From:	J McLaughlin
Cc:	Council
Subject:	Population projections; Comprehensive Plan update
Date:	Friday, January 22, 2016 2:55:14 PM
Attachments:	popWC_proj_sum.pdf

Dear Honorable County Council Members,

Thank you for your efforts to lead our County toward a desirable future, in particular through work on updating the County Comprehensive Plan. As you must understand as well as most, Comprehensive Plan updates are developed using population projections. Selecting appropriate projections is essential to updating the Comprehensive Plan in a responsible manner.

Unfortunately, the population projections you have been given are not reliable; they were developed from invalid models, using erroneous statistical analyses. Those projections, developed by a Seattle-based consultant (BERK), must not be used to update our comprehensive plan in any way. I am writing not merely to criticize, but also to offer you a better alternative. While BERK committed most statistical and modeling errors possible, I developed population projections for Whatcom County using methods appropriate for our population data and standard in population modeling. My work is described in the attached document, which I request you add to the public record on the Comprehensive Plan update. Mostly, I ask that you read my analysis and use the results to improve the Comprehensive Plan.

I did not reach these conclusions from cursory or biased review. Rather, I applied objective methods in statistical analysis and population modeling with which I have decades of professional experience. I would be happy to provide additional explanation, interpretation, or analysis at your invitation.

You have not been well served by outside consultants who lack the requisite experience in population forecasting. I have worked to meet your need for reliable population projections, at no cost to the County. I would be happy to assist you further in whatever capacity you would find useful. Regardless, I caution you against committing County resources and our home to decisions based on erroneous projections.

Thank you for your service,

John

John McLaughlin
Dept. Environmental Sciences
Huxley College of the Environment
Western Washington University
Bellingham, WA 98225-9181 USA

Whatcom County Population Growth Forecast Comparisons

Prepared by

John McLaughlin
Department of Environmental Sciences
Huxley College of the Environment
Western Washington University

21 January 2016

This is a summary of extensive work on population projections for Whatcom County's ongoing comprehensive plan update. I have done similar work on population projections for Bellingham, which has produced qualitatively identical results. As a professional population biologist with some experience analyzing population data, developing population models, and forecasting with models, I became concerned that the County has not been well served by projections provided by its consultants. Accordingly, I reviewed errors in those projections and developed more reliable projections using valid methods.

There are three general approaches to developing population projections: (1) determine values to achieve desired future conditions, (2) mechanistic modeling, and (3) statistical modeling. Approach (1) serves communities best, for diverse reasons. Approach (2) is excluded here, due to immense uncertainty in migration, which dominates population change in Whatcom County. Approach (3) is traditionally applied in Whatcom County and constituent cities. This approach fits a statistical model to population data and applies that model to forecast future population growth. Since I began following county and city projection processes in 2002, statistical models used in this approach have been invalid, developed by consultants who appear to have lacked the relevant competencies. Given the financial, social, and environmental costs associated with erroneous projections, we should not rely on projections generated by invalid models or unqualified consultants. If we choose not to use approach (1), and since we cannot use approach (2), we must apply approach (3) as thoughtfully and skillfully as possible. As summarized below, projections provided by BERK consulting do not meet these criteria. I have developed population forecasts that do, described below.

This report summarizes my work on Whatcom County population projections, in seven sections.

(1) Technical review summary of population projections developed for Whatcom County by BERK consulting.

(2) Analysis of temporal structure in county population data, to inform forecast model development.

(3) Development of reliable population forecast models for the county, fit to OFM population data.

(4) Calculation of forecast accuracy for forecast models and BERK's projections.

(5) Generation of population forecasts using forecast models, including full uncertainty.

(6) Summary of differences between projections from BERK vs. forecast models.

(7) Determine projection probability across a range of growth projections.

1 BERK's population projections and associated reports are erroneous.

Details on my technical review of BERK's work are in my comments submitted to Bellingham City Council (18 November 2013), Whatcom County Council (2 March 2014), COB Planning Director Sepler (9 April 2015), and Whatcom County Planning and Development Services (15 May 2015). Errors in BERK's analysis include the following.

(1) BERK's projections assume implicitly that population growth will occur without constraints. This assumption is invalid because many fiscal, economic, geographic, social, and environmental factors will constrain growth rates relative to prior years when such factors were not so limiting.

(2) BERK's assertion that the medium projection is "most likely" is erroneous. "Most likely" has a precise statistical interpretation, which contradicts BERK's assertion. Population data and analysis for Whatcom County show that the medium projection exceeds the "most likely" growth value by a substantial margin, greater than 13,000.

(3) BERK's assertion that their medium and high projections are comparable with historical average growth rates is incorrect. BERK's figures for "historical" average growth rates are biased values calculated using time intervals cherry-picked from periods with highest growth. BERK's biased "averages" are substantially larger than unbiased historical averages, as plotted in Figure 1.

