

# Visualizing Lack of Fit in Complex Regression Models: Adding Partial Residuals to Effect Displays

John Fox <sup>1</sup>   Sanford Weisberg <sup>2</sup>

<sup>1</sup>McMaster University  
Hamilton, Ontario, Canada

<sup>1</sup>University of Minnesota  
Minneapolis, Minnesota

Georg-August University Göttingen  
June 2014

## Introduction

- *Effect displays*, introduced by Fox (1987) for generalized linear models, visualize the response surface of complex regression models by conditioning and slicing the surface.
- *Partial-residual plots*, also called *component-plus-residual plots*, visualize lack of fit, traditionally in relatively simple regression models.
- We combine partial residuals with effect displays to visualize lack of fit in complex regression models. The combination is not perfect, but it is, we believe, useful.

# Introduction

## Example: SLID Regression

- Complex regression models are difficult to understand from the coefficients alone.
- Consider the following regression model fit, after preliminary examination of the data, to data from the 1994 Canadian Survey of Labour and Income Dynamics:

```
> mod <- lm(log(wages) ~ sex*(poly(age, 2) +  
+ I(education^2)), data=SLID)  
> summary(mod)
```

# Introduction

## Example: SLID Regression

```
> summary(mod)
```

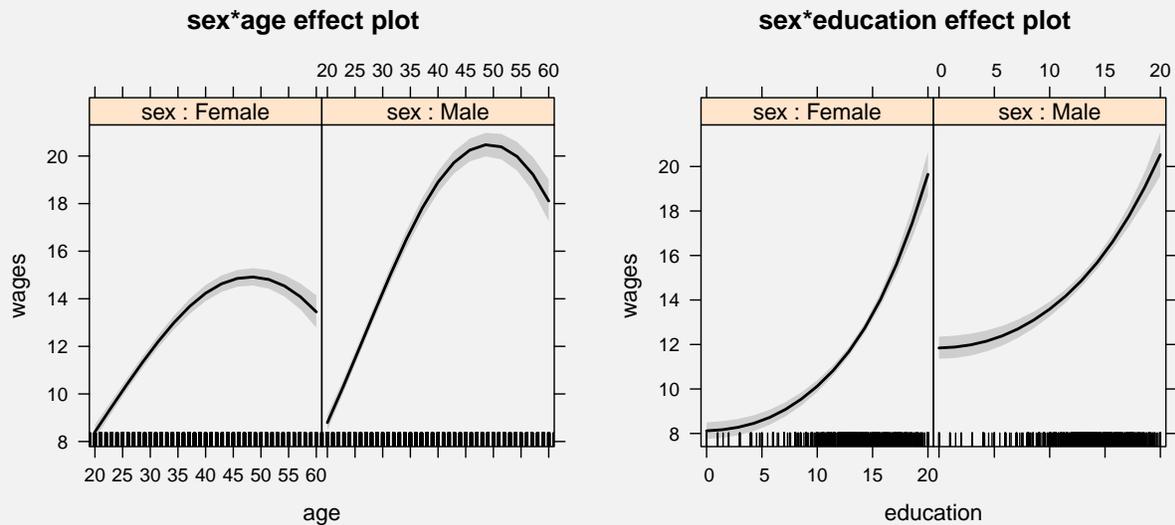
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.048900	0.024122	84.94	< 2e-16
sexMale	0.372564	0.033017	11.28	< 2e-16
poly(age, 2)1	0.832109	1.924148	0.43	0.66543
poly(age, 2)2	-20.887307	1.593706	-13.11	< 2e-16
I(education^2)	0.002210	0.000115	19.23	< 2e-16
sexMale:poly(age, 2)1	3.058961	2.666050	1.15	0.25130
sexMale:poly(age, 2)2	-8.034655	2.242862	-3.58	0.00034
sexMale:I(education^2)	-0.000834	0.000156	-5.35	0.000000091

Multiple R-squared: 0.3982, Adjusted R-squared: 0.3971

# Introduction

## Example: SLID Regression

- Effect plots of the sex  $\times$  age and sex  $\times$  education interactions are much easier to understand than the coefficients:



