

Chapter 8

AER and Time Series

Additive error regression and some time series models are similar to multiple linear regression for response plots and prediction intervals.

8.1 Additive Error Regression

Definition 8.1. The *additive error regression* (AER) model is

$$Y = m(\mathbf{x}) + e \tag{8.1}$$

where m is a real valued function and the errors e_i are iid with zero mean and finite variance σ^2 . The AER model is a 1D regression model with sufficient predictor $SP = h(\mathbf{x}) = m(\mathbf{x}) = E(Y|\mathbf{x})$. The estimated sufficient predictor $ESP = \hat{m}(\mathbf{x}) = \hat{Y}$, and the residual $r = Y - \hat{Y}$. We will usually assume that the error distribution is not highly skewed.

Definition 8.1. The response plot for the AER model is a plot of ESP versus Y . The residual plot is a plot of ESP versus r .

Rule of thumb 8.1. If the error distribution is unimodal and not highly skewed, the plotted points should follow the identity line in the response plot and the $r = 0$ line in the residual plot with a rectangular or ellipsoidal pattern. Hence the plots look like those for multiple linear regression when the error distribution is unimodal and not highly skewed. Add the identity line to the response plot. Pointwise prediction interval bands can also be added.

Remark 8.1 Prediction intervals for the AER model were given in Section 7.12.

Many regression models are special cases of the AER model. The multiple linear regression model is a special case with $m(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}$. Then AER single

index model is a special case with $m(\mathbf{x}) = g(\mathbf{x}^T\boldsymbol{\beta})$. For this model, $m(\mathbf{x})$ and $\mathbf{x}^T\boldsymbol{\beta}$ are both sufficient predictors. See Chapter 9. Nonlinear regression and nonparametric regression are also special cases. The nonlinear regression model has $m(\mathbf{x}) = g\boldsymbol{\theta}(\mathbf{x})$, a known function except the k unknown parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^T$. The additive error generalized additive model (AE GAM) has $m(\mathbf{x}) = SP = AP = \alpha + \sum_{j=1}^p S_j(x_j)$ for some (usually unknown) functions S_j . Then $ESP = EAP = \hat{\alpha} + \sum_{j=1}^p \hat{S}_j(x_j)$. The AER GAM is useful for checking the multiple linear regression model: check that each \hat{S}_j linear.

Multiple linear regression uses an inflexible hyperplane $\hat{m}(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}$. Many AER fitting methods use flexible estimators $\hat{m}(\mathbf{x})$. These flexible methods often fit outliers well so the outliers are masked. Hence, outlier detection tends to be more difficult for AER than for MLR. In the response and residual plots, look for gaps in the plot with clusters of outliers far from the bulk of the data.

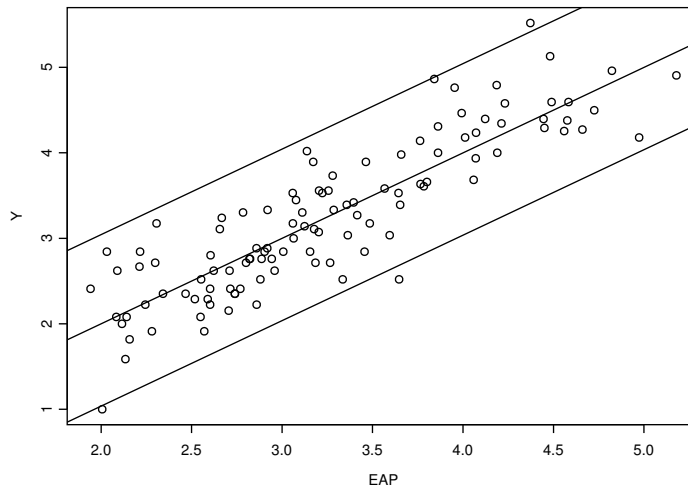


Fig. 8.1 Pointwise Prediction Interval Bands for Ozone Data

Example 8.1. Chambers and Hastie (1993, p. 251, 516) examine an environmental study that measured the four variables $Y =$ ozone concentration, solar radiation, temperature, and wind speed for $n = 111$ consecutive days. Figure 8.1 shows the response plot with the pointwise large sample 95% PI bands for the additive model. Here $\hat{m}(\mathbf{x}) =$ estimated additive predictor (EAP). Note that the plotted points scatter about the identity line in a

roughly evenly populated band, and that 3 of the 111 PIs corresponding to the observed data do not contain Y .

