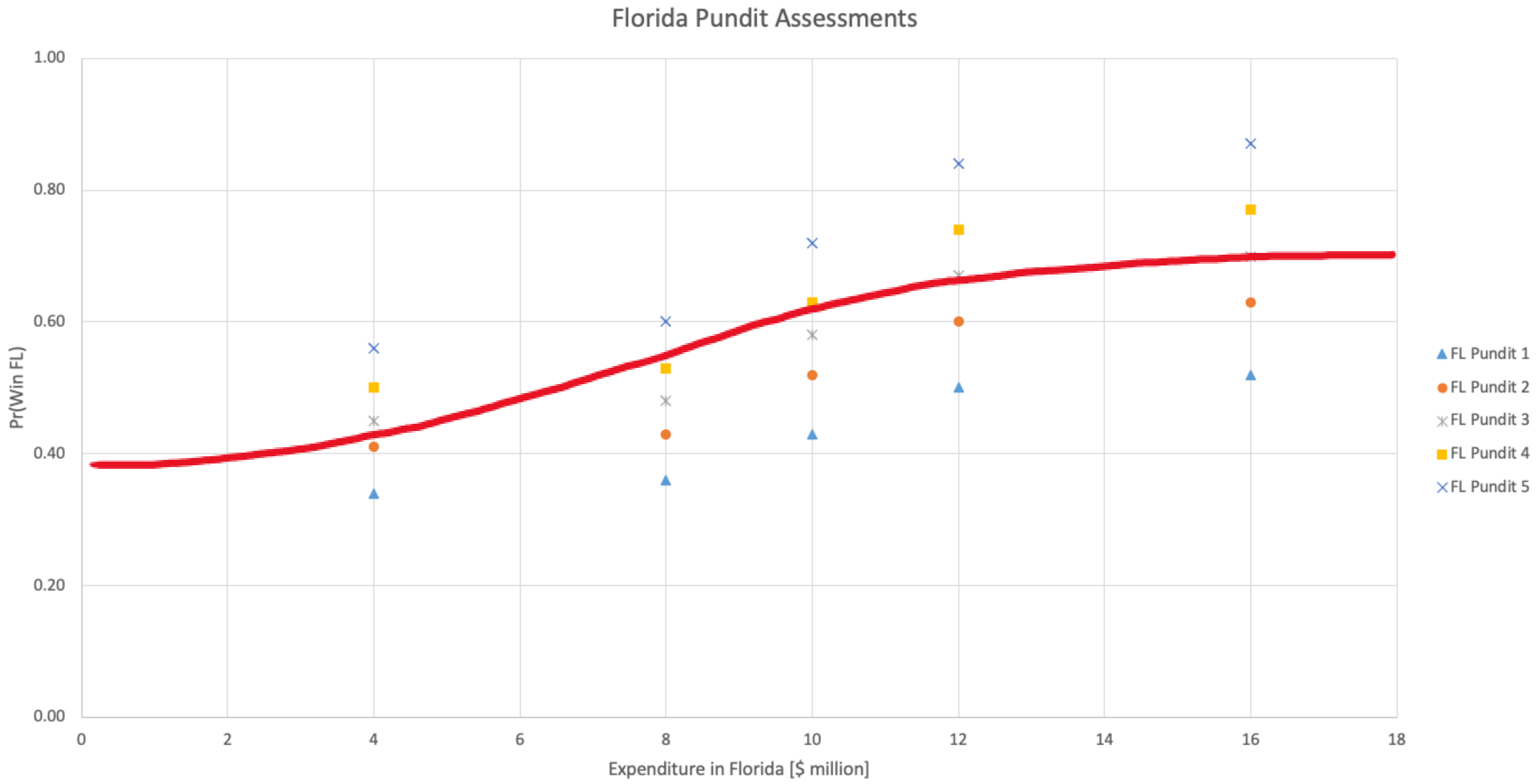
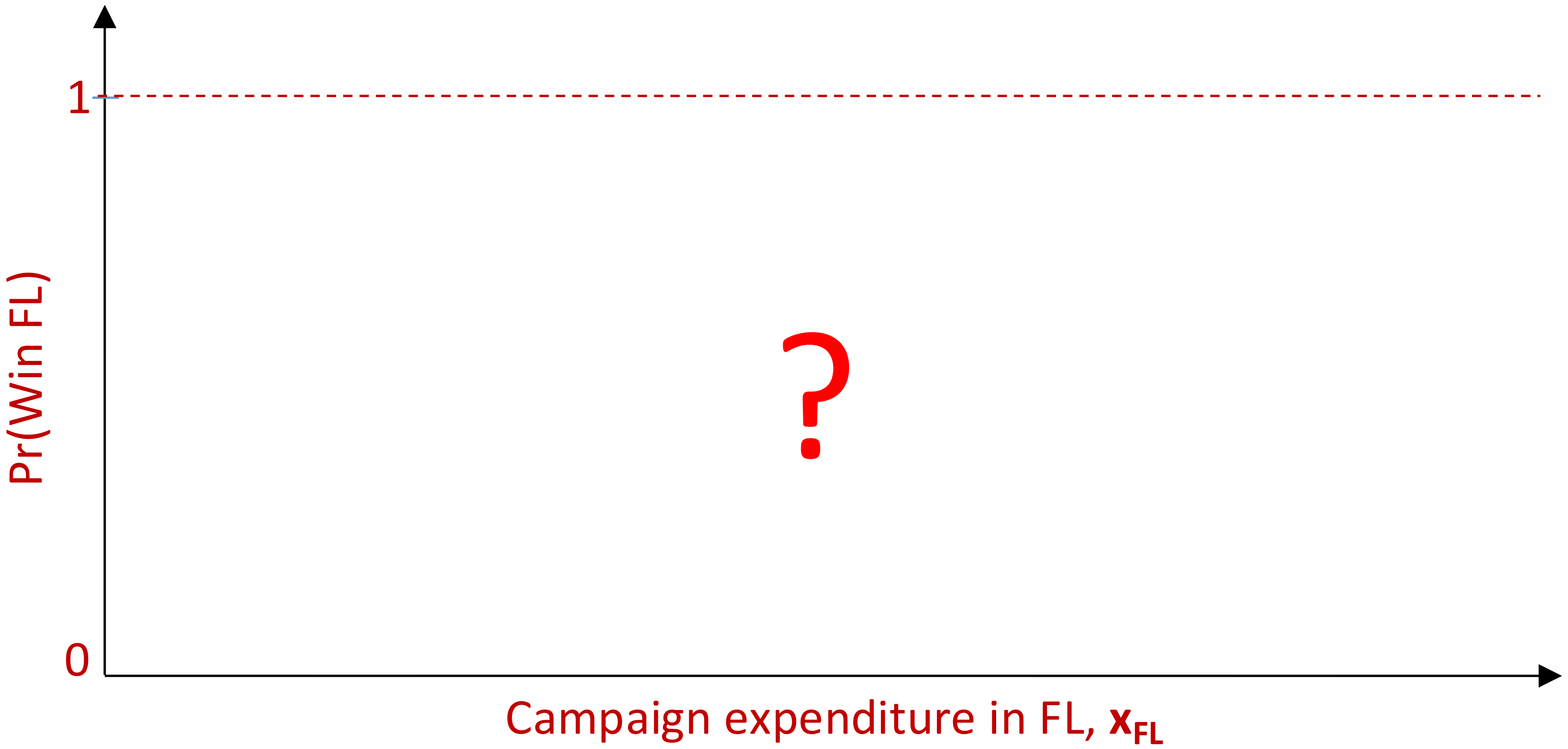
Maybe we can “fit” a curve to this data?