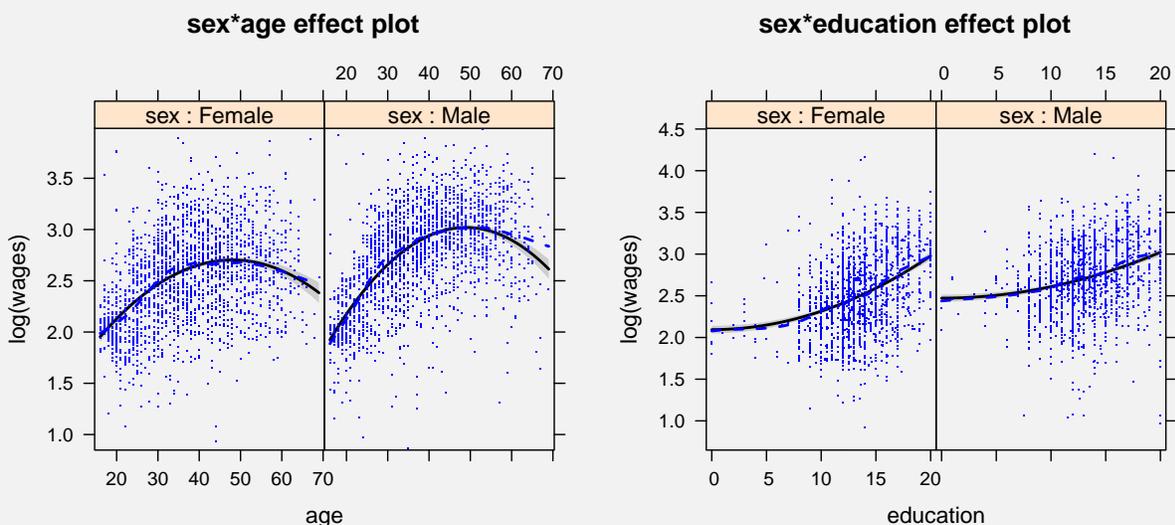
The bands represent 95% point-wise confidence limits for the conditional mean response.



# Introduction

## Example: SLID Regression

- Adding partial residuals to the effect plots reveals whether the data support the specified form of the model:



# Effect Displays

- We address the following situation: There is a *response variable*  $y$  and a set of *predictors*  $\mathbf{x} = \{x_1, \dots, x_p\}$ .
- We consider models in which the response is related to the predictors via a *linear predictor*

$$\begin{aligned}\eta(\mathbf{x}) &= \beta_0 + \beta_1 z_1 + \dots + \beta_k z_k \\ &= \mathbf{z}^T \boldsymbol{\beta}\end{aligned}$$

and where the *regression mean function*  $E(y|\mathbf{x}) = f[\eta(\mathbf{x})]$  for a known *link function*  $f(\cdot)$ .

- In this initial formulation, we assume that the model is correct, an assumption that we relax later.

# Effect Displays

- The *regressors* in the linear predictors, the  $z$ s, are pre-specified functions of the  $x$ s.
  - For example, a categorical predictor  $x$  may give rise to several dummy regressors or other “contrasts”;
  - a numeric predictor  $x$  might be represented by polynomial or regression-spline regressors;
  - when predictors interact, some regressors are functions of two or more predictors;
  - and numeric predictors may appear in the linear predictor as transformations, such as  $z = \log(x)$ .

## Effect Displays

- An *effect display* allows a subset of predictors, say,  $\mathbf{x}_e = \{x_1, \dots, x_s\}$ , to range over their combinations of values, while the remaining predictors,  $\mathbf{x}_{e^c} = \{x_{s+1}, \dots, x_p\}$ , are held to typical values, such as means,  $\mathbf{x}_{e^c}^*$ , tracing the *partial regression mean function*  $E(y|\{\mathbf{x}_e, \mathbf{x}_{e^c}^*\}) = f_e(\mathbf{x}_e)$ .
- For complex regression models in which different regressors are related by marginality, Fox (1987) suggested selecting effects that correspond to high-order terms in the model.
  - Thus, for example, in a model with linear predictor

$$\begin{aligned}\eta &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &\quad + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3\end{aligned}$$

the high-order terms are  $\{x_1, x_2\}$ ,  $\{x_1, x_3\}$ , and  $\{x_2, x_3\}$ .

- An effect display would be produced for each such term — e.g., for  $\{x_1, x_2\}$  holding  $x_3$  to its mean and using estimates of the regression coefficients,  $\hat{\beta}$ .

## Effect Displays

- We prefer a more general formulation, in which we can plot terms higher-order to those that appear in the model.
  - For example, in the model with three two-way interactions,  $\{x_1, x_2\}$ ,  $\{x_1, x_3\}$ , and  $\{x_2, x_3\}$ , we might plot  $\{x_1, x_2, x_3\}$ .
- This can be useful for visualizing models in which (as in the example) predictors “overlap” in high-order terms, and can be helpful when we look for lack of fit.
- We would not, however, plot a term — such as  $\{x_1\}$  in the example — that is marginal to a higher-order term in the model.
- This definition of effect displays is related to, but more general than, traditional *adjusted means* in analysis of covariance, and so-called *least-squares means* for categorical explanatory variables.

