POPULATION ESTIMATES & PROJECTIONS

Research Brief No. 47
August 2007

Estimating Household Size for Use in Population Estimates

By Thomas Kimpel and Theresa Lowe

THE OFFICE OF FINANCIAL MANAGEMENT'S (OFM) April 1 population estimates program develops estimates for local jurisdictions that are used for revenue allocation and program administration (RCW 43.62.020). This Brief revisits a topic developed in Research Brief No. 10 where a regression procedure was developed using administrative data to update household size—a key variable used in local population estimates based on the Housing Unit Method. Additional research into household size is warranted because the availability of Census 2000 data provides an opportunity to test the efficacy of previous models while also providing a new baseline upon which to base more recent estimates.

Housing Unit Method

The Housing Unit Method is commonly used to estimate city populations. The basic formula is shown below:

Current city housing x occupancy rate x avg. persons per occupied house = Persons in houses

\+

Current persons in group quarters (nursing homes, correctional, other facilities) = Persons in facilities

Total city population

The ability to accurately generate postcensal estimates of population is important. Each year, the state distributes a set of revenues to cities and towns based on official April 1 population estimates generated by OFM. The annual population estimates are benchmarked to the most recent federal decennial census. Housing counts are updated on the basis of new construction, demolitions, and annexations. Measures of occupancy and household size are updated on the basis of available administrative or survey data. Annual fluctuations in household size at the local level is one of several factors affecting revenue allocation to cities.

Table 1
Change in Household Size by Structure Type

	Actual PPH				Percent Change in PPH		
State of Washington	**1970**	**1980**	**1990**	**2000**	**1970-1980**	**1980-1990**	**1990-2000**
All housing	2.9737	2.6086	2.5348	2.5349	-12.28	-2.83	0.01
Single-family	3.2655	2.8729	2.7969	2.7731	-12.02	-2.65	-0.85
Multi-unit structures	1.9286	1.8633	1.8788	1.9271	-3.39	0.83	2.57
Mobile homes/specials	2.4332	2.3826	2.4134	2.4837	-2.08	1.29	2.91

Household size has changed considerably over the past few decades at both the state and local level. The information presented in Table 1 shows that average household size at the state level remained largely constant over the past decade—a departure from the downward trend of the previous two decades. Between 1990 and 2000, the average number of persons per household (PPH) for single-family housing declined at a much slower rate compared to the prior decade, however, this was offset by larger increases in PPH for multi-unit structures and mobile homes/special housing. As shown in Table 2, there is

considerable variation in PPH over time at the level of the individual county.[1] Four counties experienced an increase in household size between the decades 1980-1990 and 1990-2000—Adams, Chelan, Franklin, and Yakima. Eight counties experienced a decline in household size between 1980 and 1990 only to be followed by an increase between 1990 and 2000. Twenty-seven of the 39 counties in Washington experienced a decline in PPH over consecutive decades. Generally speaking, counties in Eastern Washington with growing Hispanic/Latino populations experienced an increase in household size in the 1990s whereas counties in the Northwest and Northeast portions of the state with growing retirement age populations experienced a decline.

Table 2
Change in Household Size by County (Ranked by 1990-2000 Percent Change)

