Stable Seasonal Pattern (SSP) Model

Thomas B. Fomby
Department of Economics
Southern Methodist University
August 2010

I. Introduction

Sometimes forecasters have to work with very small data sets. For example, suppose we owned a Wholesale Toy Distribution Company, the XYZ Toy Company, and that after 6 months of developing the full-scale of our company, we have the following 12 months of sales figures (in dollars) to work with.

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
964,977	2,699,324	884,494	1,035,007	1,930,143	1,124,814	1,098,136	1,812,798

Sep	Oct	Nov	Dec
1,095,294	1,163,039	1,920,424	1,000,743

Suppose we want to forecast our sales for the 12 months of next year. How do we go about doing this? One approach is to use the **Stable Seasonal Pattern (SSP) Model** Approach. Here is the logic of this approach.

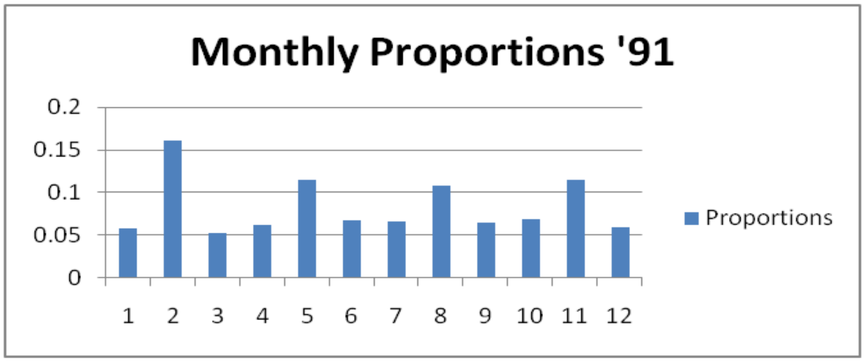
The total sales for the year (January – December) were $16,729,193. The proportion of the total sales for the year as distributed out to each month is as follows:

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
0.057682	0.161354	0.052871	0.061868	0.115376	0.067237	0.065642	0.108361

Sep	Oct	Nov	Dec
0.065472	0.069522	0.114795	0.059820

Notice that these proportions add to one. A plot of these proportions is displayed below.

Graph of
Monthly Proportions
Used in SSP Method



As you can see the proportions are not evenly distributed over the 12 months of the year (i.e. the proportions are not 0.08333 per month). Therefore, one might say that the above proportions (and thus the sales figures) display a certain **seasonal pattern**. In the present case the months of February, May, August, and November appear to be "stronger" months while the remaining months seems to reflect about the same amount of lesser business per month. So if we are going to forecast ahead the sales of our company, we should probably take this seasonal pattern into account.

One of the major assumptions of the SSP model is that the seasonal pattern represented in the graph above will be maintained (stable) in the future and, therefore, we could apply it to our forecasting problem for next year and into the future as well. All that remains for us to do is to come up with our best guess as to what next year's total sales are going to be and then we could use the above proportions to "distribute" the total among the various months of next year. Given no more information than we currently have, we are going to have to rely on someone's **domain-specific knowledge** about our industry. Suppose we contact the Executive Director of Research of the Wholesale Toy Sales Association and ask her, "What is the typical growth rate in the sales of first year Toy Wholesalers?" If she comes back with the answer that, of the 10 Toy Wholesalers that are comparable to XYZ Inc., the average first years sales growth is 5%, then we have a way of projecting next year's total sales for our company. We simply multiply our last year's total sales $16,729,193 by 1.05 to get the figure of $17,565,653. Once we have the next year's projection of the total year's sales, we can use the above proportions to distribute the forecasted sales for next year, month-by-month. Doing this we get the monthly forecasts for next year of

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
1,013,226	2,834,290	928,719	1,086,757	2,026,650	1,181,055	1,153,043	1,903,438

Sep	Oct	Nov	Dec
1,150,059	1,221,191	2,016,445	1,050,780

Of course, when we add up the above monthly projections we get our predicted next year's total of $17,565,653 because the proportions that we used in making the monthly projections add to one.

II. Notation and Implementation of the SSP Model

Now let us generalize the above simple example to the case where we have n years of monthly data, where $n \geq 2$. Consider the following tabular form of the data.

Months

Year	1	2	3	….	10	11	12
1	y_{11}	y_{12}	y_{13}	….	$y_{1,10}$	$y_{1,11}$	$y_{1,12}$
2	y_{21}	y_{22}	y_{23}	….	$y_{2,10}$	$y_{2,11}$	$y_{2,12}$
3	y_{31}	y_{32}	y_{33}	….	$y_{3,10}$	$y_{3,11}$	$y_{3,12}$
.	.	.	.	….	.	.	.
.	.	.	.	….	.	.	.
.	.	.	.	….	.	.	.
n	y_{n1}	y_{n2}	y_{n3}	….	$y_{n,10}$	$y_{n,11}$	$y_{n,12}$

Here y_{ij} $(i = 1,2,\cdots,n; j = 1,2,\cdots,12)$ represents the sales of a company in the j-th month of the i-th year. Further let us represent the totals of each of the n years as follows:

$$y_1^T = \sum_{j=1}^{12} y_{1j}\ ,\ y_2^T = \sum_{j=1}^{12} y_{2j}, \cdots, y_n^T = \sum_{j=1}^{12} y_{nj}. \qquad (1)$$

That is, the i-th year's total is represented by $y_i^T = \sum_{j=1}^{12} y_{ij}$.

