Interaction terms in MLR and Difference in Difference (DID)

Eduard Bukin

MLR with an interaction term

Interaction term

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2) + \beta_4 x_4 + \epsilon$$

Interpretation is only possible through **marginal effects**:

Marginal effect is a partial derivative of the regression equation with respect to a regressor of interest evaluated at certain value of the interaction term.

$$\frac{\partial y}{\partial x_1} = \beta_1 + \beta_3 x_2$$

$$\frac{\partial y}{\partial x_2} = \beta_2 + \beta_3 x_1$$

Marginal Effect (the slope) of x_1 depends on x_2 and vice versa.

Reporting results with marginal effects

1. Report coefficients β_1, β_2, β_3 as is.
2. Report marginal effects at mean values of the interaction term(s):
   - $\text{ME}(\bar{x}_1) = \beta_1 + \beta_3 \bar{x}_2$
   - $\text{ME}(\bar{x}_2) = \beta_2 + \beta_3 \bar{x}_1$
3. To compute standard errors for the marginal effects, we use **delta method**.

Learn more: (Angrist & Pischke, 2009, also Chapter 3)

Recall the hedonic land prices model:

$$\text{acrePrice} = \beta_0 + \beta_1 \text{crpPct} + \beta_2 \text{acres} + \beta_3 \text{improvements} \\ + \beta_4 \text{year} + \beta_5 \text{region} + e$$

- `acrePrice` - sale price in dollars per acre;
- `crpPct` - the percentage of all farm acres enrolled in CRP;
- `acres` - size of the farm in acres;
- `improvements` - share of infrastructure’s value in the land price;
- `development` - dummy variable aliased with `improvements`. Is 1 when improvements > 25% and is 0 otherwise.
- `region` - region in the state Minnesota;
- `year` - year of the land sales translation;

Loading data

```r
library(tidyverse)
library(readxl)
library(modelsummary)
library(ggeffects)
library(marginaleffects)
library(GGally)
options(modelsummary_get = "broom")
## 1. Load the data
dta <- read_excel("land-prices.xlsx") %>%
  mutate(improvements  = as.numeric(improvements),
         development = as.factor(as.integer(improvements > 25)),
         productivity  = as.numeric(productivity),
         tillable  = as.numeric(tillable),
         year = as.factor(year)) %>%
  select(acrePrice, crpPct, acres, improvements, development, region, year)

glimpse(dta)
```

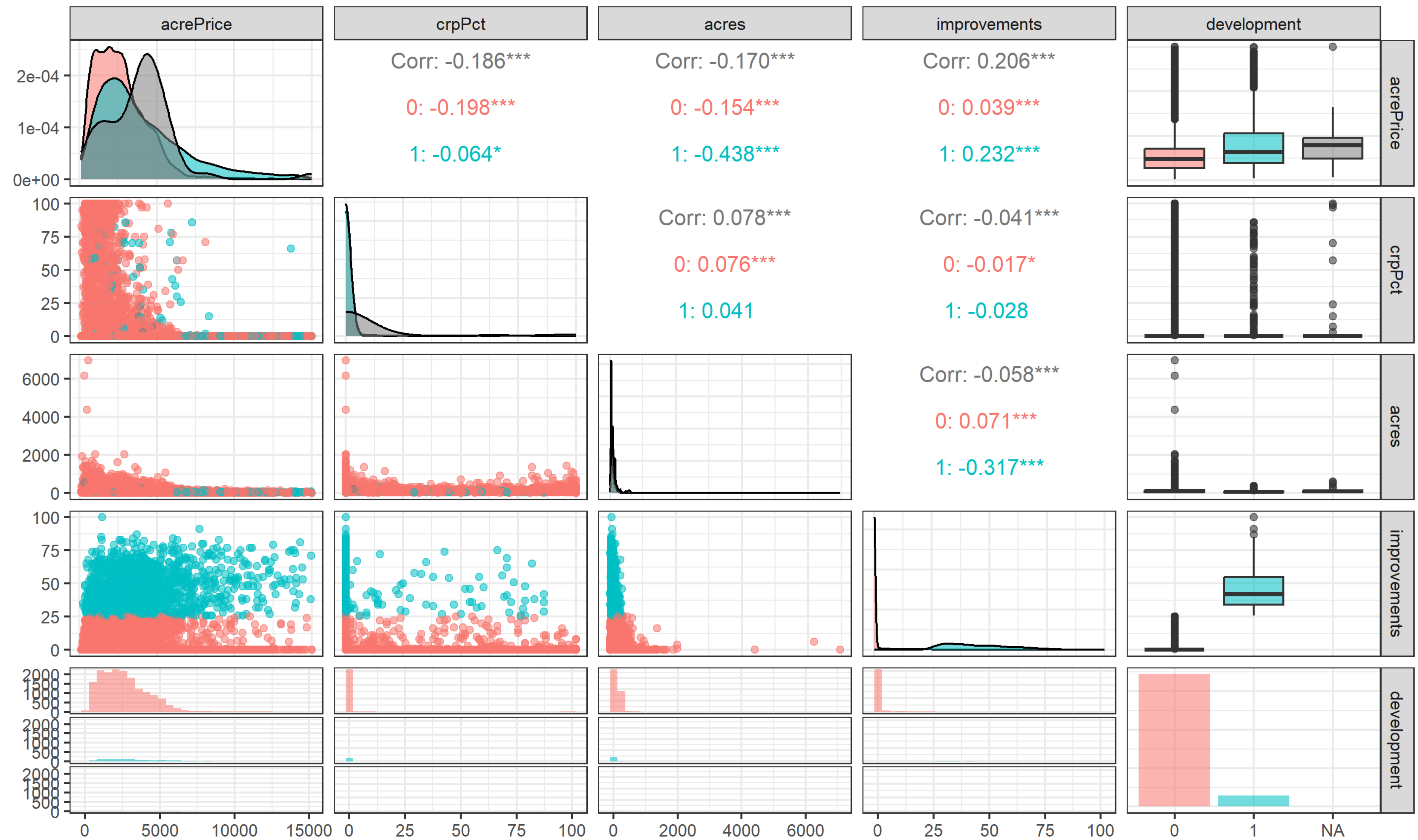
Descriptive statistics

```
datasummary(
    acrePrice + crpPct + acres + improvements + as.integer(development) ~
      (mean + sd + min + median + max) *
      Arguments(na.rm = TRUE),
    data = dta,
    output = "html"
  )
```

