

What the Numbers Say: A Digit-Based Test for Election Fraud Using New Data from Nigeria

Bernd Beber Alexandra Scacco

Columbia University

APSA Annual Meeting
Boston, August 2008

Research question

Can we detect manipulation by just looking at electoral returns?

Basic idea

Exploit human biases in number generation

Research question

Can we detect manipulation by just looking at electoral returns?

Basic idea

Exploit human biases in number generation

Two approaches in the literature:

- Covariate-based studies (Wand et al. 2001; Mebane & Sekhon 2004)
 - But what about data-poor environments?
- Digit-based tests (Mebane 2006)
 - But do digits from clean electoral results follow a Benford distribution?

Two approaches in the literature:

- Covariate-based studies (Wand et al. 2001; Mebane & Sekhon 2004)
 - But what about data-poor environments?
- Digit-based tests (Mebane 2006)
 - But do digits from clean electoral results follow a Benford distribution?

A different digit-based approach

- Focus on last digits for which possible values are equiprobable
- Does not require strong assumptions about how electoral returns are distributed
- Can be implemented in the absence of contextual data

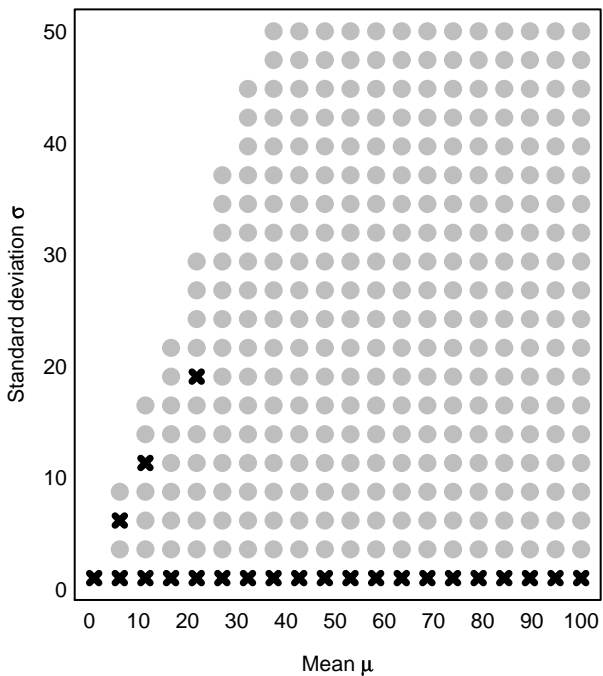
A different digit-based approach

- Focus on last digits for which possible values are equiprobable
- Does not require strong assumptions about how electoral returns are distributed
- Can be implemented in the absence of contextual data

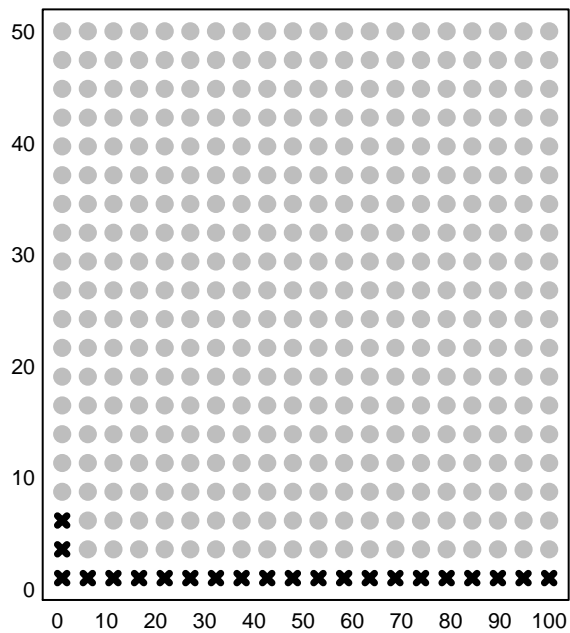
A different digit-based approach

- Focus on last digits for which possible values are equiprobable
- Does not require strong assumptions about how electoral returns are distributed
- Can be implemented in the absence of contextual data

