

Can the Accuracy of Small Area Estimates Be Increased By Ignoring Census Counts?

Ken Hodges

The Nielsen logo is a white circle containing the word "nielsen" in a lowercase, serif font. Below the word are seven small dots arranged in a horizontal line.

nielsen
.....

Background

Can the accuracy of estimates be increased by ignoring census counts?

- Notion long-coveted by data companies
 - Mass-produce small area estimates
 - Usually start with census counts
 - Estimate change to current year
 - Multiple input data sources
 - Multiple geographic levels, with control totals
- A lot of work . . . A lot of time
 - Data companies would like something simpler

Background

- Data resources improving
 - Counts of residential addresses
 - Totals equaling or exceeding total US households
 - Coded to small areas
 - Demonstrated ability to capture population change
- Data companies wonder . . .
 - Can't we just use address totals as household estimates?
- Question no longer easily dismissed

Background

- Paper compares accuracy of household estimates
- Two types of estimates
 - 1. Census-based estimates
 - 2. Address-based estimates
 - Both vs. 2010 census household counts
- Which are more accurate?
- Also:
 - Explore potential for hybrid approaches
 - Using both census-based and address-based estimates

Evaluation Strategy

- Census-based estimates
 - Start with census
 - Estimate change since 2000 census
 - Nielsen 2010 HH estimates
- Address-based estimates
 - Address counts as stand-alone household estimates
 - Nielsen Master Address File (NMAF)
 - Addresses from multiple sources
 - Steps to remove duplicate addresses
 - Also those related to
 - Businesses, seasonal occupancies, college dorms

Evaluation Strategy

- Why not USPS addresses?
 - USPS provides only counts for carrier routes
 - Have to convert to census geography
 - Additional layer of error
- Evaluated NMAF for 2009 and 2010
 - 2009: Most recent when 2010 estimates produced
 - 2010: Accuracy if you could wait a year
- Basic error measure
 - Mean Absolute Percent Error (MAPE)
- Focus on block group level
- Focus on households (not population)

Evaluations After 2000

- Census-based estimates for 2000
- Address counts for 2000
- Block Group MAPEs
 - Census-based 13.3
 - USPS 28.6
 - Supplier A 28.1
 - Supplier B 31.4
 - Supplier C 27.9
 - Supplier D 25.7
- Census-based more accurate even in rapid change areas
- In 2000
 - Little case for widespread use of address counts as estimates

Evaluations After 2010

- Census-based estimates for 2010
- Address counts for 2009 and 2010
- Block Group MAPEs
 - Census-based 10.4 (down from 13.3)
 - USPS 2009 21.9
 - USPS 2010 20.9 (down from 28.6)
 - NMAF 2009 14.7
 - NMAF 2010 12.5 (down from 25.7)
- NMAF
 - Still higher than census-based
 - But getting close

Evaluations After 2010

Block Group MAPEs by Percent Change in Households 2000-2010

Pct Change	N of BGs	Cen-based	NMAF '09	NMAF '10
-50 or more	630	213.8	101.7	70.5
-25 to -50	3,236	37.9	31.9	24.0
-10 to -25	14,389	14.5	21.3	17.4
-0 to -10	65,704	7.3	12.4	10.8
0 to 10	48,455	6.9	12.8	11.4
10 to 25	21,615	9.0	15.6	13.4
25 to 50	9,978	12.5	14.9	12.5
50 or more	8,271	23.9	16.2	12.6

- NMAF more accurate where HH change 2000-2010 was rapid
- Less accurate in stable block groups
- Address-based varies less by HH rate of change

Evaluations After 2010

Block Group MAPEs by Percent Change in Households 2000-2010

Pct Change	N of BGs	Cen-based	NMAF '09	NMAF '10
-50 or more	630	213.8	101.7	70.5
-25 to -50	3,236	37.9	31.9	24.0
-10 to -25	14,389	14.5	21.3	17.4
-0 to -10	65,704	7.3	12.4	10.8
0 to 10	48,455	6.9	12.8	11.4
10 to 25	21,615	9.0	15.6	13.4
25 to 50	9,978	12.5	14.9	12.5
50 or more	8,271	23.9	16.2	12.6