	mean	sd	min	median	max
acrePrice	2787.33	1914.04	108.00	2442.00	15000.00
crpPct	4.16	17.17	0.00	0.00	100.00
acres	112.69	128.46	1.00	80.00	6970.00
improvements	4.49	12.94	0.00	0.00	100.00
as.integer(development)	1.08	0.26	1.00	1.00	2.00

Descriptive Plots

```
dta %>%
  select(acrePrice, crpPct, acres, improvements, development) %>%
  ggpairs(aes(colour = development, alpha = 0.2))
```



Example 1. Basic regressions

```
fit1 <-
  lm(acrePrice  ~ acres + development + crpPct + region + factor(year),
     data = dta)

summ <-
  function(x, coef_omit = "reg|year", output = "html", notes = NULL, ...) {
    modelsummary(
      x,
      estimate = "{estimate}{stars} ({std.error})",
      statistic = NULL,
      output = output,
      gof_omit =  c("AIC|BIC|Log|F|RMS"),
      coef_omit = coef_omit,
      notes = notes,
      ...
    )
  }

quat_summ <- function(x, coef_omit = "reg|year", output = "markdown", ...) {
```

Interpret the baseline results

```
cust_summ(list(fit1))
```

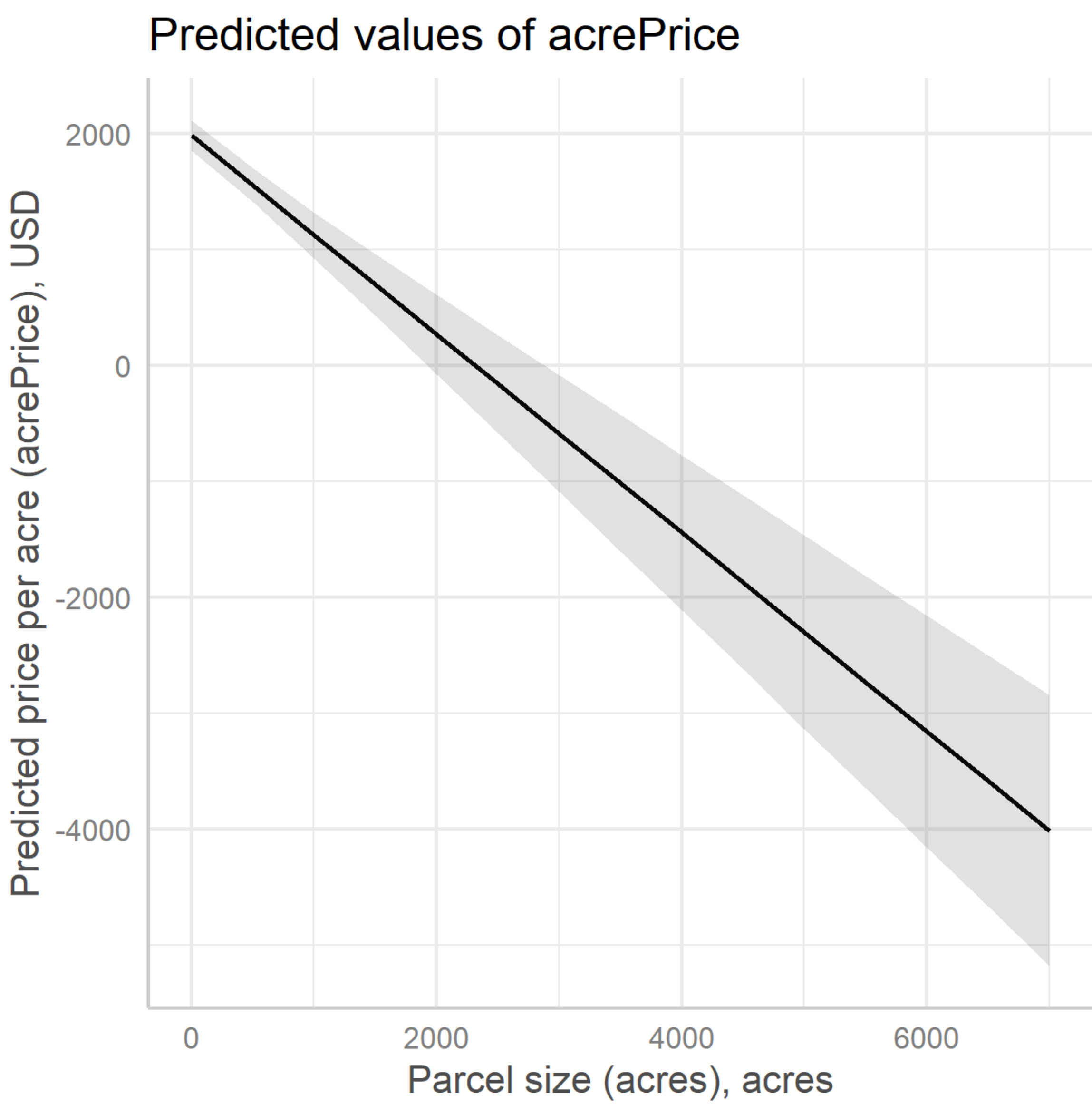
	Model 1
(Intercept)	2019.877*** (65.416)
acres	-0.857*** (0.086)
development1	1555.287*** (41.467)
crpPct	-8.892*** (0.645)
Num.Obs.	18650
R2	0.413
R2 Adj.	0.412