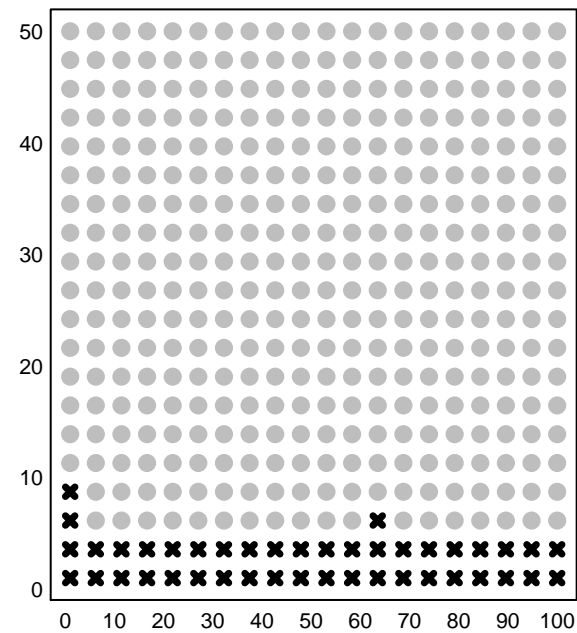
Binomial (for $\sigma^2 < \mu$), negative binomial (for $\sigma^2 > \mu$),
and Poisson (for $\sigma^2 = \mu$) distributions



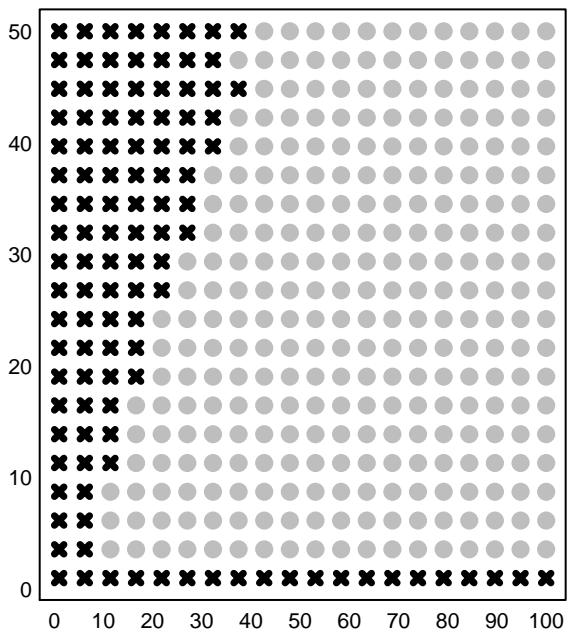
Discretized normal distribution, truncated at 0



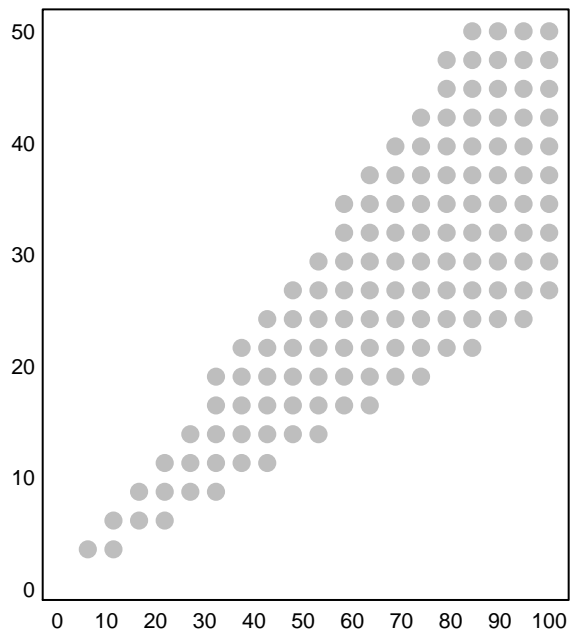
Discretized uniform distribution, truncated at 0



