ESTIMATING INTERCENSAL HOUSEHOLD SIZE IN NYC'S NEIGHBORHOODS

2019 ACS Data User Conference

Annette Jacoby, Joseph Salvo, Peter Lobo

NYC PLANNING

Objective

- Housing Unit Method for population estimation:

  Occupied HU x   PpH $^?$

- ACS PpH values too unreliable

  Swanson, David A. (2010): "Substantial massaging needs to be done to iron out the temporal instabilities and large variances found for person per household values in many areas."

- Develop intercensal household size for reliable population estimates at sub-borough levels that reflect geodemographic segmentation

  Estimates of persons per household (PpH) for New York City's 188 Neighborhood Tabulation Areas (NTAs) for 2013-2017 period

DRAFT

# A regression-based approach for intercensal PpH

- "At least for some states and counties, regression models are capable of producing more precise and less biased PPH estimates than those produced by more common methods" (West, Ahmed, Bruce, & Judson, 2009)

- Regression-based approach
  - Employ demographic rather than a statistical perspective
  - PpH varies in response to socio-demographic neighborhood characteristics or "symptomatic indicators"

NYC Context

2010 Household Size by NTA
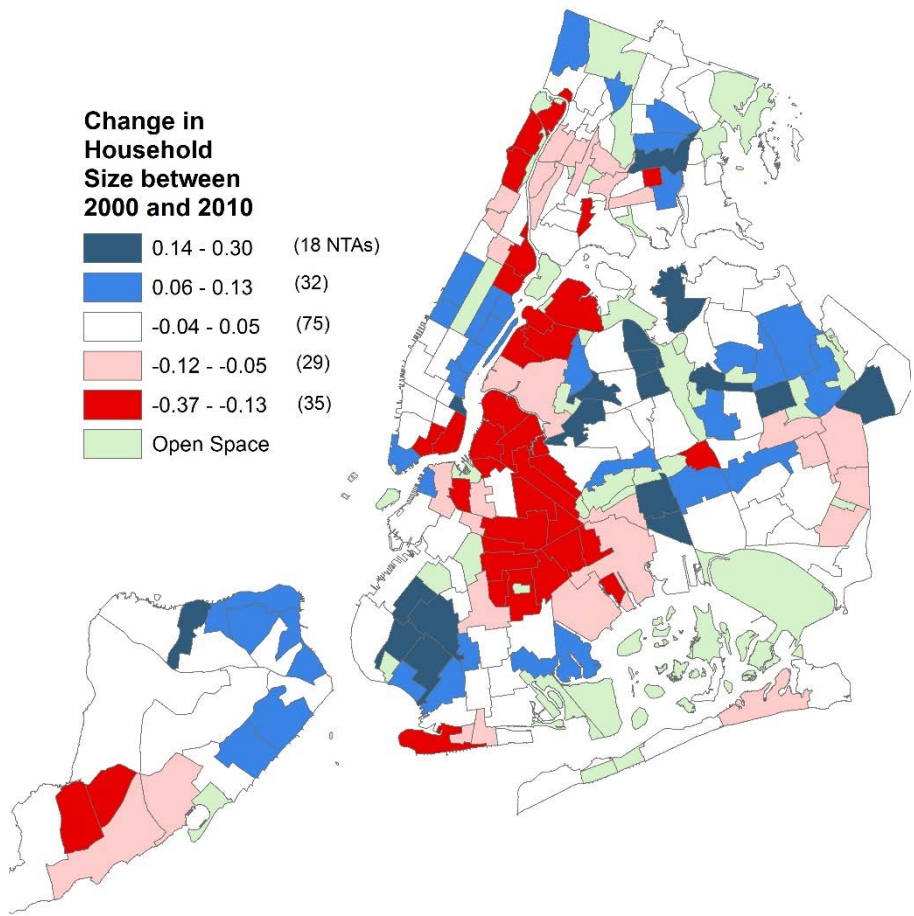**(adjusted for vacancy error in 2010) ***

- 3.22 - 4.29 (18 NTAs)
- 2.84 - 3.21 (51)
- 2.48 - 2.83 (63)
- 2.00 - 2.47 (38)
- 1.57 - 1.99 (17)
- OpenSpace