	Actual PPH			Difference in PPH		Percent Change in PPH	
County	**1980**	**1990**	**2000**	**1980-1990**	**1990-2000**	**1980-1990**	**1990-2000**
Washington	2.6086	2.5348	2.5349	-0.0738	0.0001	-2.83	0.01
Franklin	2.8817	3.0340	3.2637	0.1523	0.2296	5.29	7.57
Grant	2.7986	2.7407	2.9204	-0.0579	0.1796	-2.07	6.55
Yakima	2.7711	2.8039	2.9576	0.0328	0.1537	1.18	5.48
Chelan	2.4827	2.4863	2.6192	0.0036	0.1328	0.15	5.34
Adams	2.9113	2.9405	3.0949	0.0292	0.1544	1.00	5.25
Douglas	2.7591	2.6769	2.7554	-0.0822	0.0785	-2.98	2.93
Skagit	2.5656	2.5495	2.6032	-0.0162	0.0537	-0.63	2.11
Walla Walla	2.5411	2.4955	2.5388	-0.0456	0.0433	-1.79	1.73
Benton	2.7971	2.6516	2.6795	-0.1455	0.0278	-5.20	1.05
Clark	2.7625	2.6625	2.6900	-0.1000	0.0276	-3.62	1.04
Kittitas	2.3976	2.3251	2.3314	-0.0725	0.0063	-3.02	0.27
Grays Harbor	2.5966	2.4813	2.4826	-0.1152	0.0013	-4.44	0.05
Garfield	2.5955	2.3948	2.3911	-0.2008	-0.0037	-7.73	-0.15
Cowlitz	2.6619	2.5588	2.5531	-0.1031	-0.0057	-3.87	-0.22
Lincoln	2.5726	2.4308	2.4233	-0.1418	-0.0075	-5.51	-0.31
King	2.4868	2.3982	2.3905	-0.0886	-0.0078	-3.56	-0.32
Spokane	2.5789	2.4747	2.4646	-0.1042	-0.0101	-4.04	-0.41
Okanogan	2.6674	2.5877	2.5762	-0.0797	-0.0115	-2.99	-0.44
Pierce	2.6586	2.6231	2.6047	-0.0355	-0.0183	-1.34	-0.70
Whatcom	2.5902	2.5324	2.5114	-0.0577	-0.0211	-2.23	-0.83
Snohomish	2.7606	2.6794	2.6547	-0.0813	-0.0247	-2.94	-0.92
Mason	2.5458	2.5162	2.4891	-0.0296	-0.0271	-1.16	-1.08
Lewis	2.6732	2.5997	2.5690	-0.0735	-0.0307	-2.75	-1.18
Kitsap	2.6820	2.6469	2.6007	-0.0351	-0.0462	-1.31	-1.75
Wahkiakum	2.7724	2.4762	2.4243	-0.2962	-0.0518	-10.68	-2.09
Thurston	2.6441	2.5530	2.4987	-0.0911	-0.0543	-3.45	-2.13
Asotin	2.5662	2.4727	2.4162	-0.0935	-0.0565	-3.64	-2.28
Skamania	2.7896	2.6921	2.6120	-0.0975	-0.0801	-3.50	-2.98
Columbia	2.5254	2.4368	2.3628	-0.0886	-0.0740	-3.51	-3.04
Whitman	2.4688	2.3868	2.3115	-0.0820	-0.0752	-3.32	-3.15
Stevens	2.9070	2.7318	2.6439	-0.1752	-0.0879	-6.03	-3.22
Pacific	2.4465	2.3499	2.2711	-0.0966	-0.0788	-3.95	-3.35
Island	2.6706	2.6149	2.5223	-0.0557	-0.0926	-2.09	-3.54
Pend Oreille	2.8088	2.6030	2.5074	-0.2058	-0.0955	-7.33	-3.67
Clallam	2.5374	2.4007	2.3066	-0.1367	-0.0941	-5.39	-3.92
Klickitat	2.7211	2.6409	2.5361	-0.0802	-0.1048	-2.95	-3.97
San Juan	2.2946	2.2489	2.1587	-0.0458	-0.0902	-1.99	-4.01
Jefferson	2.4537	2.3089	2.2122	-0.1448	-0.0967	-5.90	-4.19
Ferry	2.8567	2.6978	2.4938	-0.1589	-0.2040	-5.56	-7.56

[1] When interpreting the percent change in PPH in terms of household population, it is often useful to take into account the base population of a county. Using Adams County as an example, the 5.25 percent increase in PPH between 1990 and 2000 translates to an increase in household population of 708 persons holding the number of occupied housing units constant at their 1990 values. Using similar logic, the difference for King County's 0.32 percent decrease in PPH translates to a decline in household population of 4,775 persons.

Due to the high degree of local variability in household size over time, postcensal estimates of household size readily lend themselves to the use of administrative data. Because administrative data are often collected on an annual basis, they can be leveraged to model changes in household size since the previous federal census.

Estimating Household Size

Regression models are a common technique for estimating household size or changes in household size at disaggregate levels (OFM, 1983; OFM, 2000; Smith, Nagle & Cody, 2002). In practice, the specification of household size models involves a series of tradeoffs. While theory should guide model specification, there are several drawbacks to using administrative data: 1) the data may not exist in the first place so proxy variables must be used, 2) the data may not exist at the appropriate summary level, 3) the data may not exist at the appropriate time interval, and 4) the data may simply be unreliable. Besides the selection of explanatory variables, an additional tradeoff involves the overall model structure (e.g., basic, ratio, or change model).

