ACS Multiyear PUMS Estimates and Usage: User-Created PUMS Files

Nate Ramsey
U.S. Department of Education

Research Questions

1. How accurate are concatenated single-year PUMS file estimates?

2. How often do people use multiyear PUMS files?

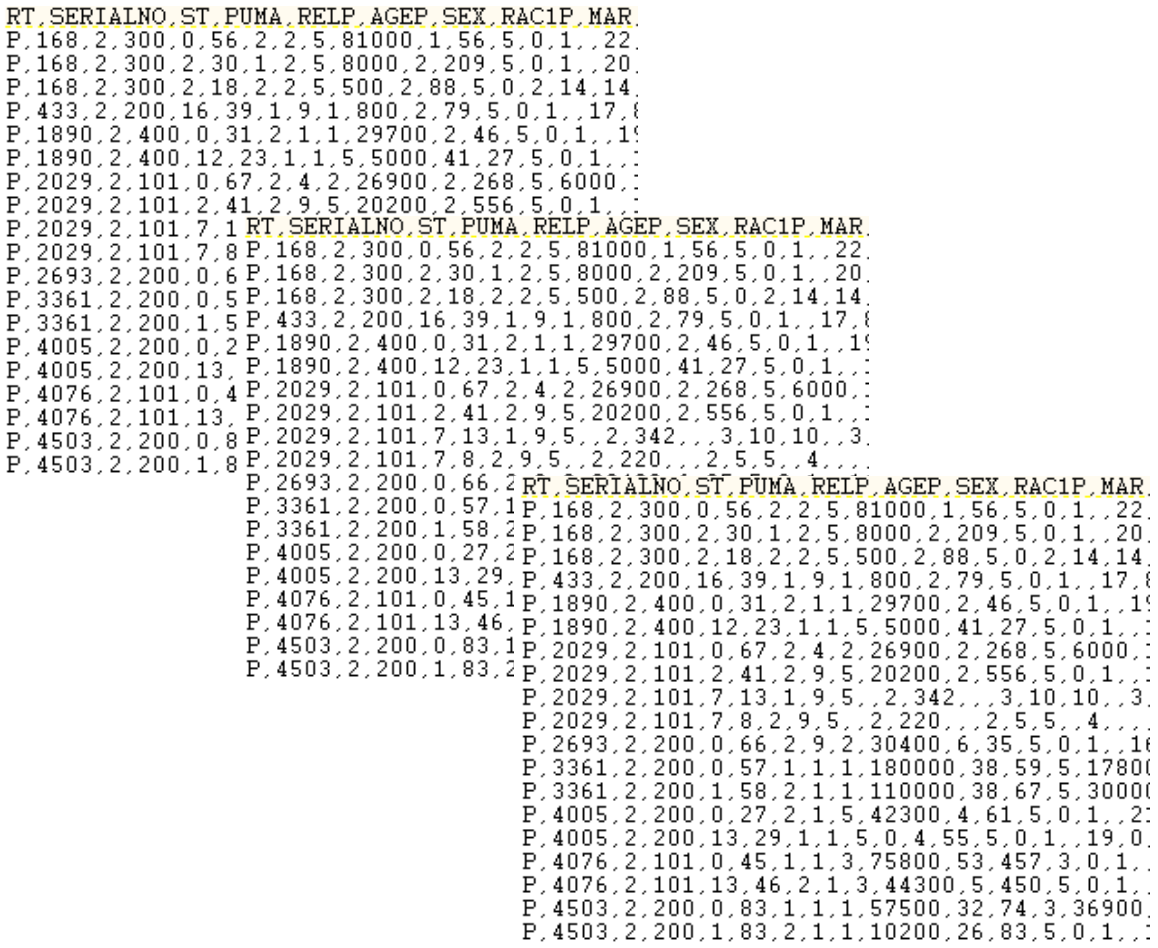
3. Can users be reasonably expected to create their own multiyear files?

