Visualizing Fit and Lack of Fit in Complex Regression Models with Predictor Effect Plots and Partial Residuals

John Fox
McMaster University

Sanford Weisberg
University of Minnesota

Abstract

Predictor effect displays, introduced in this article, visualize the response surface of complex regression models by averaging and conditioning, producing a sequence of 2D line graphs, one graph or set of graphs for each predictor in the regression problem. Partial residual plots visualize lack of fit, traditionally in relatively simple additive regression models. We combine partial residuals with effect displays to visualize both fit and lack of fit simultaneously in complex regression models, plotting residuals from a model around 2D slices of the fitted response surface. Employing fundamental results on partial residual plots along with examples for both real and contrived data, we discuss and illustrate both the strengths and limitations of the resulting graphs. The methods described in this paper are implemented in the **effects** package for R.

Keywords: interaction, nonlinearity, model misspecification, component plus residual plot, R, **effects** package.

1. Introduction

Predictor effect displays, a reinterpretation of effect displays introduced by Fox (1987) for generalized linear models, visualize the response surface of complex regression models with a linear predictor that includes main effects and interactions by averaging and conditioning, producing a sequence of 2D line graphs for the predictors in a model. *Partial residual plots*, also called *component plus residual plots*, visualize lack of fit, traditionally in relatively simple additive regression models. The properties of partial residuals plots were systematically explored by Cook (1993) and Cook and Croos-Dabrera (1998).

In the first part of this article we describe predictor effect displays, which require one or more 2D line graphs to describe the dependence of a fitted regression surface on each predictor.

This approach corresponds closely to the way most analyses are traditionally summarized based on tests and estimates. We then show how to combine partial residuals with predictor effect displays to visualize both fit and lack of fit simultaneously in complex regression models, plotting residuals from a model around 2D slices of the fitted response surface. Referencing Cook's fundamental results, we discuss and illustrate both the strengths and limitations of the resulting graphs. The extension to predictor effect displays is implemented for linear and generalized linear models of arbitrary complexity in the current version of the **effects** package for R (Fox 2003; Fox and Hong 2009; Fox, Weisberg, Friendly, and Hong 2018a), which we use to generate the illustrations in the paper. As summarized in Section 5, predictor effect displays have been extended to a wide variety of other models that include a linear predictor in the mean function.

Section 2 of the paper describes the general setting that we address and introduces predictor effect displays. We also discuss the relationship of predictor effect displays to *term effect displays*, as previously described by Fox (1987). Section 3 reviews partial residual plots, connecting them to predictor effect displays. Section 4 develops a variety of examples, using both real and contrived data, to explore the utility and limitations of adding partial residuals to effect displays. The paper concludes in Section 5 with advice about using partial residuals in effect displays to explore lack of fit in complex regression models, and compares our approach to related work.

2. Predictor effect displays

We address the following situation: There is a *response y* and a set of *p predictors* $\mathbf{x} = (x_1, \ldots, x_p)$, along with a regression model for the conditional mean $\mathsf{E}(y|\mathbf{x})$. Predictors in a parametric regression model are represented by *regressors*. For example, if x_j is a factor with k levels, then a *main effect* for x_j would be represented by $k - 1$ indicator or contrast regressors. A numeric x_j can be represented by x_j itself, by a transformation such as $\log(x_j)$, by a set of polynomial basis functions, by a spline basis, or perhaps by other regressors. The correspondence between predictors and regressors is not unique, but the methods we discuss are invariant under changes in parameterization. As is conventional, we define an *interaction term x_j:$x_{j'}$* to be the set of all pairwise products of the regressors that are derived from x_j with all the those derived from $x_{j'}$. This definition extends straightforwardly to interactions of more than two predictors, such as the three-way interaction x_j:$x_{j'}$:$x_{j''}$.

We define the *linear predictor $h(\boldsymbol{\beta}, \mathbf{x})$* to be a linear combination of regressors in the main effects and interactions created from the predictors $\mathbf{x}$, with the regression coefficients $\boldsymbol{\beta}$ providing the weights that multiply the regressors. An intercept β_0 is generally included in h with a corresponding constant regressor, that is, a column of ones. We consider only mean functions of the form

$$\mathsf{E}(y|\mathbf{x}) = \eta^{-1}\left[h(\boldsymbol{\beta}, \mathbf{x})\right] \tag{1}$$

for some known invertible *link function η*. This class of regression models includes linear and generalized linear models, additive and generalized additive models, as well as linear and generalized linear mixed models, among others.

Given a suitable estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, we write $\widehat{y}(\mathbf{x}) = \eta^{-1}[h(\widehat{\boldsymbol{\beta}}, \mathbf{x})]$ as the estimated mean function. The goal is to visualize the dependence of $h(\widehat{\boldsymbol{\beta}}, \mathbf{x})$ or of $\widehat{y}(\mathbf{x})$ on $\mathbf{x}$. The most general approach would examine a single high-dimensional display with $h(\widehat{\boldsymbol{\beta}}, \mathbf{x})$ or $\widehat{y}(\mathbf{x})$ on the "vertical" axis

and $\mathbf{x}$ on the "horizontal" axes. Although concentrating on predictors rather than regressors has reduced the dimension of the visualization problem from approximately the number of linearly independent regressors to approximately the number of linearly independent predictors plus one for the response, this graph is likely to be useful only for $p \leq 2$.

Standard practice in summarizing a regression model is to proceed predictor-by-predictor. Predictors that occur in main effects only are generally summarized by statements or estimates or tests that essentially average over, or conditionally fix, all other predictors. Predictors that occur in interactions require a more complex summary that conditions successively on the combinations of values of the other predictors in the interactions. Predictor effect displays follow this paradigm.

Suppose that we are interested in the visual summary of a particular *focal predictor* x_f in the set of predictors. We assume that the formula defining the linear predictor is hierarchical, meaning that if an interaction is present in the model, then so are all of its lower-order relatives, equivalent to the *principle of marginality* (Nelder 1977). For example, the inclusion of $x_j{:}x_{j'}$ implies that both x_j and $x_{j'}$ are in the formula. We can then partition the set of predictors $\mathbf{x} = (x_f, \mathbf{x}_1, \mathbf{x}_2)$, where $x_j \in \mathbf{x}_1$ if $x_f{:}x_j$ is in the model formula, and the subvector $\mathbf{x}_2$ contains all the remaining predictors. Either of $\mathbf{x}_1$ or $\mathbf{x}_2$ may be empty.

For a given x_f, we can always fix the values of the predictors in $\mathbf{x}_2$, if any, and plot, in linear predictor scale,

$$h\left[\widehat{\boldsymbol{\beta}}, (x_f, \mathbf{x}_1, \mathbf{x}_2^a)\right] \text{ versus } x_f \tag{2}$$

where $\mathbf{x}_2^a$ is a fixed value of $\mathbf{x}_2$, typically determined by averaging in some meaningful way. In the **effects** package, we use by default the arithmetic average for continuous predictors. For factors, we average by default over the levels of the factor, with weights given by the sample sizes at each level; this procedure is equivalent to averaging the columns of the model matrix encoding a conditionally fixed factor and is therefore invariant with respect to contrast coding. These defaults can be changed, for example to set continuous predictors equal to some meaningful value, or to average a factor over its levels using a different weighting scheme. Then the vertical values in (2) are simply

$$h\left[\widehat{\boldsymbol{\beta}}, (x_f, \mathbf{x}_1, \mathbf{x}_2^a)\right] = C + h\left[\widehat{\boldsymbol{\beta}}, (x_f, \mathbf{x}_1, \mathbf{0})\right] \tag{3}$$

for some constant C that depends on $\mathbf{x}_2^a$. Thus, choice of $\mathbf{x}_2$ affects only the *height* of the predictor effect display in linear predictor scale for x_f, but not its *shape*, and $\mathbf{x}_2$ is therefore generally unimportant for examination of the effect of x_f. In mean scale, conditioning is not entirely benign if the link function is nonlinear, as the shape of the plot can depend on $\mathbf{x}_2^a$. Understanding these plots is therefore generally simpler in linear predictor scale.

In the important special case of $\mathbf{x}_1 = \emptyset$, the empty set, x_f appears in the formula only through a main effect. If the regressor representing x_f is x_f itself, then the predictor effect display in linear predictor scale for x_f is a straight line with slope equal to the estimated coefficient corresponding to x_f, and hence the plot merely displays this estimated slope, along with an essentially arbitrary intercept. If x_f is represented some other way, for example by a transformation such as $\log(x_f)$, a polynomial, a smooth estimated using an additive or generalized additive model, or a spline basis, then the display will visualize the appropriate nonlinear effect of x_f in the linear predictor scale.

If, however, $\mathbf{x}_1$ is *not* empty, then the plot described by (2) is as a practical matter inadequate because it describes a graph with $1 + \dim(\mathbf{x}_1)$ "vertical" axes and one horizontal axis. To

reduce this high-dimensional graph to a sequence of 2D graphs, we invoke *conditioning*: For each $x_j \in \mathbf{x}_1$ define a *grid* of a few values in the range of x_j. If x_j is a factor, then the "grid" typically consists of all factor levels, while for continuous x_j, selected quantiles or values evenly spread over the range of x_j can be used to form the grid. If the jth predictor has G_j grid points, there are then $G = \prod G_j$ combinations of grid values of the predictors in $\mathbf{x}_1$.

Let $\mathbf{x}_1^g$ be one of the G sets of grid values. Then the corresponding graph in the predictor effect display for x_f is of

$$h\left[\widehat{\boldsymbol{\beta}}, (x_f, \mathbf{x}_1^g, \mathbf{x}_2^a)\right] \text{ versus } x_f. \tag{4}$$

The predictor effect display in its entirety consists of the sequence of separate 2D line graphs of (4) for each of the G choices of g. Often visualization can be simplified by overlaying some of these 2D line graphs on the same plot, creating a *multi-line* display.

2.1. Example: Infant mortality by per-capita GDP and national group

We begin by loading the **effects** package:

```
R> library("effects")


Loading required package: carData
lattice theme set by effectsTheme()
See ?effectsTheme for details.
```

Loading the **effects** package also loads the **carData** package (Fox, Weisberg, and Price 2018b), which contains a variety of regression data sets, and, if the **lattice** package (Sarkar 2008) is not loaded, sets a custom theme for lattice graphics. On some platforms, setting the lattice theme may open a trellis graphics device (see `?trellis.device`).

