University of Illinois Department of Economics
Spring 2006 Economics 471 Roger Koenker

Lecture 4
Partial Residual Plots

A useful and important aspect of diagnostic evaluation of multivariate regression models is the partial residual plot. We illustrate technique for the gasoline data of PS 2 in the next two groups of figures. In the first group of 4 figures I plot in the upper two panels the scatterplots of percapita US gasoline demand vs percapita income and the price of gasoline respectively. Note that neither of these plots look like something a reasonable person would want to fit with a straight line. Nevertheless, we need to become accustomed to the idea that the *multivariate* relationship may be approximately linear even if the bivariate relationships are not. In this case this is (encouragingly!) roughly true.

The partial residual plot is a device for representing the final step of a multivariate regression result as a bivariate scatterplot. To accomplish this slightly mysterious feat, we need somehow to "remove" the effect of the "other" variables before doing the scatterplot. The natural way of doing this is to regress the two variables of primary interest on the "other" variables of the model, and then plot the resulting residuals against one another. This can be formalized in the following way.

Consider the model

$$y = x\beta + z\gamma + u$$

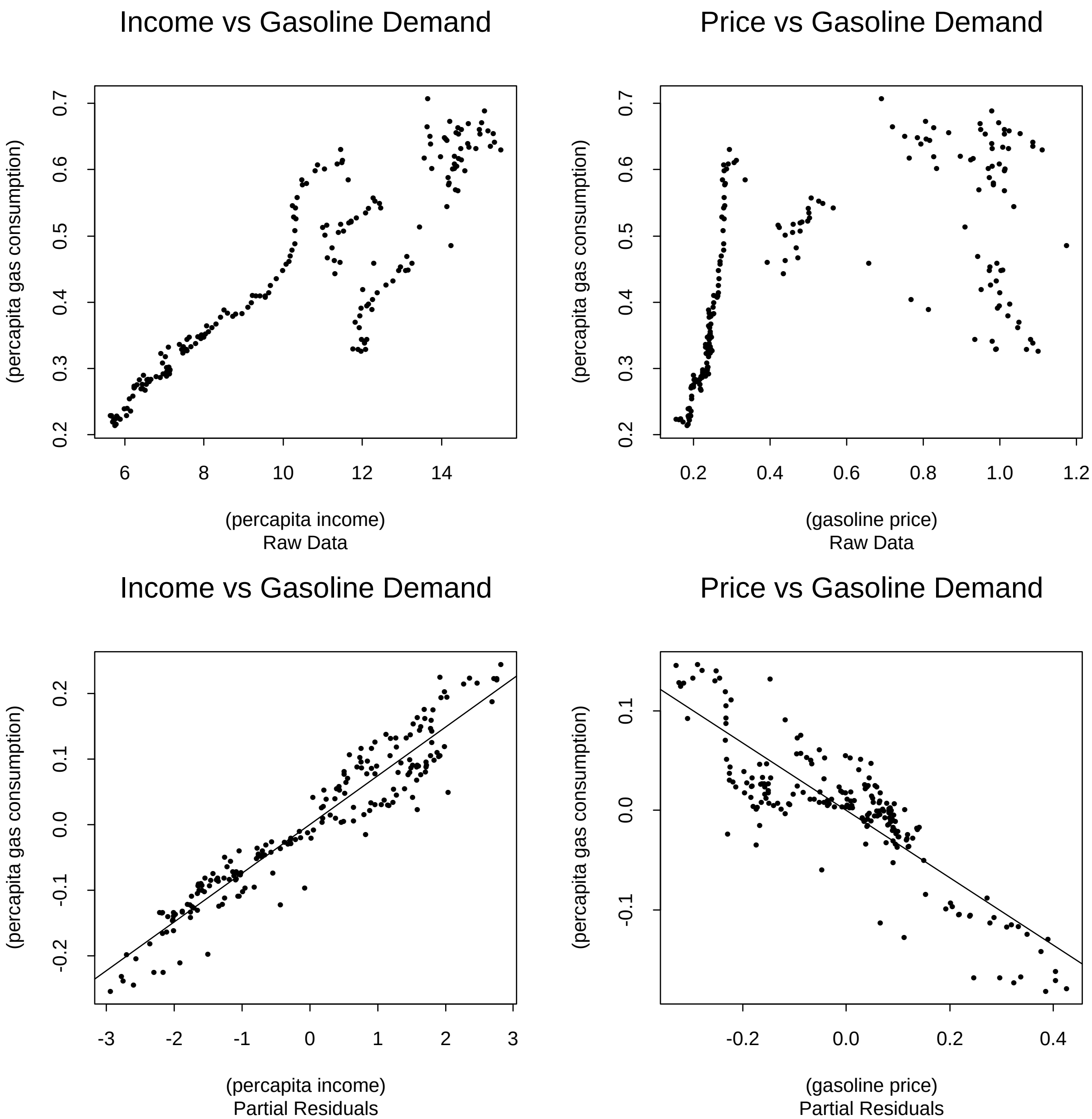
and the least squares fitted values,

$$\hat{y} = x\hat{\beta} + z\hat{\gamma}.$$

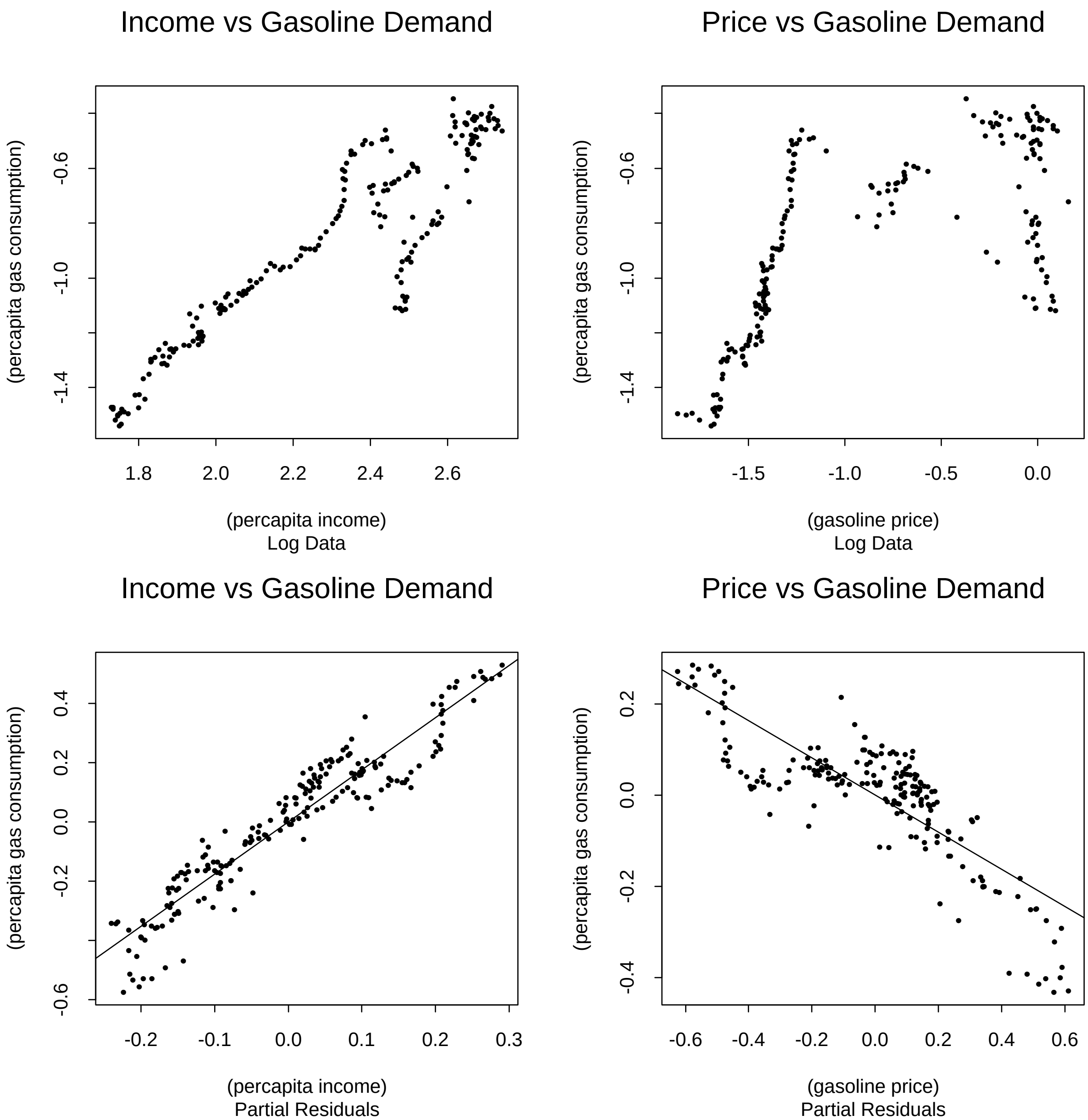
The partial residual plot carries out the regression of y on x and z in two stages: first, we regress y and z on x and compute the residuals, say $\tilde{y}$ and $\tilde{z}$: second, we regress $\tilde{y}$ on $\tilde{z}$. The coefficient obtained in the second regression is precisely the same as would be obtained by carrying out the full regression. We have seen this already in a very special case when the "variable" x is just the intercept. In fact, we can generalize the result to cases in which x stands for several covariates.

