Measuring Benefits from Improving Accuracy of the 2020 Census: Apportionment of the U.S. House of Representatives and Allocation of Federal Funds

Bruce D. Spencer[12] and Zachary H. Seeskin[23]

NCRN Virtual Seminar, December 3, 2014

Preliminary Draft - Do Not Cite or Quote

[1] Professor, Department of Statistics, and Faculty Fellow, Institute for Policy Research, Northwestern University, Evanston, IL 60208-4070, U.S.A. email: bspencer@northwestern.edu

[2] Research supported by NSF grant SES-1129475, "NCRN-SN: Census Bureau Data Programs as Statistical Decision Problems".

[3] Graduate Student, Department of Statistics, and Graduate Research Assistant, Institute for Policy Research, Northwestern University.

Overview

The Census Bureau is currently planning the 2020 decennial census.

Target sticker price: $20 billion. Will that buy sufficient accuracy?

Our approach: use statistical decision theory (cost-benefit analysis).

We develop estimates of effects of census errors on apportionment of the House of Representatives and on allocations of federal funds.

C.V. of State Population	Expected No. of House Seats to Wrong State	Expected Sum of Errors in Allocations of Federal Funds, 2021-2031 ($ Billions)	Census Costs (Not yet specified)
0.2%	0.8	4.5	high
0.4%	1.5	8.9	?
0.6%	2.1	13.4	?
0.8%	2.8	17.8	?
1.0%	3.4	22.3	low

Preliminary - Do Not Cite or Quote!

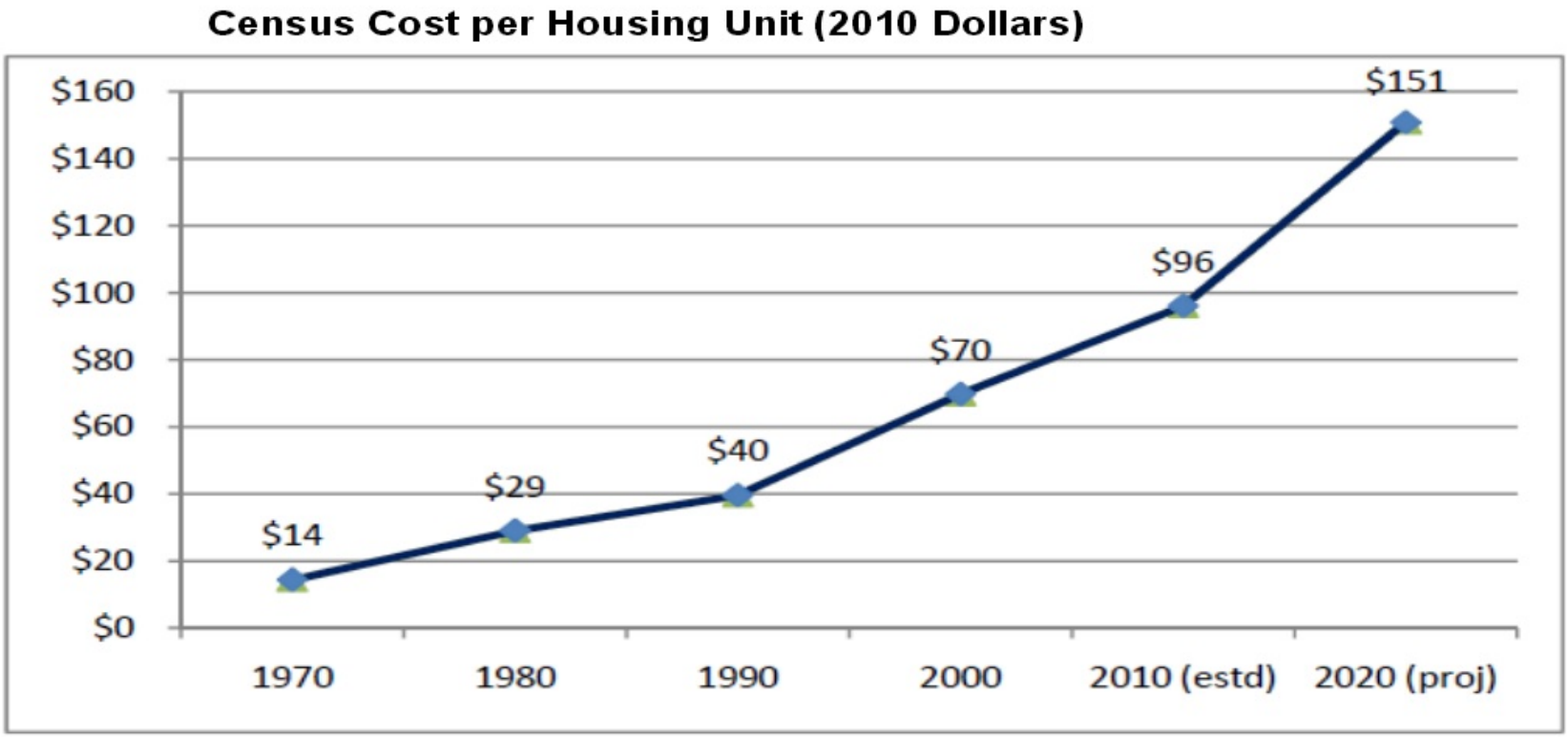
Numbers depend strongly on assumptions (to be discussed).

The U.S. Census

U.S. conducts census every 10 years to count the population.

Article 1, Section 2 of the U.S. Constitution:
"Representatives and direct Taxes shall be apportioned among the several States ... according to their respective Numbers ... **The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years.**"

Census Cost



Notes: (1) Chart reflects cost projections based on known FY 2010 savings of $1.87 billion as of October 4, 2010 (some additional adjustments may occur as cost information is finalized/verified); (2) Projected cost per housing unit for 2020 assumes no change in design and real cost growth from 1990-2000 and 2000-2010 averaged (57%); (3) Includes the costs for the MAF-TIGER Enhancement Program and American Community Survey.

From Presentation from James Christy, Regional Director, Los Angeles, U.S. Census Bureau, 2012 Annual California State Data Center Network Meeting, May 31, 2012

Census Costs per Housing Unit (2010 $)

2010	$97	(actual, corresponds to $13 billion total cost)
2020	$151	if no change in design from 2010 ($20 billion total)
	$70	recommended by NRC in 2010 ($9 billion total)
	$97	recommended by NRC in 2011 ($13 billion total; estimated "sticker price" $20 billion in 2020 dollars)

Census Error

Error in census counts arises from a variety of causes, including omissions, double counting and counting an individual in the wrong place.

2010 census estimated net undercount: -0.01%.

Black net undercount estimate: 2.1%.
White net undercount estimate: -0.8%.
Children aged 0-4 net undercount estimate: 4.6%.

Within a state, both undercount and overcount can occur.

State estimated net undercounts ranged from 1.3% (Vermont) to -1.4% (West Virginia).

Decision Question

A decision question for the Census Bureau (and the government more widely):

> **How much accuracy is needed for the 2020 state-level census counts?**

The market does not provide an answer because census data are a public good.

From the perspective of statistical decision theory, the answer depends on:

- the **value** of accuracy.
- the **cost** of accuracy.

Cost-Benefit Analysis

Cost-Benefit Analysis estimates and compares the costs and benefits of alternative programs.

Example: Higher cost census vs. Lower cost census

To assess the **benefit** from statistics, one needs to understand **data use**:

- how the statistic is used.
- what would happen if the statistic were not available or were available with different accuracy or data quality.
  - involves counterfactuals.

Three Categories of Data Use

Instrumental Use: use statistics in specific, direct ways.

Conceptual Use: use statistics for general enlightenment. Statistics affect actions indirectly and less specifically than in instrumental use. Includes uses of statistics for research, policy development, administration.

Symbolic Use: Window dressing. Use statistics to legitimize and support predetermined positions.

Conceptual uses may be very important but are difficult to identify, study and quantify.

We focus here on instrumental uses.

Instrumental Uses of the Census

Prominent instrumental uses of U.S. Census include:

- apportioning the House of Representatives.
- defining legislative districts and other governmental areas.
- distributing more than $400 billion in funds each year to state and local government for various programs.
- improving other statistics, e.g., forecasts and survey statistics (updated frames, control totals).