To develop a simple example of predictor effect displays, we use the `UN` data set in the **carData** package. UN member states and observer states were divided into three groups – African states, OECD states, and other non-African states. The response variable in the example is `infantMortality`, the infant mortality rate (infant deaths per 1000 live births) for each country, and the predictors are `ppgdp`, per-person GDP in U.S. dollars, and `group`. The data are from approximately 2011.

We want to visualize the fit of a model for infant mortality as a function of per person GDP and the three national groups, permitting `ppgdp` to interact with `group`:

```
R> m1 <- lm(log(infantMortality) ~ group * log(ppgdp), data = UN,
+    subset = rownames(UN) != "Equatorial Guinea")
R> summary(m1)


Call:
lm(formula = log(infantMortality) ~ group * log(ppgdp), data = UN,
    subset = rownames(UN) != "Equatorial Guinea")

Residuals:
     Min       1Q   Median       3Q      Max
-1.11823 -0.27109  0.02494  0.28529  1.14925
```

```
Coefficients:
                        Estimate Std. Error t value   Pr(>|t|)
(Intercept)              6.47641    1.30955   4.946 0.00000169
groupother               0.94287    1.34496   0.701   0.484155
groupafrica             -0.04564    1.36583  -0.033   0.973379
log(ppgdp)              -0.48455    0.12617  -3.840   0.000168
groupother:log(ppgdp)   -0.05885    0.13110  -0.449   0.654028
groupafrica:log(ppgdp)   0.14201    0.13766   1.032   0.303613

Residual standard error: 0.4413 on 186 degrees of freedom
  (20 observations deleted due to missingness)
Multiple R-squared:  0.8391,        Adjusted R-squared:  0.8348
F-statistic: 194.1 on 5 and 186 DF,  p-value: < 2.2e-16
```

This linear model has two predictors: the factor `group`, with three levels, and the numeric variable `ppgdp`. Both the response variable and `ppgdp` are log-transformed to linearize the partial relationship between the two, a point to which we return in Section 3. The regressors in the model include log(`ppgdp`) to represent `ppgdp`, two indicator regressors for the levels of `group`, and two product regressors for the interactions. Because the linear model uses the identity link function, the mean function and linear predictor are the same. In fitting the model to the data, we removed the African country Equatorial Guinea, for a reason that will become apparent as we further develop this example in Section 3.1.

The interactions and log transformations make it inconvenient to interpret the model directly from the coefficients, and so we turn to predictor effect plots for `group` and `ppgdp`, invoking the `predictorEffect` function in the **effects** package:

```
R> plot(predictorEffect("group", m1,
+    transformation = list(link = log, inverse = exp),
+    xlevels = list(ppgdp = 10 ^ (2 : 5))),
+    lines = list(multiline = TRUE), axes = list(x = list(rotate = 45),
+    y = list(lab = "Infant Mortality", ticks = list(at = 2 ^ (1 : 8)))),
+    confint = list(style = "auto"))
R> plot(predictorEffect("ppgdp", m1,
+    transformation = list(link = log, inverse = exp)),
+    lines = list(multiline = TRUE), axes = list(x = list(rotate = 45),
+    y = list(lab = "Infant Mortality", ticks = list(at = 2 ^ (1 : 8)))),
+    confint = list(style = "auto"))
```

The objects returned by `predictorEffect` are graphed by a corresponding `plot` method. The calls to `predictorEffect` and `plot` use several optional arguments to customize the resulting graphs:

- The `transformation` argument to `predictorEffects` reverses the log transformation of the response variable, to express the infant-mortality rate as deaths per 1000 births. The effect is *plotted* on the log(`infantMortality`) scale, where the structure of the model is linear, but the axis is *labeled* on the untransformed scale. Adding `type =`
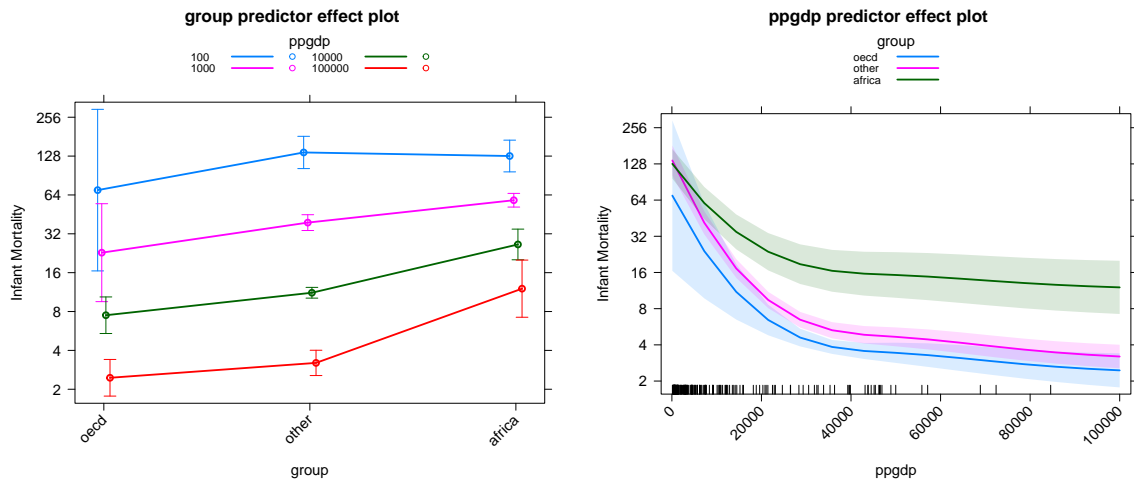
Figure 1: Predictor effect displays for the model for `infantMortality` fit to the UN data.

`"response"` to the y-axis specification list would plot the effect on the scale of the response, producing a less informative graph (try it!).

- The `xlevels` argument to `predictorEffect` sets the values to which the predictor `ppgdp` is conditionally fixed in the predictor effect display for `group`. The default is to evaluate the numeric predictor `ppgdp` on a grid of five approximately equally spaced values rounded to "nice" numbers. By supplying the `ppgdp` grid directly, we can use values that are evenly spaced on the log scale of the regressor log(`ppgpd`) rather than the default of equally spaced values in the arithmetic scale of the predictor `ppgdp`.

- The `lines` argument to `plot` specifies a multi-line graph for each predictor effect; the default is to draw separate panels for each grid value of the conditioning predictor, or combination of grid values of conditioning predictors when there are more than one.

- The `axes` argument rotates the horizontal-axis tick-labels, changes the label on the vertical axis to reflect untransformed infant mortality, and customizes the placement of the vertical-axis tick marks.

- The argument `confint = list(style = "auto")` displays 95-percent point-wise confidence intervals for the fitted effects, using error-bars for factors and bands for numeric predictors. The default in multi-line displays is to suppress confidence intervals.

See `?predictorEffect`, `?Effect`, and `?plot.eff` for details of these and other optional arguments.

The predictor effect displays are shown in Figure 1. The left panel has the focal predictor `group` on the horizontal axis. The remaining predictor, the numeric predictor `ppgdp`, interacts with `group` and hence is evaluated at the supplied grid of four values equally spaced on the log scale, with a separate line drawn for each of the grid values of `ppdgp`. Because `group` is a factor, $\widehat{y}(\mathbf{x})$, which is equivalent to the linear predictor because the link function η is the identity link, is computed only at the factor levels, indicated by the plotting symbols, which are slightly displaced horizontally to avoid overplotting. The lines joining the plotting

symbols are an aid for viewing the graph, and may be suppressed if desired. We see that, in general, infant mortality at fixed levels of `ppgdp` is lowest in the `oecd` group and highest in `africa`; at the lowest level of `ppgdp`, however, fitted `infantMorality` is slightly lower in `africa` than in the `other` group, and the confidence interval for the `oecd` group is very wide, because there are no `oecd` countries at this level of `ppgdp`. In all three groups, infant mortality declines with `ppgdp`, though less so in `africa` than in the other groups.

The display in the right panel of Figure 1 is for the effect of `ppgdp`, with separate lines for the three groups of states overlaid on the same graph. The lines are curved because `ppgdp` is represented by the regressor log(`ppgdp`) in the model. By default, a *rug plot*, showing the marginal distribution of `ppgdp`, is shown at the bottom of the graph. In this instance the inference is the same from the second plot as it is from the first, namely that infant mortality declines with per-capita GDP in all three groups, though less so in `africa` than in the other two groups, and that except at the lowest levels of `ppgdp`, infant mortality is lowest among the `oecd` states and highest in `africa` at fixed levels of `ppgdp`,

2.2. Term effects versus predictor effects

Previous discussions of effect plots (such as Fox 1987), and previous versions of the **effects** package, develop what might be called high-order term effects, or, for short, *term effects*: Term effect displays are drawn for combinations of predictors corresponding to the high-order terms in a model – that is, terms that are not marginal to any terms in the model.

Consider, for example, the model formula `y ~ a*b + a*c`. We adopt the version of the Wilkinson and Rogers (1973) notation for linear models that is used in S and R (Chambers and Hastie 1992). In this notation, `~` separates the left- and right-hand sides of the model and `*` is the crossing operator, and so, in expanded form, the model is `y ~ 1 + a + b + c + a:b + a:c`, where `y` is the response, `1` represents the intercept, `a`, `b` and `c` are the main effects of the three predictors, and `a:b` and `a:c` are interactions. The high-order terms in the model are `a:b` and `a:c`.

The `allEffects` function applied to a model with this formula produces two plots, one with `a` and `b` as the focal predictors, and the other with `a` and `c` as the focal predictors. The `plot` method for more than one focal predictor uses an algorithm to choose which predictor is plotted on the horizontal axis and which is used as a conditioning variable. If the formula has numeric predictors, then the left-most predictor in the formula is used for the horizontal axis.

For example, if `b` were the only numeric predictor and `a` and `c` were factors, then the term effect plot for `a:b` would average over `c`, have `b` on the horizontal axis, and condition on `a`. The term effect plot for `a:c` would average over `b`, and, for the horizontal axis, would use the factor with the fewest levels or the left-most factor if they have the same number of levels.