8.1.1 Response Transformations

This subsection extends the graphical method for response transformations of Section 5.2 to regression models of the form $Y_i = m(\mathbf{x}_i) + e_i$. Predictor transformations from Section 5.1 are still useful for such models.

The applicability of the AER model (8.1) can be expanded by allowing response transformations. An important class of *response transformation models* adds an additional unknown transformation parameter λ_o , such that

$$Y_i = t_{\lambda_o}(Z_i) \equiv Z_i^{(\lambda_o)} = m(\mathbf{x}_i) + e_i. \quad (8.2)$$

If λ_o was known, then $Y_i = t_{\lambda_o}(Z_i)$ would follow model (8.1). The function m depends on λ_o , the p predictors x_j are assumed to be measured with negligible error, and the zero mean constant variance errors e_i are assumed to be iid from a unimodal distribution that is not highly skewed. The power transformation and modified power transformations of Definitions 5.6 and 5.7 are again used.

A graphical method for response transformations refits the model using the same fitting method: changing only the “response” from Z to $t_\lambda(Z)$. Compute the “fitted values” \hat{W}_i using $W_i = t_\lambda(Z_i)$ as the “response.” Then a *transformation plot* of \hat{W}_i versus W_i is made for each of the seven values of $\lambda \in \Lambda_L$ with the identity line added as a visual aid. Vertical deviations from the identity line are the “residuals” $r_i = W_i - \hat{W}_i$. Then a candidate response transformation $Y = t_{\lambda^*}(Z)$ is reasonable if the plotted points follow the identity line in a roughly evenly populated band. Then take $\hat{\lambda}_o = \lambda^*$, that is, $Y = t_{\lambda^*}(Z)$ is the response transformation. Curvature from the identity line suggests that the candidate response transformation is inappropriate. After selecting the transformation, the usual checks should be made. In particular, the transformation plot for the selected transformation is a response plot, and a residual plot should also be made.

Each transformation plot is a “response plot” for the seven values of $W = t_\lambda(Z)$, and the method chooses the “best response plot” where the model (8.1) seems “most reasonable.” If more than one value of $\lambda \in \Lambda_L$ gives a linear plot, take the simplest or most reasonable transformation or the transformation that makes the most sense to subject matter experts. Also check that the corresponding “residual plots” of \hat{W} versus $W - \hat{W}$ look reasonable. The values of λ in decreasing order of importance are 1, 0, 1/2, -1 and 1/3. So the log transformation would be chosen over the cube root transformation if both transformation plots look equally good. Note that this procedure can be modified to create a graphical diagnostic for a numerical estimator $\hat{\lambda}$ of

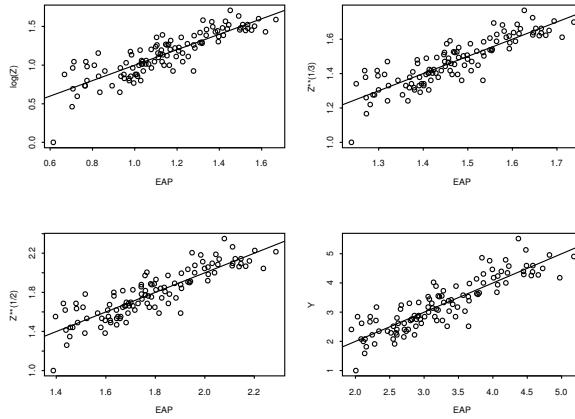


Fig. 8.2 Transformation Plots for Ozone Data

λ_o by adding $\hat{\lambda}$ to A_L . For linear models, Box and Cox (1964) is widely used. Useful powers are $\pm 1/4$, $\pm 2/3$, ± 2 , and ± 3 . Powers from numerical methods can also be added.

