

When is a Linear Model “as good” as Logistic Regression?

Very small note

Paul Johnson¹

¹CRMDA, University of Kansas

2018



Outline

- 1 Similarity of linear and logistic fits
- 2 Repeat with more dramatic slope
- 3 And the Answer Is
- 4 What about a dichotomy?

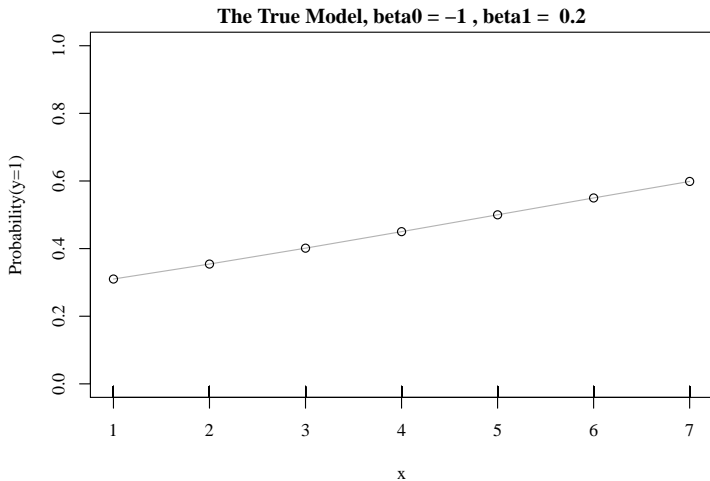
Outline

- 1 Similarity of linear and logistic fits
- 2 Repeat with more dramatic slope
- 3 And the Answer Is
- 4 What about a dichotomy?

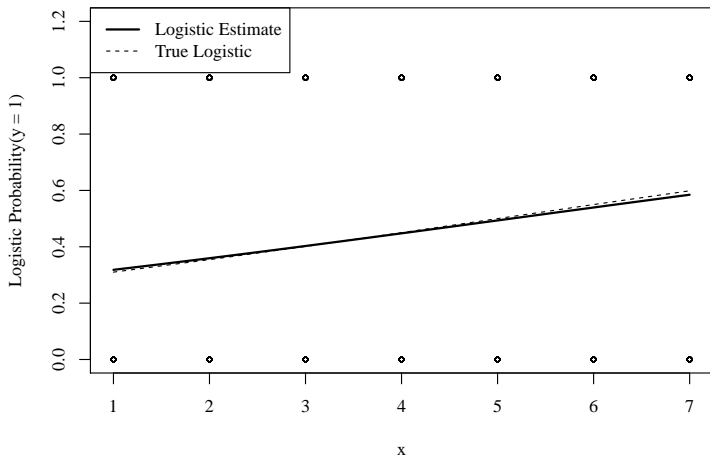
The "true" model

- This uses R (2018) to do a simple illustration, adapting a student's example from POLS707
- The predictor is coded 1 through 7 (political ideology)

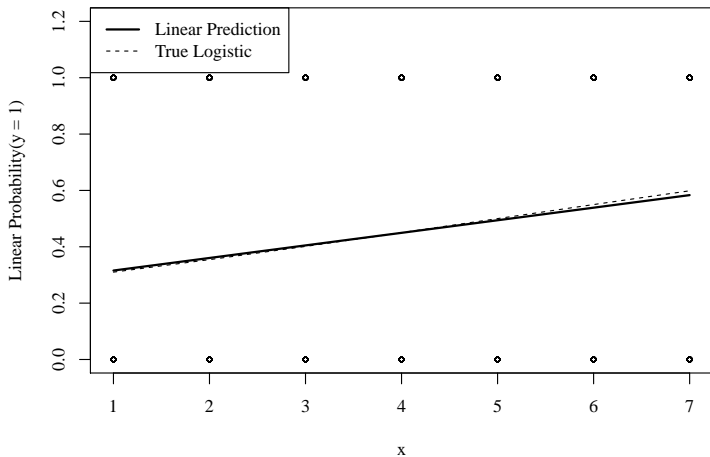
The "true" model ...



Logistic Predicted Probabilities



Linear Probability Model



Predicted values are very very close

```
## The logistic model
predictOMatic(m1, predVals = list(x=1:7), type =
  "response")
```

```

x      fit
1 1 0.3182457
2 2 0.3594214
3 3 0.4027772
4 4 0.4477077
5 5 0.4935086
6 6 0.5394186
7 7 0.5846695
```

```
## The linear model
predictOMatic(m2, predVals = list(x=1:7), type =
  "response")
```


Predicted values are very very close ...