Error in 2020 census counts causes distortions in these uses.

Our research is a partial benefit analysis of the U.S. Census studying only apportionment of the House of Representatives and allocation of federal funds.

Absolute Error Loss Functions

For both apportionment and allocation of funds, we use the easily interpretable absolute loss function.

a_i^{true} = state i's apportionment (or allocation) if census conducted perfectly
a_i^{est} = state i's apportionment (or allocation) with census error

Apportionment and monetary allocations mostly divide up fixed totals, so one state's shortfall is another's gain.

Errors in state apportionment cause distortion in average size of congressional district within state.

We study the number of malapportioned House seats (or the total of misallocated funds): $\sum_i |a_i^{est} - a_i^{true}|$.

Social loss $= \lambda \times \sum_i |a_i^{est} - a_i^{true}|$.

Value of λ is a political question, not a statistical one.

Apportionment

Each state's number of representatives is roughly proportional to its census population number. 435 seats are awarded.

Specifically, the Method of Equal Proportions:

- awards the first 50 seats, one to each state
- iteratively awards each seat 51 to 435 to the state with the maximum value of

$$\frac{P_i}{\sqrt{n_i(n_i+1)}}$$

P_i = census population in state i
n_i = number of seats already awarded to state i

Allocation of Federal Funds

Blumerman and Vidal (2009) identified 140 federal grant and assistance programs that distributed $435.7 billion in FY 2007 funds at least in part due to decennial census data.

Example: Medical Assistance Program (Medicaid) distributed $203.5 billion in FY 2007.

Medicaid allocation to state = state medical expenditures $\times$ FMAP,
FMAP (Federal Medical Assistance Percentage) =

$$1 - 0.45(\text{State PCI}/\text{US PCI})^2,$$

subject to a floor of 0.50 and a ceiling of 0.83, or

$$FMAP = min\{max[1 - 0.45(I_i/P_i)^2, 0.50], 0.83\}$$

I_i = BEA State Personal Income / U.S. Personal Income
P_i = Census State Midyear Population / U.S. Midyear Population

Steps for Analysis

1. Specify profiles of census error under different census designs

2. For allocation of funds, select a probability sample of federal programs to study

3. For allocation of funds, consider how census error propagates to errors for statistics used in allocation formulas

4. Use Monte Carlo simulation to estimate errors in apportionment and allocation

5. Summarize results with absolute error loss functions

Profiles of Census Error

Our goal is to quantify the effects of alternative profiles of census error being considered by the Census Bureau.

- We do not yet know which error profiles the Census Bureau is considering.
- Here we consider a simple set of alternative profiles.
- We provide software to the Census Bureau to apply to profiles of their choosing.
- For error profiles we examine, state populations counts are:
  - independent
  - normally distributed
  - unbiased
  - have constant coefficients of variation (c.v.) across states (0.2%, 0.4%, 0.6%, 0.8% or 1.0%)
- Sensitivity analyses suggest how effects on allocations and apportionment change with alternative error profiles.

Sampling of Programs Distributing Federal Funds

Select 8 largest allocation programs with certainty

Program	FY 2007 Obligation ($ Billions)
Medicaid	203
Unemployment Insurance	36
Highways	34
SNAP	30
TANF	16
Pell Grants	14
Title I Grants to LEAs	13
Special Education Grants to States	11
8 Largest Programs	358
All 140 Programs	436

8 of 140 programs (6%) allocate 82% of the funds.

Sampling of Programs Distributing Federal Funds

Take stratified sample of remaining programs

Stratum h	N_h	n_h	Program	FY 2007 Oblgn ($ B)	Wtd Oblgn ($ B)	Tot Wtd Oblgn ($ B)
			Head Start	6.9	10.3	
1	3	2	CHIP	5.9	8.9	19.2
			WIC	5.5	16.7	
2	6	2	Child Care Mandatory & Matching	2.9	8.7	25.3
			Child Care & Development	2.0	12.3	
3	12	2	Social Services	1.7	10.2	22.5
			English Language Acquisition	0.6	4.9	
4	16	2	Spec Ed - Infants & Families	0.4	3.5	8.4
			Nonpoint Source Implementation	0.2	9.5	
5	95	2	Title V Deliquency Prevention	0.1	3.0	12.5
Total	132	10				88.0

Statistics Used for Federal Funding

Program	Midyear Pop. Est.	Model-based Pop. Est.	ACS-based Pop. Est.	CPS Unemp. Rates	CPI-Urban	Non-Census Stats Used	New Census Stats Not Used
Medicaid	Y					Y	
CHIP	Y					Y	
Child Care Mandatory & Matching	Y						
Child Care & Development	Y					Y	
Social Services	Y						
Spec Ed - Infants & Families	Y						
Title V Deliquency Prevention	Y						
Title I - LEAs	Y	Y				Y	
Spec Ed - States	Y		Y				
WIC		Y					
English Language Acquisition			Y				
Unemployment Insurance				Y			
SNAP					Y		
Pell Grants					Y	Y	
Head Start					Y		
Highways							Y
TANF							Y
Nonpoint Source Implementation							Y

Motifs for Modeling Effects of Statistical Error

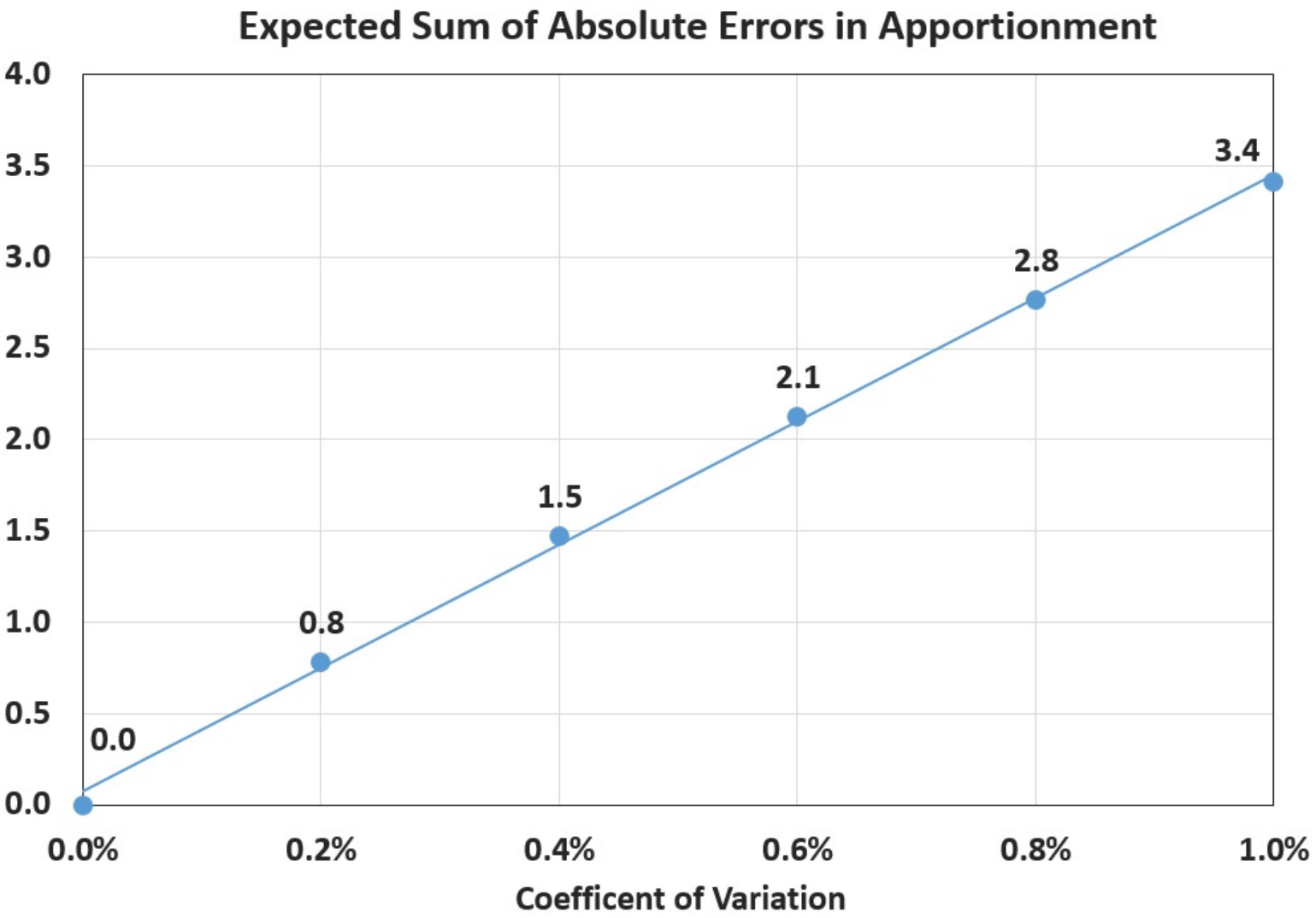
- Allocation formulas are complicated functions of census population counts and other statistics.
- We use various approximations to facilitate study of effects of census error.
- Except as noted, approximations overstate effects of census error.
- Examples of approximations
  - When allocation depends upon multiple census-based statistics, all assumed to have same relative error.
  - Error in non-census statistics ignored.
  - Effect of census error on CPI derived from differential net undercounts for owners and renters in 2010.
  - Effect of census error on unemployment rate derived from differential net undercount by age, race, sex in 2010.