* Categorization based on natural breaks
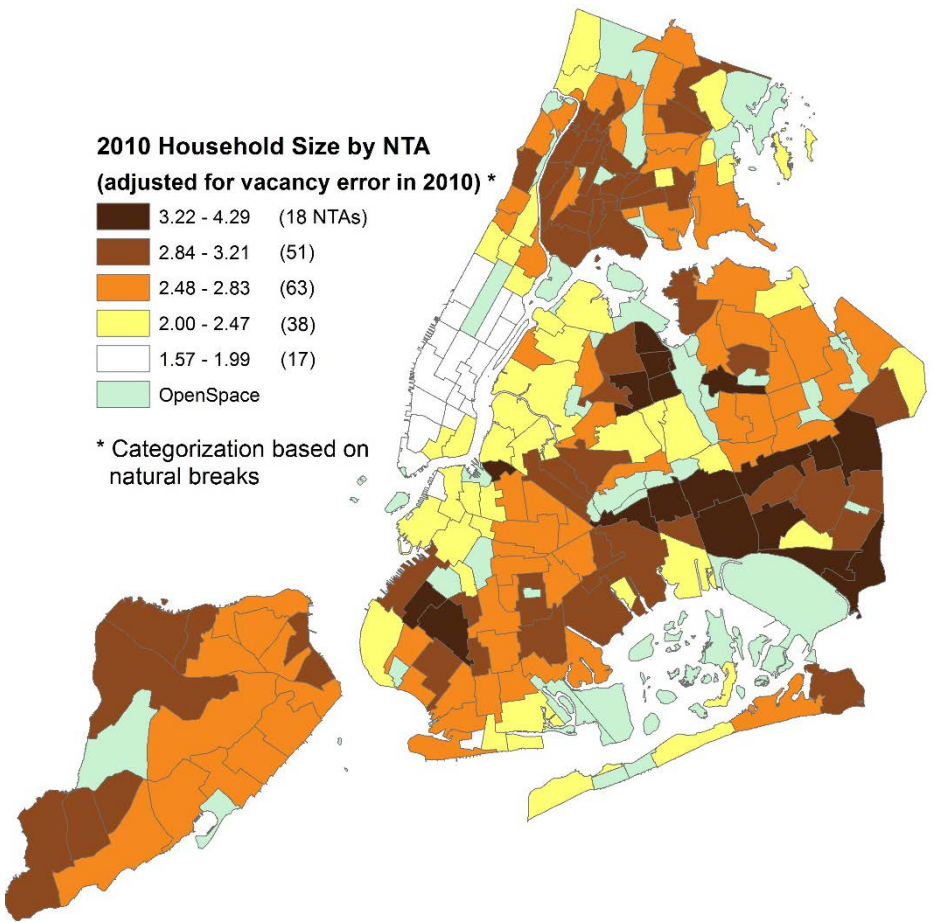
Change in Household Size between 2000 and 2010

- 0.14 - 0.30 (18 NTAs)
- 0.06 - 0.13 (32)
- -0.04 - 0.05 (75)
- -0.12 - -0.05 (29)
- -0.37 - -0.13 (35)
- Open Space

Methodology: Analytical Approach

Step 1. Two-Panel Difference Model

$$\text{PpH}_{i,2010} = \beta_0 + \text{PpH}_{i,2000} + \beta_1 \frac{X_{i,2010}^{(1)} - X_{i,2000}^{(1)}}{X_{i,2000}^{(1)}} + \cdots \beta_9 \frac{X_{i,2010}^{(9)} - X_{i,2000}^{(9)}}{X_{i,2000}^{(9)}} + \epsilon_i$$

