State-Level Design Based Weights for National Surveys

Stephen Ash and Brian Shaffer

JPSM

March 9, 2018

What do we want?

- We have a national survey that produces great national-level estimates.
- Some states or areas within the national sample have large sample sizes.
- We'd like to produce state-level estimates.
- Is there low-hanging fruit (state estimates) within the national survey?

Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau.

Two Questions

Q1. What are all of the ways that we can use to improve the estimation of state-level statistics?

Q2. How do we determine that the resultant state-level estimates are reasonable?

Surveys

- Consumer Expenditures Surveys (CE)
- American Housing Survey (AHS)
- National Crime Victimization Survey (NCVS)

Low Hanging Fruit

- CE – CA, FL, NJ, NY, TX
- AHS – CA, FL, NY, PA, TX

  In 2015 – CO, OH

  In 2017(?) – AL, MD, MN, NV, OK, VA ?
- NCVS – 22 states

Q1. What are all of the ways that we can use to improve the estimation of state-level statistics?

1) Stratification of first-stage sample
2) Adding or pooling sample **
3) Noninterview adjustments
4) Ratio adjustments **
5) Variance estimation

2) Adding or pooling sample

- Have state-sample sizes that support state estimates as part of the original sample design
- Add sample on top of national sample
  - NCVS is “boosting” the sample
- Combine multiple years of survey
  - NCVS is combining 5 years sample
- Combine surveys
  - AHS is combing the National and Metro samples

4) Ratio Adjustments

- First-stage and overall ratio adjustments
- Overall – Last step of weighting
- Both
  - Reduce the variance of resultant estimates
  - Make estimates consistent with known totals

Q2. How do we determine that the resultant state-level estimates are reasonable?

- We have done everything to improve our state estimates.
- Sometimes it’s not enough
- What is enough?

Q2. How do we determine that the resultant state-level estimates are reasonable?

- Surveys with specific reliability goals.
  - Have specific statistic of interest
  - Examples
    - NCVS – victimization rates
    - CPS – unemployment rates
- General Purpose Surveys
  - Are used to produce a variety of estimates
  - AHS – all facets of housing