## Graphing Effect Displays

- When the  $s$  predictors in an effect are continuous, the natural graph of an effect display is an  $(s + 1)$ -dimensional partial response surface.
- This works well when  $s = 1$ , in which case the effect display is a 2D graph, and can be made to work when  $s = 2$ , where the 3D effect display can be represented by a wireframe figure or similar device, but it is not fully general.
- Instead, for  $s \geq 2$  we *slice* the regression surface in the direction of a particular *focal predictor*, say  $x_1$ , setting the other predictors in the effect,  $x_2, \dots, x_s$ , to all combinations of several values selected over their respective ranges, producing a sequence of 2D plots.
- The slices are lines on the  $(s + 1)$ D wireframe representing the effect, drawn parallel to the  $x_1$  axis.
- When the conditioning predictors in an effect are all categorical or discrete numeric variables, slicing is unnecessary, because the conditioning variables then naturally partition the data.

## Graphing Effect Displays

### Example: Canadian Occupational Prestige Regression

- The data in this example appeared originally in Blishen and McRoberts (1976), and describe 98 Canadian occupations circa 1970. The data were reanalyzed by Fox and Suschnigg (1989).
- We regress the rated prestige of the occupations (from a 0–100 “thermometer” scale) on their average income (\$1000s), average education (years), and type of occupation (classified as blue collar, white collar, or professional/managerial).

# Graphing Effect Displays

Example: Canadian Occupational Prestige Regression

- We fit a model that includes interactions between between income and type of occupation and between education and type:

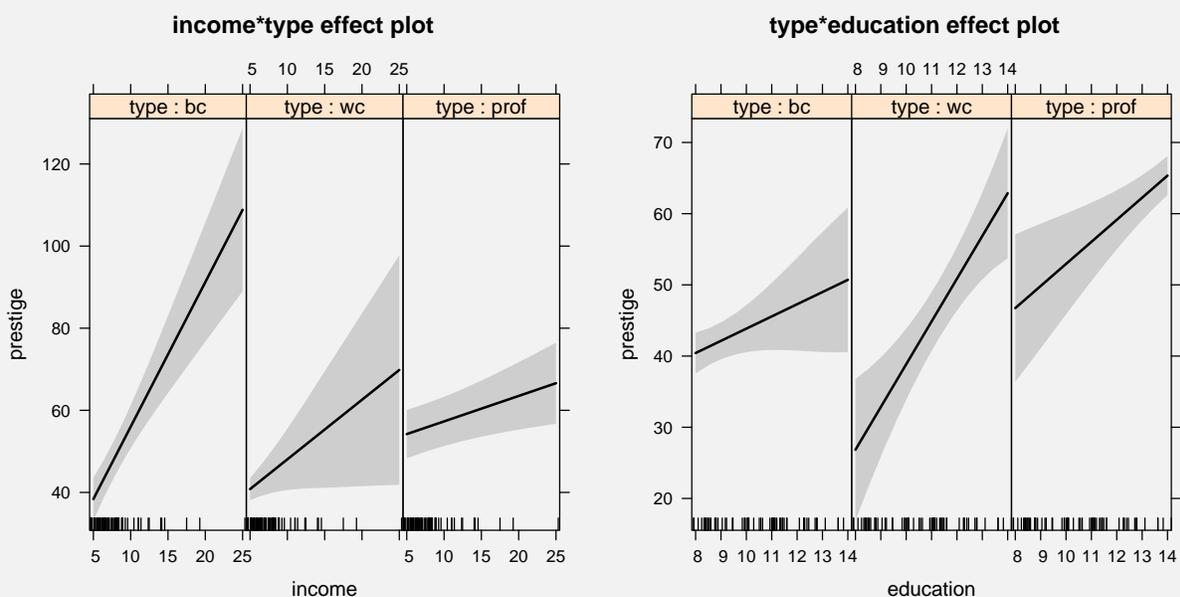
**“Type II” Analysis of Variance Table**

Source	Sum of Squares	df	F	p
Type	591.2	2	7.4	.001
Income	1131.9	1	28.3	< .001
Education	1068.0	1	26.7	< .001
Income × Type	951.8	2	11.9	< .001
Education × Type	238.4	2	3.0	.06
Residuals	3552.9	89		

# Graphing Effect Displays

Example: Canadian Occupational Prestige Regression

- Effects displays for the high-order terms in the model:



# Graphing Effect Displays

Example: Canadian Occupational Prestige Regression

- The effect display for the type  $\times$  income interaction holds education to its average level, while that for the type  $\times$  education interaction holds income to its average level.
- Plotting on the scale of the linear predictor, the values to which predictors excluded from the effect (e.g., education in the plot for {type, income}) are fixed affects the height but not the shape of the effect plot.
- In this example, the conditioning (i.e., non-focal) predictor in each effect, type of occupation, is categorical, and thus slicing is unnecessary.