(4) BERK calculated projection uncertainty using an invalid "method" that contradicts OFM statements about uncertainty and all statistical reason. BERK's application of arbitrary fudge factors without justification should be rejected as statistically fraudulent.

2 Population data for the county and city contain significant temporal structure.

The first step in statistical analysis (after clarifying one's question) is to look at the data. In the context of population projections, one of the most important patterns to look for is temporal structure in the data. Autocorrelation functions (ACFs) provide a simple way to evaluate temporal structure. The ACF for Whatcom County growth data shows significant positive correlations extending to a lag of three years (Figure 2). This could be interpreted as substantial temporal "inertia" in County growth data. Due to the magnitude and temporal reach of these autocorrelations, forecast models that include temporal structure should perform better than models that do not. In particular, the ACF suggests population growth in the near future will remain low, in contrast to BERK projections that assume immigration rates will rebound magically to values not observed since the real estate bubble a decade ago.

3 Forecast models produce projections markedly different from BERK's.

Responsible planning decisions cannot be informed by BERK's population projections, due to pervasive errors in BERK's analysis. Consequently, I developed projections independently, using standard forecasting methods for populations with variable data. Due to temporal structure in County population growth data evident in the ACF (Figure 2), I fit several autoregressive models to Whatcom County population data (archived by Washington's Office of Financial Management, OFM 2014). The simplest is a linear 1st order autoregressive model, Eq.(1). Because temporal feedbacks in population growth are nonlinear, I also fit a quadratic 1st order autoregressive model, Eq.(2). Finally, I fit a linear 3rd order autoregressive model to reflect third order temporal structure evident in the ACF (Fig. 2), Eq.(3). Projections generated by deterministic versions of the models are similar to each other, but differ substantially from BERK projections. (See section 6, below, for results from stochastic versions of the models.) Growth projections generated by these models are summarized in Table 1. Projections for the most accurate model, Eq.(2), are plotted in Figures 3 and 4.

Linear 1st order autoregressive model

$$X_t = a_0 + a_1 X_{t-1} + \varepsilon_t \tag{1}$$

where X_{t-i} is population growth in year $t-i$

a_i are fitted parameters

ε_t is normally distributed random variable, fitted to model residuals.

Quadratic 1st order autoregressive model

$$X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-1}^2 + \varepsilon_t \tag{2}$$

Linear 3rd order autoregressive model

$$X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + a_3 X_{t-3} + \varepsilon_t \tag{3}$$

4 Forecast accuracy is high for forecast models; nearly zero for BERK's projections.

I measured forecast accuracy for each model, using the standard metric for population modeling with highly variable data. Without a time machine, we cannot measure forecast accuracy against future census data. Instead, we can measure how accurately models forecast existing data. To do so, one selects a population growth value from the past as a starting point, applies the model to forecast subsequent population growth, and then compares the forecast with the actual data. Repeating this process systematically throughout the County's population data set, one obtains a record of model forecast accuracy spanning the entire range of observed population growth. Formally, this measure is known as the prediction coefficient of determination, R_p^2 (Turchin 1996).

I conducted forecast accuracy measurements for the three autoregressive models described in section 4, as well as the high, medium, and low projections provided by BERK. Results of these calculations are in Table 1, below. The forecast accuracy calculations show the autoregressive models perform very well. Their nearly ideal forecast accuracy is exceptional, given the high variability in Whatcom County growth data. In contrast, accuracy of all BERK projections is exceptionally low, equivalent to the most crude estimate using a simple average. Two of BERK's three projections are even less accurate than a simple average, indicating model failure.

5 Valid assessment of projection uncertainty must reflect uncertainty in the data.

To assess projection uncertainty, I ran simulations using each model with random inputs, selected from distributions with parameters estimated from model residuals. For each simulated year (through 2036), growth was projected with the deterministic model Eq.(1)-(3) plus a random number selected from a distribution determined by the uncertainty in the actual data. (Note: this approach is a valid way to determine projection uncertainty, which incorporates all the variability in the data. In contrast, BERK applied arbitrary fudge factors without justification to reduce uncertainty artificially.) I ran one thousand stochastic simulations for each model. Results are summarized in Table 2, below. Projection uncertainty for the most accurate model, Eq.(2), is plotted in Figures 3 and 4.

6 Five key differences between model forecasts and BERK's projections.

(1) Forecasts were generated from autoregressive models fit to Whatcom County population data, instead of BERK's simplistic and invalid assumption of unlimited growth.

(2) The models were fit in an unbiased way to all Whatcom County population data available from OFM, in contrast to BERK's references to biased "averages" calculated from high growth periods.

(3) The models incorporate temporal structure evident in Whatcom County population data. BERK projections ignore temporal structure. Worse, by assuming growth rates will jump to values far exceeding recent data, BERK projections contradict temporal structure in the data.

(4) The models include uncertainty fully, accurately, and genuinely. In contrast, BERK's projections constrain uncertainty with bias toward high growth, contrary to OFM's statements, using arbitrary fudge factors without justification.