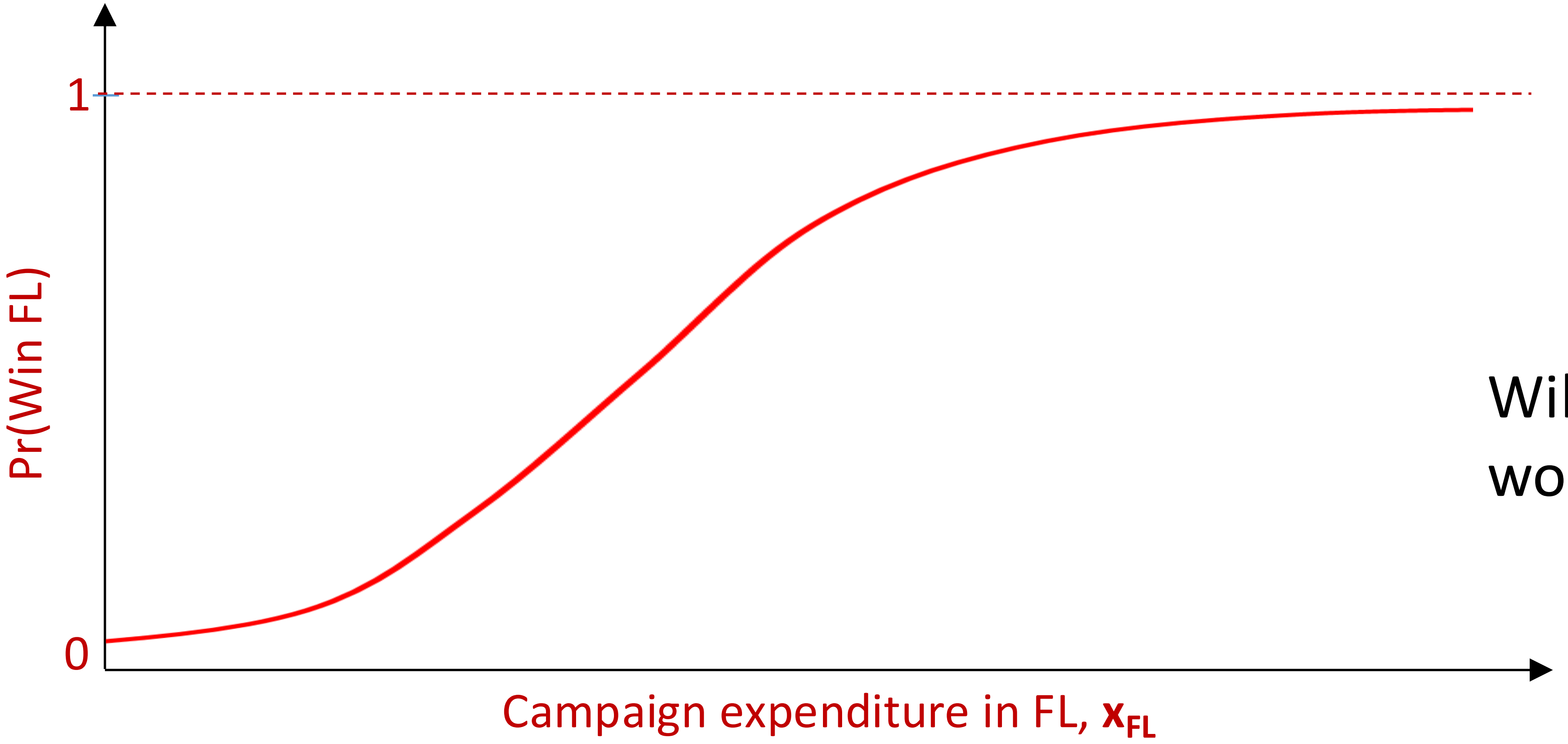


State Level Model



Propose a model
that relates campaign expenditure in FL
to the probability of winning FL

