Westat

50 years Improving lives through research

Progress in Developing Small Area Estimates of Crime Based on the National Crime Victimization Survey

Robert E. Fay (Westat), Mamadou S. Diallo (Westat), and Michael Planty (BJS)

Federal Committee on Statistical Methodology (FCSM)
Washington DC
November 4-6, 2013

Disclaimer

The results reflect the views of the authors and not necessarily those of Westat, the Bureau of Justice Statistics, or the Census Bureau.

We wish to thank Meagan Wilson and her Census Bureau colleagues for resource support enabling this research.

Outline

- Introduction
- Univariate SAE Models
- Multivariate Extension
- Application to NCVS

Introduction: The NCVS

- National Crime Victimization Survey (NCVS)
  - National household sample
  - Rotating panel survey – interviews every 6 months
  - Self-report for all persons ages 12 and over in a household for violent crime (rape, aggravated assault, simple assault, robbery)
  - Household respondent for property crime (burglary, motor vehicle theft, other theft)
  - National estimates for major domains annually
  - Confidentiality under Title 13 of the U.S. Census Bureau

Introduction: The NCVS (2)

- Crime: a local issue rather than national
  - High interest in geographic detail because crime is a local phenomenon
  - Local Police Statistics may underestimate level and nature of crime (e.g. Truman and Planty, 2012)
  - NCVS provides a vehicle to compare standardized measures of crime across geographic areas and to the nation as a whole
  - Strategy to protect confidentiality has been to release very little geographic detail (Only region is on the public file, state is not available)

Introduction: The NCVS (3)

- Bureau of Justice Statistics (BJS) has released limited set of subnational estimates (e.g. Lauritsen and Schaum, 2005)
- BJS has developed a small area estimation program that explores both direct and indirect estimation procedures (Cantor et al., 2010)
- Model based approach for estimating crime at state level presented in this talk
- Sub-state level estimates are also of interest

Models: Summary of previous work

- Case for independent variables from FBI – (Uniform Crime Report).

Aggregated crime	NCVS rate	Best UCR predictor
Violent (personal) crime	Rape/Sexual Assault	Forcible rape
	Robbery	Robbery
	Aggravated assault	Forcible rape
	Simple assault	Forcible rape
Property crime	Household burglary	Burglary
	Motor vehicle theft	Motor vehicle theft
	Theft	Larceny

- Variation by type of crime and high correlations over time (higher than 0.90)
- Review of Rao-Yu (1992, 1994) model and introduction of dynamic model as an alternative
- Fay, Planty, and Diallo (2013, JSM) summarizes the evidence and references previous work

Models: Small Area Estimation and Time series

- Sampling model: $y_{it} = \theta_{it} + e_{it}$,
- where
  - y_{it} is the direct estimator of crime rate
  - θ_{it} (crime rate to estimate) is a function of:
    - known auxiliary variables (x'_{it})
    - Area effect v_i
    - Time-area effect u_{it}
  - e_{it} is random sampling error for area i, time t, with:
    - $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{iT})' \sim N_T(0, \boldsymbol{\Sigma}_i)$,
    - does not require $\boldsymbol{\Sigma}_i$ to be diagonal.
    - $\boldsymbol{\Sigma}_i$ estimated from survey data and treated as known

Models: Rao-Yu (1992, 1994)

- Mixed model for population values, θ_{it},

$$\theta_{it} = x'_{it}\beta + v_i + u_{it}$$

where

x'_{it} is a row vector of known auxiliary variables,

β is a vector of fixed effects

v_i is a random effect for area i, $v_i \sim^{iid} N(0, \sigma_v^2)$

u_{it} is a random effect for area i, time t, with

$u_{it} = \rho u_{i,t-1} + \epsilon_{it}$, $|\rho| < 1$ and $\epsilon_{it} \sim^{iid} N(0, \sigma^2)$, stationary

Models: Rao-Yu (1992, 1994) (2)

- Rao-Yu's stationarity assumption leads to:
  - $\text{Var}(u_{it}) = \frac{\sigma^2}{1-\rho^2}$
  - Discontinuity at $\rho = 1$
- "Early" difficulties/concerns with Rao-Yu model application to NCVS:
  - Difficulties estimating ρ close to 1
  - Non-stationary times series
- Alternative model: *Dynamic Model*.