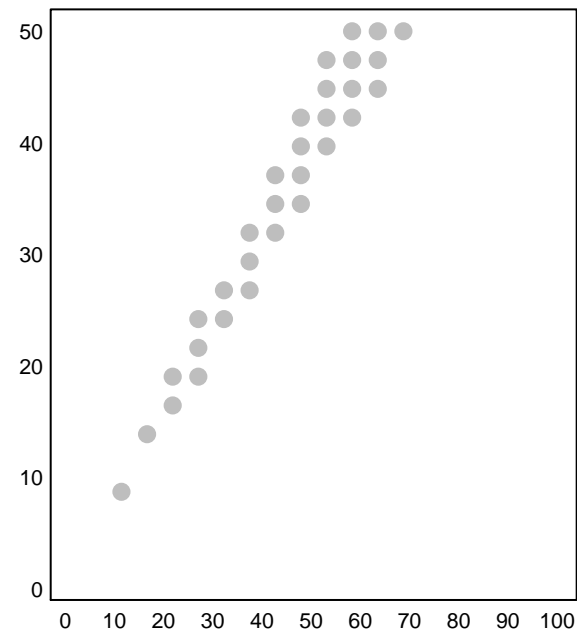
Discretized Gamma distribution



Discretized mix of distributions
(Mechanism A, Mebane 2006)

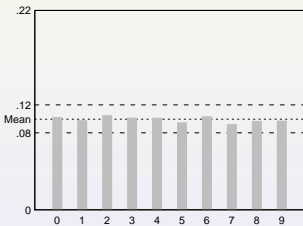


Discretized mix of distributions
(Mechanism B, Mebane 2006)