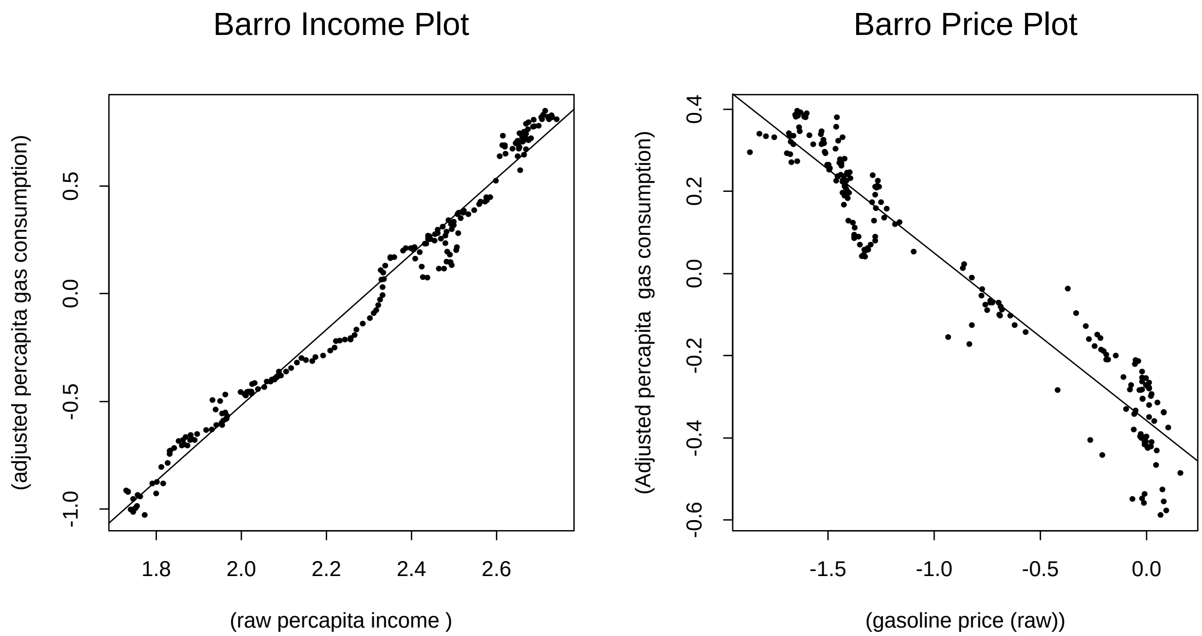
An additional feature of this approach is that the standard errors that would be computed by the last step are exactly the same as those that come out the of full regression. So not only does the scatter plot give an accurate assessment of the position of the least squares fit of the bivariate relationship, it also provides an accurate visual assessment of the precision of this estimate.

This last point inevitably recalls an amusing "fiasco of econometrics" perpetrated here by Robert Barro in his 1997 David Kinley lecture. Barro presented some "growth regressions" of the type described in his monograph with Sala i Martin. But to illustrate the results for a "general audience" he chose to spend a considerable portion of the talk showing slides of the bivariate relationship between various explanatory variables and his "national growth" variable, after controlling for the effect of other variables. Barro's approach was, however, somewhat idiosyncratic. For each of the possible z variables of interest, he computed $\tilde{y} = \hat{u} + z_i\beta_i$, that is the residuals from the full regression plus the estimated effect of the ith variable, and then this variable was centered at zero and plotted against z_i. This approach is easily shown to produce the correct point estimate of the coefficient $\hat{\beta}_i$, (it would be a useful exercise for students to show this), however the visual impression of the scatterplot is much more optimistic than one would expect to get from the partial residual plot method described above. In effect, the denominator effect in the t-statistics of the variance the residuals of the regression of z on all of X is replaced in the Barro approach by regression on only an intercept. Since the former can be very small,



compared to the later the result is a picture that has an implied standard error that appears considerably smaller than would the standard error in the full regression. In the next panel of two figures, I illustrate the effect of the Barro approach for the gasoline data. As one can see, these figures suggest a much more precise estimate of the two effects than that conveyed by the conventional partial residual plots appearing above. This is good, of course, if the object is to impress the viewer with the precision of the fit, but bad if one is interested in conveying an accurate assessment of that precision.


Income vs Gasoline Demand
(percapita gas consumption)
-1.4
-1.0
-0.6
1.8
2.0
2.2
2.4
2.6
(percapita income)
Log Data
Price vs Gasoline Demand
(percapita gas consumption)
-1.4
-1.0
-0.6
-1.5
-1.0
-0.5
0.0
(gasoline price)
Log Data
Income vs Gasoline Demand
(percapita gas consumption)
-0.6
-0.2
0.0
0.2
0.4
-0.2
-0.1
0.0
0.1
0.2
0.3
(percapita income)
Partial Residuals
Price vs Gasoline Demand
(percapita gas consumption)
-0.4
-0.2
0.0
0.2
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
(gasoline price)
Partial Residuals


Barro Income Plot
(adjusted percapita gas consumption)
-1.0
-0.5
0.0
0.5
1.8
2.0
2.2
2.4
2.6
(raw percapita income)
Barro Price Plot
(Adjusted percapita gas consumption)
-0.6
-0.4
-0.2
0.0
0.2
0.4
-1.5
-1.0
-0.5
0.0
(gasoline price (raw))