Neither of these plots corresponds to a predictor effect plot because they average over, rather than condition on, `c` in the first plot and `b` in the second plot, producing a display that is not invariant in shape with respect to the manner in which the levels of the factor `c` are averaged over in the term effect plot for `a:b`, or the typical value to which the numeric `b` is set in the term effect plot for `a:c`. In contrast, recall that averaging over or fixing the values of predictors in predictor effect plots affects only the height, and not the shape, of the effect. It is largely this invariance property that leads us to prefer predictor effects to term effects.

The most general function in the **effects** package is `Effect`, in which the predictors in an

effect are specified explicitly. `Effect` may be used to produce both predictor and term effect displays, including effect displays for terms that do not appear in the model, such as an interaction higher-order term to those in the model. For example, a plot equivalent to the predictor effect plot for `b` could be obtained with the command

```
R> plot(Effect(c("a", "b", "c"), m), x.var = "b")
```

where `m` is the regression-model object. This specification recognizes that `b` interacts with both `a` and `c`, and the `x.var` argument overrides the default procedure for determining the predictor on the horizontal axis of the graph. Predictor effect plots for the other two predictors are obtained by providing each predictor in turn as the `x.var` argument.

All predictor effect plots can be produced more conveniently using

```
R> plot(predictorEffects(m))
```

The three predictor effect plots in this example are views of the same four-dimensional surface from three different view points. In the preceding infant-mortality example, we used `predictorEffect` in preference to `predictorEffects` to exert finer-grain control over the resulting graphs.

3. Partial residual plots

Whereas predictor effect plots are designed to summarize the conditional effects of each predictor given the others in a correctly specified regression model, partial residual plots are used to visualize misspecification of the mean function attributable to continuous predictors. We begin with a *working model* given by (1) that is potentially misspecified. Suppose that $\mathbf{x}_i$ is the vector of predictors for the ith of n observations in the data, and y_i is the corresponding value of the response. The *estimated working linear predictor* for the ith observation is $h(\widehat{\boldsymbol{\beta}}, \mathbf{x}_i)$, and the corresponding *working residuals* are $e(\mathbf{x}_i) = [y_i - \widehat{y}(\mathbf{x}_i)]\eta'(\mathbf{x}_i)$, where $\eta'(\mathbf{x}_i)$ is the first derivative of η with respect to $\mathsf{E}(y|\mathbf{x}_i)$ (Cook and Croos-Dabrera 1998), which translates from the mean scale to the linear predictor scale. Partial residual plots are always drawn in the linear predictor scale and only for numeric predictors.

Paralleling the development of predictor effect displays, for a numeric focal predictor x_f, we divide the ith vector of observed predictors into $\mathbf{x}_i = (x_{fi}, \mathbf{x}_{1i}, \mathbf{x}_{2i})$. Partial residual plots are traditionally defined only when $\mathbf{x}_1 = \emptyset$. In this case, the partial residual plot for a focal predictor x_f is a graph of n points, the ith of which is

$$e(\mathbf{x}_i) + \left\{ h[\widehat{\boldsymbol{\beta}}, (x_{fi}, \emptyset, \mathbf{0})] - \widehat{\beta}_0 \right\} \text{ versus } \mathbf{x}_{fi} \tag{5}$$

where $\emptyset$ has been inserted as a placeholder for the empty value of $\mathbf{x}_1$. The term in curly braces in (5) is called a *partial regression function* and it represents the component of the fitted mean function that depends on $\mathbf{x}_{fi}$. The working residuals $e(\mathbf{x}_i)$ appear as random scatter around the partial regression function for a correctly specified model.

In certain circumstances, however, the scatter added by the residuals will be systematic. Suppose that in place of the working linear predictor in (1), the "true" linear predictor is

$$h(\boldsymbol{\beta}, \mathbf{x}) + \{t(x_f) - h[\boldsymbol{\beta}, (x_f, \emptyset, \mathbf{0})]\} \tag{6}$$

where $t(x_f)$ is a potentially nonlinear function of x_f. If all the *regressors* (not the *predictors*) in $\mathbf{x}$ are at least approximately linearly related, and the method used to estimate parameters is Fisher consistent, then a smoother fit to the partial residual plot for x_f provides a visualization of t (Cook 1993, Lemma 2.1) and possible misspecification with respect to x_f. Extension from linear models to generalized linear models is provided by Cook and Croos-Dabrera (1998).

Comparing (5) to (2), we can superimpose the partial residuals on the predictor effect plot simply by adding the constant $\widehat{\beta}_0 + C$ to the abscissa of the points in (5). The partial residuals are linearly translated, but nonlinear shapes, the main focus of the partial residuals, are unaffected.

When $\mathbf{x}_1 \neq \emptyset$, the predictor effect display consists of G 2D line plots by conditioning on $\mathbf{x}_1^g$. The points we add to the gth 2D plot are

$$e(\mathbf{x}_i) + h\left[\widehat{\boldsymbol{\beta}}, (x_{fi}, \mathbf{x}_{1i}, \mathbf{x}_2^a)\right] + C \text{ versus } x_f \tag{7}$$

for all i such that $|\mathbf{x}_{1i} - \mathbf{x}_1^g|$ is minimized over g for each element of $\mathbf{x}_1$, and the constant C is chosen to match the intercept in the predictor effect display. Cook's lemma can then be applied to each grid value separately to diagnose unmodelled curvature with respect to x_f separately for each g.

3.1. Example: Infant mortality revisited

Continuing with the UN infant mortality example in Section 2.1, we start with the response variable `infantMortality` and predictor `ppgdp` unlogged. The predictor effect plot for `ppgdp` with partial residuals is shown in Figure 2. We initially leave Equatorial Guinea in the data set.

```
R> m2 <- lm(infantMortality ~ group * ppgdp, data = UN)
R> plot(predictorEffects(m2, ~ ppgdp, partial.residuals = TRUE),
+    axes = list(x = list(rotate = 25), y = list(lim = c(0, 150))),
+    id = list(n = 1))
```

The **effects** package suppresses partial residuals for multi-line plots because of the confusion produced by overlapping residuals for different values of a conditioning predictor. Instead, the residuals are plotted with the lines corresponding to different values of the conditioning predictors, here just the predictor `group`, in separate panels. The blue line in each panel represents the fitted model, with a point-wise 95-percent confidence band shown around the fitted effect. The magenta line in each panel is a *loess nonparametric regression smooth* (Cleveland, Grosse, and Shyu 1992), using a span of 2/3 by default. The argument `id = list(n = 1)` to `plot` identifies the most unusual point in each panel, defined as the point with the largest Mahalanobis distance from the centroid of the points.

The points for `other` and `africa` are a clear mismatch for the fitted line, perhaps suggesting the log-transformations of `infantMortality` and `ppgdp` used in the initial example in Section 2.1. The right-most point in `africa`, for Equatorial Guinea, is out of line with the rest of the African states, pairing a large value of `ppgdp` with a large value of `infantMortality`. The most unusual points in the other panels, Turkey in the `oecd` group and Afghanistan in the `other` group, are not out of line with the other points in their groups to the same degree. While extreme in their values of `ppgdp` and `infantMortality`, these two countries follow the
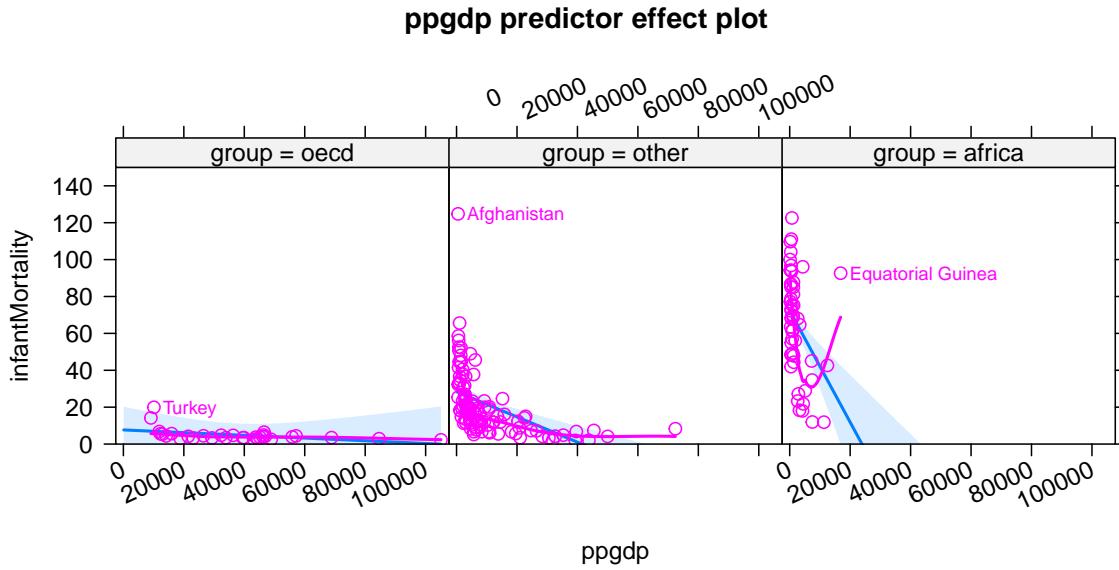
ppgdp predictor effect plot

Figure 2: Predictor effect plot with partial residuals for `ppgdp` when neither it nor the response `infantMortality` is log-transformed, labeling the most unusual point in each panel.

