

**SMALL DATA:
HOW MODERN
ECONOMETRICS CAN AID
RAPID COLLABORATION**

**Gavin
McCormick**

**UC Berkeley &
WattTime.org**

**Sage Scholars
Conference
2015**



5 YEARS



8 HOURS





5 YEARS



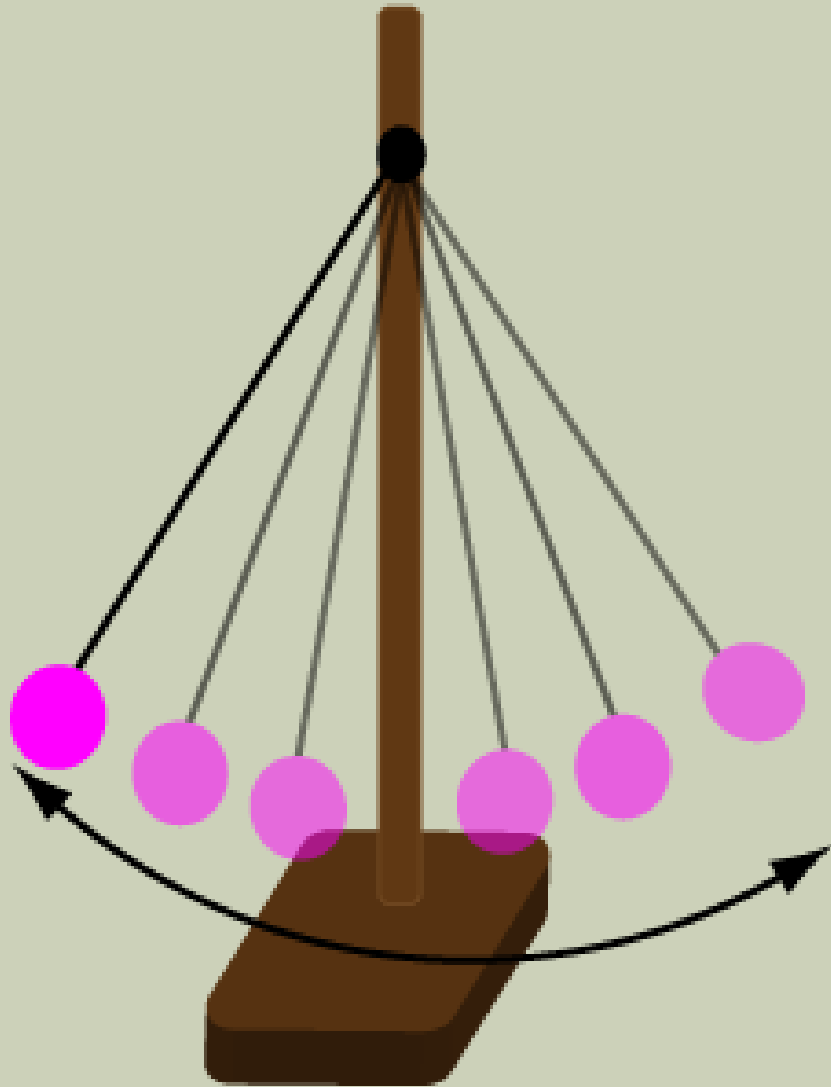
8 HOURS



Disclaimer!

THINGS ECONOMICS GOT WRONG

- Recessions
- Unemployment
- Growth
- Happiness
- Incentives
- Innovation
- Uncertainty
- Transitivity
- The economy





PLAYING “SMALL BALL”

