Multiple Linear Regression: Extensions.

A. Skripnikov[1]

[1]New College of Florida

IDC 5205

Extensions to Linear Models.

Advantages of classic linear regression models:

- in many cases they work quite well from prediction standpoint,
- they provide interpretable results

On the other hand, they make some restrictive assumptions:

1. **Additivity**: "Per 1-unit increase in X, holding other predictors constant, Y will change by β ($\equiv$ **constant**), on average."

   The effect of changes in X on the response Y is **constant**, independent of **what values** the other predictors are held at.

2. **Linearity**: Y constitutes a linear function of $X_1, X_2, \dots, X_p$,

$$Y \approx f(X_1, X_2, \dots, X_p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

Removing the Additivity Assumption: Interaction Effect.

Example. *TV* and *radio* advertisement was strongly related with *Sales* in

$$\widehat{Sales} = \hat{\beta}_0 + \hat{\beta}_1 TV + \hat{\beta}_2 radio$$

stating that, e.g.,

"$1,000$\$ increase in *TV*, holding *radio* constant, yields a $\hat{\beta}_1$ (**constant**, not dependent on *radio*) change in *Sales*, on average."

Q: What if money spent on radio actually increase the effectiveness ($\hat{\beta}_1$) of TV advertising?

E.g. for an overall advertising budget of $100,000$\$,

- instead of allocating the entire $100,000$\$ to one media (TV or radio),
- it was more effective to spend $50,000$\$ on each?

In marketing this is known as **synergy** effect, while in **statistics** - **interaction** effect (**between** *TV* and *radio* budgets in **predicting** *sales*).

Interaction Effect.

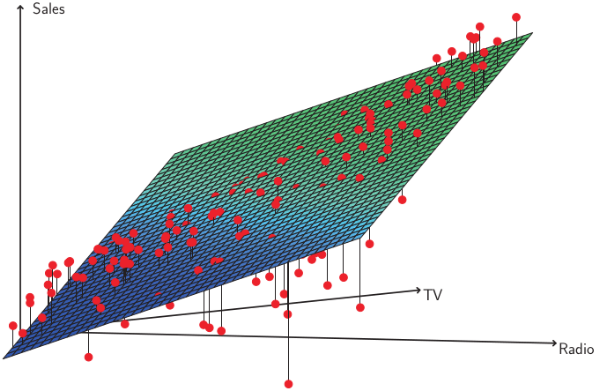


FIGURE 3.5. *For the* `Advertising` *data, a linear regression fit to* `sales` *using* `TV` *and* `radio` *as predictors. From the pattern of the residuals, we can see that there is a pronounced non-linear relationship in the data. The positive residuals (those visible above the surface), tend to lie along the 45-degree line, where TV and Radio budgets are split evenly. The negative residuals (most not visible), tend to lie away from this line, where budgets are more lopsided.*

Interaction Effect: Advertising.

To model an **interaction**, we include **interaction term** $TV \times radio$:

$$sales = \beta_0 + \beta_1 TV + \beta_2 radio + \beta_3 (TV \times radio) + \epsilon$$

Task. Now we could write:

$$sales =$$

$$= \beta_0 + \tilde{\beta}_1 TV + \beta_2 radio + \epsilon,$$

where

$$\tilde{\beta}_1 = \qquad \Longrightarrow$$

effect $\tilde{\beta}_1$ (of TV on $sales$) depends on **value of** $radio$ $\Longrightarrow$

effect of TV on $Sales$ is no longer constant $\Longrightarrow$

adjusting **radio** will change the **impact of** TV on $Sales$.

Interaction Effect: Productivity of a Factory.

A more **intuitive** example on **interaction effect**:

Example. Predict the $Y = \{\#\text{ of units}\}$ produced at a factory, based on

- $X_1 = \{\#\text{ of production lines}\}$,
- $X_2 = \{\#\text{ of workers}\}$

If no workers are available to operate the lines, increasing the # of lines will not be as effective in increasing factory's production $\Longrightarrow$

there's an **interaction** effect between X_1 (# lines) and X_2 (# workers) **in explaining** Y (# units produced)

Specifically, presume:

$$units = 1.2 + 3.4\ lines + 0.22\ workers + 1.4\ (lines \times workers)$$

Additional line increases the # of produced units by

Hence, more workers $\Longrightarrow$

Interaction Effect: Advertising.

