More than just a game: What quantitative study of sports can teach us about general principles of statistics

Andrew Gelman
Department of Statistics and Department of Political Science
Columbia University, New York

7 Mar 2016

Classic statistics problems in sports

- In-game decisions
  - Swing at the first pitch?
  - Go for it on 4th down?
- Player evaluation
  - Predict season outcome from aggregated individual statistics
  - Predictions and adjustments (age, context, . . .)
- Predictions and odds
- . . .

nineteen eighty-four

FAT BOYS

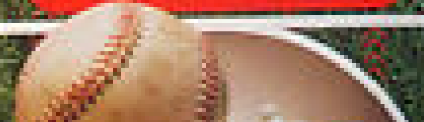

Rolling Stone
BILL MURRAY
The Rolling Stone Interview
On the Road
THE JACKSONS
SPRINGSTEEN

"A MUST FOR FANATICS"

1984

The Bill James BASEBALL ABSTRACT

100% NEW MATERIAL



Some things Bill James taught us

- Outs matter
  - On-base average
  - Caught stealing
- At what age are players at their peak?
- Beware statistical illusions
- Minor league stats predict major-league performance
- ...

Minor league stats predict major-league performance

A friend from Alaska writes:

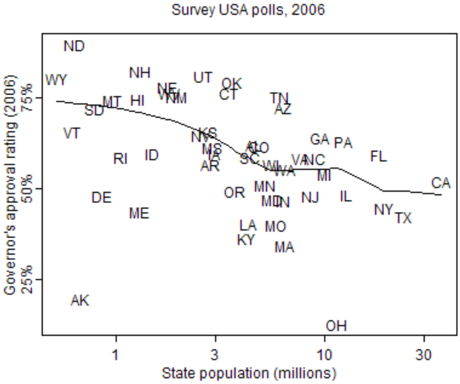
"Palin's magic formula for success has been simply been to ignore partisan crap and get down to the boring business of fixing up a broken government . . . The real significance of Gov Palin's success and her phenomenal approval ratings is that they demonstrate her genuine talent as a non-partisan."

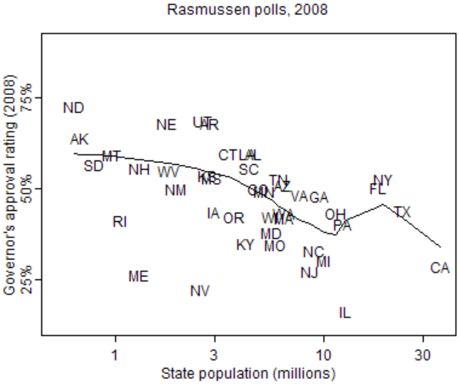
Do minor league stats predict major-league performance?

My Alaskan friend continues:

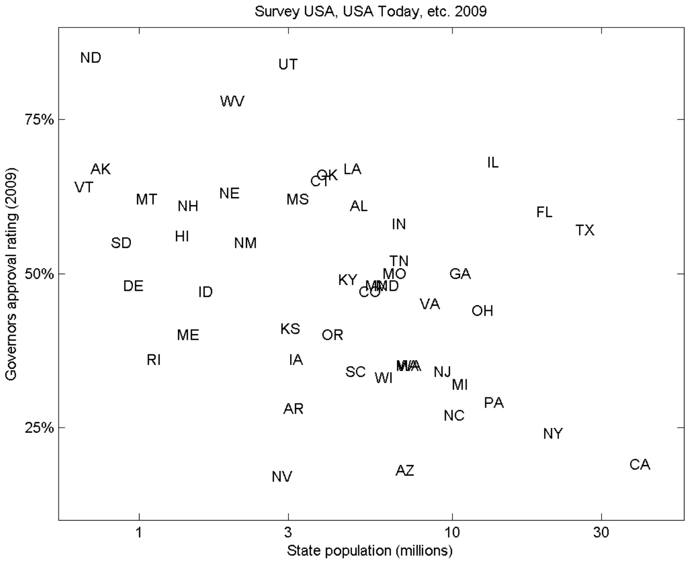
"Sarah Palin is not just popular. She is fantastically popular. Her percentage approval ratings have reached the 90s. Even now, with a minor nepotism scandal going on, she's still about 80%. . . . How does one do that? You might get 60% or 70% who are rabidly enthusiastic in their love and support, but youre also going to get a solid core of opposition who hate you with nearly as much passion. The way you get to 90% is by being boringly competent while remaining inoffensive to people all across the political spectrum."

Popularity of governors





Popular governor, small state

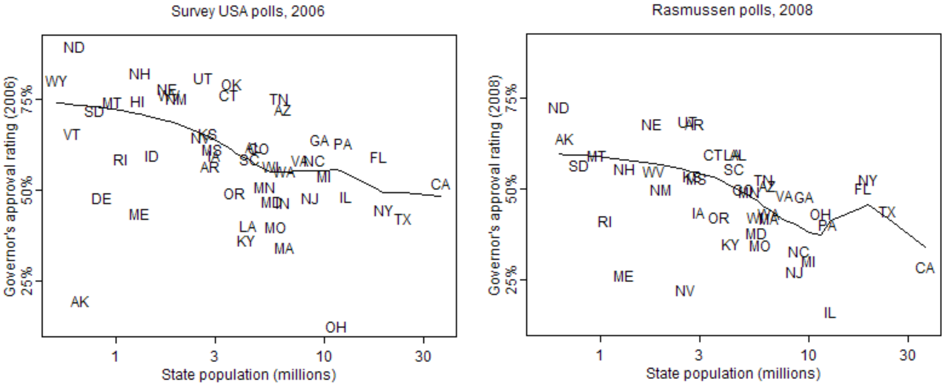


Governor popularity, state size, and the economy

We fit a linear regression ($n = 50$):

```
lm (popularity ~ c.log.statepop + c.income.change)
              coef.est coef.se
(Intercept)      48.6  2.2
c.log.statepop   -6.1  2.3
c.income.change   2.4  1.7
```

How is baseball different from politics?