In the spirit of the SSP approach, the proportion of the total year's sales from month j in the i-th year is given by

$$P_{ij} = \frac{y_{ij}}{y_i^T} \quad , \ i = 1,2,\cdots,n; j = 1,2,\cdots,12. \qquad (2)$$

Of course, we can use these proportions to judge, roughly speaking, if there is seasonality in our data or not. If $P_{ij} \approx \frac{1}{12} = 0.08333$ for each month, j, in all years, i, then there isn't a significant amount of seasonality in the data. (We will talk about some graphical methods for detecting seasonality, namely Buys-Ballot plots, and a non-parametric statistical test (Friedman's test) for seasonality later.) On the other hand, if there is a significant difference in the proportions across months within each year and if this pattern is fixed and persists from one year to the next (stable seasonal pattern), then we can use this seasonal pattern to our advantage when forecasting future values of y.

If, in fact, the monthly seasonal pattern of our data is stable over time, we can get better estimates of the stable monthly proportions, say $\pi_1, \pi_2, \cdots, \pi_{12}$, by estimating them with the average of the monthly proportions, the averages taken for given month j over the n years as in

$$\bar{P}_j = \frac{\sum_{i=1}^{n} P_{ij}}{n} = \frac{P_{1j} + P_{2j} + \cdots + P_{nj}}{n} \quad (3)$$

for $j = 1,2, \cdots, 12$. Notice that, across the 12 months, these averages add to 1, i.e.,

$$\sum_{j=1}^{12} \bar{P}_j = \bar{P}_1 + \bar{P}_2 + \cdots + \bar{P}_{12} = 1 \quad . \quad (4)$$

Now, to begin, let's assume we have only two years of monthly data, $y_{11}, y_{12}, \cdots, y_{1,12}$ and $y_{21}, y_{22}, \cdots, y_{2,12}$. What we need to do is predict next year's total y_3^T by say, $\hat{y}_3^T$, and then distribute this predicted total to next year's months by using the estimated stable monthly proportions $\bar{P}_j$, $j = 1,2, \cdots, 12$, calculated over the two years of data.

One way to predict next year's total is to use what is called "last-year's change" rule. That is, last year's change in the total from year 1 to year 2, $y_2^T - y_1^T$, when added to year two's total, y_2^T, will give us a prediction of next's year total via the formula

$$\hat{y}_3^T = y_2^T + (y_2^T - y_1^T) \quad . \quad (5)$$

Then the j-th month's predicted value for the following year can be computed as

$$\hat{y}_{3,j} = \hat{y}_3^T \cdot \bar{P}_j \qquad , j = 1,2, \cdots, 12. \quad (6)$$

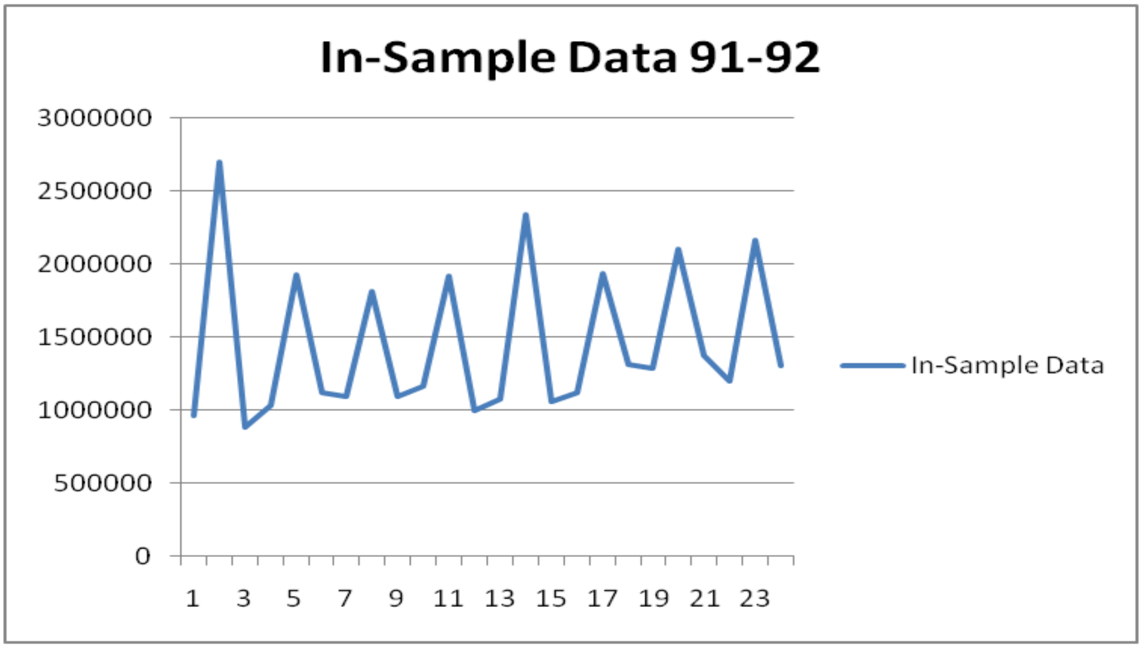
Obviously, since $\sum_{j=1}^{12} \bar{P}_j = 1$ we can see that

$$\sum_{j=1}^{12} \hat{y}_{3,j} = \hat{y}_3^T \cdot \sum_{j=1}^{12} \bar{P}_j = \hat{y}_3^T \quad (7)$$

and the monthly predictions add up to the predicted total and the predictions are congruent (internally consistent).

