



■ Population Estimates



Pam Perlich
URBPL 5010: Urban Research
University of Utah
Rev. November 27, 2006



■ Why Estimates are Necessary

- Enumerations are only every ten years
- Changes in:
 - Totals
 - Distributions
 - Characteristics
- National surveys do not cover small scale geography



■ Uses of Estimates

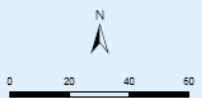
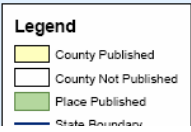
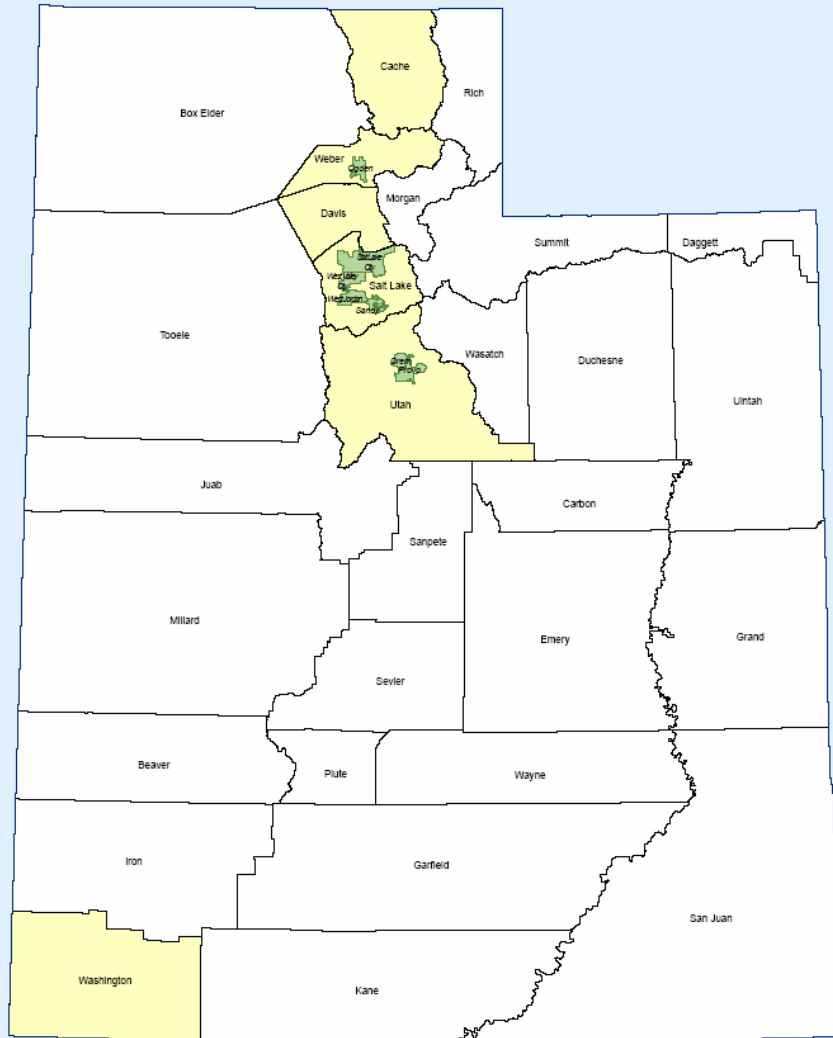
- Market analysis
- Public service planning
- Federal government allocates resources based on estimates generated by Bureau of the Census
- Civil rights enforcement
- Environment impact studies



American Community Survey (ACS)

- Nationwide survey designed to
 - Provide more frequent data than the decennial census
 - Replace the long form by 2010 census
- Conducted by the Bureau of the Census
- Data available for states, some MSAs, and some counties

Utah Counties and Places
Published in the 2005 ACS



2005 ACS Data for Utah

- Geographic areas with > 65,000 population
- State
- Cache, Weber, Salt Lake, Davis, Utah, Washington counties
- Provo-Orem MSA
 - Utah County
- SLC-Ogden MSA
 - Weber, Davis, and Salt Lake counties
- Places: Ogden, Orem, Provo, Salt Lake, Sandy, West Valley, West Jordan



■ ACS Basics

- Sample survey → confidence intervals
- Can use micro data
- Online access to summary tables
- Content very similar to long form
- <http://www.census.gov/acs/>