5

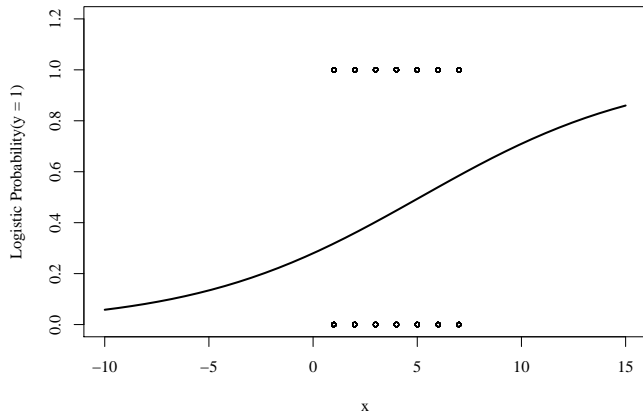
	x	fit
1	1	0.3156633
2	2	0.3602750
3	3	0.4048866
4	4	0.4494983
5	5	0.4941099
6	6	0.5387215
7	7	0.5833332

How to make sense out of that?

One guess: imagine predictor had wider range

```
plotCurves(m1, plotx = "x", ylab = "Logistic  
Probability(y = 1)", plotLegend = FALSE,  
plotxRange = c(-10, 15))
```

How to make sense out of that? ...

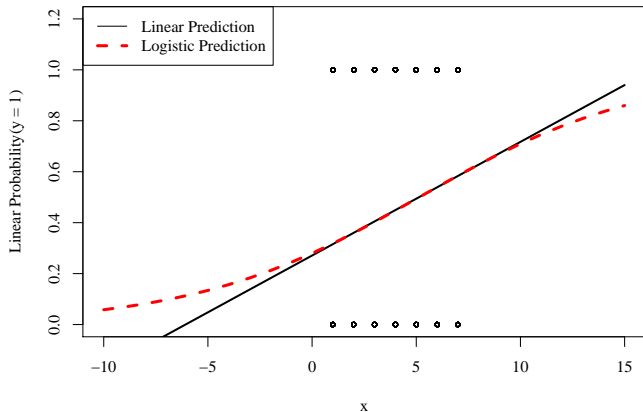


Linear model eventually goes "out of bounds"

```
plotCurves(m2, plotx = "x", ylab = "Linear
  Probability(y = 1)", plotLegend = FALSE,
  plotxRange = c(-10, 15))
```

```
plotCurves(m2, plotx = "x", ylab = "Linear
  Probability(y = 1)", plotLegend = FALSE,
  plotxRange = c(-10, 15))
m1.pom <- predictOMatic(m1, predVals = list(x =
  seq(-10, 15)))
lines(fit ~ x, m1.pom, col = "red", lty = 2, lwd
  = 3)
legend("topleft", legend = c("Linear Prediction",
  "Logistic Prediction"), lty = c(1,
  2), col=c("black", "red"), lwd = c(1,3))
```

Linear model eventually goes "out of bounds" ...

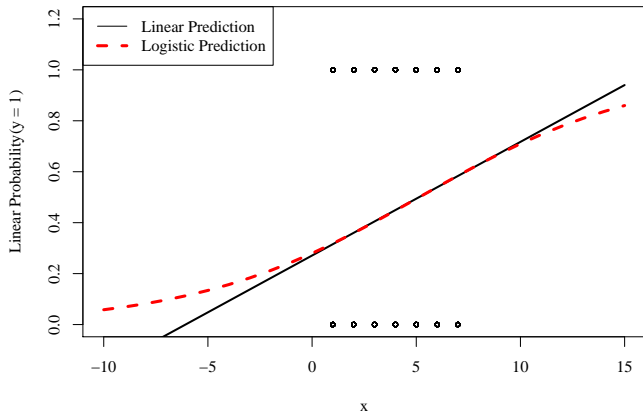


Linear and Logistic differences appear on wider range

```
plotCurves(m2, plotx = "x", ylab = "Linear
  Probability(y = 1)", plotLegend = FALSE,
  plotxRange = c(-10, 15))
```

```
plotCurves(m2, plotx = "x", ylab = "Linear
  Probability(y = 1)", plotLegend = FALSE,
  plotxRange = c(-10, 15))
m1.pom <- predictOMatic(m1, predVals = list(x =
  seq(-10, 15)))
lines(fit ~ x, m1.pom, col = "red", lty = 2, lwd
  = 3)
legend("topleft", legend = c("Linear Prediction",
  "Logistic Prediction"), lty = c(1,
  2), col=c("black", "red"), lwd = c(1,3))
```

Linear and Logistic differences appear on wider range ...