State Level Model

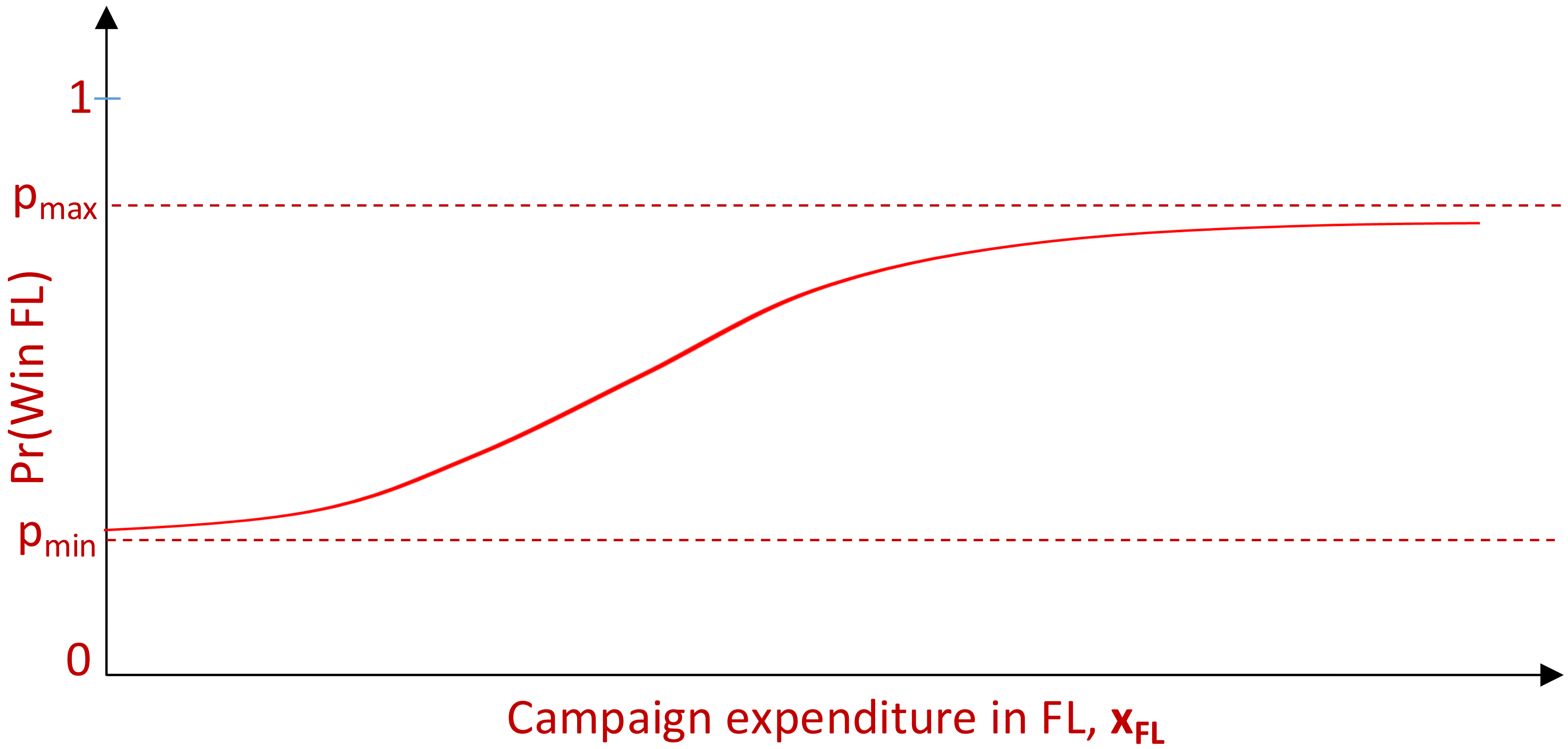


Will this work?

