An Evaluation of Persons Per Household (PPH) Estimates Generated By The American Community Survey: A Demographic Perspective

Applied Demography Conference
San Antonio, Texas USA (8-10 January 2012)

Session 1D (0920 – 1020)
Monday, January 9th
Improving Demographic Measurement Through Sample Estimates from Surveys and/or Administrative Sources

David A. Swanson
University of California Riverside
Riverside, CA 92151 USA
David.swanson@ucr.edu

George C. Hough, Jr.
Rice University
Houston, TX 77005 USA
george.hough@rice.edu

OVERVIEW

The American Community Survey (ACS) is a U.S. Census Bureau product designed to provide accurate and timely demographic and economic indicators on an annual basis for both large and small geographic areas within the United States. Operational plans call for ACS to serve not only as a substitute for the decennial census long-form, but as a means of providing annual data at the national, state, county, and sub-county levels. In addition to being highly ambitious, this approach represents a major change in how data are collected and interpreted.

OVERVIEW

Two of the major questions facing the ACS are its functionality and usability. This paper explores the latter of these two questions by examining “Persons Per Household (PPH),” a variable of high interest to demographers and others preparing regular post-censal population estimates. The data used in this exploration are taken from 18 of the counties that formed the set of 1999 ACS test sites.

OVERVIEW

The examination proceeds by first comparing 1-Year ACS PPH estimates to Census 2010 PPH values along with extrapolated estimates generated using a geometric model based on PPH change between the 1990 and 2000 census counts. Both sets of estimates are then compared to annual 2001-2009 PPH interpolated estimates generated by a geometric model based on PPH from the 2000 census to the 2010 census.

OVERVIEW

The ACS PPH estimates represent what could be called the “statistical perspective” because variations in the estimates of specific variables over time and space are viewed largely by statisticians with an eye toward sample error.

The model-based PPH estimates represent a “demographic perspective” because PPH estimates are largely viewed by demographers as varying systematically and changing relatively slowly over time, an orientation stemming from theory and empirical evidence that PPH estimates respond to demographic and related determinants.

OVERVIEW

The comparisons suggest that the ACS PPH estimates exhibit too much “noisy” variation for a given area over time to be usable by demographers and others preparing post-censal population estimates. These findings should be confirmed through further analysis and suggestions are provided for the directions this research could take.

We conclude by noting that the statistical and demographic perspectives are not incompatible and that one of the aims of our paper is to encourage the U.S. Census Bureau to consider ways to improve the usability of the 1-Year ACS PPH estimates.

Housing Unit Method (HUM)

The Housing Unit Method (HUM) is a “stock” method that describes a basic identity in the same way that the fundamental demographic equation does. In the case of the HUM, this identity is usually given as

$$P = H*(1-VR)*PPH + GQ$$

where
P = Population,
H = Housing units,
VR= Vacancy Rate (Proportion Vacant),
PPH = Average number of persons per household, and
GQ = Population residing in “group quarters” and the homeless.

Housing Unit Method (HUM)

Like the fundamental demographic identity (aka the "balancing equation"), the HUM equation can be expressed in less detail, i.e., ‘

$$P = HH*PPH + GQ,$$

where $HH=H*(1-VR)$,

or more detail - by structure type, for example.

It also can be used in combination with sample data, which opens the door to developing measures of statistical uncertainty for the estimates so produced.

DATA & METHODS

The ACS data used in this exploration are taken from 18 counties that were in the 1999 ACS test sites . These 18 counties represent the smallest pieces of geography for which ACS PPH data are available for the entire inter-censal period, 2000 to 2010.

Exhibit 1

Pima County, AZ	Madison County, MS
Jefferson County, AR	Douglas County, NE
San Francisco County, CA	Bronx County, NY
Tulare County, CA	Rockland County, NY
Broward County, FL	Franklin County, OH
Lake County, IL	Multnomah County, OR
Black Hawk County, IA	Schuylkill County, PA
Calvert County, MD	Sevier County, TN
Hampden County, MA	Yakima County, WA

DATA & METHODS

The ACS data used in this exploration are taken from 18 counties that were in the 1999 ACS test sites (See Exhibit 1). These 18 counties represent the smallest pieces of geography for which ACS PPH data are available for the entire inter-censal period, 2000 to 2010. The examination proceeds in four phases. In the first phase, we examine the accuracy of PPH estimates extrapolated from a geometric model. Here, we construct models from 1980 and 1990 census data for each of the 39 counties of Washington state and then compare the 2000 PPH estimates extrapolated from these county-specific models to the 2000 Census PPH values. Second, we compare single-year (1-Year) 2010 ACS PPH estimates for these 18 counties to the 2010 Census PPH values. Third, we compare PPH estimates extrapolated from a geometric model based on PPH change from Census 1990 to Census 2000 to Census 2010 PPH values. In the fourth and final phase we compare the accuracy of the 1-Year ACS PPH for 2001 to 2009 to PPH estimates extrapolated from the 1990-2000 based geometric model for the same years as well as PPH estimates interpolated from a 2000-2010 based geometric model.

PPH: STATISTICAL PERSPECTIVE

The ACS PPH estimates represent what could be called the "statistical perspective" because variations in the estimates of specific variables over time and space are viewed by statisticians with an eye toward sample error. Applied to an on-going survey such as the ACS, this implies that fluctuations over time are not necessarily viewed with alarm because they are due to statistical uncertainty.

In addition, In terms of summarizing a variable derived from sample data, the statistical perspective is oriented toward a range of values for an estimate (i.e., its upper and lower confidence bounds) based on sample error. That is, it is oriented toward "interval" estimates.

PPH: DEMOGRPAHIC PERSPECTIVE

The model-based PPH estimates represent a “demographic perspective” because PPH estimates are viewed by demographers as not likely to change abruptly over time. Instead, they are viewed as changing slowly over time, an orientation stemming from theory and empirical evidence that PPH estimates respond to a constellation of demographic and related determinants that taken as a whole changes slowly over time. As a consequence, even in the face of statistical uncertainty, demographers view abrupt changes in PPH over short periods of time as a problem.

In addition, the demographic perspective is oriented toward a single “value” for an estimate of a given variable in terms of summarization. That is, it is oriented toward “point” estimates.

TABLE 1. POPULATION, HOUSING AND ACS SAMPLE SIZE INFORMATION FOR THE 18 COUNTIES

ACS TEST SITE	CENSUS POPULATION		PERCENT CHANGE IN POPULATION	CENSUS HOUSING UNITS		PERCENT CHANGE IN HOUSING UNITS	N OF HOUSING UNITS INTERVIEWED IN 1-YEAR ACS	RATIO OF INTERVIEWED HOUSING UNITS TO CENSUS HOUSING UNITS
COUNTY & STATE	2000	2010	2000-2010	2000	2010	2000-2010	2010	2010
PIMA AZ	843,746	980,263	16.18%	366,737	440,909	20.22%	5,388	0.0122
JEFFERSON AR	84,278	77,435	-8.12%	34,350	33,006	-3.91%	450	0.0136
SAN FRANCISCO CA	776,733	805,235	3.67%	346,527	376,942	8.78%	4,393	0.0117
TULARE CA	368,021	442,179	20.15%	119,639	141,696	18.44%	2,136	0.0151
BROWARD FL	1,623,018	1,748,066	7.70%	741,043	810,388	9.36%	8,726	0.0108
LAKE IL	644,356	703,462	9.17%	225,919	260,310	15.22%	3,853	0.0148
BLACK HAWK IA	128,012	131,090	2.40%	51,759	55,887	7.98%	916	0.0164
CALVERT MD	74,563	88,737	19.01%	27,576	33,780	22.50%	440	0.0130
HAMPDEN MA	456,228	463,490	1.59%	185,876	192,175	3.39%	2,605	0.0136
MADISON MS	74,674	95,203	27.49%	28,781	38,558	33.97%	494	0.0128
DOUGLAS NE	463,585	517,110	11.55%	192,672	219,580	13.97%	3,261	0.0149
BRONX NY	1,332,650	1,385,106	3.94%	490,659	511,896	4.33%	5,753	0.0112
ROCKLAND NY	286,753	311,687	8.70%	94,973	104,057	9.56%	1,465	0.0141
FRANKLIN OH	1,068,978	1,163,414	8.83%	471,016	527,186	11.93%	6,878	0.0130
MULTNOMAH OR	660,486	735,334	11.33%	288,561	324,832	12.57%	4,443	0.0137
SCHUYKILL PA	150,336	148,289	-1.36%	67,806	69,323	2.24%	1,608	0.0232
SEVIER TN	71,170	89,889	26.30%	37,252	55,918	50.11%	537	0.0096
YAKIMA WA	222,581	243,231	9.28%	79,174	85,474	7.96%	1,039	0.0122

Table 2. **Accuracy Test of the Geometric Method of Estimating PPH Estimates for Counties: Washington State 2000**

Washington State PPH Values By County, 1980, 1990, and 2000

	1980 Persons Per Household	1990 Persons Per Household	2000 Persons Per Household	1980-90 Geometric Rate of Change	Estimated 2000 Persons Per Household	Absolute Error	Percent Error	Mape
STATE	2.609	2.535	2.535	-0.003	2.463	-0.072	-2.83%	2.83%
Adams	2.911	2.941	3.095	0.001	2.970	-0.125	-4.03%	4.03%
Asotin	2.566	2.473	2.416	-0.004	2.383	-0.034	-1.39%	1.39%
Benton	2.797	2.652	2.680	-0.005	2.514	-0.166	-6.19%	6.19%
Chelan	2.483	2.486	2.619	0.000	2.490	-0.129	-4.93%	4.93%
Clallam	2.537	2.401	2.307	-0.006	2.271	-0.035	-1.53%	1.53%
Clark	2.763	2.663	2.690	-0.004	2.566	-0.124	-4.61%	4.61%
Columbia	2.525	2.437	2.363	-0.004	2.351	-0.012	-0.49%	0.49%
Cowlitz	2.662	2.559	2.553	-0.004	2.460	-0.093	-3.66%	3.66%
Douglas	2.759	2.677	2.755	-0.003	2.597	-0.158	-5.74%	5.74%
Ferry	2.857	2.698	2.494	-0.006	2.548	0.054	2.16%	2.16%
Franklin	2.882	3.034	3.264	0.005	3.194	-0.069	-2.12%	2.12%
Garfield	2.596	2.395	2.391	-0.008	2.210	-0.182	-7.59%	7.59%
Grant	2.799	2.741	2.920	-0.002	2.684	-0.236	-8.09%	8.09%
Grays Harbor	2.597	2.481	2.483	-0.005	2.371	-0.112	-4.49%	4.49%
Island	2.671	2.615	2.522	-0.002	2.560	0.038	1.51%	1.51%
Jefferson	2.454	2.309	2.212	-0.006	2.173	-0.040	-1.79%	1.79%
King	2.487	2.398	2.391	-0.004	2.313	-0.078	-3.25%	3.25%
Kitsap	2.682	2.647	2.601	-0.001	2.612	0.012	0.44%	0.44%
Kittitas	2.398	2.325	2.331	-0.003	2.255	-0.077	-3.29%	3.29%
Klickitat	2.721	2.641	2.536	-0.003	2.563	0.027	1.06%	1.06%
Lewis	2.673	2.600	2.569	-0.003	2.528	-0.041	-1.59%	1.59%
Lincoln	2.573	2.431	2.423	-0.006	2.297	-0.127	-5.22%	5.22%
Mason	2.546	2.516	2.489	-0.001	2.487	-0.002	-0.09%	0.09%
Okanogan	2.667	2.588	2.576	-0.003	2.510	-0.066	-2.56%	2.56%
Pacific	2.447	2.350	2.271	-0.004	2.257	-0.014	-0.62%	0.62%
Pend Oreille	2.809	2.603	2.507	-0.008	2.412	-0.095	-3.80%	3.80%
Pierce	2.659	2.623	2.605	-0.001	2.588	-0.017	-0.64%	0.64%
San Juan	2.295	2.249	2.159	-0.002	2.204	0.045	2.10%	2.10%
Skagit	2.566	2.550	2.603	-0.001	2.534	-0.070	-2.68%	2.68%
Skamania	2.790	2.692	2.612	-0.004	2.598	-0.014	-0.54%	0.54%
Snohomish	2.761	2.679	2.655	-0.003	2.601	-0.054	-2.04%	2.04%
Spokane	2.579	2.475	2.465	-0.004	2.375	-0.090	-3.65%	3.65%
Stevens	2.907	2.732	2.644	-0.006	2.567	-0.077	-2.90%	2.90%
Thurston	2.644	2.553	2.499	-0.004	2.465	-0.034	-1.35%	1.35%
Wahkiakum	2.772	2.476	2.424	-0.011	2.212	-0.213	-8.77%	8.77%
Walla Walla	2.541	2.496	2.539	-0.002	2.451	-0.088	-3.47%	3.47%
Whatcom	2.590	2.532	2.511	-0.002	2.476	-0.035	-1.41%	1.41%
Whitman	2.467	2.387	2.312	-0.003	2.309	-0.002	-0.09%	0.09%
Yakima	2.771	2.804	2.958	0.001	2.837	-0.121	-4.08%	4.08%