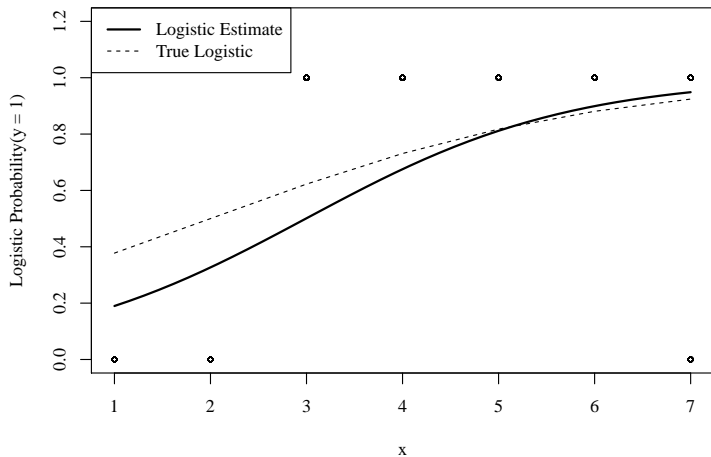
Outline

- 1 Similarity of linear and logistic fits
- 2 Repeat with more dramatic slope
- 3 And the Answer Is
- 4 What about a dichotomy?

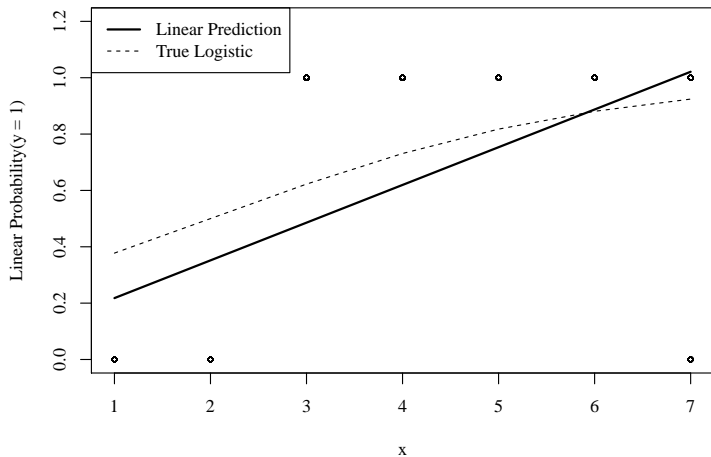
The "true" model with higher beta coefficient



Logistic Predicted Probabilities



Linear Probability Model



Predicted values are not so very very close

```
## The logistic model
predictOMatic(m1, predVals = list(x=1:7), type =
  "response")
```

```

x      fit
1 1 0.1898391
2 2 0.3266963
3 3 0.5011813
4 4 0.6753790
5 5 0.8116101
6 6 0.8992027
7 7 0.9486457
```

```
## The linear model
predictOMatic(m2, predVals = list(x=1:7), type =
  "response")
```

Predicted values are not so very very close ...

5

	x	fit
1	1	0.2178671
2	2	0.3517437
3	3	0.4856204
4	4	0.6194971
5	5	0.7533738
6	6	0.8872505
7	7	1.0211271

Outline

- 1 Similarity of linear and logistic fits
- 2 Repeat with more dramatic slope
- 3 And the Answer Is
- 4 What about a dichotomy?

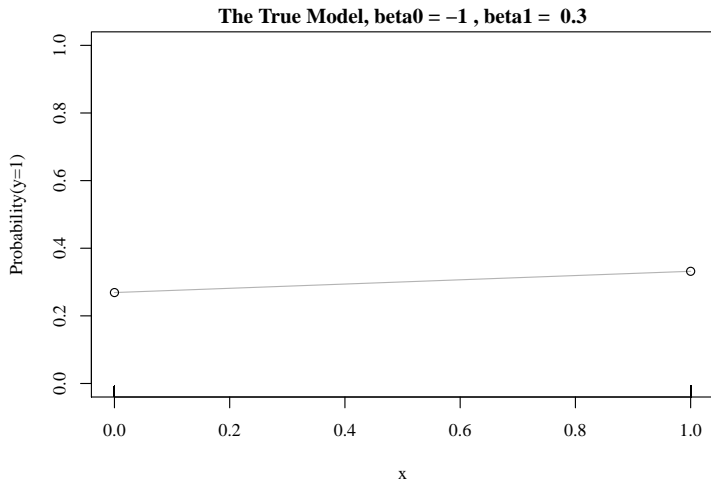
When does the OLS model do "as well"?

- If the "range" of the predictor is narrow, then there is not much curvature visible within the data range.
- If the slope coefficient is small, then the S shaped curve is very elongated.