Multiyear and Concatenated Estimate Differences

Multiyear ACS PUMS files allow users to more simply produce more accurate estimates

PUMS Concatenation Process

Step 1



Step 2

÷ **M**

Divide weight values or estimate by M (M = # of 1-year files)

Population Totals for Nation and States

3-Year Files: 2005-2007, 2006-2008, 2007-2009, 2008-2010, 2009-2011

5-Year Files: 2005-2009, 2006-2010, 2007-2011

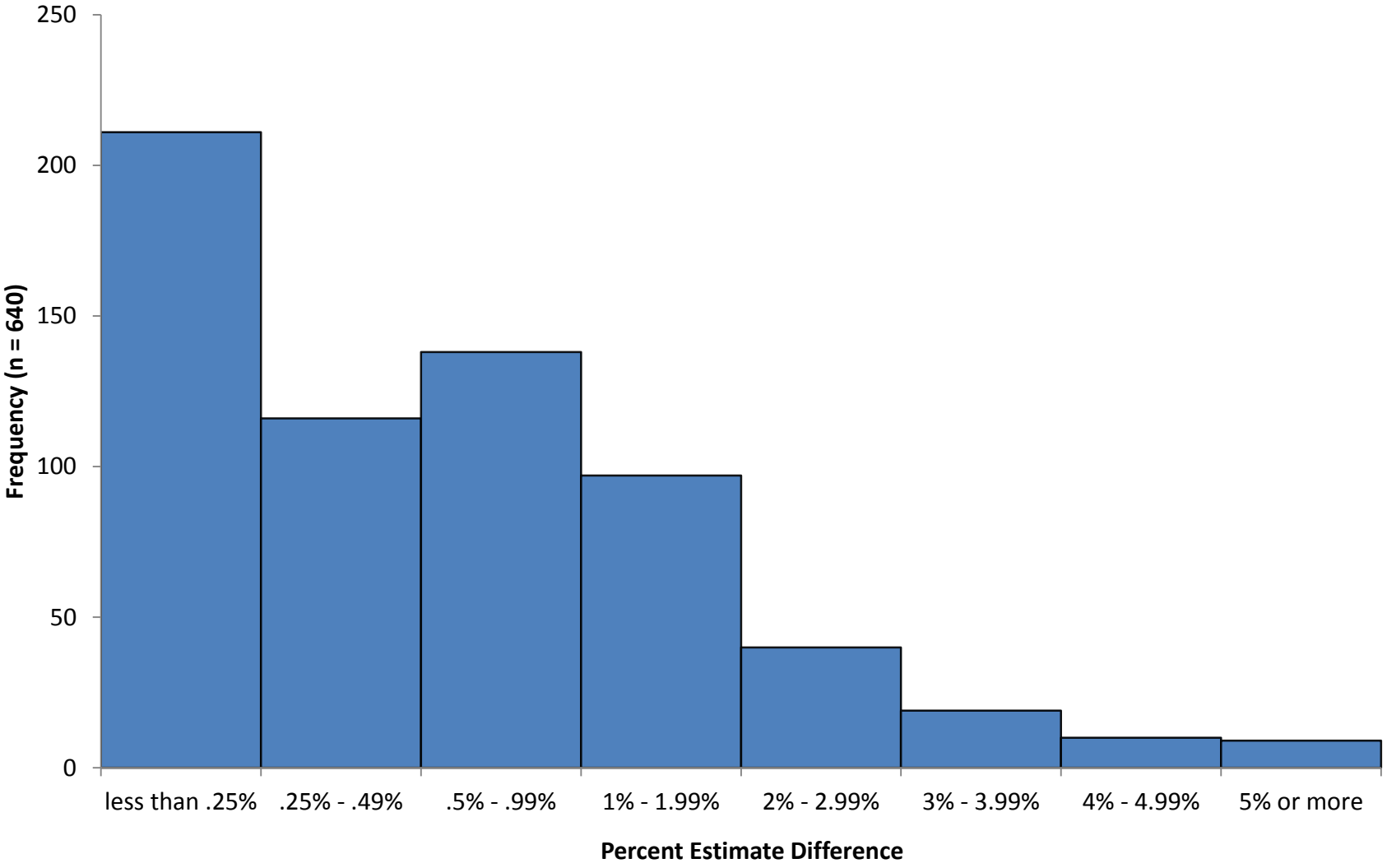
Geographies: US, California, Kentucky, Wyoming