- One goal (winning) vs. two (winning and policy)
- Ability vs. being in right place at right down
- Single vs. different environments
- Different individual career options
- Small N vs. large N

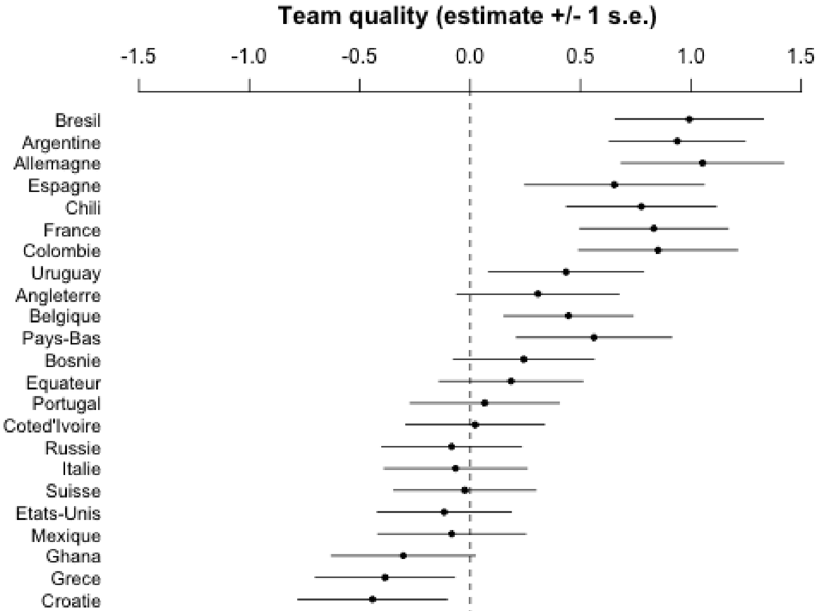
But even the Master is not perfect

"Are athletes special people? In general, no, but occasionally, yes. Johnny Pesky at 75 was trim, youthful, optimistic, and practically exploding with energy. You rarely meet anybody like that who *isn't* an ex-athlete—and that makes athletes seem special."
— Bill James, 2001.

- Do the math:

$$\Pr(\text{athlete} \,|\, \text{75 and energetic}) \propto \Pr(\text{athlete}) \Pr(\text{75 and energetic} \,|\, \text{athlete})$$

Rating soccer teams



worldcup2012.txt

```
Bresil 3 Croatie 1
Mexique 1 Cameroun 0
Bresil 0 Mexique 0
Cameroun 0 Croatie 4
Cameroun 1 Bresil 4
Croatie 1 Mexique 3
Espagne 1 Pays-Bas 5
Chili 3 Australie 1
Espagne 0 Chili 2
Australie 2 Pays-Bas 3
Australie 0 Espagne 3
Pays-Bas 2 Chili 0
Colombie 3 Grece 0
Coted'Ivoire 2 Japon 1
Colombie 2 Coted'Ivoire 1
Japon 0 Grece 0
Japon 1 Colombie 4
Grece 2 Coted'Ivoire 1
Uruguay 1 CostaRica 3
Angleterre 1 Italie 2
Uruguay 2 Angleterre 1
```

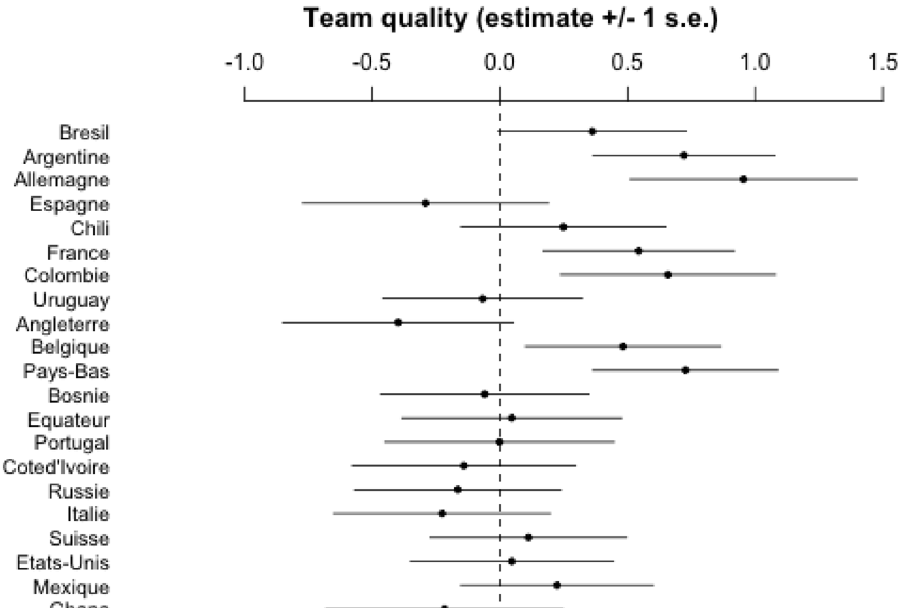
soccerpowerindex.txt

```
Bresil
Argentine
Allemagne
Espagne
Chili
France
Colombie
Uruguay
Angleterre
Belgique
Pays-Bas
Bosnie
Equateur
Portugal
Coted'Ivoire
Russie
Italie
Suisse
Etats-Unis
Mexique
```