Previously, OFM modeled the change in PPH from 1980 to 1990 at the county level as a function of the change in public school enrollment grades K-8, the change in the sum of births four years prior to the prediction date, and the change in persons age 65 and over. To control for the size effect, all of the independent variables were divided by the number of housing units. An evaluation of the predicted values of PPH change from 1990 to 2000 against actual Census data shows that the model did not perform as well as expected. In particular, the model did not adequately capture positive change in household size and tended to overstate the magnitude of the declines.

For the present analysis, an alternative model specification was developed. First, it was deemed necessary to control for the existing level of household size. This predictor variable was expected to have a negative impact on household size since the more persons there are per household, the greater the propensity for change. Second, it was decided that the variable for births should be extended to 14 years prior to the prediction date, thus eliminating the need for school enrollment information which is believed to be less reliable.[2] The cumulative number of births proxies for the number of children in households. A variable measuring the change in Hispanic/Latino population was included to capture any effects related to this specific population subgroup. Both the change in cumulative births and Hispanic population were expected to have a positive impact on household size. Similar to the previous model, a variable representing the number of elderly persons per household was included. It was expected that an increase in the number of elderly persons would reduce household size because an aging population generates more one-person households. The final model specification is shown in Table 3 below.

Table 3
Regression Equation for Predicting Change in PPH

Dependent variable (x):	1990-2000 change in PPH			
Independent variables (y):	(1) 1990 PPH	(2) 1990-2000 change in the sum of births 14 years prior to the prediction date per all housing	(3) 1990-2000 change in Hispanic population per all housing	(4) 1990-2000 change in persons age 65+ per all housing
Regression Statistics:	N=39	R^2 Adj. = 0.8192	F = 44.03	

[2] The school enrollment data is subject to considerable error because 1) school district boundaries do not conform to county boundaries and 2) the data do not capture children who drop out of school.

Of the explanatory variables included in the model, the change in births 14 years prior to the prediction date, change in Hispanic population, and change in persons 65 and over are all statistically significant at the 95 percent level of confidence. The existing level of PPH in 1990 is not significant, however, theory suggests that it should remain in the model specification. The signs of the regression coefficients are all in the expected direction and the relative magnitudes of the coefficients appear reasonable. Diagnostic tests for multicollinearity and heteroscedasticity were negative.

In addition to a change model, a basic model was also run using data for a single point in time, i.e., the prediction of household size per se—not change in household size. Although this model had high explanatory power, the predicted household size values provided very poor estimates of the change in household size.[3] As such, a county-level change model was deemed more appropriate. Other independent variables that were tested included an adjusted school enrollment variable[4] and median income. Neither of these variables performed well.

The results of the model predicting change in PPH from 1990-2000 are shown in Table 4. Relative to the actual value of PPH in 2000, 92.3 percent of counties have a prediction error within +/- 5 percent. The largest positive errors are associated with Franklin, Yakima, and Adams counties while the largest negative errors are associated with Ferry, Stevens, and Pend Oreille counties. To put the amount of error in more realistic terms, 25.0 percent of Washington counties have a prediction error greater than $1/10^{th}$ of a person per household and 48.7 percent have a prediction error greater than $1/20^{th}$ of a person per household. While these errors may appear to be relatively small at first glance, they can have a dramatic effect on population estimates for populous areas.[5]

The predicted level of PPH change by county through year 2007 is presented in Table 5 and depicted graphically in Figure 1. Estimates for six of Washington's 39 counties show an increase in PPH relative to 2000 values. These consist of Franklin, Adams, Chelan, Grant, Yakima, and Douglas counties. The largest gainer is Franklin County at just over $1/12^{th}$ of a person per household. Eight counties are estimated to show a decline in PPH by at least $1/10^{th}$ of a person. These include Skamania, Island, Klickitat, Stevens, Lincoln, San Juan, Pend Oreille, and Ferry counties. Ferry County is expected to decline by just under $1/5^{th}$ of a person per household. Consistent with 1990-2000 trends, counties with growing Hispanic/Latino populations are associated with the greatest increases in PPH whereas counties with growing retirement age populations are associated with the largest declines.