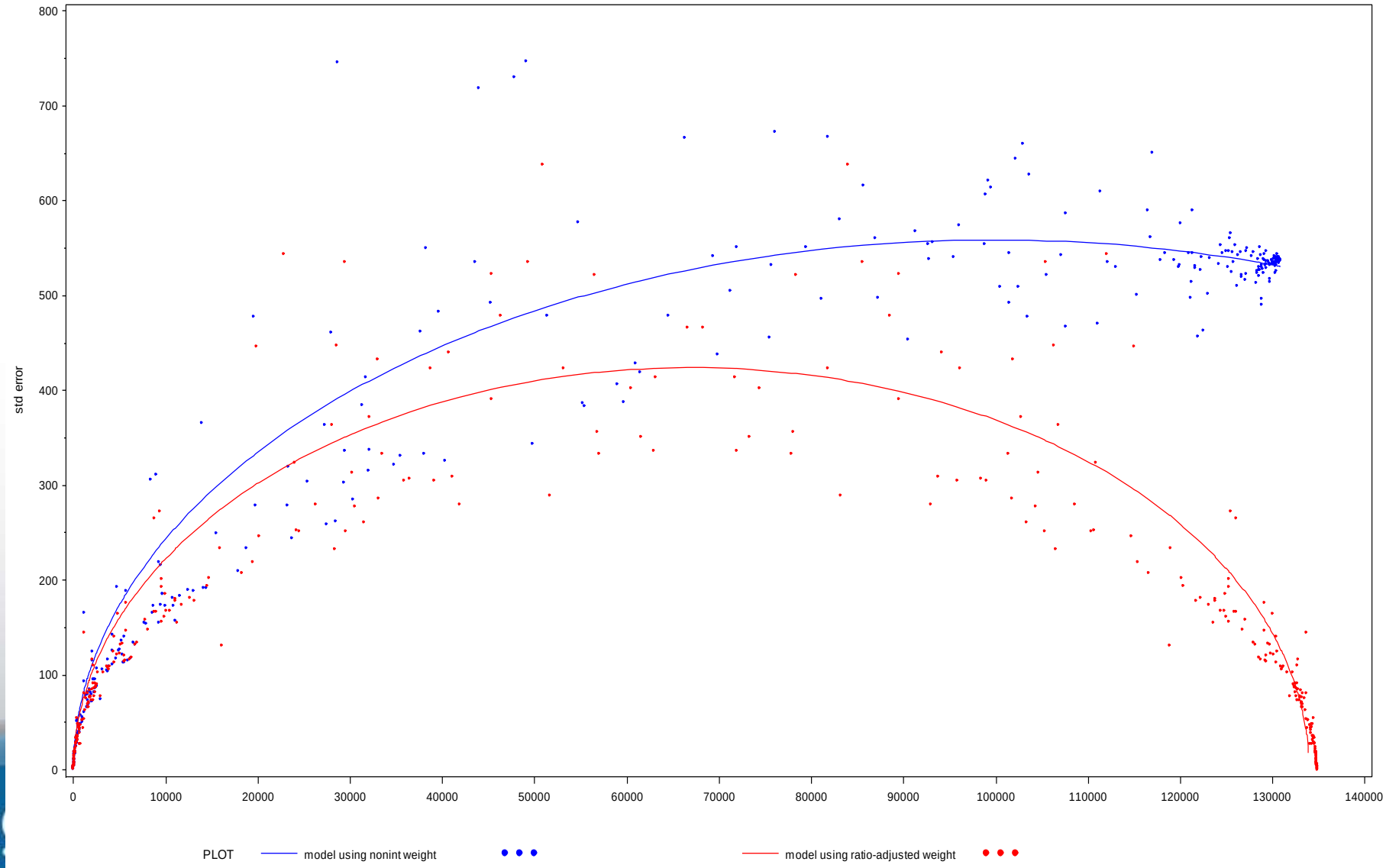
Generalized Variance Function (GVF)

- Can be used to summarize the variability of a general set of estimates
- For a general set of estimates, get

$$\left(\hat{N}_d, \hat{v}\left(\hat{N}_d\right)\right)$$

- Form the model $\hat{v}\left(\hat{N}_d\right) = a\hat{N}_a^2 + b\hat{N}_a$

GVF – w/ and w/o Ratio Adjustment



Evaluate CVs of many estimates

- CVs are usually used to judge the reliability of individual estimates
  - CV of 5% or less – excellent estimate
  - CV of 15% or less – a reasonable estimate
  - CV of 30% or more – no good at all
- How can we express the GVF in terms of CVs and sample sizes?

Method 1 – Transform the Model

- Transform the variance of the GVF to a CV.

$$\hat{v}\left(\hat{N}_d\right) \rightarrow c\hat{v}\left(\hat{N}_d\right) = \sqrt{\hat{v}\left(\hat{N}_d\right)} / \hat{N}_d$$

Transform the estimate to a sample size

$$\hat{N}_d \rightarrow \overline{\hat{n}}_d = \hat{N}_d / \overline{w}$$

where $\overline{w} = \sum_{k \in s} w_k / \sum_{k \in s} 1$

Method 2 – Transform Points then Model the Points

Transform the variance of each estimate.

