MULTIPLE REGRESSION ANALYSIS
For the Case of Two Regressors

In the following notes, least-squares estimation is developed for multiple regression problems with *two* explanatory variables, here called *regressors* (such as in the "Fast Food" example developed in class). The advantage of developing this estimation procedure explicitly for the case of $k = 2$ is that it allows the key ideas of *partial regression plots* and *partial residual plots* to be developed as well.

1. Least-Squares Estimation for *k* = 2

Given independent samples $(y_i, x_{1i} x_{2i})$, $i = 1,..,n$ from the linear model:

(1.1) $\quad Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^2\right)$,

the least-squares estimates $\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right)$ are obtained by *minimizing* the function:

(1.2) $\quad S(\beta_0, \beta_1, \beta_2) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2$

The quadratic function *S* is easily seen to be convex in $(\beta_0, \beta_1, \beta_2)$, so that the unique minimum of *S* is found by setting each of its partial derivatives to zero. For β_0 this yields:

(1.3)
$$0 = \left.\frac{\partial S}{\partial \beta_0}\right|_{\beta_0 = \hat{\beta}_0} = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}\right)(-1)$$
$$\Rightarrow \sum_i y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_i x_{1i} + \hat{\beta}_2 \sum_i x_{2i}$$
$$\Rightarrow \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2$$
$$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$$

In particular this is seen to imply that (as in simple regression), the mean point $(\bar{y}, \bar{x}_1, \bar{x}_2)$ must satisfy the estimated regression equation exactly. Next for β_1 we obtain:

(1.4)
$$0 = \left.\frac{\partial S}{\partial \beta_1}\right|_{\beta_1 = \hat{\beta}_1} = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}\right)(-x_{1i})$$
$$\Rightarrow \sum_i y_i x_{1i} = \hat{\beta}_0 \sum_i x_{1i} + \hat{\beta}_1 \sum_i x_{1i}^2 + \hat{\beta}_2 \sum_i x_{2i} x_{1i}$$
$$= \left(\bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2\right) \sum_i x_{1i} + \hat{\beta}_1 \sum_i x_{1i}^2 + \hat{\beta}_2 \sum_i x_{2i} x_{1i}$$
$$\Rightarrow \sum_i (y_i - \bar{y}) x_{1i} = \hat{\beta}_1 \sum_i (x_{1i} - \bar{x}_1) x_{1i} + \hat{\beta}_2 \sum_i (x_{2i} - \bar{x}_2) x_{1i}$$

Hence, recalling that $0 = \sum_i (y_i - \bar{y}) = \sum_i (x_{1i} - \bar{x}_1) = \sum_i (x_{2i} - \bar{x}_2)$, it follows that

(1.5) $\sum_i (y_i - \bar{y}) x_{1i} = \sum_i (y_i - \bar{y}) x_{1i} - \bar{x}_1 \sum_i (y_i - \bar{y}) = \sum_i (y_i - \bar{y})(x_{1i} - \bar{x}_1)$

and similarly that

(1.6) $\sum_i (x_{1i} - \bar{x}_1) x_{1i} = \sum_i (x_{1i} - \bar{x}_1)(x_{1i} - \bar{x}_1)$

(1.7) $\sum_i (x_{2i} - \bar{x}_2) x_{1i} = \sum_i (x_{2i} - \bar{x}_2)(x_{1i} - \bar{x}_1)$

Thus if we now introduce the following simplifying "sum-of-products" notation:

(1.8) $s_{jy} = \sum_i (y_i - \bar{y})(x_{ji} - \bar{x}_j) \quad , \; j = 1,2$

(1.9) $s_{jh} = \sum_i (x_{ji} - \bar{x}_j)(x_{hi} - \bar{x}\ ) \; , \; j,h = 1,2$

then it follows that the first-order condition (1.4) can be written simply as

(1.10) $s_{11}\hat{\beta}_1 + s_{12}\hat{\beta}_2 = s_{1y}$

The argument for the partial of S with respect to β_2 is completely parallel, so that the *normal equations* for $(\hat{\beta}_1, \hat{\beta}_2)$ are given by (1.10) and