[3] For example, 1990 census data were used as the base for predicting county-level household size for 2000. Change in household size was developed by subtracting 1990 county household size from the 2000 predicted values. The change in household size was considerably less accurate than in any of the change models evaluated.

[4] The adjusted school enrollment variable was comprised of total enrollment grades K-12 (public and private headcounts plus public running starts) multiplied by the ratio of total enrollment to the number of school age children aged 5-17 at the last census.

[5] Using Snohomish County as an example, the amount of prediction error (-0.77 percent) translates to a decrease in household population of 4,606 persons based upon Census 2000 values for occupied housing units

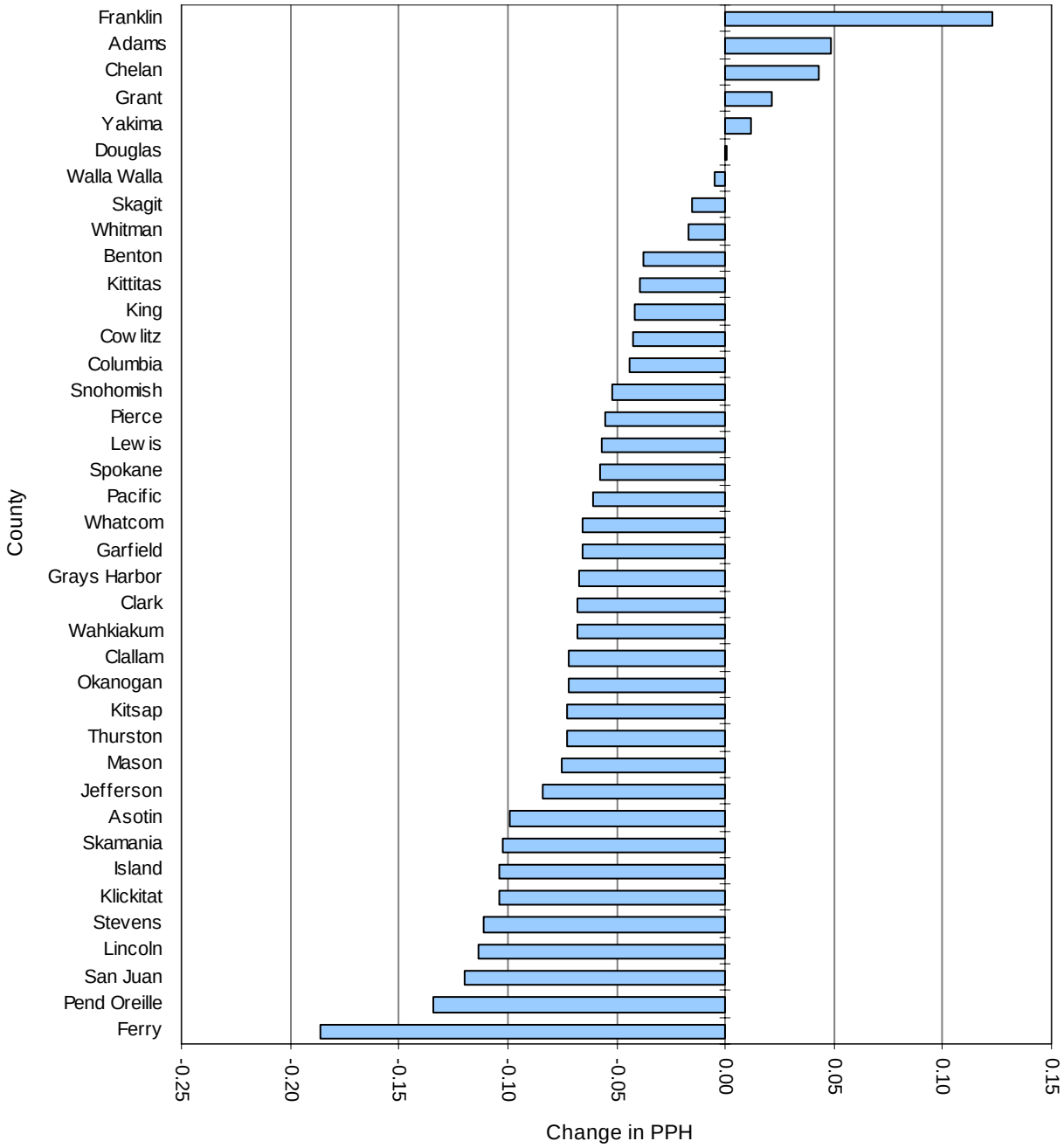
Table 4
Actual and Predicted PPH by County (Ranked by Level of Model Error)