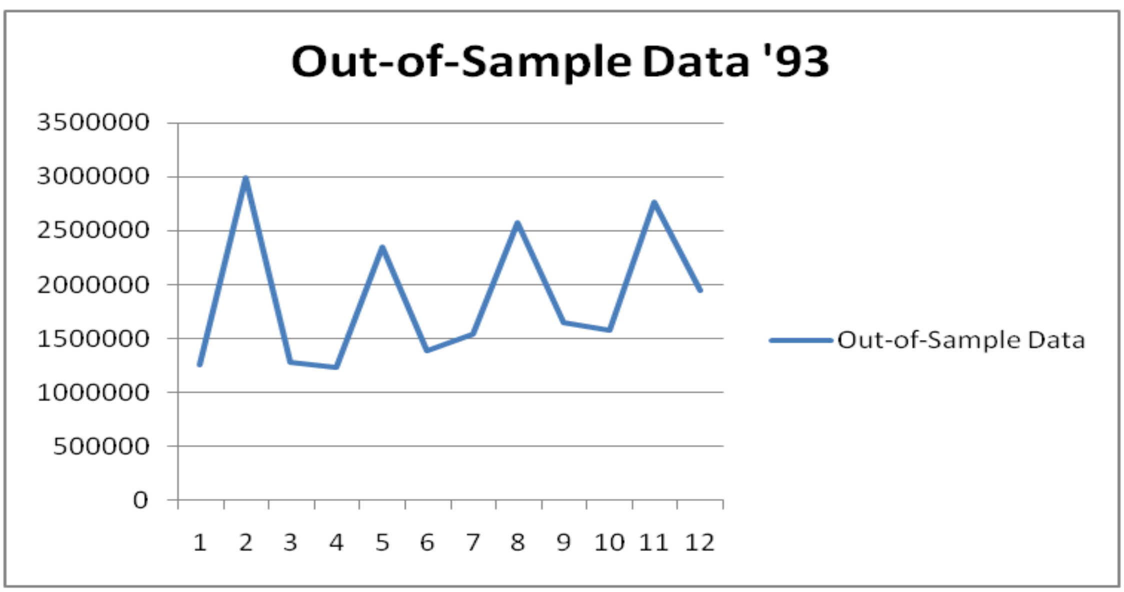
For example, consider the following 1991 – 1993 monthly data taken from the Plano Sales Tax Revenue data base. Let's call this the **in-sample data set**.

In-Sample Observations
January 1991 – December 1992



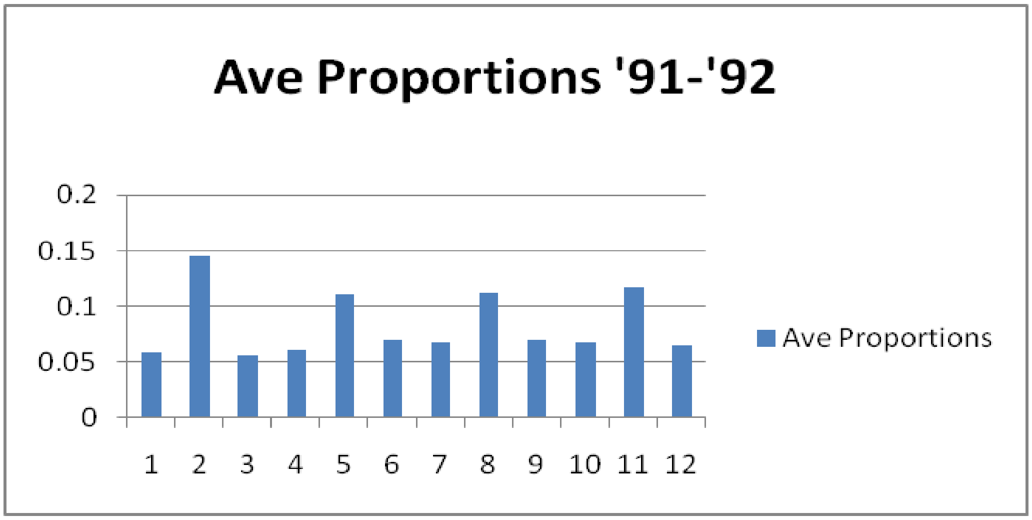
Suppose we wish to predict the monthly tax revenues for Plano for the next year (1994). The actual sales that were realized in 1994 are plotted below. How well will the SSP model forecast these values given only the data from 1991 – 1992? Let us call the below 1994 observations the **out-of-sample observations**.

Out-of-Sample Observations
January 1993 – December 1993



From the EXCEL file "Plano SSP-91-93.xls" we get the following results:
The first and second year totals are $y_1^T = 16{,}729{,}193$ and $y_2^T = 18{,}283{,}868$. The average proportions (taken over the two years) are $\bar{P}_1 = 0.058259$, $\bar{P}_2 = 0.144699$, $\bar{P}_3 = 0.05549$, $\bar{P}_4 = 0.061587$, $\bar{P}_5 = 0.110736$, $\bar{P}_6 = 0.069331$, $\bar{P}_7 = 0.067958$, $\bar{P}_8 = 0.111578$, $\bar{P}_9 = 0.070349$, $\bar{P}_{10} = 0.067611$, $\bar{P}_{11} = 0.116611$, and $\bar{P}_{12} = 0.065491$. These average proportions are plotted below and clearly indicate the presence of seasonality in the Plano Sales Tax Revenue data. The months February, May, August, and November stand out as being relatively strong months for tax receipts as compared to the rest of the months in the year.