ACS User Verification File Characteristics		
Total population	Age 20-24	Owner occupied units
Housing unit population	Age 25-34	Renter occupied units
GQ population	Age 35-44	Owned with a mortgage
GQ institutional population	Age 45-54	Owned free and clear
GQ noninstitutional population	Age 55-59	Rented for cash
Total males	Age 60-64	No cash rent
Total females	Age 65-74	Total vacant units
Age 0-4	Age 75-84	For rent
Age 5-9	Age 85 and over	For sale only
Age 10-14	Total housing units	All Other Vacant
Age 15-19	Total occupied units	

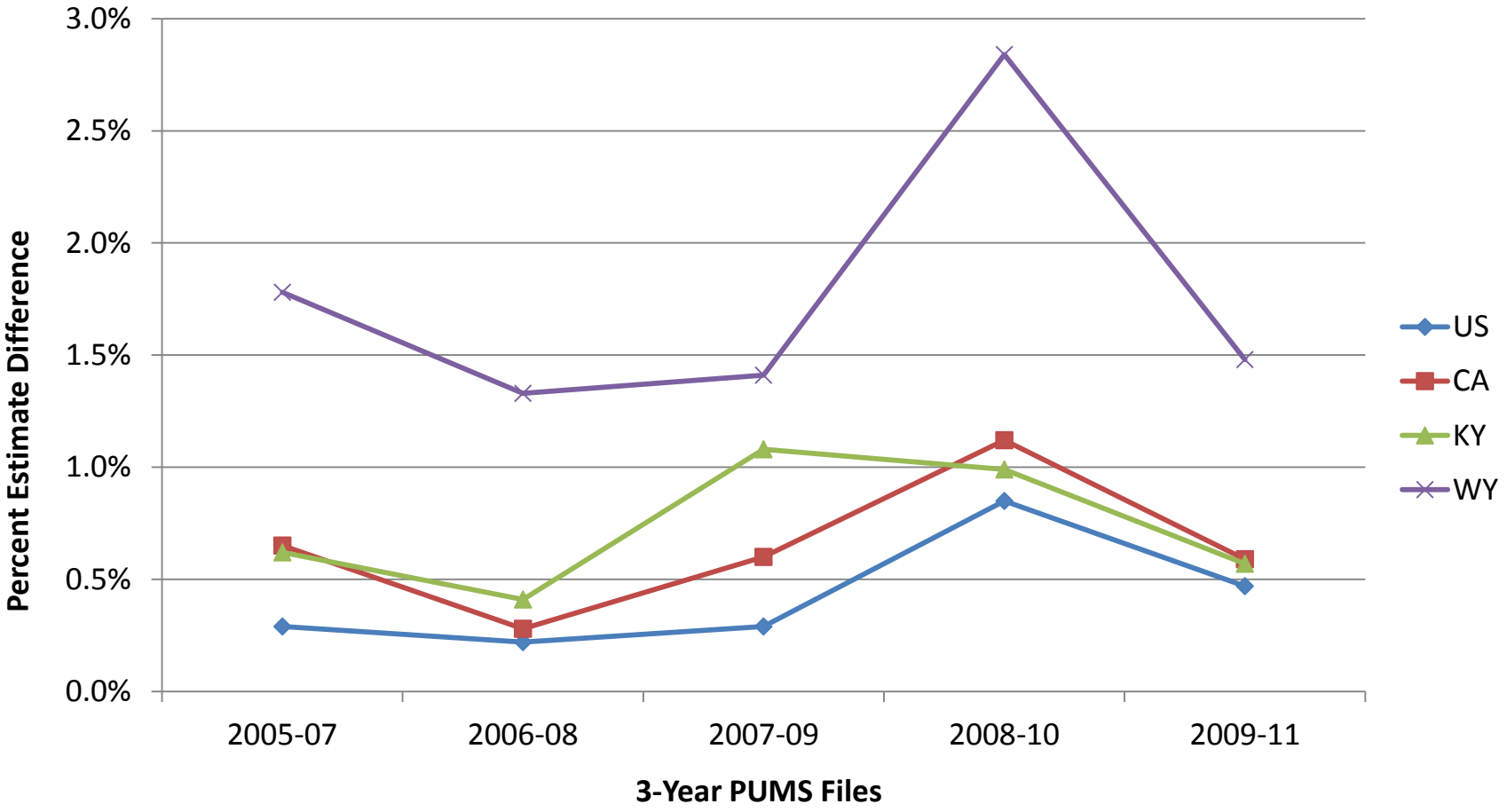
3-Year File Population Totals: Kentucky, 2009-2011