(1.11) $s_{21}\hat{\beta}_1 + s_{22}\hat{\beta}_2 = s_{2y}$

1.1 Solutions for the Beta Estimates

These two linear equations can be solved simultaneously to obtain the desired estimates$(\hat{\beta}_1, \hat{\beta}_2)$. First observe from (1.11) that

(1.12) $\hat{\beta}_2 = \left(s_{2y} - s_{21}\hat{\beta}_1\right) \big/ s_{22}$

Substituting this into (1.10) yields the relation

(1.13) $\left(s_{11} - \frac{s_{12}s_{21}}{s_{22}}\right)\hat{\beta}_1 = \left(s_{1y} - \frac{s_{12}s_{2y}}{s_{22}}\right)$

which can be reduced to yield the following expression for $\hat{\beta}_1$:

$$(1.14)\quad \hat{\beta}_1 = \frac{s_{22}s_{1y} - s_{12}s_{2y}}{s_{11}s_{22} - s_{12}s_{21}}$$

A completely parallel argument yields the following expression for $\hat{\beta}_2$:

$$(1.15)\quad \hat{\beta}_2 = \frac{s_{11}s_{2y} - s_{21}s_{1y}}{s_{11}s_{22} - s_{12}s_{21}}$$

Finally, the estimate $\hat{\beta}_0$ is obtained by substituting (1.14) and (1.15) into (1.3).

1.2 Regression with Uncorrelated Regressors

One important relation between simple and multiple regression can be seen from this explicit solution. In terms of our present notation, recall that the *sample correlation* between variables x_1 and x_2 is given by

$$(1.16)\quad r(x_1,x_2) = \frac{\sum_i (x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{\sqrt{\sum_i (x_{1i}-\bar{x}_1)^2}\sqrt{\sum_i (x_{2i}-\bar{x}_2)^2}} = \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}}$$

and hence that the variables x_1 and x_2 are *uncorrelated* if and only if $s_{12} = 0$. Next observe that if the *slope coefficients* for the *simple* regressions of y on x_1 and x_2 are denoted respectively by $\hat{\beta}_{y1}$ and $\hat{\beta}_{y2}$, then by definition

$$(1.17)\quad \hat{\beta}_{y1} = \frac{\sum_i (y_i-\bar{y})(x_{1i}-\bar{x}_1)}{\sum_i (x_{1i}-\bar{x}_1)^2} = \frac{s_{1y}}{s_{11}}$$

and

$$(1.18)\quad \hat{\beta}_{y2} = \frac{\sum_i (y_i-\bar{y})(x_{2i}-\bar{x}_2)}{\sum_i (x_{2i}-\bar{x}_2)^2} = \frac{s_{2y}}{s_{22}}$$

Hence if x_1 and x_2 are *uncorrelated* then the slope estimates, $\hat{\beta}_1$ and $\hat{\beta}_2$, in the multiple regression of y on x_1 and x_2 are seen to take the simple form:

$$(1.19)\quad \hat{\beta}_1 = \frac{s_{22}s_{1y} - 0}{s_{11}s_{22} - 0} = \frac{s_{1y}}{s_{11}} = \hat{\beta}_{y1}$$

and

$$(1.20)\quad \hat{\beta}_2 = \frac{s_{11}s_{2y} - 0}{s_{11}s_{22} - 0} = \frac{s_{2y}}{s_{22}} = \hat{\beta}_{y2}$$

Thus when regressors are *uncorrelated*, multiple regression is in fact equivalent to *simple regressions on each variable separately*. More generally, the degree to which the slope coefficients in multiple regression differ from those is separate simple regressions depends on the *degree of correlation* between the regressors. This is in fact true for *all* multiple regressions. If a given regressor, x_j, is uncorrelated with all the others, then its slope estimate, $\hat{\beta}_j$, will be the same in both simple and multiple regressions on y.