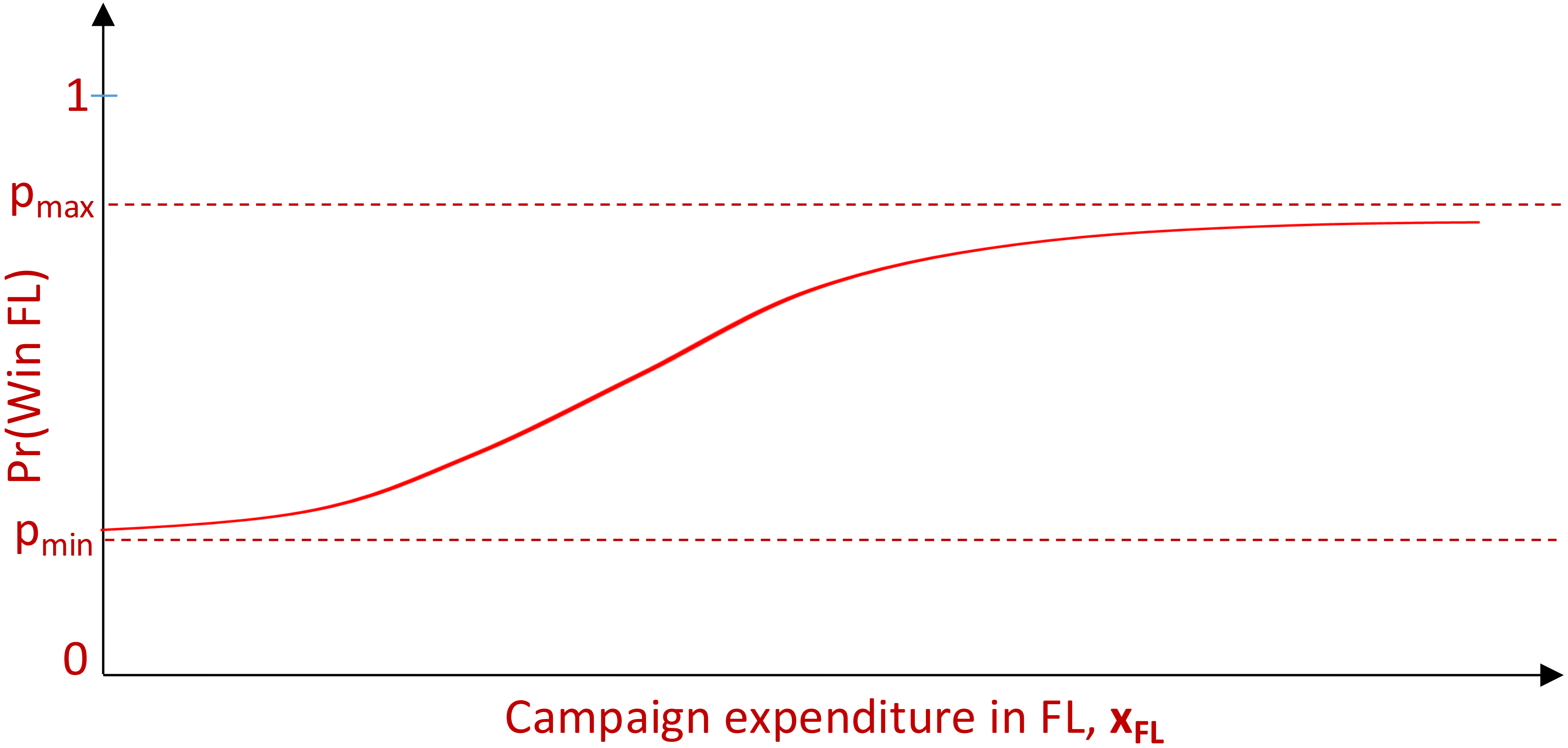
Recall:
Logistic curve

$$\boxed{Pr(Win\ FL) = \frac{1}{1 + e^{-(b_0 + b_1 x_{FL})}}}$$

A Better State Level Model?