Example. Back to *Advertising* data:

$$Sales = \beta_0 + \beta_1 TV + \beta_2 radio + \beta_3 (TV \times radio) + \epsilon \tag{1}$$

$\beta_3 = \{$change in **effect** of TV ads per 1-unit increase in radio ads (or vice-versa)$\}$

Task: In terms of model (1) parameters, formulate hypotheses to answer

"Is there a significant interaction between TV and radio?"

	Coefficient	Std. error	t-statistic	p-value
Intercept	6.7502	0.248	27.23	< 0.0001
TV	0.0191	0.002	12.70	< 0.0001
radio	0.0289	0.009	3.24	0.0014
TV×radio	0.0011	0.000	20.73	< 0.0001

Conclusion?

Interaction Effect: Advertising.

Task: According to model (1) and table on previous slide, proceed to

1. **(See *R* code)** Fit the model, write down the **fitted** equation.
2. Interpret effect that 1000$ increase in *TV* budget has on *sales*.
3. Interpret effect that 1000$ increase in *radio* budget has on *sales*.

Interaction Effect: Advertising.

Task: According to fitted model

$$\hat{sales} = 6.75 + 0.019\ TV + 0.029\ radio + 0.0011\ (TV \times radio)$$

- Interpret the interaction coefficient $\hat{\beta}_3 = 0.0011$.

- Interpret coefficients $\hat{\beta}_1 = 0.0191$ and $\hat{\beta}_2 = 0.029$.

Interaction Effect: Categorical Variable.

We checked interaction between two **quantitative** variables.

Q: How about interaction between **quantitative** and **categorical** variable?

Example. In *Credit* data set, let's use customers'

- *income* (quantitative) and
- *student* status (categorical)

to model the credit balance.

In the absence of an interaction term, the model becomes

where

$$D_{stud} = \begin{cases} 1, & i^{th} \text{ person is a student} \\ 0, & i^{th} \text{ person not a student} \end{cases}$$

Example (ct'd). This results in:

$$balance_i = \beta_1 income_i + \begin{cases} \beta_0 + \beta_2 \cdot 1 + \epsilon_i, & i^{th} \text{ person is a student} \\ \beta_0 + \beta_2 \cdot 0 + \epsilon_i, & i^{th} \text{ person not a student} \end{cases}$$

This amounts to fitting two parallel lines (**see** *R* **code**):

- one for students: $\widehat{balance}_i = \hat{\beta}_1 income_i + \ldots$
- one for non-students: $\widehat{balance}_i = \hat{\beta}_1 income_i + \ldots$

Limitaton?? Same slope $\hat{\beta}_1$ means that the effect income has on credit balance does **not** depend on whether or not the person is a student.

Interaction Effect: Categorical Variable.

This limitation is addressed by adding **interaction** term $\boldsymbol{income \times D_{stud}}$:

$$balance_i = \beta_0 + \beta_1 income_i + \beta_2 D_{stud} + \beta_3 (income_i \times D_{stud}) + \epsilon_i, \ \epsilon_i \sim_{ind} N(0, \sigma^2) \quad (2)$$

and the model now becomes

$$balance_i = \begin{cases} & , \quad i^{th} \text{ person - stud} \\ & , \quad i^{th} \text{ person - not stud} \end{cases}$$

$$= \begin{cases} & , \quad i^{th} \text{ person - stud} \\ & , \quad i^{th} \text{ person - not stud} \end{cases}$$

have two regression lines, but now with potentially differing slopes:

- $Slope = \beta_1 + \beta_3$ for students,
- $Slope = \beta_1$ for non-students

Interaction Effect: Categorical Variable.

