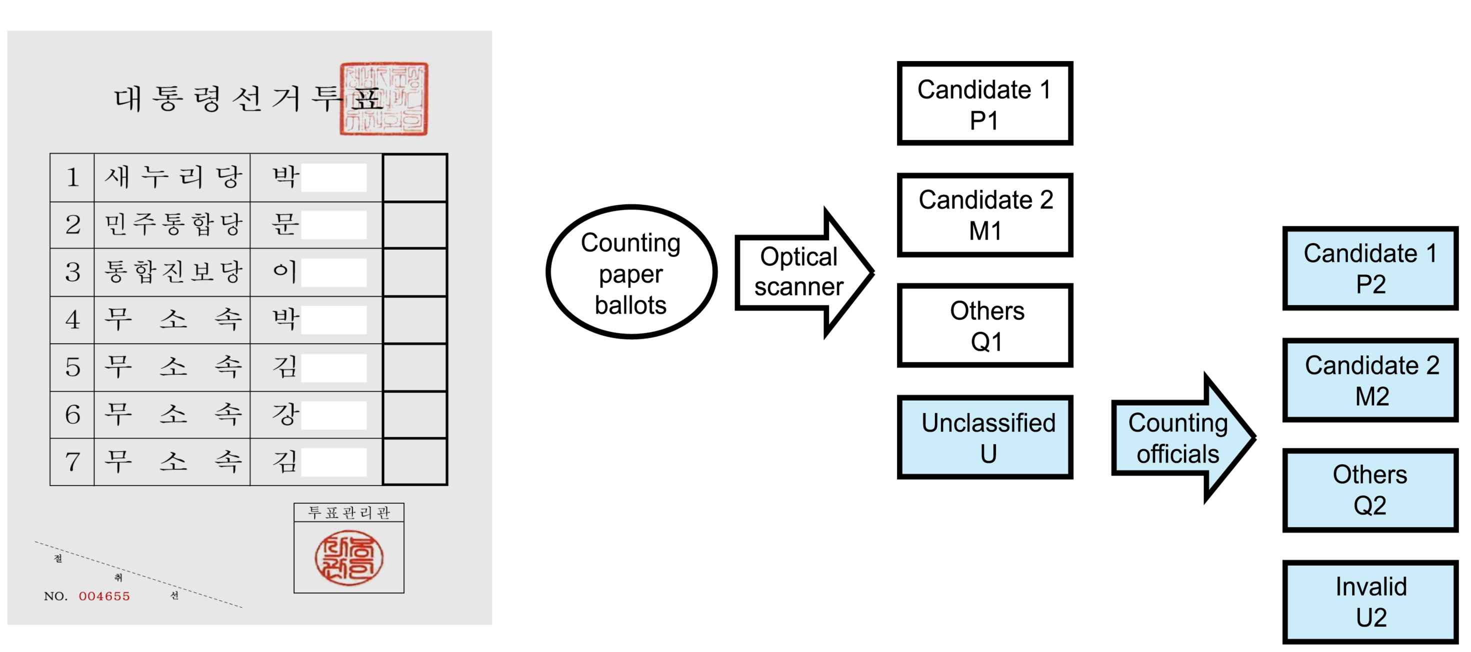
A Master Plan 1.5 Using Optical Scan Counters: An Analysis of the 2012 Presidential Election Data in South Korea

HeeKyoung Chun[1], Pierre-Jérôme Bergeron[2], HyunSeung Kim[3], OuJoon Kim[3], Hwashin Hyun Shin [4*]
[1]GeorgiaSouthern University [2]University of Ottawa [3]Project Boo Inc. [4]Queen's University [*]corresponding author

Abstract

- Optical scan counters in election have been used to count votes, sorting them into classified and unclassified ballots. The unclassified ballots are manually counted by electoral officials.
- We propose a measure **K**, to test between-candidate relative inequality whose valid ballots are unclassified by the optical counters. We prove **E[K]=1** and Var(K) depends on electoral district size and unclassified rate.
- The 2012 presidential election in South Korea was a close election with 52% vs 48% for the top two candidates.
- We found **K=1.5** nationally, which indicates a **strong discrepancy** between *machine counted* and *human counted* ballots.
- Both **systematic unintentional bias** and **intentional manipulation** were considered to explain this. Simulation results from the latter scenario with rigged optical counters are quite close to actual vote results, suggesting the election might have been affected by a national manipulation as optical counters are vulnerable programmable devices.
- The proposed statistical methods for K contribute to securing accurate vote counting in elections worldwide where the optical counters are used.