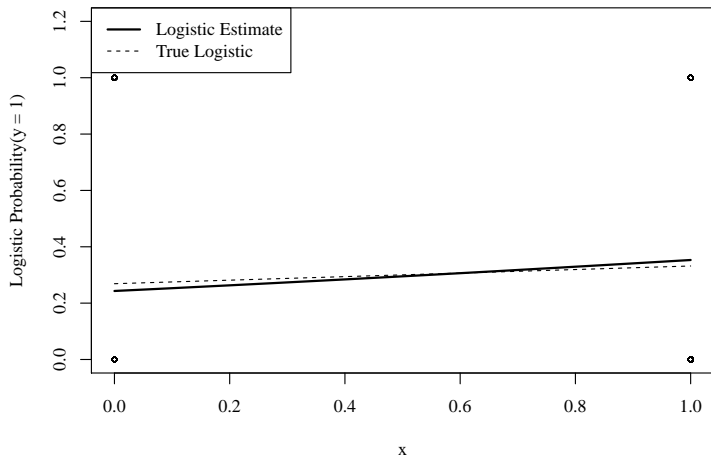
Outline

- 1 Similarity of linear and logistic fits
- 2 Repeat with more dramatic slope
- 3 And the Answer Is
- 4 What about a dichotomy?

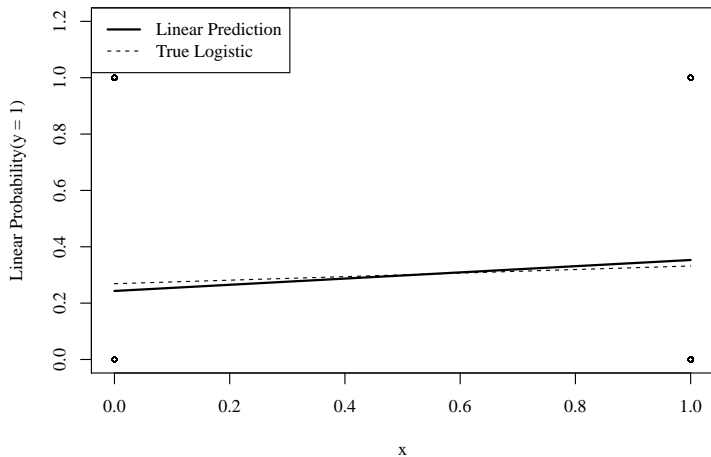
Predicted values are not so very very close



Logistic Predicted Probabilities



Linear Probability Model



Predicted values same

```
## The logistic model
predictOMatic(m1, predVals = list(x=1:7), type =
  "response")
```

```

x      fit
1 1 0.3530612
2 2 0.4810929
3 3 0.6116566
4 4 0.7279461
5 5 0.8196787
6 6 0.8853519
7 7 0.9291736
```

```
## The linear model
predictOMatic(m2, predVals = list(x=1:7), type =
  "response")
```

Predicted values same ...

5

	x	fit
1	1	0.3530612
2	2	0.4629852
3	3	0.5729092
4	4	0.6828331
5	5	0.7927571
6	6	0.9026811
7	7	1.0126050

References

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing.

Session

```
sessionInfo()
```

```
R version 3.5.1 (2018-07-02)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 18.04.1 LTS

Matrix products: default
BLAS: /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.7.1
LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.7.1

locale:
 [1] LC_CTYPE=en_US.UTF-8          LC_NUMERIC=C
      LC_TIME=en_US.UTF-8
 [4] LC_COLLATE=en_US.UTF-8      LC_MONETARY=en_US.UTF-8
      LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8        LC_NAME=C              LC_ADDRESS=C
[10] LC_TELEPHONE=C              LC_MEASUREMENT=en_US.UTF-8
      LC_IDENTIFICATION=C

attached base packages:
[1] stats      graphics  grDevices  utils      datasets  methods    base

other attached packages:
[1] rockchalk_1.8.119  stationery_0.98.5.5
```

Session ...

```

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.17      knitr_1.20        magrittr_1.5      splines_3.5.1
      kutils_1.49     MASS_7.3-50
 [7] mnormt_1.5-5     lattice_0.20-35  pbivnorm_0.6.0   xtable_1.8-2
      minqa_1.2.4     carData_3.0-1
[13] stringr_1.3.1    plyr_1.8.4        tools_3.5.1      grid_3.5.1
      nlme_3.1-137    htmltools_0.3.6
[19] lme4_1.1-17     digest_0.6.15    rprojroot_1.3-2  lavaan_0.6-1
      Matrix_1.2-14  zip_1.0.0
[25] nloptr_1.0.4     evaluate_0.10.1  rmarkdown_1.10   openxlsx_4.1.0
      stringi_1.2.3  compiler_3.5.1
[31] backports_1.1.2  stats4_3.5.1     foreign_0.8-70

```