2. Partial Regression Plots

Given these estimation results for the $k = 2$ case, we are in a position to analyze the properties of partial regression plots. If for $j = 1,2$ we let $y^{(j)}$ denote the *residual values* of y in a simple regression on explanatory variable x_j, and similarly, if $x_h^{(j)}$ denotes the *residual values* of the other explanatory variable x_h ($h \neq j$) in a simple regression on x_j, then by definition, the *partial regression plot* for x_1 in a multiple regression of y on (x_1, x_2) is a plot of $y^{(2)}$ against $x_1^{(2)}$. The key feature of this plot is that the beta estimate, $\hat{\beta}_1$, for x_1 in the *multiple* regression is precisely the same as the slope coefficient, $\hat{\beta}_1^{(2)}$, in a *simple* regression of $y^{(2)}$ on $x_1^{(2)}$. We are now in a position to show this explicitly.

2.1 Equivalence of Betas

To show that $\hat{\beta}_1$ and $\hat{\beta}_1^{(2)}$ are identical, we begin by observing that the simple regression of y on x_2 yields a *slope estimate*, $\hat{\alpha}_{y2}$, of the form

$$(2.1)\quad \hat{\alpha}_{y2} = \frac{\sum_i (y_i - \bar{y})(x_{2i} - \bar{x}_2)}{\sum_i (x_{2i} - \bar{x}_2)^2} = \frac{s_{2y}}{s_{22}}$$

with corresponding *intercept estimate*, $\hat{\alpha}_0$, given by

$$(2.2)\quad \hat{\alpha}_0 = \bar{y} - \hat{\alpha}_{2y} x_2$$

Hence the i^{th} *residual value*, $y_i^{(2)}$, is given for all $i = 1,..,n$ by

$$(2.3)\quad y_i^{(2)} = y_i - \hat{\alpha}_0 - \hat{\alpha}_{y2} x_{2i} = y_i - (\bar{y} - \hat{\alpha}_{y2}\bar{x}_2) - \hat{\alpha}_{y2} x_{2i}$$

$$= (y_i - \bar{y}) - \hat{\alpha}_{y2}(x_{2i} - \bar{x}_2)$$

$$= (y_i - \bar{y}) - \frac{s_{2y}}{s_{22}}(x_{2i} - \bar{x}_2)$$

Similarly, the regression of x_1 on x_2 yields *slope estimate*

$$(2.4) \quad \hat{\alpha}_{12} = \frac{\sum_i (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sum_i (x_{2i} - \bar{x}_2)^2} = \frac{s_{21}}{s_{22}}$$

and by the same argument, thus yields the *residual values*:

$$(2.5) \quad x_{1i}^{(2)} = x_{1i} - (\bar{x}_1 - \hat{\alpha}_{12}\bar{x}_2) - \hat{\alpha}_{12}x_{2i}$$

$$= (x_{1i} - \bar{x}_1) - \hat{\alpha}_{12}(x_{2i} - \bar{x}_2)$$

$$= (x_{1i} - \bar{x}_1) - \frac{s_{21}}{s_{22}}(x_{2i} - \bar{x}_2)$$

Notice also that (as in every regression) the *sample means*, $\bar{y}^{(2)}$ and $\bar{x}_1^{(2)}$, of these residual values are identically *zero*. For example

$$(2.6) \quad \bar{x}_1^{(2)} = \tfrac{1}{n}\sum_i x_{1i}^{(2)} = \tfrac{1}{n}\sum_i \left(x_{1i} - (\bar{x}_1 - \hat{\alpha}_{12}\bar{x}_2) - \hat{\alpha}_{12}x_{2i}\right)$$
$$= \tfrac{1}{n}\sum_i x_{1i} - (\bar{x}_1 - \hat{\alpha}_{12}\bar{x}_2)\tfrac{n}{n} - \hat{\alpha}_{12}\tfrac{1}{n}\sum_i x_{2i}$$
$$= \bar{x}_1 - (\bar{x}_1 - \hat{\alpha}_{12}\bar{x}_2) - \hat{\alpha}_{12}\bar{x}_2 = 0$$