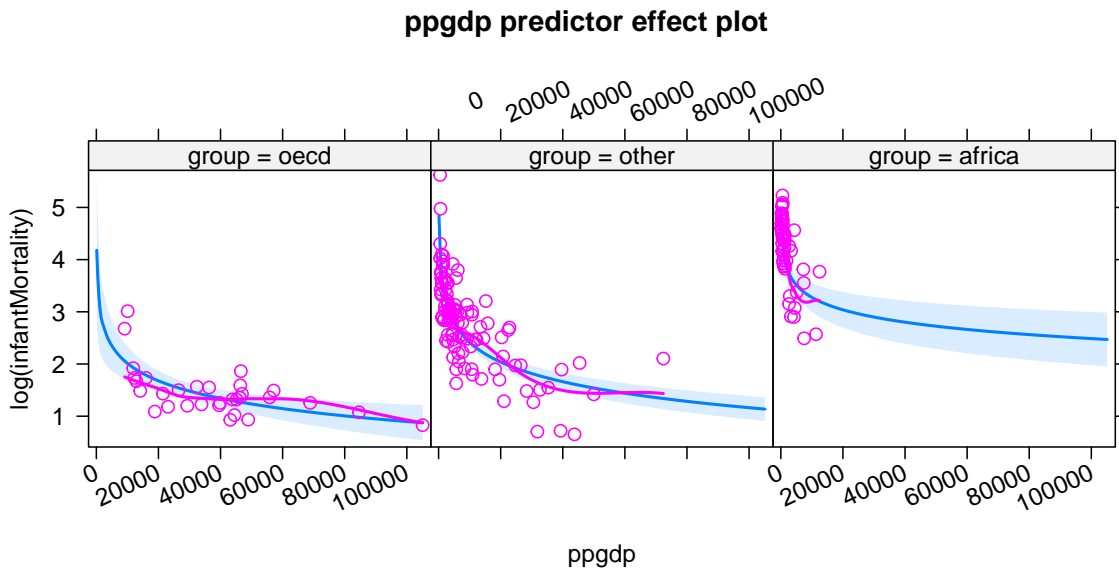**ppgdp predictor effect plot**

Figure 3: Predictor effect plot with partial residuals for `ppgdp` in the model with `ppgdp` and the response `infantMortality` log-transformed.

general pattern of the data. Recall that we removed Equatorial Guinea, but not Turkey or Afghanistan, in the model fit in Section 2.1.

The partial residuals have the added benefit of highlighting that while `ppgdp` is both relatively high and highly variable in the `oecd` group, it has relatively small variation in `africa`, where it is concentrated in very low values. The `other` group is intermediate. When both the response variable and `ppgdp` are log-transformed, as in model `m1` in Section 2.1, the partial-residual

plots are much more satisfactory (see Figure 3):

```
R> plot(predictorEffects(m1, ~ ppgdp, partial.residuals = TRUE),
+    axes = list(x = list(rotate = 25)))
```

3.2. Conditioning on continuous predictors

When $\mathbf{x}_1$ includes continuous numeric predictors, as in the first example in the next section, the assignment of partial residuals to one of the grid of conditioning values introduces additional variation, because the linear predictor is evaluated at $(x_{fi}, \mathbf{x}_{1i}, \mathbf{x}_2^a)$ rather than at the grid value $(x_{fi}, \mathbf{x}_1^g, \mathbf{x}_2^a)$. That is, there is a potential extra source of variability in the plot due to conditioning. If we assume that the value of this difference has a symmetric distribution about zero, then, from Cook (1993, Lemma 2.1), the *unadjusted partial residual plot* (7) visualizes $t(x_f)$ with extra variation. If the difference is *not* symmetrically distributed, as is likely, for example, for extreme values of the continuous predictors in $\mathbf{x}_1$, then bias may be introduced.

A predictor effect display in linear predictor scale with *partial residuals adjusted for conditioning* includes the points given for the gth plot by

$$ e(\mathbf{x}_i) + h\left[\widehat{\boldsymbol{\beta}}, (x_{fi}, \mathbf{x}_1^g, \mathbf{x}_2^a)\right] + C \text{ versus } x_f \tag{8} $$

substituting the grid values $\mathbf{x}_1^g$ for the data values $\mathbf{x}_{1i}$ of the conditioning predictors. This plot also visualizes $t(x_f)$ for each g under the same conditions as the unadjusted version, but the visualization may be sharper. The adjusted version is implemented in the **effects** package.

The requirement of linearly related regressors for the usefulness of partial residual plots may be restrictive on its face, particularly in problems with $\mathbf{x}_1 \neq \emptyset$. Because we are conditioning on $\mathbf{x}_1 = \mathbf{x}_1^g$, however, linearly related regressors are only required within a fixed value of $\mathbf{x}_1$. Moreover, experience suggests that only fairly strong nonlinear relationships among the regressors prove to be problematic.

4. More examples

4.1. Volunteering for a psychological experiment

Cowles and Davis (1987) conducted a study on volunteering for a psychological experiment, in which the subjects were students in an introductory psychology course. The authors of the study collected data on the students' gender, on the personality dimensions extraversion and neuroticism, each of which ranges potentially from zero to 24, and on the students' willingness to volunteer for an experiment. Of the 1421 students for whom data were collected, 597 were willing to serve as volunteers. The data are in the `Cowles` data frame in the **carData** package:

```
R> summary(Cowles)
```

```
  neuroticism     extraversion        sex         volunteer
 Min.   : 0.00   Min.   : 2.00   female:780   no :824
```

```
1st Qu.: 8.00    1st Qu.:10.00    male  :641    yes:597
Median :11.00    Median :13.00
Mean   :11.47    Mean   :12.37
3rd Qu.:15.00    3rd Qu.:15.00
Max.   :24.00    Max.   :23.00
```

Cowles and Davis expected extraversion and neuroticism to interact in affecting volunteering, leading to the following logistic regression model:

```
R> library("car")
R> mod.cowles.1 <- glm(volunteer ~ sex + neuroticism * extraversion,
+    data = Cowles, family = binomial)
R> summary(mod.cowles.1)

Call:
glm(formula = volunteer ~ sex + neuroticism * extraversion, family = binomial,
    data = Cowles)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-1.4749  -1.0602  -0.8934   1.2609   1.9978

Coefficients:
                         Estimate Std. Error z value   Pr(>|z|)
(Intercept)             -2.358207   0.501320  -4.704 0.00000255
sexmale                 -0.247152   0.111631  -2.214    0.02683
neuroticism              0.110777   0.037648   2.942    0.00326
extraversion             0.166816   0.037719   4.423 0.00000975
neuroticism:extraversion -0.008552   0.002934  -2.915    0.00355

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1933.5  on 1420  degrees of freedom
Residual deviance: 1897.4  on 1416  degrees of freedom
AIC: 1907.4

Number of Fisher Scoring iterations: 4

R> Anova(mod.cowles.1)

Analysis of Deviance Table (Type II tests)

Response: volunteer
                         LR Chisq Df  Pr(>Chisq)
sex                        4.9184  1    0.026572
neuroticism                0.3139  1    0.575316
extraversion              22.1372  1 0.000002538
neuroticism:extraversion   8.6213  1    0.003323
```
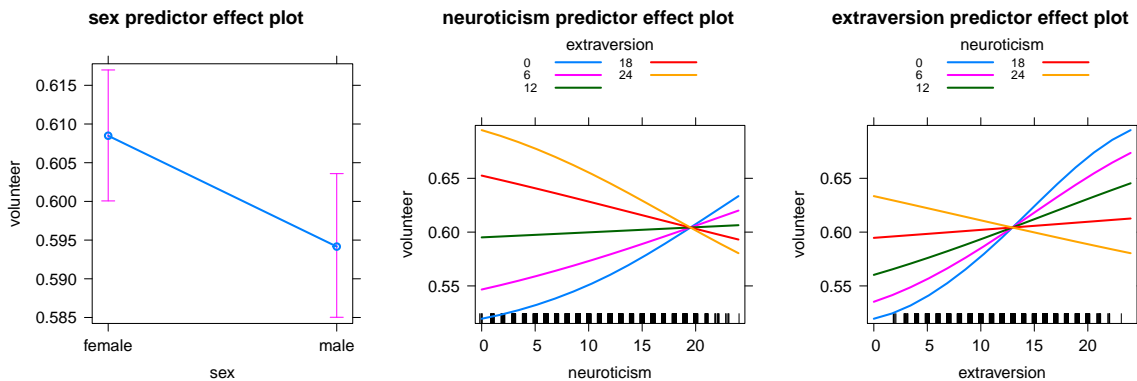
Figure 4: Predictor effect displays for Cowles and Davis's logistic regression for volunteering for a psychological experiment.

We use the `Anova` function in the **car** package (Fox and Weisberg 2011) to obtain Type II tests for the terms in the model. As expected, the interaction between `neuroticism` and `extraversion` has a small *p* value, and some evidence for a difference between the sexes is also apparent.

The predictor effect displays in mean scale (i.e., the probability scale) can all be drawn simultaneously by the `predictorEffects` function, as shown in Figure 4:

```
R> plot(predictorEffects(mod.cowles.1,
+    xlevels = list(extraversion = seq(0, 24, by = 6),
+    neuroticism = seq(0, 24, by = 6))),
+    axes = list(y = list(type = "response")),
+    lines = list(multiline = TRUE), rows = 1, cols = 3)
```

The lines in the predictor effect plots for `neuroticism` and `extroversion` are not straight because of the conversion from linear predictor (logit) to mean (probability) scale, obtained by specifying the argument `axes = list(y = list(type = "response"))` to `plot`. As before, we obtain multi-line plots for the continuous predictors by `lines = list(multiline = TRUE)`. We use the `xlevels` argument to `predictorEffects` to exert control over the values of these predictors. The `rows` and `cols` arguments to `plot` specify that the meta-array of effect displays should be arranged horizontally. By default, confidence intervals around the estimated effects are suppressed in multi-line plots; as before, they could be turned on by `confint = list(style = "auto")`. The `predictorEffects` function can also be used for a subset of predictors; see the function's help page.

The effect plot for `sex` is little more than a visualization of the regression coefficient for this factor, with females somewhat more likely than males to volunteer, and because the difference in estimated probabilities is small, the change to mean scale suggests that this visualization would apply for any meaningful averaging over the remaining predictors. The other two displays are two views of the same 3D surface, because both have $\mathbf{x}_2 = (\texttt{sex})$ fixed in the same way. The second display suggests clearly that as `neuroticism` increases, the probability of volunteering increases for subjects with low `extraversion`, but decreases for subjects with high `extraversion`. The third display, with `extraversion` on the horizontal axis, shows that the probability of volunteering generally increases with `extraversion`, at a very high rate
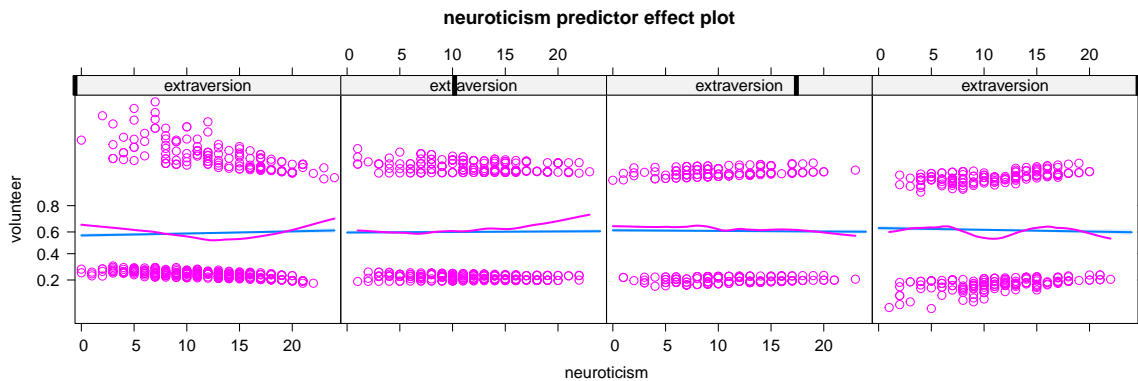
Figure 5: Predictor effect display for `neuroticism` showing partial residuals.

when `extraversion` is low, and a much lower rate when `neuroticism` is high; at the highest level of `neuroticism`, the relationship becomes negative. In this instance, both displays of the interactions can be useful, as they emphasize somewhat different stories.