➢ Data and Variables for 2000-2010 Change

Elderly %

Head of Non-Family Household under 65 %

Non-Hispanic White %

Non-Hispanic Black %

Non-Hispanic Asian %

Hispanic %

Renter-Households %

Housing Additions 2000-2010

Birth Rates

Methodology: Analytical Approach

Step 1. Two-Panel Difference Model

$$\text{PpH}_{i,2010} = \beta_0 + \text{PpH}_{i,2000} + \beta_1 \frac{X_{i,2010}^{(1)} - X_{i,2000}^{(1)}}{X_{i,2000}^{(1)}} + \cdots \beta_9 \frac{X_{i,2010}^{(9)} - X_{i,2000}^{(9)}}{X_{i,2000}^{(9)}} + \epsilon_i$$

➢ Data and Variables for 2000-2010 Change

 Elderly %

 Head of Non-Family Household under 65 %

 Non-Hispanic White %

 Non-Hispanic Black %

 Non-Hispanic Asian %

 Hispanic %

 Renter-Households %

 Housing Additions 2000-2010

 Birth Rates

Methodology: Analytical Approach

Step 1. **Two-Panel Difference Model**

$$\text{PpH}_{i,2010} = \beta_0 + \text{PpH}_{i,2000} + \beta_1 \frac{X_{i,2010}^{(1)} - X_{i,2000}^{(1)}}{X_{i,2000}^{(1)}} + \cdots \beta_9 \frac{X_{i,2010}^{(9)} - X_{i,2000}^{(9)}}{X_{i,2000}^{(9)}} + \epsilon_i$$

Step 2. **Projecting Household Size for 2013-2017**

- The coefficients ß$_1$ - ß$_9$ are applied to housing and population changes from 2008-2012 to 2013-2017

Methodology: Analytical Approach

Step 1. Two-Panel Difference Model

$$\text{PpH}_{i,2010} = \beta_0 + \text{PpH}_{i,2000} + \beta_1 \frac{X_{i,2010}^{(1)} - X_{i,2000}^{(1)}}{X_{i,2000}^{(1)}} + \cdots \beta_9 \frac{X_{i,2010}^{(9)} - X_{i,2000}^{(9)}}{X_{i,2000}^{(9)}} + \epsilon_i$$
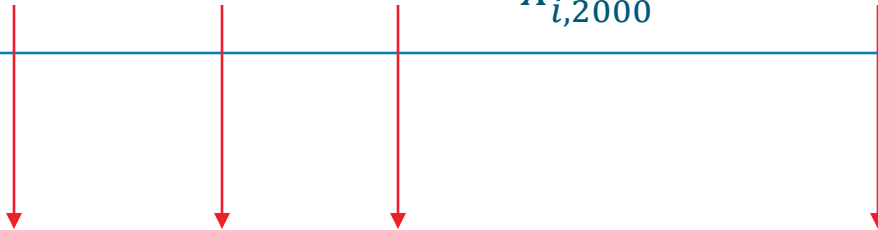
Step 2. Projecting 2014 Household Size