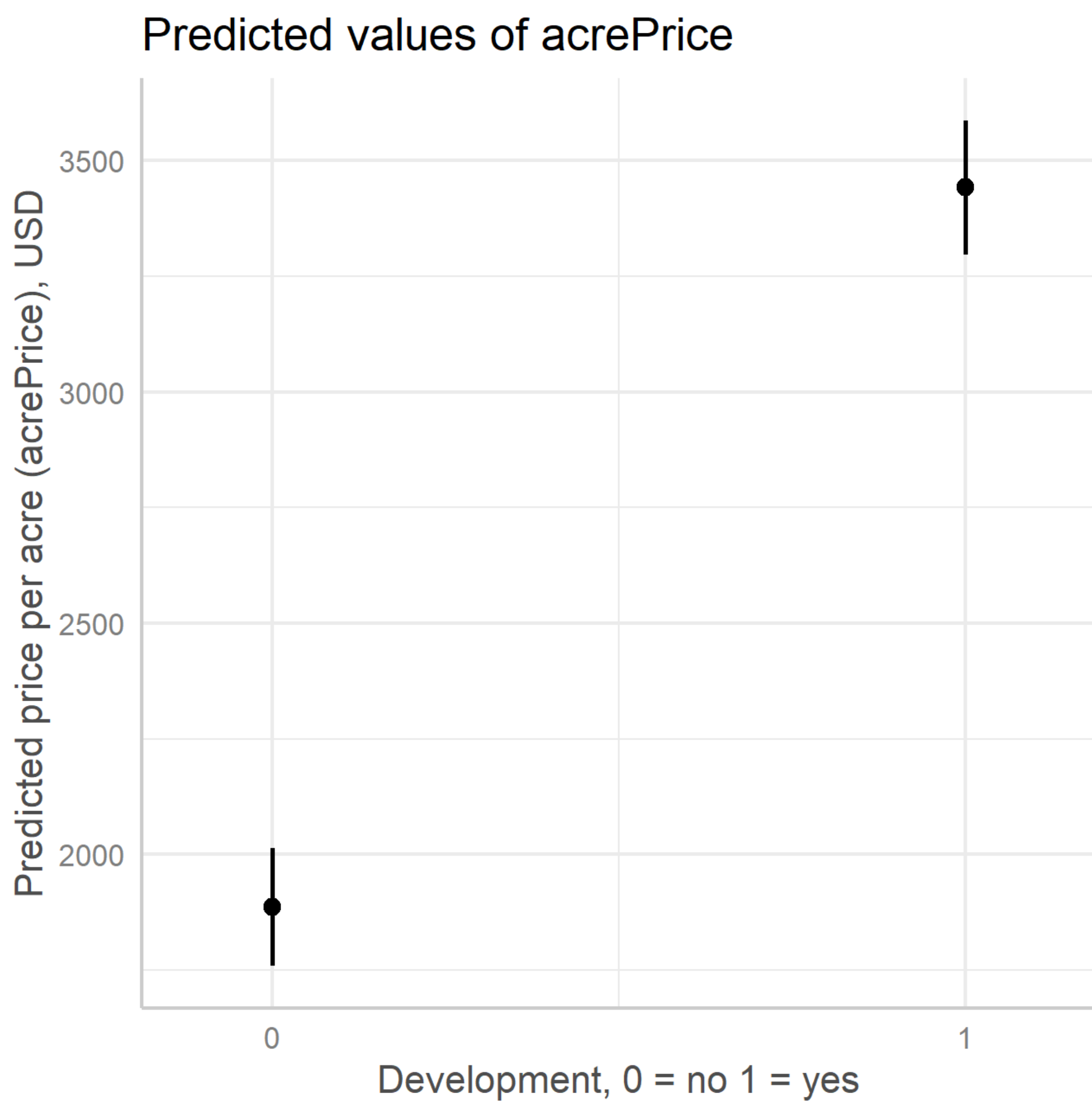
Note: ^^ Model 1: acrePrice ~ acres + development + crpPct + region + factor(year)

