^2	0.841	0.841
Residual Std. Error	0.999 (df = 997)	1.000 (df = 996)
F Statistic	2,635.024*** (df = 2; 997)	1,756.031*** (df = 3; 996)

Note: *p<0.1; **p<0.05; ***p<0.01

In simple linear models with no interaction terms, the default out for the coefficient estimates in Table 1 model (1) is simple and directly interpretable as

the expected change in household size for a unit change in age. In particular, a unit change in age is associated with a household that is 0.997 bigger. This is the *unconditional marginal effect* [Leeper, 2017] and it is constant across all the observations and levels of all other variables.

Figure 1 shows the comparison of *variable prediction* and unconditional marginal effect of effect of age on household size, based on model estimates in Table 1. Since the model has no interactions, the relationship between the predicted household size and age is linear, hence the marginal effect of age is the slope ($\frac{\Delta \text{hhsize}}{\Delta \text{age}}$) of the variable effect line and can be calculated irrespective of the values of wealth index.
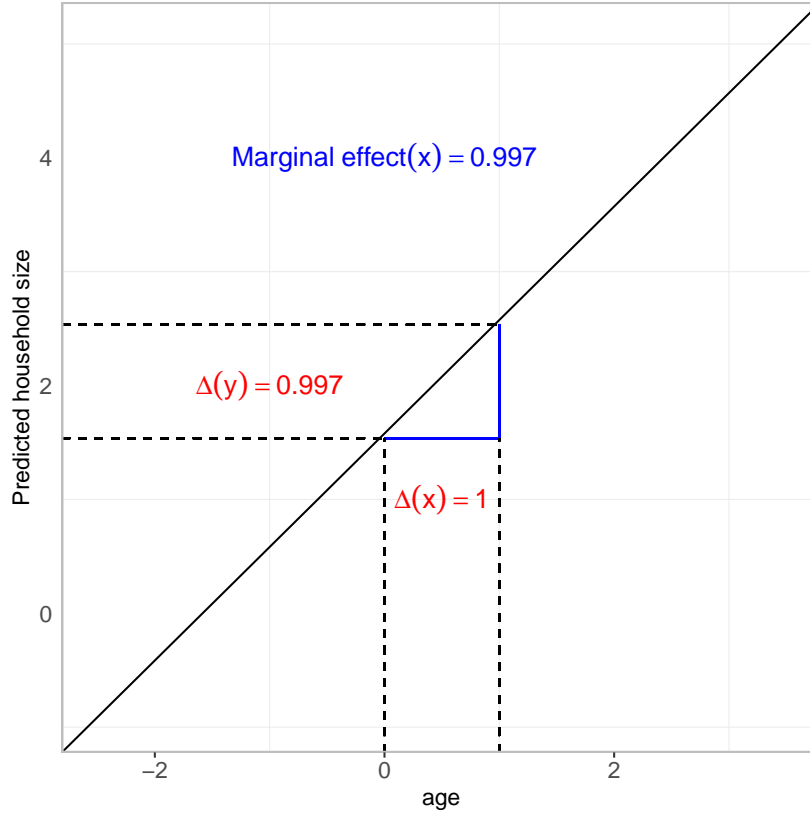

Figure 1: Variable effect plot and marginal effect of age of household head on household size. For a linear model with no interaction, the marginal effect is the slope of the variable effect line.

4

Variable effect, on the other hand, is the expected household size for a particular age, holding wealth index at its mean. In particular, the purpose and goal of a variable effect seems fairly straightforward; for specified values of (a) focal predictor(s), we want to give a point estimate and confidence intervals for the prediction of the model for a "typical" (= random sample over the multivariate distribution of non-focal parameters) individual with those values of the predictors.

On the other hand, when model specification involves other kinds of terms such as variable interactions, log and power transformations, etc., the coefficient estimates cannot easily and directly communicate the relationship between the outcome and the independent variable of interest because of the multiple coefficients of the same variable. For instance, in model (2), the interaction between age and wealth index is added, hence the coefficient estimate for age (0.996) can only be regarded as unconditional marginal effect when effect of wealth index (and thus `age:wealthindex`) is zero, which is never the case. In other words, for models involving interactions terms, looking at a single estimated coefficient in isolation and failing to take into account the estimates involving the interaction terms may lead to inaccurate interpretations [Brambor et al., 2006, Leeper, 2017].