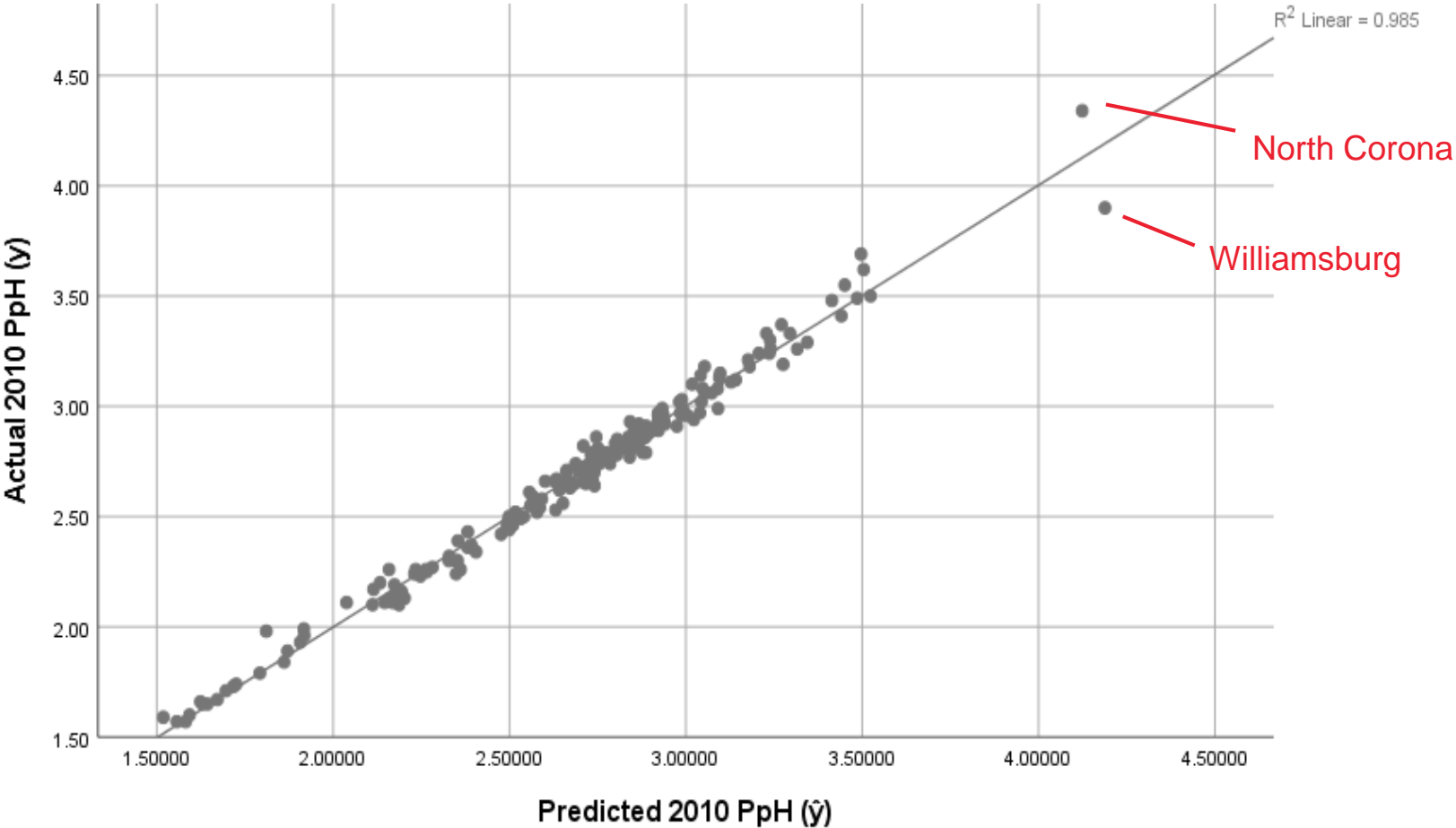- The coefficients ß1 - ß9 are applied to housing and population changes from 2008-2012 to 2013-2017