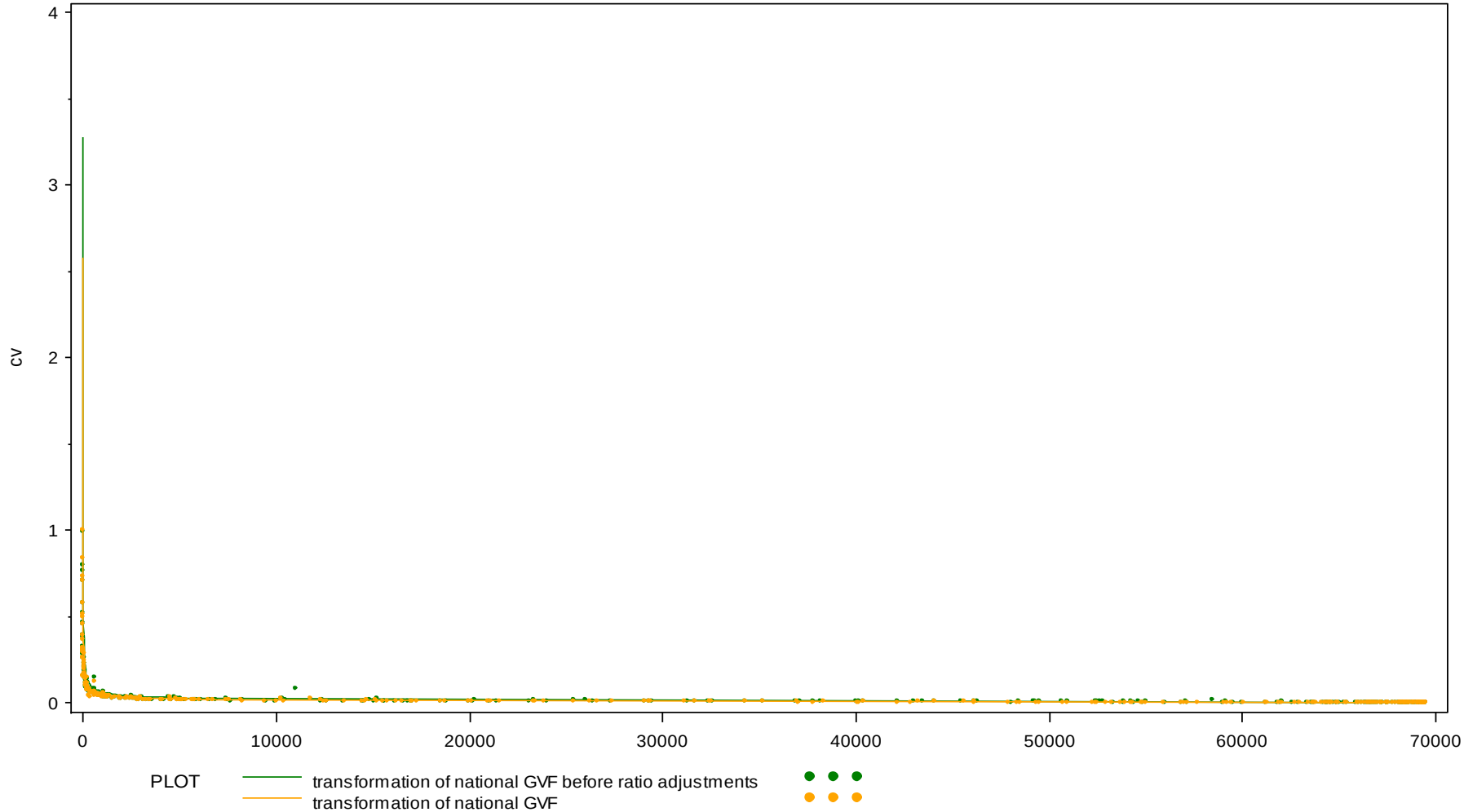
$$\hat{v}\left(\hat{N}_d\right) \rightarrow c\hat{v}\left(\hat{N}_d\right) = \sqrt{\hat{v}\left(\hat{N}_d\right)} / \hat{N}_d$$