Average Monthly Proportions Taken over the Years 1991 and 1992



Now using "last year's change" rule, we get next year's predicted total,

$$\hat{y}_3^T = y_2^T + (y_2^T - y_1^T) \quad (8)$$

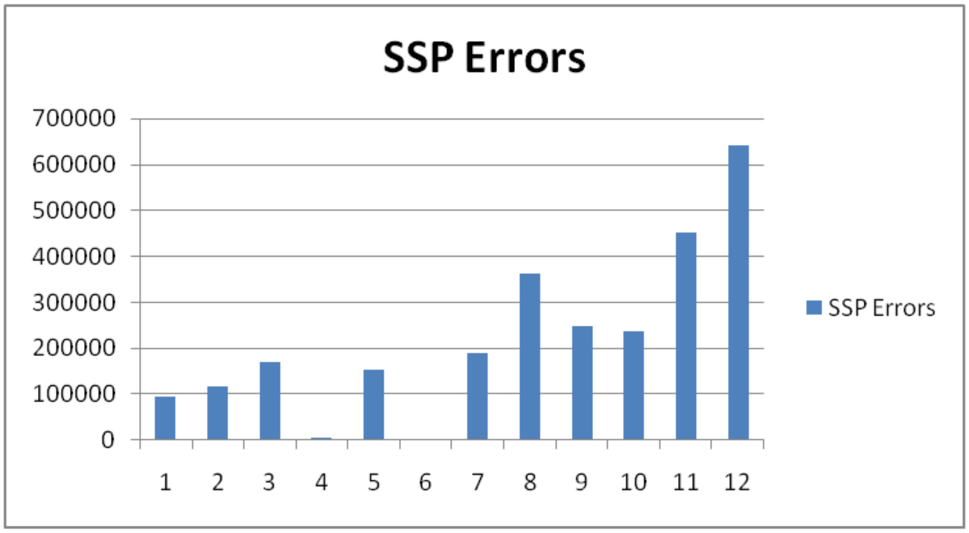
$$= 18{,}283{,}868 + (18{,}283{,}868 - 16{,}729{,}193)$$

$$= 19{,}838{,}543.$$

Once we apply the above average monthly proportions to $\hat{y}_3^T$, we get the following predictions for 1993 monthly tax receipts in Plano: $\hat{y}_{31} = 1{,}155{,}783$, $\hat{y}_{32} = 2{,}870{,}612$, $\hat{y}_{33} = 1{,}100{,}839$, $\hat{y}_{34} = 1{,}221{,}793$, $\hat{y}_{35} = 2{,}196{,}850$, $\hat{y}_{36} = 1{,}381{,}380$, $\hat{y}_{37} = 1{,}348{,}191$, $\hat{y}_{38} = 2{,}213{,}545$, $\hat{y}_{39} = 1{,}395{,}622$, $\hat{y}_{3,10} = 1{,}341{,}299$, $\hat{y}_{3,11} = 2{,}313{,}385$, and $\hat{y}_{3,12} = 1{,}299{,}244$. Notice these monthly forecasts sum to the forecasted year total for 1993, namely, 19,838,543, by construction.

Comparing these predictions with the actual values in the out-of-sample data set, we plot below the errors of our forecasts where the errors are defined by e = $y_{i,j} - \hat{y}_{i,j}$, where the $y_{i,j}$ represent the actual monthly values for 1993.

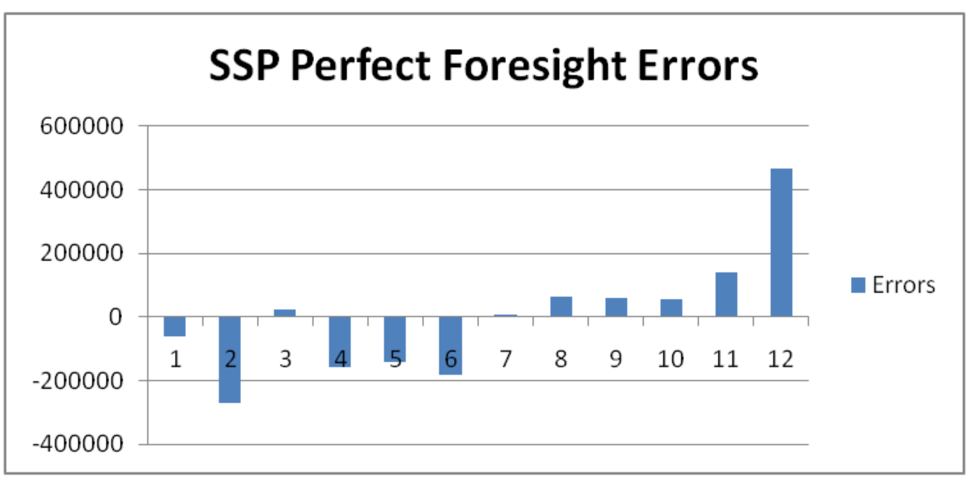
Errors of Forecasts
For 1993
Using SSP and "Last-Year's Change" Rule



As it turns out, we "under-forecasted" each of the 1993 month's sales tax revenues. This is due, of course, to the fact that we under-forecasted the total for 1993. We forecasted it to be 19,838,543 using the "last-period's change" rule, while the actual year total turned out to be 22,512,213. The 1993 tax revenues for Plano turned out to be much stronger that we anticipated.

If, in fact, we had forecasted the next year's total perfectly (i.e. had perfect foresight) our errors instead would have been much smaller as depicted in the graph below:

Errors of Forecasts
1993
Using SSP but Having Perfect Foresight on Next Year's Total



To gauge the accuracy of our forecasts we need to develop some forecasting accuracy measures. Here are some measures that are frequently used by forecasters. Here $e_i = y_i - \hat{y}_i$ represents the error of the forecast in the i-th instance with m being the number of forecasts made.