COUNTY LEVEL SUMMARY STATISTICS

Mean Error	MAPE	MALPE	N ABS % Error >10
-0.068	2.97%	-2.60%	0

TABLE 3. COMPARISON OF ACS SINGLE YEAR 2010 PPH ESTIMATES TO 2010 CENSUS PPH VALUES

Area	Estimate	Margin of Error	Census	Percent Difference
PIMA, AZ	2.71	0.03	2.46*	10.16%
JEFFERSON, AR	2.48	0.1	2.49	-0.40%
SAN FRANCISCO, CA	2.46	0.03	2.26*	8.85%
TULARE, CA	3.37	0.05	3.36	0.30%
BROWARD, FL	2.67	0.03	2.52*	5.95%
LAKE, IL	2.99	0.04	2.82*	6.03%
BLACK HAWK, IA	2.36	0.07	2.38	-0.84%
CALVERT, MD	2.94	0.11	2.85	3.16%
HAMPDEN, MA	2.56	0.03	2.49*	2.81%
MADISON, MS	2.64	0.05	2.61	1.15%
DOUGLAS, NE	2.53	0.03	2.49*	1.61%
BRONX, NY	2.82	0.02	2.77*	1.81%
ROCKLAND, NY	3.02	0.05	3.07	-1.63%
FRANKLIN, OH	2.47	0.03	2.38*	3.78%
MULTNOMAH, OR	2.38	0.03	2.35	1.28%
SCHUYLKILL, PA	2.27	0.06	2.35*	-3.40%
SEVIER, TN	2.74	0.15	2.52*	8.73%
YAKIMA, WA	2.93	0.06	2.97	-1.35%
			MAPE	**3.51%**

* Census PPH value is outside of the 90% margin of error of the ACS PPH Estimate.

TABLE 4. COMPARISON OF EXTRAPOLATED GEOMETRIC MODEL-BASED 2010 PPH ESTIMATES TO 2010 CENSUS PPH ESTIMATES*

Area	Estimate	Census	Percent Difference
PIMA, AZ	2.45	2.46	-0.40%
JEFFERSON, AR	2.48	2.49	-0.22%
SAN FRANCISCO, CA	2,31	2.26	-2.21%
TULARE, CA	3.45	3.36	2.63%
BROWARD, FL	2.55	2.52	1.36%
LAKE, IL	2.91	2.82	3.20%
BLACK HAWK, IA	2.39	2.38	0.48%
CALVERT, MD	2.81	2.85	-1.29%
HAMPDEN, MA	2.44	2.49	-1.91%
MADISON, MS	2.60	2.61	-0.31%
DOUGLAS, NE	2.43	2.49	-2.37%
BRONX, NY	2.82	2.77	1.83%
ROCKLAND, NY	2.99	3.07	-2.60%
FRANKLIN, OH	2.31	2.38	-2.83%
MULTNOMAH, OR	2.38	2.35	1.28%
SCHUYLKILL, PA	2.25	2.35	-4.05%
SEVIER, TN	2.38	2.52	-5.40%
YAKIMA, WA	3.13	2.97	5.36%
		MAPE	**2.25%**

***Models are county specific and based on 1990-2000 trends in PPH values**

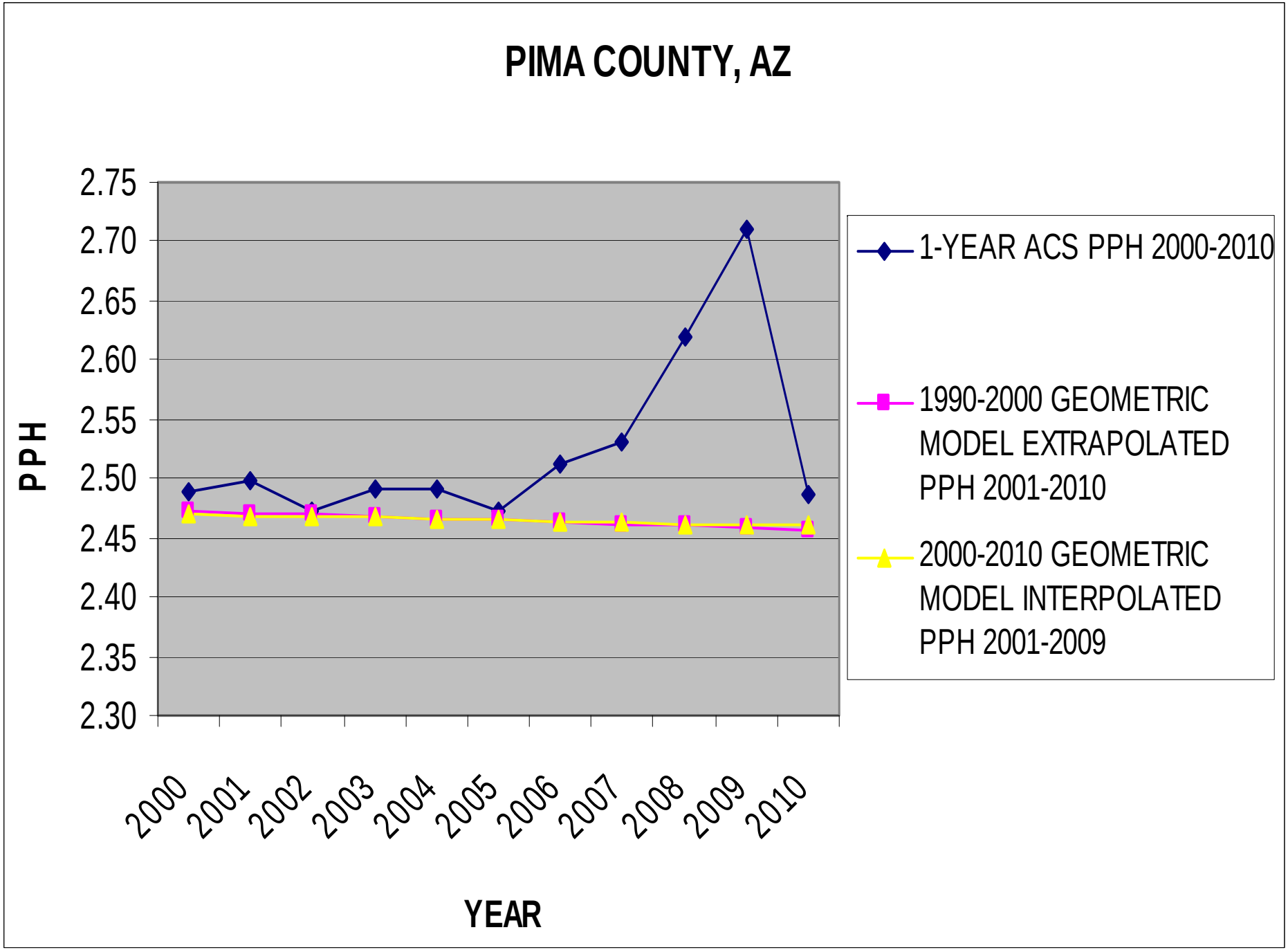
Results

We also used the census 2010 PPH values as a basis for comparing the accuracy of the 1-Year 2010 ACS PPH estimates to the accuracy of PPH estimates generated by the geometric method for all of the 807 counties for which ACS data are available. The latter were developed in the same manner as the estimates discussed in Table 2: the 1990 to 2000 trends in PPH values were extrapolated to 2010 using the geometric model.

