Direct Parameter Recentering & Rescaling

Doug Hemken

The Problem (recap)

- Re-calculating coefficients for models

  - ***As if*** the data have been recentered, rescaled, or both

  - ***Without*** actually transforming the data

The Problem (recap)

- There is a classic formula for ***standardized*** coefficients

  - $\beta_0^z = 0$
  - $\beta_i^z = \frac{\sigma_{x_i}}{\sigma_y}\beta_i$

  - Widely implemented in software
  - Of limited use for interactions and polynomials

Example: set up

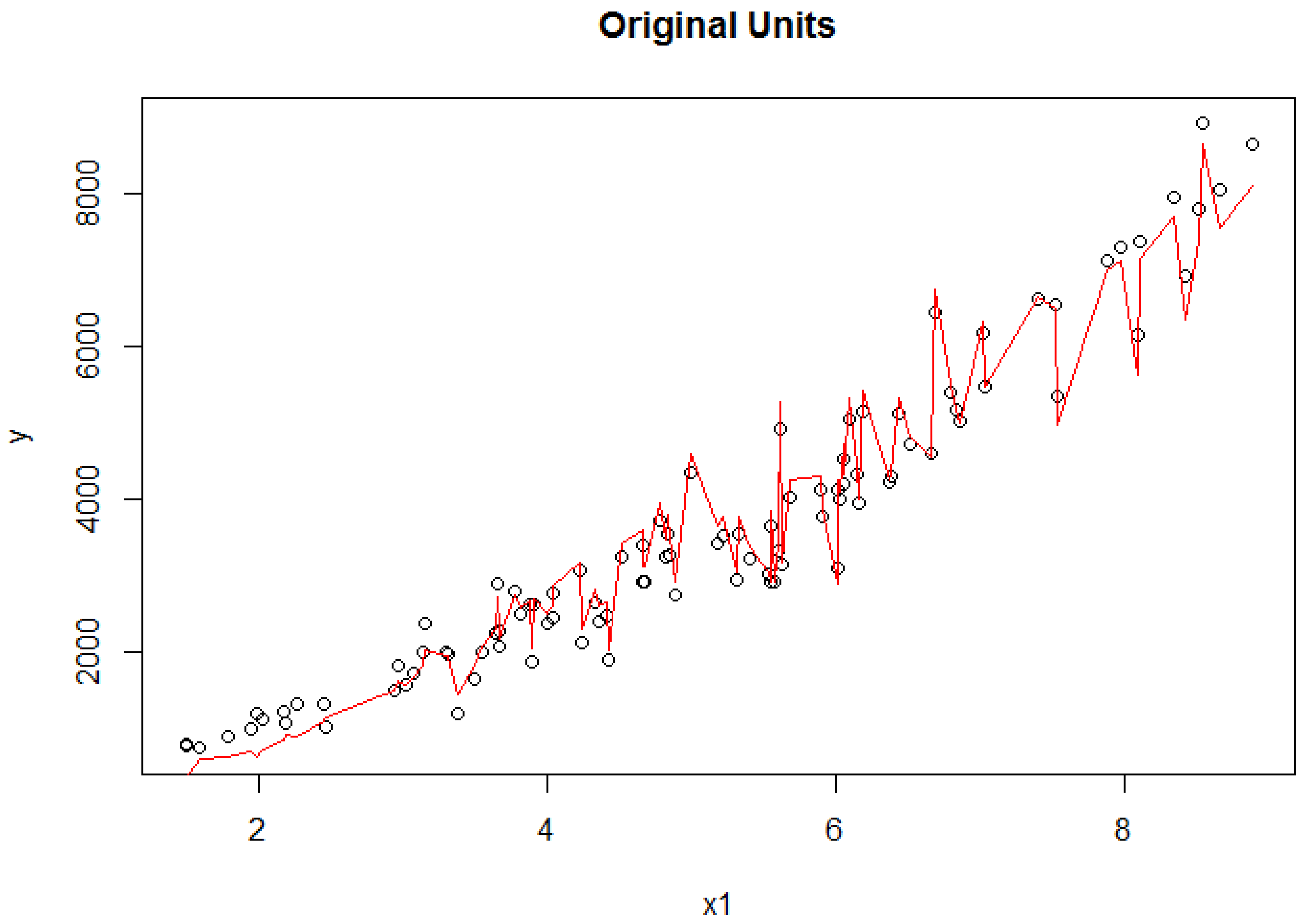
- Simulated data and model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_c x_1 x_2$

```
orig <- lm(y ~  x1 * x2, data=example)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4840.09     847.54   5.711 1.26e-07 ***
x1          -1428.67     169.61  -8.423 3.58e-13 ***
x2           -600.20      84.62  -7.093 2.23e-10 ***
x1:x2         235.31      16.76  14.038  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 231 on 96 degrees of freedom
Multiple R-squared:  0.9864,    Adjusted R-squared:  0.986
F-statistic:  2319 on 3 and 96 DF,  p-value: < 2.2e-16
```