Models: Dynamic model (Fay, 2012)

- Modification to random effects of Rao-Yu:

$$\theta_{it} = x'_{it}\beta + \rho^{t-1} v_i^* + u_{it}^*$$

$v_i^* \sim^{iid} N(0, \sigma_{v*}^2)$: random area effect at $t = 1$

$$\begin{cases} u_{i1}^* = 0 \\ u_{it}^* = \rho u_{i,t-1}^* + \epsilon_{it}, & \text{for } t > 1 \end{cases}$$

- ρ is not constrained
- $\epsilon_{it} \sim^{iid} N(0, \sigma^2)$
- v_i^*, $\epsilon_{it,}$ and $\boldsymbol{e_i}$ are mutually independent

Multivariate extension

- Interest in both crimes by type and in aggregate:

  Property crime = burglary + auto theft + (other) theft

  Violent crime (type) = (rape + aggravated assault) + simple assault + robbery

  Violent crime (perpetrator) = intimate partner violence + crime by strangers + crime by all others

Multivariate extension (2)

- Modeling choices:
  - common ρ across crime components involved in one multivariate model
  - common correlation matrix $\boldsymbol{R}$ (more details on next slide)
- Advantage:
  - Potential gains in efficiency from multivariate approach, especially if random effects are highly correlated with each other but sampling correlations are low
  - Total crime = sum of crime components (not true for separate univariate models)

Multivariate extension (3)

- Suppose $\boldsymbol{\theta}_{it} = (\theta_{it1}, \theta_{it2} \ldots)'$, then

$$\theta_{itk} = x_{itk}^{'}\beta_k + \rho^{t-1}v_{ik}^{*} + u_{itk}^{*},\ k = crime\ component$$

with

$$\boldsymbol{v}_i^* = (v_{i1}^*, v_{i2}^* \ \ldots)' \sim^{iid} N\left(0, \Sigma_{v*} = (\boldsymbol{\sigma}_{v*}^2)'\mathbf{R}\,\boldsymbol{\sigma}_{v*}^2\right)$$

$$u_{i1k}^* = 0$$

$$u_{itk}^* = \rho u_{i,t-1,k}^* + \epsilon_{itk}, \quad \text{for } t > 1$$

- $\boldsymbol{\epsilon}_{it} = (\epsilon_{it1}, \epsilon_{it2} \ldots)' \sim^{iid} N\left(0, \Sigma = (\boldsymbol{\sigma}^2)'\mathbf{R}\,\boldsymbol{\sigma}^2\right)$
- $\boldsymbol{v}_i^*$ $\boldsymbol{\epsilon}_{it,}$ and sampling error $\boldsymbol{e_i}$ are mutually independent
- $\boldsymbol{\sigma}_{v*}^2 = (\sigma_{v*1}^2, \sigma_{v*2}^2, \ldots)'$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2 \ldots)'$

Multivariate extension (4)

- General treatment of EBLUP in Rao (2003) sufficient for Multivariate Dynamic Model:
  - REML/MLE parameter estimation
  - MSE estimation (REML)
- Also programmed multivariate Rao-Yu (MLE and REML)
- Authors implemented in R functions

Multivariate extension (5): Simulation setup

- Consistent with both dynamic and Rao-Yu model
- M=32 areas
- T=10 years
- ρ = .95
- 5,000 replicates

Two related variables, each with

- σ^2 = .08775
- σ_v^2 = 0
- uncorrelated sampling errors σ_e^2 = 1.
- rank(X) = 17 with time and area group effects

Correlation of random effects, $R_{12} = .6667$

Multivariate extension (6): Simulation results

MSE and % difference	Dynamic	Rao-Yu
Individual components		
Univariate model	.268	.265
Multivariate model	.249	.249
Improvement	6.8%	5.8%
Estimated sum		
Univariate model	.677	.675
Multivariate model	.655	.656
Improvement	3.3%	2.8%