In the following example, the plots show $t_\lambda(Z)$ on the vertical axis. The label “EAP” of the horizontal axis is for the fitted values that result from using $t_\lambda(Z)$ as the “response” in the software.

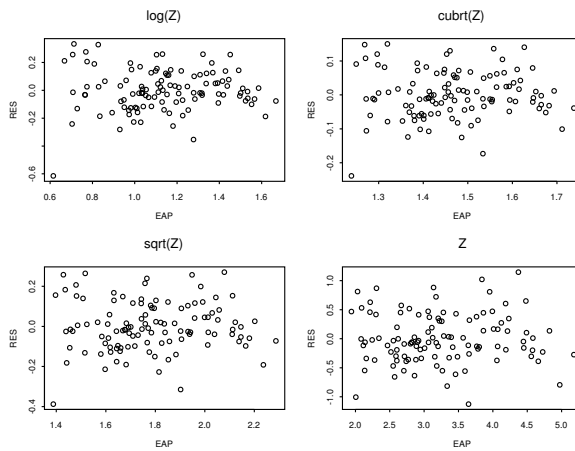


Fig. 8.3 Residual Plots for Ozone Data

Response transformations for the AE GAM $Y = AP + e$ are among the most difficult for regression models with additive errors since AE GAMs are very flexible and tend to fit more than one candidate response transformation well. Rule out poor models with transformation and residual plots. For each remaining competing model, check the \hat{S}_j and whether any of the predictors can be deleted.

Example 8.2. Chambers and Hastie (1993, p. 251, 516) examine an environmental study that measured the four variables $Z = \text{ozone concentration, solar radiation, temperature and wind speed}$ for 111 consecutive days. Additive models were fit using Z and $Z^{1/3}$ as the response. Figure 8.2 shows the four best transformation plots, and Figure 8.3 shows the corresponding residual plots. The plotted points scatter about the identity line and $r = 0$ line in roughly evenly populated bands except possibly the case that appears in the lower left corner. No transformation $Y = Z$ may be best since the predictor *solar radiation* does not seem to be needed for this model, and the other transformations fit the case in the lower left corner poorly.

Similar graphical methods for response transformations can be used for time series, which are covered briefly in the next section.

8.2 Time Series

8.3 Summary

8.4 Complements

See Olive (2007, 2013a) and Pelawa Watagoda and Olive (2020) for prediction intervals for AER. The graphical response transformation method is due to Olive (2013b).

Experimental design models are often AER models. Response transformations for such models are given in Olive (2017a, § 5.4).

8.5 Problems

8.1. When doing a PI or CI simulation for a nominal $100(1 - \delta)\% = 95\%$ interval, there are m runs. For each run, a data set and interval are generated, and for the i th run $Y_i = 1$ if μ or Y_f is in the interval, and $Y_i = 0$, otherwise. Hence the Y_i are iid Bernoulli($1 - \delta_n$) random variables where $1 - \delta_n$ is the true probability (true coverage) that the interval will contain μ or Y_f . The observed coverage (= coverage) in the simulation is $\bar{Y} = \sum_i Y_i/m$. The

variance $V(\bar{Y}) = \sigma^2/m$ where $\sigma^2 = (1 - \delta_n)\delta_n \approx (1 - \delta)\delta \approx (0.95)0.05$ if $\delta_n \approx \delta = 0.05$. Hence

$$SD(\bar{Y}) \approx \sqrt{\frac{0.95(0.05)}{m}}.$$

If the (observed) coverage is within $0.95 \pm kSD(\bar{Y})$ the integer k is near 3, then there is no reason to doubt that the actual coverage $1 - \delta_n$ differs from the nominal coverage $1 - \delta = 0.95$ if $m \geq 1000$ (and as a crude benchmark, for $m \geq 100$). In the simulation, the length of each interval is computed, and the average length is computed. For intervals with coverage $\geq 0.95 - kSD(\bar{Y})$, intervals with shorter average length are better (have more precision).

a) If $m = 5000$ what is $3 SD(\bar{Y})$, using the above approximation? Your answer should be close to 0.01.

b) If $m = 1000$ what is $3 SD(\bar{Y})$, using the above approximation?