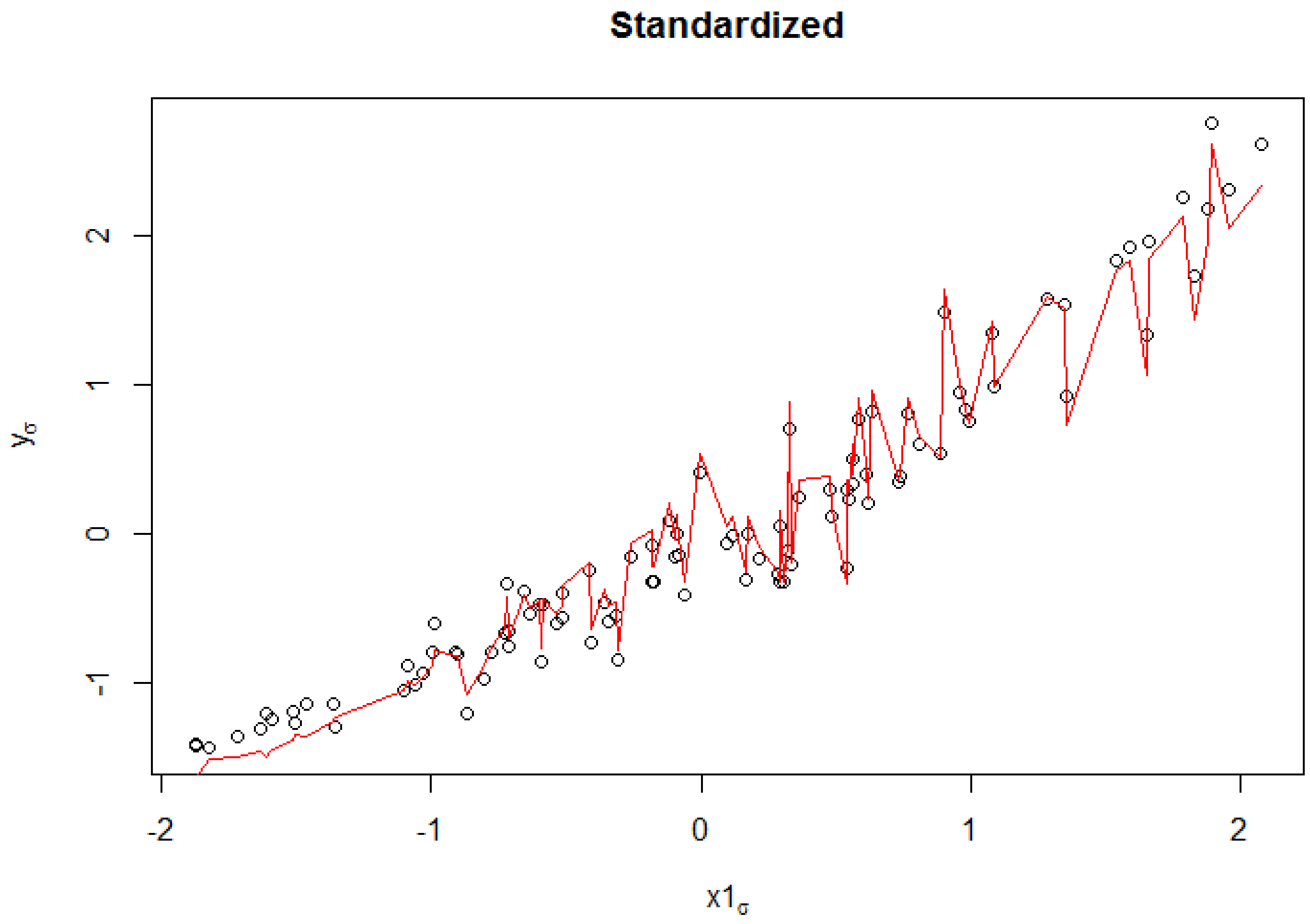
Example: compare coefficients

	Original	Classic Formula	Recalculated
(Intercept)	4840.09	0.00	-0.04
x1	-1428.67	-1.37	0.89
x2	-600.20	-0.27	0.26
x1:x2	235.31	0.23	0.20

Two graphs



Two graphs



Of note

- Unchanged
  - Relative positions of data points
  - Relative position of predictions
- Changed
  - x- and y- axes
- Change of basis

Change of basis, *data* ⇒ Change of basis, *parameters*

- ⇒ there is a linear transformation for our parameters
  - Can also transform variance-covariance matrices

- How do we get from data space to parameter space?
  - The column space of the parameters is a subspace of the outer product of the column space of the data
  - We construct our parameter change of basis as an outer product of the data changes of basis

Single variable

$$y = \beta_0 + \beta_1 x$$

$$Y = X\beta$$

- Recenter, $x_\delta = x - \mu$
    - Let $C = \begin{bmatrix} 1 & \mu \\ 0 & 1 \end{bmatrix}$
- $\beta^\delta = C \times \beta$
- Rescale, $x_\sigma = x/\sigma$
    - Let $S = \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix}$
- $\beta^\sigma = S \times \beta$
- Standardize
- $\beta^z = S \times C \times \beta$

Two variables, re-centering

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$