Population Estimates: Bureau of the Census

- National: Age, Race, Sex, Ethnicity
 - Most detail of all estimates
- Counties
 - Components of change
 - Age, race, ethnicity, sex
- All incorporated places: totals
 - Distributive housing method allocates county totals
- State – sum of the counties

U.S. Census Bureau

p o p u l a t i o n e s t i m a t e s

Estimates Data Analysis Graphics Gallery Estimates Topics Geographic Topics Archives Related Topics

[census](#) > [population estimates](#) > [topics](#) > [methodology] > overview

methodology

The methodologies presented below are drawn from printed reports and from unprinted sources. Each methodology statement pertains to the most recent data

- ◆ [National Population Estimates](#)
- ◆ [National Population Estimates by Age, Sex, Race, and Hispanic Origin](#)
- ◆ [State and County Population Estimates](#)
- ◆ State Population Estimates by Age, Sex, Race, and Hispanic Origin
 - ◇ PART ONE: [State Population Estimates by Age and Sex](#)
 - ◇ PART TWO: [State Population Estimates by Age, Sex, Race, and Hispanic Origin](#)
- ◆ [County Population Estimates by Age, Sex, Race, and Hispanic Origin](#)
- ◆ [Cities and Towns Population Estimates](#)
- ◆ [Housing Unit Estimates](#)
- ◆ [Puerto Rico Commonwealth](#)
- ◆ [Puerto Rico Municipios](#)

Source: U.S. Census Bureau, Population Division
Maintained By: Demographic Internet Staff
Contact: pop@census.gov

[People](#) | [Estimates](#) | [Projections](#) | [Glossary](#) | [Help](#)

Census Bureau Links: [Home](#) • [Search](#) • [Subjects A-Z](#) • [FAQs](#) • [Data Tools](#) • [Catalog](#) • [Census 2000](#) • [Quality](#) • [Privacy Policy](#) • [Contact](#)

U S C E N S U S B U R E A U

Helping You Make Informed Decisions



National Estimates: BOC

- 2000 Census enumeration (STOCK)

FLOWS:

- Plus births
- Minus deaths
- Plus net international migration
(based on survey and other data)
- Plus net movement of U.S. Armed Forces



County Estimates: BOC

- Method: administrative records component of change
- Separate treatment of
 - Household and group quarters
 - Under and over 65 years of age
 - International and internal migration



County Estimates: BOC

- Internal migration
 - < 65 years: IRS data
 - 65+ years: Medicare data
- International migration
 - Distribute national totals

Estimation Methods (Siegel Ch. 9)

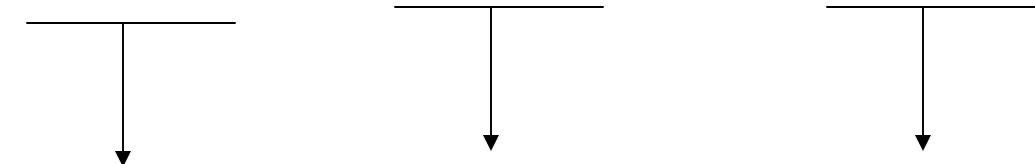
1. Census & population registers
 2. Direct estimates based on sample surveys
 - Estimate characteristics, not total population
 3. Mathematical extrapolation and proration / apportionment
 - Constant proportion or growth rate
 - “These methods produce estimates that have rather large errors over a decade as compared with more analytic approaches.” Siegel, p. 405
 4. Demographic Analysis Methods
 5. Ratio-regression or other regression
- } Discussed in detail below

Demographic Analysis

- Use of symptomatic data to estimate migration flows
- ***Housing unit method***
- ***Censal ratio method:*** vital rates
- ***Component Methods***
 - Administrative records
 - School enrollment

Housing Unit Method

$$P_t = OH_t \times PPH_t + GQ_t$$


Occupied Housing Units Persons per Household Group Quarters Population



■ Housing Unit Method

- Household population
 - Changes in occupied housing stock
 - Demolition and construction
 - Building permits – lags
 - Vacancy rates
 - Changes in persons per household
- Group quarters population
 - Customized approaches
- Detailed worksheet, Siegel, Ch. 9.

