Investigating the Use of Holt-Winters Time Series Model for Forecasting Population at the State and Sub-State Levels

By

Amanda Walters

Qian Cai

Demographics and Workforce Section
Weldon Cooper Center for Public Service
University of Virginia

February 21, 2008

Investigating the Use of Holt-Winters Time Series Model for Forecasting Population at the State and Sub-State Levels

Amanda Walters, Qian Cai

University of Virginia

Introduction

Much of the research concerning the evaluation of time series models for population forecasting has been focused on ARIMA models at the national and sub-national levels, but few have evaluated the accuracy of these models at the sub-state level. Fewer still have explored the use of the Holt-Winters time series model at the state and local levels.

Most of the research evaluating the accuracy of the Holt-Winters time series model for population forecasting at the local level has been primarily limited to Europe. The most relevant study was conducted in the Netherlands; researchers at Erasmus University evaluated forecasts for over 800 municipalities using 154 different models, including time series models and extrapolation techniques (Openshaw et al., 1983). They found that the best forecasts are likely to be provided by either a Holt-Winters model, or a ratio-correction model, or a low order exponential-smoothing model.

Although few studies in the U.S. have developed and evaluated the Holt-Winters time series models for sub-national and sub-state areas, a recent study considered one Holt-Winters time series model. In the study, forecasts from the Holt-Winters model, as well as several ARIMA models, were evaluated in terms of precision and bias at the state level (Tayman et al., 2007). Although the Holt-Winters model was considered, the primary focus of the study was to evaluate the accuracy of the prediction intervals provided by time series models.

In this paper, we develop and evaluate the accuracy of the Holt-Winters time series model for population forecasting at the state and sub-state level. Using a series of annual population estimates from 1960 to 2005, we construct population forecasts for states in the U.S. and for Virginia's planning districts and counties. Additionally, we compare the accuracy of the Holt-Winters time series model with three types of population forecasting methods: trend extrapolation techniques, ARIMA time series models, and the cohort component method.

As a final note, we use the following forecasting terminology (Rayer et al., 2005) in this paper:

Base Year: the year of the earliest population size used to make a forecast.
Launch Year: the year of the latest population size used to make a forecast.

Target Year: the year for which the population size forecasted.
Base Period: the interval between the base year and launch year.
Forecast Horizon: the interval between the launch year and target year.

Methods

Using a set of annual population estimates from 1960 to 2005 for states in the U.S. and for Virginia's planning districts and counties, we evaluated the accuracy of the Non-Seasonal Holt-Winters model, as well as trend extrapolation techniques and ARIMA time series models. As a basis of comparison, we also evaluated the accuracy of the cohort component method for forecasts produced by two different sources: the U.S. Census Bureau and the Virginia Employment Commission. Each base period was 20 years and the forecast horizon varied from 5 to 20 years.

Virginia's counties were chosen for evaluation because we believe they are highly representative of counties in the United States. Virginia's 134 counties and county-equivalent cities have a variety of growth rates and patterns and also comprise over four percent of all counties in the United States. Evaluation was also done on Virginia's 21 planning districts, each of which consists of several counties and county-equivalent cities. The planning districts were chosen for evaluation in addition to the state's counties because we wanted to evaluate the accuracy of the Holt-Winters model at an intermediate sub-state level.

Forecast accuracy was measured in two ways. After comparing each forecast to the census count for the relevant target year, we calculated the mean absolute percentage error (MAPE) and the mean algebraic percentage error (MALPE). MAPE is the average when the direction of error is ignored and MALPE is the average when the direction of error is not ignored. The main difference between these two measures is that MAPE measures precision, while MALPE is a measure of bias (Tayman et al., 2007).

Holt-Winters Modeling

The Holt-Winters model is an exponential smoothing model. The simplest form of exponential smoothing involves assigning weights to a time series so that more recent observations are given more weight than older observations. More specifically, simple exponential smoothing weighs past observations with exponentially decreasing weights to forecast future values. The Holt-Winters model builds upon the simplest form of exponential smoothing and takes into account the possibility of a time series exhibiting some form of trend or seasonality, both of which are updated by simple exponential smoothing (NIST, 2006). Since the data used in this study did not exhibit any type of seasonality, the seasonal model will not be discussed in this paper.