Figure 5 is the predictor effect plot for `neuroticism` in linear predictor (logit) scale with the partial residuals shown:

```R
R> plot(predictorEffects(mod.cowles.1, ~ neuroticism,
+    partial.residuals = TRUE), lattice = list(layout = c(4, 1)))
```

The `lattice` argument to `plot` sets the **lattice** package `layout` argument, producing a plot with four panels arranged in one row (with the unusual column, row order standard for the **lattice** `layout` argument). The conditionally fixed values of `extraversion` increase from left to right across the range of this predictor, as indicated by the black line in the strip at the top of each panel.

For this logistic regression, the vertical axis is on the logit scale, and the default in the **effects** package is to label tick-marks on this axis with values of the inverse link function applied to the logits – that is, with corresponding probabilities. Because this is a diagnostic plot, we have not bothered to customize the location of the tick-marks on the vertical axis. The partial residuals are given by the magenta open circles, and the magenta line is the loess smooth of the partial residuals, with default span of 2/3. As before, the blue lines, which are straight on the logit scale, represent the fitted model, with the 95-percent point-wise confidence envelope around the fit superimposed. Robust smooths for non-Gaussian generalized linear models can result in substantial bias in the fitted curve (Landwehr, Pregibon, and Shoemaker 1980), and so a non-robust loess smoother is used. The general agreement of the smooths with the fitted effect suggests that the model reasonably represents the data.

As an additional check, we fit an alternative model to Cowles and Davis's data, in which each of `neuroticism` and `extraversion` is represented by a five-degree-of-freedom natural regression spline. The resulting model uses 25 degrees of freedom for the interaction, along with five degrees of freedom for each of the neuroticism and extraversion main effects, and is consequently much more flexible than the original model with a linear-by-linear interaction. A likelihood-ratio test comparing the new model to the original one fails to reveal significant lack of fit in the original model, and the original model is strongly preferred by both the AIC and BIC:

```
R> library("splines")
R> mod.cowles.2 <- glm(volunteer ~ sex + ns(neuroticism, 5) *
+    ns(extraversion, 5), data = Cowles, family = binomial)
R> anova(mod.cowles.1, mod.cowles.2, test = "Chisq")


Analysis of Deviance Table


Model 1: volunteer ~ sex + neuroticism * extraversion
Model 2: volunteer ~ sex + ns(neuroticism, 5) * ns(extraversion, 5)
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1416     1897.4
2      1384     1857.0 32   40.468   0.1448


R> cbind(AIC(mod.cowles.1, mod.cowles.2), BIC(mod.cowles.1, mod.cowles.2))


              df      AIC df      BIC
mod.cowles.1   5 1907.440  5 1933.736
mod.cowles.2  37 1930.972 37 2125.560
```

4.2. Canadian occupational prestige data

Blishen and McRoberts (1976) assembled data on the prestige, income level, and education level of males in 102 Canadian occupations, with the purpose of developing a prediction equation for occupational prestige based on income and education. We analyze similar data here, although the income and education scores in our data set are for *all* occupational incumbents, rather than just for men. These data were also analyzed by Fox and Suschnigg (1989). The prestige scores are average ratings for the occupations in a national survey conducted in the mid-1960s (Pineo and Porter 1967). The income and education scores are averages from the 1971 Canadian census. We classified 98 of the occupations by type: blue collar, white collar, and professional or managerial. Four of the occupations – "athletes," "newsboys," "babysitters," and "farmers" – did not fit into this classification and are dropped from our analysis. The Canadian occupational prestige data are in the data frame `Prestige` in the **carData** package:

```
R> summary(Prestige)

   education         income          women          prestige
 Min.   : 6.380   Min.   :  611   Min.   : 0.000   Min.   :14.80
 1st Qu.: 8.445   1st Qu.: 4106   1st Qu.: 3.592   1st Qu.:35.23
 Median :10.540   Median : 5930   Median :13.600   Median :43.60
 Mean   :10.738   Mean   : 6798   Mean   :28.979   Mean   :46.83
 3rd Qu.:12.648   3rd Qu.: 8187   3rd Qu.:52.203   3rd Qu.:59.27
 Max.   :15.970   Max.   :25879   Max.   :97.510   Max.   :87.20
     census        type
 Min.   :1113   bc  :44
 1st Qu.:3120   prof:31
```

```
 Median :5135   wc  :23
 Mean   :5402   NA's: 4
 3rd Qu.:8312
 Max.   :9517
```

Similar to the analysis by Blishen and McRoberts, we will begin by fitting an additive linear model with continuous numeric predictors `income` and `education`, and factor predictor `type`. Blishen and McRoberts's original analysis did not, however, include the predictor `type`. We reorder the levels of `type` from their default alphabetical ordering to their natural ordering:

```
R> Prestige$type <- factor(Prestige$type, levels = c("bc", "wc", "prof"))
R> mod.prestige.1 <- lm(prestige ~ income + education + type,
+    data = Prestige)
R> summary(mod.prestige.1)

Call:
lm(formula = prestige ~ income + education + type, data = Prestige)

Residuals:
     Min      1Q   Median      3Q      Max
-14.9529  -4.4486   0.1678   5.0566  18.6320

Coefficients:
             Estimate Std. Error t value    Pr(>|t|)
(Intercept) -0.6229292  5.2275255  -0.119       0.905
income       0.0010132  0.0002209   4.586 0.000014049
education    3.6731661  0.6405016   5.735 0.000000121
typewc      -2.7372307  2.5139324  -1.089       0.279
typeprof     6.0389707  3.8668551   1.562       0.122

Residual standard error: 7.095 on 93 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.8349,        Adjusted R-squared:  0.8278
F-statistic: 117.5 on 4 and 93 DF,  p-value: < 2.2e-16

R> Anova(mod.prestige.1)

Anova Table (Type II tests)

Response: prestige
          Sum Sq Df F value       Pr(>F)
income    1058.8  1 21.0339 0.0000140492
education 1655.5  1 32.8882 0.0000001205
type       591.2  2  5.8721     0.003966
Residuals 4681.3 93
```

The ANOVA table for the model reveals that all three terms have very small p values, suggesting that all three predictors may be useful.
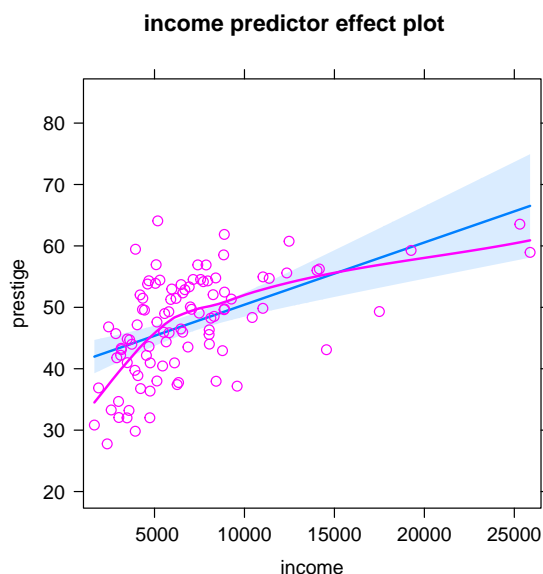
income predictor effect plot

Figure 6: Predictor effect display with partial residuals for `income` in the additive regression of `prestige` on `income`, `education`, and `type` of occupation.

An effect plot with partial residuals for `income` in this additive model is, except for the scaling of the vertical axis, a traditional partial residual plot:

```
R> plot(predictorEffects(mod.prestige.1, ~ income,
+    partial.residuals = TRUE))
```

The resulting graph, shown in Figure 6, reveals apparent nonlinearity in the partial regression of `prestige` on `income`.

An alternative story, however, is told by the term effect plot for `income` and `type` of occupation, which is higher-order than the terms actually in the model, and which can be computed using the `Effect` function in the **effects** package, producing Figure 7:

```
R> plot(Effect(c("income", "type"), mod.prestige.1,
+    partial.residuals = TRUE), partial.residuals = list(span = 0.9),
+    axes = list(x = list(rotate = 25)), lattice = list(layout = c(3, 1)))
```

We use a large span of 0.9 for the loess smoothers in this graph because dividing the data by the levels of the factor `type` leaves relatively few cases in each panel of the graph. Although the relationship between `prestige` and `income` in each panel appears positive and reasonably linear, the assumption that the slopes are equal in the panels is questionable, with an apparently larger slope for blue-collar occupations, a smaller slope for professional and managerial occupations, and an intermediate slope for white-collar occupations.