Censal Ratio: Vital Rates Method

- For the decennial year, calculate
 - Births to population ratio
 - Deaths to population ratio
- Post censal
 - Use vital records to calculate population
 - Average the two results
- Not often used
 - Ignores age waves
 - Assumes constant age structure



Component Method: Administrative Records

- Births and deaths from vital records
- Migration estimated from symptomatic data
 - IRS
 - Medicare
 - LDS membership
 - Others
- Assumes ratio of variable to population is constant.



■ School Enrollment Method

- Births and deaths from vital records
- School enrollment grades 1-8 in year 1
- Apply survival rates
- Compare to school enrollment grades 2-9 in year 2
- Calculate implied student age migration
- Scale up to total population

Ratio Regression or Other Regression Models

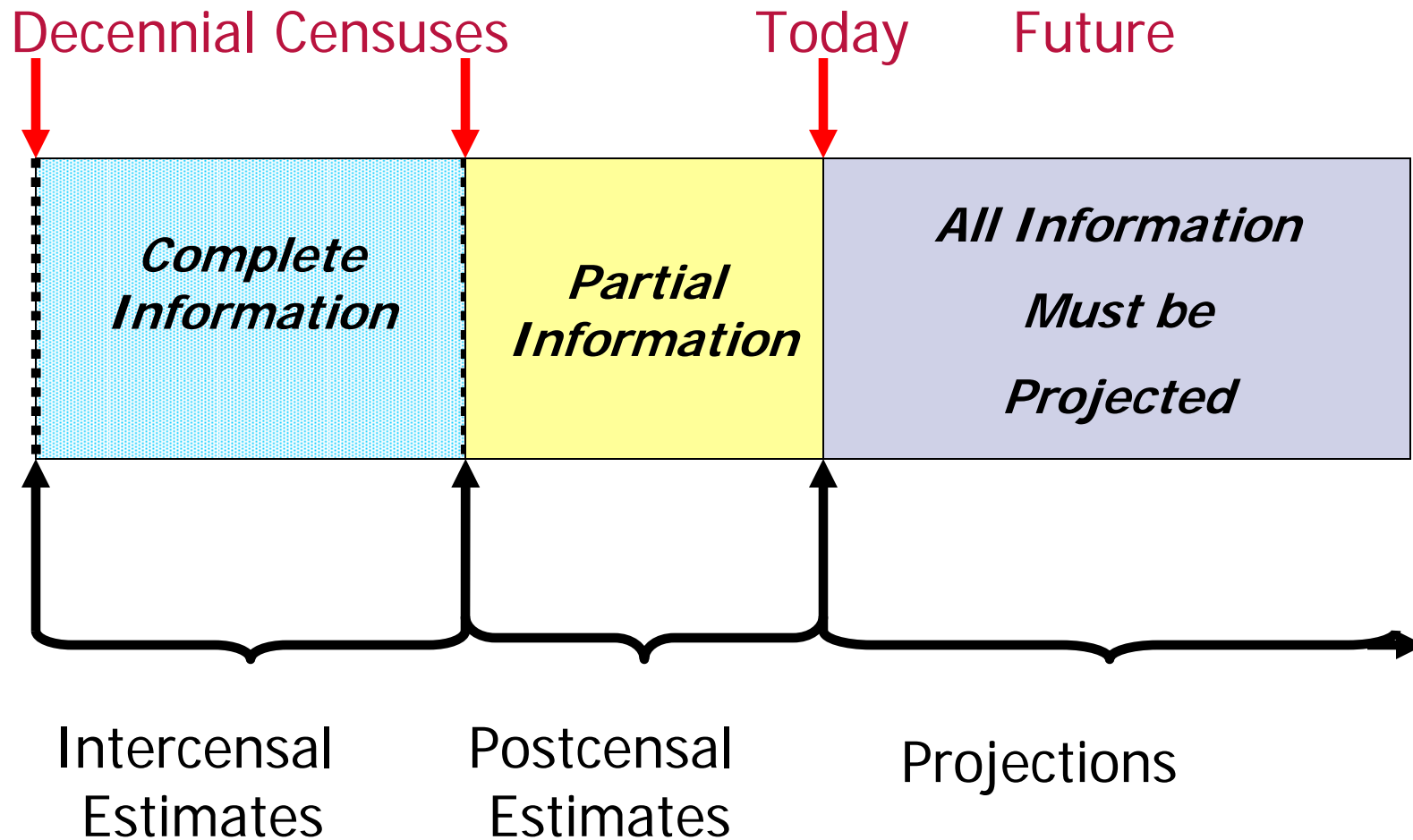
- Build time series multiple regression models
- County shares of state:
 - Population, births, deaths, school enrollment
 - Include all counties
 - Modeling shifting shares of all counties simultaneously over time.
- Multicollinearity is a problem
 - Independent variables are interrelated