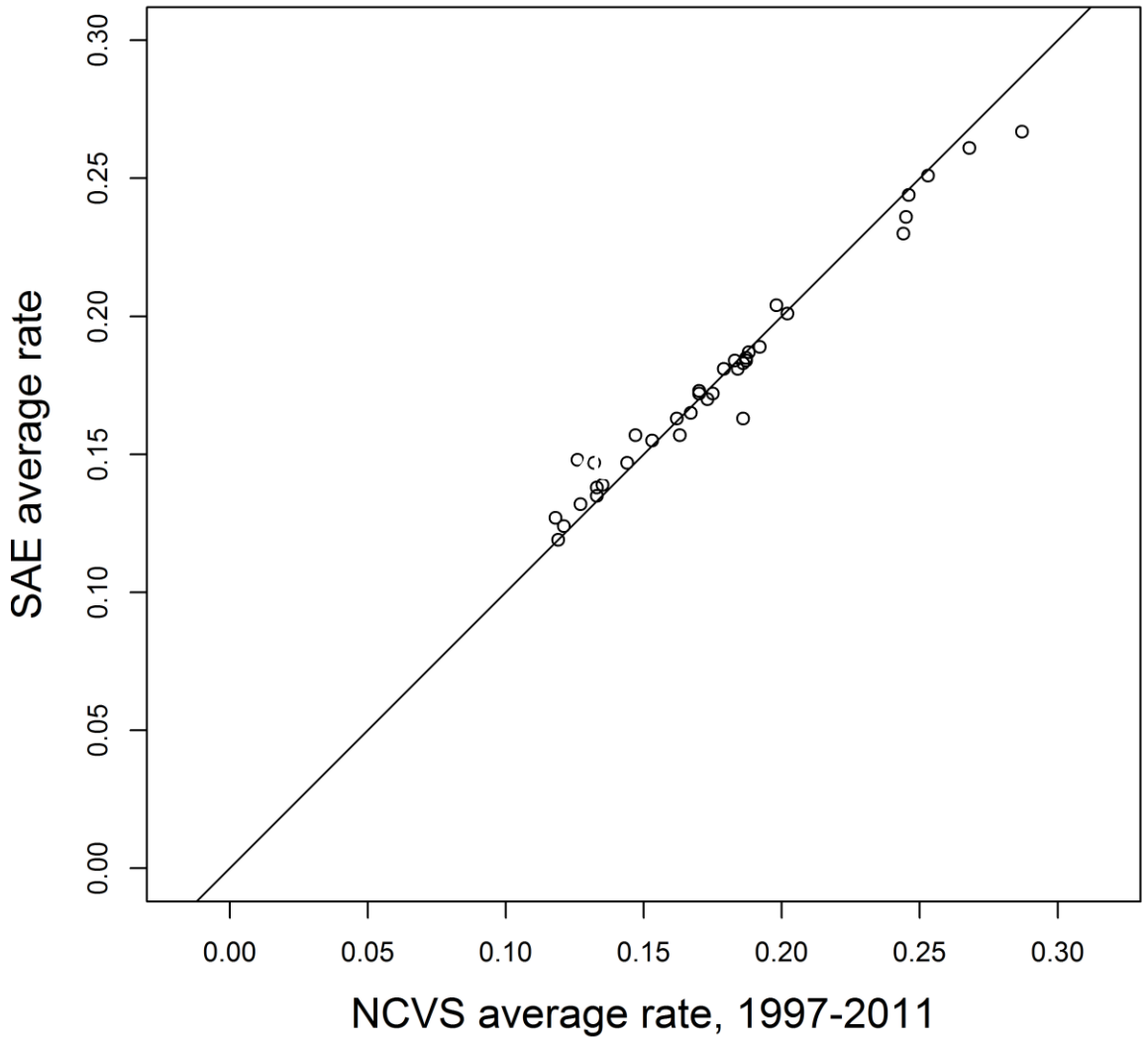
Application to NCVS

Aggregated crime	Crime components	Comments
Property crime	▪ burglary ▪ **all theft**	
All theft	▪ auto theft ▪ (other) theft	"all theft" proportionally adjusted to agree with "all theft" in "property crime" model.
Violent crime (perpetrator)	▪ strangers ▪ **non-strangers**	
Non-strangers	▪ intimate partner violence ▪ all other non-strangers	"non-strangers" proportionally adjusted to agree with "non-strangers" in "violent crime" model
Violent crime (type)	▪ rape + aggregated assault ▪ simple assault ▪ Robbery	"violent Crime (type)" proportionally adjusted to agree "violent crime (perpetrator)"

Application to NCVS (2)

- Produced state-level estimates for 1997-2011
- Modeled on annual basis, but summarized to three-year average rates, 1997-1999, …, 2009-2011
- Benchmarked to ACS state estimates of population and household to produce estimates of total consistent with published NCVS totals. (Small adjustments.)