Results: House Apportionment

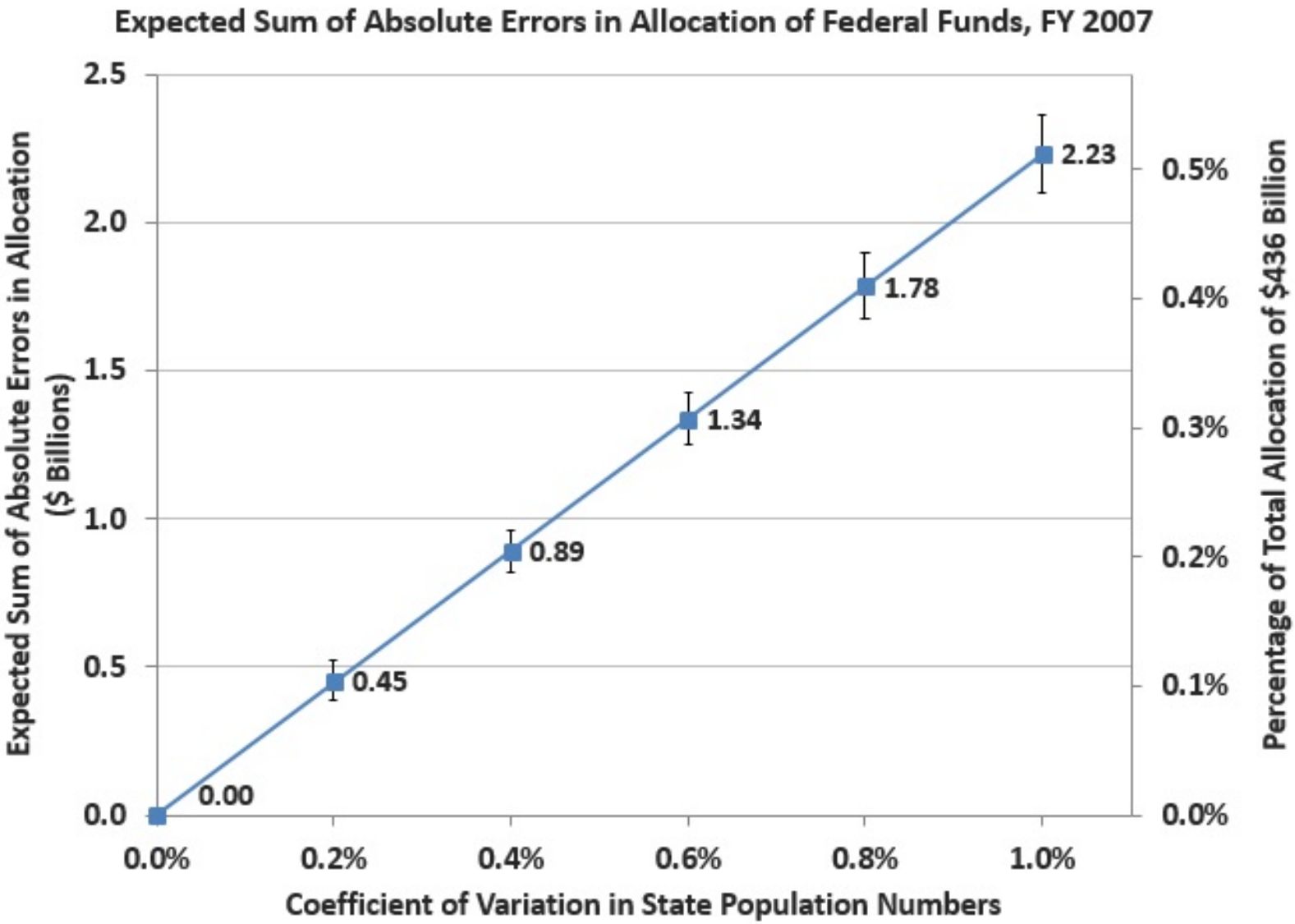
Probability Distributions of Malapportioned House Seats

C.V.	Prob. of ≥ 2 Seats Wrong	Prob. of ≥ 4 Seats Wrong	Prob. of ≥ 6 Seats Wrong	Prob. of ≥ 8 Seats Wrong	Expected Seats Wrong
0.2%	37%	2%	0%	0%	0.8
0.4%	63%	10%	1%	0%	1.5
0.6%	79%	24%	3%	0%	2.1
0.8%	88%	41%	8%	1%	2.8
1.0%	94%	57%	17%	2%	3.4