Estimates of Detailed Demographic Characteristics

- Cohort component method
 - Detailed beginning population
 - Age and survive
 - Apply appropriate fertility rates
 - Estimate in and out migration
 - Sum to new population
- Like a projection but with partial data

Estimates and Projections





■ Evaluation of Methods

- Create a series using each method
- Generate estimates for 1990 – 2000
- Compare estimates to the 2000 enumeration.
- Compute errors.

Measures of Error: MAPE

Mean Absolute Percent Error, unweighted

$$MAPE_u = \frac{\left| \sum P_i \right|}{n}$$



Sum of absolute value
of the percent errors
for each of the n areas
in estimate

Measures of Error: MAPE

Mean Absolute Percent Error, weighted by population of areas

$$MAPE_w = \frac{\left(\sum |P_i| \times PopCount\right)}{\left(\sum PopCount \times 100\right)}$$



Sum of absolute value of the percent errors for each of the n areas in estimate times the decennial enumeration

Percent errors of large areas are more heavily weighted

Generally weighted MAPE < unweighted MAPE

Measures of Error: RMSE

Root Mean Squared Error

$$RMSE = \sqrt{\left(\sum P_i^2\right) \div n}$$

Sum of squared percent errors. Divide by the number of areas in estimate. Take the square root of the result.

Greater weight to larger errors.



**Revised Population Estimates for the 1990s:
Utah Population Estimates Committee**

Pam Perlich
Bureau of Economic and Business
Research
University of Utah

August 27, 2001

Intercensal Estimates Revision

- Utah Population Estimates Committee Work Group:
 - Peter Donner, Governor's Office of Planning & Budget
 - Ken Jensen, Utah Dept. of Workforce Services
 - Pam Perlich, Bureau of Econ. & Business Research, U of U
 - Tom Williams, Utah State Tax Commission
- Pamela S. Perlich, Ph.D, "Revised Population Estimates for the 1990s," *Utah Business and Economic Review*, Volume 61, Numbers 5 and 6.
- http://www.business.utah.edu/bebr/bebrFiles/May_Jun2001.pdf



Population Estimates & Census Counts

- Decennial Census – Bureau of the Census
 - Every decade the population is enumerated
- Utah Population Estimates Committee
 - Prepares annual county population estimates
 - Uses symptomatic data in combination with vital records
 - The estimates for each decade are revised to incorporate Census counts



■ Utah Population: April 1, 2000

- Census count = 2,233,169
- UPEC estimate = 2,151,956
- Underestimated by 3.6% or 81,213 persons

Population Change

$$\text{Pop}_{(2000)} = \text{Pop}_{(1990)} + \text{NatInc}_{(1990-2000)} + \text{NetMig}_{(1990-2000)}$$

