A 2016 View of 2020 Census Quality, Costs, Benefits

Bruce D. Spencer

Department of Statistics and Institute for Policy Research
Northwestern University

May 10, 2016, NSF-Census Research Network, Census Bureau

Research supported by NSF grant SES-1129475
Overview of Steps in Cost-Benefit Analysis of 2020 Census

1. Specify a census design
2. Forecast performance of census operational programs
3. Forecast census cost
4. Forecast accuracy for key state & national statistics
5. Forecast benefits from census accuracy for key uses
Census Uses

• Apportionment of House of Representatives
  o Required by U.S. Constitution, Article I, Section 2
  o Number of house seats proportional to state share of population

• Allocation of funds by formula
  o Census statistics used for $ trillions / decade
  o Allocations based on shares of population

• If all states have same relative errors, little if any effect on apportionment or allocations.
Census Uses

• Census population totals are important
  o population control totals for calibrating surveys
  o planning
  o market research
  o pluralistic society
Benefits Arise from Census Uses

• Societal benefits increase with accuracy of census stats

• Increased accuracy of state shares of population causes
  o more accurate & equitable apportionment of
    ▪ House of Representatives & Electoral College

• Increased accuracy of population totals for subgroups
  o improved estimates of rates
  o improved population controls for surveys
  o improved policy analysis and planning, e.g.,
    ▪ how many people need services
    ▪ how much human capital
Census Operational Programs

• Census planning and design involves making choices concerning alternative census operational programs, e.g.,
  o in-office address canvassing (as contrasted to field)
  o use of internet for self-response (versus paper)
  o use of admin. recs. & 3rd party data for non-response follow-up (NRFU)

• Focus of much (most?) 2020 Census research
  o increase efficiency of census operations
  o maintain or reduce the cost of the 2020 census relative to 2010
Program Output Parameters & Accuracy Parameters

• Outputs of census operational programs are characterized in part by *program output parameters*.
  
  o E.g., of housing units designated for NRFU,
    
    ▪ what fraction can be classified as vacant on basis of admin recs, w/o need for field follow-up?

• Also characterized by *program accuracy parameters*.
  
  o E.g., of the aforementioned housing units classified as vacant from admin recs,
    
    ▪ what fraction are occupied and hence erroneously classified?
Forecasting Program Output & Accuracy Parameters

• Cannot now know output & accuracy parameter values for 2020.

• Parameter values must be forecasted based on
  o prior census experiences (e.g., evaluations of 2010 census operations)
  o information about social & technological developments post 2010
  o tests and studies for the 2020 census
  o expert judgement.
Forecasting Program Output & Accuracy Parameters

• Probabilistic forecasting
  o Census Bureau researchers use probability distributions to represent beliefs about parameters
  o Probabilistic forecasting is an excellent approach
    ▪ enhances production of aggregate forecasts
    ▪ consistent treatment of uncertainty
Probabilistic Forecasts of Program Output Parameters

- Researchers are specifying **marginal probability distributions** for output parameters based on specification of
  - mean
  - 5\textsuperscript{th} percentile
  - 95\textsuperscript{th} percentile
  - shape of distribution (normal, triangular, uniform, etc.)

- It is important to consider joint distributions, not only marginal distributions.
Probabilistic Forecasts of Program Output Parameters

• Often, the average value of a product is unequal to the product of the averages.

*If I have a moral, it is this: Do not multiply lightly.*

How Many Program Output & Accuracy Parameters?

- Number of output parameters varies by program, e.g.,
  - 6 for part of NRFU based on admin recs (AR)
  - 24 for the part of NRFU based on field work w/o AR
  - # of correlations to specify is quadratic in # of output parameters.
  - For cost estimation in 2016, costs at the national level may suffice.
  - As 2020 nears, will Bureau need output parameters at finer geo levels,
    - e.g., Regional Office
How Many Program Output & Accuracy Parameters?

- To date, forecasting program accuracy less developed

- Need accuracy parameters by

  - demographic subgroup at national level
  - state level

- **Variability** in accuracy parameters across states is consequential

- With 50 or 51 states,

  - many parameters for marginal distributions
  - so many correlations among parameters that parameterized models of correlation will be needed.
Small Area Estimation Methods

- Small area estimation methods can help forecasting program accuracy parameters at the state level.