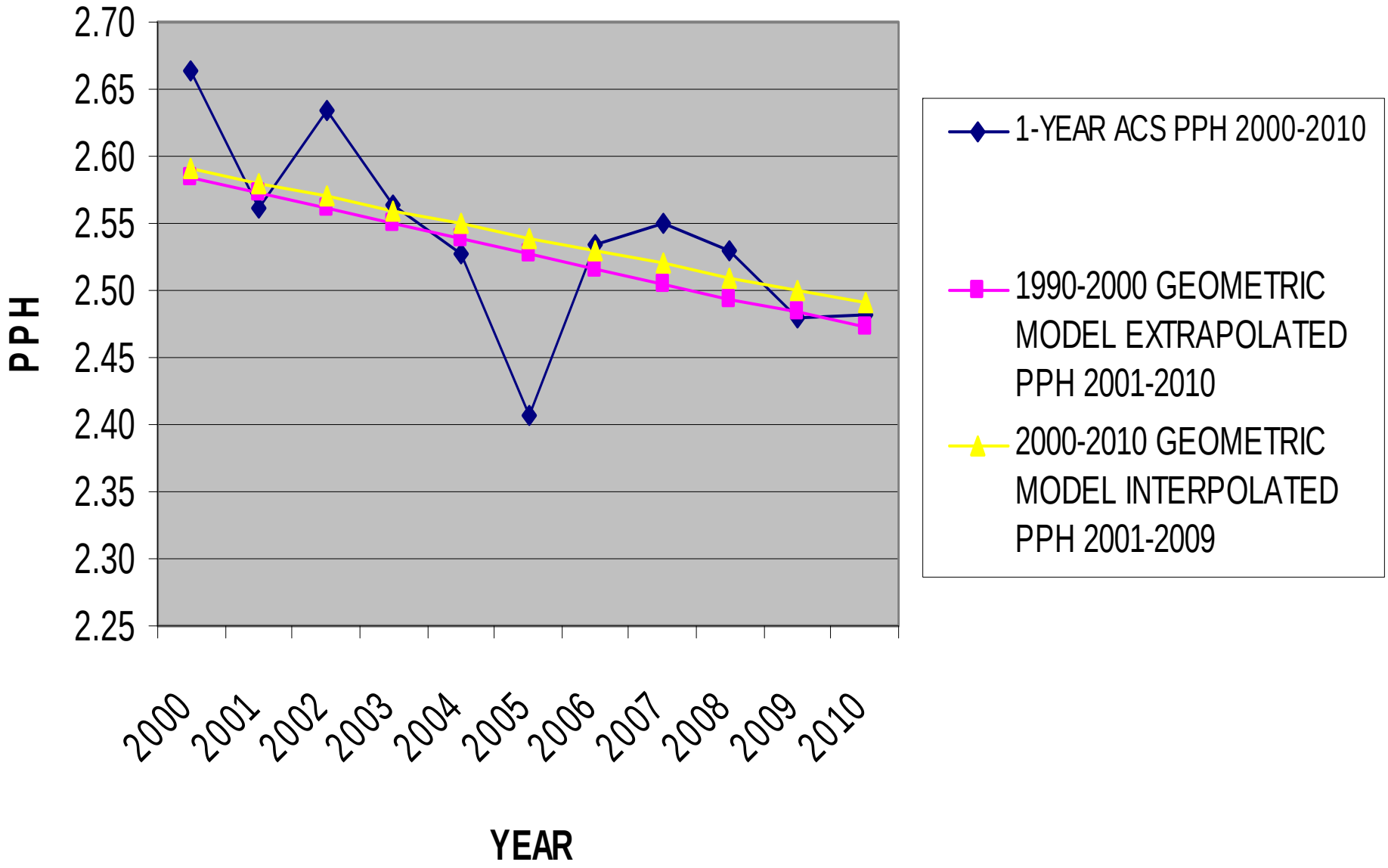
Results

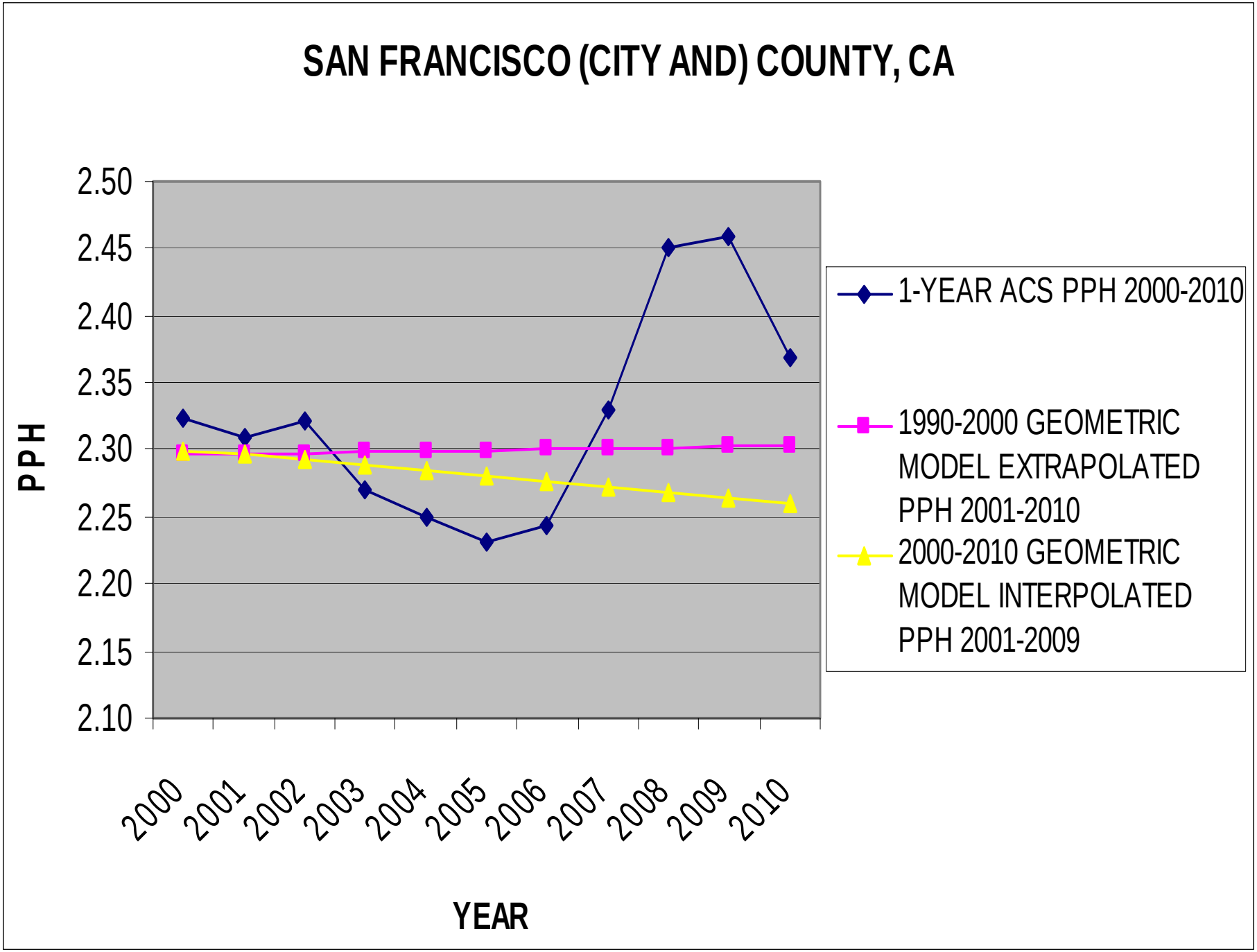
At 6.85%, the Mean Absolute Percent Error (MAPE) of the ACS PPH estimates is higher than the MAPE for the geometric model, 5.83%, indicating that the ACS is less accurate than the geometric model not only for the 18 test counties, but for all counties.

We also found that the 90 percent margins of error provided by the Census Bureau for the 2010 1-Year ACS PPH estimates contained the 2010 census PPH values in only 64% (515) of the 807 counties. This is a better showing than the 39% observed for the 18 test counties, but one would intuitively expect it to be higher than 64% for the entire universe of ACS counties in that 90% margins of error are used.

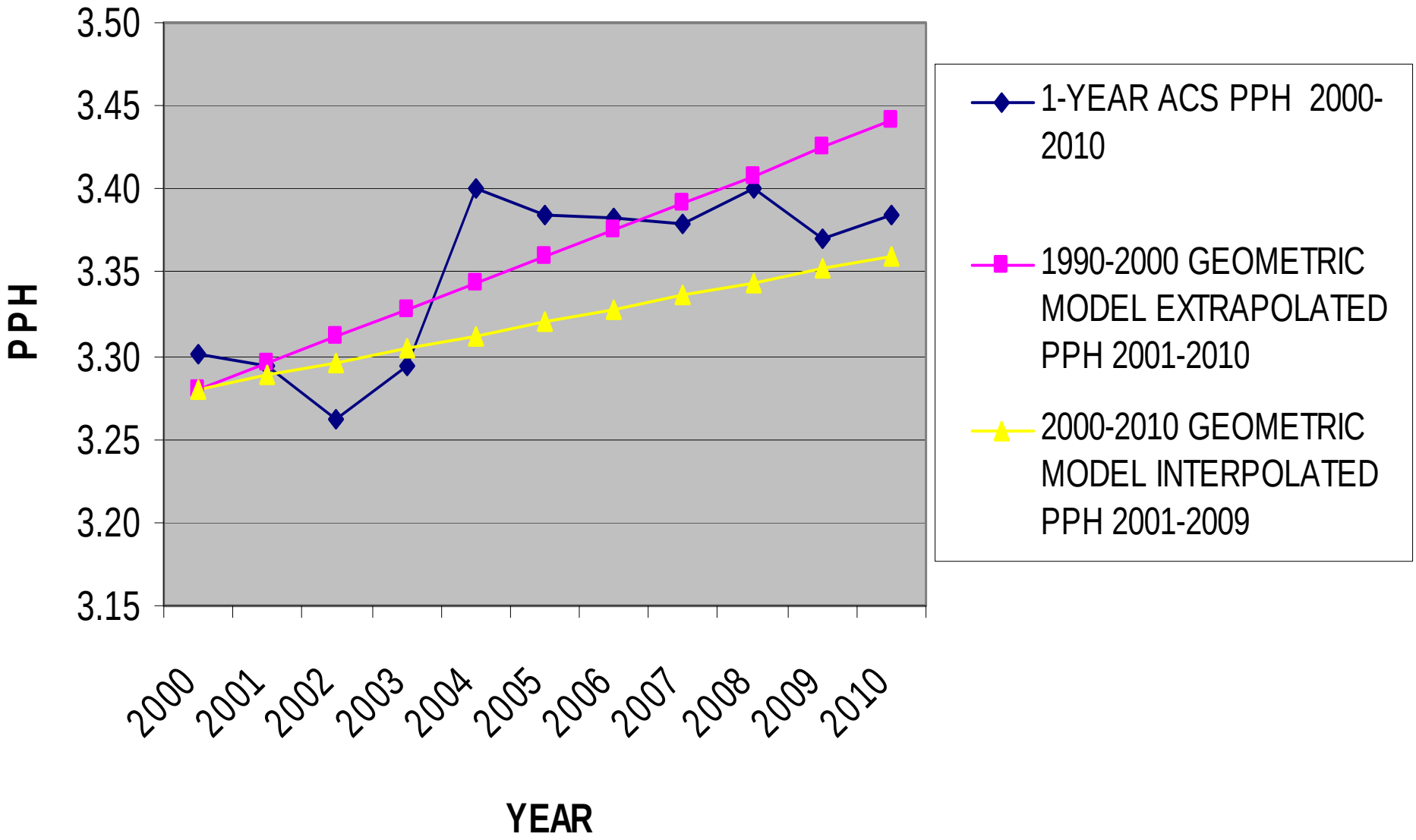

PIMA COUNTY, AZ
PPH
2.75
2.70
2.65
2.60
2.55
2.50
2.45
2.40
2.35
2.30
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
YEAR
1-YEAR ACS PPH 2000-2010
1990-2000 GEOMETRIC MODEL EXTRAPOLATED PPH 2001-2010
2000-2010 GEOMETRIC MODEL INTERPOLATED PPH 2001-2009