Hence in the simple regression of $y^{(2)}$ on $x_1^{(2)}$, it follows that the slope coefficient, $\hat{\beta}_1^{(2)}$, must have the form:

$$(2.7) \quad \hat{\beta}_1^{(2)} = \frac{\sum_i y_i^{(2)} x_{1i}^{(2)}}{\sum_i \left(x_{1i}^{(2)}\right)^2}$$

But (recalling that $s_{12} = s_{21}$) we then see from (2.3) and (2.5) that the numerator of (2.7) is given by

$$(2.8) \quad \sum_i y_i^{(2)} x_{1i}^{(2)} = \sum_i \left[(y_i - \bar{y}) - \frac{s_{2y}}{s_{22}}(x_{2i} - \bar{x}_2)\right]\left[(x_{1i} - \bar{x}_1) - \frac{s_{21}}{s_{22}}(x_{2i} - \bar{x}_2)\right]$$

$$= \sum_i (y_i - \bar{y})(x_{1i} - \bar{x}_1) - \frac{s_{2y}}{s_{22}} \sum_i (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) -$$

$$\frac{s_{21}}{s_{22}} \sum_i (y_i - \bar{y})(x_{2i} - \bar{x}_2) + \frac{s_{2y}}{s_{22}} \frac{s_{21}}{s_{22}} \sum_i (x_{2i} - \bar{x}_2)^2$$

$$= s_{1y} - \frac{s_{2y}}{s_{22}} s_{12} - \frac{s_{21}}{s_{22}} s_{2y} + \frac{s_{2y}}{s_{22}} \frac{s_{21}}{s_{22}} s_{22}$$

$$= s_{1y} - 2\frac{s_{2y}s_{21}}{s_{22}} + \frac{s_{2y}s_{21}}{s_{22}}$$

$$= s_{1y} - \frac{s_{2y}s_{21}}{s_{22}} = \frac{s_{22}s_{1y} - s_{21}s_{2y}}{s_{22}}$$

Similarly, the denominator of (2.7) is given by

$$(2.9) \quad \sum_i \left(x_{1i}^{(2)}\right)^2 = \sum_i \left[(x_{1i} - \bar{x}_1) - \frac{s_{21}}{s_{22}} (x_{2i} - \bar{x}_2) \right]^2$$

$$= \sum_i (x_{1i} - \bar{x}_1)^2 - 2\frac{s_{21}}{s_{22}} \sum_i (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) + \left(\frac{s_{21}}{s_{22}}\right)^2 \sum_i (x_{2i} - \bar{x}_2)^2$$

$$= s_{11} - 2\frac{s_{21}}{s_{22}} s_{12} + \left(\frac{s_{21}}{s_{22}}\right)^2 s_{22} = s_{11} - 2\frac{s_{21}s_{12}}{s_{22}} + \frac{s_{21}s_{12}}{s_{22}}$$

$$= s_{11} - \frac{s_{21}s_{12}}{s_{22}} = \frac{s_{11}s_{22} - s_{21}s_{12}}{s_{22}}$$

Hence, taking the ratio of (2.8) and (2.9) we may conclude from (2.7) together with (1.13) that

$$(2.10) \quad \hat{\beta}_1^{(2)} = \frac{s_{22}s_{1y} - s_{12}s_{2y}}{s_{11}s_{22} - s_{12}s_{21}} = \hat{\beta}_1$$

It should be clear that the above argument also applies to the partial regression of y on x_2 given x_1 [i.e., the simple regression of $y^{(1)}$ on $x_2^{(1)}$] and shows that

$$(2.11) \quad \hat{\beta}_2^{(1)} = \frac{s_{11}s_{2y} - s_{21}s_{1y}}{s_{11}s_{22} - s_{12}s_{21}} = \hat{\beta}_2$$

This same result holds for all regressors, x_j, in a general multiple regression ($j = 1,..,k \geq 2$), and provides an important alternative interpretation of the associated beta coefficients, $\hat{\beta}_j$. In particular, $\hat{\beta}_j$ is an estimate of the *partial linear relationship* between y and x_j once the effects of all other explanatory variables have been *removed*.