- Covariates could include
  - demographic variables
  - indicators of quality of AR / 3rd party data
  - historical info about census/survey quality in the state
  - others

- **Caveat:** synthetic estimation models & simple shrinkage methods prone to underestimate variation across states

- **Caveat:** expert judgements often reflect overconfidence.
Forecasting Cost & Accuracy of a Census Design

- Forecasting Cost

- Forecasting Accuracy – focus on
  - national numbers of people in different subgroups
    - forecasted from program accuracy parameter forecasts
  - state population sizes or shares
    - need to forecast 50 x 1 bias vector and a 50 x 50 covariance mx
    - too many parameters to forecast directly, so use indirect/small-area methods
Major Steps in Costs-Benefit Analysis of 2020 Census Design

1. Specify configuration of census operational programs
2. Prob. forecast of program output & accuracy params.
   a. Separately
   b. In combination (as a census design)
3. Probabilistic forecast of census cost
4. Probabilistic forecast of census accuracy, including
   c. Vector of biases $\mathbf{B}$ in state & national population
   d. Covariance matrix $\mathbf{V}$ of state & national population
   e. Distributional family (e.g., normal, Student’s $t$)
Major Steps in Costs-Benefit Analysis of 2020 Census Design

5. Forecast benefits from census accuracy
   f. House apportionment
   g. Allocation of federal funds
   h. Other uses
Benefits from More Accurate Apportionment of Congress and Allocation of Federal Funds

The following analysis and results are based on Zach Seeskin’s Northwestern Ph.D. thesis

- Models for Accuracy of State Population Shares
  - Relative biases specified for large states and small states separately
  - Coeffs. of variation (CV) specified separately for large & small states
  - Constant correlation across states
Benefits from More Accurate Apportionment of Congress and Allocation of Federal Funds

- Student’s $t$ distribution with various d.f.
  - Has smaller expected absolute error than normal, ceteris paribus
- Response surface model was fitted to predict mean absolute errors in apportionment and allocation as functions of these specifications.
Benefits from More Accurate Apportionment of Congress and Allocation of Federal Funds

- 4 Simple Models for Accuracy of State Population Shares
  - Base case: indep. normal, zero bias, constant CV
  - Correlated case: same as base case except
    - correlation of .5 for all pairs of states
  - Accurate small states case: same as base case except
    - smallest 25 states & DC have zero error
    - pop-weighted ave. of relative RMSE for remaining states equals CV for base case
Benefits from More Accurate Apportionment of Congress and Allocation of Federal Funds

- Differential bias case: independent normal with
  - equal RMSE
  - bias equal in magnitude to CV
  - bias for small states and large states have opposite signs
Effects of Census Accuracy on House Apportionment

- Used simulation to assess distributions of malapportioned seats
- Mean of normal prior distribution for state sizes based on state population forecasts for 2020,
- Variance based on average squared error for 2010 forecasts, with assumption of independence.
Effects of Census Accuracy on House Apportionment
Effects of Census Accuracy on House Apportionment

- The graph shows expected #s of malapportioned seats
- The 95\textsuperscript{th} percentile of # of malapportioned seats can be about 50% higher than the expected number.
  - E.g., for base case, 95\textsuperscript{th} percentile is about 20 versus mean of 13.3.
- Based on 2010 CCM, either 4 or 6 House seats were malapportioned following 2010 census
  - Corresponds to CV of 1% - 2% for 2020.
Effects of Census Accuracy on Funds Allocation

- Blumerman & Vidal (2009) identified fed. grant / direct assistance programs
  - 140 of them in FY 2007
  - alloc. $446.4 billion using Census pop. & income data
  - Stratified sample of 18 programs from the 140
    - 8 largest selected with certainty
      - accounted for 80% of the money allocated
      - effects of census inaccuracy were modeled and estimated for each selected progr
Effects of Census Accuracy on Funds Allocation
Quantifying Benefits in Dollars

• How much is it worth paying to reduce the number of malapportioned House seats by 2 or 10 or 20?
  ▪ How much is it worth paying to reduce the expected misallocations of funds over 10 years by $50 billion?
  ▪ These are political questions.
  ▪ Statisticians can help explain the questions and can provide technical assistance to help answer them.
  ▪ I do not view statistical expertise as being sufficient for providing answers, however.
Other Uses

- Examples
  - denominators for rates
  - control totals for a public and private population surveys in the U.S.
- Other uses important as well.
- Little known about benefits from improved accuracy for such uses.
Other Uses

- In general, little is known about direct uses of the census for research & planning.
- The limited number of variables collected by the census suggests that such uses are more indirect than direct, via denominators and survey controls.
Summary

- A high level overview of the steps in forecasting census costs and benefits was presented.
- Technical challenges in forecasting census program operation output parameters and accuracy parameters were identified.
Summary

- Relevant methodologies were suggested
  - to meet those challenges and
  - to formulate forecasts for error distributions for census population numbers.

- Once those error distributions are specified, the benefits arising from their use in House apportionment and allocation of federal funds can be quantified using existing analytical results.