Fitted vs area and development

```r
library(ggeffects)
ggpredict(fit1, terms = "acres") %>%
  plot() +
  ylab("Predicted price per acre (acrePrice
  xlab("Parcel size (acres), acres")
```

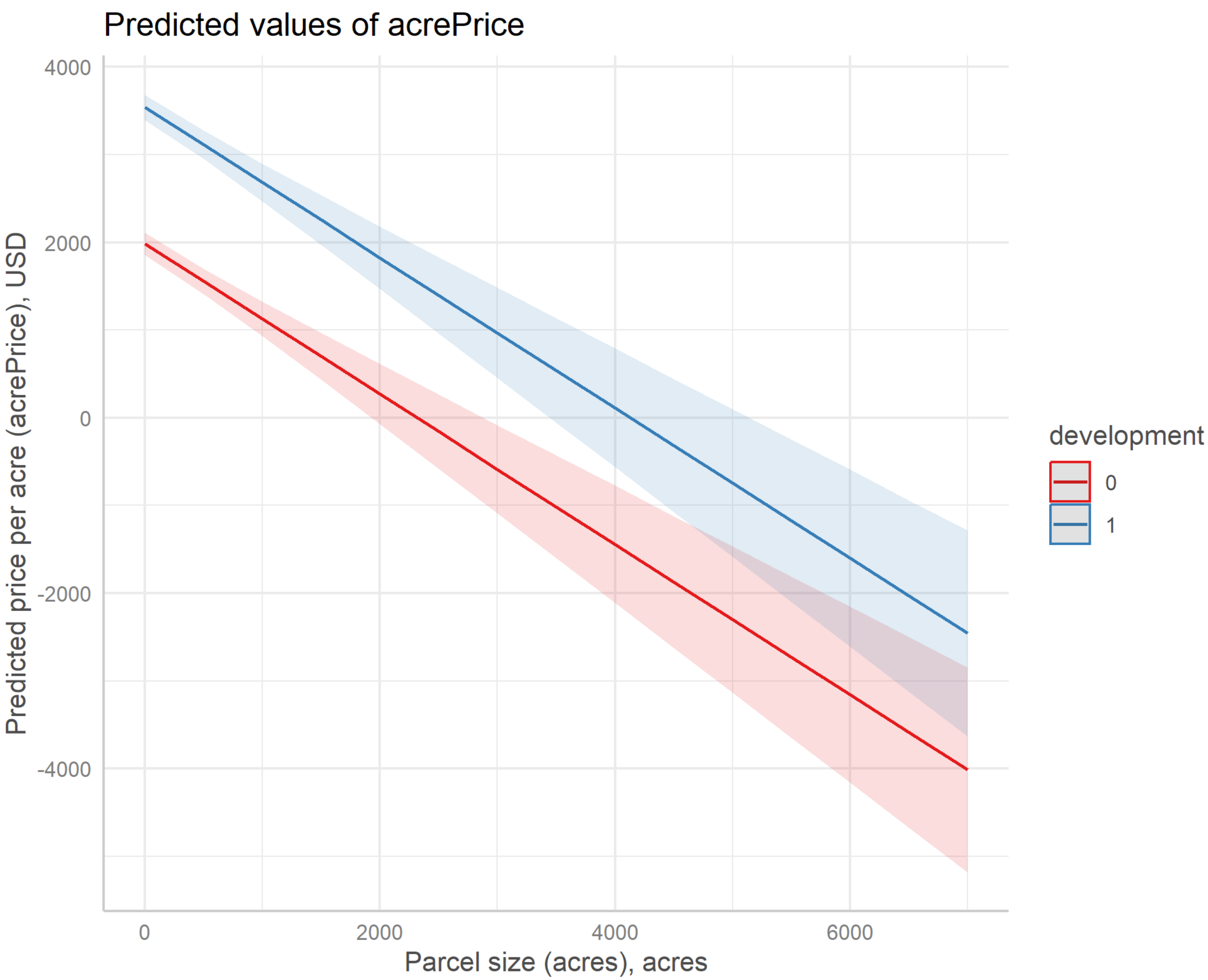
```r
ggpredict(fit1, terms = "development") %>%
  plot() +
  ylab("Predicted price per acre (acrePrice
  xlab("Development, 0 = no 1 = yes")
```





Fitted vs area and development (2)

```
ggpredict(fit1, terms = c("acres", "development")) %>%
  plot() +
  ylab("Predicted price per acre (acrePrice), USD") +
  xlab("Parcel size (acres), acres")
```



- Without an interaction term, dummy variables simply cause shifts in the outcome variable.