# Partial-Residual Plots

- Although *partial-residual plots* were traditionally applied to simple additive regression models, Cook (1993) presented a more general framework that we will adapt, with partial success, to effect displays.
- Extending our notation, suppose that we fit the following tentative or “working” model to the data on  $i = 1, \dots, n$  independent observations:

$$y_i = b_0 + \mathbf{b}_1^T \mathbf{z}_{i1} + \mathbf{b}_2^T \mathbf{z}_{i2} + e_i$$

where, following Cook, we use Latin rather than Greek letters for regression coefficients because we regard the working model as possibly wrong.

- In the context of an effect display, the regressors in  $\mathbf{z}_2$  are functions of the focal and conditioning predictors  $\mathbf{x}_e$ , possibly jointly with (some of) the other predictors  $\mathbf{x}_{e^c}$ , while the regressors in  $\mathbf{z}_1$  are functions *only* of predictors in  $\mathbf{x}_{e^c}$ .

## Partial-Residual Plots

- For example, consider the working model

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} \\ + b_{12}x_{i1}x_{i2} + b_{13}x_{i1}x_{i3} + b_{23}x_{i2}x_{i3}$$

- If the effect of interest is  $\mathbf{x}_e = \{x_1, x_2\}$ , then  $\mathbf{x}_{e^c} = \{x_3\}$ ,  $\mathbf{z}_{i1} = (x_{i3})$ , and  $\mathbf{z}_{i2} = (x_{i1}, x_{i2}, x_{i1}x_{i2}, x_{i1}x_{i3}, x_{i2}x_{i3})^T$ .

## Partial-Residual Plots

- Again following Cook, but using our notation, suppose that the true regression is given by

$$y_i = \alpha_0 + \boldsymbol{\alpha}_1^T \mathbf{z}_{i1} + g_e(\mathbf{z}_{i2}) + \varepsilon_i$$

where  $E(\varepsilon_i | \mathbf{z}_{i1}, \mathbf{z}_{i2}) = 0$ , and where  $g_e(\mathbf{z}_2)$  then expresses the true partial dependence of  $y$  on the regressors in  $\mathbf{z}_2$ .

- *Partial residuals* for the regressors in  $\mathbf{z}_2$  are defined as

$$e_i^{(2)} = \hat{\mathbf{b}}_2^T \mathbf{z}_{i2} + \hat{e}_i$$

where  $\hat{\mathbf{b}}_2$  are estimates of the coefficients  $\mathbf{b}_2$  in the working model and  $\hat{e}_i$  are residuals from the working model.

## Partial-Residual Plots

- Cook shows that the partial residuals will reflect the true partial regression function  $g_e(\cdot)$  — that is consistently estimate  $g_e(\mathbf{z}_{i2}) + \varepsilon_i$  — if
  - 1 the working model is correct, i.e.,  $g_e(\mathbf{z}_{i2}) = \mathbf{b}_2^T \mathbf{z}_{i2}$ ;
  - 2 or, if the working model is incorrect, the regressors in  $\mathbf{z}_{i1}$  are linearly related to those in  $\mathbf{z}_{i2}$ . Nonlinear relationships between  $\mathbf{z}_{i1}$  and  $\mathbf{z}_{i2}$  may “leak” into the partial residuals so that their pattern imperfectly reflects  $g_e(\cdot)$ .
- The second condition is unlikely to hold in adapting partial residuals to effect displays, so leakage may be a problem, but we still may be able to approximate  $g_e(\cdot)$  sufficiently accurately for practical purposes.
- Cook and Croos-Dabrera (1998) extend these results to partial-residual plots for generalized linear models employing canonical links (such as binomial logit models). For GLMs, partial residuals are computed using working residuals.

## Adding Partial Residuals to Effect Displays

- Adding partial residuals to effect displays presents the following complications:
  - 1 Effects include the regression constant,  $\hat{b}_0$ . This is an inessential change because it affects only the height of the display.
  - 2 Predictors not in the effect,  $\mathbf{x}_{e\mathcal{C}}$ , are set to typical values  $\mathbf{x}_{e\mathcal{C}}^*$ .
    - Setting  $\mathbf{x}_{e\mathcal{C}} = \mathbf{x}_{e\mathcal{C}}^*$  contributes not only to  $\mathbf{z}_1$ , which affects only the height of the display, but also in certain instances to the values of some of the regressors  $\mathbf{z}_2$  at which the effect is evaluated.
    - Regressors that are (partial) functions of  $\mathbf{x}_{e\mathcal{C}}$  are not in general equal to the values of these regressors at the observed values  $\mathbf{x}_{e\mathcal{C},i}$ .
    - Adding partial residuals to an effect display therefore requires adjusting the fit computed at the observations. As we will show presently, slicing has a similar implication.