3 Statistical background

To get an intuition of how conditioning on the mean values of the non-focal predictors work, suppose we are interested in the variable predictions of a particular predictor (hence forth referred as *focal* predictor otherwise *non-focal*), x_f, from the set of predictors. To keep it simple, assume that the model has no interaction terms. Then the idea is to *anchor* the values of non-focal predictors to some particular values. For example fixing the values of *non-focal* predictor(s) at some typical values – typically determined by averaging (for now) in some meaningful way, for example, arithmetic mean for continuous and average over the levels of the factors for categorical non-focal predictors. The easiest way to achieve this is by constructing $\mathbf{X}^\star$ by averaging the columns of non-focal variables in model matrix $\mathbf{X}$, and together with appropriately chosen values of focal predictor(s).

Consider a simple linear model with linear predictor $\eta = \mathbf{X}\boldsymbol{\beta}$ and let $g(\boldsymbol{\mu}) = \boldsymbol{\eta}$ be an identity link function (in the case of simple linear model), where $\boldsymbol{\mu}$ is the expected value of response variable y. Let $\hat{\boldsymbol{\beta}}$ be the estimate of $\boldsymbol{\beta}$, together

with the estimated covariance matrix $\Sigma = V(\hat{\boldsymbol{\beta}})$ of $\hat{\boldsymbol{\beta}}$. Let $\mathbf{X}^*$ be the model matrix, inheriting most of its key properties, for example transformations on predictors and interactions from the model matrix, $\mathbf{X}$. Then the prediction $\hat{\boldsymbol{\eta}}^\star = \mathbf{X}^\star \hat{\boldsymbol{\beta}}$ is the variable effect for the focal predictor in question.

An alternative, save for later, formulation of variable effect involves, expressing the linear predictor as the sum of the focal and non-focal predictor linear predictor. In particular, $\eta^\star(x_f, \bar{x}_{\{n\}}) = \beta_f x_f + \sum \beta_{\{n\}} \bar{x}^\star_{\{n\}}$, where $\bar{x}^\star_{\{n\}}$ are the appropriately averaged entries of non-focal predictors and $x_{\{n\}}$ is a vector of values of the non-focal predictors for a particular observation.

Generally, in generalized linear models with other link functions, the predictions are first generated in linear predictor scale, $\hat{\boldsymbol{\eta}}^*$, and then transformed to the original response scale using $g^{-1}(\hat{\boldsymbol{\eta}}^*)$.

4 Uncertainty propagation

What about the confidence intervals (CI)? The limits of the confidence intervals are points, not mean values. In principle, every observation/value of focal predictors has a different CI. The traditional way to compute variances for predictions is $\sigma^2 = \text{Diag}(\mathbf{X}^\star \Sigma \mathbf{X}^{\star\top})$ [Lenth and Lenth, 2018, Fox and Hong, 2009], so that the confidence intervals are $\eta \pm q\sigma$, where q is an appropriate quantile of Normal or t distribution. This approach incorporates all the uncertainties – including the uncertainty due to non-focal predictors. But what if we are only interested in the uncertainty as a result of the focal predictor, so that the confidence intervals are $\eta \pm q\sigma_f$?

Currently, commonly used R packages for constructing predictions do not exclude the uncertainties resulting from the *non-focal* predictors when computing the CIs. A non-trivial way to exclude uncertainties associated with non-focal predictors in some of these packages is to provide a user defined variance-covariance matrix with the covariances of *non-focal* terms set to 0 – *zeroing-out* variance-covariance matrix. This only works when the input predictors are *centered* prior to model fitting, in case of numerical predictors, and even much complicated when the predictors are categorical. We first describe this variance-covariance based approach and then discuss our proposed method which is based on *model centering* and does not require input predictors to be scaled prior to model fitting. Model center is found by averaging the variables over the model matrix, $\mathbf{X}^\star$.

4.1 Variance-covariance

The computation of $\hat{\eta}^{\star}$ remains the same as described above. However, to compute σ, Σ is modified by *zeroing-out* (the variance-covariance of all non-focal predictors are assigned zero) variances of non-focal terms. Although this is the simplest approach, it requires centering of continuous predictors prior to model fitting and proper handling of categorical predictors. One way to center predictors is to create a design matrix, $\mathbf{X}$, and then compute $\mathbf{X}_c = \mathbf{X} - \bar{\mathbf{X}}$.