JEFFERSON COUNTY, AR

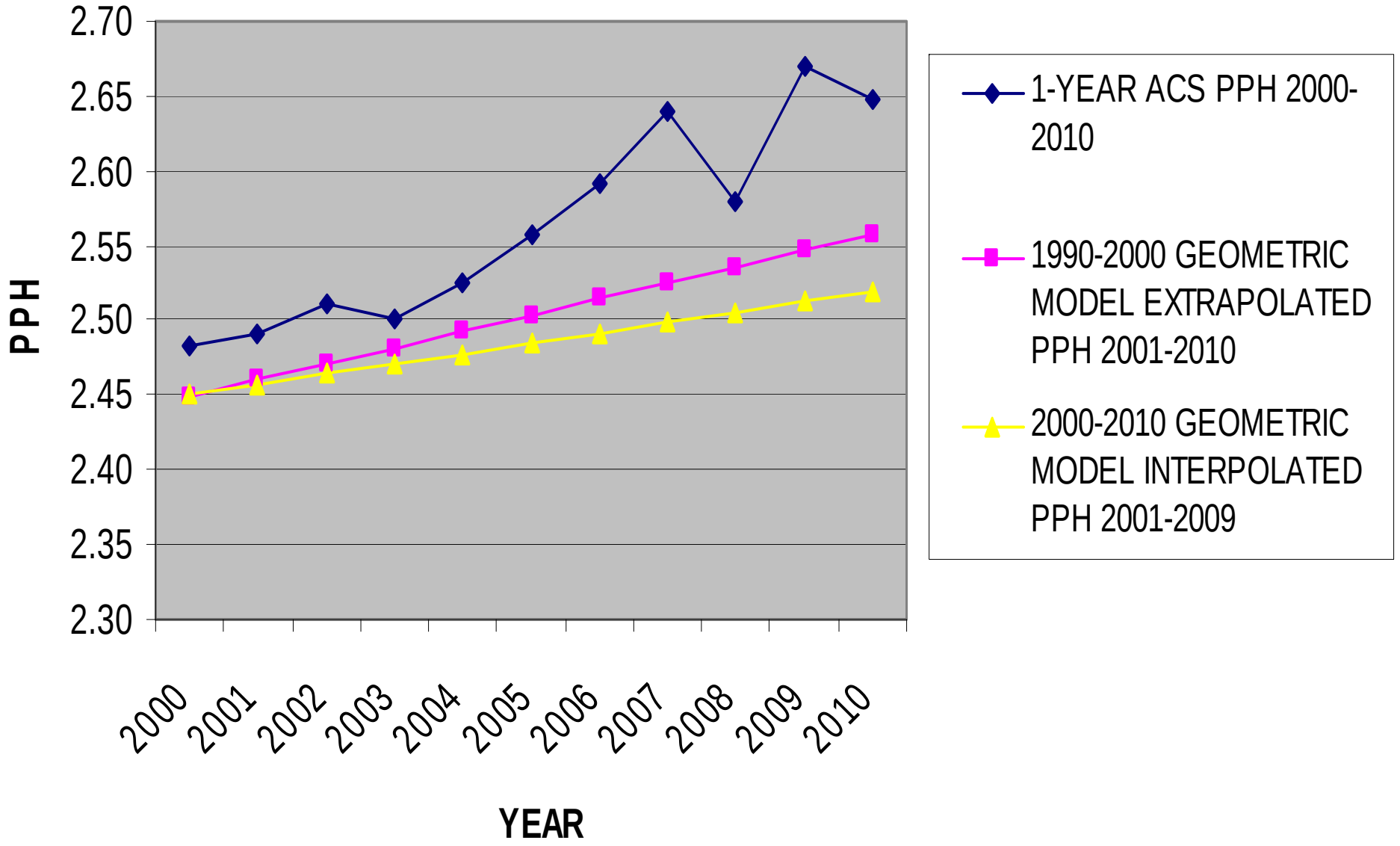


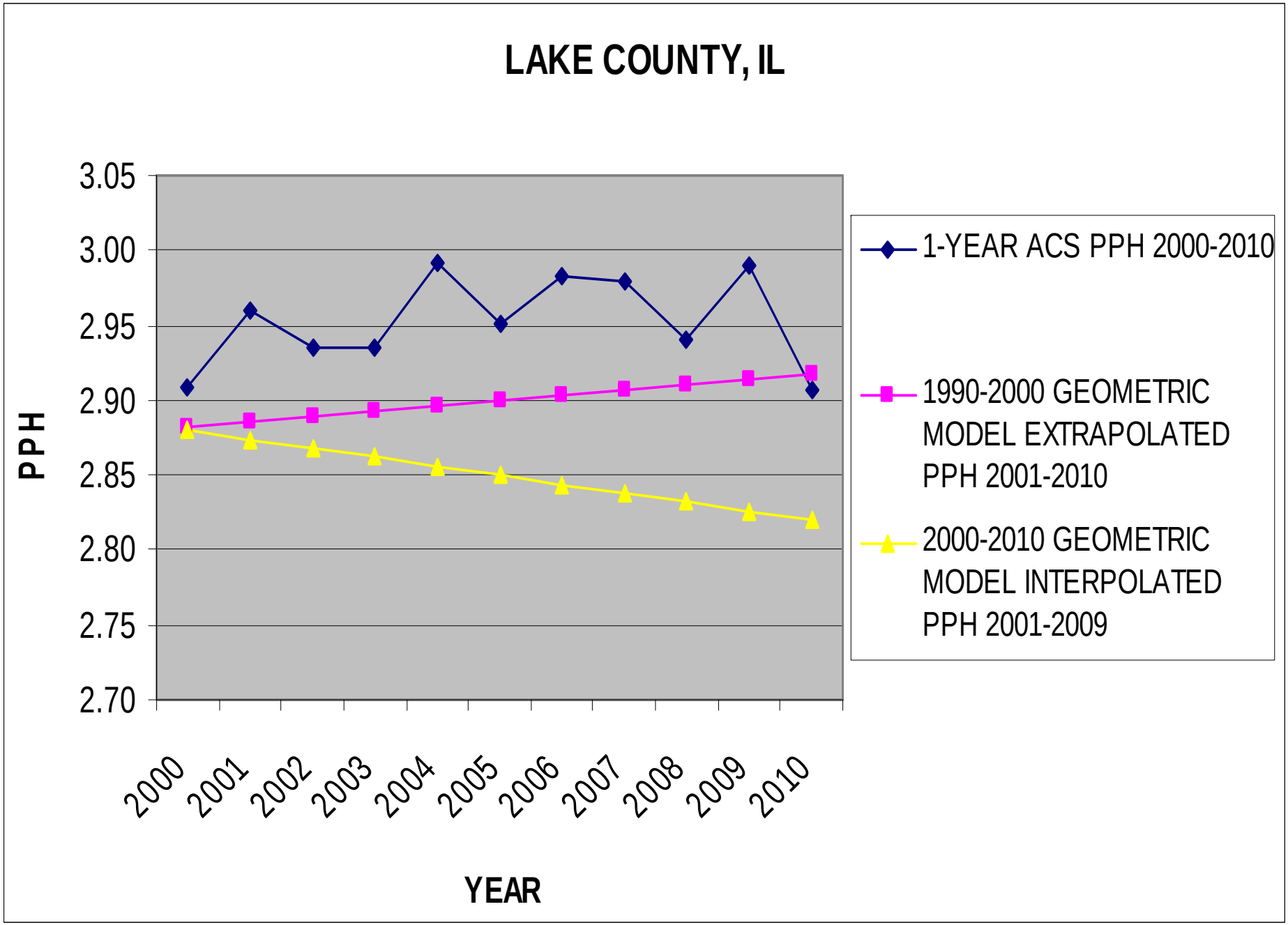

SAN FRANCISCO (CITY AND) COUNTY, CA
PPH
2.50
2.45
2.40
2.35
2.30
2.25
2.20
2.15
2.10
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
YEAR
1-YEAR ACS PPH 2000-2010
1990-2000 GEOMETRIC MODEL EXTRAPOLATED PPH 2001-2010
2000-2010 GEOMETRIC MODEL INTERPOLATED PPH 2001-2009