OLS Regression Results

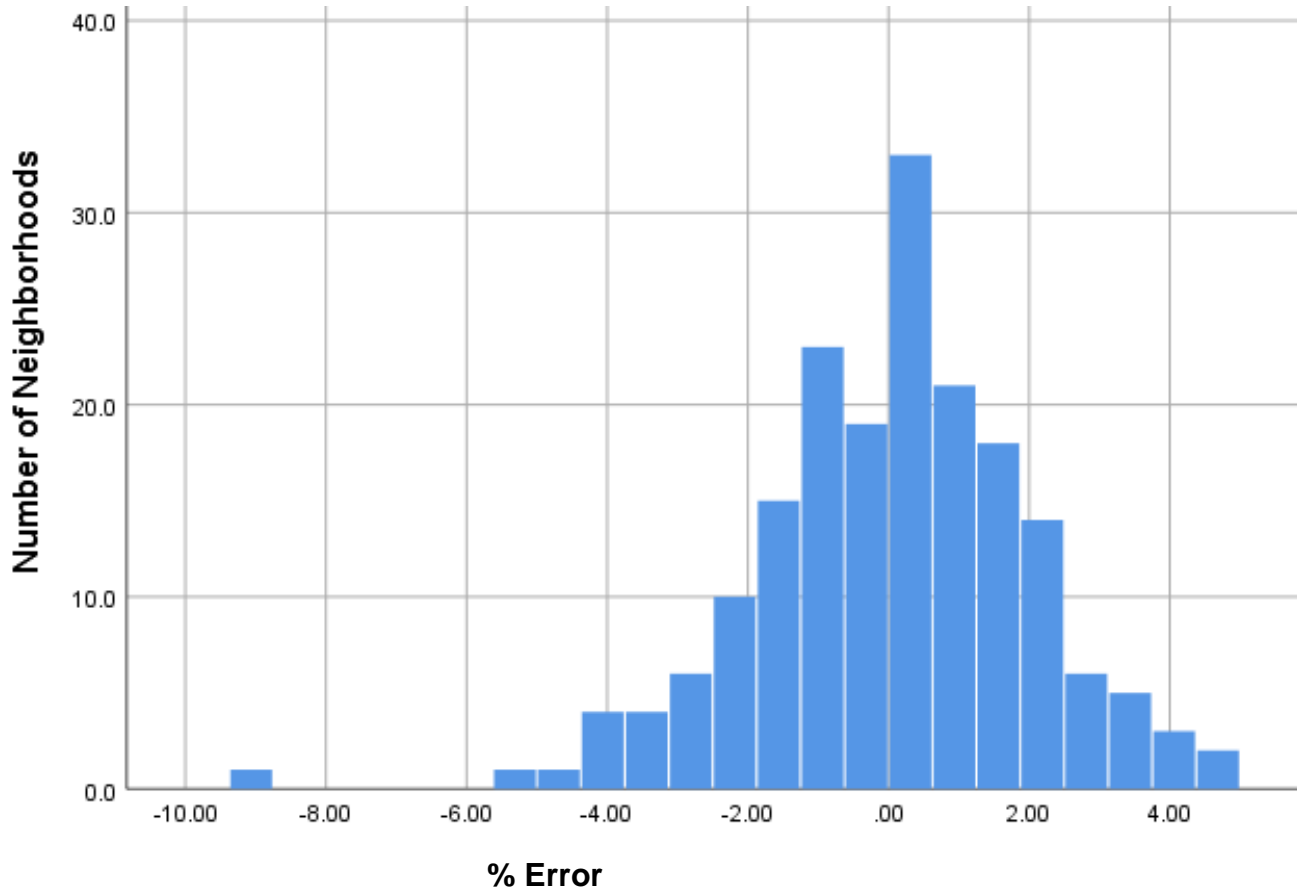|	Unstandardized Coefficients		Standardized Coefficients	
R²: 0.985	**B**	**Std. Error**	**Beta**	**Sig.**
(Constant)	2.682	.008		.000
PpH	.980	.011	.981	.000
△ Elderly 2000-2010	-.028	.003	-.101	.000
△ Hispanic 2000-2010	.006	.001	.069	.000
△ Asian 2000-2010	.008	.001	.081	.000
△ Birthrate 2000-2010	.003	.002	.013	.174
△ Renter 2000-2010	.004	.002	.028	.013
△ Head of Non-Family Household under 65	-1.987	.159	-.141	.000