Example 2. Interaction term with a binary variable

```
fit2 <-
  lm(acrePrice  ~ acres * development + crpPct + region + factor(year),
     data = dta)
cust_summ(list(fit1, fit2))
```

	Model 1	Model 2
(Intercept)	2019.877*** (65.416)	1997.055*** (64.413)
acres	-0.857*** (0.086)	-0.652*** (0.085)
development1	1555.287*** (41.467)	2978.491*** (71.424)
crpPct	-8.892*** (0.645)	-9.073*** (0.635)
acres × development1		-18.053*** (0.743)
Num.Obs.	18650	18650
R2	0.413	0.431
R2 Adj.	0.412	0.430

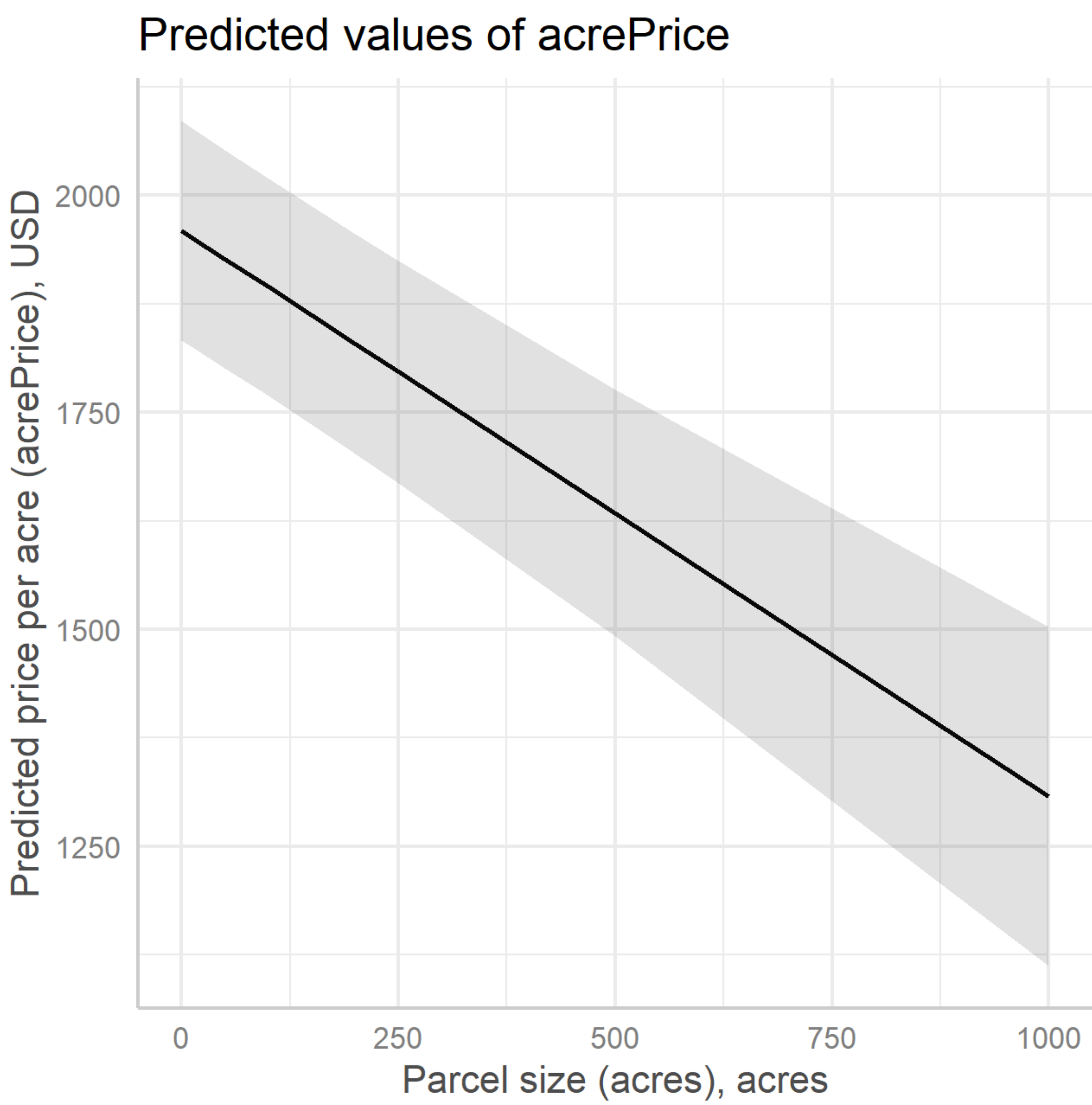
Note: ^^ Model 1: acrePrice ~ acres + development + crpPct + region + factor(year)
Model 2: acrePrice ~ acres * development + crpPct + region + factor(year)