2.2 A Simulated Example

As one illustration of this concept, a random sample (y_i, x_{1i}, x_{2i}) , $i = 1,..,100$ was simulated using the following *linear model*:

$$(2.12)\quad Y = \beta_0 + \beta_1 \ln(x_1) + \beta_2 x_2 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) ,$$

with $\beta_0 = 3,\ \beta_1 = 4,\ \beta_2 = 2, \sigma = 2$ and $x_1 > 0$. A multiple regression was then run, under the assumption that the "true" linear model is given by (1.1) above. Hence the first regressor in this model is actually misspecified as x_1 rather than $\ln(x_1)$. The results of this regression are shown below:

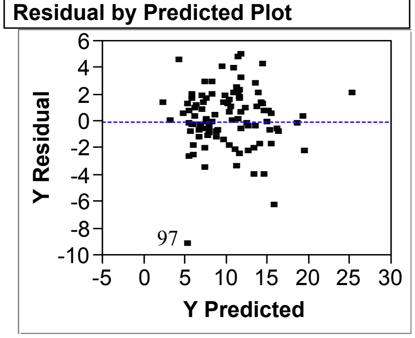
Summary of Fit

RSquare	0.766023
RSquare Adj	0.761198
Root Mean Square Error	2.224162
Mean of Response	10.1184
Observations (or Sum Wgts)	100

Parameter Estimates

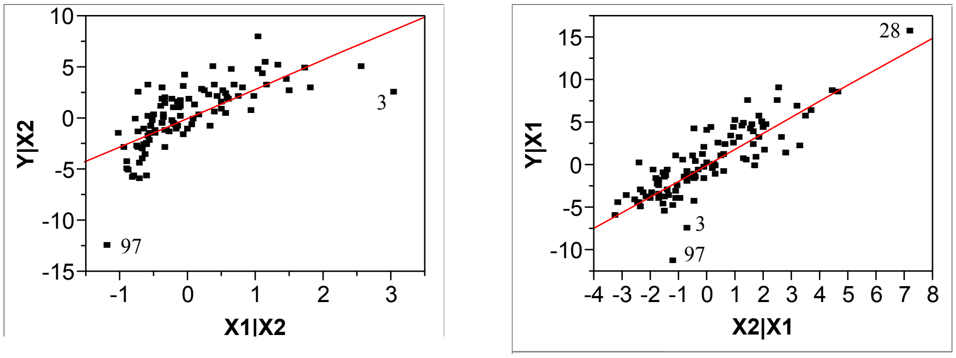
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.499853	0.640262	-0.78	0.4369
X1	2.8505178	0.286232	9.96	<.0001
X2	1.8718112	0.117053	15.99	<.0001

Residual by Predicted Plot



These results show that both regressors are highly significant, and that the root mean squared error, $s = 2.22$, is not too different from the actual residual standard deviation, $\sigma = 2$. Moreover, aside from one possible outlier (point 97), the residuals also look reasonable. But while the estimate $\hat{\beta}_2 = 1.87$ is reasonably close to the true value, $\beta_2 = 2$, the estimate $\hat{\beta}_1 = 2.85$ for the misspecified regressor is (not surprisingly) further from the true value, $\beta_1 = 4$.