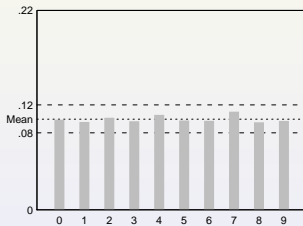


Empirical baseline: Sweden 2002

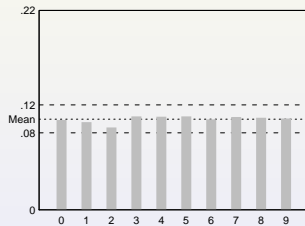
SAP results



Share of last digits
MSP results

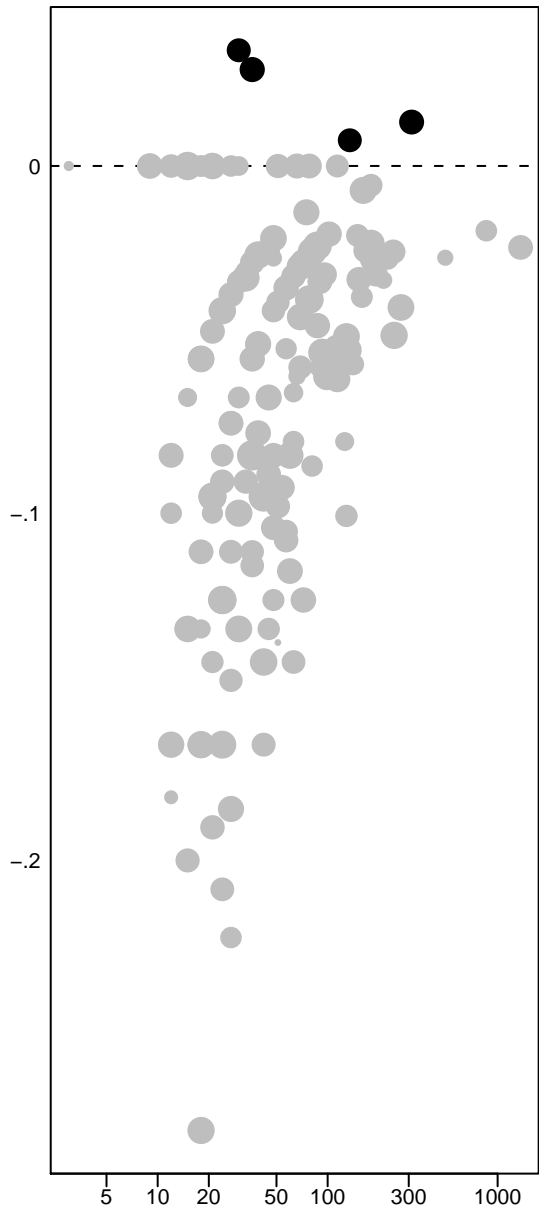


Number of registered voters

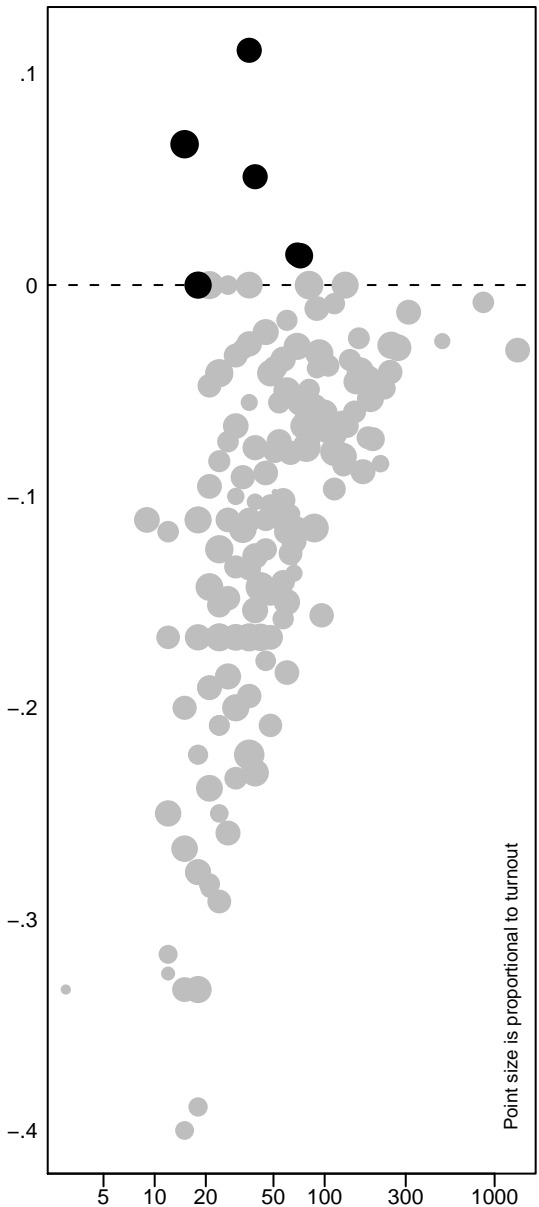


Multiple columns

5th pct of simulated frequency with which last two digits are identical, minus empirical frequency

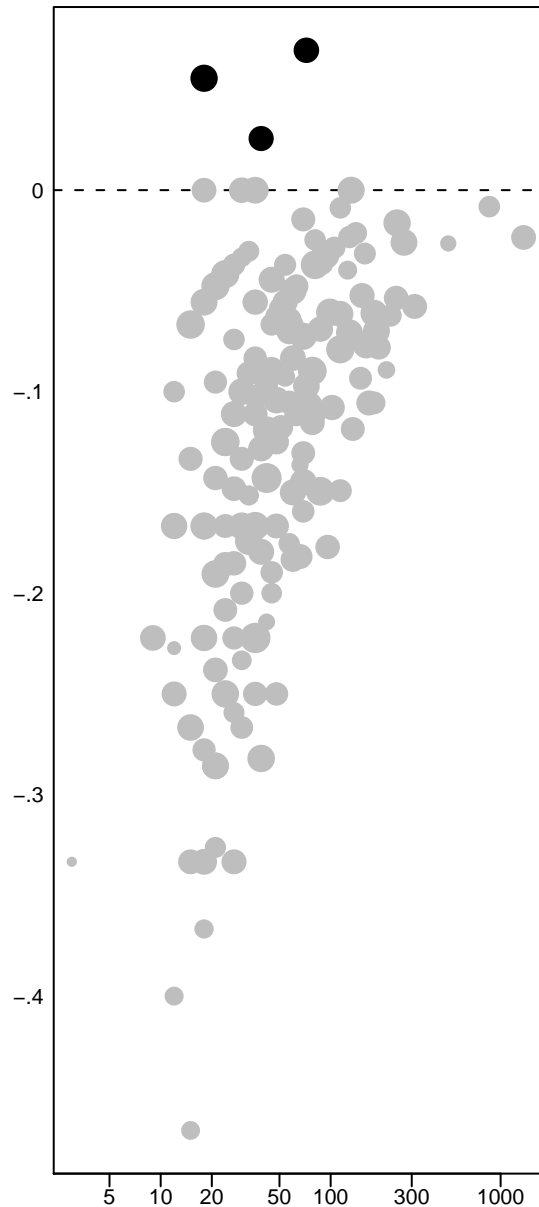


Empirical frequency minus 95th pct of simulated frequency with which distance between last and penultimate digit is one



Point size is proportional to turnout

5th pct of simulated frequency with which distance between last and penultimate digit is greater one, minus empirical frequency