(5) The autoregressive models perform well, with very high forecast accuracy ($R_p^2 \geq 0.85$). In contrast, BERK projections have such low forecast accuracy that they should not be considered forecasts at all. Forecast accuracy of BERK projections is equivalent to using the historical average growth rate to project future growth rates. On average, winter days in Bellingham are rainy and summer days are sunny. Those averages are not weather forecasts, nor should they be treated as such. For similar reasons, BERK projections should not be used to forecast population growth. Two BERK projections have negative forecast accuracy. A negative value for forecast accuracy indicates a failed model, which must not be used to guide planning in any way.

7 Projection probability and risk of over-projection vs. under-projection.

Population projections and projection methods used by BERK are erroneous and unreliable, as analysis above demonstrates. I developed credible and reliable projections using appropriate methods, described above. Your decision to adopt a projection should be informed by knowledge of the probability that actual growth will exceed or fall below that projection. To provide you that knowledge, I have estimated probabilities of over- and under-projection associated with all projections, including mine and BERK's low, medium, and high projections. Those probabilities and methods for estimating them are described below.

Forecast accuracy of the best autoregressive model, Eq.(2), is very high (Table 1). Accordingly, I applied that model with full uncertainty in the data to simulate Whatcom County population growth to 2036. After running 10,000 independent simulations, I determined the probability associated with all possible growth projections. Results are summarized in Figure 5, below. Technically, the figure is equivalent to a "probability density function," which should be interpreted as follows. Heights of vertical bars on the graph form a curve that is approximately bell-shaped. Highest probability values lie in the middle of the curve, with probability declining rapidly away from the middle. The area under the curve, or sum of heights of all bars, equals one (100% probability). Probability that growth will be smaller than a given projection is equal to the sum of all heights to the left of that projection. Probability that growth will exceed a given projection is equal to the sum of all heights to the right of that projection.

Figure 5 shows BERK's high projection will almost certainly exceed actual growth, with probability 96.5%. Conversely, there is only a 3.5% chance that growth will exceed the high projection. The medium projection also will exceed growth with high probability, 78.9%. There is only a small chance (21.1%) of under-projection with the medium projection. Probabilities of over- and under-projection are more balanced for BERK's low projection, 45.3% and 54.7% respectively.

You can apply Figure 5 to determine the probability that any projection will be high or low. If you would like assistance doing so, if you want more precise probability values for a projection than can be obtained from the graph alone, or if you have any other questions regarding population projections, I would be happy to help at your request.

<u>Translating probability into risk</u>.
For decision-makers, risk is defined as the cost of a future outcome multiplied by the probability it will occur.

$$\text{Risk} = \text{Cost} \times \text{Probability} \qquad (4)$$

Your decision to choose a given population projection involves two kinds of risks: risks associated with under- and over-projection. Total risk associated with a decision is the sum of both kinds of risks, described below, where $C()$ denotes cost and $P()$ denotes probability.

$$\text{Total Risk} = C(\text{underproject}) \times P(\text{underproject}) + C(\text{overproject}) \times P(\text{overproject}) \qquad (5)$$

Above I have provided you values for probabilities relevant to each projection. Planning staff can provide infrastructure cost estimates, which are greatest for projections that require UGA expansions. To make a fully informed decision, you also will need cost estimates for all infrastructure associated with growth, social costs for each projection, and environmental costs for each projection. While infrastructure costs and impacts of under-projection may be problematic, they can be mitigated. In contrast, most social and environmental costs of over-projection are irreversible. For example, further UGA expansion will cause extinction of wild Chinook salmon in the Nooksack basin within several decades with high probability (McLaughlin 2014). Salmon extinction would be an irreversible impact.

You also should include opportunity costs of each projection you consider. If you decide against expanding UGA boundaries now, you can exercise that option in the future. Once UGA boundaries are expanded, however, retracting them becomes very difficult even when growth falls short of projections. Indeed, growth in the near term is likely to remain well below projections recommended to you. As described in Section 3, Whatcom County and Bellingham population data show protracted inertia in growth patterns. Low growth values observed in recent years are likely to persist into the near future. Projections using the most accurate model show average growth of 2490/year over the next ten years. That is 80% of the medium projection and just 62% of BERK's high projection. In ten years, you will have completed your next Comprehensive Plan update and associated population projections. You would preserve the most options for the future by choosing a low projection now, and revising it as needed during the next update.

Conclusion: Projections under consideration, derived from BERK consulting, should be discarded for conceptual reasons: we should not make important decisions using erroneous projections or invalid models. BERK's projections also should be rejected for reasons of poor performance: their forecast accuracy is unacceptably low, equivalent or worse to using a simple average. Models that incorporate temporal structure observed in Whatcom County population data perform much better, with nearly ideal forecast accuracy. Coincidentally, projections generated by those models are lower than all but the lowest BERK projection. These low projections truly are "most likely" and provide the most responsible basis for comprehensive planning.