The model in Stan

```
parameters {
  real b;
  real<lower=0> sigma_a;
  real<lower=0> sigma_y;
  vector[nteams] eta_a;
}
transformed parameters {
  vector[nteams] a;
  a <- b*prior_score + sigma_a*eta_a;
}
model {
  eta_a ~ normal(0,1);
  for (i in 1:ngames)
    sqrt_dif[i] ~ student_t(df, a[team1[i]]-a[team2[i]],sigma_y)
}
```

Estimates not using prior ranking



Lessons from World Cup example

- Model score differential, not simple wins and losses—even if your only goal is to predict wins and losses
- Same thing in education (model test scores rather than pass/fail) and elections (model vote share not win/loss)
- Jump in and fit a model, then check its fit to data
- Combine sources of information
- Compare different fits graphically

Does it hurt to have a team full of superstars?

SCIENTIFIC AMERICAN™

SUBSCRIBE

MIND

The Surprising Problem of Too Much Talent

A new finding from sports could have implications in business and elsewhere

Roderick Swaab and colleagues suggests there is a limit to the benefit top talents bring to a team. Swaab and colleagues compared the amount of individual talent on teams with the teams' success, and they find striking examples of more talent hurting the team.

Research Article

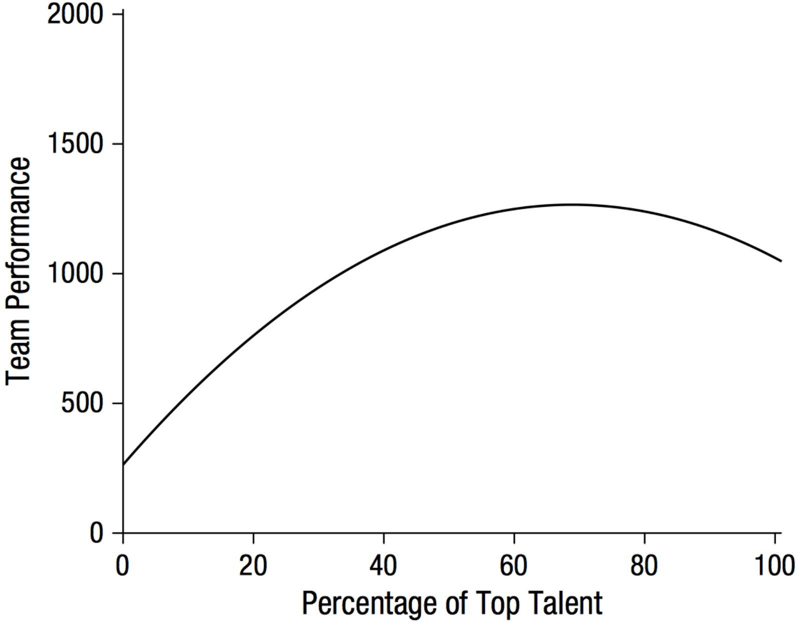
Psychological Science
2014, Vol. 25(8) 1581–1591

DOI: 10.1177/0956797614537280
pss.sagepub.com

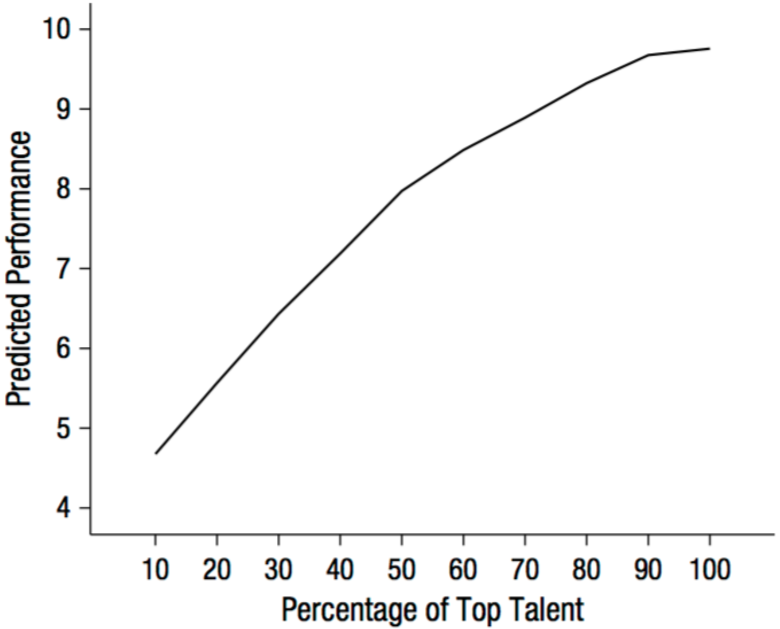
The Too-Much-Talent Effect: Team Interdependence Determines When More Talent Is Too Much or Not Enough

Roderick I. Swaab[1], Michael Schaerer[1], Eric M. Anicich[2], Richard Ronay[3], and Adam D. Galinsky[2]
[1]Organisational Behaviour Area, INSEAD, Fontainebleau, France; [2]Management Department, Columbia University; and [3]Department of Social and Organizational Psychology, VU University Amsterdam