$$2,233,169 = 1,722,850 + 298,340 + 211,979$$

UPEC: 130,776

$$81,213 / 211,979 = 38\%$$

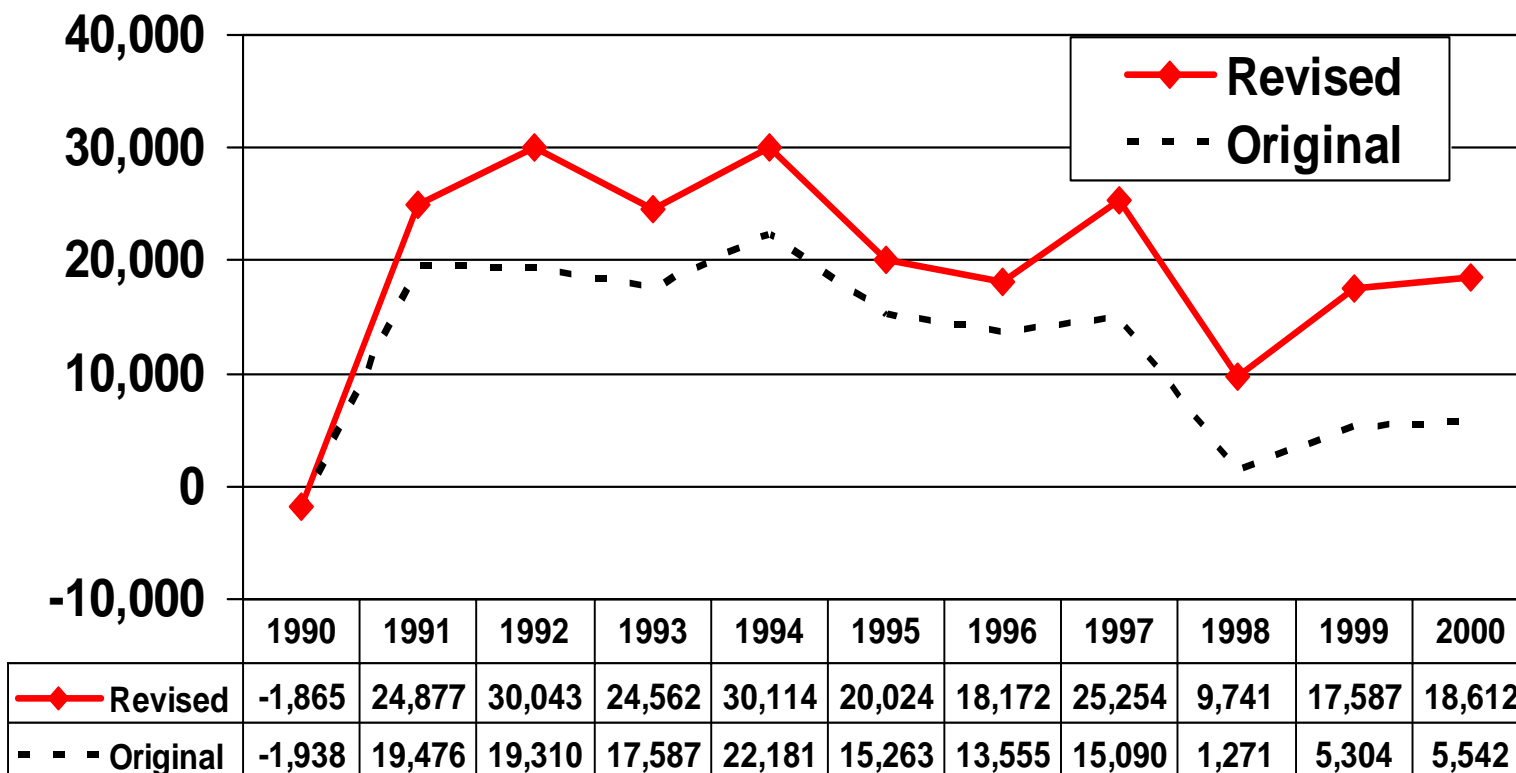
UPEC missed 38% of net in-migration to the state



■ UPEC Intercensal Revisions

- Distribute across time and space the 81,000 people we missed in the 1990s estimates
- Evaluate the accuracy of UPEC methods
- Consider past revision methodologies
- Determine a methodology for the revisions

Utah Net Migration: Original and Revised





■ UPEC Methods

- School Enrollment: Component Method II
- LDS Church Membership
- IRS Exemptions
- UPEC: The results of all three methods are combined judgmentally with additional data such as building permits, employment, utility hook-ups, etc.



■ Revision Method

- County population estimates were produced using each of the four methods.
- The least error method was identified for each county.
- This “best method” series was adjusted to terminate at the Census 2000 count.