References

Efron B, and RJ Tibshirani. 1993. *An Introduction to the Bootstrap.* Chapman & Hall/CRC Press, Boca Raton, FL.

McLaughlin J. 2014. "Sustaining wild Chinook salmon: climate change dictates shrinking development." Presented at Washington Higher Education Sustainability Conference, Bellingham, WA.

Turchin P. 1996. Nonlinear time-series modeling of vole population fluctuations. *Res.Popul.Ecol.* 38:121-132.

Washington OFM. 2014. Historical estimates of April 1 population and housing for the state, counties, and cities. Olympia, WA.
[online] http://www.ofm.wa.gov/pop/april1/hseries/default.asp
(accessed 19 March 2015)

Table 1: Forecast accuracy results for autoregressive models and BERK projections. Prediction accuracy was estimated using ordinary cross-validation with one-step predictions (Efron and Tibshirani 1993). Models were used to predict population growth for the next year in the data set, given population growth in the prior year(s). These predictions were repeated for all years in the OFM data set for Whatcom County. Prediction accuracy was measured with prediction coefficient of determination (Turchin 1996), R_p^2.

$$R_p^2 = 1 - \frac{\sum_i \left(L_i - \hat{L}_i\right)^2}{\sum_i \left(L_i - \overline{L}\right)^2}$$

where L_i is log-transformed population growth in year i,

$\hat{L}_i$ is log-transformed predicted population growth in year i,

$\overline{L}$ is average log-transformed population growth.

Potential R_p^2 values range from 1, indicating perfect prediction, to $-\infty$. Models with R_p^2=0 are equivalent to using average population growth as a simple predictor containing no information about temporal patterns in the data. Models with $R_p^2 < 0$ cannot provide useful population forecasts.

Model	**Total growth 2014-2036**	**Annual Growth (average)**	**Prediction Accuracy, R_p^2**
1st order, quadratic	59,052	2,567	0.859
3rd order, linear	59,581	2,590	0.851
1st order, linear	59,235	2,575	0.846
BERK low	56,086	2,439	–0.213
BERK medium	68,111	2,961	–0.049
BERK high	86,149	3,746	0.043

Table 2: Population forecasts generated by models described in Equations (1)-(3). Population growth for the years 2014-2036 was simulated 1000 times using the stochastic version of each model. Uncertainty was included as a normally distributed random variable (ε_t), with mean and standard deviation fit to model residuals. Growth averages across all simulations for each model are in **bold** font. Uncertainty is expressed as 95% confidence limits for these averages and as averages plus or minus one standard deviation.

Model		**Total growth 2014-2036**	**Average Annual Growth**
1^{st} order, quadratic	Mean + SD	87,571	3,807
	Upper 95%CL	60,822	2,644
	Mean	**59,052**	**2,567**
	Lower 95% CL	57,283	2,491
	Mean – SD	30,534	1,328
3^{rd} order, linear	Mean + SD	88,493	3,848
	Upper 95%CL	61,375	2,668
	Mean	**59,581**	**2,590**
	Lower 95% CL	57,787	2,512
	Mean – SD	30,670	1,333
1^{st} order, linear	Mean + SD	87,652	3,811
	Upper 95%CL	60,998	2,652
	Mean	**59,235**	**2,575**
	Lower 95% CL	57,472	2,499
	Mean – SD	30,819	1,340