- $\beta^{\delta} = (C_2 \otimes C_1) \times \beta = \begin{bmatrix} 1 & \mu_1 & \mu_2 & \mu_1\mu_2 \\ 0 & 1 & 0 & \mu_2 \\ 0 & 0 & 1 & \mu_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \beta$

Two variables, re-scaling

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$

- $\beta^{\sigma} = (S_2 \otimes S_1) \times \beta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \boldsymbol{\sigma_1} & 0 & 0 \\ 0 & 0 & \boldsymbol{\sigma_2} & 0 \\ 0 & 0 & 0 & \boldsymbol{\sigma_1 \sigma_2} \end{bmatrix} \beta$

Implementation

- Consider
  - Less-than-full factorial models
    - Missing lower-order terms (no-intercept, nested terms)
  - Polynomial terms, collecting terms
  - Untransformed variables, direct sums
  - Category coding/contrasts

- Progress
  - All of the pieces, but not fully integrated yet.

Advantages

- Requires only sufficient statistics, not the full data
- Reduces the total computation
- Reduces numerical error

On numerical error

- Simulation study
  - Predicted values from recentered models should be unchanged
    - Error measured as normed difference in predicted values

  - 3-way model, appropriate to the data
  - 100,000 simulations, each model N=100

  - The same (recentered) data is used to calculate predicted values
    - One model is re-estimated
    - The other has directly transformed coefficients

Kernel density of normed errors
(red=direct, black=re-estimate)