A Better State Level Model



A better model:
Scaled logistic curve

$$\boxed{Pr(Win\ FL) = p_{min} + \frac{(p_{max} - p_{min})}{1 + e^{-(b_0 + b_1 x)}}}$$

A Better State Level Model

Florida	Probability of Winning Florida for Expenditure Level				
	$4 million	$8 million	$10 million	$12 million	$16 million
FL Pundit 1	0.34	0.36	0.43	0.50	0.52
FL Pundit 2	0.41	0.43	0.52	0.60	0.63
FL Pundit 3	0.45	0.48	0.58	0.67	0.70
FL Pundit 4	0.50	0.53	0.63	0.74	0.77
FL Pundit 5	0.56	0.60	0.72	0.84	0.87

How can we estimate the parameters p_{min}, p_{max}, b_0, and b_1 from the above data?

$$\boxed{Pr(Win\ FL) = p_{min} + \frac{(p_{max} - p_{min})}{1 + e^{-(b_0 + b_1 x)}}}$$

A Better State Level Model

Florida	Probability of Winning Florida for Expenditure Level				
	$4 million	$8 million	$10 million	$12 million	$16 million
FL Pundit 1	0.34	0.36	0.43	0.50	0.52
FL Pundit 2	0.41	0.43	0.52	0.60	0.63
FL Pundit 3	0.45	0.48	0.58	0.67	0.70
FL Pundit 4	0.50	0.53	0.63	0.74	0.77
FL Pundit 5	0.56	0.60	0.72	0.84	0.87

How can we estimate the parameters p_{min}, p_{max}, b_0, and b_1 from the above data?

$$\boxed{Pr(Win\ FL) = p_{min} + \frac{(p_{max} - p_{min})}{1 + e^{-(b_0 + b_1 x)}}}$$

We can use <u>optimization to find the best-fit curve</u> for the pundit data (like how we found the best-fitting line in Linear Regression)

Pundit Estimation

For any choice of p_{max}, p_{min}, b_0 and b_1, we can calculate the sum of squared *errors* of that pundit model. We'd like to choose p_{max}, p_{min}, b_0 and b_1 to minimize this error.

$$\min_{p_{min},\, p_{max},\, b_0,\, b_1} \mathit{Total\ Prediction\ Error}$$

$$\Downarrow$$

$$\min_{p_{min},\, p_{max},\, b_0,\, b_1} \mathit{Sum\ of}\ (\mathit{pundit}\ - \mathit{model})^2$$

$$\Downarrow$$

$$\min_{p_{min},\, p_{max},\, b_0,\, b_1} \sum_{i=1}^{n} \left(y_i - \left(p_{min} + \frac{(p_{max} - p_{min})}{1 + e^{-(b_0 + b_1 x_i)}} \right) \right)^2$$

Pundit Estimation

For any choice of p_{max}, p_{min}, b_0 and b_1, we can calculate the sum of squared *errors* of that pundit model. We'd like to choose p_{max}, p_{min}, b_0 and b_1 to minimize this error.

$$\min_{p_{min}, p_{max}, b_0, b_1} Total\ Prediction\ Error$$

$$\Downarrow$$

$$\min_{p_{min}, p_{max}, b_0, b_1} Sum\ of\ (pundit\ - model)^2$$

$$\Downarrow$$

$$\min_{p_{min}, p_{max}, b_0, b_1} \sum_{i=1}^{n} \left(y_i - \left(p_{min} + \frac{(p_{max} - p_{min})}{1 + e^{-(b_0 + b_1 x_i)}} \right) \right)^2$$

Aside: Regression is a non-linear optimization problem! Recall that we find the parameters that minimize the sum of squared errors.