The two *partial regression plots* for this regression are shown below, where in terms of the notation above: $y^{(2)}$ = Y|X2 , $x_1^{(2)}$ = X1|X2, $y^{(1)}$ = Y|X1 , and $x_2^{(1)}$ = X2|X1. In addition, simple regressions have been carried out in each plot to show that the resulting slope coefficients are indeed the same as those in the multiple regression above. Notice also that the intercepts of these regressions are in fact identically zero (subject to



Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.679e-15	0.221278	0.00	1.0000
X1|X2	2.8505178	0.284768	10.01	<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-5.96e-15	0.221278	-0.00	1.0000
X2|X1	1.8718112	0.116454	16.07	<.0001

rounding error), reflecting the fact that regression lines always go through mean points, which in this case are identically zero (as noted above).

By comparing these two partial regression plots it is also clear that the underlying misspecification is now rather evident. In particular, while the partial regression of y on x_2 (the plot on the right) is reasonably linear, the partial regression of y on x_1 (the plot on the left) shows a distinct nonlinearity. Moreover, this nonlinearity appears to be concave, indicating that perhaps a concave transformation of x_1 might yield better results. One difficulty here is that the x-axis in not in terms of the variable, x_1, but rather is in terms of the *residual* of x_1 when regressed on x_2. Hence it can often be difficult to identify the best transformation of x_1 based on this partial regression plot alone. Note in particular that since least-squares residual values are necessarily negative as well as positive, the standard concave transformations, $\sqrt{(\cdot)}$ and $\ln(\cdot)$, are not even defined for these residuals.

But practitioners have found that partial regression plots are generally very useful in identifying those data points that are most influential in determining the values of the beta estimates. For example, the multiple-regression outlier (point 97) noted above is now seen to be an outlier in both the partial regression plots, and hence to be potentially influential for both $\hat{\beta}_1$ and $\hat{\beta}_2$. However, there are some noticeable differences in these plots with respect to other points. Perhaps the best example is point 28, which is seen to be far to the right in the partial regression plot for x_2, and hence to have high leverage for $\hat{\beta}_2$. But in the partial regression plot for x_1, this point is so close to the center of the point scatter that it is hard to single out, and clearly has little effect on $\hat{\beta}_1$. Another example is point 3, which is seen to be more influential for $\hat{\beta}_1$ than $\hat{\beta}_2$. Hence, one of the main uses of partial-regression plots is to gain further information about the differential influences of possible outliers.

3. Partial Residual Plots

The results of section 1 also permit a detailed analysis of partial residual plots. Given beta estimates $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ in a multiple regression of y on (x_1, x_2), if we now let $(e_i : i = 1,..,n)$ denote the residuals obtained from this regression and let

$$(3.1) \quad e_i^{(j)} = y_i - \hat{\beta}_0 - \hat{\beta}_h x_{hi} = \hat{\beta}_j x_{ji} + e_i \ , \quad j,h = 1,2\,(j \neq h)$$

denote the i^{th} *residual for the* j^{th} *regressor* (i.e., the residual with respect to all variables other than x_j) then the *partial residual plot* for x_j is simply a plot of $e^{(j)}$ against x_j. As with partial regression plots above, a key feature of partial residual plots is that the simple linear regression of $e^{(j)}$ on x_j produces an estimated slope coefficient, $\hat{\beta}^{(j)}$, that is identical to $\hat{\beta}_j$. Our first objective is to show this, and then to compare partial residual plots with partial regression plots in terms of the above example.

3.1 Equivalence of Betas

Again we focus only on the case $j = 1$, and observe that the mean of $e^{(1)}$ is given by

$$(3.2) \quad \begin{aligned} \bar{e}^{(1)} &= \tfrac{1}{n}\sum_i e_i^{(1)} = \tfrac{1}{n}\sum_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_2 x_{2i}\right) \\ &= \tfrac{1}{n}\sum_i y_i - \hat{\beta}_0 \tfrac{n}{n} - \hat{\beta}_2 \tfrac{1}{n}\sum_i x_{2i} \\ &= \bar{y} - \hat{\beta}_0 - \hat{\beta}_2 \bar{x}_2 \end{aligned}$$