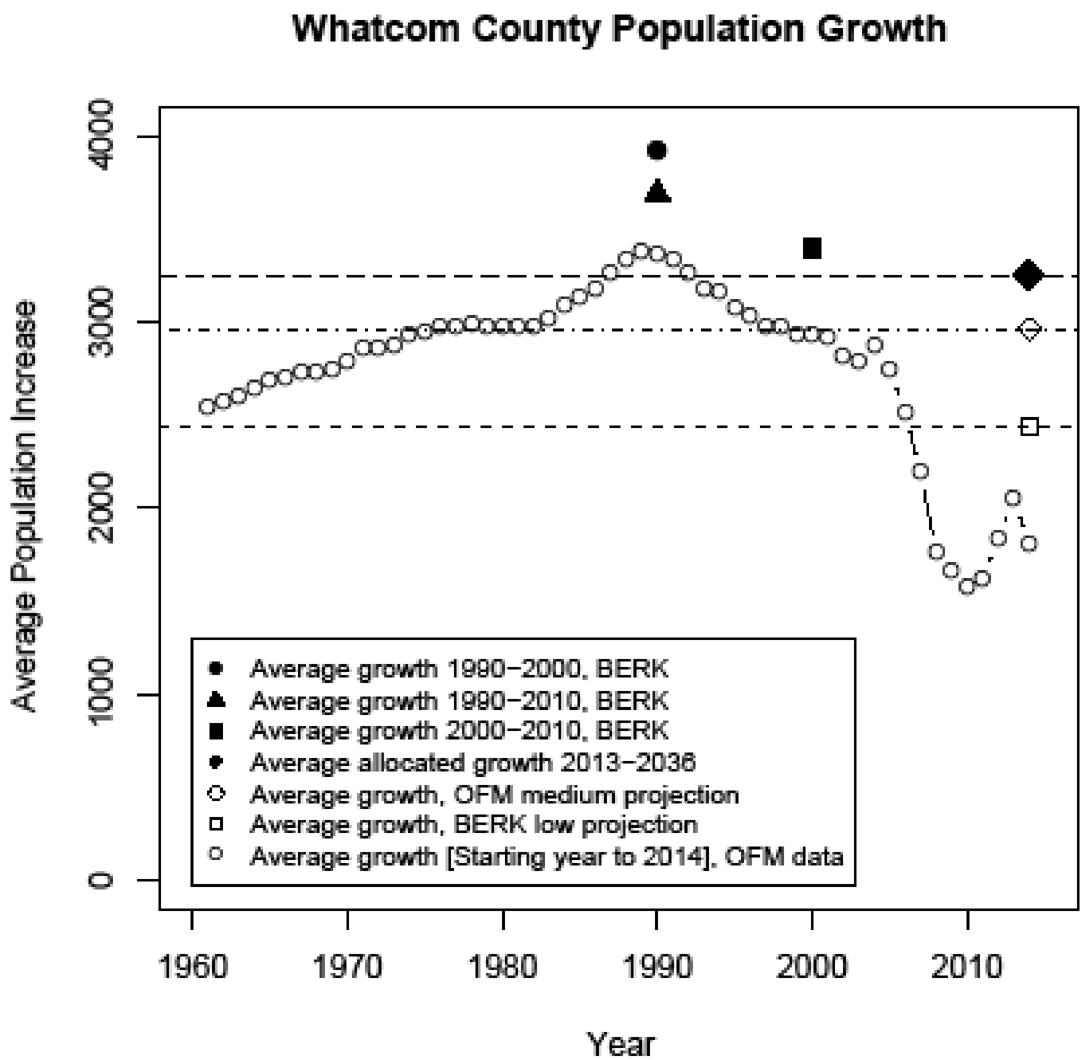

Figure 1. Whatcom County annual population growth: historical averages and projections. Solid symbols (circle, triangle, square) plot averages during high growth years, and the provisional "allocated projection" (solid diamond). Open circles plot historical averages calculated from the year on the x-axis to 2014. Solid diamond, open diamond, and open square plot BERK's high ("allocated"), medium, and low projections, respectively. Dashed lines extend horizontally through those projections, to facilitate comparisons with historical averages. Whatcom County population data are from OFM (2014).