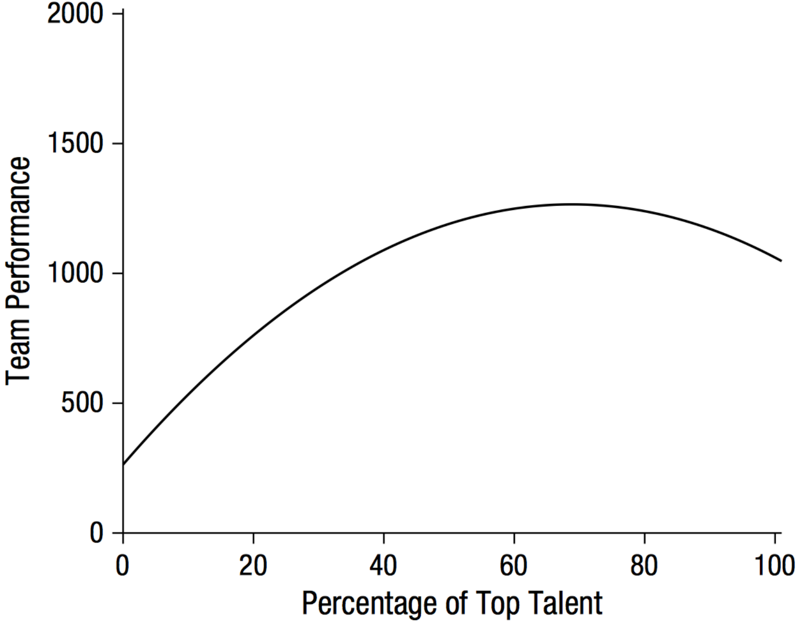
The curve they fit to the data



What ordinary people expected to see



The curve they fit to the data

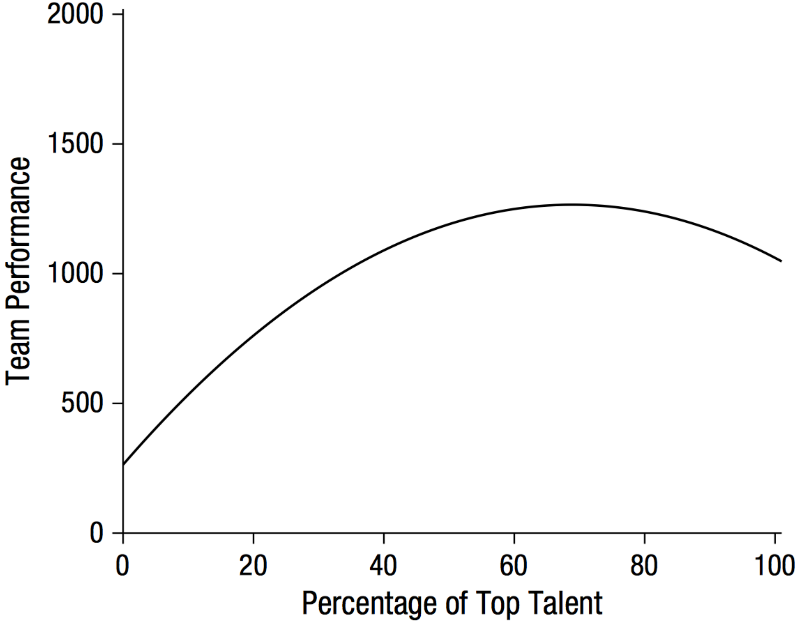


The range of the data

Table 1. Descriptive Statistics and Correlations in Study 2

Measure	*M*	*SD*
1. Team performance (points)	393.30	320.12
2. Top-talent percentage	7%	16%
3. Roster size	18.53	6.79
4. Games played	8.90	4.65

The curve they fit to the data ...



. . . the data!

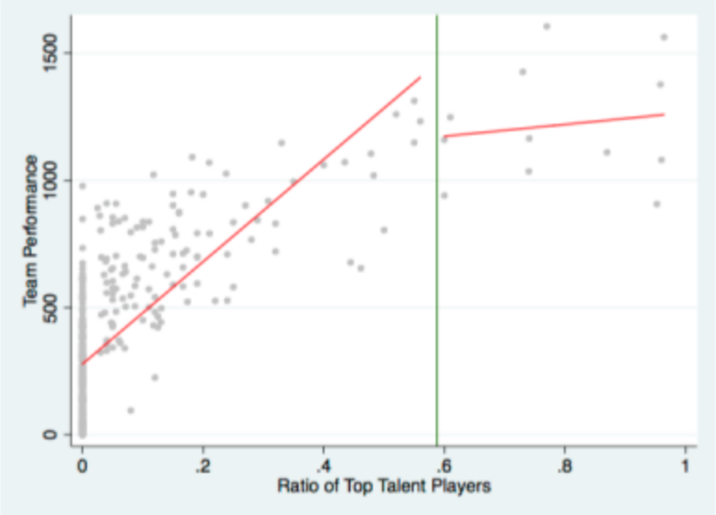


Figure 3d. Soccer performance – top talent (Top 20+ Clubs). S&N test reveals that the first slope is significant and positive ($p \leq .001$) and that the second slope is not significant ($p = .53$).