Results: House Apportionment



Results: Allocation of Federal Funds

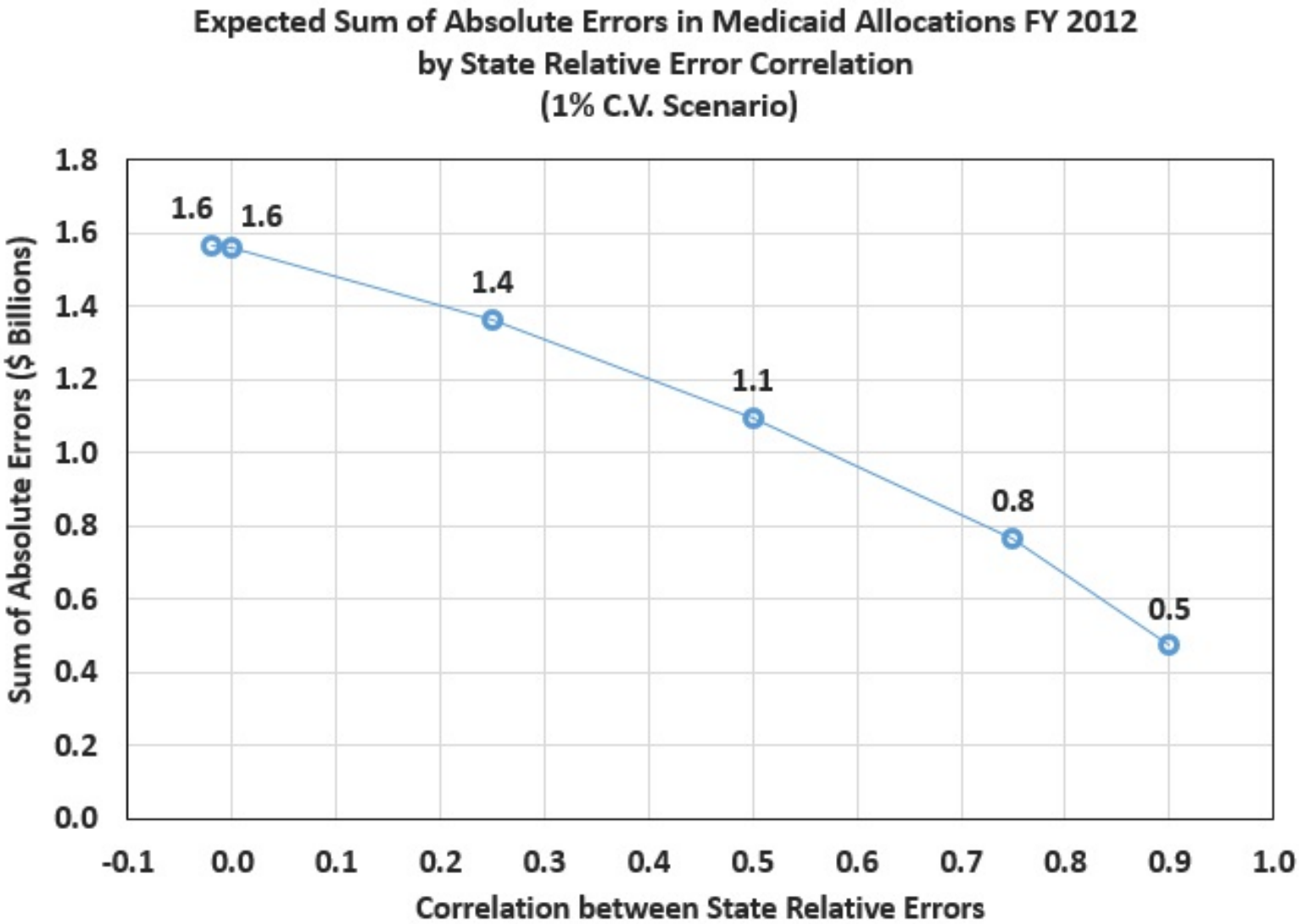


Federal Funds Results by Program

Expected Weighted Sum of Absolute Errors in FY 2007 Allocations Due to Census Error ($ Millions)

Program	0.2% CV	0.4% CV	0.6% CV	0.8% CV	1.0% CV
Medicaid	$240.7	$501.4	$764.8	$1,027.9	$1,290.2
Unemployment Insurance	$41.5	$82.9	$124.4	$165.9	$207.4
Title I - LEAs	$38.5	$77.8	$116.9	$156.0	$195.1
Child Care & Development	$50.3	$65.0	$84.7	$106.5	$129.3
WIC	$16.9	$33.8	$50.6	$67.5	$84.4
Social Services	$15.8	$31.6	$47.4	$63.2	$79.0
Spec Ed - States	$13.0	$22.7	$32.6	$42.7	$53.0
Child Care Mandatory & Matching	$7.8	$15.6	$23.4	$31.2	$39.0
English Language Acquisition	$7.3	$14.6	$22.0	$29.3	$36.6
CHIP	$5.6	$11.8	$18.2	$24.6	$31.0
SNAP	$5.6	$11.2	$16.8	$22.3	$27.9
Spec Ed - Infants & Families	$5.0	$9.9	$14.9	$19.8	$24.8
Title V Deliquency Prevention	$4.7	$9.5	$14.2	$19.0	$23.7
Pell Grants	$2.3	$4.5	$6.8	$9.0	$11.3
Total	$454.8	$892.3	$1,337.6	$1,784.8	$2,232.4
Total for 10 Years ($ Billions)	**$4.5**	**$8.9**	**$13.4**	**$17.8**	**$22.3**

Effects Highly Sensitive to Correlated Errors



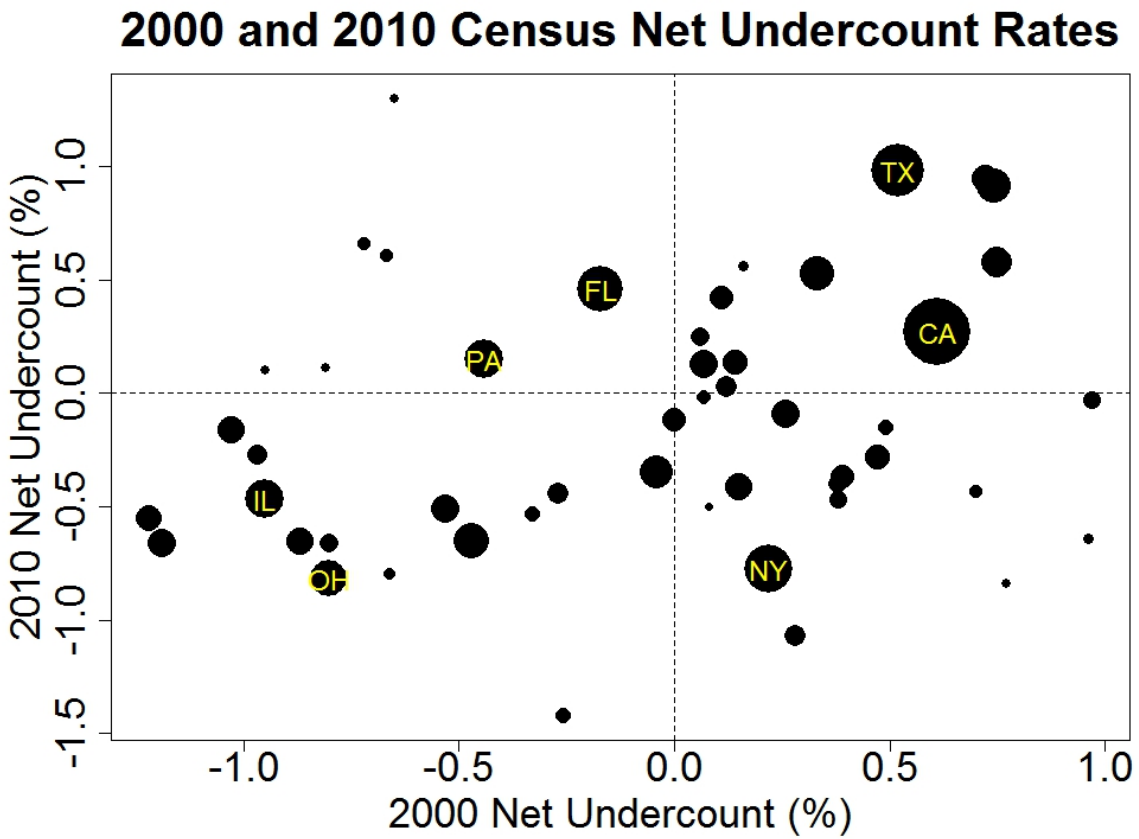
Expected Abs. Errors Decrease for Heavy-Tailed Dist.

A t dist. has heavier tails than the normal distribution.
A t dist. with 4 degrees of freedom has standard deviation $\sqrt{2}$.

Expected Sums of Absolute Errors (1% CV Scenario)

Distribution	Medicaid ($ Billions)	Apportionment
$N(0,1)$	1.6	3.3
$t_4/\sqrt{2}$	1.4	3.0
% Difference	-10.1%	-9.6%