Background: Votes Sorting Process: Classified versus Unclassified



Motivation Example

Voting results for three specific districts from three presidential elections in 2002, 2007, 2012

	Votes from the classified (sorted by op-scanners)		Votes from the unclassified (sorted by counting officials)		
Election year and district	candidate 1 (P_C)[a]	candidate 2 (M_C)[a]	candidate 1 (P_U)[b]	candidate 2 (M_U)[b]	$K=K_U/K_C$
16th in 2002, district G	36.3%	56.5%	33.7%	50.3%	1.04
17th in 2007, district N	23.1%	47.7%	25.1%	50.7%	1.02
district Y	16.9%	59.6%	17.5%	59.4%	1.04
18th in 2012, district G	40.2%	59.4%	39.9%	43.7%	1.35
district N	46.3%	53.3%	45.0%	36.0%	1.44
district Y	51.9%	47.8%	54.4%	36.6%	1.37

[a] P_C= (votes for candidate 1/total votes); M_C= (votes for candidate 2/total votes) from classified
[b] P_U= (candidate 1/total votes); M_U= (candidate 2/total votes) from unclassified; $K_U= P_U/M_U$

A Proposed Measure for Between-Candidate Relative Inequality

$K=K_U/K_C$ = (P2/M2)/(P1/M1) for any electoral district.
- P1 & M1: votes counts from the classified votes
- P2 & M2: votes counts from the unclassified votes

- **Known fact:** $X \sim B(n, p)$, $E\left[\frac{1}{X+1}\right] = \frac{1}{(n+1)p} \simeq \frac{1}{np}$ for large n

- **Theoretical expectation of K: E[K]=1**

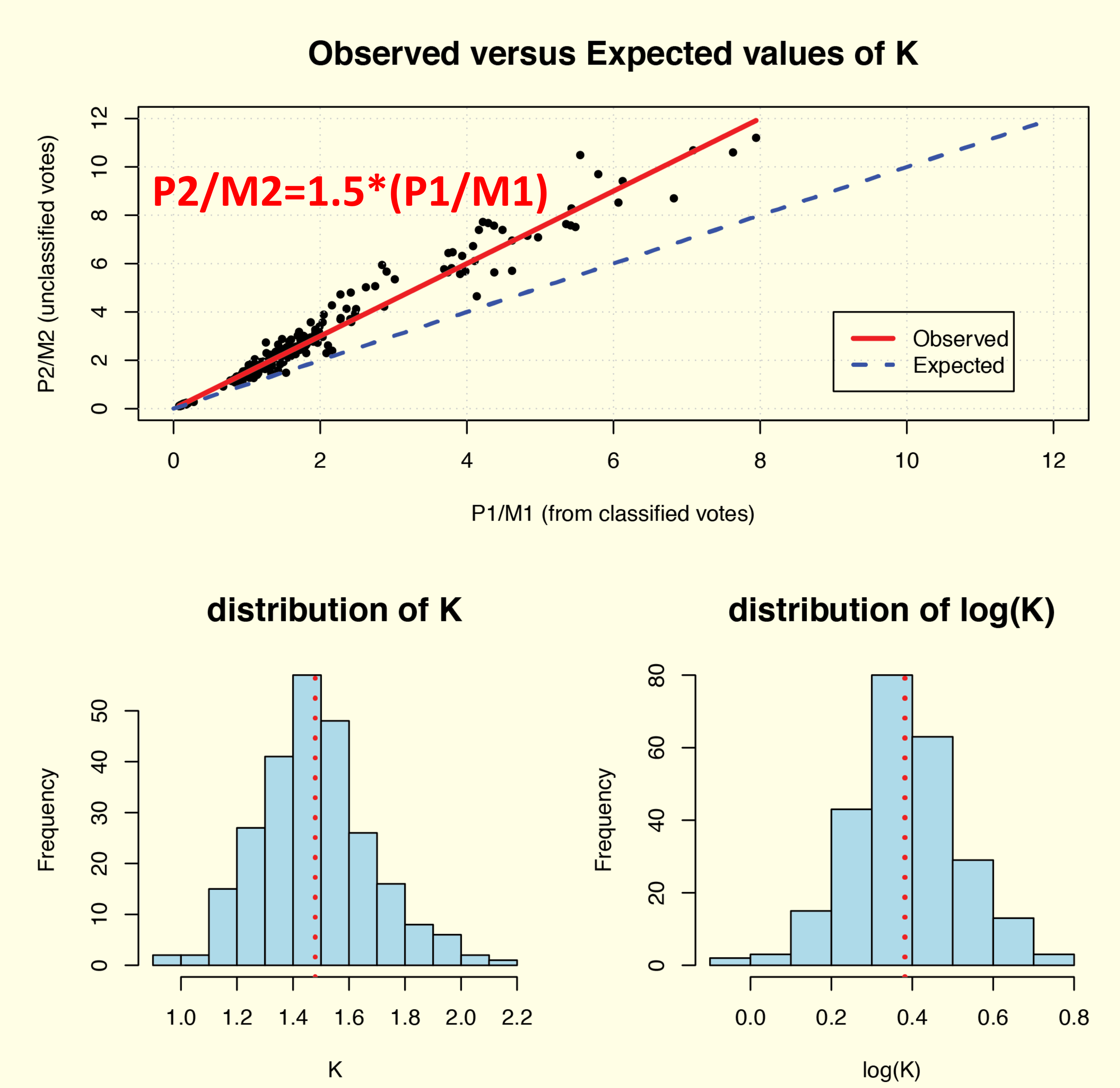
$$E[K] = E\left[\frac{P2/M2}{P1/M1}\right] = E\left[\frac{P2}{P1}\right]\cdot E\left[\frac{M1}{M2}\right] = \frac{r\cdot(1-r)}{(1-r)\cdot r} = 1$$