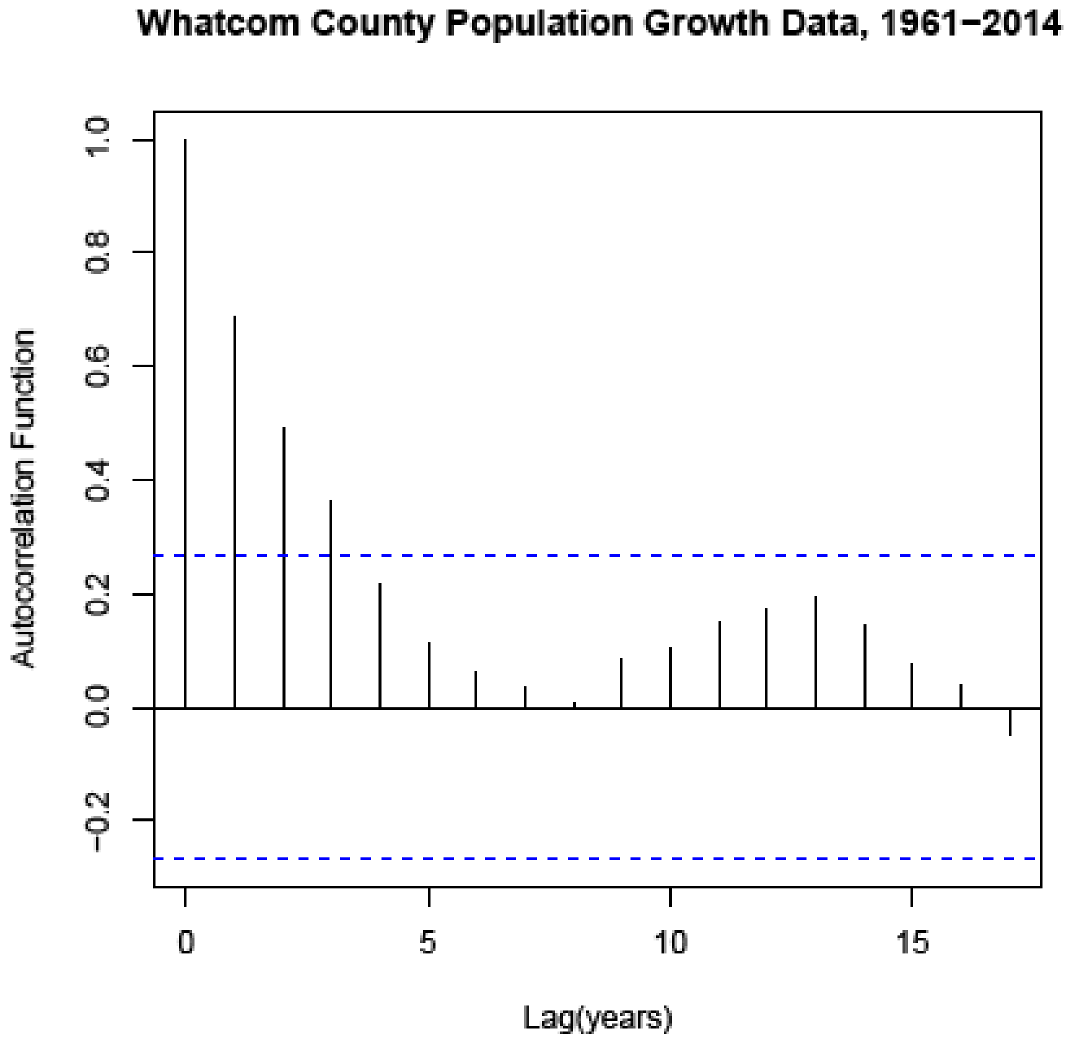

Figure 2. Autocorrelation function (ACF) for Whatcom County population growth data, obtained from Washington's Office of Financial Management (OFM 2014). Correlation between growth in the current year with growth in prior years is plotted as vertical lines. The abscissa (X-axis) represents the interval between years for which correlation was calculated. Correlations above the upper dashed line are statistically significant. The ACF equals one at zero lag because all years are perfectly correlated with themselves. The ACF for Whatcom County growth data shows significant positive correlation extending to a three year lag.