State	Characteristic	Multiyear	Concatenated	% Difference
Kentucky	Total population	4,344,553	4,342,989	0.04%
Kentucky	Housing unit population	4,218,755	4,217,910	0.02%
Kentucky	GQ population	125,798	125,079	0.58%
Kentucky	GQ institutional population	70,619	69,395	1.76%
Kentucky	GQ noninstitutional population	55,179	55,684	0.91%
Kentucky	Total males	2,136,267	2,132,320	0.19%
Kentucky	Total females	2,208,286	2,210,669	0.11%
Kentucky	Age 0-4	280,519	282,400	0.67%
Kentucky	Age 5-9	284,560	281,331	1.15%
Kentucky	Age 10-14	286,060	283,840	0.78%
Kentucky	Age 15-19	295,552	296,909	0.46%
Kentucky	Age 20-24	293,965	296,603	0.89%
Kentucky	Age 25-34	563,028	567,636	0.81%
Kentucky	Age 35-44	577,051	575,839	0.21%
Kentucky	Age 45-54	640,333	638,770	0.24%

3-Year File Percent Estimate Differences: United States, California, Kentucky, Wyoming



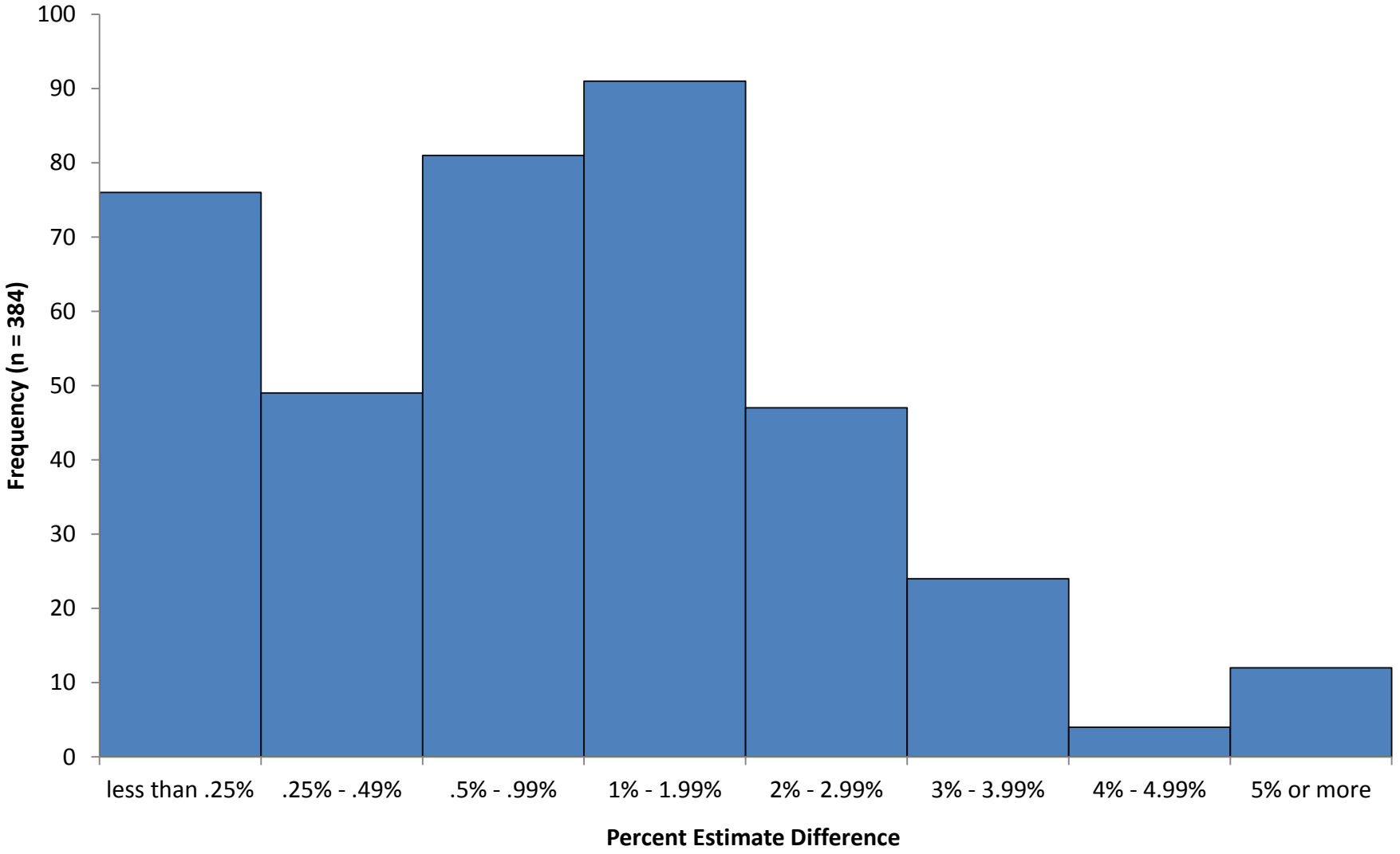
Average Estimate Differences: 3-Year PUMS Files