Mean Error (ME) = $(\sum_{i=1}^{m} e_i)/m$ (9)

Mean Absolute Error (MAE) = $(\sum_{i=1}^{m} |e_i|)/m$ (10)

Mean Absolute Percentage Error (MAPE) = $100 \cdot (\sum_{i=1}^{m} (|e_i|/y_i))/m$ (11)

Mean Squared Error (MSE) = $(\sum_{i=1}^{m} e_i^2)/m$ (12)

Root Mean Squared Error (RMSE) = $\sqrt{(\sum_{i=1}^{m} e_i^2)/m}$ (13)

If the Mean Error is negative, we have, on average, "over-forecasted" our target variable. On the other hand, if the Mean Error is positive, we have, on average "under-forecasted" our target variable. This tells us if we are biased upward (ME < 0) or biased downward (ME > 0). The other forecasting accuracy measures (MAE, MAPE, MSE, and RMSE) tell us how accurate our forecasts are. **The smaller the values of these measures the better the accuracy of our forecasts.** MAPE and RMSE are frequently used accuracy measures in forecasting error comparisons. MAPE measures the average percentage error of our forecasts while RMSE provides us with an accuracy measure that can easily be compared to the scale of the variable we are forecasting, y.

For example, in our attempt to forecast Plano's sales tax revenues for 1993 using 1991 – 1992 data and the SSP approach we get the following forecasting accuracy measures:

SSP

ME = 222,805.8
MAE = 222,805.8
MAPE = 11.50012
RMSE = 285, 086.3

Compare these accuracy measures with the results that we would have gotten had we had had perfect foresight in our prediction of the next year's (1993) total:

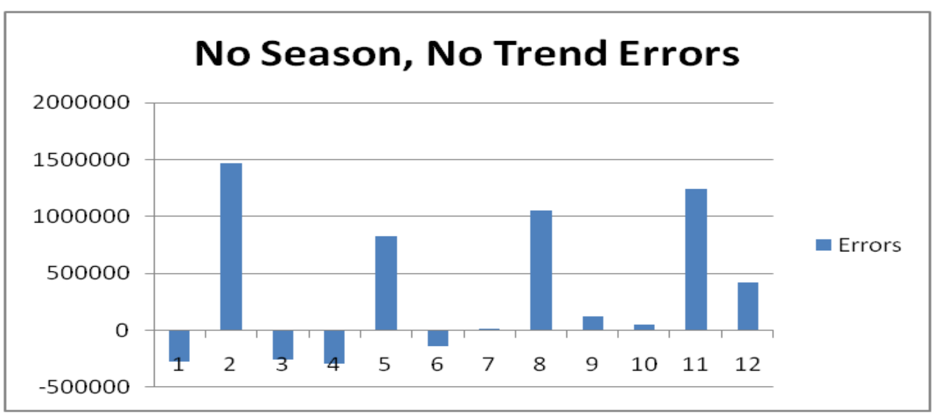
SSP
With Perfect-Foresight 1993 Year Total

ME = 7.76E-11 (essentially zero)
MAE = 135,842.3
MAPE = 7.247023
RMSE = 183,556.2

Therefore, if we had had perfect foresight in predicting the 1993 total, we would have gone from an 11.50% error to a 7.247% error. This goes to show you how important an accurate next year's total is for producing accurate forecasts using the SSP approach.

Unfortunately, with only a few years of observations on yearly totals, our forecasts of the next year's total can be, in many instances, quite inaccurate. Thus, when using the SSP approach on short time series, **it most certainly benefits the forecaster to consult an expert who has domain-specific knowledge about the forecasting problem at hand**. In this specific example the Plano City Manager and his/her staff could be very useful sources of such domain-specific information. Suppose that the City Manager told you at the end of 1992 that anecdotal information from Plano merchants was indicating that the tax revenues for the coming year (1993) was going to be stronger than the previous two years and that instead of growing at a 8.5% rate implied by the "last-period's change" rule, the rate of growth would more likely be 10%. Obviously, we have to be careful in accepting subjective forecasts like this as they may be subject to overly optimistic or overly pessimistic views coming from the "street." One pragmatic approach you might take is to average your statistical prediction for next year's total, 19,838,543 with a total of 18,283,868*(1.10) = 20,112,254.8 getting an **"add-factored" predicted total** of (19,838,543 + 20,112,254.8)/2 = 19,975,398.9. In this instance, using the add-factored predicted total would have improved the accuracy of the predictions of 1993 monthly tax revenues but, unfortunately, not as much as in the "perfect foresight" case described above.

Another useful comment is the following. **In forecasting problems, failing to recognize the salient features of the time series we are working with (trend, seasonality, and cycle) can be very costly in terms of reduced forecasting accuracy**. We will see this phenomenon exhibited over and over again in this course. Suppose that we are totally naïve and assume (naively) that we have **no seasonality** in our data (thus using the monthly proportion of 0.08333 for all months) and that our data has **no trend** in it. In this case we use as our next year's predicted total, last year's total value. In other words, let us assume that we predict next year's total by using the rule $\hat{y}_3^T = y_2^T$, "last-period's value" rule. In this case, our forecasting performance would have been abysmal. See the below graph of the errors of forecast.



In particular, the costs arising from ignoring trend and season in the SSP method are measured by the following forecasting accuracy measures:

ME = 352,362.1
MAE = 512,213.4
MAPE = 23.22189
RMSE = 701,755.3

Instead of having a 11.5% Mean Absolute Percentage Error (MAPE) in our forecasts as first proposed, we have a 23.22 % Mean Absolute Percentage Error (MAPE) arising from this naïvete with respect to seasonality and trend. This is a more than doubling of the percentage error in forecasting the sales tax revenues for 1993! Obviously, in terms of producing accurate forecasts, we need to be especially sensitive to whether we have trend and/or seasonality our data and, if they are present, make allowance for them in our forecasting procedures. We will see the same applies to detecting cycles in our data, but that is a topic for later discussion as the SSP approach does explicitly address the possibility of cyclicity in our data.