Estimating parameters in Excel

CURVE FITTING USING NONLINEAR OPTIMIZATION

"DECISIONS"

p_min	0.45
p_max	0.70
b_0	-10.16
b_1	1.02

OBJECTIVE

Objective function	0.249	<-

$$p_{min} + \frac{(p_{max} - p_{min})}{1 + e^{-(b_0 + b_1 x)}}$$

CONSTRAINTS *None*

CALCULATIONS

Observation	Data		Model	Residual	Square of
i	x_i	y_i (= FL pundits)	Prediction		Residual
1	4.00	0.34	0.45	0.11	0.01
2	4.00	0.41	0.45	0.04	0.00
3	4.00	0.45	0.45	0.00	0.00
4	4.00	0.50	0.45	-0.05	0.00
5	4.00	0.56	0.45	-0.11	0.01
6	8.00	0.36	0.48	0.12	0.01
7	8.00	0.43	0.48	0.05	0.00
8	8.00	0.48	0.48	0.00	0.00
9	8.00	0.53	0.48	-0.05	0.00
10	8.00	0.60	0.48	-0.12	0.01

Parameter Estimation Results

- Suppose we do this for each state and obtain the following parameter estimates:

Parameter	State		
	FL	**OH**	**PA**
p_{min}	0.45	0.50	0.40
p_{max}	0.70	0.60	0.80
b_0	-10	-5	-7
b_1	1	1	1