TULARE COUNTY, CA

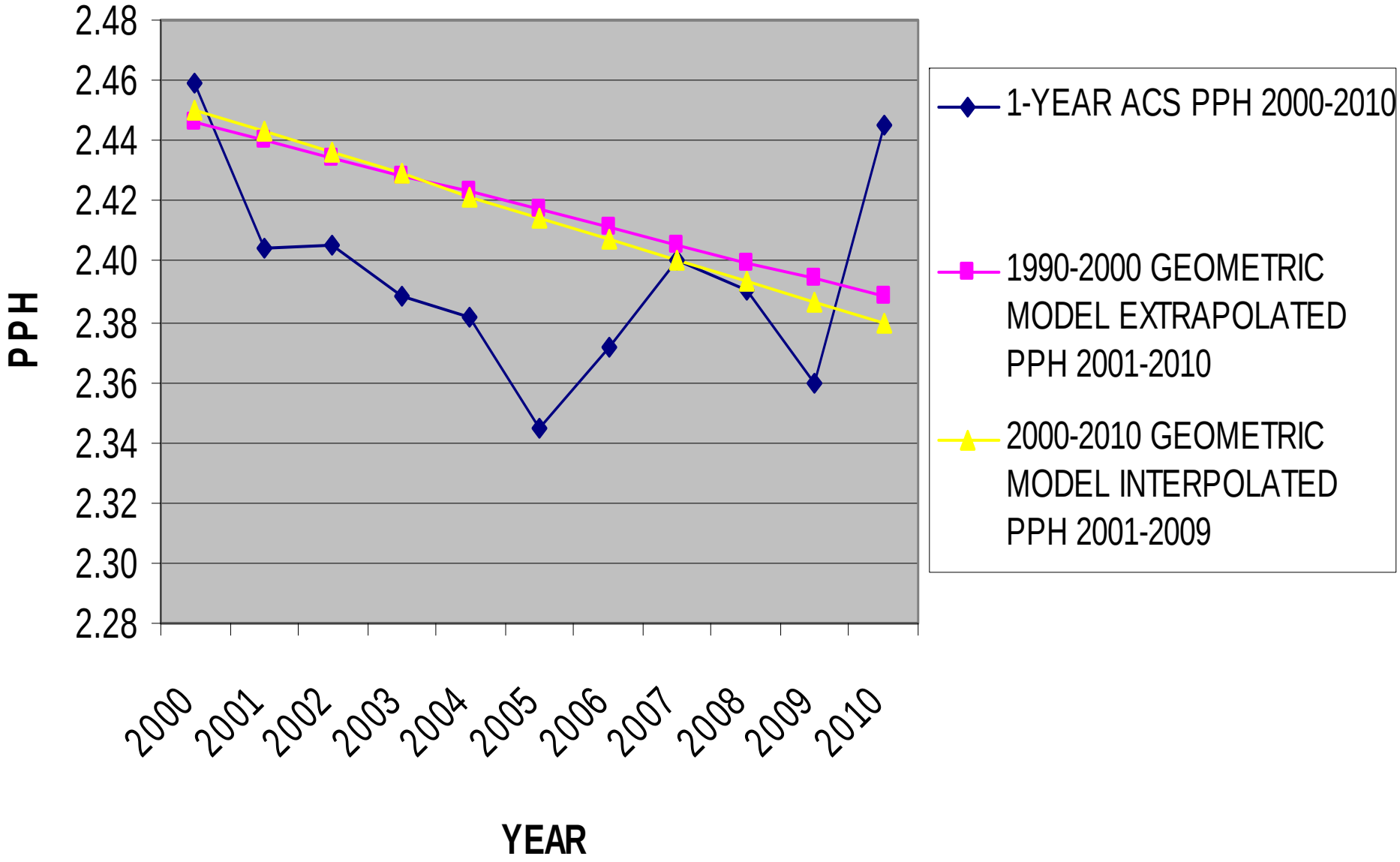


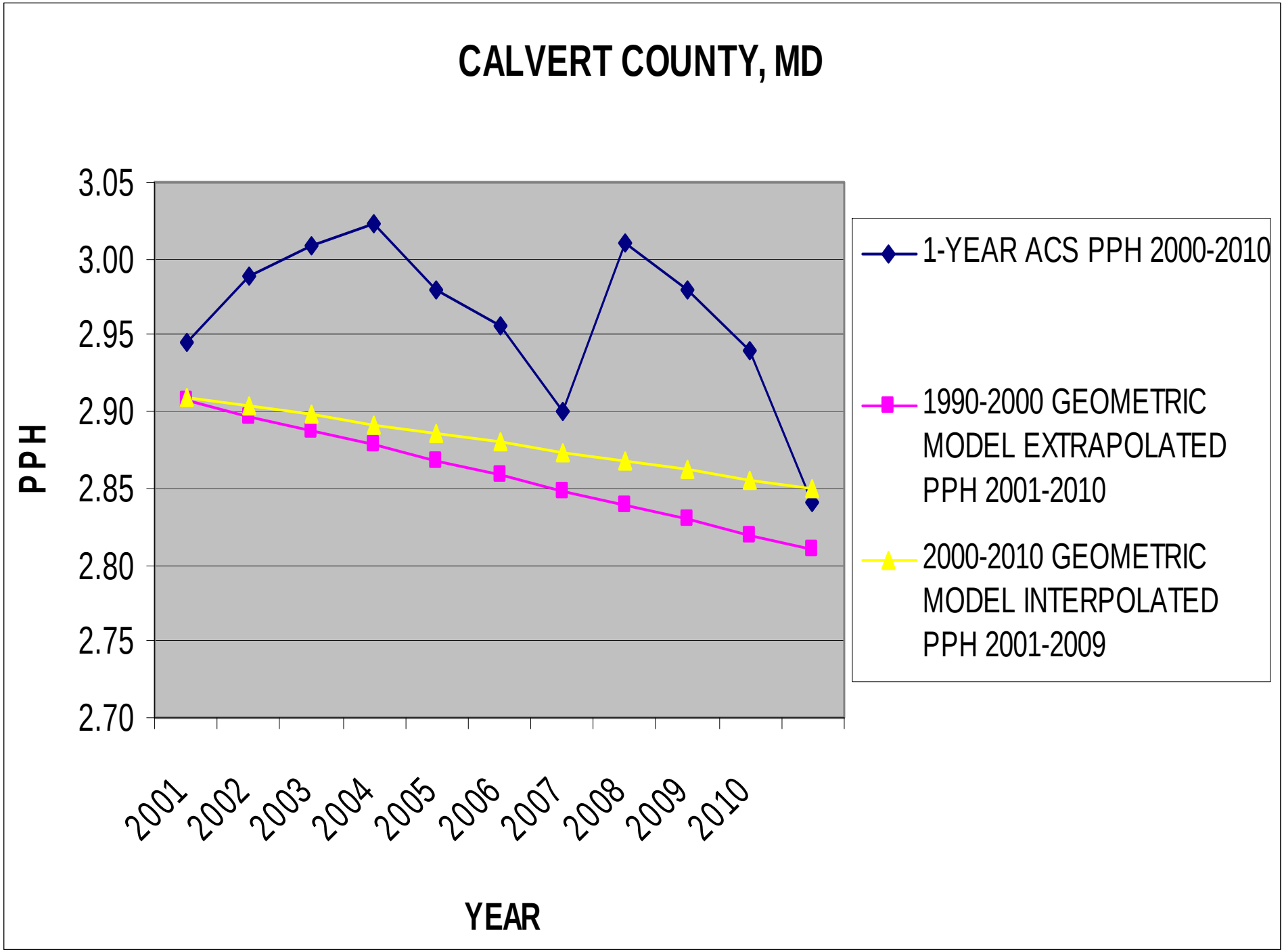
BROWARD COUNTY, FL



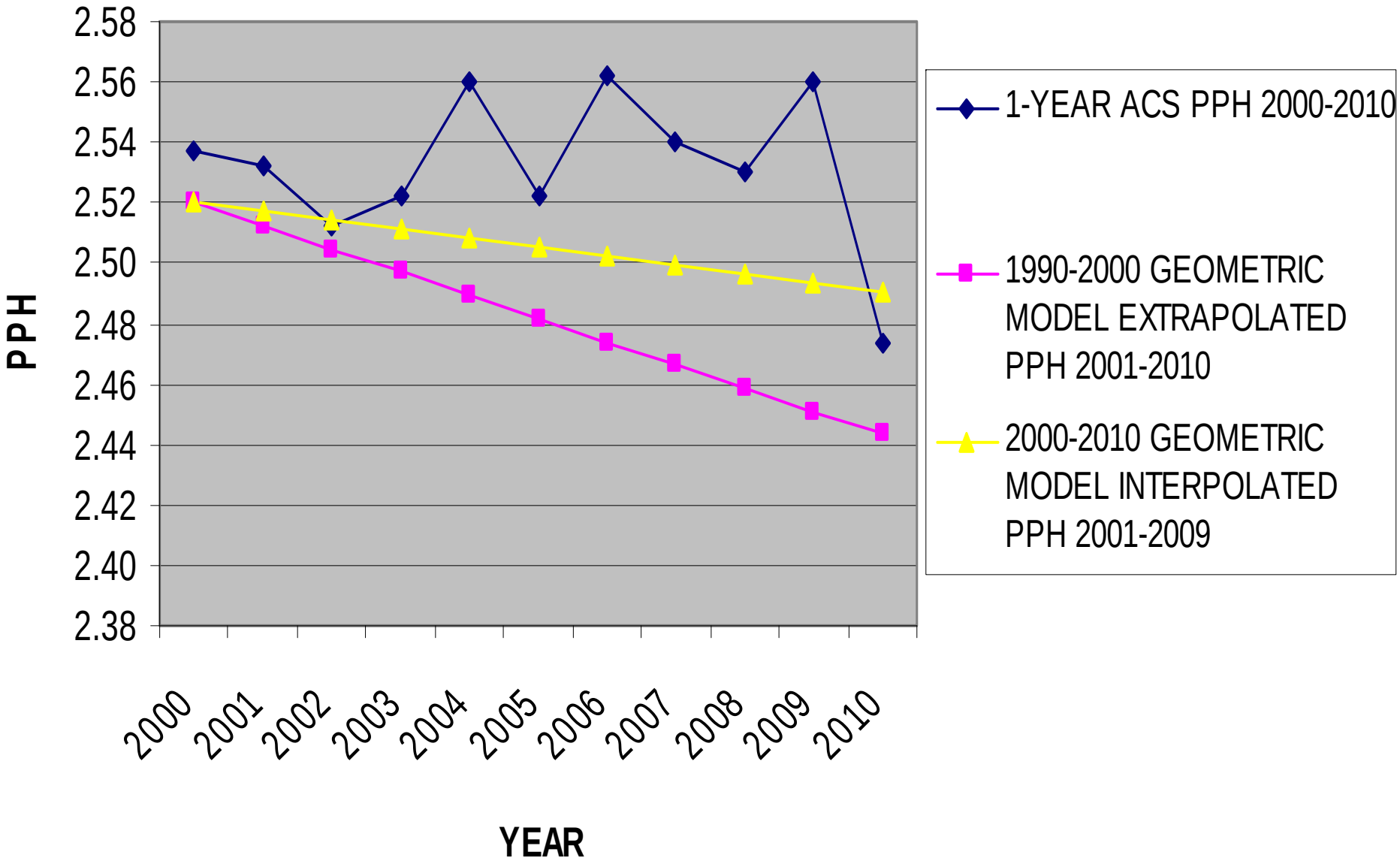

LAKE COUNTY, IL
PPH
3.05
3.00
2.95
2.90
2.85
2.80
2.75
2.70
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
YEAR
1-YEAR ACS PPH 2000-2010
1990-2000 GEOMETRIC MODEL EXTRAPOLATED PPH 2001-2010
2000-2010 GEOMETRIC MODEL INTERPOLATED PPH 2001-2009

BLACK HAWK COUNTY, IA

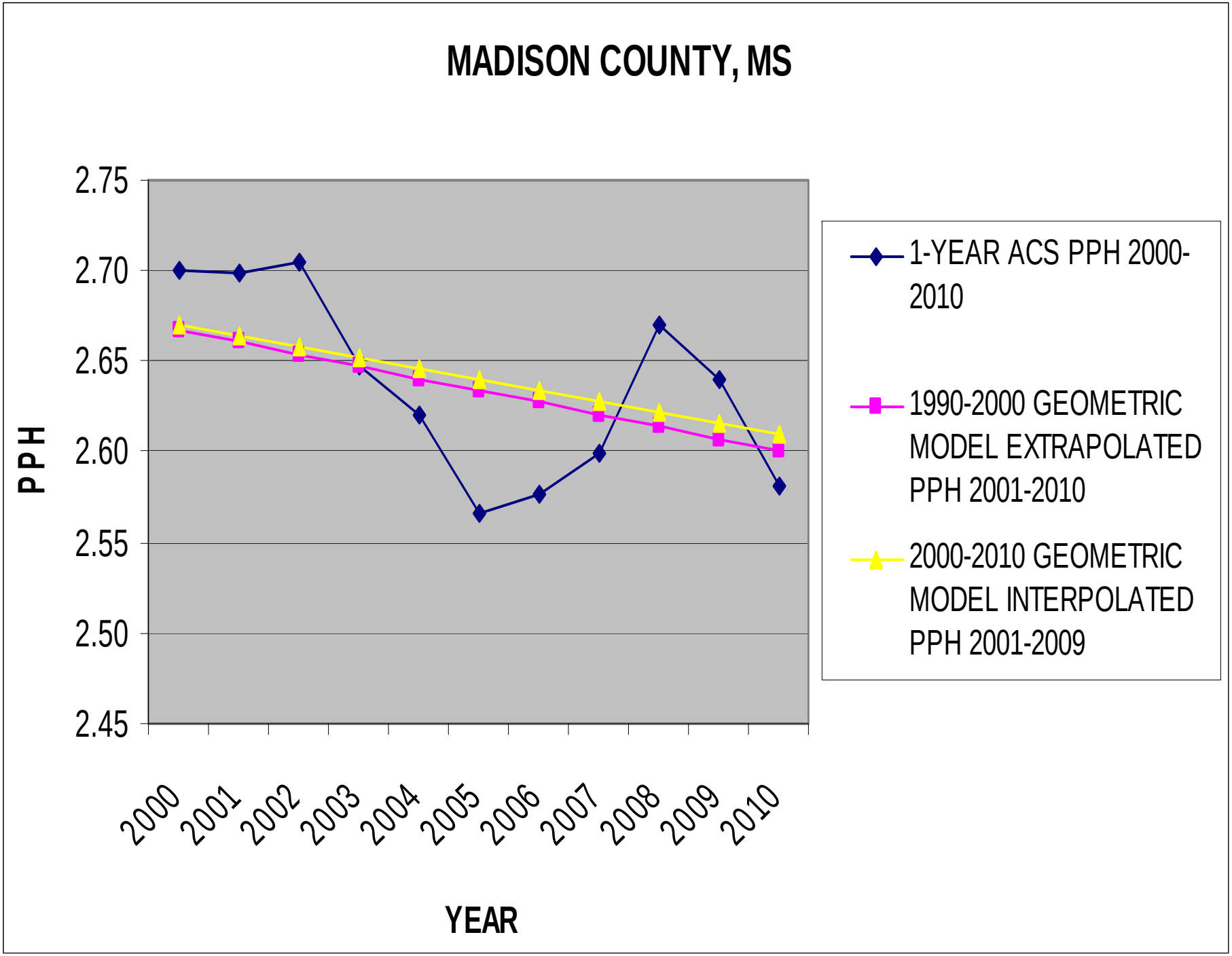



CALVERT COUNTY, MD
PPH
3.05
3.00
2.95
2.90
2.85
2.80
2.75
2.70
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
YEAR
1-YEAR ACS PPH 2000-2010
1990-2000 GEOMETRIC MODEL EXTRAPOLATED PPH 2001-2010
2000-2010 GEOMETRIC MODEL INTERPOLATED PPH 2001-2009