## Adding Partial Residuals to Effect Displays

- 3 To create a sequence of 2D plots, non-focal (conditioning) predictors in the effect  $\mathbf{x}_{e \setminus f} = \{x_2, \dots, x_s\}$  are sliced at a grid of values,

$$\mathbf{x}_{e \setminus f}^m = \{\mathbf{x}_2^m, \dots, \mathbf{x}_s^m\}, \text{ where } \mathbf{x}_j^m = \{x_j^{(1)}, \dots, x_j^{(M_j)}\}.$$

- The values on the grid do not correspond to the observed values of the non-focal predictors,  $\mathbf{x}_{e \setminus f, i}$ , so we round the non-focal predictors to the sliced values to which they are closest, which we denote  $\mathbf{x}_{e \setminus f, i}^m$ . The result is to allocate each observation to one panel of the display.
- The “smearing” that occurs by rounding  $\mathbf{x}_{e \setminus f, i}$  to  $\mathbf{x}_{e \setminus f, i}^m$  will generally cause the estimate of the partial regression function corresponding to the effect to be biased, with the degree of bias dependent upon the extent of rounding, the nonlinearity of the regression function in the direction of the smear, and the asymmetry of the distribution of  $\mathbf{x}_{e \setminus f}$  in the smear around  $\mathbf{x}_{e \setminus f, i}^m$ .
- Discrete predictors in  $\mathbf{x}_{e \setminus f}$ , however, don’t require slicing: The observed values of these predictors are used.

## Adding Partial Residuals to Effect Displays

- 4 Consider a working model of the form, e.g.,

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + b_{23} x_{i2} x_{i3} + e_i$$

- For partial residuals for the effect  $\{x_2, x_3\}$  not to be contaminated by leakage, the regression of  $x_1$  on each of  $x_2$ ,  $x_3$ , and  $x_2 x_3$  must be linear.
- But if the first two regressions are linear, almost surely the third is not.
- This problem isn’t unique to the example: Cook’s second condition is generally not plausible for partial residuals in effect plots.
- In our experience, however, nonlinear relationships must be very strong before leakage is serious.

# Partial-Residual Effect Plots: Examples

## Canadian Occupational Prestige Regression

- Let us begin by fitting a simple additive model to the Canadian occupational prestige data:

$$\text{prestige} = b_0 + b_1\text{income} + b_2\text{education} + b_3\text{wc} + b_4\text{prof} + e$$

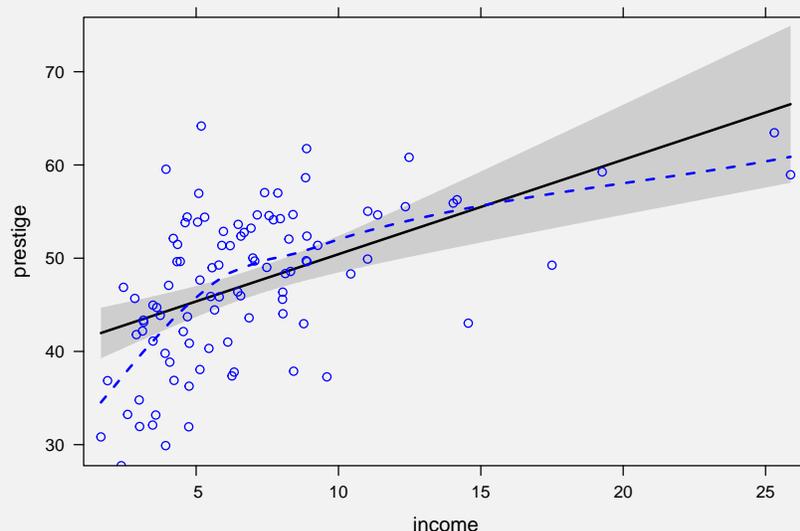
where *wc* and *prof* are dummy regressors for type of occupation. But for the inclusion of occupational type, this is similar to the model originally fit to the data by Blishen and McRoberts (1976).

- Regarding this as a working model, the effect display for income is (aside from the scaling of the vertical axis) a traditional partial-residual plot:

# Partial-Residual Effect Plots: Examples

## Canadian Occupational Prestige Regression

income effect plot



- The plot appears to reveal nonlinearity in the partial relationship of prestige to income controlling for education.