In mathematical terms, the Holt-Winters (non-seasonal) model is:

$$S_t = \alpha Y_t + (1 - \alpha)(S_{t-1} + b_{t-1}) \qquad 0 \le \alpha \le 1$$
$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \qquad 0 \le \gamma \le 1$$

where S_t is the smoothed value, α and γ are smoothing constants, b_t is the trend, and Y_t is the actual value of the time series. The speed at which the older responses are dampened is a function of α and γ. When α and γ are close to 1, the dampening effect is quick; when α and γ are close to 0, the dampening effect is slow (NIST, 2006). For this study, α and γ are determined by minimizing the squared prediction error and the starting values for S and b are:

$$S_3 = Y_2$$
$$b_3 = Y_2 - Y_1$$

ARIMA Modeling

The general ARIMA model is expressed as ARIMA (p,d,q), which represents the order of the autoregressive term, degree of differencing, and the order of the moving average term, respectively. An autoregressive model is a linear regression of the current value of the series against one or more prior values of the series and a moving average is a linear regression of the current value of the series against white noise (NIST, 2006). Differencing is a type of transformation used if the time series is non-stationary.

We considered several different ARIMA models for this study. All time series were difference once, as one difference is usually sufficient (NIST, 2006). The optimal ARIMA model was chosen by Akaike's Information Criterion (AIC). For each individual area, all possible models were fitted to data from the base period and the AIC value was computed for each model fit. Akaike's Information Criterion measures goodness of fit; the smaller the AIC, the better the model fitting. Therefore, the model with the smallest AIC value was deemed optimal and forecasts were based on that particular model. The formula for Akaike's Information Criterion is as follows:

$$-2LL + 2p$$

where LL is the log-likelihood and p is the number of parameters fitted in the model.

Trend Extrapolation Techniques

There are several different types of trend extrapolation techniques. The most commonly used ones are linear, modified linear, share-of-growth, shift-share, exponential, and constant. For this study, we considered the linear and exponential extrapolation techniques. The linear technique assumes that the population will increase or decrease by the same number of persons in each future decade as the average per decade increase or decrease observed during the base period. The exponential technique assumes that the population will grow or decline exponentially in each future decade as during the base period. In mathematical terms, the linear and exponential techniques can be written as:

Linear: $P = O_1 + (x / y)(O_1 - O_n)$

Exponential: $P = O_1 \exp[(\ln(O_1 / O_n) / y)x]$

where x is the length of the forecast horizon, y is the length of the base period, P is the target year, O_1 is the population in the launch year, and O_n is the population in the base year (Rayer et al., 2005).

Cohort Component Method

In the cohort component method, components of change, such as fertility, mortality, and migration, are projected separately for each birth cohort. For the U.S. Census Bureau's projections, the base population is advanced each year by using projected survival rates and net migration by single year of age, sex, race, and Hispanic origin. Each year, a new birth cohort is added to the population by applying the projected fertility rates by race and Hispanic origin to the female population. The components of change are then individually applied to each of the resulting race/ethnic groups (Campbell, 1996).

Results

Results by Geographic Area

(a) Mean Absolute Percentage Error

Our research shows that population forecast errors for smaller areas are not necessarily higher than those for bigger areas. Although forecast errors seem to increase as the geographic area projected becomes smaller, the differences in accuracy are negligible. The differences in mean absolute percentage error (MAPE) between state projections and planning district projections did not exceed two percentage points for 5- and 10-year forecast horizons. Furthermore, for exactly half of these projections, the MAPE value for the planning districts was smaller than that of the states. As for county level projections, we found that these projections were consistently less accurate than state and planning district level projections; however, differences in MAPE values did not exceed five percentage points. These results are summarized in the table below.

Figure 1. All states and Virginia's planning districts and counties: mean absolute percentage error (MAPE) by base period and target year.

Model	Base Period	Forecast Horizon	Target Year	State	VA Planning District	VA County
ARIMA (p,1,q)	1966-1985	5	1990	2.5	3.4	4.3
	1976-1995	5	2000	2.6	1.6	3.1
	1961-1980	10	1990	4.2	6.4	8.5
	1971-1990	10	2000	6.3	5.1	7.7
	1961-1980	20	2000	6.6	11.2	14.0
Holt-Winters	1966-1985	5	1990	2.4	2.9	3.7
	1976-1995	5	2000	2.3	1.6	4.4
	1961-1980	10	1990	5.5	7.2	10.1
	1971-1990	10	2000	7.1	5.6	8.2
	1961-1980	20	2000	9.2	13.9	16.7
Linear	1966-1985	5	1990	3.3	3.8	4.4
	1976-1995	5	2000	2.9	1.6	3.3
	1961-1980	10	1990	4.6	6.3	8.3
	1971-1990	10	2000	5.3	4.9	7.9
	1961-1980	20	2000	6.9	10.9	13.5
Exponential	1966-1985	5	1990	3.6	3.3	3.8
	1976-1995	5	2000	2.4	2.1	3.7
	1961-1980	10	1990	4.8	5.7	8.3
	1971-1990	10	2000	4.6	4.7	8.2
	1961-1980	20	2000	5.9	8.9	14.2