Adding the linear `income`-by-`type` interaction to the model, as suggested by Figure 7, confirms this impression:

```
R> mod.prestige.2 <- lm(prestige ~ type * income + education,
+    data = Prestige)
R> anova(mod.prestige.1, mod.prestige.2)
```
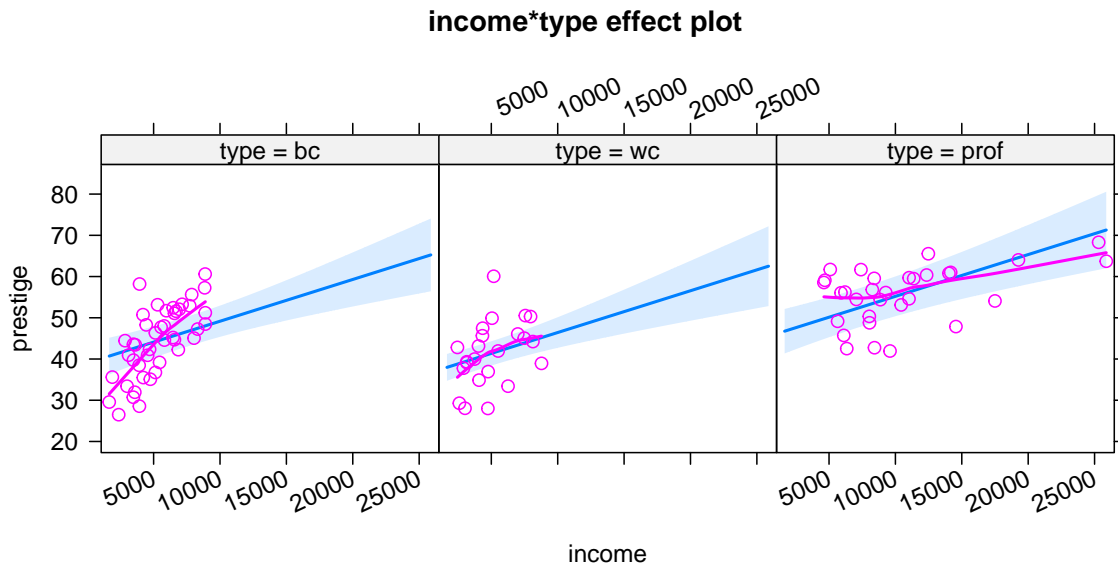
income*type effect plot

Figure 7: Term effect display with partial residuals for the predictors `income` and `type` in the additive regression of `prestige` on `income`, `education`, and `type` of occupation.

```
Analysis of Variance Table

Model 1: prestige ~ income + education + type
Model 2: prestige ~ type * income + education
  Res.Df    RSS Df Sum of Sq      F      Pr(>F)
1     93 4681.3
2     91 3791.3  2    890.02 10.681 0.00006809
```

Of course, the test for the interaction needs to be taken with a grain of salt, in that we added the interaction to the model after examining the data.

Figure 8 is the term effect plot for `income` and `type` (equivalent to the predictor effect plot for `income`) in the model that includes the income-by-type interaction:

```
R> plot(Effect(c("income", "type"), mod.prestige.2,
+    partial.residuals = TRUE), partial.residuals = list(span = 0.9),
+    axes = list(x = list(rotate = 25)), lattice = list(layout = c(3, 1)))
```

The nonlinearity apparent in the partial residual plot for `income` in the additive model in Figure 6 was induced by the relationship between `income` and occupational `type`, together with the unmodelled `income`-by-`type` interaction: Blue-collar occupations, for which the `income` slope is steep, are clustered at lower incomes, while professional occupations, for which the `income` slope is smaller, tend to have higher incomes. In addition to supporting the respecified regression, Figure 8 makes a useful pedagogical point about precision of estimation of the regression surface: The confidence envelopes show that the fitted regression is sensibly imprecisely estimated where there are no data.
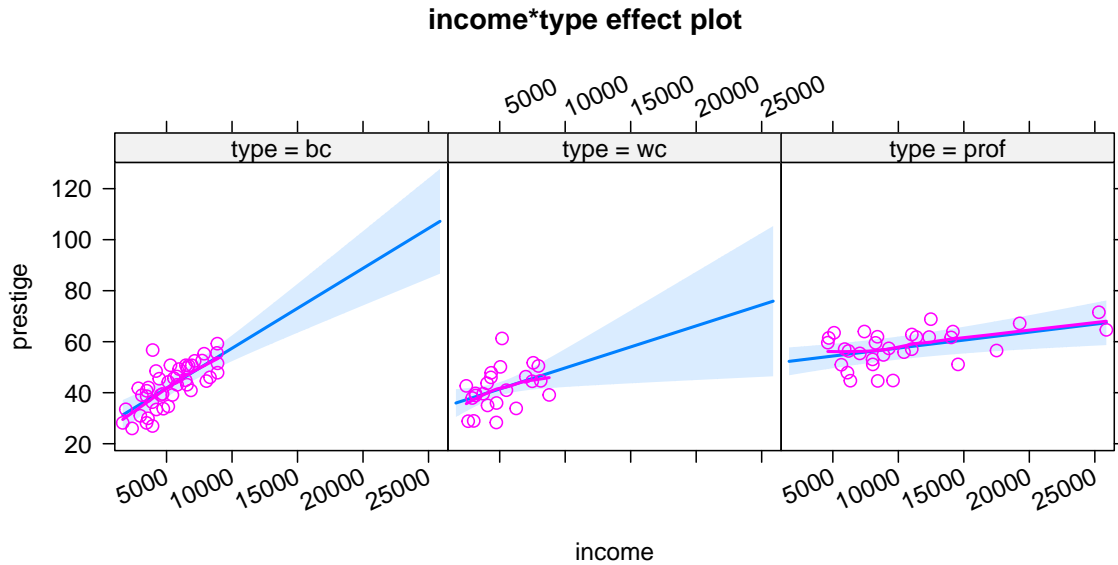
income*type effect plot

Figure 8: Term effect display with partial residuals for `income` and `type` in the model incorporating the `income:type` interaction.

4.3. Contrived regression data

We will analyze contrived data generated according to the following setup:

- We sample $n = 5000$ observations from a trivariate distribution for predictors x_1, x_2, and x_3, with uniform margins on the interval $[-2, 2]$, and with a prespecified bivariate correlation ρ between each pair of predictors. The method employed, described by Schumann (2009) and traceable to results reported by Pearson (1907), produces predictors that are nearly linearly related. Using 5000 observations allows us to focus on essentially asymptotic behavior of partial residuals in effect plots while still being able to discern individual points in the resulting graphs.

- We then generate the response y according to the model

$$y = \beta_0 + h\left(\boldsymbol{\beta}, \{x_1, x_2, x_3\}\right) + \varepsilon \tag{9}$$

where $\varepsilon \sim \mathrm{N}(0, 1.5^2)$. The regression function $h(\cdot)$ varies from example to example.

A variety of contrived examples generated in this manner, along with R functions for flexibly generating simulated data, are included in a vignette in the **effects** package.

In a sense, the example developed in this section and the examples in the vignette are unnecessary, because the results obtained are generally predictable from Cook's theoretical analysis of partial-residual plots, discussed in Section 3. We nevertheless think that these examples are useful for illustrating the application of Cook's analysis to partial-residual effect plots and for cultivating judgment about how to interpret these plots.

We consider a true model that combines nonlinearity and interaction, $\mathsf{E}(y|\mathbf{x}) = x_1^2 + x_2 x_3$; the predictors are moderately correlated, with $\rho = 0.5$. We then fit the incorrect working model $y \sim x_1 + x_2 + x_3$ to the data, producing the predictor effect displays with partial residuals in
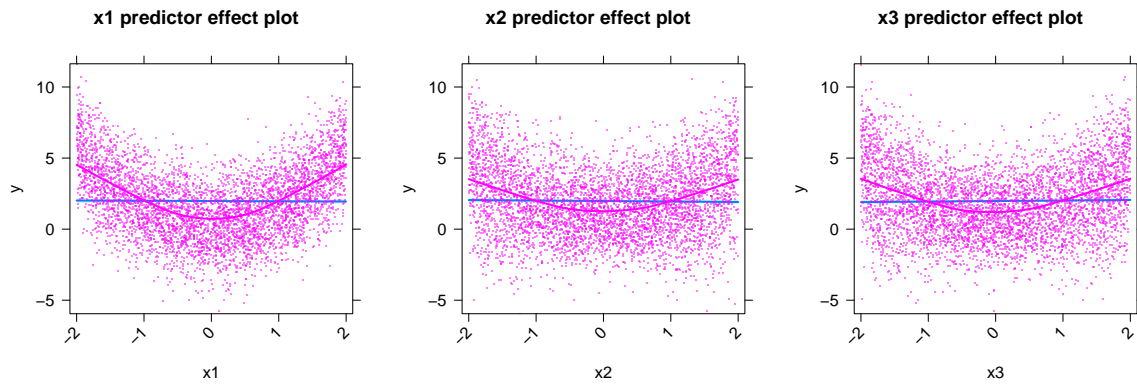
Figure 9: Effect displays with partial residuals for the predictors x_1, x_2, and x_3 in the incorrect model $y \sim x_1 + x_2 + x_3$ fit to data generated with the mean function $\mathsf{E}(y|\mathbf{x}) = x_1^2 + x_2 x_3$, with moderately correlated predictors.
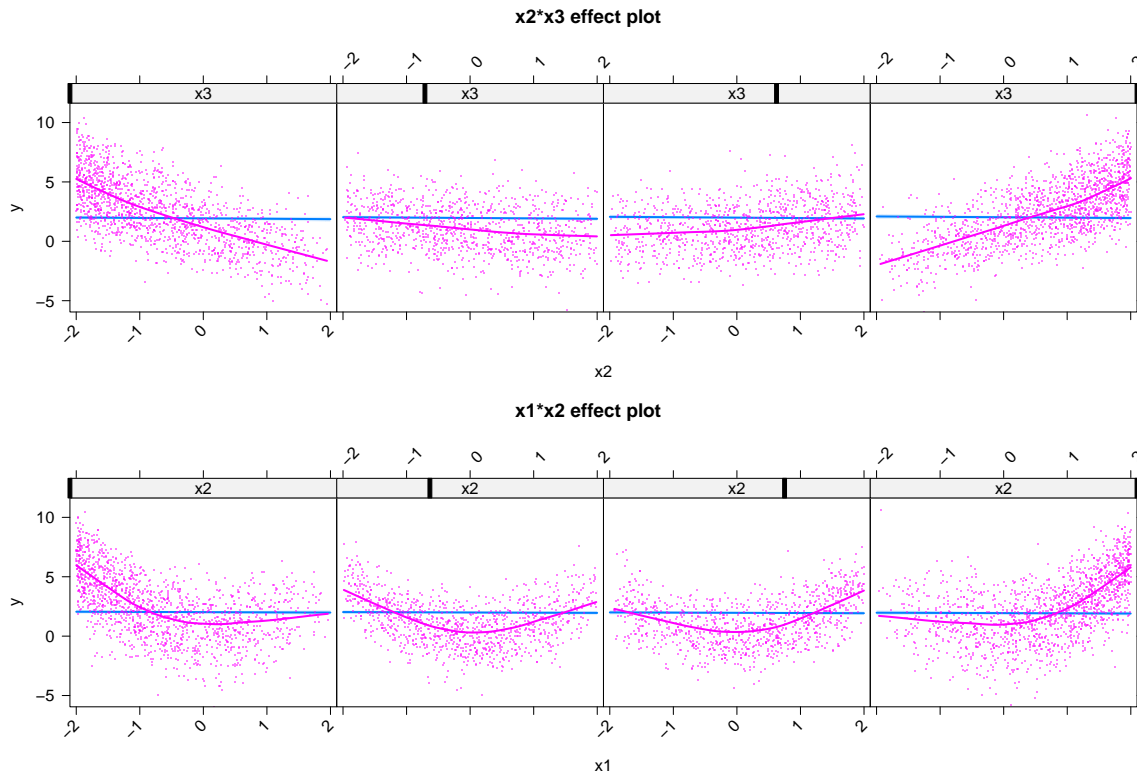

Figure 10: Term effect displays with partial residuals for $\{x_2, x_3\}$ (top) and for $\{x_1, x_2\}$ (bottom), the first of which corresponds to the missing $x_2{:}x_3$ interaction in the model generating the data.