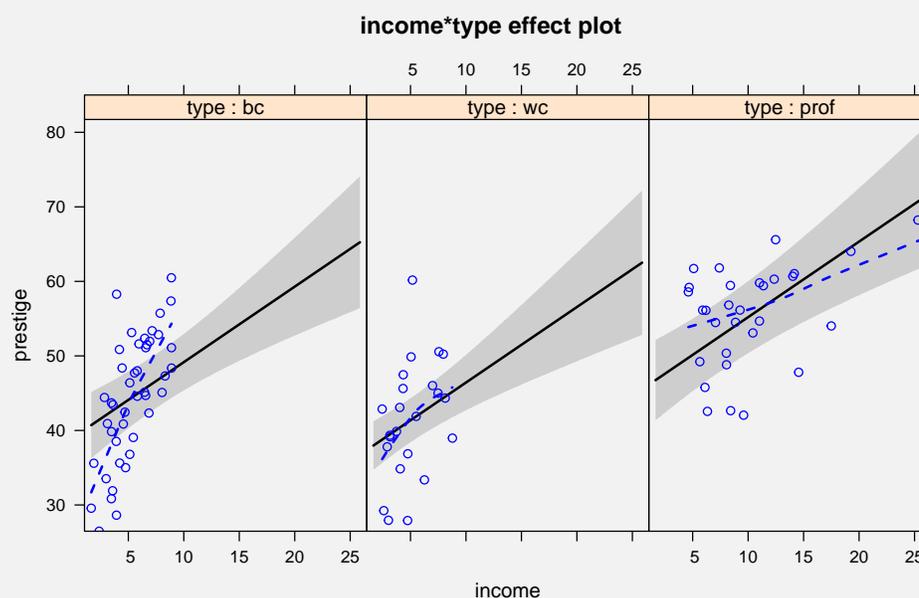
# Partial-Residual Effect Plots: Examples

## Canadian Occupational Prestige Regression

- An alternative is to plot the term  $\{\text{income, type}\}$ , which reveals the unmodeled interaction between income and type (along with information about where the data are located):

# Partial-Residual Effect Plots: Examples

## Canadian Occupational Prestige Regression

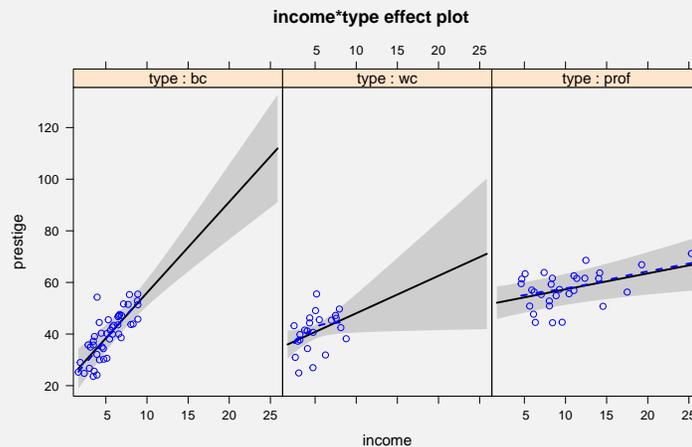


The broken line in each panel is a nonparametric-regression ("loess") smooth of the points in the plot.

# Partial-Residual Effect Plots: Examples

## Canadian Occupational Prestige Regression

- Fitting the model with interactions between income and type and between education and type produces the following partial-residual effect plot:



- In addition to validating the form of the model, this plot shows how the relative precision of the estimated effect is influenced by the location of the data.



# Partial-Residual Effect Plots: Examples

## Contrived Data Exploring Leakage

- We next construct two examples for contrived data, the second of which raises the possibility of leakage.
- The regressors are generated by the following R code:

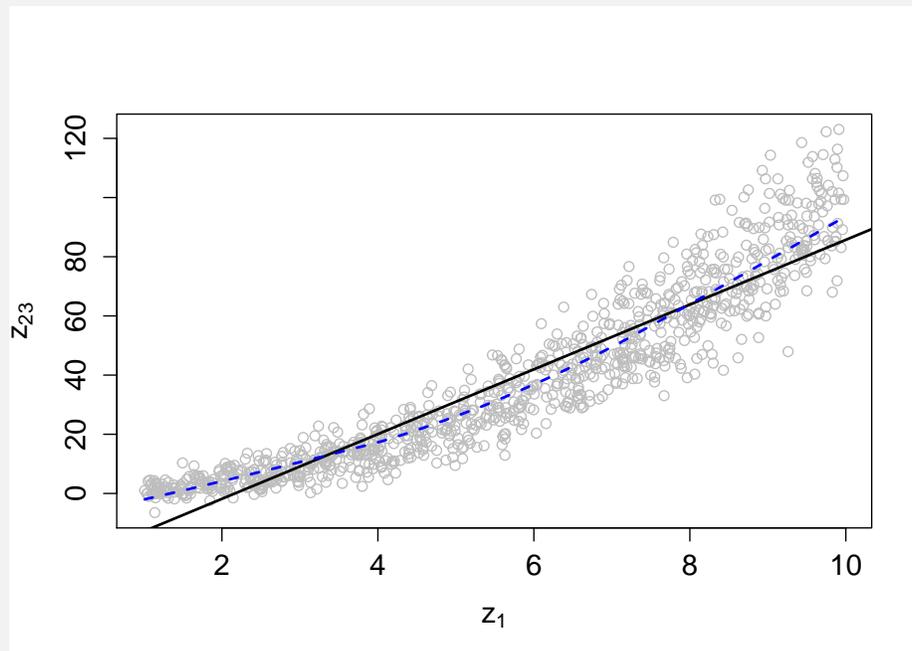
```
> set.seed(12345) # for reproducibility
> z1 <- runif(1000, 1, 10)
> z2 <- z1 + rnorm(1000)
> z3 <- z1 + rnorm(1000)
> z23 <- z2*z3
```