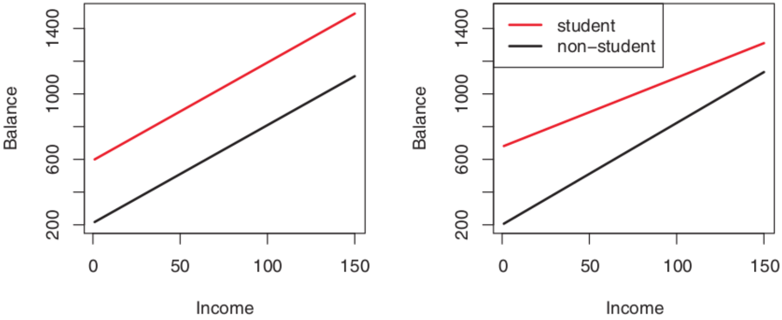


FIGURE 3.7. *For the* Credit *data, the least squares lines are shown for prediction of* balance *from* income *for students and non-students.* Left: *The model (3.34) was fit. There is no interaction between* income *and* student. Right: *The model (3.35) was fit. There is an interaction term between* income *and* student.

- Witnessing non-parallel lines after including interaction term points to **(POTENTIAL)** presence of an interaction effect.

Example (ct'd). For the model

$$balance_i = \beta_0 + \beta_1 income_i + \beta_2 D_{stud} + \beta_3 (income_i \times D_{stud}) + \epsilon_i,$$

proceed to

- Formulate hypotheses to test for interaction.

- **Fit the model in** R, write down the fitted model equation. Subsequently, provide conclusion to signif test for interaction.

Example (ct'd). For the fitted model

$$\widehat{balance} = 200.623 + 6.218\ income + 476.676\ D_{stud} - 2\ (income_i \times D_{stud}),$$

- Interpret the interaction coefficient ($\hat{\beta}_3 = -2$).

- Interpret coefficients $\hat{\beta}_1 = 6.218, \quad \hat{\beta}_2 = 476.676.$

DON'T Test **Main** Effects in Models **WITH** Interaction.

In equation

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + \epsilon_i, \ \epsilon_i \sim_{i.i.d} N(0, \sigma^2)$$

the effects β_1 and β_2 are known as **main effects**, and are considered **marginal** to the **interaction term**.

Rule: If variables happen to have a strong interaction effect, DO NOT test or interpret their main(!) effects.

Example (ct'd). Judging by interpretations of $\hat{\beta}_1$ and $\hat{\beta}_2$ on previous slide: did these sound extremely useful, and "all-encompassing"? Explain.

DON'T Test **Main** Effects in Models **WITH** Interaction.

Rule (ct'd): If the **interactions** are **not significant**, you should

1. **drop** them, and
2. only **then** interpret the **main effects**.

Example (ct'd, in *R*). For *Credit* example,

1. Is the interaction term significant? If not, proceed to drop it and re-fit the model without it. Provide the resulting model fit.

2. Interpret the $\hat{\beta}_1$ and $\hat{\beta}_2$ terms in the re-fitted model. Are the interpretations more useful & "all-encompassing" now?

Modeling Interaction with $K > 2$ Levels.

Interactions easily extend to categorical predictors with $\boldsymbol{K > 2}$ levels:

We require one **interaction term** for a product of each **dummy variable** with a **quantitative explanatory variable**

Example. In *Carseats* data,

- *ShelveLoc* has 3 levels, encoded via 2 dummies:

$$D_{GoodLoc}, D_{MedLoc}$$

- *Income* is a quantitative variable,

Then, to include interaction between *ShelveLoc* and *Income*, we add

$$Income \times D_{GoodLoc}, Income \times D_{MedLoc}$$

Modeling Interaction with $K > 2$ Levels: Example.

Example (cont'd, with R). For

$$Sales \sim Income + ShelveLoc + Income{:}ShelveLoc$$

1. Write the **full modeling** equation with all interaction terms.

2. Having fitted the model in R, write the **fitted** equation.

3. Comment on significance of interaction terms. Generally, what should be our following steps?

Interpreting Models With Interactions: Separate Equations.

How to interpret Models with Interactions?

Method #1: Write out the implied regression equation for **each group**.

Example (ct'd).