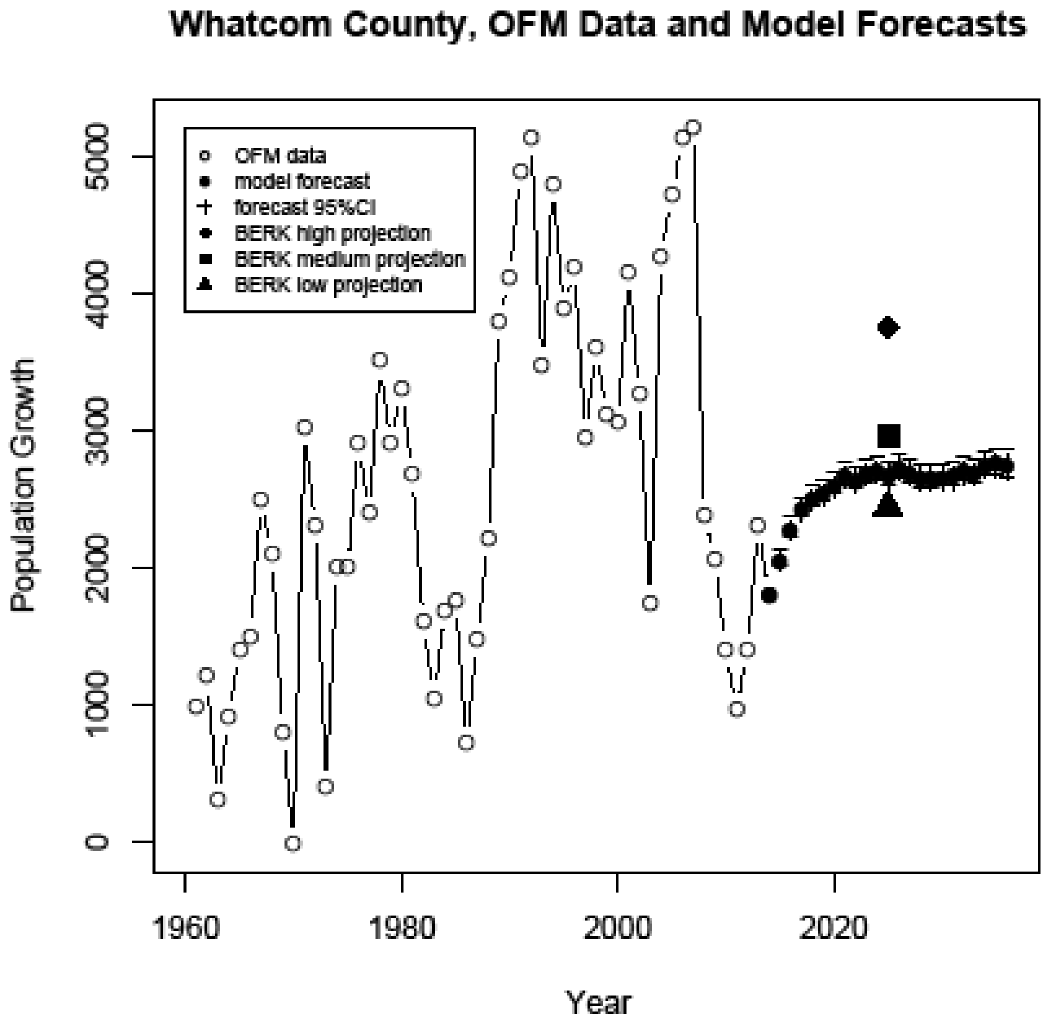

Figure 3. Population growth forecasts generated by the quadratic 1st order autoregressive model, Eq.(2). Population growth for the years 2014-2036 was simulated 1000 times using the stochastic version of the model. Uncertainty was included as a normally distributed random variable (ε_t), with mean and standard deviation fit to model residuals. Uncertainty is represented as 95% confidence limits for these averages, plotted with error bars. For comparison, average annual growth projections by BERK are plotted as a solid diamond (◆), square (■), and triangle (▲) for high, medium, and low projections, respectively.

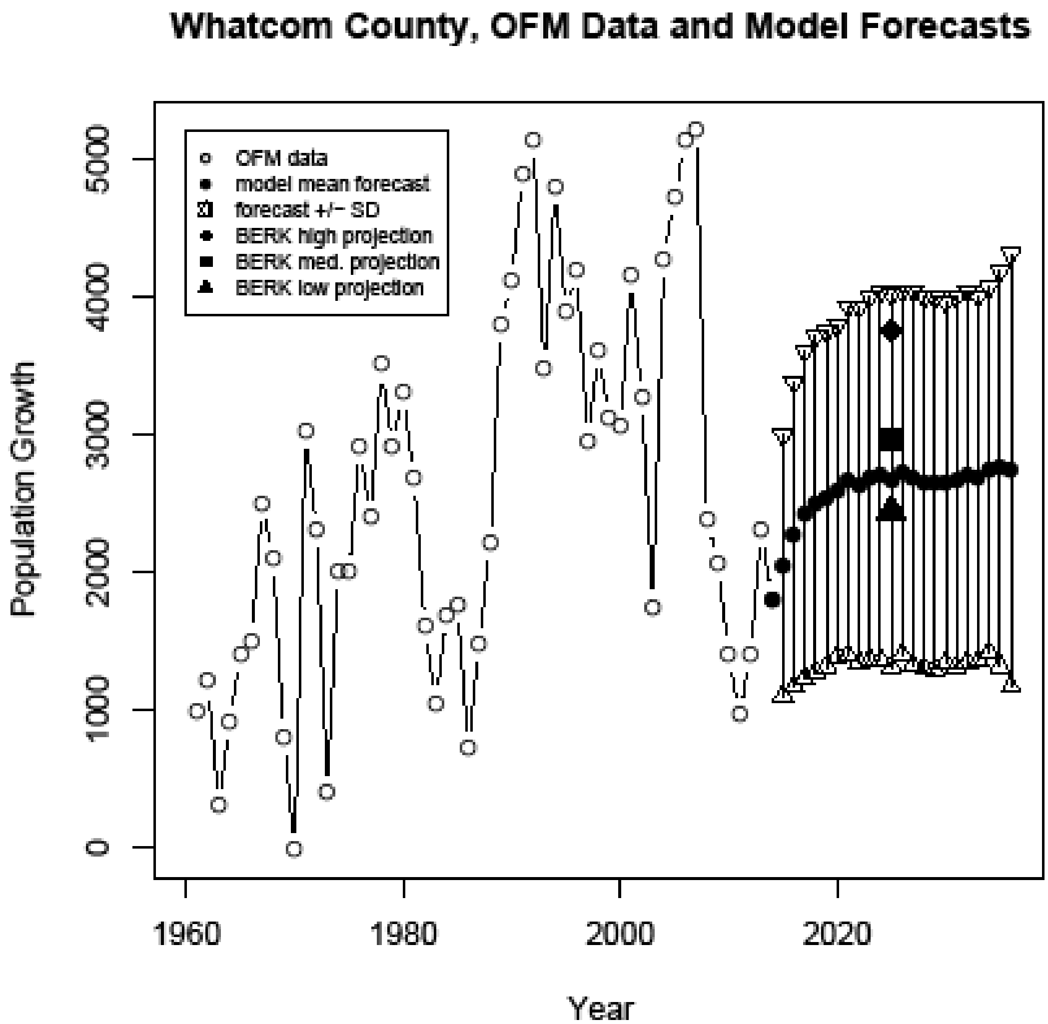

Figure 4. Population growth forecasts generated by the quadratic 1st order autoregressive model, Eq.(2). Population growth for the years 2014-2036 was simulated 1000 times using the stochastic version of the model. Uncertainty was included as a normally distributed random variable (ε_t), with mean and standard deviation fit to model residuals. Uncertainty is represented for each year as the average plus (▽) and minus (△) one standard deviation of the 1000 simulations t. For comparison, average annual growth projections by BERK are plotted as a solid diamond (◆), square (■), and triangle (▲) for high, medium, and low projections, respectively.