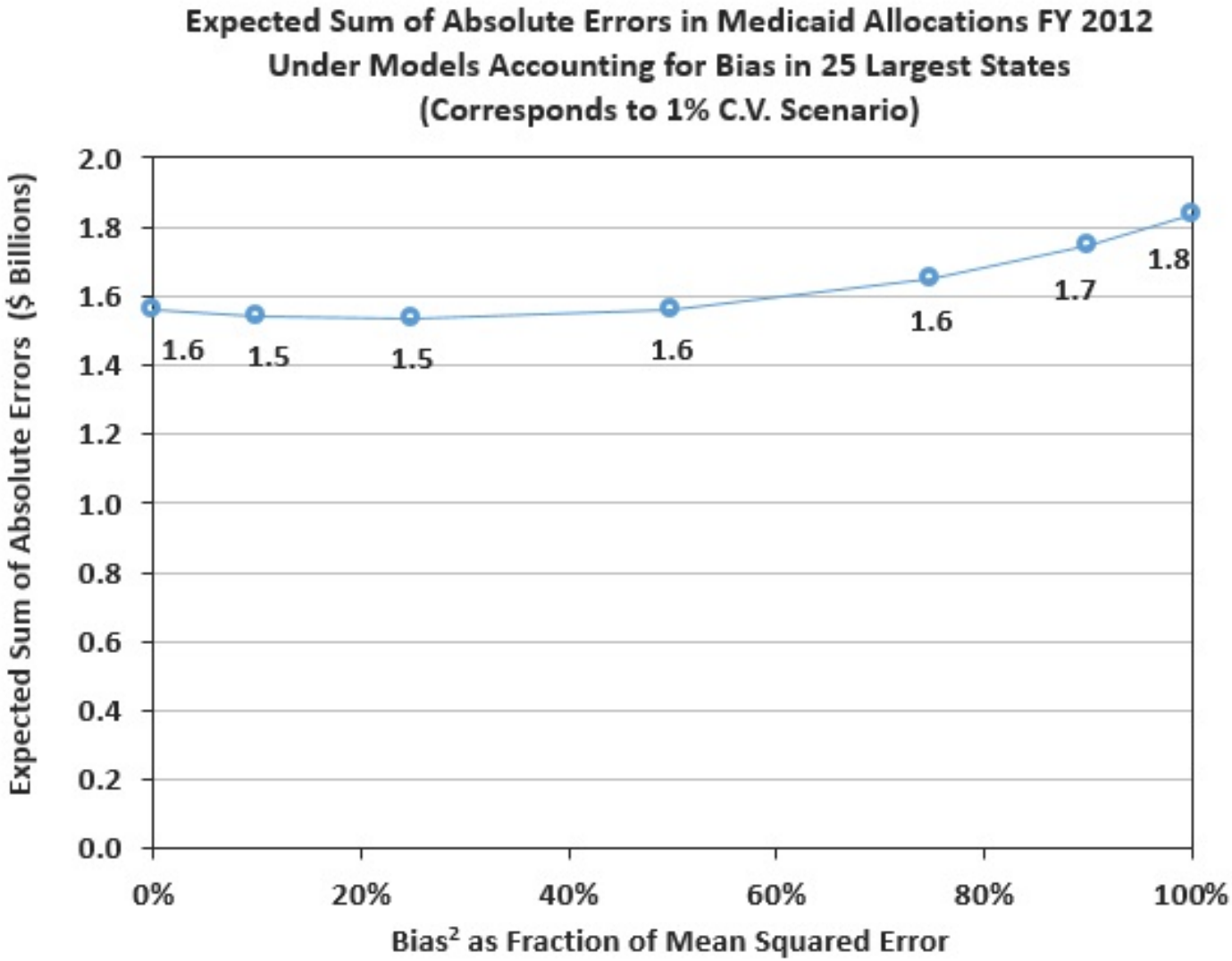
Evidence for Systematic Errors in Counts for Larger States