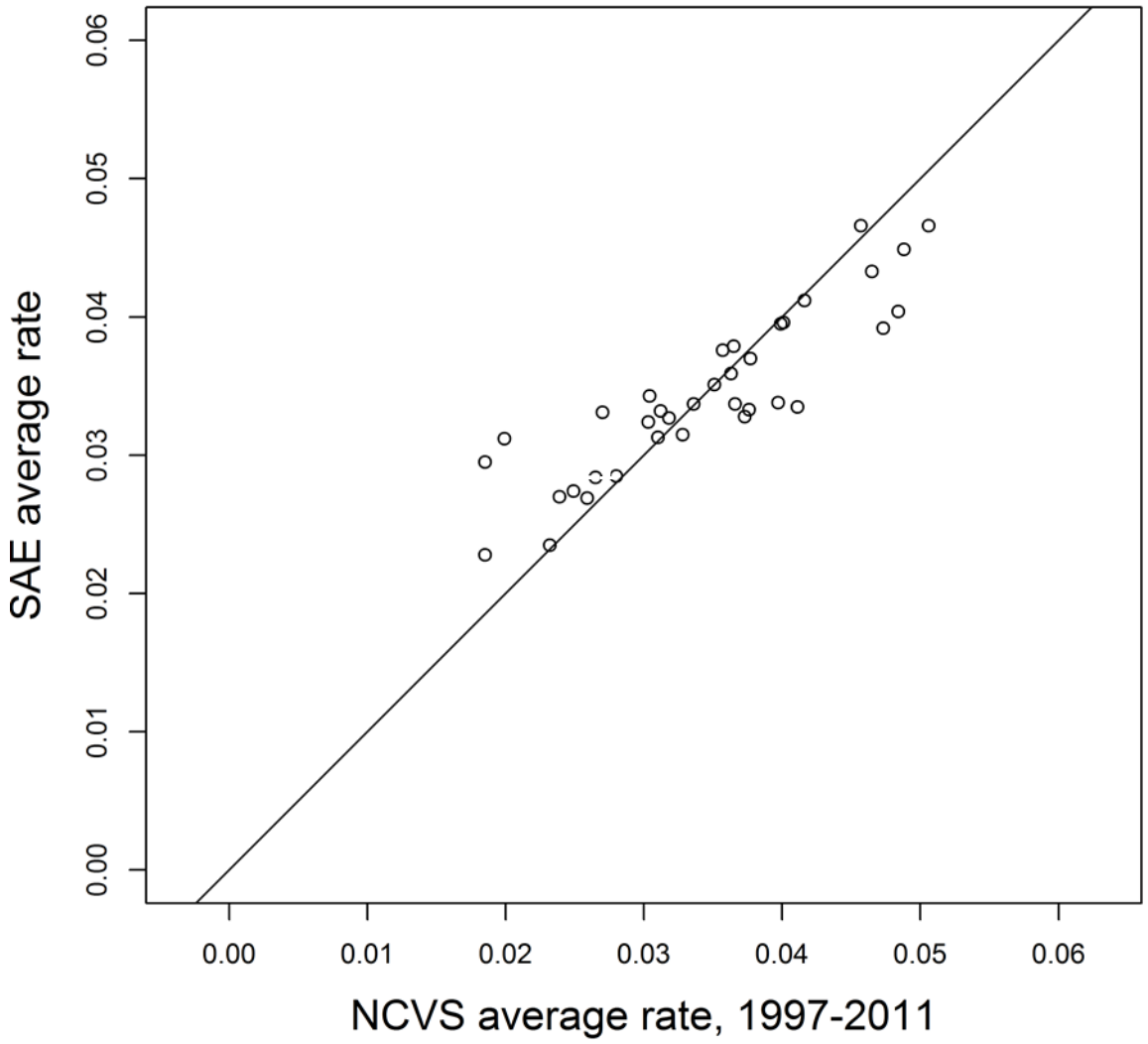
15-year averages of property crime by state



Westat

50 years Improving lives through research

15-year averages of violent crime by state



Westat

50 years

Improving lives through research

References

- Cantor, D., Krentzke, T., Stukel, D., and Rizzo, L. (2010), "NCVS Task 4 Report: Summary of Options Relating to Local Area Estimation," issued by Westat to the Bureau of Justice Statistics, May 19, 2010.

- Lauritsen, Janet L., Schaum, Robin L. "Crime and Victimization in the Three Largest Metropolitan Areas, 1980-2003". *Technical Report*. NCJ 208075, Washington, DC: United States Department of Justice, Bureau of Justice Statistics.

- Rao, J.N.K. (2003), *Small Area Estimation,* John Wiley & Sons, Hoboken, NJ.

- Rao, J.N.K. and Yu, M. (1992), "Small Area Estimation Combining Time Series and Cross-Sectional Data," *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 1-9.

- _____(1994), "Small Area Estimation by Combining Time Series and Cross-Sectional Data," *Canadian Journal of Statistics,* 22, 511-528.

- Truman, J.L. and Planty, M. (2012), "Criminal Victimization: 2011," NCJ 239437, BJS website, Oct. 2012.

Contact information

fay_b@westat.com

MamadouDiallo@westat.com

Michael.Planty@usdoj.gov

Westat

50 years

Improving lives through research