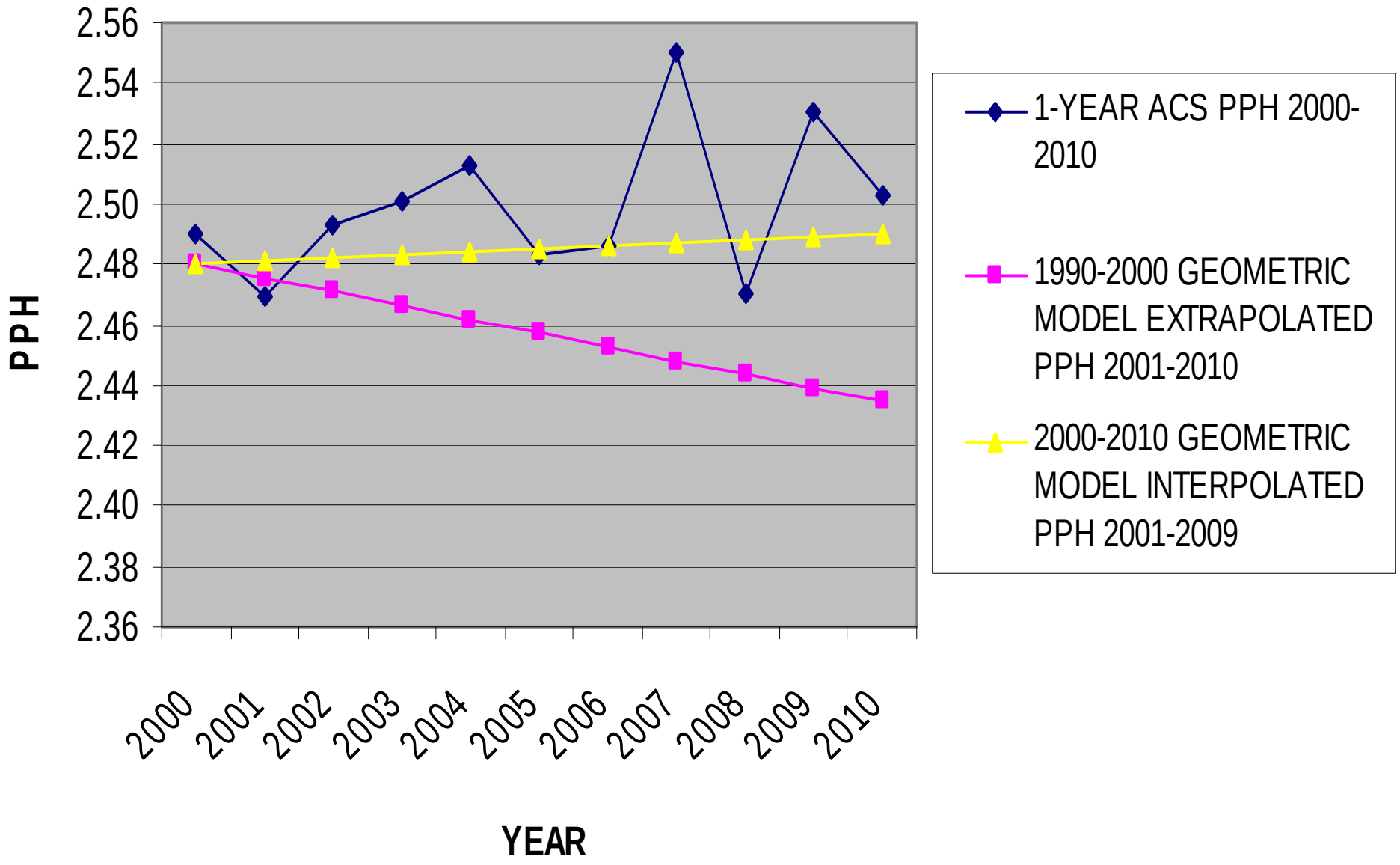
HAMPDEN COUNTY, MA



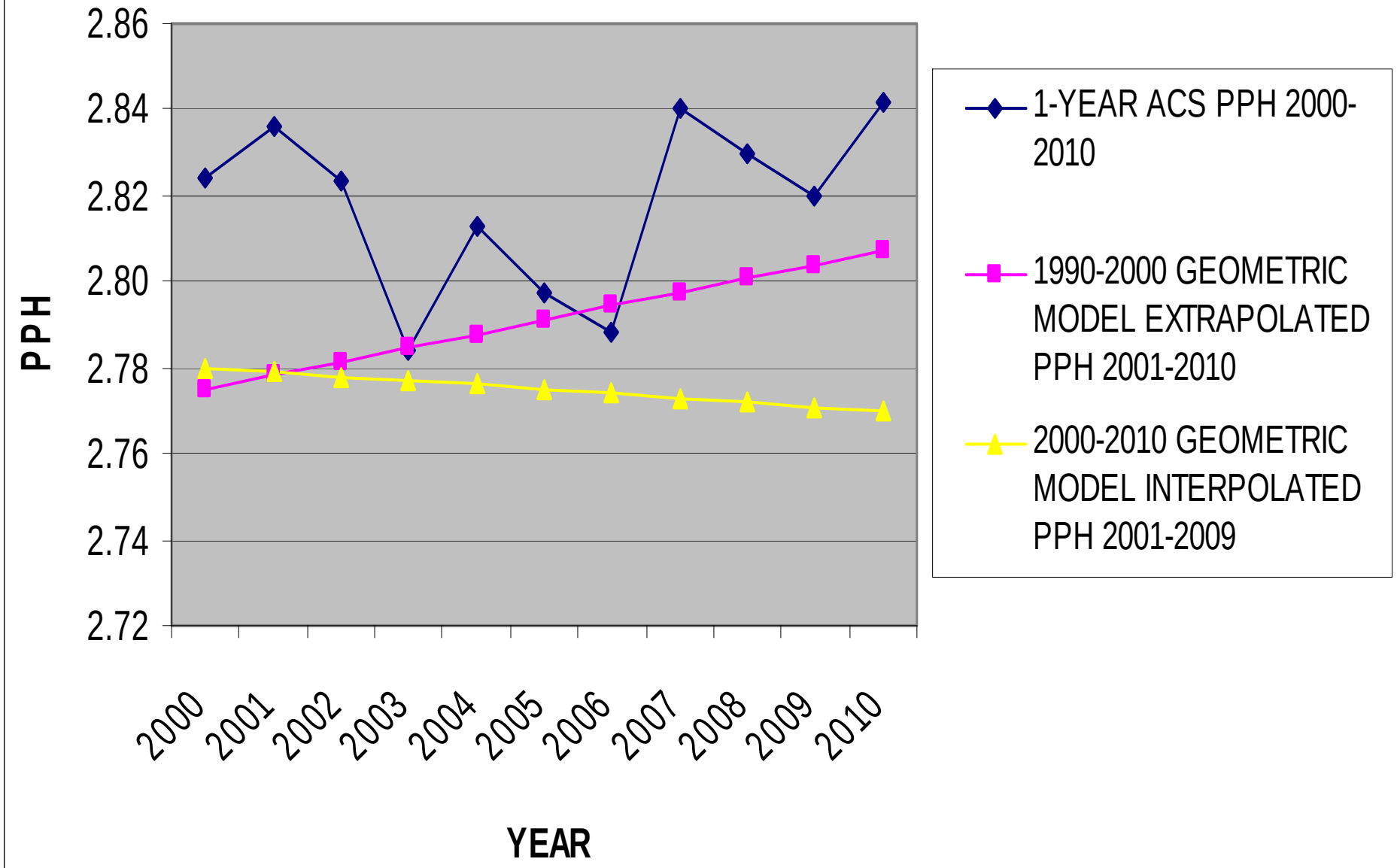
MADISON COUNTY, MS



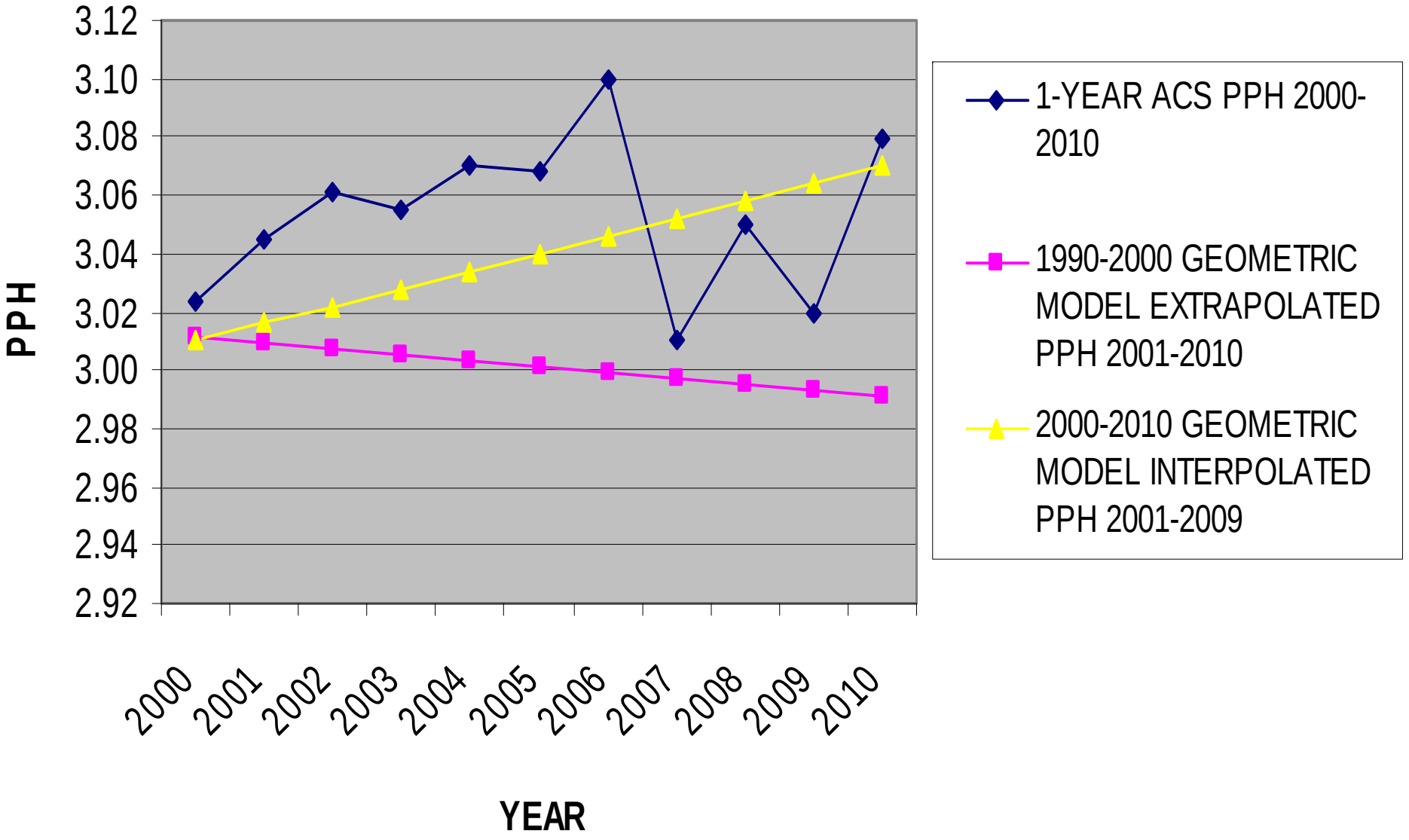
DOUGLAS COUNTY, NE

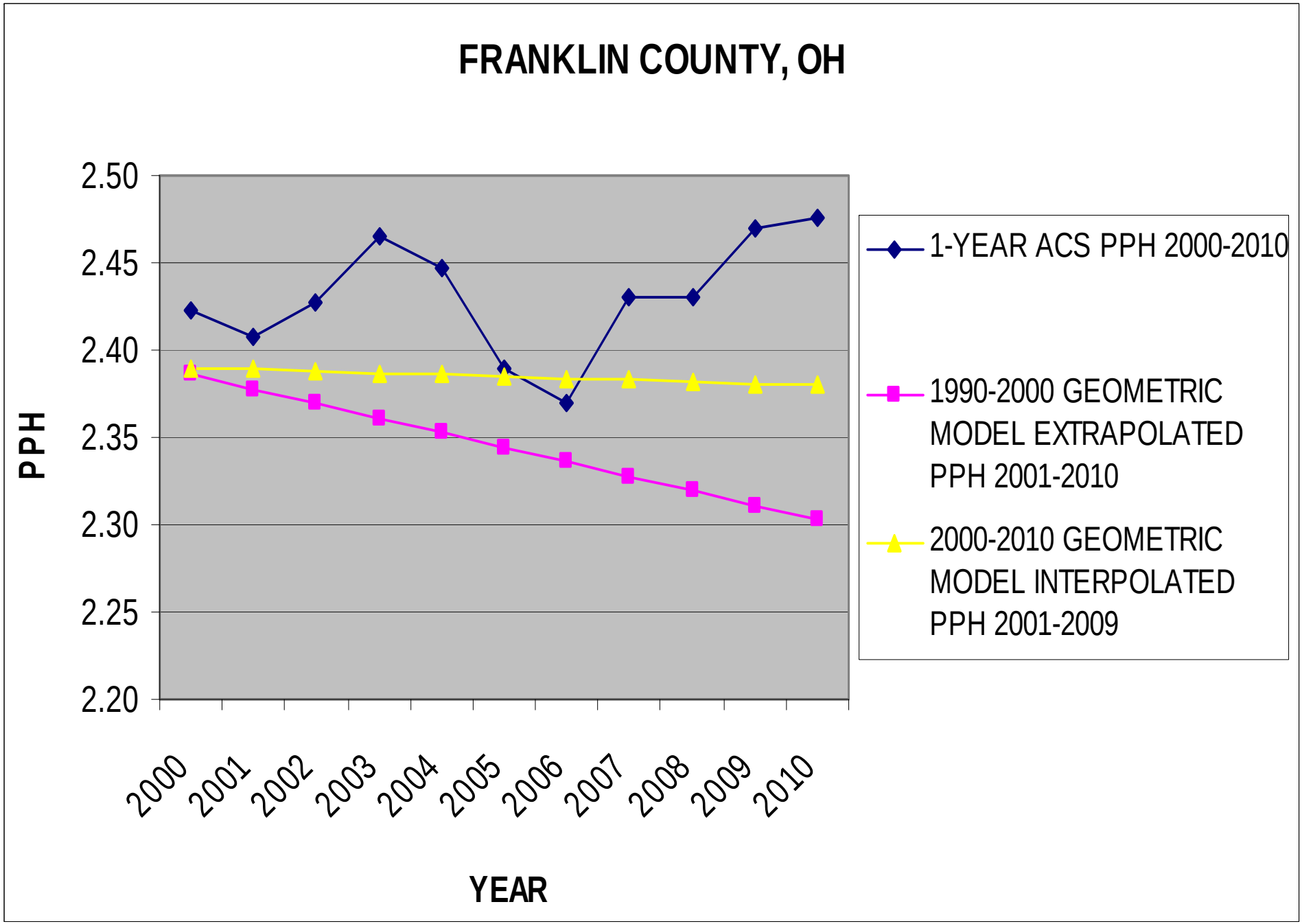


BRONX COUNTY, NY

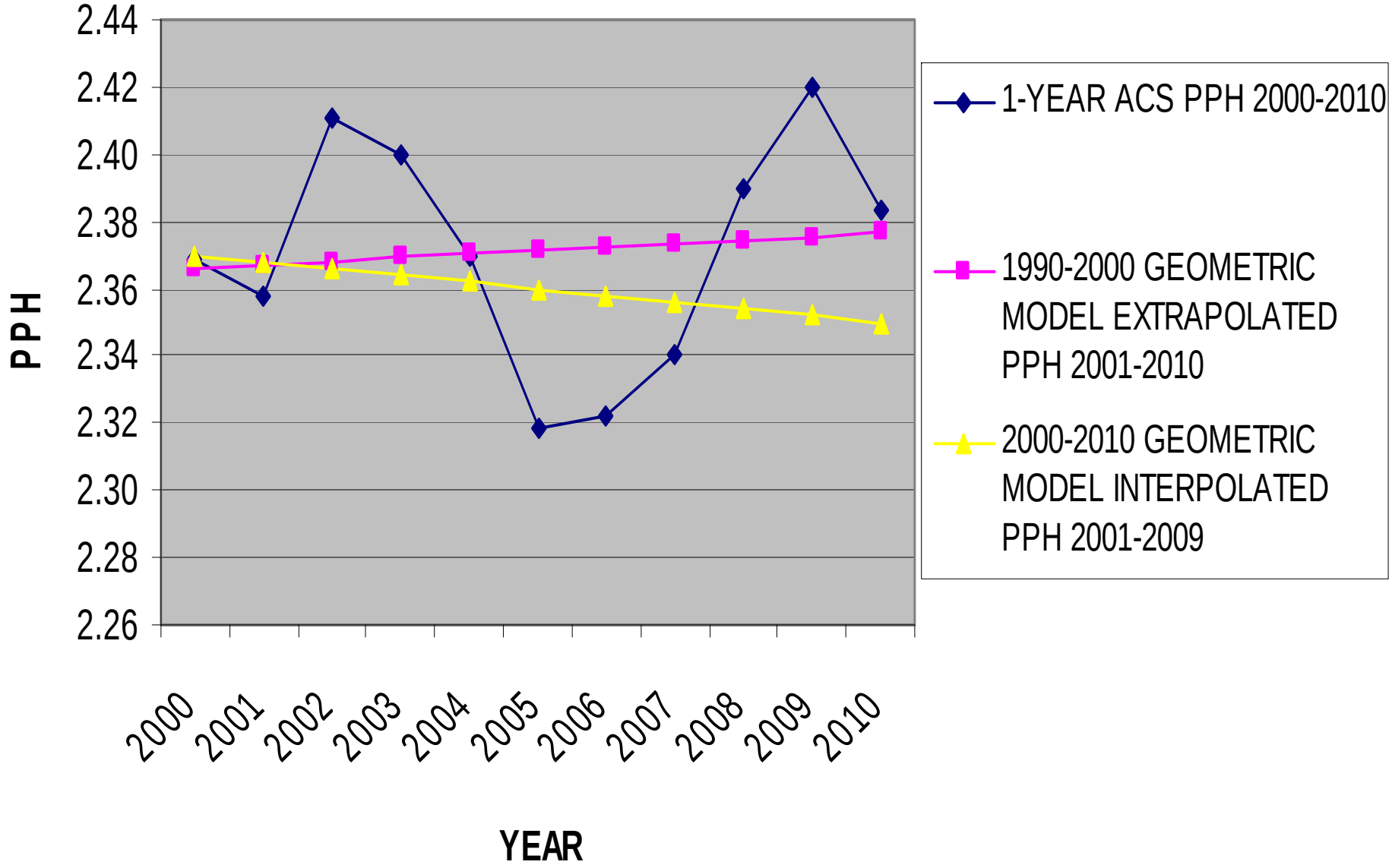


ROCKLAND COUNTY, NY