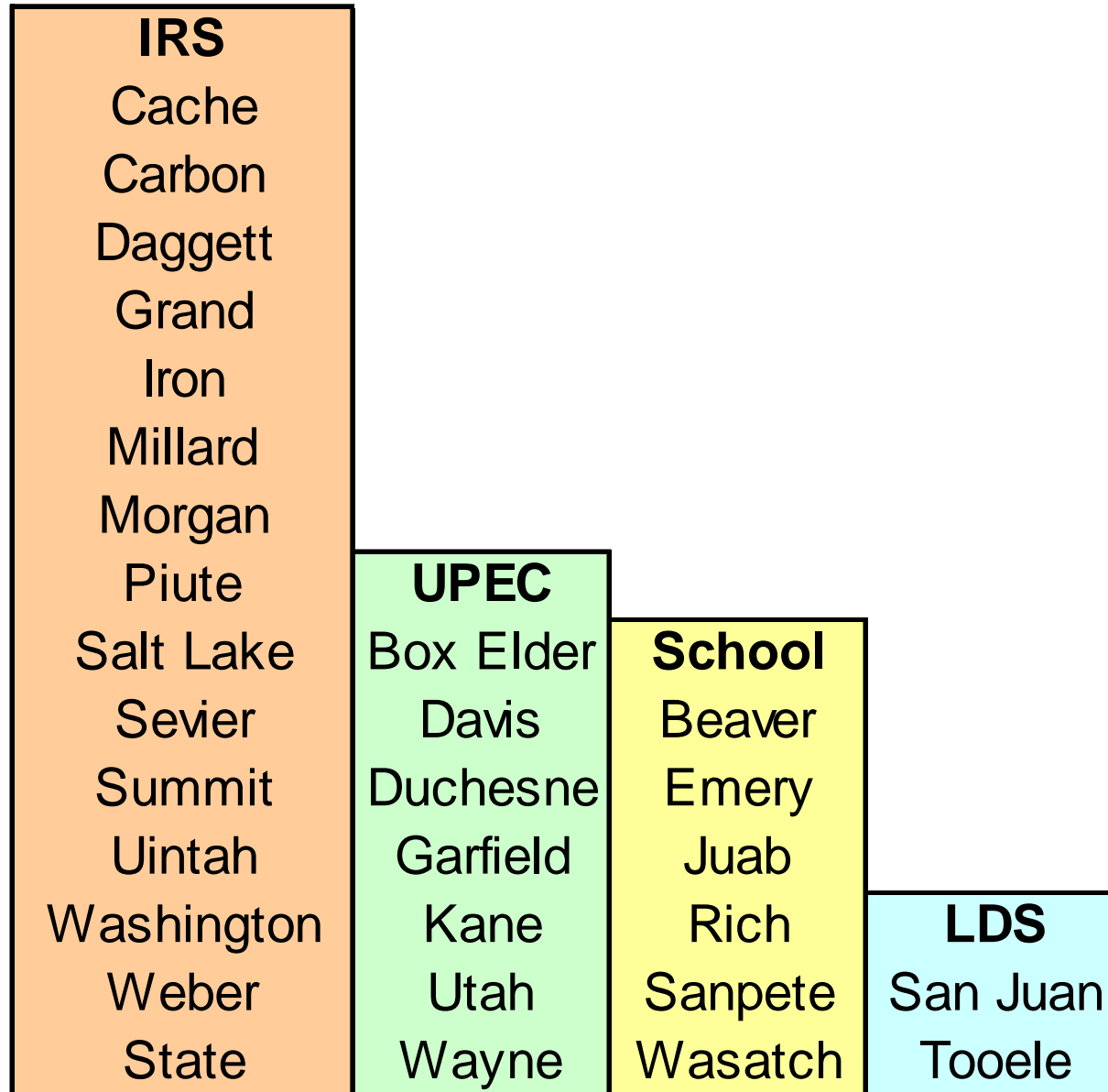
Accuracy of Methods

- All methods underestimated the population.
- IRS Exemption Method was by far the most accurate underestimating the state total by 3,733 people or 0.02 percent.
- IRS Exemption Method was also more accurate than the estimates generated by the Bureau of the Census.
- UPEC Method was second most accurate followed by the School Enrollment and finally the LDS Membership Method.