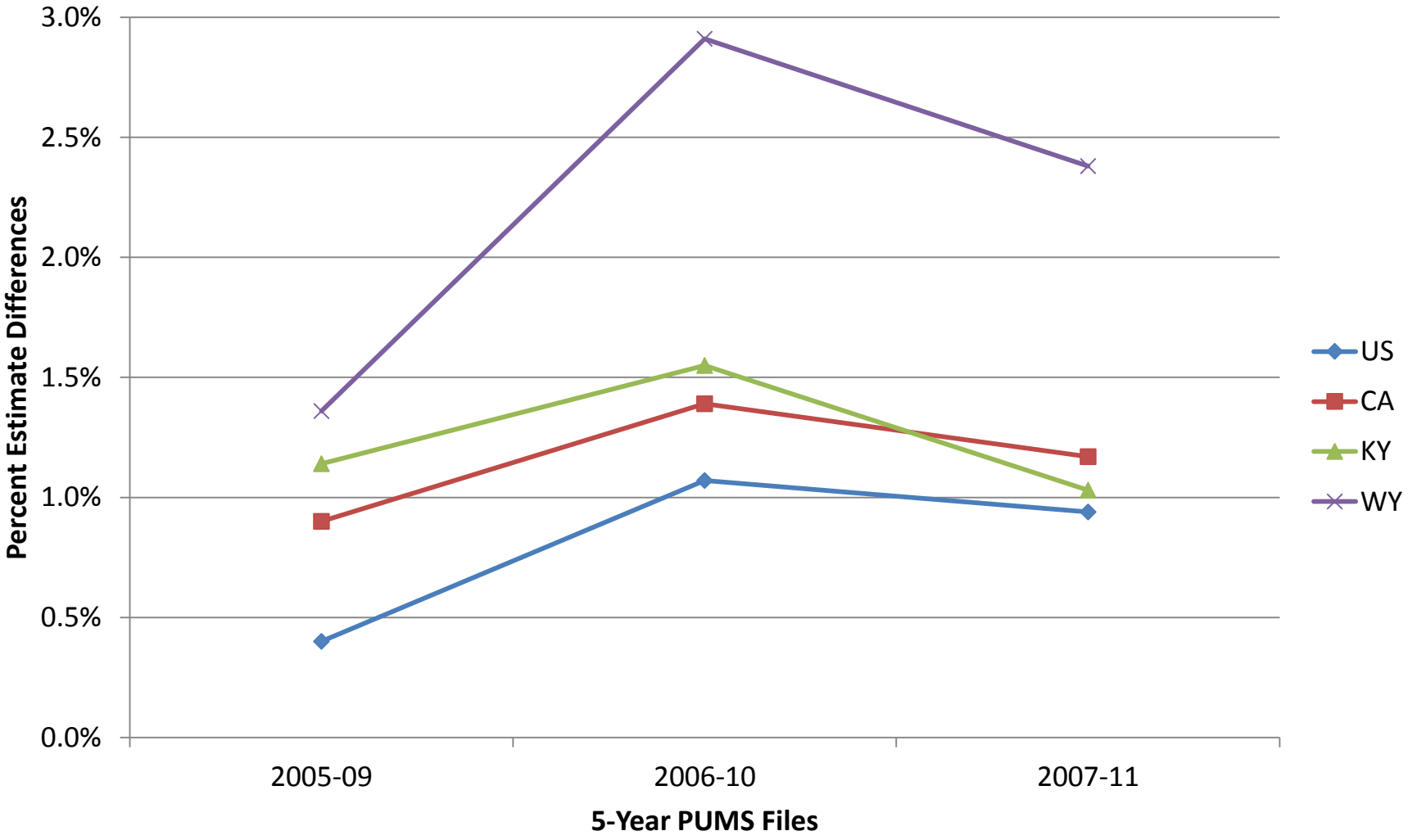
5-Year Population Totals: Kentucky, 2007-2011