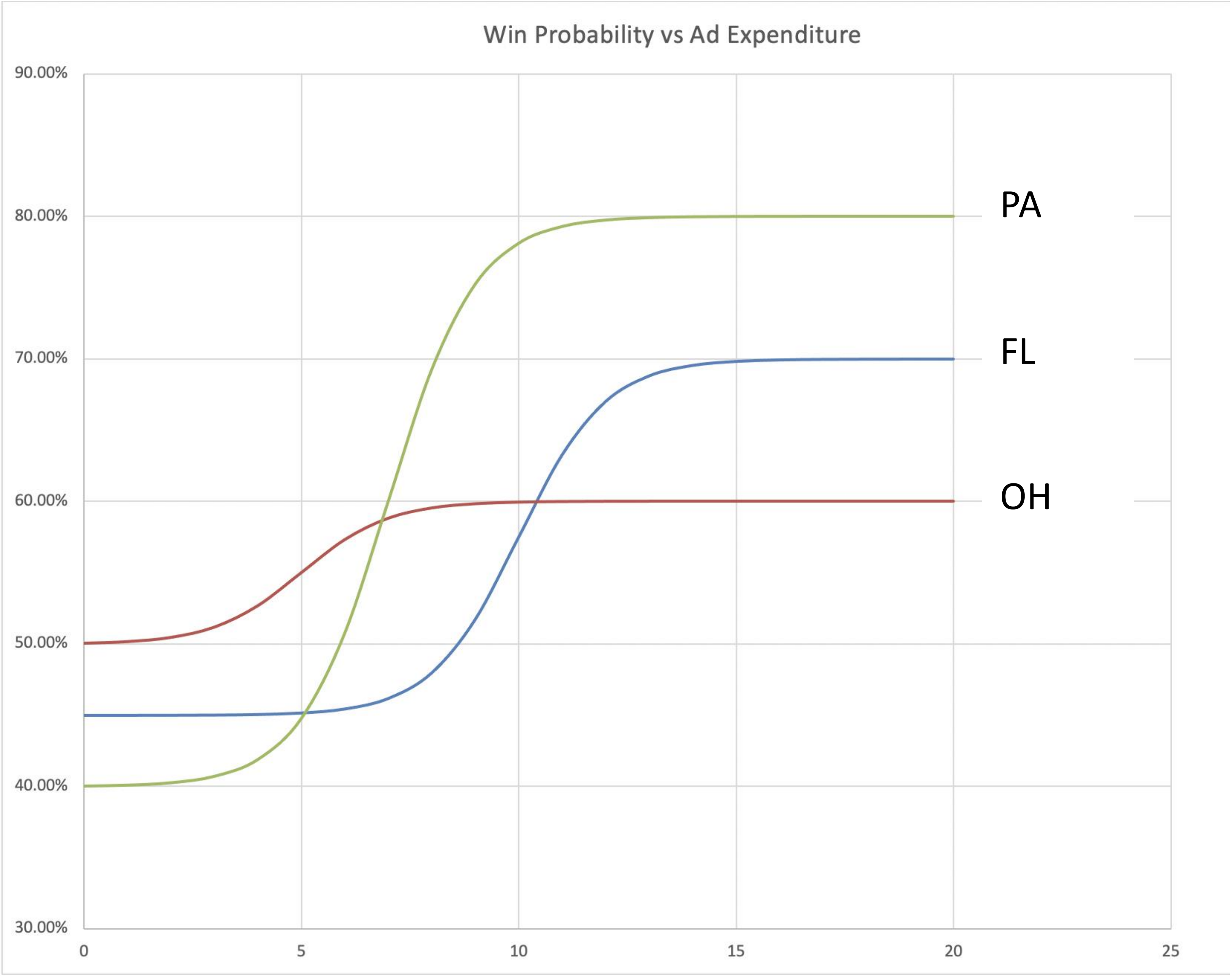
Parameter Estimation Results

Parameter	State		
	FL	**OH**	**PA**
p_{min}	0.45	0.50	0.40
p_{max}	0.70	0.60	0.80
b_0	-10	-5	-7
b_1	1	1	1

Pr(Win FL) Pr(Win OH) Pr(Win PA)

$$\left(0.45 + \frac{(0.70 - 0.45)}{1 + e^{-(x_{FL} - 10)}}\right)\left(0.50 + \frac{(0.60 - 0.50)}{1 + e^{-(x_{OH} - 5)}}\right)\left(0.40 + \frac{(0.80 - 0.40)}{1 + e^{-(x_{PA} - 7)}}\right)$$

Predictive Models for FL, OH and PA



Formulating an Optimization Model

Let’s Formulate the Optimization Problem

Decision variables

Objective function

Constraints

Decision Variables

Decision variables

x_{FL} : \$millions to spend in Florida
x_{OH} : \$millions to spend in Ohio
x_{PA} : \$millions to spend in Pennsylvania

Objective function

Constraints

Objective Function

<u>Decision variables</u>

x_{FL} : \$millions to spend in Florida
x_{OH} : \$millions to spend in Ohio
x_{PA} : \$millions to spend in Pennsylvania

<u>Objective function</u>

What's the Objective Function?

<u>Constraints</u>

Objective Function

Decision variables

x_{FL} : $millions to spend in Florida
x_{OH} : $millions to spend in Ohio
x_{PA} : $millions to spend in Pennsylvania

Objective function

Maximize Pr(Winning the Election)

Constraints

Given our models, what's the probability of winning?

x_{FL} *FL expenditure* → **FL Model** → *Win FL prob* → Pr (Win FL)

$$0.45 + \frac{(0.70 - 0.45)}{1 + \mathrm{e}^{-(x_{FL} - 10)}}$$

x_{OH} *OH expenditure* → **OH Model** → *Win OH prob* → Pr (Win OH)

$$0.50 + \frac{(0.60 - 0.50)}{1 + e^{-(x_{OH} - 5)}}$$

x_{PA} *PA expenditure* → **PA Model** → *Win PA prob* → Pr (Win PA)

$$0.40 + \frac{(0.80 - 0.40)}{1 + \mathrm{e}^{-(x_{PA} - 7)}}$$

Probability of Election Win?

Paths to Victory

Path	Electoral Votes
Win FL and OH and PA	29 + 18 + 20 = 67
Win FL and OH, lose PA	29 + 18 = 47
Win FL and PA, lose OH	29 + 20 = 49

[Win FL] *<u>and</u>* *[don’t lose both PA and OH]*

$$\Pr(\text{Winning the Election}) = \Pr(\text{Win FL})\ [\ 1 - \Pr(\text{Lose OH}) \times \Pr(\text{Lose PA})\ ]$$