Figure 9, for the predictors x_1, x_2, and x_3, which appear additively in the working model, and the term effect displays in Figure 10 for $\{x_2, x_3\}$ and $\{x_1, x_2\}$, corresponding respectively to the incorrectly excluded $x_2{:}x_3$ term and the correctly excluded $x_1{:}x_2$ interaction.

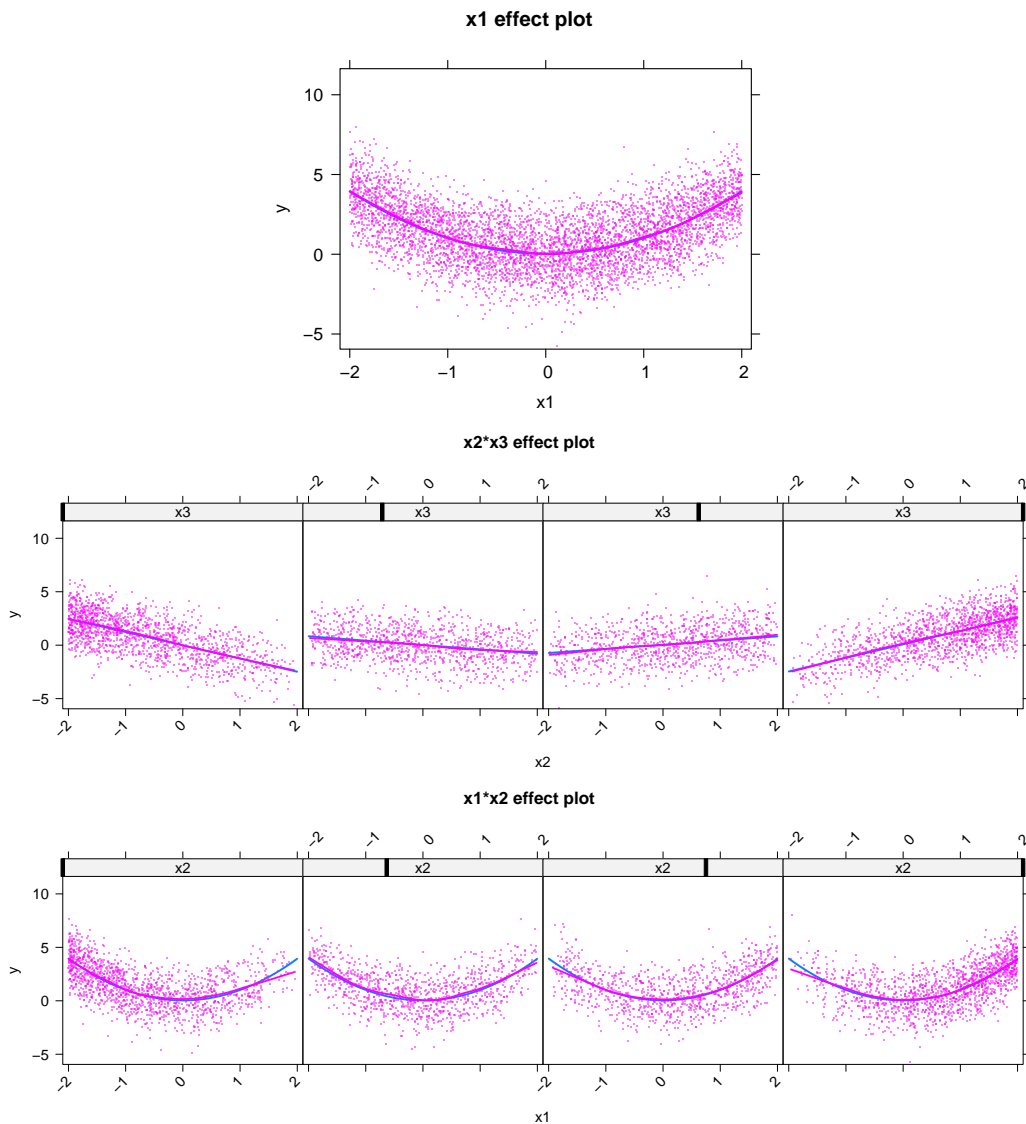The nonlinearity in the partial relationship of y to x_1 shows up clearly. The nonlinearity

Figure 11: Effect displays with partial residuals for x_1 and $\{x_2, x_3\}$, which correspond to the terms in the model generating *and* fitted to the data, $y \sim x_1^2 + x_2 * x_3$, and for $\{x_1, x_2\}$, which corresponds to an interaction that is not in the model.

apparent in the plots for x_2 and x_3 is partly due to contamination with x_1, but largely to the unmodelled interaction between x_2 and x_3, coupled with the correlation between these predictors. A similar phenomenon was noted in our analysis of the Canadian occupational prestige data in Section 4.2, where the unmodelled interaction between `type` and `income` induced nonlinearity in the partial relationship of `prestige` to `income`. The plot corresponding to the missing x_2:x_3 term (in the top panel of Figure 10) does a good job of detecting the unmodelled interaction, and curvature in this plot is slight. The plot for the x_1:x_2 term (in the bottom panel of Figure 10), a term neither in the true model nor in the working model, primarily reveals the unmodelled nonlinearity in the partial relationship of y to x_1.

If we fit the correct model, $y \sim x_1^2 + x_2 * x_3$, to the data, we obtain the plots shown in Figure 11.

As theory suggests, the partial residuals in these effect displays validate the model, supporting the exclusion of the $x_1{:}x_2$ interaction, the linear-by-linear interaction between x_2 and x_3, and the quadratic partial relationship of y to x_1.

5. Discussion

Graphical methods play a central role in many aspects of statistical data analysis. Their use roughly divides into three phases: an *exploratory phase*, in which an analyst examines data graphically for expected and unexpected structure (Tukey 1977); an *analysis phase*, in which graphs are used as an aid in formulating and assessing the adequacy of statistical models fit to the data; and a *presentation phase*, in which graphs provide summaries of an analysis that may be shared with others. Predictor effect plots are straightforward summary graphs for each predictor in a regression model. These plots are analogous to the usual numeric summaries of a fitted model, providing a separate explanation of the role of each predictor in a regression model after conditioning on all other relevant predictors.

The contribution of this article and the associated software in the **effects** package is two-fold:

1. We introduce predictor effect displays as an alternative to term effect displays. Predictor effect displays correspond more naturally to how researchers interpret the results of complex regression models, are simpler to describe formally, and have improved invariance properties relative to term effect displays.

2. Although effect displays, including effect plots with partial residuals, are related to other approaches for interpreting complex regression models, and although the general scheme employed using two-dimensional conditioning plots is not entirely original (see below for both of these points), the conceptualization described in this paper and its implementation in the **effects** package are novel in certain respects and more general than alternative approaches.

Partial residuals in effect plots can help to detect incorrectly specified models and point toward their improvement. If the model is correctly specified, then partial residuals for predictor effects, for the high-order terms of the model, and for effects of higher-order to those included in the model, should confirm the correctness of the model. On the other hand, if the model is incorrectly specified, then partial residual plots should not be interpreted naïvely, because a failure in one part of the model can contaminate plots for other combinations of predictors. For example, as we have shown, failure to model an interaction can appear as nonlinearity in a partial residual plot for one of the predictors entering the unmodelled interaction; and unmodelled nonlinearity in one predictor can also appear in the partial residuals for other predictors that are correlated with it. Awareness of these potential artifacts increases the utility of partial-residual effect plots in improving complex regression models. For example, if multiple issues are detected in partial residual plots, it is generally sensible to address them one at a time, rechecking at each step.

Displays similar to effect plots are also available in a number of other implementations.

- In R, the **visreg** (Breheny and Burchett 2018) package is most similar to **effects**, but it provides only for conditioning on *specific levels* of a factor rather than *averaging over* them, as is done in the **effects** package. The **visreg** package also seems to be limited

to two-factor interactions, excluding the possibility of plotting higher-order terms, so problems with more than one interaction may not be properly displayed.

- The `margins` and `marginsplot` programs in Stata (StataCorp. 2015) create displays that are similar to effect plots, except averaging or conditioning is over the empirical distribution of the regressors rather than the predictors, which can lead to invariance problems. As far was we can see, partial residuals cannot be added to a margins plot.

- Least-squares means, a generalization of adjusted means in analysis of covariance, introduced by Fisher 1936, as implemented in SAS (SAS Institute Inc. 2012) and the **lsmeans** package for R (Lenth 2016, 2018b), are capable of displaying interactions among factors, and in certain instances least-squares means coincide with effect displays. Partial residuals are not relevant to displays of least-squares means, however. The **lsmeans** package has been superseded by the **emmeans** package (Lenth 2018a).

For a linear predictor with only main effects, adding partial residuals to an effect plot is straightforward, and provides little that is new. For example, the plots produced by the `gam` functions in the **mgcv** (Wood 2017) and **gam** (Hastie 2018) packages are effect plots with partial residuals added. In an early general article on Trellis displays, Becker, Cleveland, and Shyu (1996) include a graph (their Figure 6) that they describe as a partial residual plot. Rather than fitting an explicit model to the data, however, they subtract marginal means for one factor from the data in a three-way classification with one case per cell, and then plot the resulting values against the other two factors. This procedure works because the data are balanced, and is equivalent to fitting a one-way ANOVA for one of the three factors. The procedure is not general, however, and the plotted values would not typically be termed partial residuals.