Accuracy Measures of Four UPEC Methods: Evaluation of the 1990s

| | Methods | | | |
|------------------------------------|--------------|--------|-------|-------|
| | IRS | School | LDS | UPEC |
| <i>Measures</i> | | | | |
| State Level Total Error | -0.2% | -4.1% | -7.8% | -3.6% |
| County Errors: Population Weighted | 5.72 | 61.19 | 78.47 | 19.80 |
| County Errors: MAPE | 3.4% | 10.4% | 7.8% | 4.6% |
| <i>Ranking</i> | | | | |
| State Total Error | 1 | 3 | 4 | 2 |
| County Errors: Population Weighted | 1 | 3 | 4 | 2 |
| County Errors: MAPE | 1 | 4 | 3 | 2 |
| Overall Ranking | 1 | 3 | 4 | 2 |

Best Methods for Counties



**Underestimation
< 3 Percent**

| | |
|------------|--------|
| Daggett | -19.5% |
| Summit | -8.7% |
| Washington | -8.0% |
| Tooele | -7.1% |
| Wasatch | -6.9% |
| Rich | -6.9% |
| San Juan | -6.8% |
| Sanpete | -5.4% |
| Salt Lake | -5.2% |
| Weber | -4.2% |
| Iron | -4.1% |
| State | -3.6% |
| Millard | -3.6% |

Within 3 Percent

| | |
|------------------|--------------|
| Cache | -2.8% |
| Garfield | -2.3% |
| Emery | -2.3% |
| Box Elder | -1.7% |
| Beaver | -1.6% |
| Utah | -1.6% |
| Uintah | -1.1% |
| Davis | 0.0% |
| Juab | 0.2% |
| Duchesne | 0.7% |
| Kane | 1.3% |
| Sevier | 1.5% |
| Wayne | 2.0% |
| Carbon | 2.8% |

**Overestimation
> 3 Percent**

| | |
|--------|-------|
| Morgan | 3.9% |
| Grand | 5.2% |
| Piute | 16.5% |

Accuracy by County

- Salt Lake County, home to 40 percent of Utah residents, accounted for 58 percent (47,069 persons) of the error.
- Combined estimation errors for Salt Lake, Weber, Washington, Utah, Tooele, Summit, and Cache account for 94 percent of the error.
- The Davis County estimate was the most precise, missing the total of 238,994 by 72, which is an error of -0.2 percent.
- Percent estimation error was not correlated with the size of the county. This is unusual.

Accuracy by County

- Four of the smallest counties were among the ten counties with the largest percentage estimation error.
- Among the ten counties with the ***smallest*** estimation errors were Wayne, Beaver, Kane and Juab. Each of these is among the ten smallest counties in the state.
- Summit, Washington, Tooele, and Salt Lake (which together have 47 percent of the state's population) were among the ten counties with the largest percentage estimation errors.

Accuracy Measures of UPEC Estimates by Decade: 1950s through 1990s

| | Decades | | | | |
|------------------------------------|---------|-------|--------------|-------|--------------|
| | 1950s | 1960s | 1970s | 1980s | 1990s |
| <i>Measures</i> | | | | | |
| State Level Total Error | 2.3% | 2.0% | -3.5% | 0.7% | -3.6% |
| County Errors: Population Weighted | 5.13 | 4.18 | 11.95 | 9.41 | 19.80 |
| County Errors: MAPE | 5.4% | 4.5% | 5.3% | 4.7% | 4.6% |
| <i>Ranking</i> | | | | | |
| State Total Error | 3 | 2 | 4 | 1 | 5 |
| County Errors: Population Weighted | 2 | 1 | 4 | 3 | 5 |
| County Errors: MAPE | 5 | 1 | 4 | 3 | 2 |
| Overall Ranking | 3 | 1 | 4 | 2 | 5 |

Accuracy by Decade

- In the relatively slower growth decades (1950s, 1960s, 1980s):
 - Estimates were generally more accurate.
 - Smaller percentage errors were associated with larger counties.
- In the decades of rapid growth (1970s, 1990s):
 - Estimates were generally less accurate.
 - There was no correlation between county size and percent errors.
- Natural resource counties have been difficult to estimate.

Revision Methodologies: 1950s & 1960s

- New series was created using UPEC methods
- Natural increase data was used in combination with school enrollment method, simple linear growth, and labor force changes
- Series were forced to match the Census counts.
- Series were forced to match the state totals in the Bureau of the Census revisions.
- County totals were independent of the Bureau of the Census revisions.

Revision Methodologies: 1970s

- A completely new series was produced using methods not normally utilized by UPEC.
- Ratios of symptomatic data to the population series were calculated for 1960 and 1970.
- A multiple regression time series model was built to generate estimates.
- Series was forced to match the Census counts.
- Series was forced to match the state totals in the Bureau of the Census revisions.
- County totals were independent of the Bureau of the Census revisions.