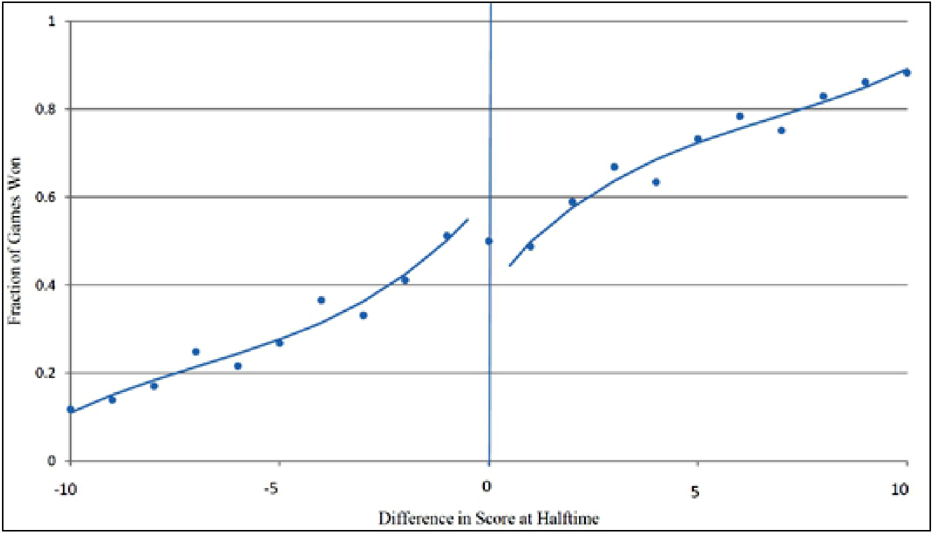
Lessons from superstars example

- Plot data and model together
- Don't overinterpret statistical significance
- Be careful with nonlinear models

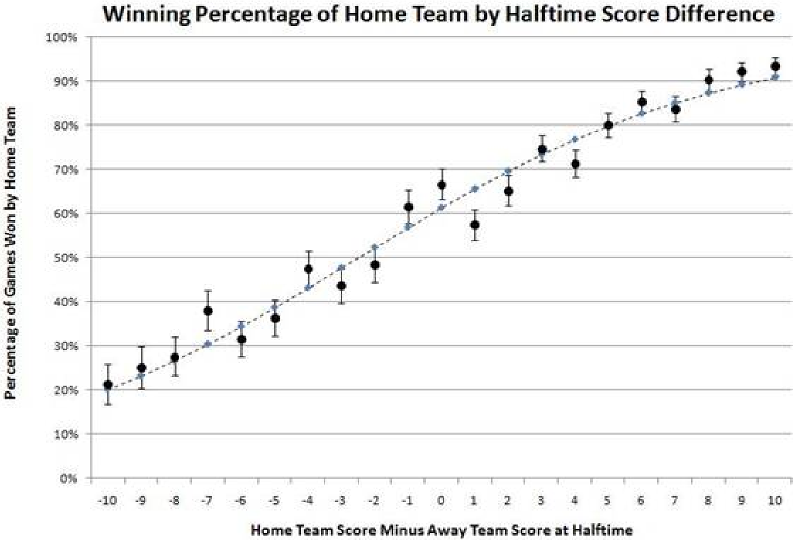
Halftime motivation in basketball

- Economists Jonah Berger and Devin Pope:
  "Analysis of over 6,000 collegiate basketball games illustrates that being slightly behind increases a team's chance of winning. Teams behind by a point at halftime, for example, actually win more often than teams ahead by one. This increase is between 5.5 and 7.7 percentage points . . . "
- But . . . in their data, teams that were behind at halftime by 1 point won 51.3% of the time
- Approx 600 such games; thus, std. error is $0.5/\sqrt{600} = 0.02$
- Estimate ± 1 se is $[0.513 \pm 0.02] = [0.49, 0.53]$
- So where did they get "5.5 and 7.7 percentage points"??

Halftime motivation in basketball: the data and the fitted 5th-degree polynomial



The data without the 5th-degree polynomial



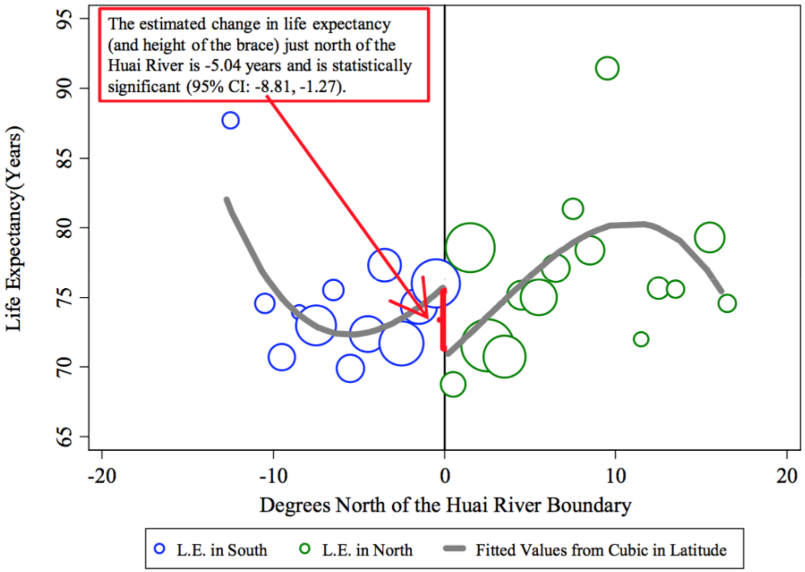


Fig. 3. The plotted line reports the fitted values from a regression of life expectancy on a cubic in latitude using the sample of DSP locations, weighted by the population at each location.