State	Characteristic	Multiyear	Concatenated	% Difference
Kentucky	Total population	4,316,043	4,307,788	0.19%
Kentucky	Housing unit population	4,190,755	4,186,736	0.10%
Kentucky	GQ population	125,288	125,713	0.34%
Kentucky	GQ institutional population	70,185	66,482	5.57%
Kentucky	GQ noninstitutional population	55,103	54,571	0.98%
Kentucky	Total males	2,119,386	2,111,759	0.36%
Kentucky	Total females	2,196,657	2,196,030	0.03%
Kentucky	Age 0-4	279,966	284,305	1.53%
Kentucky	Age 5-9	282,683	278,178	1.62%
Kentucky	Age 10-14	282,711	279,005	1.33%
Kentucky	Age 15-19	296,912	297,197	0.10%
Kentucky	Age 20-24	291,385	285,779	1.96%
Kentucky	Age 25-34	561,384	568,805	1.30%

5-Year File Percent Estimate Differences: United States, California, Kentucky, Wyoming



Average Estimate Differences: 5-Year PUMS Files



Derived Estimates and Population Totals for PUMAs

3-Year Files: 2009-2011

Geographies: 54 Selected Public Use Microdata Areas (PUMAs)

Social Characteristics
Population in households with a relationship of spouse
Female population 15 years and over that are divorced
Population 3 years and over enrolled in college or graduate school
Population 25 years and over with a graduate or professional degree
Civilian population 18 years and over that are civilian veterans
Population 1 year and over with a residence 1 year ago of a different house in US
Population that are foreign born
Population 5 years and over that speak a language other than English at home
Economic Characteristics
Civilian labor force that is unemployed
Workers 16 years and over commuting to work by car, truck or van-carpooled
Civilian employed population 16 years and over in service occupations
Occupied housing units with income and benefits between $15,000-$24,999
Occupied housing units with Food Stamp/SNAP benefits in past 12 months
Housing Characteristics
Housing units built 1939 or earlier
Housing units with 1 bedroom
Occupied housing units with house heating fuel from electricity
Owner-occupied housing units with a property value of less than $50,000

Derived Estimates and Population Totals for PUMAs

Concatenated and Multiyear Estimates for Bend, OR						
	Derived Estimates (%)			Population Totals		
Characteristic	Multiyear	Concat.	Difference	Multiyear	Concat.	Difference
Relationship of spouse	22.15%	22.08%	0.07%	35,113	35,095	0.05%
Females that are divorced	13.35%	13.56%	0.21%	8,706	8,944	2.66%
Enrolled in college/grad school	23.45%	24.27%	0.82%	8,765	9,011	2.73%
Graduate or professional degree	10.8%	10.82%	0.02%	11,882	12,009	1.05%
Civilian veterans	12.45%	12.46%	0.01%	15,447	15,629	1.16%
Residence 1yr ago different house	17.21%	17.41%	0.2%	27,015	27,415	1.46%
Foreign born	4.57%	4.45%	0.12%	7,235	7,074	2.28%
Speak non-English language	6.16%	6.17%	0.01%	9,182	9,241	0.63%
Civilian labor force unemployed	9.33%	9.37%	0.04%	11,797	11,977	1.50%
Commuters by car, truck or van	10.38%	10.13%	0.25%	6,136	6,025	1.84%
Employed in service occupations	21.23%	21.33%	0.1%	14,735	14,911	1.18%
Income/benefits $15,000-$24,999	11.13%	11.46%	0.33%	7,183	7,361	2.42%
Food Stamp/SNAP benefits	16.41%	16.33%	0.08%	10,585	10,485	0.96%
Housing units built 1939 or earlier	3.61%	3.48%	0.13%	2,895	2,772	4.43%
1 bedroom households	5.77%	5.71%	0.06%	4,631	4,554	1.69%
House heating fuel from electricity	45.51%	45.31%	0.2%	29,359	29,098	0.90%
Property value less than $50,000	4.2%	4.22%	0.02%	1,765	1,768	0.16%
Average difference			**0.16%**			**1.60%**

