On the Ratio-Correlation Method of Subnational Population Estimation

David A. Swanson
University of California, Riverside

Jeff Tayman
University of California, San Diego

History

- Snow (1911) – regression for postcensal population estimates
- Crosetti and Smith (1954) – ratio-correlation
- Schmitt and Grier (1966) – difference-correlation
- Namboodiri and Lalu (1971) – average of simple regression models
- Swanson and Tedrow (1989) – rate-correlation
- Swanson and Beck (1994) – lagged ratio-correlation

Background

- Relate changes in symptomatic indicators to changes in population
  - births, deaths, school enrollment, employment, registered voters, tax returns, voter registration
  - total population, population 65+
- Geographic Hierarchy
  - Counties within states most common
  - Any nested geographic system suitable
- Requires independent population estimate for higher level of geography

Ratio-Correlation Model

- Most widely used form of regression methods

$$P_{i,t} = a_0 + \sum(b_j)^* S_{i,j,t} + \varepsilon_i$$

$$P_{i,t} = (P_{i,t} / \sum P_{i,t}) / (P_{i,t-z} / \sum P_{i,t-z})$$

$$S_{i,j,t} = (S_{i,t} / \sum S_{i,t})_j / (S_{i,t-z} / \sum S_{i,t-z})$$

- Ratio of shares (subarea to parent) between censuses
- Regression used to estimate a_o and b_j coefficients
- Solve equation using $(S_{i,t+k} / \sum S_{i,t+k})_j$
- Combines synthetic and censal ratio techniques in a regression context

Ratio-Correlation Model: Washington State Counties

$$P_{i,t} = 0.195 + (0.0933*\text{Voters}) + (0.3362*\text{Autos}) + (0.3980*\text{Enroll})$$
$$[p<.001] \quad [p=0.14] \quad [p<.001] \quad [p<.001]$$

where

$$P_{i,t} = (P_{i,2000}/\sum P_{i,2000}) / (P_{i,1990}/\sum P_{i,1990})$$

$$S_{i,1,t} = (\text{Voters}_{i,2000}/\sum \text{Voters}_{i,2000}) / (\text{Voters}_{i,1990}/\sum \text{Voters}_{i,1990})$$

$$S_{i,2,t} = (\text{Autos}_{i,2000}/\sum \text{Autos}_{i,2000}) / (\text{Autos}_{i,1990}/\sum \text{Autos}_{i,1990})$$

$$S_{i,3,t} = (\text{Enroll}_{i,2000}/\sum \text{Enroll}_{i,2000}) / (\text{Enroll}_{i,1990}/\sum \text{Enroll}_{i,1990})$$

$$R^2 = 0.794$$

$$\text{adj } R^2 = 0.776$$

Ratio-Correlation Model Shortcomings

- Inconsistency between calibration and postcensal estimate time periods
- Temporal instability of regression coefficients
- Multicollinearity
- Lag between symptomatic data and postcensal estimate dates
- Use of different symptomatic variables limits comparability of models and estimates
- Measurement error
- Spatial autocorrelation

Alternatives to Ratio-Correlation Model

- Rate-Correlation model
- Difference-Correlation model
- Combining symptomatic indicators and sample surveys
- Ridge regression
- Averaging estimate from simple regression models

Uncertainty in Population Estimates

- Almost all information on estimate error based on retrospective or post-hoc analysis
- Post hoc analysis not provide information directly relevant for current estimates
- Postcensal estimates have error, but typically only a single number is presented
- Direct measures of error are useful
  - Quickly see trustworthiness of estimates
  - Users are entitled to this assessment; a single number gives a false sense of security

Regression Models: Estimate Uncertainty

- Provide inferential tools to develop a direct quantification of uncertainty
  - Measures sampling variability
  - Measures lack of fit between estimate and population regression line
- Treat observations as coming from a super-population
- Treat upper and lower limits (confidence band) as an interval estimate for a parameter

66% Confidence Bands, 2010 Estimates: Selected Counties in Washington State

	Lower Limit	Point Estimate	Upper Limit	2010 Census	Outside Interval Lower	Outside Interval Upper
Adams	19,223	20,006	20,790	18,728	x	
Chelan	71,078	74,172	77,265	72,453		
Clark	429,504	445,660	461,816	425,363	x	
Franklin	72,086	75,116	78,146	78,163		x
Garfield	2,191	2,304	2,418	2,266		
Grant	89,121	92,596	96,071	89,120	x	
King	1,886,466	1,966,293	2,046,121	1,931,249		

% Outside All Counties **38%**

Conclusions

- Regression methods have strong advantages for making population estimates
- Long history of successful use
- Alternative approaches are available to overcome limitations
- Future areas of research and application
  - Spatial regression modeling
  - Uncertainty based on regression modeling

On the Ratio-Correlation Method of Subnational Population Estimation

David A. Swanson
University of California, Riverside

Jeff Tayman
University of California, San Diego