4.2 Centered model matrix

Consider centered model matrix $\mathbf{X}_c^{\star} = \{\mathbf{X}_f^{\star}, \mathbf{X}_{\{n\}}^{\star} - \bar{\mathbf{X}}_{\{n\}}^{\star}\}$. It follows that the non-focal terms in $\mathbf{X}_c^{\star}$ are all zero. Consequently the uncertainty due to non-focal predictors are all zeroed-out in the computation of σ. More generally, centered design matrix, $\mathbf{X}_c^{\star}$, impacts on the estimated value of the intercept and its associated variance but not the slopes. Thus since non-focal terms in $\mathbf{X}_c^{\star}$ are all zero, it does not matter what their corresponding values are in the variance-covariance matrix. Hence, we can compute CIs from non-centered predictors (in other words, fitted models with predictors in their natural scales).

4.3 Handling higher order terms

In models containing higher order terms such as interactions, splines, polynomials, etc., we adopt the method implemented in **effects** package [Fox and Hong, 2009] which involves combining the interactions with the main effects marginal to the interaction.

4.4 Simulation examples

In this section, we illustrate variable effects (together with the associated CI) in the context of linear models involving continuous and categorical predictors with or without interactions. We compare the expected to the known simulation "truth".

4.5   Continuous predictors

Consider simulation example described in Section 2, with slight modifications to incorporate interaction between predictors. The first case, is the model with no interaction:

$$\text{hh size} = \beta_0 + \beta_{\text{A}}\text{Age}_i + \beta_{\text{W}}\text{Wealthindex}_i + \epsilon_i$$
$$\text{Age}_i \sim \text{Normal}(0.2, 1)$$
$$\text{Wealthindex}_i \sim \text{Normal}(0, 1)$$
$$\epsilon_i \sim \text{Normal}(0, 1)$$
$$\beta_0 = 1.5$$
$$\beta_{\text{A}} = 0.2$$
$$\beta_{\text{W}} = 2$$
$$i = 1, \cdots, 1000$$

Secondly, we add interaction terms between the non-focal predictors (`wealth index` and `household expenditure`) but not with the focal predictor `age`:

$$\text{hh size} = \beta_0 + \beta_{\text{A}}\text{Age}_i + \beta_{\text{W}}\text{Wealthindex}_i + \beta_{\text{E}}\text{Expenditure}_i$$
$$+ \beta_{\text{WE}}\text{Wealthindex}_i * \text{Expenditure}_i + \epsilon_i$$
$$\text{Age}_i \sim \text{Normal}(0.2, 1)$$
$$\text{Wealthindex}_i \sim \text{Normal}(0, 1)$$
$$\text{Expenditure}_i \sim \text{Normal}(0, 1)$$
$$\epsilon_i \sim \text{Normal}(0, 1)$$
$$\beta_0 = 1.5$$
$$\beta_{\text{A}} = 0.8$$
$$\beta_{\text{W}} = 2$$
$$\beta_{\text{E}} = 1.5$$
$$\beta_{\text{WE}} = 1$$
$$i = 1, \cdots, 1000$$

And lastly, we now consider model with interaction terms between the focal predictor (`age`) and one of the non-focal predictor (`wealth index`):

$$\text{hh size} = \beta_0 + \beta_A \text{Age}_i + \beta_W \text{Wealthindex}_i + \beta_E \text{Expenditure}_i$$
$$+ \beta_{AW} \text{Age}_i * \text{Wealthindex}_i + \epsilon_i$$
$$\text{Age}_i \sim \text{Normal}(0.2, 1)$$
$$\text{Wealthindex}_i \sim \text{Normal}(0, 1)$$
$$\text{Expenditure}_i \sim \text{Normal}(0, 1)$$
$$\epsilon_i \sim \text{Normal}(0, 1)$$
$$\beta_0 = 1.5$$
$$\beta_A = 0.2$$
$$\beta_W = 2$$
$$\beta_E = 1.5$$
$$\beta_{AW} = 1$$
$$i = 1, \cdots, 1000$$

We first fitted all the three models and in each case, compared three different variable effects together with the corresponding CIs – 1) uncertainty due to non-focal predictors included (everything), 2) uncertainty due to non-focal predictors removed using variance-covariance matrix (zero-vcov), and 3) uncertainty due to non-focal predictors removed using centered model matrix (centered mm). The results are summarised in Figure 2.