8.2. The smoothing spline simulation compares the PI lengths and coverages of 3 large sample 95% PIs for $Y = m(x) + e$ and a single measurement x . Values for the first PI were denoted by *scov* and *slen*, values for 2nd PI were denoted by *ocov* and *olen*, and values for third PI by *dcov* and *dlen*. The average degrees of freedom of the smoothing spline was recorded as *adf*. The number of runs was 5000. The *len* was the average length of the PI and the *cov* was the observed coverage. One student got the following results shown in Table 4.2.

Table 8.1 Results for 3 PIs

error	95%	PI	95%	PI	95%	PI		
type	n	slen	olen	dlen	scov	ocov	dcov	adf
5	100	18.028	17.300	18.741	0.9438	0.9382	0.9508	9.017

For the PIs with coverage ≥ 0.94 , which PI was the most precise (best)?

R Problems Some *R* code for homework problems is at (<http://parker.ad.siu.edu/Olive/robRhw.txt>).

Warning: Use a command like `source("G:/rpack.txt")` to download the programs. See Preface or Section 11.2. Typing the name of the `rpack` function, e.g. `regbootsim3`, will display the code for the function. Use the `args` command, e.g. `args(regbootsim3)`, to display the needed arguments for the function.

8.6. A problem with response and residual plots is that there can be a lot of black in the plot if the sample size n is large (more than a few thousand). A variant of the response plot for the additive error regression model $Y = m(\mathbf{x}) + e$ would plot the identity line, the two lines parallel to the identity line corresponding to the Section 7.12 large sample $100(1 - \delta)\%$ prediction intervals for Y_f that depends on \hat{Y}_f . Then plot points corresponding

to training data cases that do not lie in their $100(1 - \delta)\%$ PI. We will use $\delta = 0.01$, $n = 100000$, and $p = 8$.

a) Copy and paste the commands for this part into *R*. They make the usual response plot with a lot of black. Do not include the plot in *Word*.

b) Copy and paste the commands for this part into *R*. They make the response plot with the points within the pointwise 99% prediction interval bands omitted. Include this plot in *Word*. For example, left click on the plot and hit the *Ctrl* and *c* keys at the same time to make a copy. Then paste the plot into *Word*, e.g., get into *Word* and hit the *Ctrl* and *v* keys at the same time.

c) The additive error regression model is a 1D regression model. What is the sufficient predictor $= h(\mathbf{x})$?

8.7. The Rousseeuw and Leroy (1987, p. 26) Belgian telephone data has response $Y = \text{number of international phone calls}$ (in tens of millions) made per year in Belgium. The predictor variable $x = \text{year}$ (1950-1973). From 1964 to 1969 total number of minutes of calls was recorded instead, and years 1963 and 1970 were also partially effected. Hence there are 6 large outliers and 2 additional cases that have been corrupted.

a) The simple linear regression model is $Y = \alpha + \beta x + e = SP + e$. Copy and paste the *R commands* for this part to make a response plot of $ESP = \hat{Y} = \hat{\alpha} + \hat{\beta}x$ versus Y for this model. Include the plot in *Word*.

b) The additive error GAM is $Y = \alpha + S(x) + e = AP + e$ where S is some unknown function of x . The *R commands* make a response plot of $EAP = \hat{\alpha} + \hat{S}(x)$ versus Y for this model. Include the plot in *Word*.

c) The simple linear regression model is a special case of the additive error GAM with $S(x) = \beta x$. The additive error GAM is a special case of the additive error regression model $Y = m(x) + e$ where $m(x) = \alpha + S(x)$. The response plots for these three models are used in the same way as the response plot for the multiple linear regression model: if the model is good, then the plotted points should cluster about the identity line with no other pattern. Which response plot is better for showing that something is wrong with the model? Explain briefly.