Hence the slope coefficient in the simple regression of $e^{(1)}$ on x_1 is by definition of the form:

$$(3.3) \quad \begin{aligned} \hat{\beta}^{(1)} &= \frac{\sum_i \left(e_i^{(1)} - \bar{e}^{(1)}\right)\left(x_{1i} - \bar{x}_1\right)}{\sum_i \left(x_{1i} - \bar{x}_1\right)^2} \\ &= \frac{\sum_i \left[\left(y_i - \hat{\beta}_0 - \hat{\beta}_2 x_{2i}\right) - \left(\bar{y} - \hat{\beta}_0 - \hat{\beta}_2 \bar{x}_2\right)\right]\left(x_{1i} - \bar{x}_1\right)}{s_{11}} \\ &= \frac{\sum_i \left[\left(y_i - \bar{y}\right) - \hat{\beta}_2\left(x_{2i} - \bar{x}_2\right)\right]\left(x_{1i} - \bar{x}_1\right)}{s_{11}} \\ &= \frac{\sum_i \left(y_i - \bar{y}\right)\left(x_{1i} - \bar{x}_1\right) - \hat{\beta}_2 \sum_i \left(x_{2i} - \bar{x}_2\right)\left(x_{1i} - \bar{x}_1\right)}{s_{11}} = \frac{s_{1y} - \hat{\beta}_2 s_{12}}{s_{11}} \end{aligned}$$

But recall from the first normal equation (1.10) for the multiple regression of y on (x_1, x_2) that

$$(3.4) \qquad s_{11}\hat{\beta}_1 + s_{12}\hat{\beta}_2 = s_{1y} \;\Rightarrow\; s_{1y} - \hat{\beta}_2 s_{12} = s_{11}\hat{\beta}_1$$

Hence by substituting (3.4) into (3.3) we may conclude that

$$(3.5) \qquad \hat{\beta}^{(1)} = \frac{s_{11}\hat{\beta}_1}{s_{11}} = \hat{\beta}_1$$

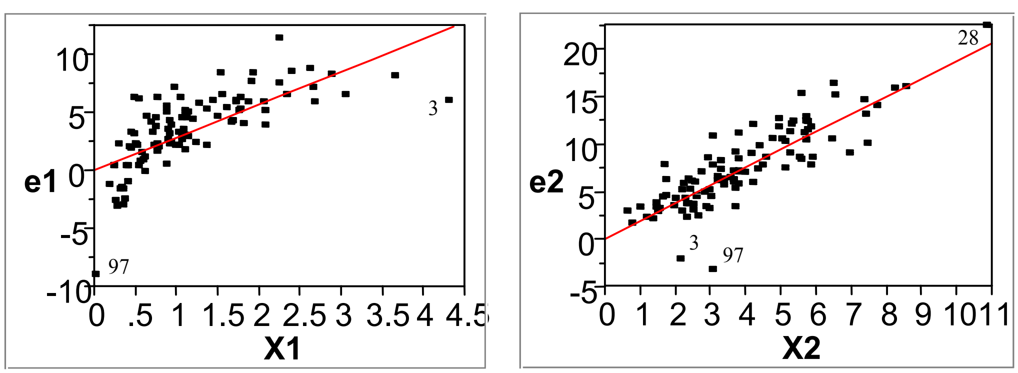
and the desired equivalence is established. As before, the same argument shows that

$$(3.6) \qquad \hat{\beta}^{(2)} = \hat{\beta}_2$$

More generally, this identity continues to hold for all variables x_j in a general multiple regression ($j = 1,..,k \geq 2$).