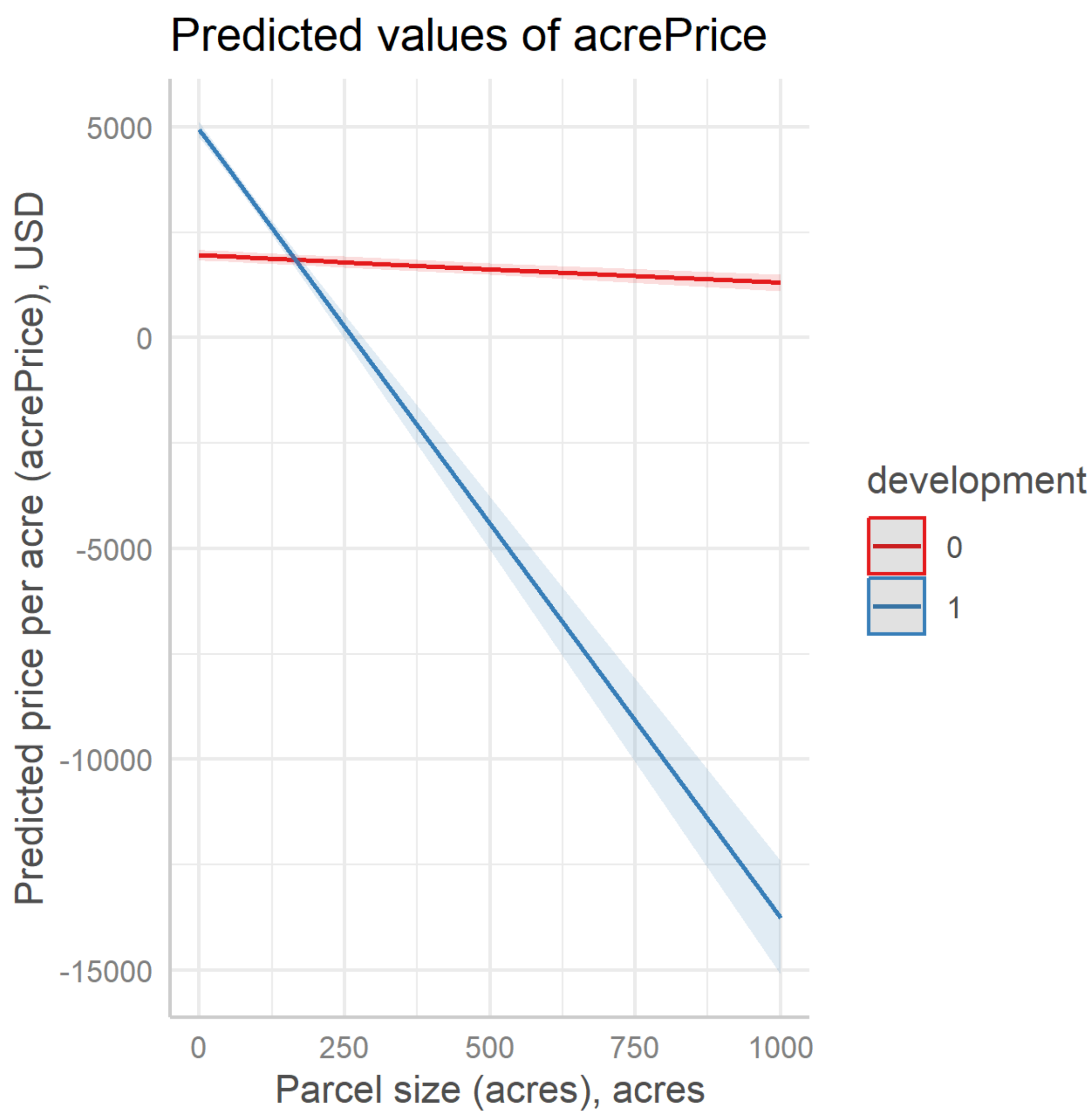
Fitted vs development with an interaction term

Fitted vs area with an interaction term

```
ggpredict(fit2, terms = c("acres [0, 10, 5(
  plot() +
  ylab("Predicted price per acre (acrePrice
  xlab("Parcel size (acres), acres")
```

```
ggpredict(fit2, terms = c("acres [0, 10, 5(
  plot() +
  ylab("Predicted price per acre (acrePrice
  xlab("Parcel size (acres), acres")
```





Marginal Effect `acres` and `development` at means

```
library(marginaleffects)
modelsummary(fit2)
modelsummary(marginaleffects(fit2))
```

	Coef. as is	M.E. at means
acres	−0.652*** (0.085)	−2.017*** (0.097)
crpPct	−9.073*** (0.635)	−9.073*** (0.635)
development1	2978.491*** (71.424)	944.727*** (47.947)
acres * development1	−18.053*** (0.743)	
Num.Obs.	18650	18650
R2	0.431	0.431
R2 Adj.	0.430	0.430

Marginal Effect `acres` and `development` (visually)

Example 3. Interaction with a continuous variable

```
fit3 <-
  lm(acrePrice  ~ acres * improvements + crpPct + region + factor(year),
     data = dta)
cust_summ(list(fit1, fit2, fit3), output = "markdown")
```