Figure 2a represents variable effects for the model with no interaction, Figure 2b shows variable effects for model with non-focal predictors interacting with each other, and Figure 2c shows variable effects for the model with focal and non-focal predictors interacting.

For perfect predictions, we expect the yellow and grey dotted lines to overlay each other and they should intersect with the black solid line and the vertical dotted grey line at the same point (or very close), Figure 2a. In other words, at the model center (mean of the focal predictor), generally, we expect the narrower CIs. If we properly anchor the model at its center (or any other appropriate value, assuming no interactions), we would expect the variance at the anchor to be zero, i.e., the CIs bands crosses at that point (see zero-cov and centered mm in Figure 2). In Figure 2b, the predicted overall mean is slightly higher than the observed mean (yellow horizontal line is slightly above the horizontal grey line), and vice versa in Figure 2c.
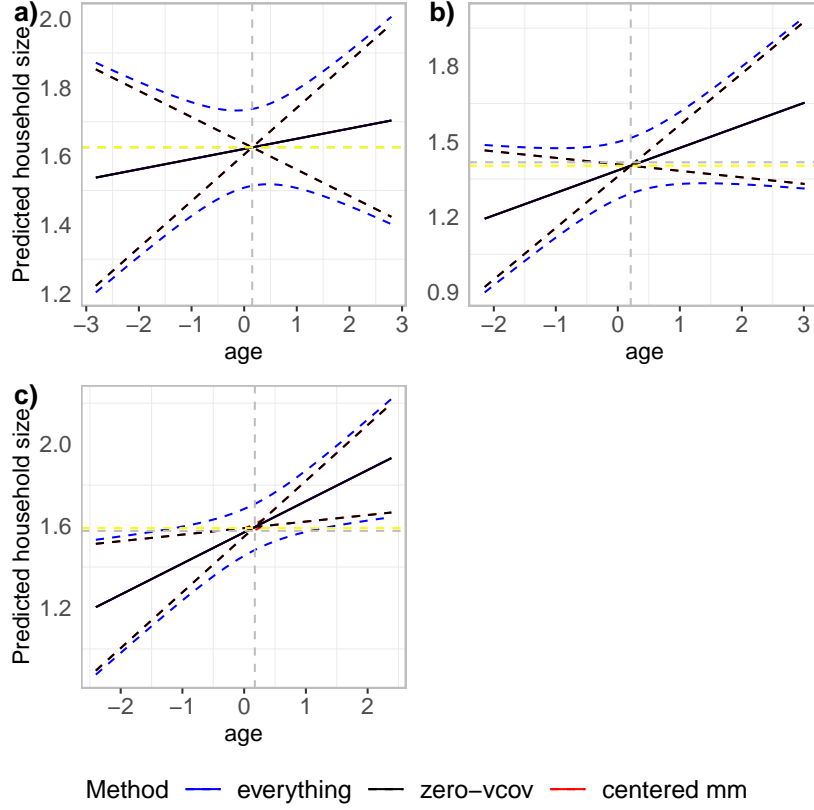
9

Figure 2: A comparison of variable effects and their corresponding CIs. Figure a) shows model with no interaction, Figure b) shows model with non-focal predictors interacting with each other, and Figure c) shows model with focal and non-focal predictors interacting. The dotted yellow and grey horizontal lines are the expected and observed/true means, respectively. The vertical grey line is the mean of the focal predictor (model center). The solid black, red (non-focal uncertainty excluded) and blue lines (all uncertainty included) are the variable effects while their corresponding CIs are the dotted black, red and blue lines, respectively.

5   Bias correction

References

Thomas Brambor, William Roberts Clark, and Matt Golder. Understanding interaction models: Improving empirical analyses. *Political analysis*, 14 (1):63–82, 2006.

John Fox and Jangman Hong. Effect displays in r for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software*, 32(1):1–24, 2009.

Thomas J Leeper. Interpreting regression results using average marginal effects with r's margins. *Reference manual*, 32, 2017.

Thomas J Leeper, Jeffrey Arnold, Vincent Arel-Bundock, and Maintainer Thomas J Leeper. Package 'margins'. *accessed December*, 5:2019, 2017.

Russell Lenth and Maintainer Russell Lenth. Package 'lsmeans'. *The American Statistician*, 34(4):216–221, 2018.