# Partial-Residual Effect Plots: Examples

Contrived Data Exploring Leakage

- This induces a nonlinear relationship between the interaction regressor  $z_{23}$  and  $z_1$ :



# Partial-Residual Effect Plots: Examples

Contrived Data Exploring Leakage

- In the first instance, we generate data and fit the correct model, so no leakage should occur:

```
> y <- 5*z1 + z2 + z3 + z23 + rnorm(1000, 0, 25)
> mod <- lm(y ~ z1 + z2 + z3 + z23)
> summary(mod)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.1335	3.1399	-0.679	0.497
z1	6.0005	1.1013	5.448	6.41e-08 ***
z2	0.6168	0.9387	0.657	0.511
z3	0.1785	1.0093	0.177	0.860
z23	1.0753	0.1078	9.972	< 2e-16 ***

---

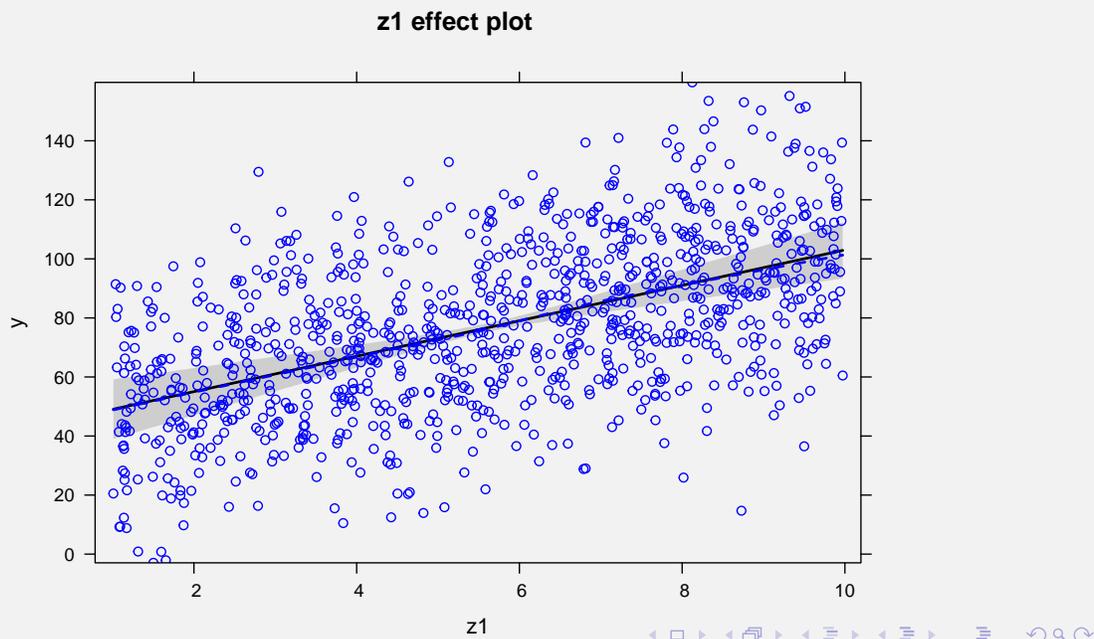
Residual standard error: 23.73 on 995 degrees of freedom

Multiple R-squared: 0.8104, Adjusted R-squared: 0.8096

# Partial-Residual Effect Plots: Examples

Contrived Data Exploring Leakage

- The partial residuals in the effect plot for  $z_1$  should (and do) confirm that the partial relationship of  $y$  to  $z_1$  is linear:



Fox, Weisberg ( Göttingen)

Visualizing Lack of Fit

June 2014

31 / 36

# Partial-Residual Effect Plots: Examples

Contrived Data Exploring Leakage

- Now consider the following data, in which the partial relationship of  $y$  to  $z_1$  is nonlinear:

```
> y <- (z1 - mean(z1))^2 + z2 + z3 + 0.1*z23 +  
+ rnorm(1000, 0, 5)
```

- We mistakenly model this relationship as linear:

```
> mod <- lm(y ~ z1 + z2 + z3 + z23)
```

```
> summary(mod)
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	24.86596	0.80713	30.808	<2e-16	***
z1	-0.04135	0.28310	-0.146	0.884	
z2	-3.10522	0.24130	-12.869	<2e-16	***
z3	-2.98198	0.25945	-11.494	<2e-16	***
z23	0.82045	0.02772	29.597	<2e-16	***

---

Multiple R-squared: 0.7139, Adjusted R-squared: 0.7128

Fox, Weisberg ( Göttingen)

Visualizing Lack of Fit

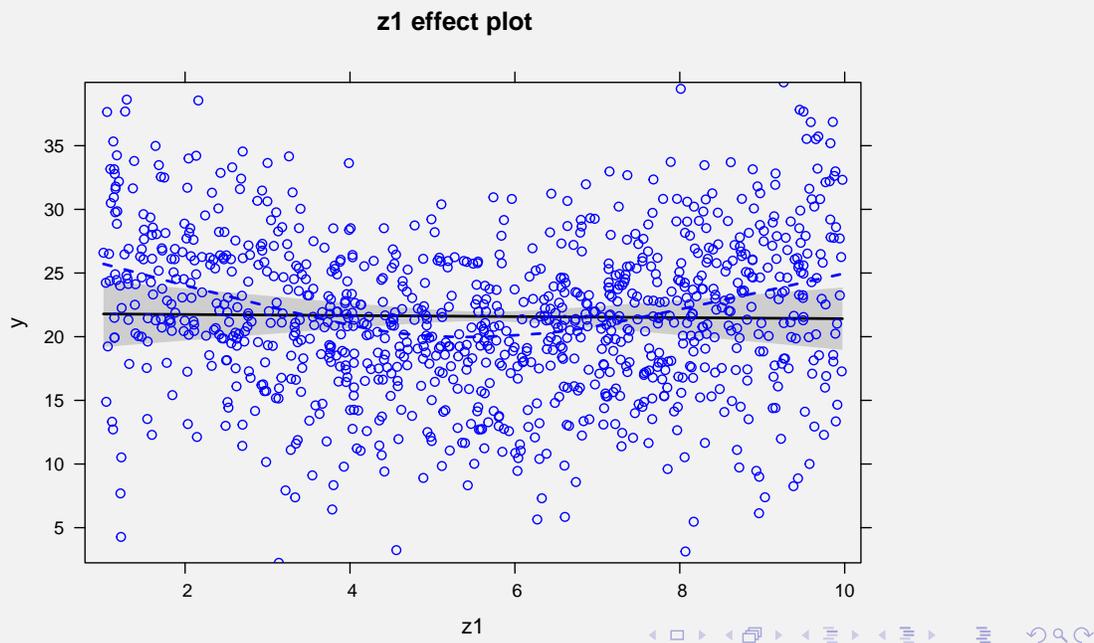
June 2014

32 / 36

# Partial-Residual Effect Plots: Examples

Contrived Data Exploring Leakage

- Although leakage is a potential issue here, the partial residuals in the effect plot for  $z_1$  accurately reflect the unmodeled nonlinearity:



# Partial-Residual Effect Plots: Summary

- It is possible to add partial residuals to effect plots to explore lack of fit in a model fit tentatively to the data.
- When there is more than one continuous predictor in an effect, the display slices the partial response surface corresponding to the effect.
- To obtain partial residuals, we round predictors in an effect, other than the focal predictor defining the horizontal axis, to sliced values. This generally will produce bias in the partial residuals.
- Effect plots in complex models will often fail to meet Cook's conditions for consistency in partial-residual plots, a problem that likely will not often compromise their usefulness in practice.
- The methods described in this talk are implemented in a general way in the development version of the **effects** package for R (Fox, 2003; Fox et al., 2014).

## References

- Blishen, B. R. and McRoberts, H. A. (1976). A revised socioeconomic index for occupations in Canada. *Canadian Review of Sociology and Anthropology*, 13:71–79.
- Cook, R. D. (1993). Exploring partial residual plots. *Technometrics*, 35:351–362.
- Cook, R. D. and Croos-Dabrera, R. (1998). Partial residual plots in generalized linear models. *Journal of the American Statistical Association*, 93(442):730–739.
- Fox, J. (1987). Effect displays for generalized linear models. In Clogg, C. C., editor, *Sociological Methodology 1987 (Volume 17)*, pages 347–361. American Sociological Association, Washington, D. C.
- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8(15):1–27.

## References

- Fox, J. and Suschnigg, C. (1989). A note on gender and the prestige of occupations. *Canadian Journal of Sociology*, 14:353–360.
- Fox, J., Weisberg, S., Friendly, M., and Hong, J. (2014). *effects: Effect Displays for Linear, Generalized Linear, Multinomial-Logit, Proportional-Odds Logit Models and Mixed-Effects Models*. R package version 3.0-0.