III. Predicting Next Year's Total Using the Least Squares Approach

We can generalize the prediction rule (5) – (6) when we have more than 2 year of data. Assume we have n > 2 yearly totals, namely, $y_i^T, i = 1,2,\cdots,n$. To predict next year's total we could use linear regression to do the job. Consider the following hypothetical annual data along with a time series plot of it and the "fitted" values coming from a least squares fit of the 5 yearly totals. The hypothetical Year Totals are as follows:

Year	Totals
1	10.65
2	23.77
3	33.90
4	35.33
5	38.50

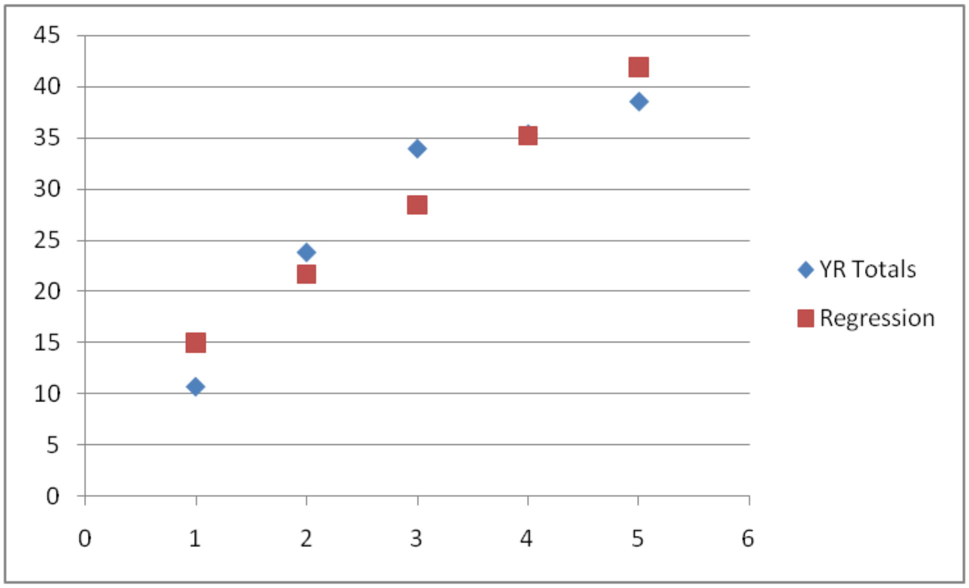
Using the "regression" facility in the Data Analysis Pack of EXCEL we obtain the following least squares fitted line of the data. Its formula is given by

$$\hat{y}_i^T = 8.252 + 6.726 * i \quad , \quad i = 1,2,\cdots,n \, . \qquad (14)$$

The below fitted values of the line (Red Squares) are obtained by plugging successively the values of 1, 2, 3, 4, and 5 for i into the above formula (14). The intercept of the regression line is 8.252 and its slope is 6.726. It follows quite straightforwardly that the next year's predicted total can be obtained by setting i = 6 and using (14) to predict the 6-th years' total as in

$$\hat{y}_6^T = 8.252 + 6.726 * (6) = 45.908 \ .$$

Plot of 5 Year Totals (Blue Diamonds) and
5 Corresponding Points on a Least Squares Line (Red Squares)



In general, the least squares regression line of yearly totals can be shown to be given by the formula

$$\hat{y}_i^T = a + b * i \tag{15}$$

where

$$b = \frac{\sum_{i=1}^{n} i y_i^T - \frac{(n+1)}{2}\sum_{i=1}^{n} y_i^T}{\frac{(n+1)n(n-1)}{12}} \tag{16}$$

and

$$a = \bar{y}^T - b\bar{\imath} = \bar{y}^T - b\left(\frac{n+1}{2}\right) \tag{17}$$

where $\bar{y}^T = \sum_{i=1}^{n} y_i^T / n$ is the mean of the yearly totals and $\bar{\imath} = (n+1)/2$ is the mean of the year indices (i) running from 1 to n. Following this approach, the least squares' prediction of next year's ((n+1) year) total is given by

$$\hat{y}_{n+1}^T = a + b(n+1) \quad . \tag{18}$$

Finally, the prediction equations (5) and (6) can be generalized as using the next year's total being predicting by (18) and the monthly predictions for next year be generated by the formula

$$\hat{y}_{n+1,j} = \hat{y}_{n+1}^T \cdot \bar{P}_j \quad , j = 1,2,\cdots,12 \tag{19}$$

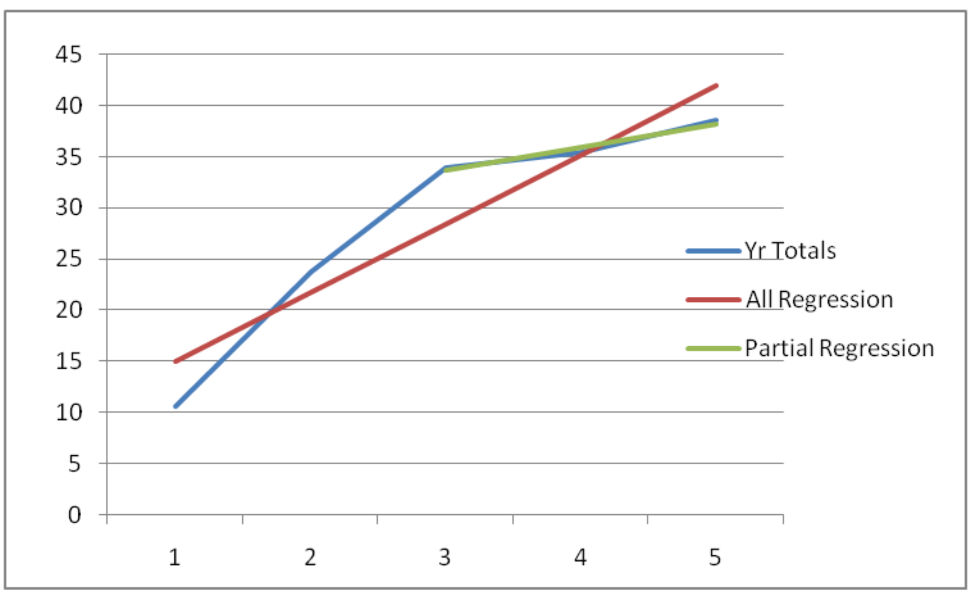
where $\bar{P}_j$ is determined as in equation (3) above. Note that in the case when only two yearly total observations are available to us ($n = 2$), "last year's change" rule and the prediction formula (5) for next year's total is equivalent to the least squares prediction formula (15).