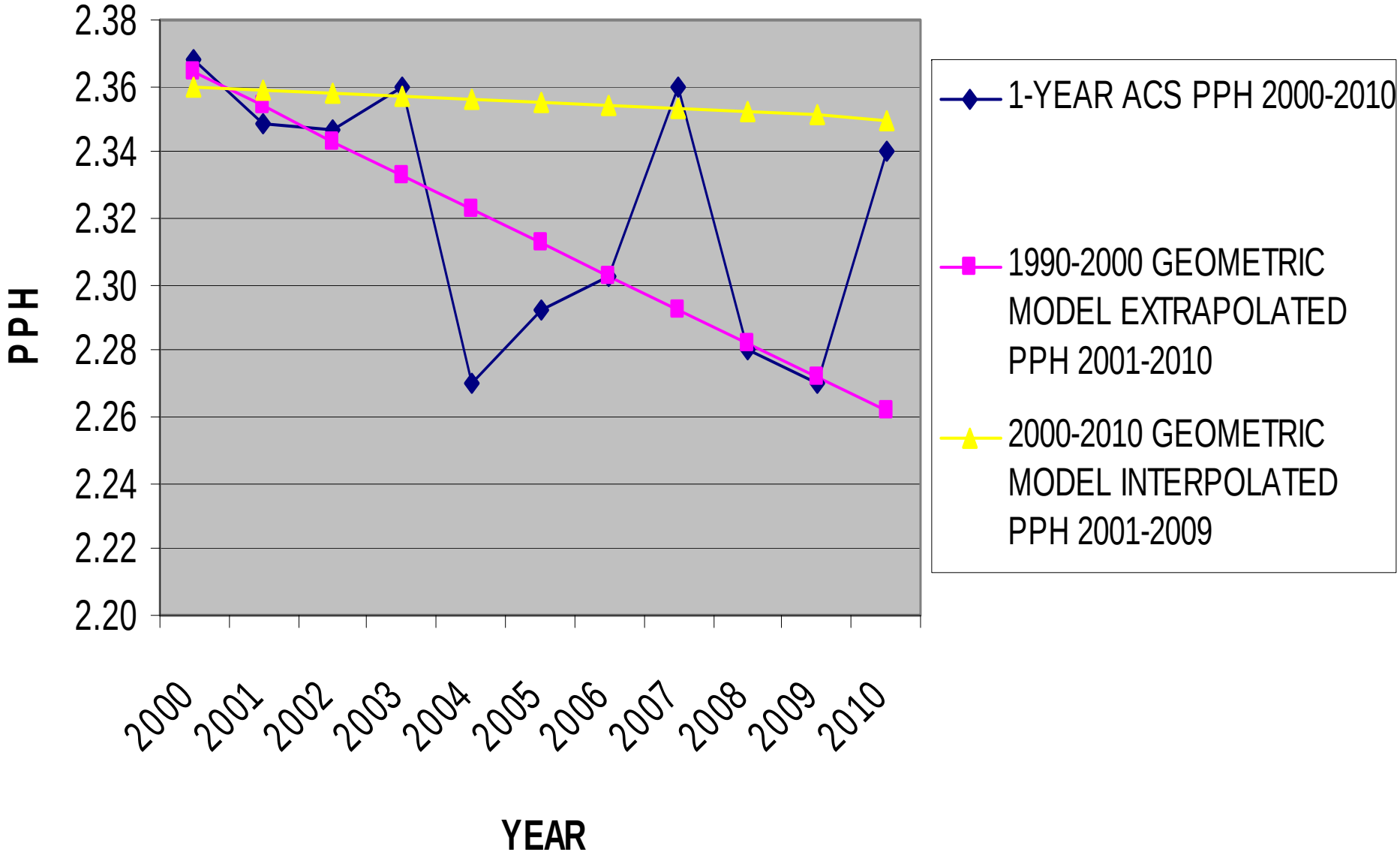



FRANKLIN COUNTY, OH
PPH
2.50
2.45
2.40
2.35
2.30
2.25
2.20
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
YEAR
1-YEAR ACS PPH 2000-2010
1990-2000 GEOMETRIC MODEL EXTRAPOLATED PPH 2001-2010
2000-2010 GEOMETRIC MODEL INTERPOLATED PPH 2001-2009

