Forecasting with SAS

Linda Lieu

Advanced Analytics Intern, TransUnion Canada

Golden Horseshoe SAS User Group (GHSUG)

Burlington, Ontario, Canada

May 2015

Time Series

- **A sequence of observations over a time period.**

- **Classical Characteristics**
  - ➢ **Trend**
  - ➢ **Seasonality**
  - ➢ **Cycles**
  - ➢ **Residuals**

- **Applications**
  - ➢ **Economics**
  - ➢ **Finance**
  - ➢ **Medicine**
  - ➢ **Science**

Agenda

- **Time Series (observe the trend)**

- **Features of Proc Forecast**

- **Methods (determine the best method to forecast the data)**

- **Breakdown of the Code**

- **Applications**

- **Example: Healthcare Labour Force in Ontario from 2009 - 2013**

- **Quirks**

- **Advanced Forecasting Methods**

Features of Proc Forecast

- Fast and automatic forecasts for many series in one step as long as the series is organized by **variables (VAR)** or **groups (BY)**.

- Does not require a thorough statistical analysis of the time series.

- Forecast a series based on the historical data (one variable), and does not consider other variables that may affect the time series in question.

> The BY statement is often used for cross-sectional studies.

```
proc forecast /* Options */ ;
        by /* Forecasting the data set in separate groups.
                 For Example, Female vs Male's Internet Consumptions over the past 12 months */;
        id /* Identify the time interval variable */;
        var /* Select the variable set to forecast */;
run;
```

General Options

data=*<dataset set to forecast>*
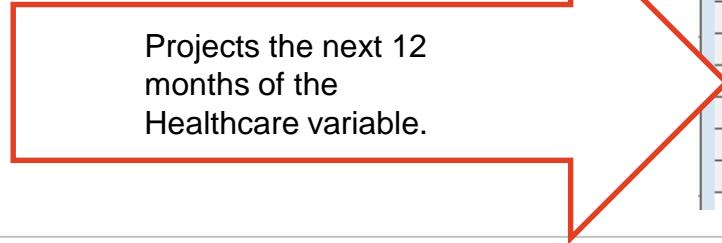**lead**=*<desired number of forecast values>*
interval=*<frequency of the time series (useful if you want to forecast by year instead of by month)>*
Out=*<outputs the parameter and forecast estimates>*
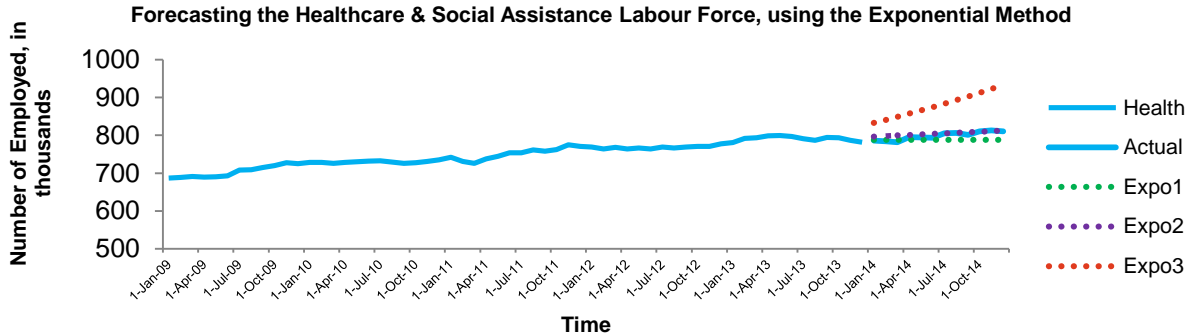**Outest**=*<outputs the statistical measure>*

Projects the next 12 months of the Healthcare variable.