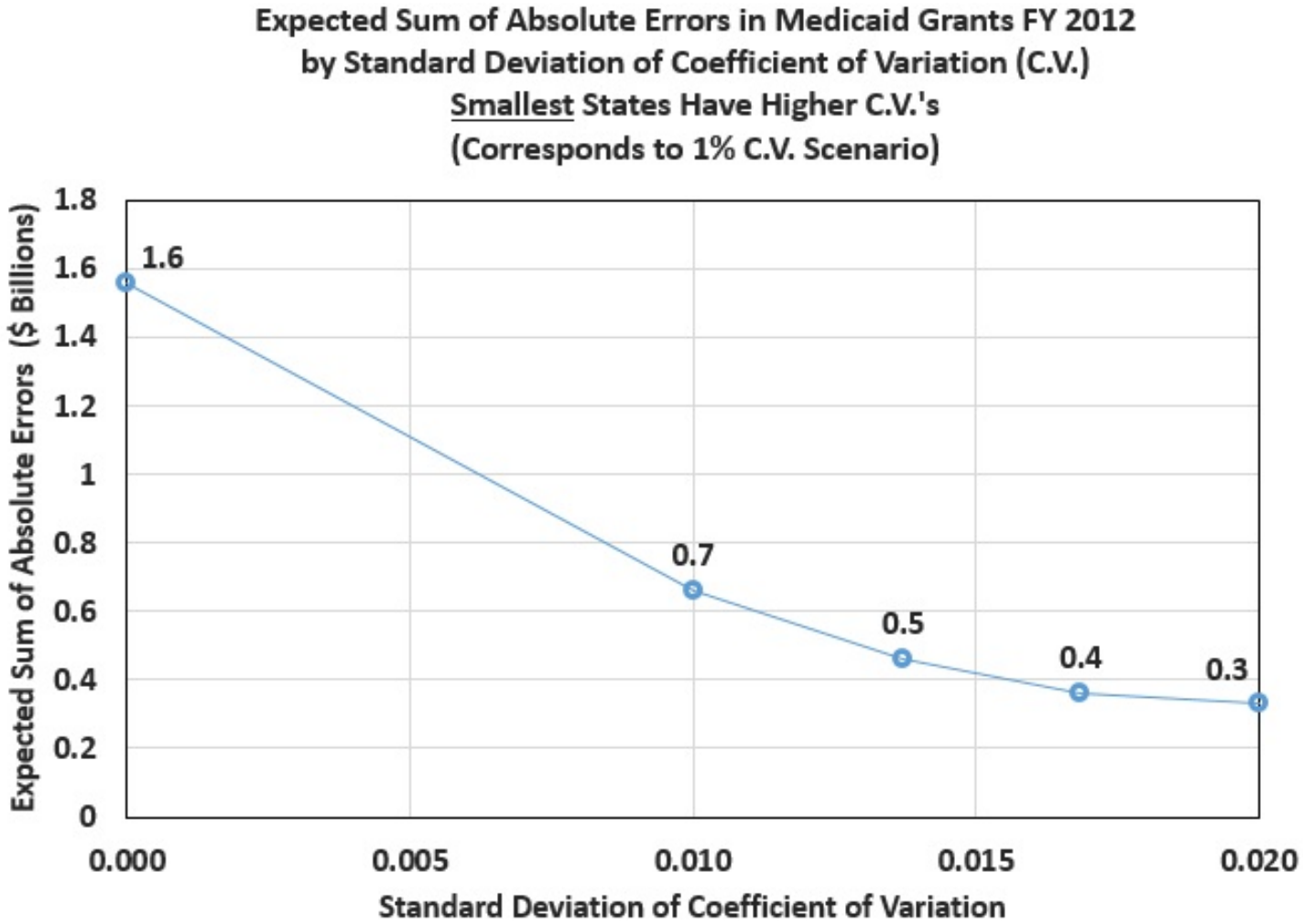
Unweighted Correlation: 0.19

Correlation Weighted by Population: 0.54

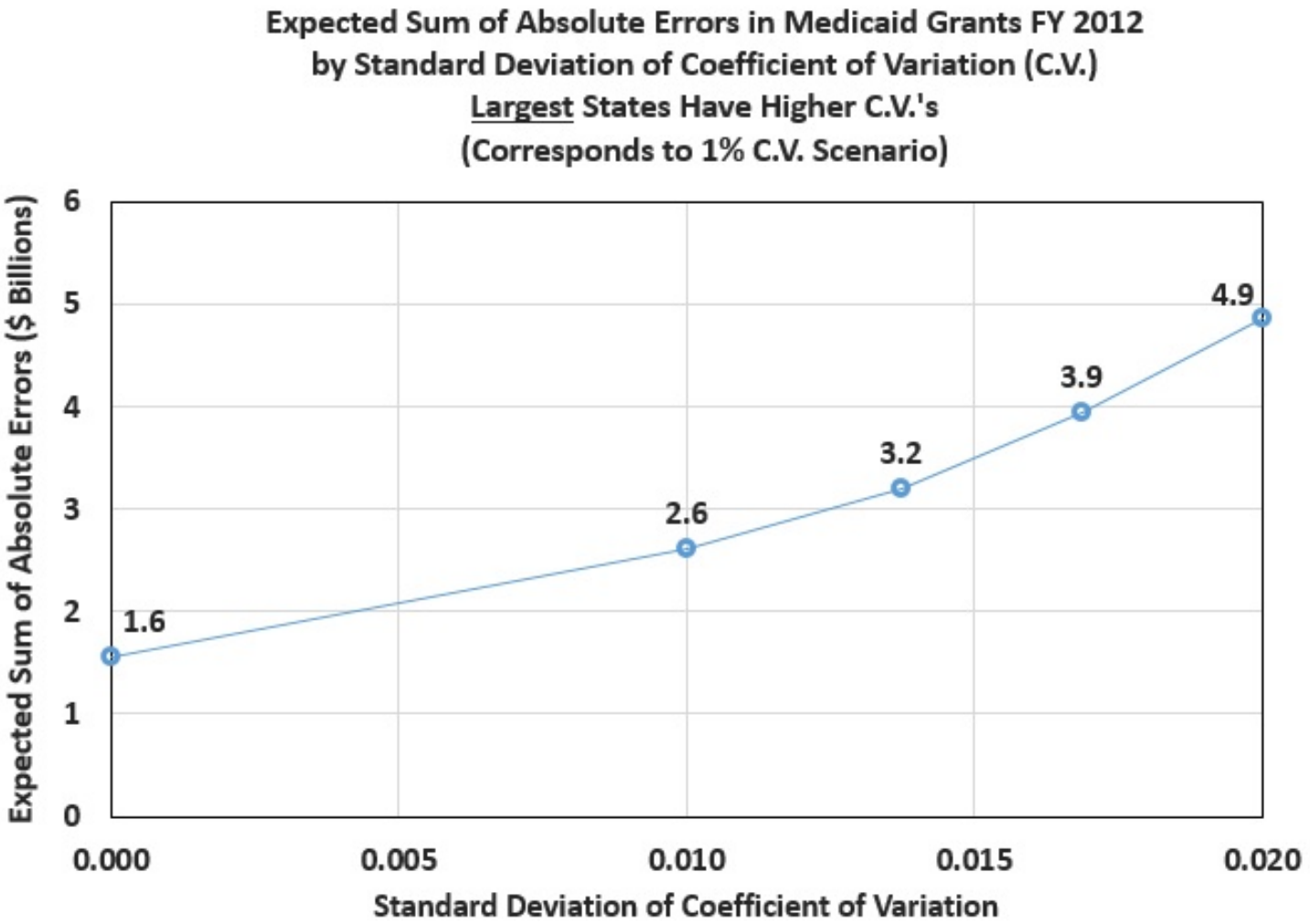
Effects Less Sensitive to Share of MSE Due to Bias



Effects Overstated if Small States Have Higher C.V.'s



Effects Understated if Large States Have Higher C.V.'s



Effects Usually Overstated if Non-Census Errors Ignored

Our analysis treats non-census statistics as fixed and without error.

Theory suggests that this approach usually leads to overstating the effect of census error on funds allocation.

When understatement occurs, the understatement is generally less than 1/3 as large as the potential overstatement.

Partial Cost-Benefit Analysis

PRELIMINARY - DO NOT CITE OR QUOTE

We have estimated effects of census error on apportionment and funds allocation.

Further specification of census error to be carried out with Census Bureau input.

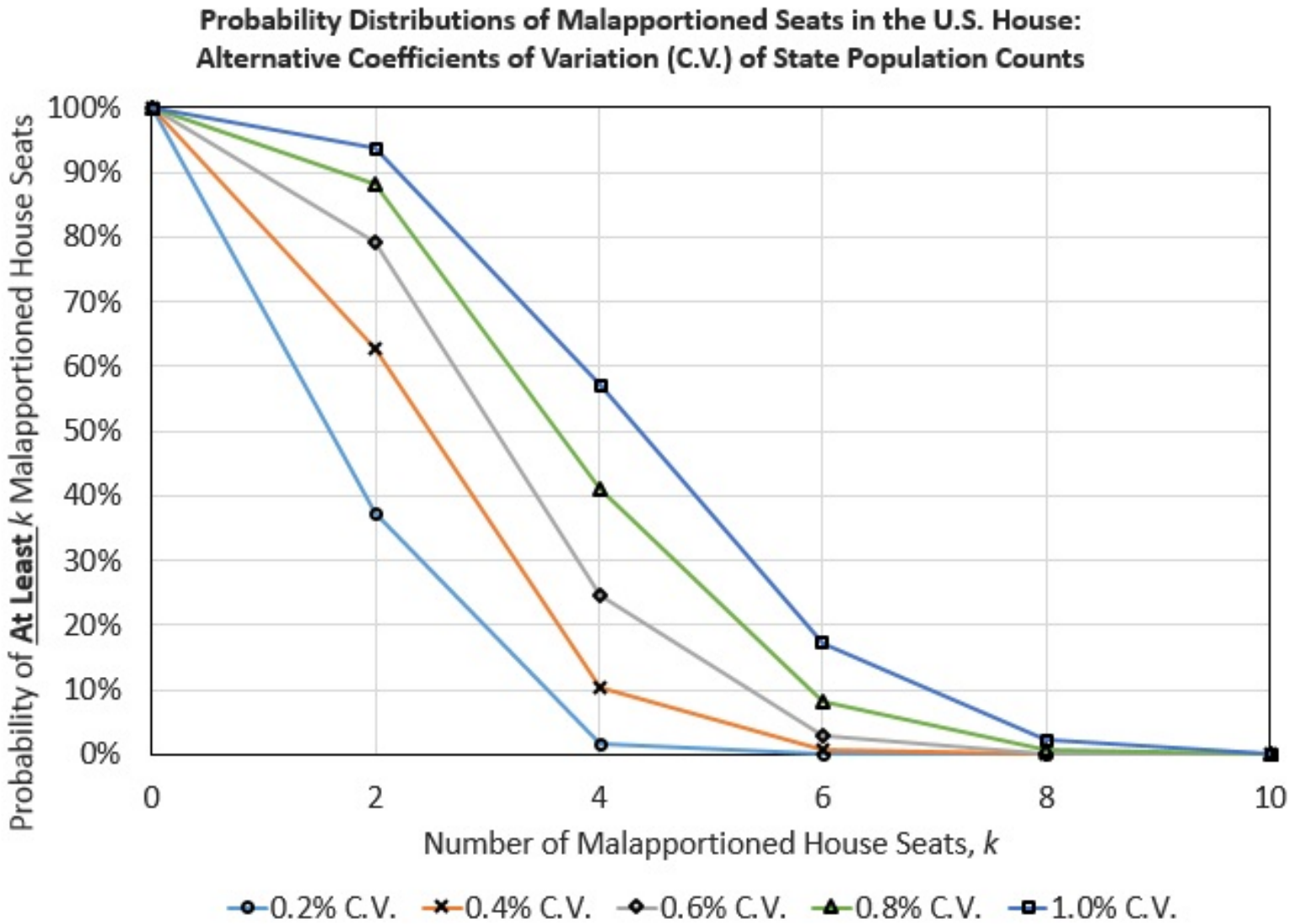
State C.V.	House Seats in Error	Errors in Fund Allocations ($ Bil.)	Census Costs
0.2%	0.8	4.5	high
0.4%	1.5	8.9	?
0.6%	2.1	13.4	?
0.8%	2.8	17.8	?
1.0%	3.4	22.3	low

What is the value to society of high accuracy in allocations?
How much is it worth spending to reduce expected errors in House apportionment by 2 seats?
How much is it worth spending to reduce expected errors in allocations by $10 billion?
These political questions require political input.

Improved apportionment / allocations only need justify a portion of census costs.
Other uses harder to measure, but not to be ignored.

Thank you!