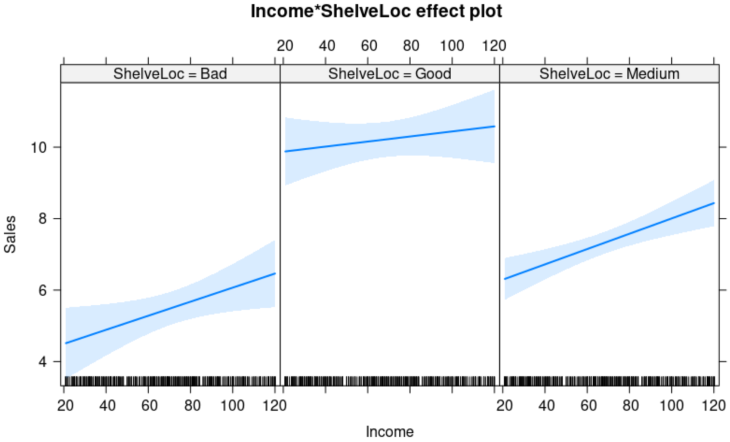
4. Write down separate **fitted** equations for each shelve location.

5. Via the equations in **p. 4**, interpret the differences across shelve locations with regards to effects of income on sales.

Interpreting Models With Interactions: Effect Displays.

Method #2: **Effect displays** that examine model's **high-order terms**.

Example. Effect display for *Income* $\times$ *ShelveLoc* interaction:



Other predictors are **fixed** at their **average**, or **majority** (if categ.), values.

Interpreting Models With Interactions: Effect Displays.

Method #2: **Effect displays** that examine model's **high-order terms**.

Example (ct'd). If you'd like, you could put all lines on the same plot:

Hypothesis Tests for Interactions.

Example (*Credit* cont'd). Our full modeling equation is

$$Sales_i = \beta_0 + \beta_1 Income_i + \beta_2 D_{LocGood,i} + \beta_3 D_{LocMed,i} +$$

$$\beta_4 (Income_i \times D_{LocGood,i}) + \beta_5 (Income_i \times D_{LocMed,i}) + \epsilon_i, \ \epsilon_i \sim_{i.i.d.} N(0, \sigma^2)$$

Q #1: What are the hypotheses if one wants to test for significance of

- shelve location variable?

- *Income* $\times$ *ShelveLoc* interaction?

Q #2: To address the above hypotheses, which test can be used?

Hypothesis Tests for Interactions: Incremental F-test.

Task. Outline **incremental F-test** for significance of interaction ("null" vs "full" model, F-statistic, degrees of freedom for F-distribution).

CAUTION REMARKS: Interaction $\neq$ Association.

CAUTION: Interaction and correlation of explanatory variables are empirically and logically distinct phenomena.

- Explanatory variables A & B can **interact** in explaining response Y, while not being related to one another statistically (e.g. correlated).
- Vice versa: Explanatory variables A & B can be related to one another statistically, but not interact in explaining response Y.
- Interaction refers to the relationship between a **combination** of explanatory variables A & B (e.g. $A \times B$) and **response variable** Y, **NOT** to the **direct relationship** between the explanatory variables A & B themselves.

Please see R **code** for examples of

- *Interaction* $\nRightarrow$ *Association*,
- *Association* $\nRightarrow$ *Interaction*.

Removing Linearity Restriction: Polynomial Regression.

Interaction terms tackled the restriction of **additivity** for linear models.

In order to alleviate the other restriction - **linearity** -

$$Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

we introduce **polynomial** regression.

Polynomial regression involves adding non-linear terms via powers of predictor variables, e.g.

- "Cubic" regression:

  $$Y \approx \beta_0 + \beta_{11} X_1 + \beta_{12} X_1^2 + \beta_{13} X_1^3 + \beta_{21} X_2 + \beta_{22} X_2^2 + \beta_{23} X_2^3 + \dots$$

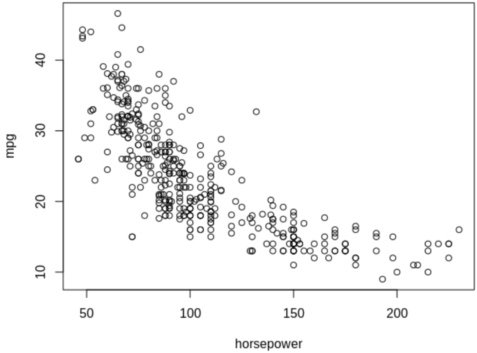
  or

- "Square-root" regression:

  $$Y \approx \beta_0 + \beta_{11} X_1 + \beta_{12} \sqrt{X_1} + \beta_{21} X_2 + \beta_{22} \sqrt{X_2} + \dots$$

Non-linear Relationships: Polynomial Regression.