The New York Times

Asia Pacific

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

AFRICA AMERICAS **ASIA PACIFIC** EUROPE MIDDLE EAST

Pollution Leads to Drop in Life Span in Northern China, Research Finds

Proceedings of the National Academy of Sciences of the United States of America

Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy

Yuyu Chen[a,1], Avraham Ebenstein[b,1], Michael Greenstone[c,d,1,2], and Hongbin Li[e,1]

This paper's findings suggest that an arbitrary Chinese policy that greatly increases total suspended particulates (TSPs) air pollution is causing the 500 million residents of Northern China to lose more than 2.5 billion life years of life expectancy. The quasi-experimental empirical approach is based on China's Huai River policy, which provided free winter heating via the provision of coal for boilers in cities north of the Huai River but denied heat to the south. Using a regression discontinuity design based on distance from the Huai River, we find that ambient concentrations of TSPs are about 184 $\mu g/m^3$ [95% confidence interval (CI): 61, 307] or 55% higher in the north. Further, the results indicate that life expectancies are about 5.5 y (95% CI: 0.8, 10.2) lower in the north owing to an increased incidence of cardiorespiratory mortality. More generally, the analysis suggests that long-term exposure to an additional 100 $\mu g/m^3$ of TSPs is associated with a reduction in life expectancy at birth of about 3.0 y (95% CI: 0.4, 5.6).

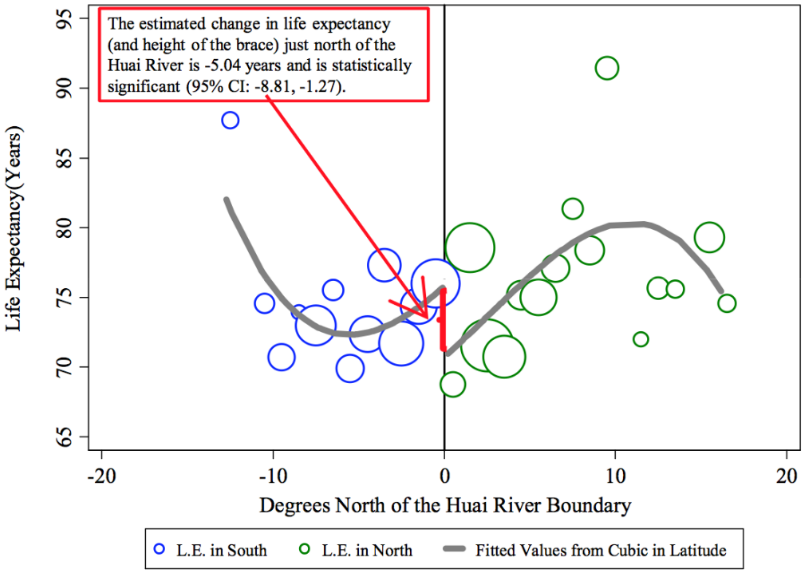


Fig. 3. The plotted line reports the fitted values from a regression of life expectancy on a cubic in latitude using the sample of DSP locations, weighted by the population at each location.

"Robustness checks"

Table S9

Robustness checks of choice of functional form for latitude

	Linear & Controls	Quadratic & Controls	Cubic & Controls	Quartic & Controls	Quintic & Controls
	(1)	(2)	(3)	(4)	(5)
Panel 1: Impact of "North" on the Listed Variable, Ordinary Least Squares					
TSP (100 μg/m^3)	2.89***	2.63***	1.84***	1.95***	1.52**
	(0.56)	(0.49)	(0.63)	(0.59)	(0.72)
Life Expectancy (years)	-1.62	-1.29	-5.52**	-5.67**	-5.43*
	(1.66)	(1.68)	(2.39)	(2.36)	(2.94)

- Recent study: “a win in the 10 days before Election Day causes the incumbent to receive an additional 1.61 percentage points of the vote in Senate, gubernatorial, and presidential elections”
- County-level vote analysis

"Will Ohio States football team decide who wins the White House?"

"The key to victory could come down to . . . Florida, Ohio, and Virginia. On Oct. 27th, a little more than a week before the election, the Ohio State Buckeyes have a big football game against Penn State. The University of Florida Gators have a huge match up against the University of Georgia Bulldogs. If the election remains razor close, these games in these two key battleground states could affect who sits in the White House for the next four years. Can you imagine Ohio State head coach Urban Meyer getting a late night call from the Obama campaign suggesting a particular blitz package? Or maybe Romney has some advice for how the Gators can bottle up Georgia's running game. The decision of whether to punt or go for it on that crucial fourth down could affect the job prospects of more than just the football teams coaching staff." — Tyler Cowen and Kevin Grier, Slate.com

Understanding the college football and elections study

- Assume that a win in the two weeks before the election gives the incumbent an extra 2% of the vote
- What does this imply about the election?
- Not so much . . .
  - Diffusion: A shift of $\pm 1\%$ of the vote in the county containing Columbus, Ohio, corresponds to a 0.1% shift in the state
  - Averaging: In the two weeks before the 2012 election, there were 4 major-college football games involving Ohio teams, and many more in Florida, Virginia, Colorado, etc.

Applying the statistical significance filter

- Results that are statistically significant are likely to be overestimates
- Downgrade the estimated 2% effect