One must be careful, however, in blindly using the formulae (18) and (19) to forecast next year's total. For example, take a closer look at the above data. Starting at observation i=3, it appears that the rate of increase in the 4-th and 5-th years has dropped off. If we feel that this "slowdown in growth" (i.e. change in the slope of the regression) is permanent, maybe we should only use the last three observations in determining a least squares line to predict the next year's total. If we do this we get the least squares line

$$\hat{y}_i^T = 26.71 + 2.3 * i \quad , \; i = 1,2,\cdots,n \tag{20}$$

This "partial" regression line predicts next year's total to be 26.71 + 2.3*(6) = 40.51 which is certainly less than the 45.908 figure obtained by using all of the yearly totals to fit the regression line. See the below graph that plots the two alternative regression lines (Red Squares versus Green Triangles). **Again, domain-specific knowledge is important here.** If we know from experts in the field that growth has permanently slowed, we should use the latter estimate based on (20). On the other hand, if experts feel that growth will soon return to its old trend then we should use the formula (14) which uses all of the yearly totals data.

Two Different Possibilities of Trend
(Red Line = growth as usual, Green Line = growth slowdown)
(Blue Line = original yearly totals)



IV. Conclusion

The Stable Seasonal Pattern (SSP) model is easy to understand and easy to apply. As we will see, when we compare this method with two other popular "simple" methods, namely exponential smoothing (ES) and the deterministic trend/seasonal (DTDS) model, the SSP model holds up pretty well in terms of performance.

It does have its drawbacks, however. The SSP model's accuracy in making monthly predictions is directly affected by the accuracy of the predicted yearly totals. If the predicted yearly totals are inaccurate, then most certainly the SSP model's forecasts will be inaccurate. Also the SSP model will most certainly perform poorly in the presence of seasonal patterns that are not stable over time. The SSP model has two additional drawbacks. First, even though the point forecasts coming from a SSP model may be accurate, due to the non-linear nature of the forecasting formulae, confidence intervals for the forecasts are not readily available. Second, the SSP model does not have an explicit way of taking into account cycles that might be present in the data as might be implied by the Plano tax revenues being affected by macroeconomic forces occurring outside of Plano's local economy. Fortunately, these last two issues can be addressed by more sophisticated approaches as in the Box-Jenkins (BJ) and Unobservable Components (UC) models that we will subsequently study.

V. Appendix: Applying Friedman's "Two-Analysis of Variance by Ranks Test" to Test for Seasonality in Time Series Data

As mentioned before, it is pretty important for a forecaster to determine the salient characteristics of a time series so that proper modeling choices can be made to ensure better forecasting results. In this Appendix we are going to discuss a non-parametric statistical test for seasonality in time series data. The test is an adaptation of Friedman's two-way analysis of variance by ranks test. See, for example, Daniel (1978). The major assumption of the test is that the data consist of b mutually independent samples (blocks) of size k, k representing the number of treatments brought to bear on each block. In our case the "blocks" are represented by years $(i = 1,2,\cdots,n)$ and the "treatments" by months $(j = 1,2,\cdots,12)$.

In our case, consider the following two-way table of monthly proportions:

Months

Year	1	2	3	….	10	11	12
1	P_{11}	P_{12}	P_{13}	….	$P_{1,10}$	$P_{1,11}$	$P_{1,12}$
2	P_{21}	P_{22}	P_{23}	….	$P_{2,10}$	$P_{2,11}$	$P_{2,12}$
3	P_{31}	P_{32}	P_{33}	….	$P_{3,10}$	$P_{3,11}$	$P_{3,12}$
.	.	.	.	….	.	.	.
.	.	.	.	….	.	.	.
.	.	.	.	….	.	.	.
n	P_{n1}	P_{n2}	P_{n3}	….	$P_{n,10}$	$P_{n,11}$	$P_{n,12}$

Here P_{ij} $(i = 1,2,\cdots,n; j = 1,2,\cdots,12)$ represents the proportions of a given year's sales of a company in the j-th month of the i-th year. The null and alternative hypotheses of the Friedman test in this context are, respectively,

H_0: The monthly proportions of yearly totals are the same for all years (i.e., $P_{ij} = \frac{1}{12}$ for all $i = 1,2,\cdots,n; j = 1,2,\cdots,12$) and therefore there is **no seasonality in the data.**

H_1: **There is seasonality in the data** in that there are at least two monthly proportions that are not equal to each other.