Error Percentages 1

MALPE: 0.012
MAPE: 1.152

Error Percentages 2

MALPE: 0.010
MAPE: 1.147

ACS
2.67

2.59

2.57

Projected
2.59

Decennial
Census

2.59

2000

2010
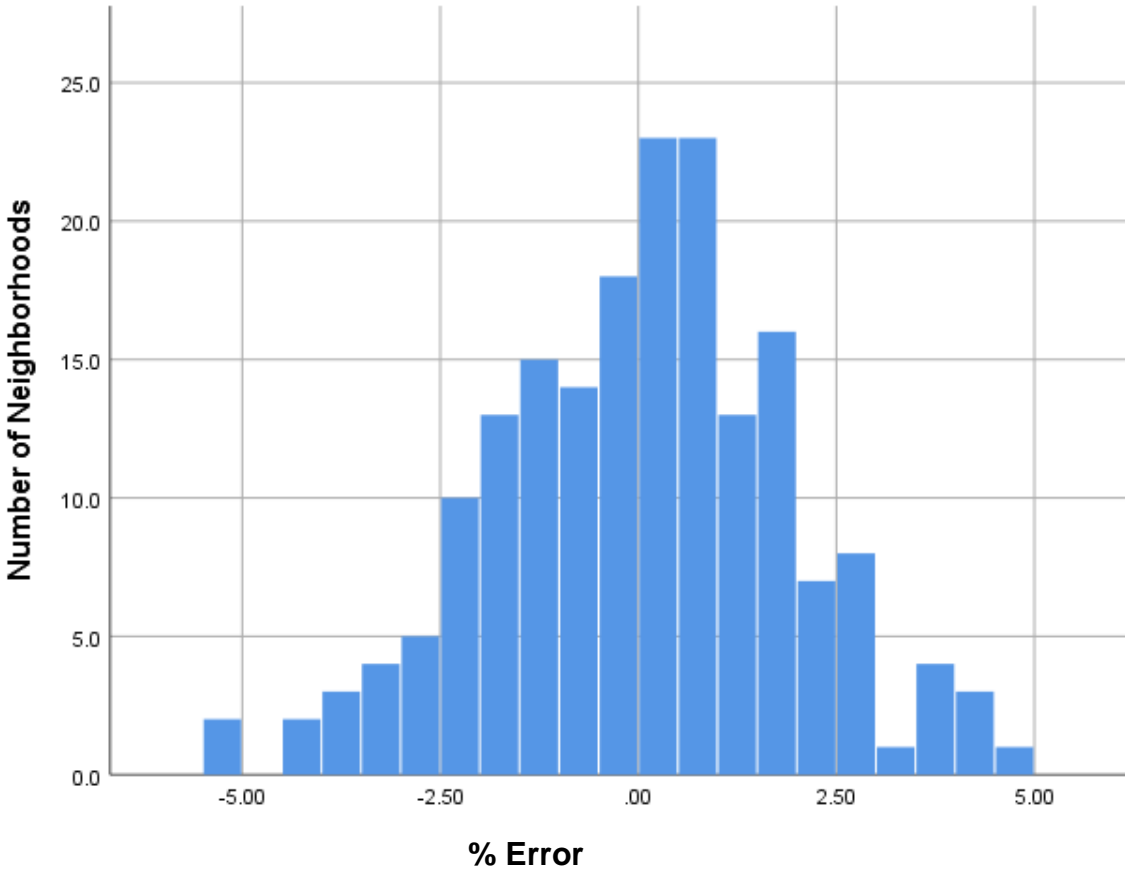
2009　2010　2011　2012　2013　2014　2015　2016　2017 *
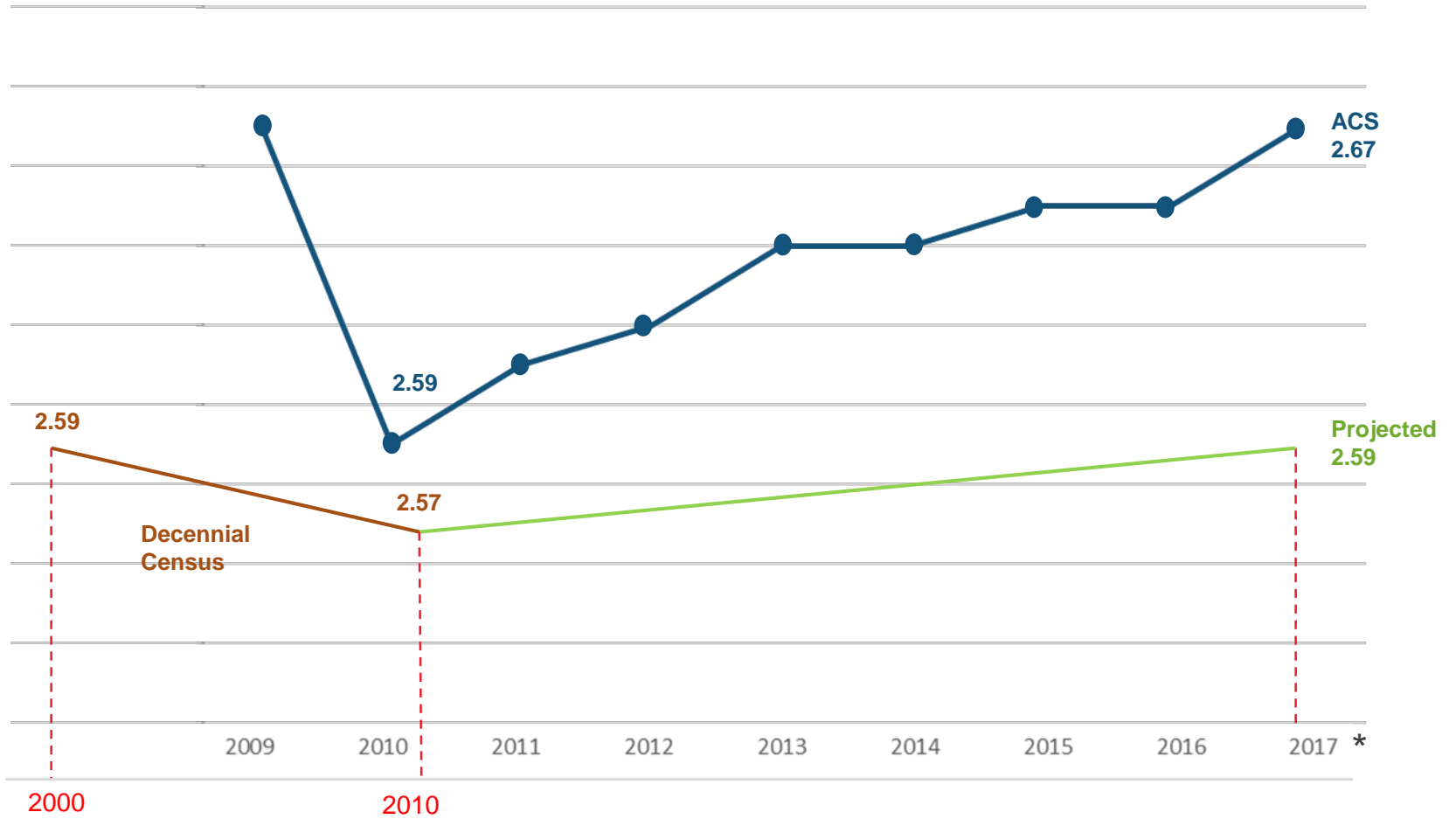
* The 2009 to 2017 axis represents the end-points for 5-year periods.

Conclusion

- Determine if the error distribution is strongly influenced by outliers and what to do with outliers
  - Statistical cutoffs to identify extreme outliers

- Overall, very predictive model

- No one perfect method for population estimates — instead, different procedures need to be used in conjunction

Contact:

Annette Jacoby

ajacoby@planning.nyc.gov