The functions in the **effects** package rely on the presence of a linear predictor in a regression model, and are therefore not suitable for less structured approaches to regression, such as regression trees. For this case, Friedman (2001) suggested plots obtained by averaging the estimate of $\hat{y}(\mathbf{x})$ over the empirical distribution of the predictors. Goldstein, Kapelner, Bleich, and Pitkin (2015) call these *individual conditional expectation* or *ICE* plots, and have implemented them in the **ICEbox** package (Goldstein *et al.* 2015) for R. These plots do not use a linear predictor and are therefore likely to be harder to interpret than predictor effect plots in problems for which the latter are appropriate.

The new ideas and software described in this article were not developed in a vacuum. In particular, we owe a debt to the general notion of conditioning plots (Cleveland 1993, 1994) and to their implementation in Trellis graphics (Becker and Cleveland 1996). In particular, the manner in which we handle the computation and display of partial residuals is loosely inspired by "shingles" in Trellis graphics, although it does not use shingles (overlapping sub-ranges for a continuous variable) in the literal sense. We also clearly lean heavily on the theoretical results concerning partial residuals developed by Cook (1993) and Cook and Croos-Dabrera (1998).

Predictor effect plots are reasonably easy to apply to a variety of modeling frameworks that use a linear predictor. In the **effects** package for R, we have included methods for linear, multivariate linear, and generalized linear models fit by the standard `lm` and `glm` functions and by the `svyglm` function in the **survey** package (Lumley 2004); linear models fit by generalized least squares using the `gls` function in the **nlme** package (Pinheiro, Bates, DebRoy,

Sarkar, and R Core Team 2018); multinomial regression models fit by `multinom` in the **nnet** package (Venables and Ripley 2002); ordinal regression models using `polr` from the **MASS** package (Venables and Ripley 2002) and `clm` and `clm2` from the **ordinal** package (Christensen 2018); linear and generalized linear mixed models using the `lme` function in the **nlme** package (Pinheiro *et al.* 2018) and the `lmer` and `glmer` functions in the **lme4** package (Bates, Mächler, Bolker, and Walker 2015); and latent class models fit by `poLCA` in the **poLCA** package (Linzer and Lewis 2011). We also include a generic method that may work with models fit by other functions that employ a linear predictor. Partial residuals, however, are not available, or even useful, for many of these classes of models. At present, we provide partial residuals for models of arbitrary complexity fit by `lm`, `glm`, `lmer`, `glmer`, and `lme`.

Acknowledgments

References

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using **lme4**." *Journal of Statistical Software*, **67**(1), 1–48. `doi:10.18637/jss.v067.i01`.

Becker RA, Cleveland WS (1996). *S-PLUS Trellis User's Manual*. MathSoft, Inc., Seattle.

Becker RA, Cleveland WS, Shyu MJ (1996). "The Visual Design and Control of Trellis Display." *Journal of Computational and Graphical Statistics*, **5**(2), 123–155. `doi:10.2307/1390777`.

Blishen BR, McRoberts HA (1976). "A Revised Socioeconomic Index for Occupations in Canada." *Canadian Review of Sociology and Anthropology*, **13**(1), 71–79. `doi:10.1111/j.1755-618x.1976.tb00759.x`.

Breheny P, Burchett W (2018). **visreg**: *Visualization of Regression Models*. R package version 2.5-0, URL `https://CRAN.R-project.org/package=visreg`.

Chambers JM, Hastie TJ (1992). "Statistical Models." In JM Chambers, TJ Hastie (eds.), *Statistical Models in S*, pp. 13–44. Wadsworth, Pacific Grove.

Christensen RHB (2018). **ordinal**: *Regression Models for Ordinal Data*. R package version 2018.8-25, URL `https://CRAN.R-project.org/package=ordinal`.

Cleveland WS (1993). *Visualizing Data*. Hobart Press, Summit.

Cleveland WS (1994). *The Elements of Graphing Data*. Revised edition. Hobart Press, Summit.

Cleveland WS, Grosse E, Shyu WM (1992). "Local Regression Models." In JM Chambers, TJ Hastie (eds.), *Statistical Models in S*, chapter 8, pp. 201–309. Wadsworth & Brooks/Cole, Pacific Grove.

Cook RD (1993). "Exploring Partial Residual Plots." *Technometrics*, **35**(4), 351–362. doi:10.2307/1270269.

Cook RD, Croos-Dabrera R (1998). "Partial Residual Plots in Generalized Linear Models." *Journal of the American Statistical Association*, **93**(442), 730–739. doi:10.1080/01621459.1998.10473725.

Cowles M, Davis C (1987). "The Subject Matter of Psychology: Volunteers." *British Journal of Social Psychology*, **26**(2), 97–102. doi:10.1111/j.2044-8309.1987.tb00769.x.

Fisher RA (1936). *Statistical Methods for Research Workers.* 6th edition. Oliver and Boyd, Edinburgh.

Fox J (1987). "Effect Displays for Generalized Linear Models." In CC Clogg (ed.), *Sociological Methodology 1987*, volume 17, pp. 347–361. American Sociological Association, Washington, DC.

Fox J (2003). "Effect Displays in R for Generalised Linear Models." *Journal of Statistical Software*, **8**(15), 1–27. doi:10.18637/jss.v008.i15.

Fox J, Hong J (2009). "Effect Displays in R for Multinomial and Proportional-Odds Logit Models: Extensions to the **effects** Package." *Journal of Statistical Software*, **32**(1), 1–24. doi:10.18637/jss.v032.i01.

Fox J, Suschnigg C (1989). "A Note on Gender and the Prestige of Occupations." *Canadian Journal of Sociology*, **14**(3), 353–360. doi:10.2307/3340609.

Fox J, Weisberg S (2011). *An R Companion to Applied Regression.* 2nd edition. Sage, Thousand Oaks.

Fox J, Weisberg S, Friendly M, Hong J (2018a). **effects***: Effect Displays for Linear, Generalized Linear, and Other Models.* R package version 4.1-0, URL https://CRAN.R-project.org/package=effects.

Fox J, Weisberg S, Price B (2018b). **carData***: Companion to Applied Regression Data Sets.* R package version 3.0-2, URL https://CRAN.R-project.org/package=carData.

Friedman JH (2001). "Greedy Function Approximation: A Gradient Boosting Machine." *The Annals of Statistics*, **29**(5), 1189–1232. doi:10.1214/aos/1013203451.

Goldstein A, Kapelner A, Bleich J, Pitkin E (2015). "Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation." *Journal of Computational and Graphical Statistics*, **24**(1), 44–65. doi:10.1080/10618600.2014.907095.

Hastie T (2018). **gam***: Generalized Additive Models.* R package version 1.16, URL https://CRAN.R-project.org/package=gam.

Landwehr JM, Pregibon D, Shoemaker AC (1980). "Some Graphical Procedures for Studying a Logistic Regression Fit." In *Proceedings of the Business and Economics Statistics Section, American Statistical Association*, pp. 15–20.

Lenth RV (2016). "Least-Squares Means: The R Package **lsmeans**." *Journal of Statistical Software*, **69**(1), 1–33. doi:10.18637/jss.v069.i01.

Lenth RV (2018a). **emmeans**: *Estimated Marginal Means, aka Least-Squares Means*. R package version 1.3.0, URL https://CRAN.R-project.org/package=emmeans.

Lenth RV (2018b). **lsmeans**: *Least-Squares Means*. R package version 2.27-62, URL https://CRAN.R-project.org/package=lsmeans.

Linzer DA, Lewis JB (2011). "**poLCA**: An R Package for Polytomous Variable Latent Class Analysis." *Journal of Statistical Software*, **42**(10), 1–29. doi:10.18637/jss.v042.i10.

Lumley T (2004). "Analysis of Complex Survey Samples." *Journal of Statistical Software*, **9**(8), 1–19. doi:10.18637/jss.v009.i08.

Nelder JA (1977). "A Reformulation of Linear Models." *Journal of the Royal Statistical Society A*, **140**(1), 48–77. doi:10.2307/2344517.

Pearson K (1907). *Mathematical Contributions to the Theory of Evolution.—XVI. On Further Methods of Determining Correlation*. Drapers' Company Research Memoirs. Biometric Series. IV. Cambridge University Press, London.

Pineo PC, Porter J (1967). "Occupational Prestige in Canada." *Canadian Review of Sociology and Anthropology*, **4**(1), 24–40. doi:10.1111/j.1755-618x.1967.tb00472.x.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2018). **nlme**: *Linear and Nonlinear Mixed Effects Models*. R package version 3.1-137, URL https://CRAN.R-project.org/package=nlme.

Sarkar D (2008). **lattice**: *Multivariate Data Visualization with R*. Springer-Verlag, New York.

SAS Institute Inc (2012). *LSMEANS Statement, SAS/STAT 9.3 User's Guide*. URL http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_introcom_a0000003362.htm.

Schumann E (2009). "Generating Correlated Uniform Variates." Accessed: 2015-05-21, URL http://comisef.wikidot.com/tutorial:correlateduniformvariates.

StataCorp (2015). "Margins." Accessed: 2015-04-09, URL http://www.stata.com/manuals13/rmargins.pdf.

Tukey JW (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. 4th edition. Springer-Verlag, New York. doi:10.1007/978-0-387-21706-2.

Wilkinson GN, Rogers CE (1973). "Symbolic Description of Factorial Models for Analysis of Variance." *Journal of the Royal Statistical Society C*, **22**(3), 392–399. doi:10.2307/2346786.

Wood SN (2017). *Generalized Additive Models: An Introduction with R.* 2nd edition. Chapman & Hall/CRC.

Affiliation:

John Fox
Department of Sociology
McMaster University
Hamilton, Ontario, Canada L8S 4M4
E-mail: jfox@mcmaster.ca
URL: http://socserv.socsci.mcmaster.ca/jfox

Sanford Weisberg
School of Statistics
University of Minnesota
Minneapolis, MN 55455, United States of America
E-mail: sandy@umn.edu
URL: http://users.stat.umn.edu/~sandy/