Lessons from football and elections example

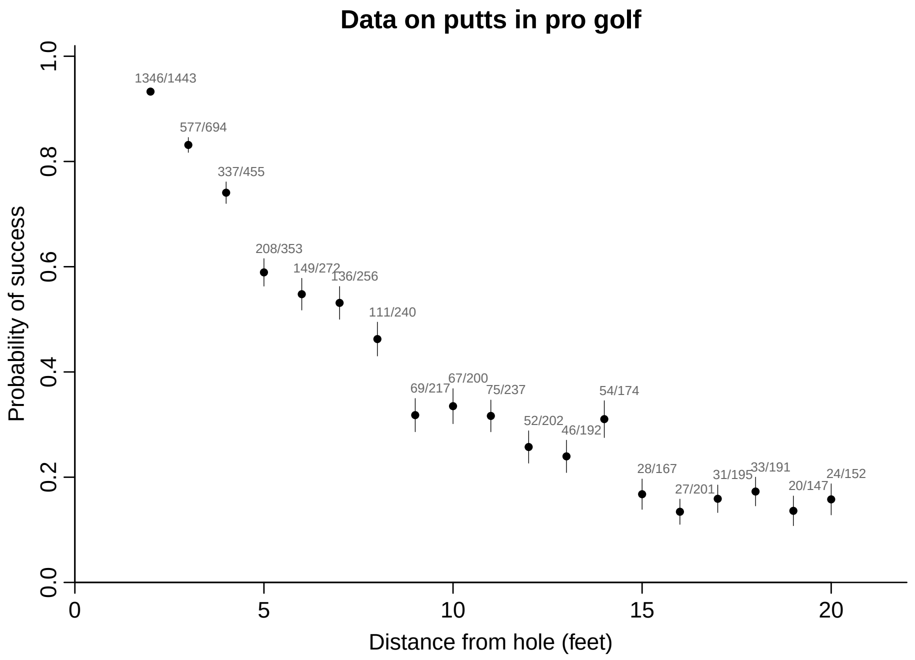
- Map parameter estimates to the real world
- Published estimates are biased upward

Same issues arise in policy analysis

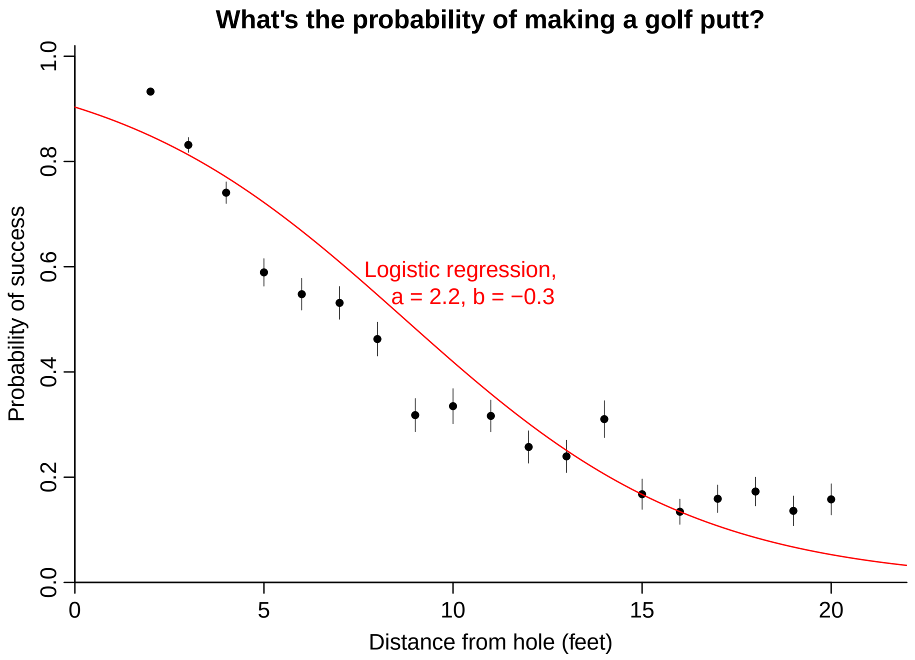
Labor Market Returns to Early Childhood Stimulation: a 20-year Followup to an Experimental Intervention in Jamaica

Paul Gertler, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, Sally Grantham-McGregor

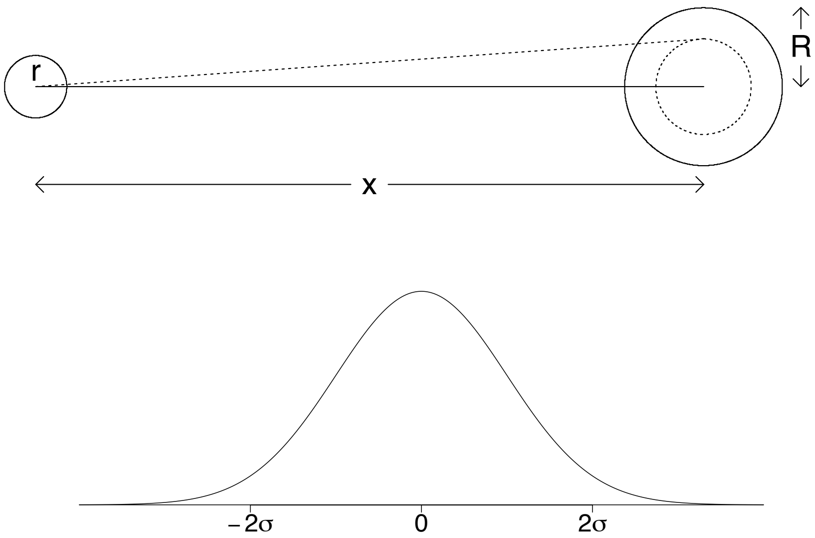
We find large effects on the earnings of participants from a randomized intervention that gave psychosocial stimulation to stunted Jamaican toddlers living in poverty. The intervention consisted of one-hour weekly visits from community Jamaican health workers over a 2-year period that taught parenting skills and encouraged mothers to interact and play with their children in ways that would develop their children's cognitive and personality skills. We re-interviewed the study participants 20 years after the intervention. Stimulation increased the average earnings of participants by 42 percent. Treatment group earnings caught up to the earnings of a matched non-stunted comparison group. These findings show that psychosocial stimulation early in childhood in disadvantaged settings can have substantial effects on labor market outcomes and reduce later life inequality.