- **Standard approach:**
 - Examine data on A and B
 - Result: effect of A on B

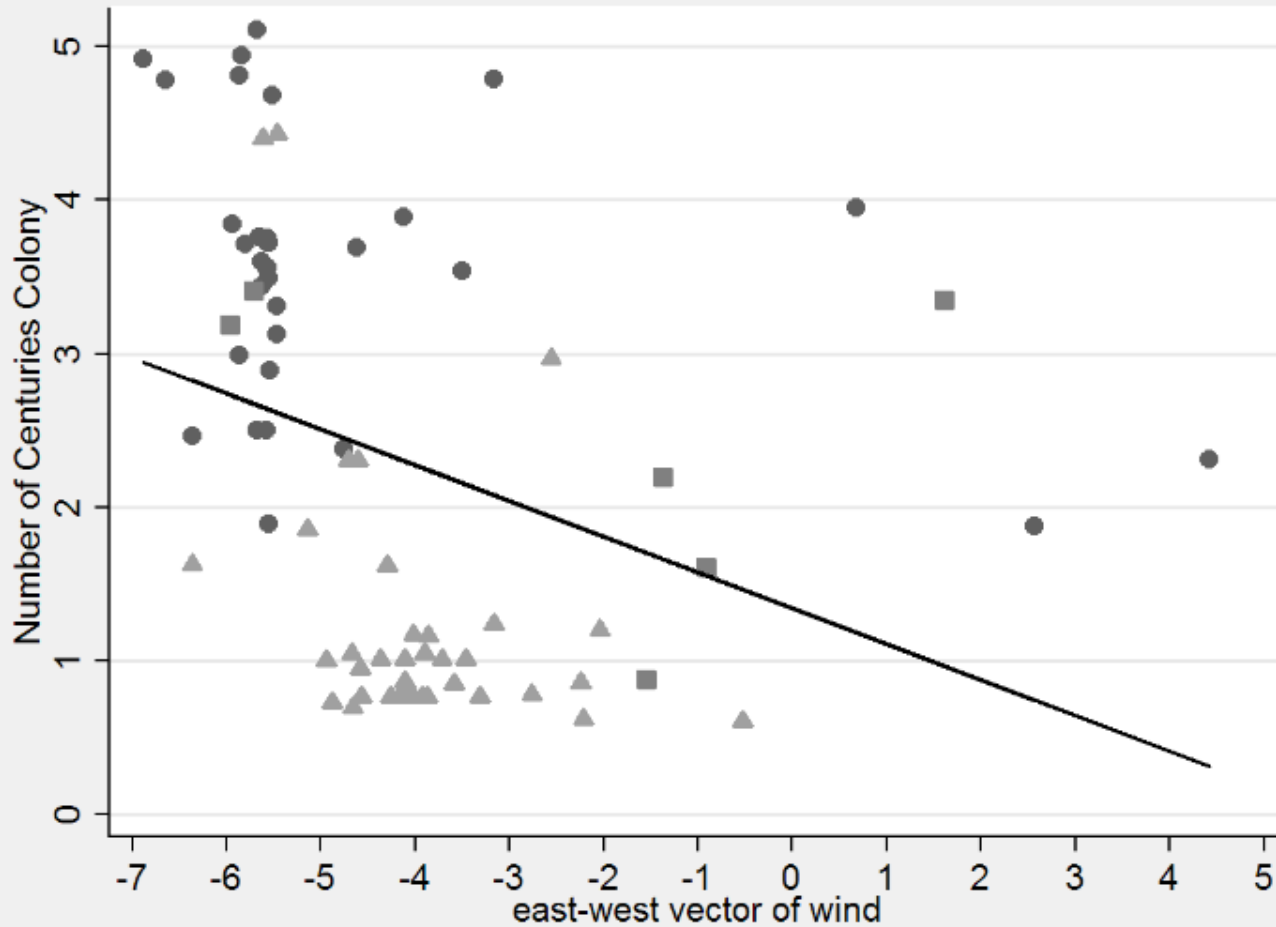
PLAYING “SMALL BALL”

- **Standard approach:**
 - Examine data on A and B
 - Result: effect of A on B
- **Big Data approach:**
 - Examine everything
 - Then mine for correlations
 - Result: many questionable effects

PLAYING “SMALL BALL”

- **Standard approach:**
 - Examine data on A and B
 - Result: effect of A on B
- **Big Data approach:**
 - Examine everything
 - Then mine for correlations
 - Result: many questionable effects
- **“Small Data” approach:**
 - First mine for randomness
 - Then examine whatever’s random
 - Result: few certain effects

Colonialism vs. Wind Direction



FERYR AND
SACERDOTE
(2008)

**startups should
fail fast**

THE BEST STARTUP...

- Is a science experiment
- That concludes quickly
- And cheaply

THE BEST STARTUP...

- Is a science experiment
- That concludes quickly
- And cheaply
- Is simple

THE BEST STARTUP...

- Is a science experiment
- That concludes quickly
- And cheaply
- Is simple
- Is a huge impact or nothing

THE BEST STARTUP...

- Is a science experiment
- That concludes quickly
- And cheaply
- Is simple
- Is a huge impact or nothing
- Probably wasn't your plan

THE WATTTIME MODEL OF COLLABORATIVE INNOVATION

- 1) Combine any 2+ large public datasets
- 2) Define some outcomes of interest
- 3) Throw out all nonrandom variation
- 4) Check for strong causal relationships
- 5) If exists: found a startup! If not: try again next weekend
- 6) Since no capital + strong impact, bootstrap to maintain independence

SAMPLE MEDICAL APPLICATIONS

If...	Study...	Vs...
Hospital incentives are mis-aligned	Treatments	Health
Occupy shuts down a port	Air pollution	Asthma
Walmart makes a food desert	Diet	Health
Medicare is easier at 65	Expensive meds	Health
A city exceeds ozone limits	Air pollution	Asthma

GENERALIZABLE MEDICAL RESEARCH TECHNIQUE?

Hospitals vary in how they are reimbursed

```
graph TD; A[Hospitals vary in how they are reimbursed] --> B[Leads to random variation in treatment decisions]; B --> C[Quasi-experimental variation in treatments]; C --> D[Can consistently quantify effect of many treatments];
```

Leads to random variation in treatment decisions

Quasi-experimental variation in treatments

Can consistently quantify effect of many treatments

THANKS!

For further information:

- Agrist & Pischke (2010), “The Credibility Revolution in Empirical Economics”
- Agrist and Pischke, Mostly Harmless Econometrics
- Banerjee and Duflo, Poor Economics
- EcoHack.org
- WattTime.org
- gavin.mccormick@berkeley.edu