The test is non-parametric in the sense that the test statistic to be developed below is based on the distribution of the ranks of the proportions of the months within each year. The first step in constructing the test is to convert the row proportions into ranks, the smallest proportion being assigned a rank of 12 and the largest proportion being assigned a rank of 1. (Without any effect on the outcome of the test, the ranks could be constructed in ascending order instead of descending order with the smallest proportion being assigned the rank of 1 and the largest proportion being assigned a rank of 12.) For example, the ranking of the proportions in the 1991 – 1992 Plano Sales Tax Revenue data are

Ranks of Proportions
By Month Within Each Year (1991 and 1992)
In the Plano Sales Tax Revenue Data

Months

Year	1	2	3	4	5	6	7	8	9	10	11	12
1	11	1	12	9	2	6	7	4	8	5	3	10
2	11	1	12	10	4	6	8	3	5	9	2	7
R_j	22	2	24	19	6	12	15	7	13	14	5	17

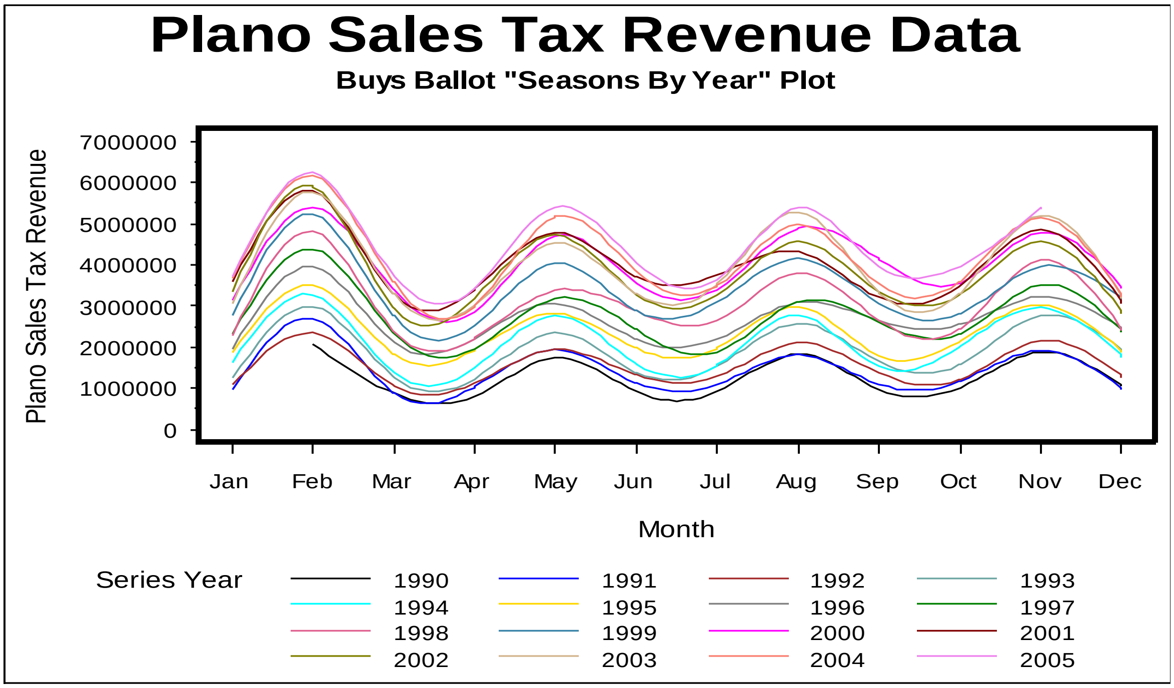
Now if H_0 is true and all months have identical proportions, the rank that appears in a particular column when the data is displayed as in the above is merely a matter of chance and therefore neither small or large ranks should tend to show a "preference" for particular columns (months). The test statistic to test for the absence (H_0) or presence (H_1) of a preference for particular months is constructed as follows:

$$\times_r^2 = \frac{1}{13n}\left(\sum_{j=1}^{12} R_j^2\right) - 39n \qquad (21)$$

where R_j is the sum of the ranks for the j-th column (month) and n is the number of years for which we have monthly data. Under the assumed truth of the null hypothesis, the above statistic has a chi-square distribution with r = 11 degrees of freedom. Therefore, we reject H_0 if the observed $\times_r^2$ is greater than or equal to the critical value of $\times_{11,(1-\alpha)}^2$, otherwise we accept H_0. At the one percent level this critical value is $\times_{11,(0.99)}^2 = 24.725$, at the 5 percent level it is $\times_{11,(0.95)}^2 = 19.675$, and at the 10 percent level it is $\times_{11,(0.90)}^2 = 17.275$. Of course, the probability value of the observed $\times_r^2$ can be used for the same purpose. If the observed probability value is less than the chosen level of significance, we reject H_0, otherwise, we accept H_0.

For example, using the 1991 – 1992 Plano Sales Tax Revenue data and the ranks above, the observed chi-squared value is $\times_r^2$ = 20.38562 with a probability value of p = 0.0403. Therefore, we reject the null hypothesis that that there is no seasonality in the Plano data and accept the alternative hypothesis that there is seasonality in the data at the 5% level. It should be noted that the same above yearly ranks could have been calculated from the raw data (sales) since the proportions within each year, say i, share a common denominator,y_i^T. The above tabular presentation used proportions P_{ij} because the independence of the blocks (years) in the Friedman test is more likely to hold in the proportions P_{ij} than in the observations y_{ij} (especially if they have trend in them). But, in practice, ranking the originally observations in lieu of the proportions is fine.

In addition to the above formal statistical test for seasonality, one can use Buys-Ballot plots to visually diagnosis the presence or absence of seasonality. Look at the file "Buys Ballot Plots.doc" for a discussion of Buys-Ballot plots. Also see the SAS program "Plano_Plot_Jacob_Williamson.sas" that produces the following "Season-By-Year" Buys-Ballot Plot for the entire Plano data set. Notice the parallelism of the seasons (months) by year. Across all years, the months February, May, August, and November are always "strong" months within the year.



Reference:

Daniel, Wayne W. (1978), Applied Nonparametric Statistics (Houghton Mifflin Co., Boston), Section 7.1 “The Friedman two-way analysis of variance by ranks”, pp. 224 – 228.