- NMAF more accurate where HH change 2000-2010 was rapid
- Less accurate in stable block groups
- Address-based varies less by HH rate of change

Evaluations After 2010

Block Group MAPEs by Percent Change in Households 2000-2010

Pct Change	N of BGs	Cen-based	NMAF '09	NMAF '10
-50 or more	630	213.8	101.7	70.5
-25 to -50	3,236	37.9	31.9	24.0
-10 to -25	14,389	14.5	21.3	17.4
-0 to -10	65,704	7.3	12.4	10.8
0 to 10	48,455	6.9	12.8	11.4
10 to 25	21,615	9.0	15.6	13.4
25 to 50	9,978	12.5	14.9	12.5
50 or more	8,271	23.9	16.2	12.6

- NMAF more accurate where HH change 2000-2010 was rapid
- Less accurate in stable block groups
- Address-based varies less by HH rate of change

Evaluations After 2010

Percent of BGs with NMAF Error Less Than Census-Based Error

Pct Change	N of BGs	NMAF '09	NMAF '10
All BGs	172,278	44.5	50.8
-50 or more	630	82.4	89.2
-25 to -50	3,236	64.2	76.9
-10 to -25	14,389	38.5	51.3
-0 to -10	65,704	39.7	46.5
0 to 10	48,455	43.7	47.3
10 to 25	21,615	45.4	49.5
25 to 50	9,978	54.2	61.7
50 or more	8,271	73.8	80.6

- NMAF more accurate in almost half of block groups
- 74 percent of rapid growth BGs
- Even 40+ percent of stable BGs

Evaluations After 2010

Percent of BGs with NMAF Error Less Than Census-Based Error

Pct Change	N of BGs	NMAF '09	NMAF '10
All BGs	172,278	44.5	50.8
-50 or more	630	82.4	89.2
-25 to -50	3,236	64.2	76.9
-10 to -25	14,389	38.5	51.3
-0 to -10	65,704	39.7	46.5
0 to 10	48,455	43.7	47.3
10 to 25	21,615	45.4	49.5
25 to 50	9,978	54.2	61.7
50 or more	8,271	73.8	80.6

- NMAF more accurate in almost half of block groups
- 74 percent of rapid growth BGs
- Even 40+ percent of stable BGs

Evaluations After 2010

Percent of BGs with NMAF Error Less Than Census-Based Error

Pct Change	N of BGs	NMAF '09	NMAF '10
All BGs	172,278	44.5	50.8
-50 or more	630	82.4	89.2
-25 to -50	3,236	64.2	76.9
-10 to -25	14,389	38.5	51.3
-0 to -10	65,704	39.7	46.5
0 to 10	48,455	43.7	47.3
10 to 25	21,615	45.4	49.5
25 to 50	9,978	54.2	61.7
50 or more	8,271	73.8	80.6

- NMAF more accurate in almost half of block groups
- 74 percent of rapid growth BGs
- Even 40+ percent of stable BGs

What Evaluations Tell Us

- Census-based estimates still have overall advantage
 - Lower overall MAPE
- But address-based much improved
 - More complete coverage
 - Better geocoding
- NMAF more accurate in many BGs
 - Readily show dramatic difference with census
 - Clearly outperform in rapid change BGs
- Cannot justify complete switch to census-based
- Can we improve accuracy with hybrid approach?
 - Census-based estimates for some BGs
 - Address-based estimates for others?