(b) Mean Algebraic Error

We also found that population forecasts for smaller geographic areas generally exhibit similar biases to those for larger areas. Although the direction of the bias, in most cases, differed by geographic area, the degree of bias did not differ significantly. In fact, the difference in the absolute value of the mean algebraic percentage error (MALPE), which measures the degree of bias, did not differ more than three percentage points between geographic areas for 5- and 10-year forecast horizons. Furthermore, only 33 percent of the MALPE values for the state projections were larger in absolute value than those for the planning districts and counties. These results are summarized in the table below.

Figure 2. All states and Virginia's planning districts and counties: mean algebraic percentage error (MAPE) by base period and target year.

Model	Base Period	Forecast Horizon	Target Year	State	VA Planning District	VA County
ARIMA (p,1,q)	1966-1985	5	1990	-0.8	1.9	1.8
	1976-1995	5	2000	2.2	0.9	1.0
	1961-1980	10	1990	-0.7	1.6	1.1
	1971-1990	10	2000	5.2	4.4	5.0
	1961-1980	20	2000	3.5	5.7	6.0
Holt-Winters	1966-1985	5	1990	-0.3	1.9	1.4
	1976-1995	5	2000	1.8	-0.9	-1.3
	1961-1980	10	1990	-1.0	2.6	-0.4
	1971-1990	10	2000	6.1	3.2	3.3
	1961-1980	20	2000	3.2	8.6	3.8
Linear	1966-1985	5	1990	-1.3	0.9	0.9
	1976-1995	5	2000	2.6	0.5	0.7
	1961-1980	10	1990	-1.2	0.0	-0.2
	1971-1990	10	2000	4.2	3.3	4.0
	1961-1980	20	2000	2.6	4.0	4.1
Exponential	1966-1985	5	1990	-2.5	0.2	-0.1
	1976-1995	5	2000	1.8	-0.4	-0.5
	1961-1980	10	1990	-3.7	-1.9	-2.9
	1971-1990	10	2000	2.1	1.0	1.0
	1961-1980	20	2000	-3.6	-0.6	-2.4

Results by Model Type

(a) Mean Absolute Percentage Error

Our research shows that overall accuracy of time series models is comparable to trend extrapolation techniques. As shown in Figure 1, the MAPE values for the ARIMA and Holt-Winters models are very similar to the MAPE values for the linear and exponential trend extrapolation techniques. In fact, the differences in MAPE values for any two models does not exceed 5 percentage points, with the average difference being about 1 percentage point.

Although the differences in MAPE values are relatively small, the Holt-Winters model seems to be the most accurate model for the 5-year projection period and the exponential model seems to be the most accurate model for the 10-year projection period. MAPE values were calculated for a total of six 5-year projection periods and six 10-year projection periods. Five of the six MAPE values for the 5-year projection periods were the lowest for the Holt-Winters model and four of the six MAPE values for the 10-year projection period were the lowest for the exponential model.

Our research also shows that the overall accuracy of time series models is comparable to cohort component models in addition the trend extrapolation techniques, regardless of geographic area. In the 1990's, both the U.S. Census Bureau and the Virginia Employment Commission (VEC) created a series of population projections using the cohort component method. The U.S. Census Bureau's projections were based on 1994 data and state projections were made in 5-year intervals for 1995 through 2025. The VEC's projections were based on 1990 data and projections were made for Virginia's planning districts and counties in 10-year intervals for 2000 through 2030. In the figure below, we have calculated the accuracy of the cohort component method for projections of the most recent Census year made by both the U.S Census Bureau and the VEC.

Figure 3. All states and Virginia's planning districts and counties: mean absolute percentage error (MAPE) for the cohort component method.