Results: House Apportionment



Federal Funds Results by Program

Expected Sum of Absolute Errors in FY 2007 Allocations Due to Census Error (Percent of Total Funds)

Program	0.2% CV	0.4% CV	0.6% CV	0.8% CV	1.0% CV
Medicaid	0.12%	0.25%	0.38%	0.51%	0.63%
Unemployment Insurance	0.12%	0.23%	0.35%	0.46%	0.58%
Title I - LEAs	0.30%	0.61%	0.91%	1.22%	1.52%
Child Care & Development	0.41%	0.53%	0.69%	0.87%	1.05%
WIC	0.10%	0.20%	0.30%	0.41%	0.51%
Social Services	0.15%	0.31%	0.46%	0.62%	0.77%
Spec Ed - States	0.12%	0.21%	0.30%	0.40%	0.49%
Child Care Mandatory	0.09%	0.18%	0.27%	0.36%	0.45%
English Language Acquisition	0.15%	0.30%	0.44%	0.59%	0.74%
CHIP	0.06%	0.13%	0.20%	0.28%	0.35%
SNAP	0.02%	0.04%	0.06%	0.07%	0.09%
Spec Ed - Infants & Families	0.14%	0.28%	0.43%	0.57%	0.71%
Title V Deliquency Prevention	0.16%	0.31%	0.47%	0.62%	0.78%
Pell Grants	0.02%	0.03%	0.05%	0.07%	0.08%

Modeling Effects of Census Error on Funding: Unemployment

Under Extended Benefits and the temporary Emergency Unemployment Compensation, the federal government provided additional funding for Unemployment Insurance (UI) in the recent recession.

Extra funding provided when a state's unemployment rate exceeds certain triggers.

We model how federal funding for UI would change based on census error affecting whether or not state unemployment rates exceed triggers.

We adjust estimates for the probability of another recession and relate estimates to state coefficients of variation.

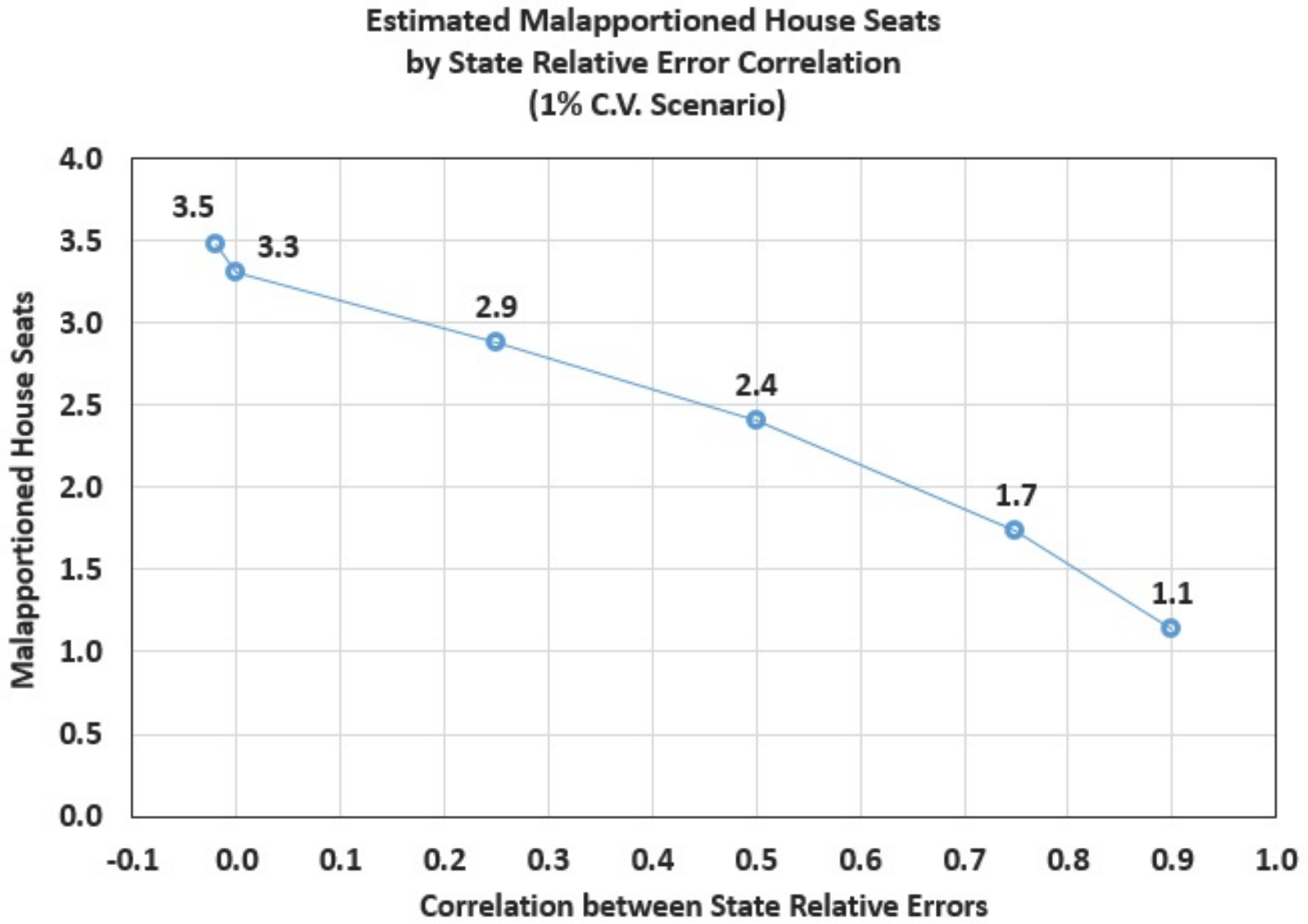
Modeling Effects of Census Error on Funding: Poverty Thresholds

Eligibility for and amount of grants for SNAP, Pell Grants and Head Start depend on poverty thresholds, which are revised annually based on CPI-U.

- For SNAP, we use income statistics and SNAP grant information to estimate the effect of census error on SNAP funding.
- For Pell Grants, we consider how total federal funding is affected by census error through CPI-U.
- We believe Head Start funding is not sensitive to poverty thresholds.

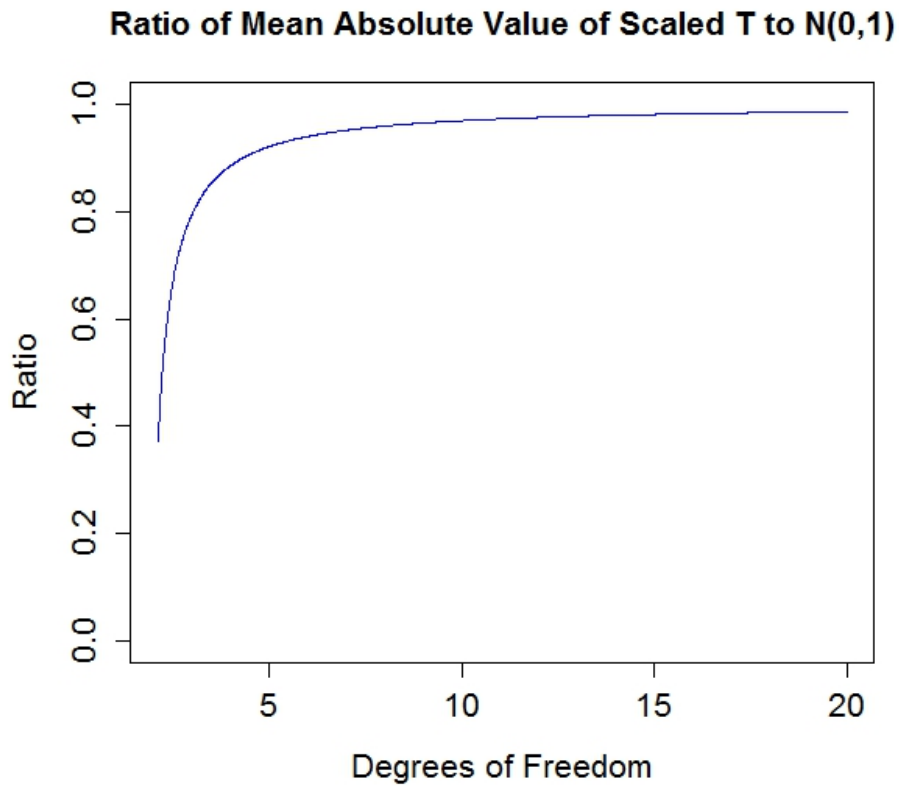
We relate estimates to state coefficients of variation.