Get the sample size of each estimate $\hat{n}_d = \sum_{k \in s_d} 1$

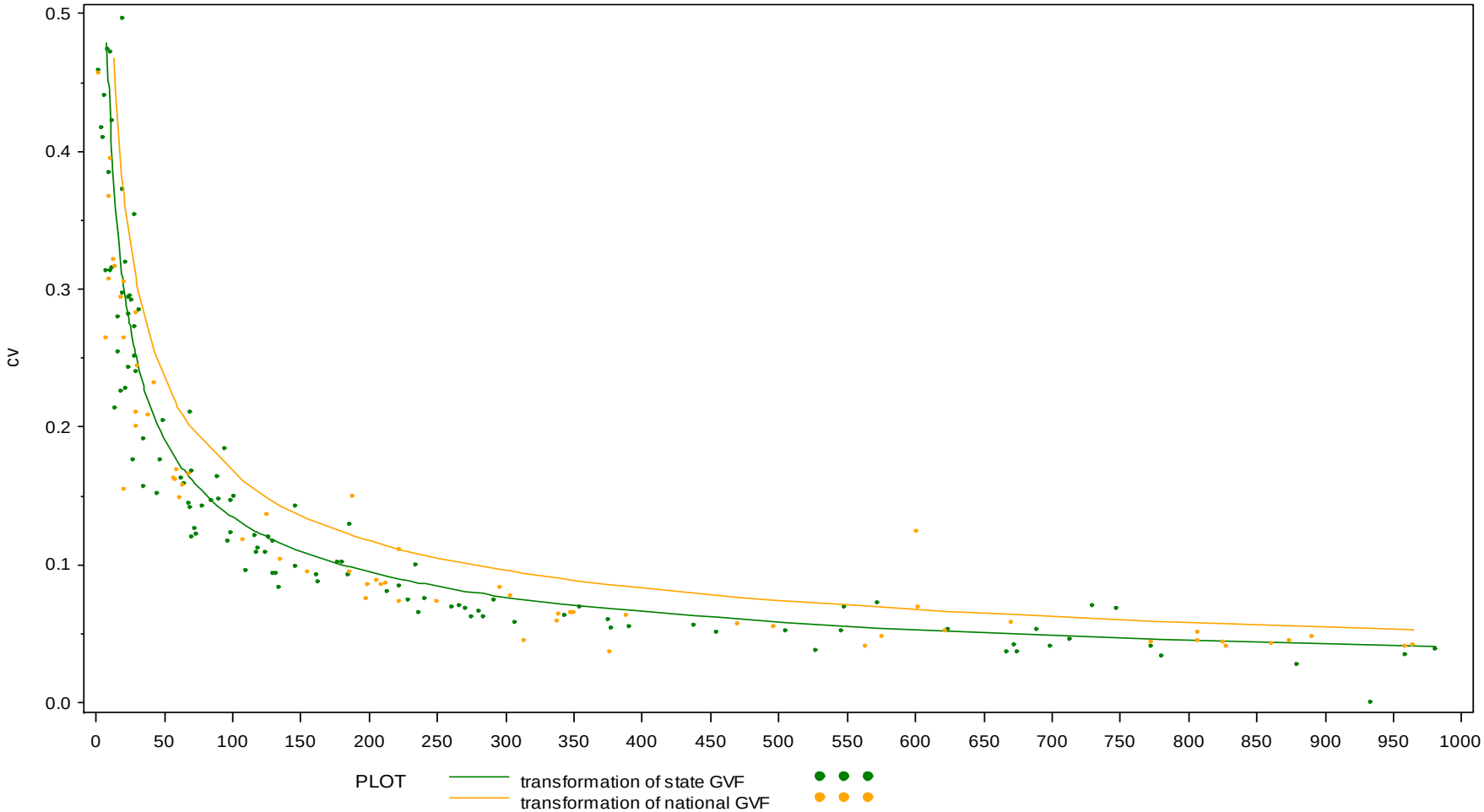
Model the points $\left(\hat{n}_d, c\hat{v}\left(\hat{N}_d\right)\right)$

as $\widehat{cv}\left(\hat{N}_d\right) = a + b\hat{n}_d^{-\frac{1}{2}} + c\hat{n}_d^{-1} + d\hat{n}_d^{-\frac{3}{2}}$