Number of wards in municipality (logged)

Cognitive biases in number generation

Human subjects tend to:

- Select some numbers more often than others
- Avoid repetition of digits
- Prefer adjacent digits

(Chapanis 1953, Rath 1966, Boland & Hutchinson 2000)

Cognitive biases in number generation

Human subjects tend to:

- Select some numbers more often than others
- Avoid repetition of digits
- Prefer adjacent digits

(Chapanis 1953, Rath 1966, Boland & Hutchinson 2000)

Cognitive biases in number generation

Human subjects tend to:

- Select some numbers more often than others
- Avoid repetition of digits
- Prefer adjacent digits

(Chapanis 1953, Rath 1966, Boland & Hutchinson 2000)

Cognitive biases in number generation

Human subjects tend to:

- Select some numbers more often than others
- Avoid repetition of digits
- Prefer adjacent digits

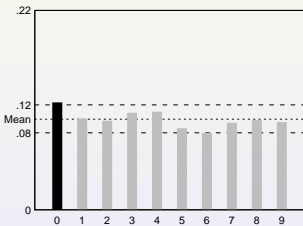
(Chapanis 1953, Rath 1966, Boland & Hutchinson 2000)

- Data for 2003 presidential election from original ward enumeration sheets used by local authorities in Plateau state
- In Awka, Anambra state, “barely any polling stations were provided with a results-sheet [...]; presumably these were being filled in elsewhere” (The Economist, April 21, 2007)
- No covariates

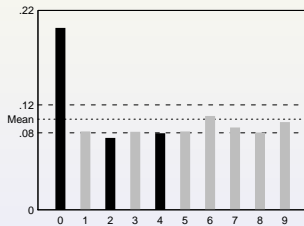
- Data for 2003 presidential election from original ward enumeration sheets used by local authorities in Plateau state
- In Awka, Anambra state, “barely any polling stations were provided with a results-sheet [...]; presumably these were being filled in elsewhere” (The Economist, April 21, 2007)
- No covariates

- Data for 2003 presidential election from original ward enumeration sheets used by local authorities in Plateau state
- In Awka, Anambra state, “barely any polling stations were provided with a results-sheet [...]; presumably these were being filled in elsewhere” (The Economist, April 21, 2007)
- No covariates

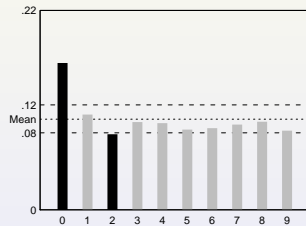
PDP results



Share of last digits
Total vote counts



Number of registered voters

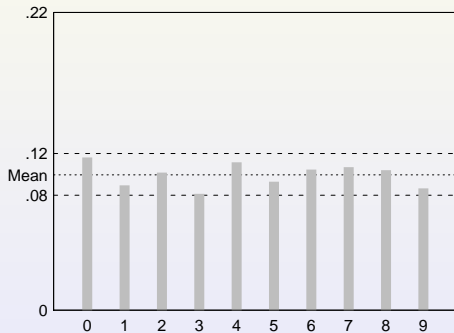
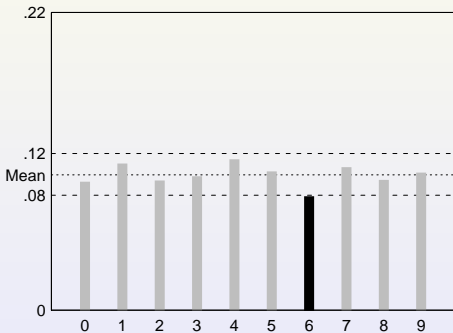


An intermediate case: Chicago 1924 & 1928

Share of last digits Chicago 1924

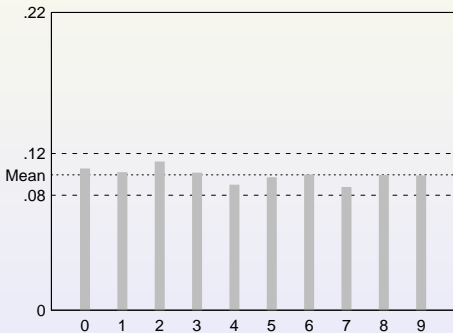
Davis

Coolidge



Share of last digits Chicago 1928

Smith



Hoover