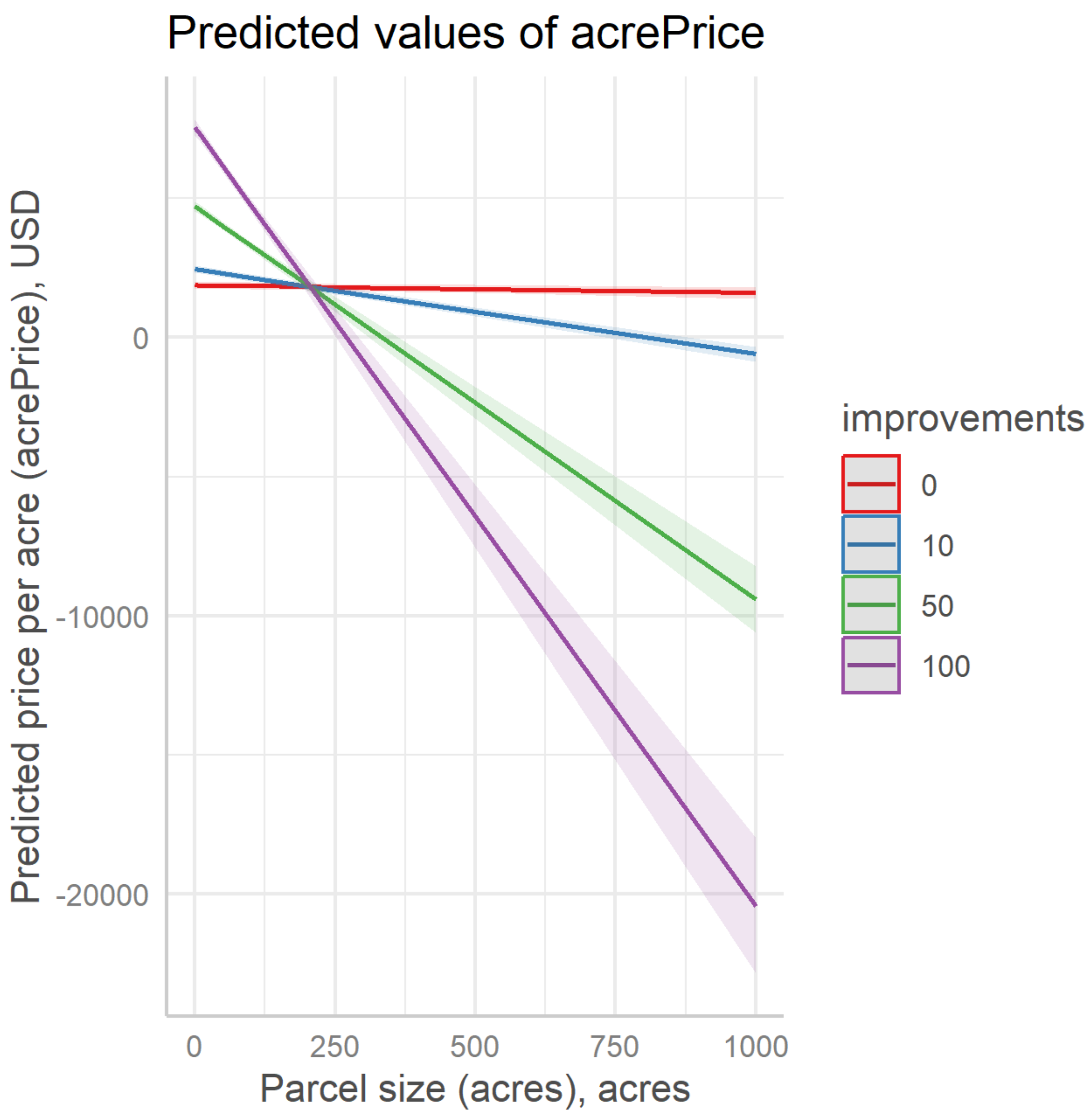
	Model 1	Model 2	Model 3
(Intercept)	2019.877*** (65.416)	1997.055*** (64.413)	1923.131*** (64.085)
acres	-0.857*** (0.086)	-0.652*** (0.085)	-0.276** (0.089)
development1	1555.287*** (41.467)	2978.491*** (71.424)	
crpPct	-8.892*** (0.645)	-9.073*** (0.635)	-8.980*** (0.631)
acres × development1		-18.053*** (0.743)	
improvements			56.695*** (1.307)
acres × improvements			-0.277*** (0.014)
Num.Obs.	18650	18650	18650
R2	0.413	0.431	0.438
R2 Adj.	0.412	0.430	0.438