Projections	Launch Year	Forecast Horizon	Target Year	State	VA Planning District	VA County
Census Bureau	1994	6	2000	2.7	-	-
VEC	1990	10	2000	-	5.6	7.1

In order to evaluate U.S. Census Bureau's projections against projections made by the Holt-Winters or ARIMA time series models, we needed to obtain time series projections at the state level for a 6-year forecast horizon. Using the base period of 1975-1994, we develop time series projections at the state level that are analogous to the Census Bureau's projections (Figure 3). The resulting MAPE values were 2.7 for the Holt-Winters model and 3.0 for the ARIMA model (Figure 4). Thus, in terms of forecasting state level population in the short-term, the overall accuracy of the Holt-Winters and ARIMA time series models is comparable to that of the cohort component model.

Figure 4. States: mean algebraic percentage error (MAPE) by base period and target year.

Projections	Base Period	Forecast Horizon	Target Year	MAPE
ARIMA	1975-1994	6	2000	3.0
Holt-Winters	1975-1994	6	2000	2.7
Census Bureau	1994*	6	2000	2.7

*Launch Year

As for the VEC's projections, the overall accuracy does not differ significantly from the overall accuracy of the time series projections. At the planning district level, MAPE is 5.6 for the cohort component model, 5.6 for the Holt-Winters model, and 5.1 for the ARIMA model for a 10-year forecast of the population in 2000. For projections at the county level with the same target year and forecast horizon, MAPE is 7.1 for the cohort component model, 8.2 for the Holt-Winters model, and 7.7 for the ARIMA model (Figures 1, 3). Thus, even at the smallest geographic area, we found that the accuracy of time series models differs from that of the cohort component model, at the most, by only 1.1 percentage points.

Furthermore, we examined an additional state to verify the accuracy of both the cohort component and time series models at the county level. The Office of Financial Management (OFM) in Washington produces county population projections for the state using the cohort component model. The set of projections used for this study were developed in 1995 and county projections were made Washington's counties in 5-year intervals for 2000 through 2010 and 1-year intervals for 2011 through 2020. In order to compare the accuracy of time series models, as well as the cohort component method, for forecasting population at the county level across different states, we chose a forecast horizon of 10 years. Since the launch year for OFM's projections is 1995, we will have to use estimates rather than Census counts to evaluate the accuracy of a 10-year forecast. Although this method is not ideal, it should be sufficient to verify that both the cohort component and time series methods will result in MAPE values of roughly 7 to 9 percent.

Using county population estimates obtained through OFM, we developed time series projections for each of the thirty-nine counties in Washington state using both the Holt-Winters and ARIMA models. The results are summarized in the figure below.

Figure 5. Washington's counties: mean algebraic percentage error (MAPE) by base period and target year.

Projections	Base Period	Forecast Horizon	Target Year	MAPE
ARIMA	1976-1995	10	2005	5.2
Holt-Winters	1976-1995	10	2005	8.8
OFM	1995*	10	2005	8.6

*Launch Year

The results seem to indicate that time series models perform consistently across states. As shown in Figure 5, the accuracy of the Holt-Winters and ARIMA models for Washington does not change significantly from that of Virginia. More importantly, the results demonstrate that the accuracy of the Holt-Winters and ARIMA time series models do not differ significantly from the accuracy of the cohort component model for Virginia, as well as Washington. In fact, the MAPE value for the ARIMA model is approximately 3.4 percentage points less than the resulting MAPE value for the cohort component model (Figure 5).

(b) Mean Algebraic Percentage Error

In terms of mean algebraic percentage error, no particular model seems to be exceptionally biased. However, the exponential model is the only model that, on the whole, has a negative bias; two-thirds of the MALPE values for the exponential model are negative, whereas only one-third or less of the MALPE values for the other models are negative (Figure 2).

As for the cohort component method, forecasts seem to exhibit a small, but insignificant, positive bias. Figure 6, below, displays the MALPE results for the U.S. Census Bureau and the VEC's projections based on cohort component method. Relative to the bias of the ARIMA and the Holt-Winters time series models, the cohort component method has similar MALPE values at the planning district county level. For instance, the MALPE value is 4.7 for the cohort component model, 3.3 for the Holt-Winters model, and 5.0 for the ARIMA model for a 10-year forecast of the population in 2000. As for state level forecasts, the MALPE value is 1.4 for the U.S. Census Bureau's projections, 1.9 for the Holt-Winters model, and 2.4 for the ARIMA model for a 6-year forecast of the population in 2000.

Figure 6. All states and Virginia's planning districts and counties: mean algebraic percentage error (MALPE) for the cohort component method.