	Actual PPH		Predicted PPH		
County	**1990**	**2000**	**2000**	**Difference**	**Percent Error**
Washington	2.6086	2.5348			
Franklin	3.0340	3.2637	3.5231	0.2524	7.95
Yakima	2.8039	2.9576	3.1327	0.1795	5.92
Adams	2.9405	3.0949	3.2623	0.1663	5.41
Grant	2.7407	2.9204	3.0625	0.1479	4.87
Chelan	2.4863	2.6192	2.7264	0.1068	4.09
Douglas	2.6769	2.7554	2.8014	0.0484	1.67
Walla Walla	2.4955	2.5388	2.5755	0.0361	1.45
Skagit	2.5495	2.6032	2.6301	0.0297	1.03
San Juan	2.2489	2.1587	2.1707	0.0055	0.56
Okanogan	2.5877	2.5762	2.5726	-0.0053	-0.14
Kittitas	2.3251	2.3314	2.3261	-0.0054	-0.23
King	2.3982	2.3905	2.3803	-0.0055	-0.42
Benton	2.6516	2.6795	2.6646	-0.0228	-0.55
Clark	2.6625	2.6900	2.6728	-0.0126	-0.64
Whatcom	2.5324	2.5114	2.4930	-0.0163	-0.73
Snohomish	2.6794	2.6547	2.6342	-0.0154	-0.77
Grays Harbor	2.4813	2.4826	2.4567	-0.0274	-1.04
Pierce	2.6231	2.6047	2.5694	-0.0349	-1.35
Lewis	2.5997	2.5690	2.5336	-0.0350	-1.38
Mason	2.5162	2.4891	2.4510	-0.0388	-1.53
Thurston	2.5530	2.4987	2.4596	-0.0391	-1.57
Kitsap	2.6469	2.6007	2.5598	-0.0401	-1.57
Cowlitz	2.5588	2.5531	2.5127	-0.0355	-1.58
Spokane	2.4747	2.4646	2.4216	-0.0411	-1.74
Whitman	2.3868	2.3115	2.2681	-0.0413	-1.88
Klickitat	2.6409	2.5361	2.4820	-0.0565	-2.13
Island	2.6149	2.5223	2.4677	-0.0516	-2.16
Jefferson	2.3089	2.2122	2.1625	-0.0518	-2.25
Lincoln	2.4308	2.4233	2.3663	-0.0591	-2.35
Pacific	2.3499	2.2711	2.2170	-0.0543	-2.38
Asotin	2.4727	2.4162	2.3582	-0.0553	-2.40
Garfield	2.3948	2.3911	2.3305	-0.0679	-2.53
Wahkiakum	2.4762	2.4243	2.3623	-0.0535	-2.56
Skamania	2.6921	2.6120	2.5433	-0.0697	-2.63
Clallam	2.4007	2.3066	2.2380	-0.0703	-2.98
Stevens	2.7318	2.6439	2.5505	-0.1007	-3.53
Ferry	2.6978	2.4938	2.3907	-0.1071	-4.14
Pend Oreille	2.6030	2.5074	2.4037	-0.1004	-4.14
Columbia	2.4368	2.3628	2.2465	-0.1204	-4.92

Table 5
Predicted Change in PPH by County: 2000-2007 (Ranked by Difference)