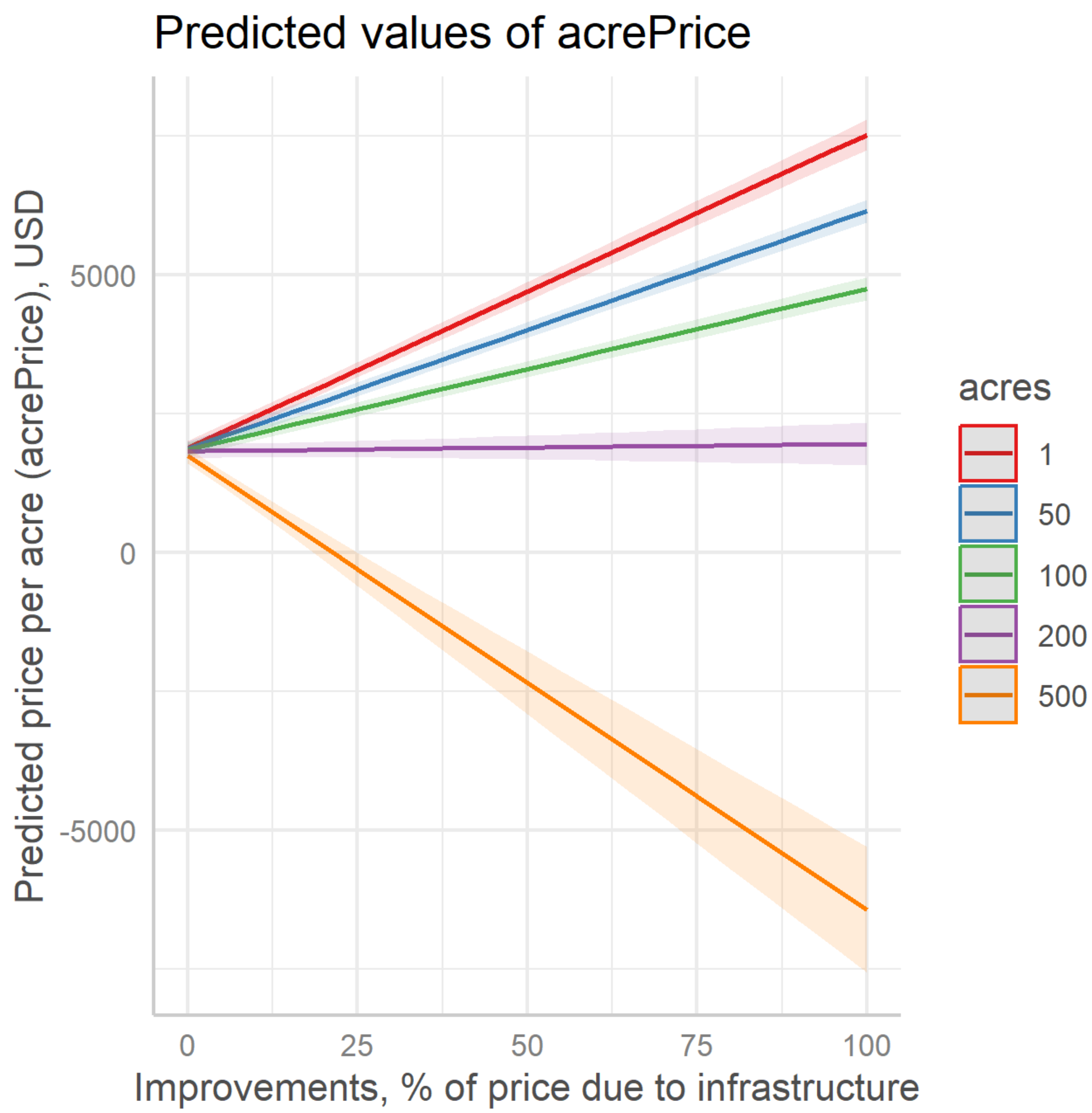
Note: ^^ Model 1: acrePrice ~ acres + development + crpPct + region + factor(year)
Model 2: acrePrice ~ acres * development + crpPct + region + factor(year)
Model 3: acrePrice ~ acres * improvements + crpPct + region + factor(year)

Fitted vs `area * improvements`

```r
ggpredict(fit3, terms = c("acres [0, 10, 50
  plot() +
  ylab("Predicted price per acre (acrePrice
  xlab("Parcel size (acres), acres")
```

```r
ggpredict(fit3, terms = c("improvements", "
  plot() +
  ylab("Predicted price per acre (acrePrice
  xlab("Improvements, % of price due to in
```





Marginal Effect `acres * improvements` (visually)

Reporting all regressions results

	Model 1 (no interaction)	Model 2 (Area*Development)	Model 3 (Area*Improvements)
Area, acres (*)	−0.857* (0.350)	−2.017*** (0.097)	−1.520*** (0.089)
crpPct	−8.892*** (1.998)	−9.073*** (0.635)	−8.980*** (0.631)
Development status, dummy (*)		944.727*** (47.947)	
Share infrastructure in land price, % (*)			25.490*** (0.984)
Num.Obs.	18650	18650	18650
R2	0.413	0.431	0.438
R2 Adj.	0.412	0.430	0.438

(*) marginal effects of the coefficients are reported at means of the corresponding interaction terms. Robust standard errors clustered at region are reported in brackets. For marginal effects, standard errors are estimated using delta method.

Takeaway and homework

Takeaway and homework

- Marginal effects.
- Difference between fitted values and marginal effects.
- Homework:
  - Reproduce code from the slides.
  - Follow pre-recorded materials with extra calculations.

References

References

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics*. Princeton University Press. http://doi.org/10.1515/9781400829828