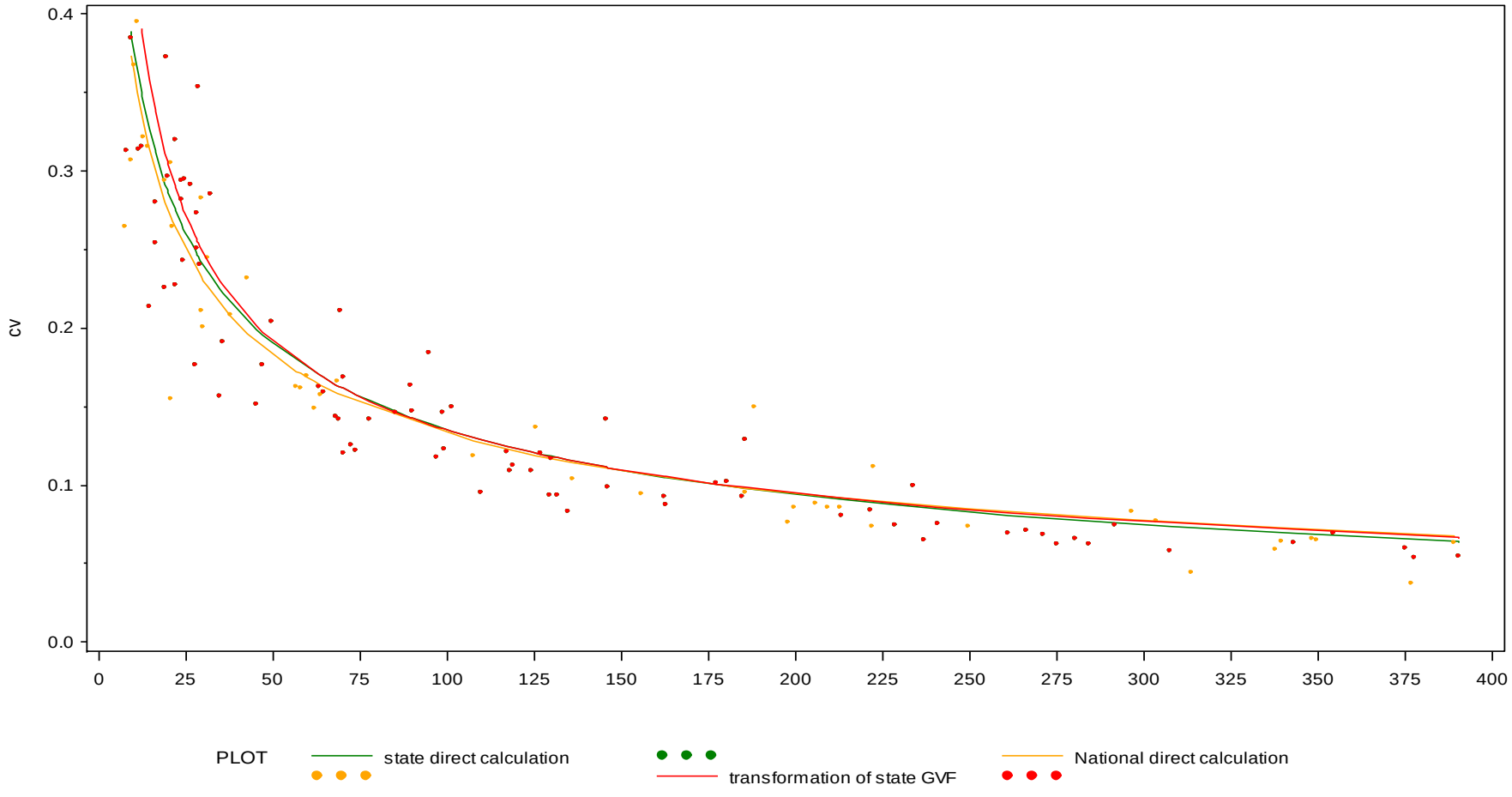
Method 1
CV for average sample size



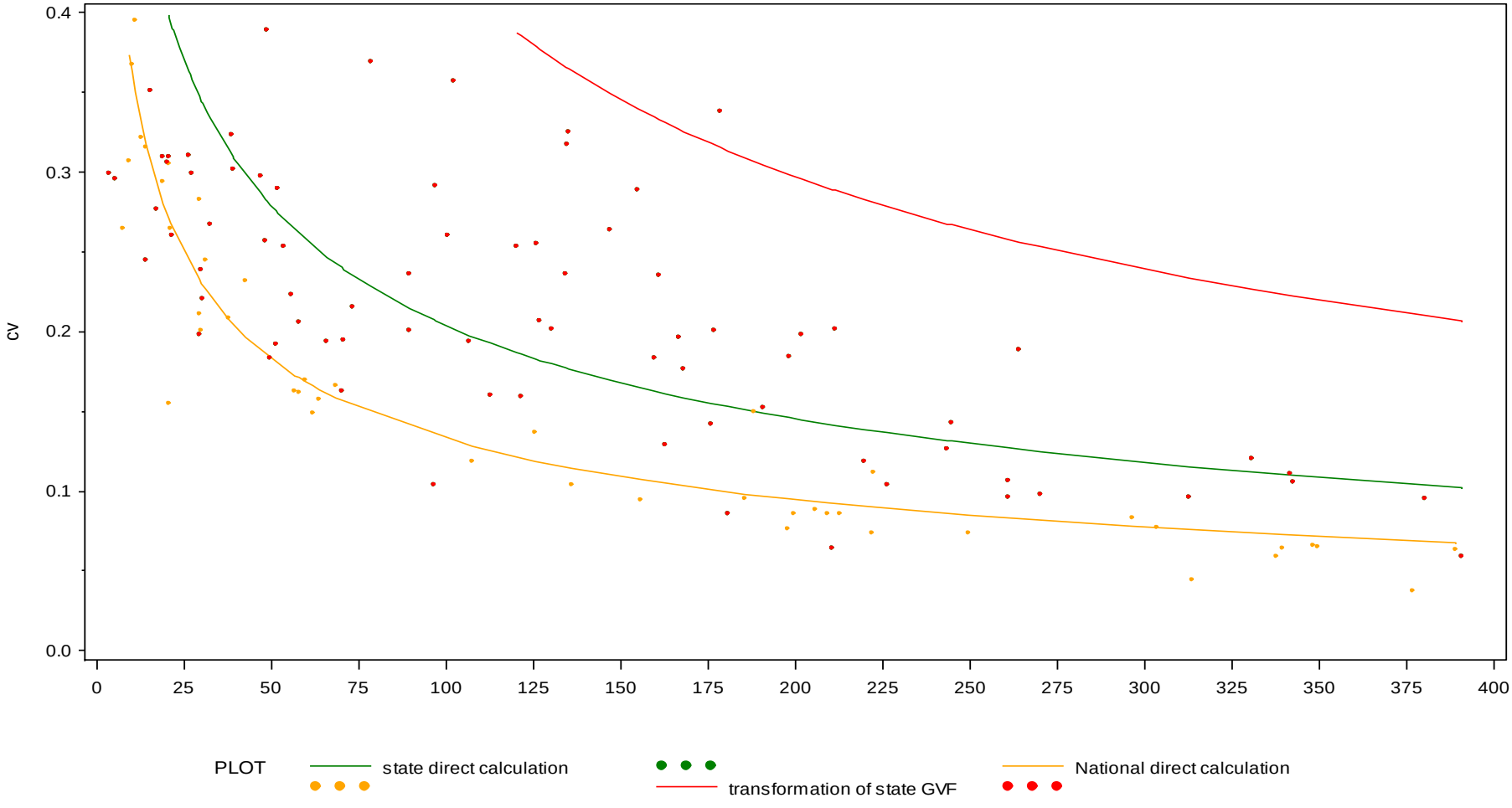
Method 1
CV for average sample size



Good State
CV for average sample size



Bad State
CV for average sample size



Average Sample Sizes by State and CV

State	Coefficient of Variation (CV)		
	< 30%	< 15%	< 5%
*California	17	67	574
*Texas	20	78	646
*Florida	30	116	880
*New York	13	53	408
*Pennsylvania	23	89	642
Illinois	30	160	574
New Jersey	13	49	344
Massachusetts	38	196	nd
Georgia	56	470	nd

Conclusions

- Creative methods are being applied to improve state-level estimates for many surveys.
- Methods are needed for evaluating the subnational estimates.
- Next steps: more work is needed on variance estimation.