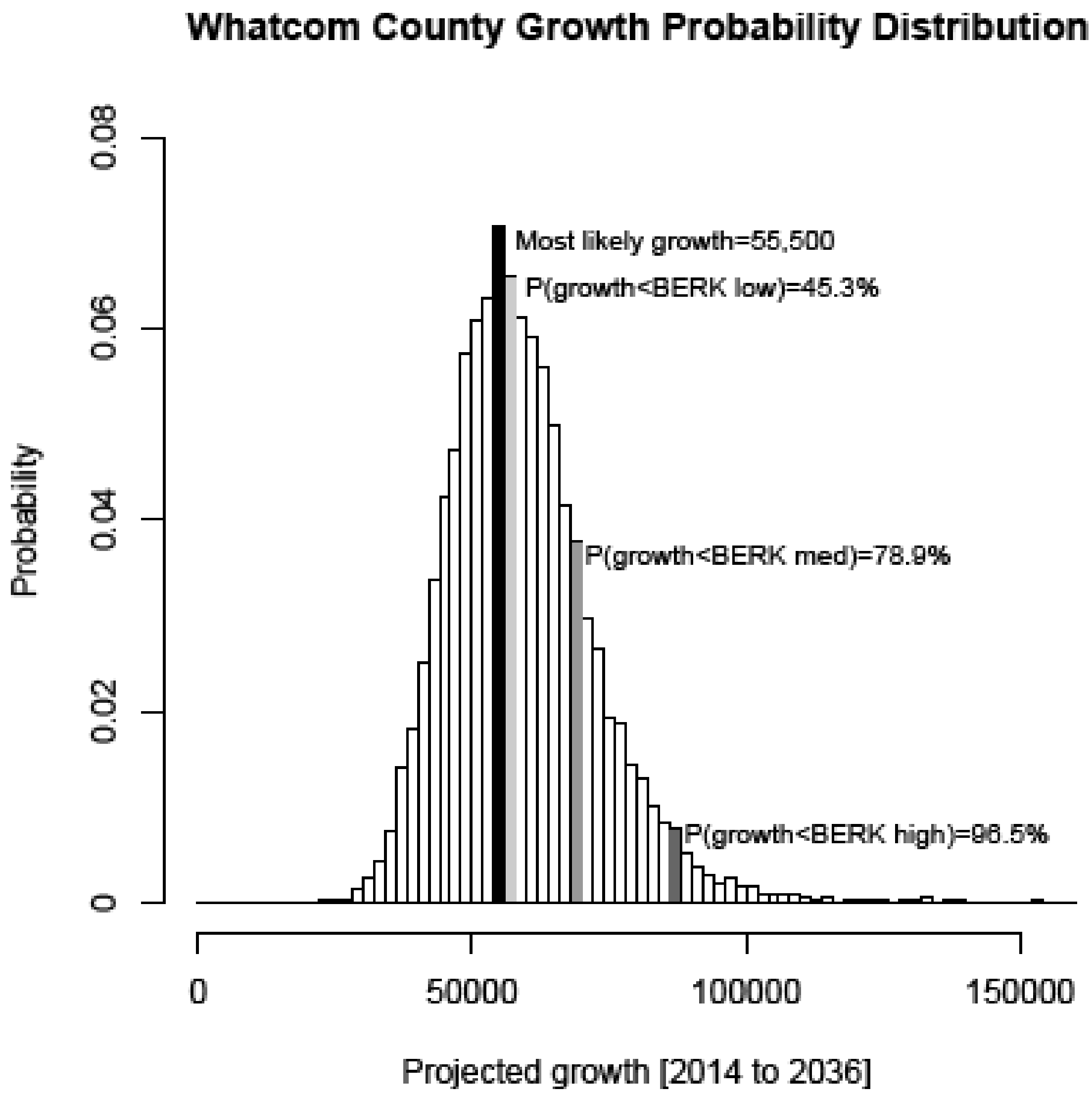

Figure 5. Probability distribution for Whatcom County population growth projections. Probabilities were determined using 10,000 stochastic simulations of a the quadratic 1st order autoregressive model, Eq.(2), fit to Whatcom County population data archived by the WA Office of Financial Management. Model uncertainty was included in simulations as an additive random variable with a distribution fit to model residuals. The horizontal axis is total projected population growth through the year 2036. Vertical bars plot the probability of growth falling within each 2000 growth increment. The most likely projection (55,500) lies within the highest bar, shaded in black. Low, medium, and high projections provided by BERK consulting lie within bars shaded light, medium, and dark gray, respectively. Estimated probability that BERK low, medium, and high projections will exceed actual population growth is 45.3%, 78.9%, and 96.5%, respectively, as indicated by text adjacent to shaded bars.