Data on putts in pro golf
Probability of success
Distance from hole (feet)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
1346/1443
577/694
337/455
208/353
149/272
136/256
111/240
69/217
67/200
75/237
52/202
46/192
54/174
28/167
27/201
31/195
33/191
20/147
24/152

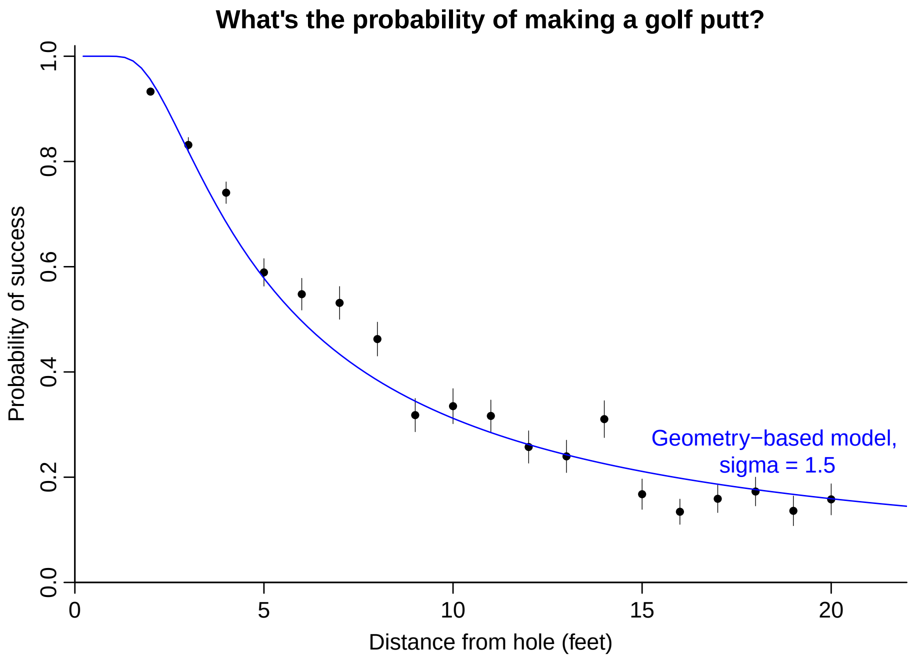

What's the probability of making a golf putt?
Probability of success
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Distance from hole (feet)
Logistic regression,
a = 2.2, b = −0.3

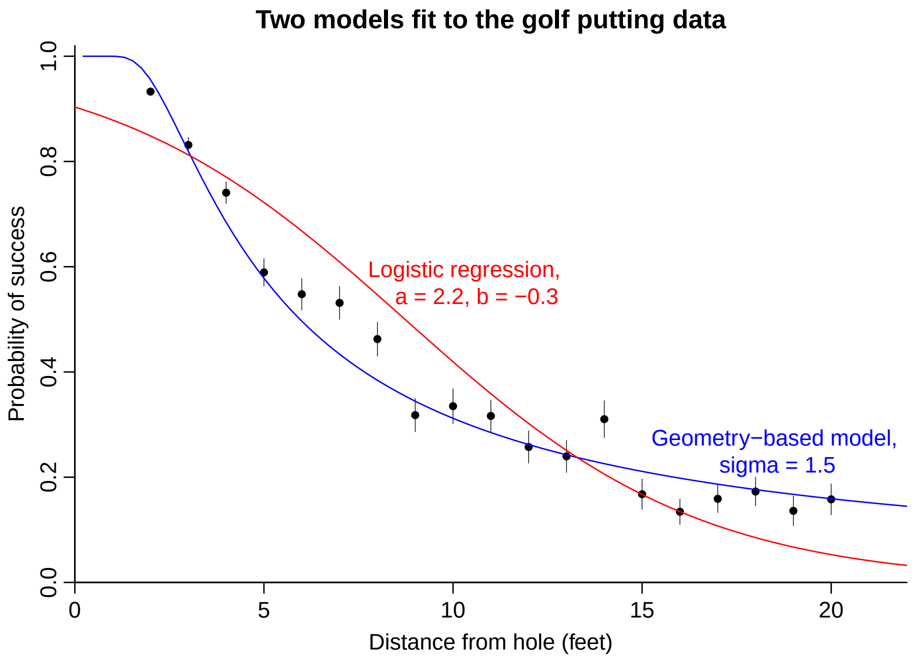
Geometry-based model



Stan code

```
data {
  int J;
  int n[J];
  real x[J];
  int y[J];
  real r;
  real R;
}
parameters {
  real<lower=0> sigma;
}
model {
  real p[J];
  for (j in 1:J)
    p[j] <- 2*Phi(asin((R-r)/x[j]) / sigma) - 1;
  y ~ binomial(n, p);
}
```


What's the probability of making a golf putt?
Probability of success
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Distance from hole (feet)
Geometry−based model,
sigma = 1.5


Two models fit to the golf putting data
Probability of success
Distance from hole (feet)
Logistic regression,
a = 2.2, b = −0.3
Geometry−based model,
sigma = 1.5
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20

Lessons from golf example

- You can (sometimes) fit and interpret a theory-based model
- Plot data and model together
- Compare different fits graphically

Summary

- God is in every leaf of every tree
- Statistics in sports:
  - Variation, uncertainty, regularity, decision making . . . life!
- Science in sports:
  - Physiology, motivation, competition . . . life!