Effects Highly Sensitive to Correlated Errors

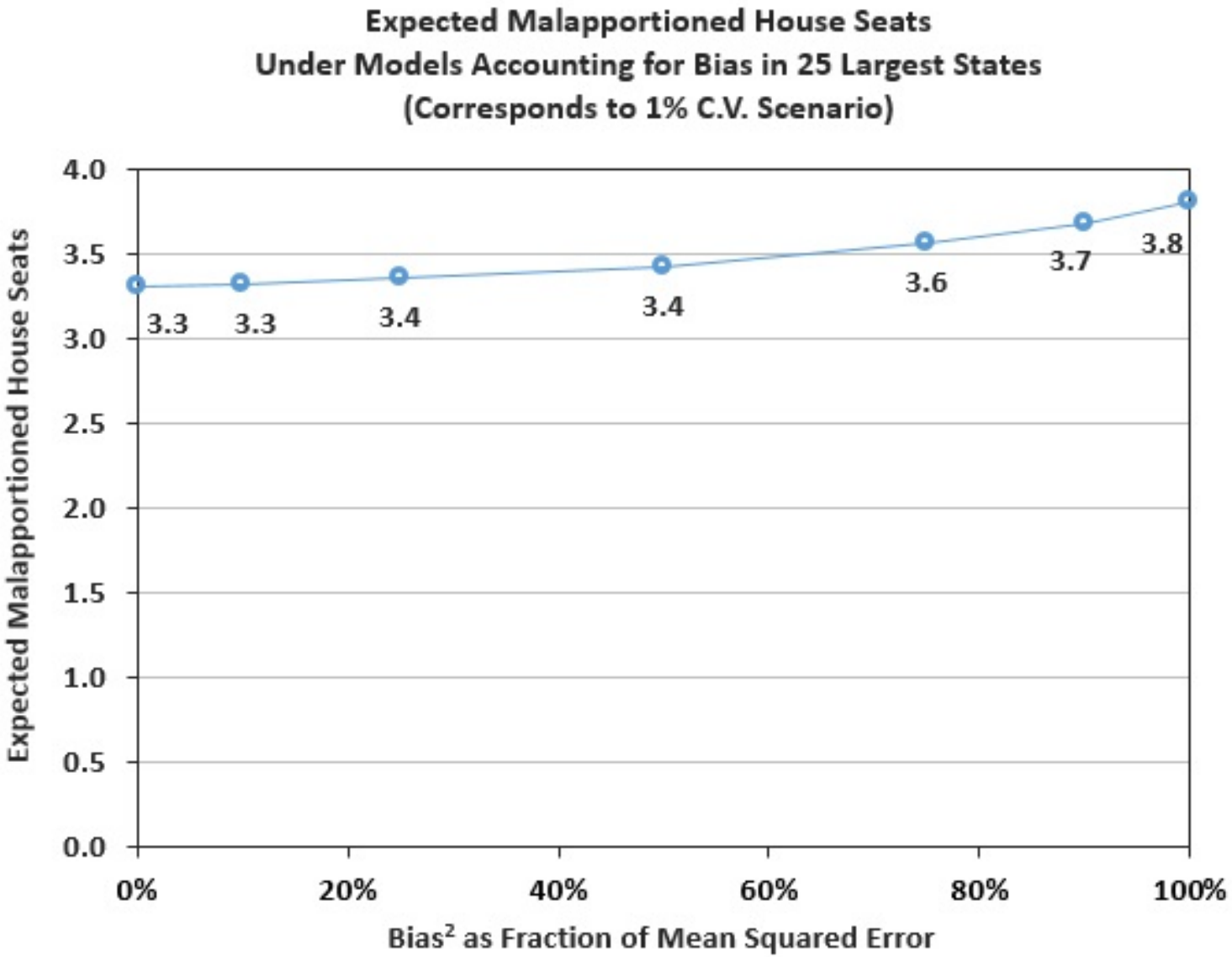


Expected Abs. Errors Decrease for Heavy-Tailed Dist.

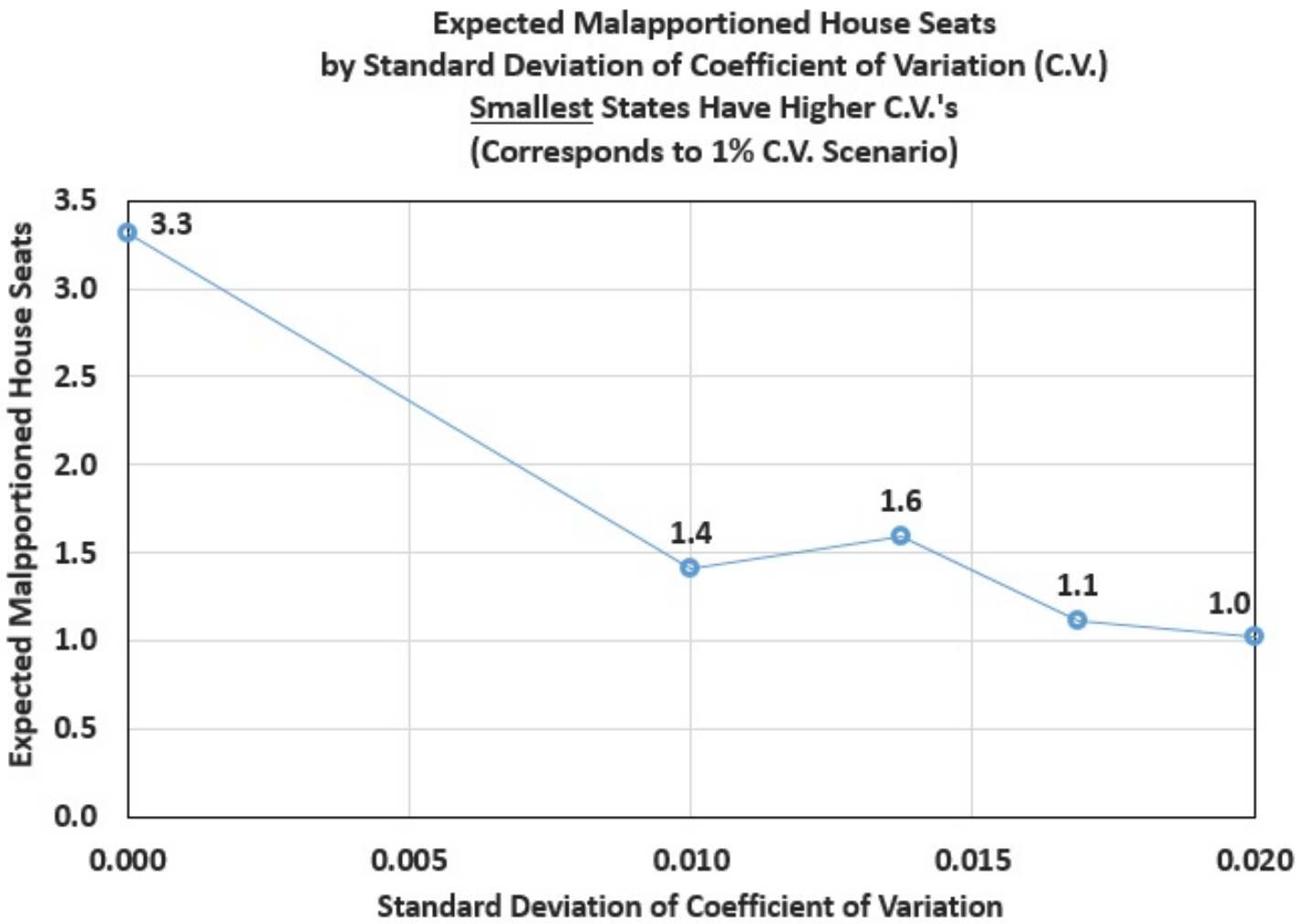
The mean absolute error of the ratio of a $t_\nu / \sqrt{\frac{\nu}{\nu-2}}$ to a $N(0,1)$ is $\frac{\Gamma(\frac{\nu+1}{2})\sqrt{2\nu-4}}{\Gamma(\frac{\nu}{2})(\nu-1)}$ for $\nu > 2$.



Effects Less Sensitive to Share of MSE Due to Bias



Effects Overstated if Small States Have Higher C.V.'s



Effects Understated if Large States Have Higher C.V.'s