P1~ B(P, 1-r), P2~B(P,r), where P1+P2=P fixed.
M1~B(M,1-r), M2~(M,r), where M1+M2=M fixed.
$r=Pr(U_M)=Pr(U_P)$, rate of the unclassified votes

- **Variance of K depending on electoral district size and r**

$$Var(K) \cong \frac{(P+M)r(1-r)+1}{PM(r(1-r))^2} \cong \frac{(P+M)}{PMr(1-r)}$$

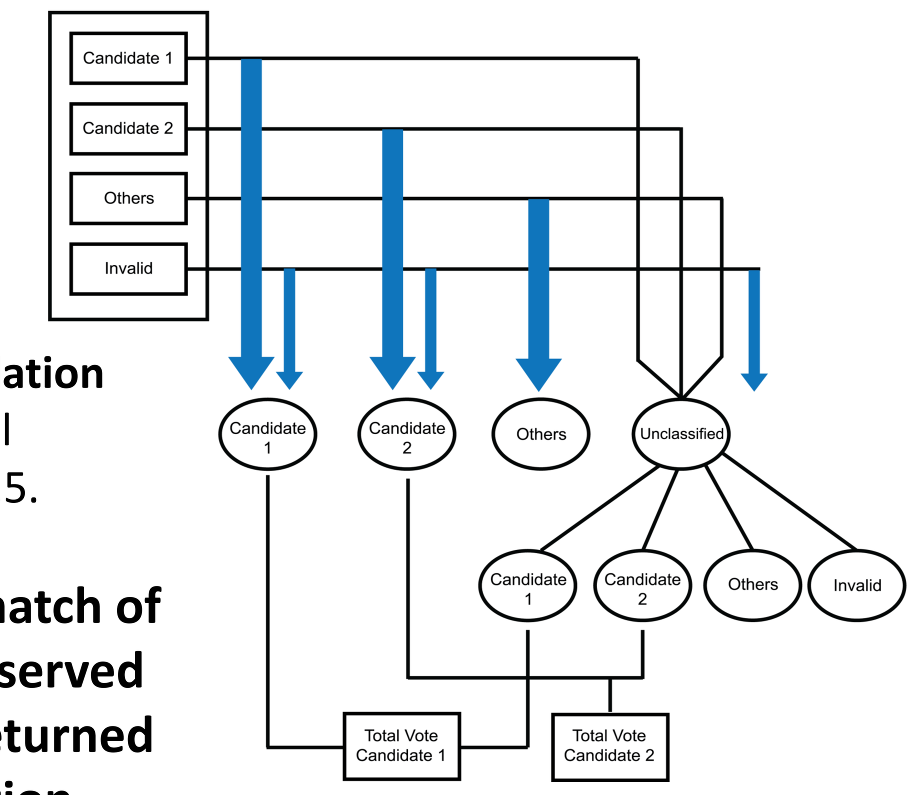
National Model: K=1.5 (Linear Regression with R^2=0.98)



Simulations with Two Scenarios

- **Scenario 1: unintentional systematic bias**
  Assume a consistent op-scan performance across the country:
  P2 ~ B(P, r+β), where M2 ~ B(M, r).
- **Scenario 2: intentional systematic manipulation**
  Apply conditional Bernoulli and multinomial distributions to sorting algorithm with K=1.5.

➔ **Scenario 1 resulted in an overall mismatch of the variability of unclassified rates observed at district level, whereas scenario 2 returned comparable results to the actual election outcomes at both district and national levels.**



Simulation Process (Scenario 2)

Results: Actual vs. Simulated Votes

251 districts combined	Total number of votes	Total unclassified votes	Votes from the classified candidate 1	Votes from the classified candidate 2	Votes from the unclassified candidate 1	Votes from the unclassified candidate 2	Invalid votes
Actual	29,827,252	1,111,165	14,782,150	13,828,239	586,632	397,505	112,360
Simulated 1	29,827,252	1,229,495	14,683,046	13,797,770	685,822	427,442	112,570
Simulated 2	29,827,252	1,080,700	14,787,440	13,857,352	594,739	370,907	111,117

Merits of the Relative Ratio K

- The proposed method can be more effective than auditing when the op-scan counters are used.
- The K-value can be examined for some electoral districts individually for local, regional, or national levels.
- Easy and economic for time and cost required

Discussions on Op-scanners and Fair Election

- This study demonstrates a potential serious loophole in using the op-scan counters, which can be **error-free but not manipulation-free**
- Op-scan counters can generate serious misclassifications **by a pre-programmed algorithm**, not by random mechanical malfunctions.
- Serious errors can go undetected if results are not audited effectively, resulting in election winner change.

➔ Solutions for detecting election frauds:
- ✓ Auditing a well curated paper trail against the electronic results or a random sample of the ballot boxes
- ✓ The proposed measure (K) to detect between candidate relative inequality