■ Revision Methodologies: 1980s

- The original UPEC series was scaled so that the county totals were consistent with the decennial Census.
- This series was further scaled to match the state totals in the Bureau of the Census revisions.
- County totals were independent of the Bureau of the Census revisions.



■ Revision Methodologies: 1990s

- A series was created for each method.
- The least error method/series was selected for each county.
- This series was further scaled to match the state totals in the Bureau of the Census revisions.

Revision Methodologies: 1990s

$$UPEC_{t, j, z} = Q_{j, z} \times POP_{t, j, z}$$

UPEC is the revised series

POP is the series prior to scaling

Time=t

County=j

Method=z

County Estimation Errors by Decade

| | 1950s | 1960s | 1970s | 1980s | 1990s |
|-------------------|-------|-------|-------|-------|-------|
| Salt Lake | 1.8% | 2.9% | -6.0% | -1.0% | -5.2% |
| Utah | 4.0% | -3.4% | -5.9% | 3.8% | -1.6% |
| Davis | 0.8% | 1.0% | -3.8% | 0.6% | 0.0% |
| Weber | 3.2% | 5.2% | 1.5% | 1.1% | -4.2% |
| Cache | 4.1% | 5.3% | 0.4% | 3.6% | -2.8% |
| <i>Washington</i> | -1.0% | -5.4% | -9.6% | -4.4% | -8.0% |
| Box Elder | 5.5% | -1.1% | 2.0% | 5.3% | -1.7% |
| Tooele | -1.9% | 8.3% | -2.6% | 5.1% | -7.1% |
| Iron | 3.3% | -2.5% | 0.9% | -4.6% | -4.1% |
| Summit | 0.0% | 5.1% | -8.4% | -8.2% | -8.7% |
| <i>Uintah</i> | -3.5% | -4.8% | -2.5% | -5.5% | -1.1% |
| Sanpete | 5.9% | 0.9% | 5.3% | 4.9% | -5.4% |
| <i>Carbon</i> | 0.5% | 4.3% | 3.0% | 5.2% | 2.8% |
| <i>Sevier</i> | 8.1% | 1.0% | 3.8% | 3.8% | 1.5% |

County Estimation Errors by Decade

| | 1950s | 1960s | 1970s | 1980s | 1990s |
|---------------|---------------|--------------|--------------|--------------|---------------|
| Wasatch | 1.9% | -1.7% | -2.5% | 0.3% | -6.9% |
| Duchesne | 3.4% | -2.8% | 0.4% | -0.3% | 0.7% |
| San Juan | -15.4% | 2.2% | 23.5% | 3.0% | -6.8% |
| Millard | 8.2% | 10.0% | 5.3% | 15.0% | -3.6% |
| Emery | 2.7% | 2.0% | -1.8% | 7.7% | -2.3% |
| Grand | -8.1% | 17.6% | -0.6% | -3.2% | 5.2% |
| Juab | 13.0% | 2.2% | 5.6% | -1.9% | 0.2% |
| <i>Morgan</i> | 9.1% | 2.6% | 10.4% | 7.8% | 3.9% |
| Kane | 11.1% | 4.2% | 2.6% | -6.0% | 1.3% |
| Beaver | 7.0% | 2.6% | 1.1% | 0.9% | -1.6% |
| Garfield | 1.4% | -3.2% | 10.1% | 1.7% | -2.3% |
| Wayne | 11.8% | 6.7% | 0.0% | -5.3% | 2.0% |
| Rich | 0.0% | -12.5% | -7.3% | 6.5% | -6.9% |
| Piute | -3.4% | 7.7% | 20.0% | 19.2% | 16.5% |
| Daggett | 18.2% | 0.0% | 6.7% | -0.1% | -19.5% |

Conclusions

- No Surprise – estimation is more difficult in times of rapid change
- Surprise – Large areas may have greater relative estimation errors, especially in times of relatively rapid growth
- The large errors associated with the LDS and School Methods indicate structural change of the population.



■ Conclusions

- IRS method is more directly tied to the economy and has proven much more accurate for the 1990s.
- The use of a housing method might also provide an economic view of population change.
- UPEC could use a “top down” state model in combination with the “bottom up” county methodologies to better capture regional growth.



■ Unresolved Questions

- Why have the systematic estimation errors in the five counties occurred?
- Why were we so close in Davis County?