Projections	Launch Year	Forecast Horizon	Target Year	State	VA Planning District	VA County
Census Bureau	1994	6	2000	1.4	-	-
VEC	1990	10	2000	-	4.5	4.7

Discussion

A substantial amount of the current research in the area of population forecasting has been devoted to examining time series models. However, time series models are considerably more difficult to apply nor have been found to be more accurate than simpler methods (Tayman et. al., 2007). For these reasons, most demographers rely on the cohort component method or trend extrapolation techniques for developing population projections. Nevertheless, they offer some major advantages compared to other methods: they provide prediction intervals and require less time and data to develop forecasts.

In this study, we developed two individual time series models, the Holt-Winters and ARIMA models, applied them to state and sub-state data, and evaluated the accuracy of the resulting forecasts. We evaluated these forecasts by comparing them to the Census Bureau's population counts for the corresponding target years. Furthermore, we evaluated the accuracy of these forecasts relative to forecasts developed using the cohort component and trend extrapolation models. Although the results of this study are based on a small projection timeframe, a number of interesting patterns were evident.

One of the most interesting patterns that emerged was that population forecast errors for smaller areas are not necessarily higher than those for bigger areas. This was especially true when comparing state level and planning district level forecast accuracy. Although we found that county level projections for Virginia are consistently less accurate than state level, this result may change if we were to include all U.S counties in the calculation of MAPE. Nonetheless, time series projections at the county level appear to be sufficiently accurate, relative to current projections produced by the U.S. Census and state agencies using the cohort component method.

Another interesting result from this study concerns the Holt-Winters time series model. Few researchers in area of demography have explored the possibility of using the Holt-Winters model for population forecasting; however, we believe that this paper provides preliminary evidence that Holt-Winters model provides both feasible and accurate population forecasts at the state and sub-state levels. Our research shows that the accuracy of both the Holt-Winters and ARIMA models is comparable to the accuracy of forecasts produced by the much more time-consuming cohort component method. In addition to being less time-consuming, the software needed to develop the Holt-Winters and ARIMA projections is accessible from the internet with no cost. Therefore, time-series models are cost-effective and time-effective than the cohort component method.

Although time series models may be preferable to the cohort component model for total population projections, we want to emphasize the fact that there will still be a need for the cohort component model in order to develop age and sex specific projections. However, total population forecasts are commonly used for planning and diagnostic purposes at the local level (Rayer et. al., 2005). Given that the U.S. Census Bureau only produces projections at the state level, it is up to the individual states to develop projections for their own counties and other sub-state entities and the methods outlined in this paper provide an accurate, as well as cost- and time-effective way of producing population forecasts at the local level.

In terms of the Holt-Winters model versus the ARIMA model, we recommend using the Holt-Winters model in most cases. Given their similar performance, the Holt-Winters model is more intuitive. It makes sense that more recent observations should be given more weight than observations further in the past. Furthermore, model specification, which is required when applying ARIMA models to population data, requires a high level of expertise and substantial time commitment (Tayman et al., 2007). However, with the Holt-Winters model and statistical software used in this paper, no specification is required.

Yet, many questions remain to be answered before we can draw any solid conclusions. Would shorter or longer base periods result in more accurate forecasts for the Holt-Winters model? Are the prediction intervals that accompany the Holt-Winters forecasts realistic and accurate? Is the accuracy of prediction intervals affected by the size of the geographic area? Can combining the Holt-Winters and ARIMA models used in this study improve forecast accuracy? Nonetheless, we believe that with further research focused on the Holt-Winters time series model, demographers and decision makers will have a practical and intuitive method for producing accurate total population forecasts that can be applied to state and local data.

References

Campbell, P.R. 1996. Population Projections for States by Age, Sex, Race, and Hispanic Origin: 1995 to 2025. U.S. Census Bureau. Retrieved January 14, 2008, from: http://www.census.gov/population/www/projections/ppl47.html.

Openshaw, S., & Van Der Knaap, G.A. 1983. Small area population forecasting: some experience with British models. *Journal of Economic and Social Geography*, 74(4), 291-304.

NIST/SEMATECH. 2006. *e-Handbook of Statistical Methods*. Retrieved January 14, 2008, from: http://www.istl.nist.gov/div898/handbook.

Rayer, S., Smith, S.K., and Tayman, J. 2005. Prediction Intervals for County Population Forecasts. Paper presented at the annual meeting of the Southern Demographic Association, Oxford, MS.

Tayman, J., Smith, S.K., and Lin, J. 2007. Precision, bias, and uncertainty for state population forecasts: an exploratory analysis of time series models. *Population Research and Policy Review*, 26, 347-369.