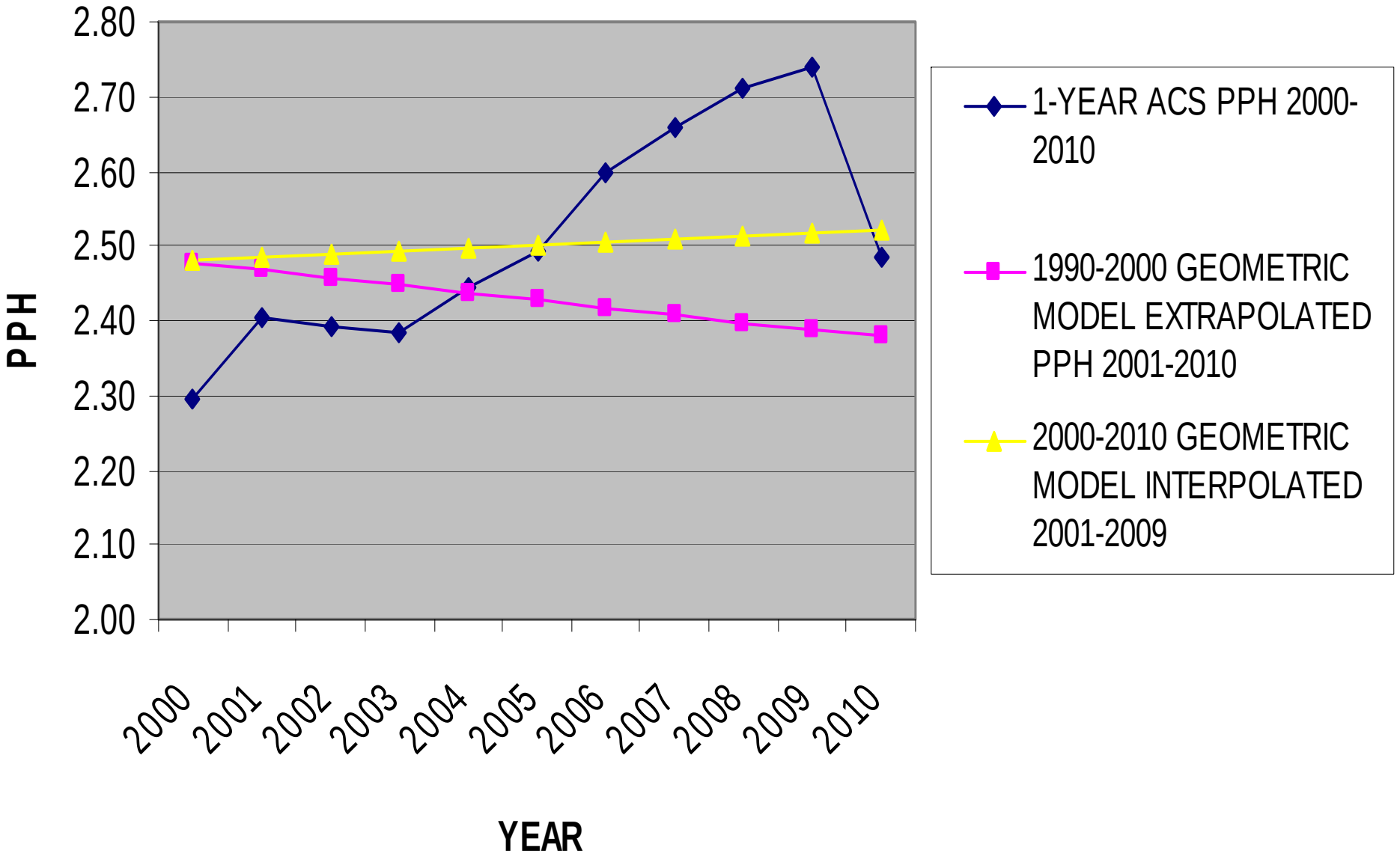
MULTNOMAH COUNTY, OR

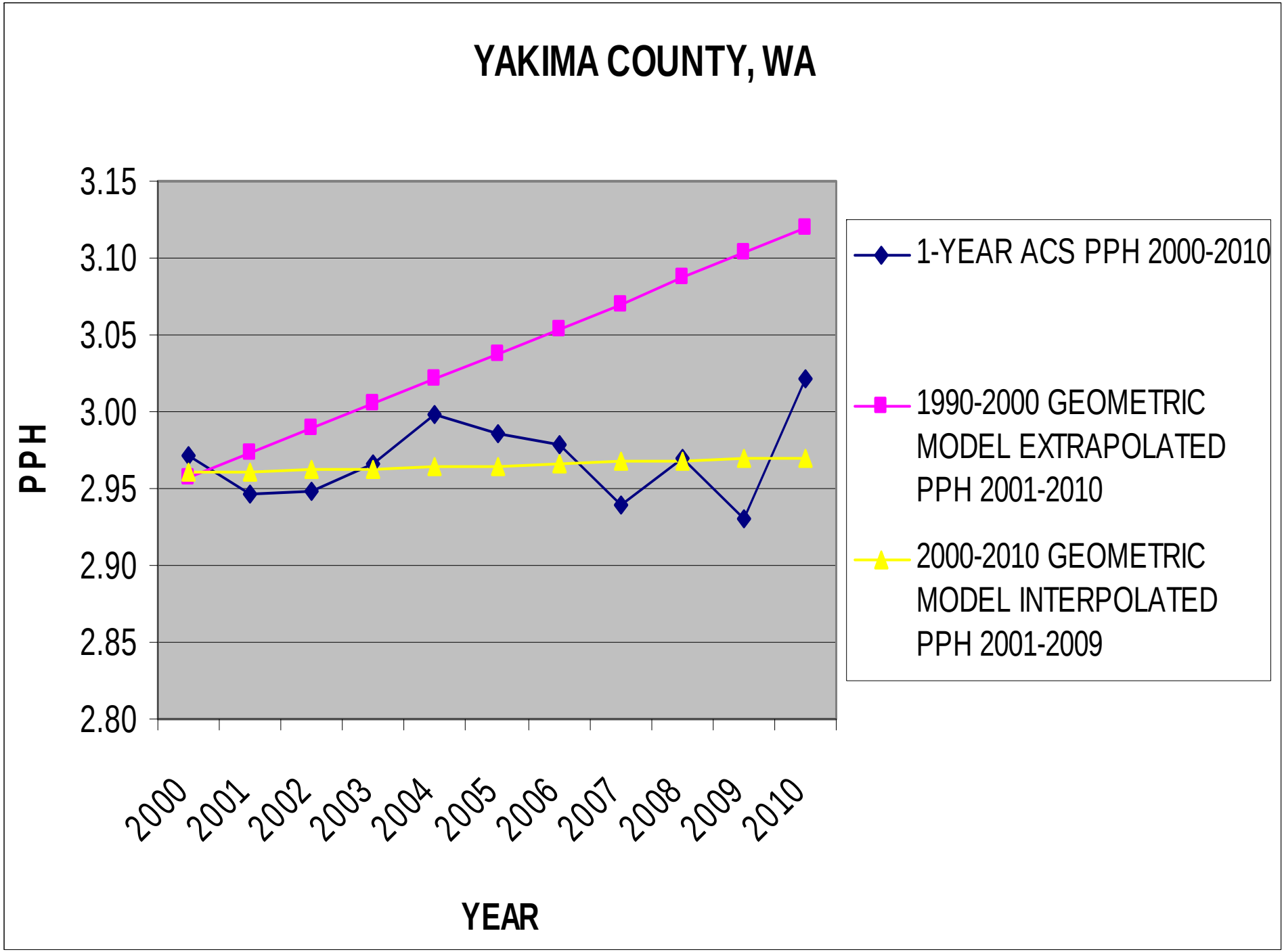


SCHUYLKILL COUNTY, PA



SEVIER COUNTY, TN




YAKIMA COUNTY, WA
PPH
3.15
3.10
3.05
3.00
2.95
2.90
2.85
2.80
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
YEAR
1-YEAR ACS PPH 2000-2010
1990-2000 GEOMETRIC MODEL EXTRAPOLATED PPH 2001-2010
2000-2010 GEOMETRIC MODEL INTERPOLATED PPH 2001-2009

CONCLUSION

The ACS provides annual PPH estimates that are subject to sample (and non-sample) error. This means that they can fluctuate from year to year in a given population, which reflects a "statistical perspective." Demographers, however, view PPH as a population characteristic that has determinants. Such that PPH is viewed as changing systematically and slowly over time - the "Demographic Perspective."

CONCLUSION

The comparisons suggest that the ACS PPH estimates exhibit too little slow systematic change and too much “noisy” variation for a given area over time to be usable by demographers and others preparing annual post-censal population estimates with the HUM.

In regard to the importance of the PPH estimates changing in a systematic manner over time, our experience in producing and defending estimates makes us appreciate model-based PPH estimates because the changes are easily understood by stakeholders. This is an important point.

Based on what we see in the temporal instability of ACS PPH values, we would have difficulty defending their use to stakeholders. This is especially the case when, as noted earlier is often the case, when these estimates are used to allocate resources.

CONCLUSION

Our finding that the 1-Year ACS PPH estimates are not particularly usable for purposes of making HUM-based population estimates at the county level is preliminary in nature. More work not only needs to be done to confirm this finding, but also to examine ways the ACS PPH estimates might be modified so that they could become more useful.

CONCLUSION

With this in mind, our suggestions for further analysis include: (1) conducting a broader scale comparison, taking into account the full range of counties; (2) examining 1-Year ACS PPH estimates that are not controlled; and (3) making adjustments (e.g., smoothing a series and then extrapolating it) to 1-Year ACS PPH estimates, perhaps in conjunction with the multi-year estimates, such that more systematic temporal change can be obtained.

CONCLUSION

Once this was done, then, depending on the results, the assessment could proceed to other geographies, such as places, census tracts, and block groups. In addition, it may be worthwhile to look at the difference between the 2010 ACS and the 2010 census PPH numbers in terms of total error, which suggests that the differences could be decomposed into variance and bias as was attempted by Mulry and Spencer (1993) in regard to estimating 1990 census error.