Derived Estimates and Population Totals for PUMAs

Summary Statistics from Comparisons for the Fifty-Four Selected PUMAs				
	Derived Estimates (%)		Population Totals	
Characteristic	Average Difference	Max Difference	Average Difference	Max Difference
Relationship of spouse	0.23%	0.69%	1.37%	6.50%
Females that are divorced	0.17%	0.62%	1.46%	5.11%
Enrolled in college/graduate school	0.36%	1.58%	1.73%	4.82%
Graduate or professional degree	0.07%	0.35%	1.10%	4.29%
Civilian veterans	0.09%	0.44%	0.97%	4.55%
Residence 1yr ago of a different house	0.13%	0.63%	1.11%	3.46%
Foreign born	0.18%	0.6%	4.30%	26.24%
Speak non-English language at home	0.21%	0.8%	2.71%	11.91%
Civilian labor force that is unemployed	0.08%	0.23%	1.63%	4.92%
Commuters to work -carpooled	0.15%	0.43%	1.76%	4.58%
Employed in service occupations	0.18%	0.73%	1.21%	3.78%
Income and benefits $15,000-$24,999	0.31%	1.02%	2.35%	7.08%
Food Stamp/SNAP benefits	0.16%	0.57%	1.27%	4.52%
Housing units built 1939 or earlier	0.08%	0.34%	1.16%	4.43%
1 bedroom households	0.11%	0.4%	1.40%	5.22%
House heating fuel from electricity	0.16%	0.38%	1.29%	6.02%
Property value of less than $50,000	0.10%	0.41%	1.29%	4.47%
Overall	0.16% (avg)	1.58% (max)	1.65% (avg)	26.24% (max)

Multiyear PUMS Publications

Multiyear PUMS Publications

Methodology

Google Scholar

- Publications include “articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites.”
- Ranks publications according to “where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature.”
- Keyword search; Select publications from beginning (not a random sample); N = 129

Multiyear PUMS Publications

Findings

	Single-Year Files	Multiyear Files
Publications (n=129)	115 (89%)	14 (11%)

	Not Concatenated	Concatenated
Single-Year Files (n=115)	106 (92%)	9 (8%)

Multiyear PUMS Publications

Findings

	National	State	PUMA
Geography (n=129)	86 (67%)	54 (42%)	28 (22%)

	Academic Journal	Research/Advocacy Organization
Publication Type (n=129)	74 (57%)	55 (43%)

Multiyear PUMS Publications

Findings

	Census (FTP/AFF/DF)	IPUMS
Access Source (n=129)	71 (55%)	58 (45%)

	Immigration/ Hispanic	Age	Race/Ancestry	Employment/ Occupation/ Industry
Topic (n=129)	53 (41%)	28 (22%)	23 (18%)	21 (16%)

Concatenation Issues and User Support

Concatenation issues/support

Concatenation

- Conceptually straightforward
  - Adding cases/rows together
  - Already required for US files
- Technically straightforward
  - Adjusting weights/estimates should be easy for microdata users
  - Simpler than merging person and housing files

2011 PUMS ReadMe Example Code

Current US File Concatenation

```
data population;
set file1 file2;
run;
```

Proposed File Concatenation

```
data population3yr;
set file1 file2 file3;
run;
```

Current Population-Housing Merge

```
proc sort data=population;
by serialno;
run;
proc sort data=housing;
by serialno;
run;
data combined;
merge population (in=pop) housing;
by serialno;
if pop;
run;
```

Concatenation issues/support

Data and Metadata Changes

- Break in series / Add or delete variables
- Metadata changes
- Dollar-denominated values

Conclusions

- A typical difference between multiyear and concatenated estimates is around 1%-1.5%
  - Small populations/geos, can mean larger differences
  - More combined years, use more caution
- Multiyear PUMS publications are a fraction of single-year publication
  - Users are already concatenating files
- Concatenation process safely within bounds of current requirements of using PUMS products