Average of Census-Based and NMAF

Block Group MAPEs: Average of Census-Based and NMAF

Pct Change	N of BGs	Cen-based	Av w/ NMAF '09	Av w/ NMAF '10
All BGs	172,278	10.4	10.6	9.6
-50 or more	630	213.8	152.6	136.5
-25 to -50	3,236	37.9	32.7	28.4
-10 to -25	14,389	14.5	16.0	14.0
-0 to -10	65,704	7.3	8.3	7.6
0 to 10	48,455	6.9	7.9	7.3
10 to 25	21,615	9.0	9.9	9.1
25 to 50	9,978	12.5	11.3	10.1
50 or more	8,271	23.9	17.2	14.9

- Overall MAPE higher (for 2009)
- Avg with NMAF 2010 looks good . . . But not an option
- What about selective substitution of NMAF?

Perfect NMAF Substitution

BG MAPEs: Substitute NMAF in All BGs Where More Accurate in 2010

Pct Change	N of BGs	Cen-based	Sub NMAF '09	Sub NMAF '10
All BGs	172,278	10.4	6.3	5.6
-50 or more	630	213.8	75.0	51.2
-25 to -50	3,236	37.9	23.4	17.8
-10 to -25	14,389	14.5	10.4	8.8
-0 to -10	65,704	7.3	4.7	4.3
0 to 10	48,455	6.9	4.3	4.0
10 to 25	21,615	9.0	5.8	5.4
25 to 50	9,978	12.5	7.6	6.7
50 or more	8,271	23.9	11.9	9.1

- This is the unattainable ideal
- MAPE reduced for all categories of change
- How can we determine where to substitute NMAF?

Substitution Based on Actual Change

BG MAPEs: Substitute NMAF if Change LT -25 or GE 50

Pct Change	N of BGs	Cen-based	Sub NMAF '09	Sub NMAF '10
All BGs	172,278	10.4	9.5	9.1
-50 or more	630	213.8	101.7	70.5
-25 to -50	3,236	37.9	31.9	24.0
-10 to -25	14,389	14.5	14.5	14.5
-0 to -10	65,704	7.3	7.3	7.3
0 to 10	48,455	6.9	6.9	6.9
10 to 25	21,615	9.0	9.0	9.0
25 to 50	9,978	12.5	12.5	12.5
50 or more	8,271	23.9	16.5	12.8

- Another ideal (we don't know actual change)
- Improvement less pronounced
- Sub in only 12,137 BGs
- Sub by Change is inefficient – even if actual change known

Substitution Based on Estimated Change

BG MAPEs: Sub Based on Change in Available Resources*

Pct Change	N of BGs	Cen-based	Sub NMAF '09	Sub NMAF '10
All BGs	172,278	10.4	10.2	10.0
-50 or more	630	213.8	184.7	180.3
-25 to -50	3,236	37.9	37.9	37.5
-10 to -25	14,389	14.5	14.5	14.4
-0 to -10	65,704	7.3	7.3	7.3
0 to 10	48,455	6.9	7.0	7.0
10 to 25	21,615	9.0	9.2	9.1
25 to 50	9,978	12.5	12.6	12.3
50 or more	8,271	23.9	19.1	17.5

- * Selection explained in paper
- Only modest reduction in overall MAPE
- Useful reduction in rapid change BGs
- Accuracy maintained in stable BGs

Conclusions

- Can we improve accuracy by ignoring census?
 - If overall MAPE “No”
 - For many BGs “Yes”
 - Almost half of all block groups
- The challenge
 - How to tell which BGs are in that half
- Hybrid approach has potential
 - Modest improvement to overall accuracy
 - More notable improvement in rapid change areas
 - Areas of interest to many users

Conclusions

- Address-based estimates are independent from census
- Error varies less by rate of change
 - Error for rapid growth BGs
 - About same as error for stable BGs
- Error also varies less by time since last census
 - Census-based: error low after census and tops out 10 years later
 - Address-based: error about same throughout decade
- Implication
 - Census-based MAPE of 10.4 - about as high as it will get
 - Address-based MAPE of 12.5 - about as low as it will get
 - 2010 improvements not relevant now
- In short

Conclusions

- Take address counts seriously . . .
- But don't throw census counts out yet



Thank You

ken.hodges@nielsen.com