County	2000 Actual	2007 Predicted	Difference
Franklin	3.2637	3.3865	0.1228
Adams	3.0949	3.1433	0.0484
Chelan	2.6192	2.6621	0.0429
Grant	2.9204	2.9415	0.0211
Yakima	2.9576	2.9693	0.0117
Douglas	2.7554	2.7564	0.0010
Walla Walla	2.5388	2.5335	-0.0052
Skagit	2.6032	2.5876	-0.0156
Whitman	2.3115	2.2950	-0.0166
Benton	2.6795	2.6416	-0.0379
Kittitas	2.3314	2.2919	-0.0395
King	2.3905	2.3491	-0.0414
Cowlitz	2.5531	2.5105	-0.0426
Columbia	2.3628	2.3191	-0.0436
Snohomish	2.6547	2.6027	-0.0520
Pierce	2.6047	2.5492	-0.0555
Lewis	2.5690	2.5126	-0.0564
Spokane	2.4646	2.4068	-0.0579
Pacific	2.2711	2.2103	-0.0608
Whatcom	2.5114	2.4457	-0.0656
Garfield	2.3911	2.3251	-0.0659
Grays Harbor	2.4826	2.4151	-0.0675
Clark	2.6900	2.6218	-0.0682
Wahkiakum	2.4243	2.3561	-0.0683
Clallam	2.3066	2.2349	-0.0717
Okanogan	2.5762	2.5043	-0.0719
Kitsap	2.6007	2.5283	-0.0724
Thurston	2.4987	2.4257	-0.0730
Mason	2.4891	2.4142	-0.0749
Jefferson	2.2122	2.1281	-0.0841
Asotin	2.4162	2.3174	-0.0987
Skamania	2.6120	2.5098	-0.1022
Island	2.5223	2.4185	-0.1037
Klickitat	2.5361	2.4319	-0.1042
Stevens	2.6439	2.5331	-0.1108
Lincoln	2.4233	2.3100	-0.1132
San Juan	2.1587	2.0392	-0.1195
Pend Oreille	2.5074	2.3733	-0.1342
Ferry	2.4938	2.3080	-0.1858

Figure 1
Predicted Change in PPH by County: 2000-2007



A comparison of household size values generated by OFM, the Puget Sound Regional Council (PSRC), and the U.S. Census Bureau's American Community Survey (ACS) is presented in Table 6 (PSRC, 2005; US Census Bureau, 2005). The data are shown through year 2005—the most recent year for which the ACS data are available. OFM is predicting a decline in PPH across the four counties in the Seattle metropolitan region of approximately 1/20th of a person per household. In contrast both the PSRC and the ACS show much more variation in the amount of PPH decline across the counties. Direct comparison of county-level change in household size across OFM, PSRC, and ACS should be undertaken with a degree of caution since the values represent outcomes from two different estimate procedures and a survey. In a sense, this would be like comparing three different varieties of apples. It suffices to say that all three entities predict a decline in household size across each of the counties, although the amount decline is subject to debate.

Table 6
Comparison of Household Size Estimates: 2005

		Predicted			Difference (from Census 2000)		
County	**Census 2000**	**OFM 2005**	**PSRC 2005**	**ACS 2005**	**OFM 2005**	**PSRC 2005**	**ACS 2005**
King	2.3905	2.3481	2.3781	2.3533	-0.0469	-0.0124	-0.0372
Kitsap	2.6007	2.5390	2.5330	2.5077	-0.0696	-0.0677	-0.0930
Pierce	2.6047	2.5531	2.5834	2.5937	-0.0583	-0.0213	-0.0110
Snohomish	2.6547	2.6059	2.6362	2.5604	-0.0549	-0.0185	-0.0942

The three household size estimation methodologies can and should be evaluated from an accuracy standpoint upon release of data from the 2010 federal census. Additional research in this area is needed as there is presently no single model specification or set of variables that population analysts/demographers agree upon.

This research brief, and others, should serve as a reminder that many of the components of population estimates are approximations—as are the resulting populations. OFM uses the Housing Unit Method as one of three estimation procedures to develop county populations. Generally the most accurate population estimates come from using several procedures and understanding the biases in each.

References:

1. Office of Financial Management Population Section (1983- Dec.). Estimating 1980 PPH by population size for cities and towns in the State of Washington. Special Report No. 82, 1-15.

2. Office of Financial Management Population Unit (2000- Aug.). Developing trends in household size for use in population estimates. Research Brief No. 10, 1-6.

3. Smith, S. K., Nogle, J., & Cody, S. (2002). A regression approach to estimating the average number of persons per household. Demography, 39 (4), 697-712.

4. Puget Sound Regional Council (2005). Population, Households, and Housing Estimates. Retrieved August 14, 2007 from Puget Sound Regional Council web site: http://www.psrc.org/data/dem/pop2005.xls.

5: U.S. Census Bureau (2005). 2005 American Community Survey. Retrieved August 14, 2007 from American FactFinder web site: http://factfinder.census.gov.