3.2 Continuation of Example

Given this theoretical property of partial residual plots, we now apply this concept to the example in section 2.2 above. Below are the two residual regression plots for this case, where $e^{(1)}$ = **e1** and $e^{(2)}$ = **e2**. A comparison with the partial regression plots shows that



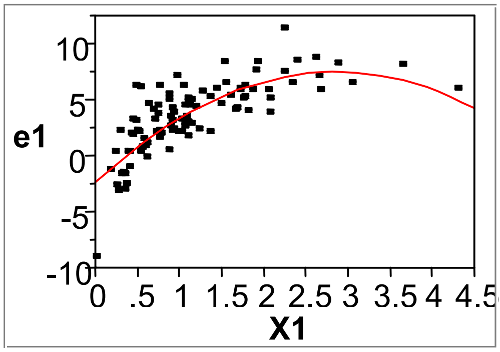
Parameter Estimates

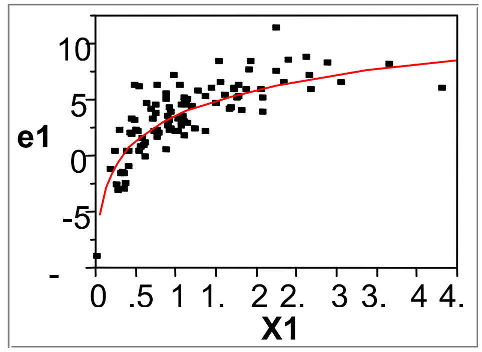
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.485e-14	0.395193	0.00	1.0000
X1	2.8505178	0.282142	10.10	<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-2.66e-15	0.502004	-0.00	1.0000
X2	1.8718112	0.11538	16.22	<.0001

these two sets of point patterns (including the influential points 3, 97, and 28) are very similar. The key difference for analysis is that here the *x*-axis is directly in terms of the regressor of interest. This means that the shape of the scatter plot can be used to fit candidate respecifications of this regressor. In the present case, a comparison of quadratic and log fits is shown. Here it is clear both visually and in terms of adjusted R-square that





Polynomial Fit Degree=2

e1 = -0.735595 + 4.128211 X1 - 1.2147809 (X1-1.16053)^2

Summary of Fit

RSquare	0.642581
RSquare Adj	0.635211
Root Mean Square Error	1.899926

Transformed Fit to Log

e1 = 3.6309514 + 3.3013465 Log(X1)

Summary of Fit

RSquare	0.699451
RSquare Adj	0.696384
Root Mean Square Error	1.733318

the log fit is superior to the quadratic fit. This log specification can be operationalized by simply constructing a new column "lnX1" in JMPIN, and running a multiple regression of Y on lnX1 and X2. The results of this regression are shown below. A comparison of

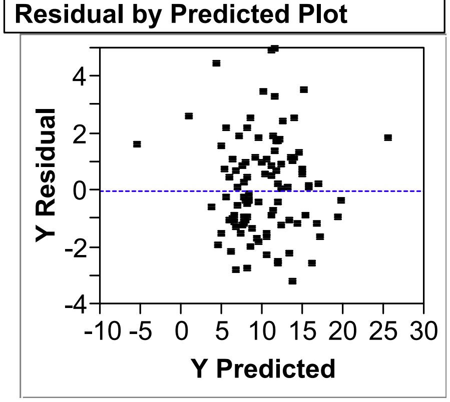
Summary of Fit

RSquare	0.856468
RSquare Adj	0.853508
Root Mean Square Error	1.742025
Mean of Response	10.1184
Observations (or Sum Wgts)	100

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.1845392	0.395533	8.05	<.0001
lnX1	3.2978242	0.22094	14.93	<.0001
X2	1.8580392	0.091346	20.34	<.0001

Residual by Predicted Plot



these results with the initial regression show that the adjusted R-square has increased by almost 10 percentage points. Moreover, the residual plot no longer shows any serious outliers, and appears to be consistent with the Gauss-Markov Assumptions. In the present simulation example, this improvement is hardly surprising since the partial residual analysis above has in fact led us to the exact model from which the data was simulated.

While one cannot expect to be so fortunate in practice, this example does serve to illustrate the potential power of partial residual plots for improving the specifications of regressors in multiple regressions. Note also that since the resulting point pattern is usually similar in nature to that of the partial regression plots, it is often possible to identify influential points for each regressor. (However, it has been the experience of practitioners that partial regression plots are generally the best tool for this purpose.)