Example. *Auto* data set has observations of miles per gallon (*mpg*) and *horsepower* for multiple cars.



Q: Does the relationship appear linear? Not quite?
See *R* **code**.

Non-linear Relationships: Polynomial Regression.

Example (cont'd). To address visible non-linearity of $mpg \sim horsepower$ relationship, let's include **quadratic polynomial** term(s):

$$mpg =$$

Note: this is still a **linear** model with respect to **parameters** $\boldsymbol{\beta_0, \beta_1, \beta_2}$.

Just pretend as if:

- $X_1 = horsepower$,
- $X_2 = horsepower^2$

then we get a classic multiple linear regression equation:

$$mpg =$$

Non-linear Relationships: Polynomial Regression.

Example (ct'd). Our polynomial regression model:

$$mpg = \beta_0 + \beta_1 horsepower + \beta_2 horsepower^2 + \epsilon$$

Q: What hypotheses correspond to testing if there's a **quadratic** relationship between *mpg* and *horsepower*?

The results of conducting individual significance tests:

	Coefficient	Std. error	t-statistic	p-value
`Intercept`	56.9001	1.8004	31.6	< 0.0001
`horsepower`	-0.4662	0.0311	-15.0	< 0.0001
`horsepower`^{2}	0.0012	0.0001	10.1	< 0.0001

Q: Is there a quadratic relationship between *mpg* and *horsepower*?

Example (ct'd).

Q: Given that quadratic polynomial regression led to such improvement, why not use

- Cubic regression?

- Fifth degree regression?

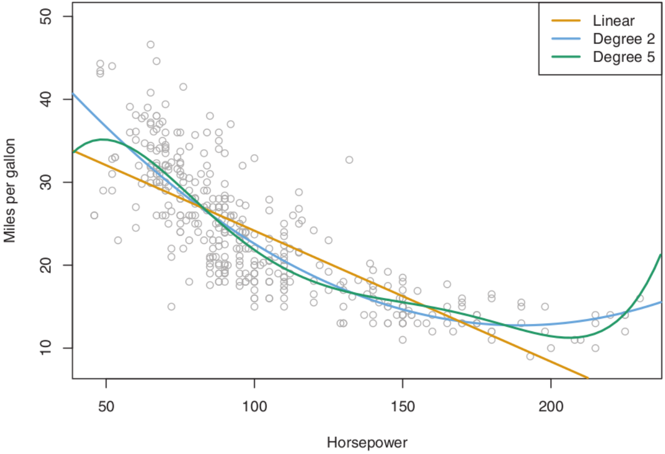
Reason #1: Higher polynomial degree $\Longrightarrow$ tougher interpretation of the relationship between predictor and response.

Task: Attempt a "per 1-unit" interpretation of quadratic or cubic regression, see how that goes.

Polynomial Regression: Why not cubic? $4^{th} - 5^{th}$ degree?

Reason #2: Polynomials of high degrees lead to fits that are:

- overly wiggly,
- indicative of **overfitting** (fitting noise, not true relationship/signal)



See *R* code.

Principle of Marginality.

Principle of Marginality

Model dealing with **high-order terms** (e.g. $X_1 X_2$, or X_1^2) should normally also include the **"lower-order relatives" of that term** (e.g. X_1, X_2).

Examples:

- interaction term $X_2 X_3 \implies$ also include X_2 and X_3
- cubed term $X_1^3 \implies$ also include X_1 and X_1^2
- interaction term $X_1 X_2 X_3 \implies$??

Models that **violate** the principle of marginality, while interpretable, are **not broadly applicable**. Details left out, but if interested - check this out:

`https://stats.idre.ucla.edu/stata/faq/`
`what-happens-if-you-omit-the-main-effect-in-a-regression-mode`