VIEWTABLE: Health.Forecast_14_expo3

	Time	Type of Observation	Number of Periods into the Forecast	Health care and social assistance (62)
1	01JAN2014	FORECAST	1	833.17206726
2	01FEB2014	FORECAST	2	840.84700985
3	01MAR2014	FORECAST	3	848.80076221
4	01APR2014	FORECAST	4	857.03332436
5	01MAY2014	FORECAST	5	865.54469628
6	01JUN2014	FORECAST	6	874.33487797
7	01JUL2014	FORECAST	7	883.40386944
8	01AUG2014	FORECAST	8	892.75167069
9	01SEP2014	FORECAST	9	902.37828171
10	01OCT2014	FORECAST	10	912.28370251
11	01NOV2014	FORECAST	11	922.46793309
12	01DEC2014	FORECAST	12	932.93097344

TransUnion

Case Scenario: The Ontario Healthcare and Social Assistance Labour Force[1]

- **Brief Summary**

- **What's the purpose of selecting this time series modeling?**

- **Who would benefit from forecasting this time series?**

- **Source of the Data & Conditions**

Forecasting the Healthcare & Social Assistance Labour Force, using the Exponential Method

Y-axis: Number of Employed, in thousands (500–1000)
X-axis: Time (1-Jan-09 to 1-Oct-14)

Legend:
- Health
- Actual
- Expo1
- Expo2
- Expo3

Trial and errors based on method and trend.

TransUnion

Smoothing Exponential Methods

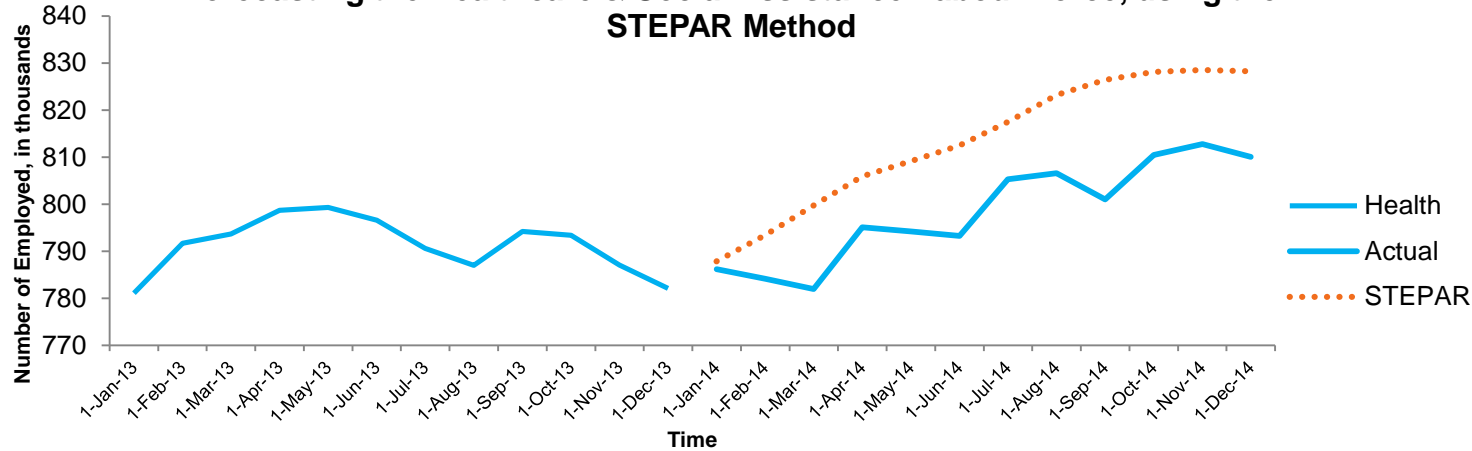|	SINGLE	DOUBLE	TRIPLE
Requirement	• The mean of the series decreases over time • Constant model	• Must be approximately linear trend	• Must be at least a quadratic model

Code	`proc forecast data=health.five lead=12`
	`    interval=month out=health.forecast_expo1`
	`    method=expo trend=1;`
	`    id time;`
	`    var &v1; run;`
		`proc forecast data=health.five lead=12`
		`    interval=month out=health.forecast_expo1`
		`    method=expo trend=2;`
		`    id time;`
		`    var &v1; run;`
			`proc forecast data=health.five lead=12`
			`    interval=month out=health.forecast_expo1`
			`    method=expo trend=3;`
			`    id time;`
			`    var &v1; run;`

Options	• Smaller smoothing weights are recommended for series with stable trends whereas larger weights should be used for unstable series.		

Forecasting the Healthcare & Social Assistance Labour Force, using the Exponential Method

Actual Value		Forecasted Values			Error Percentage		
Time	2014	Expo1	Expo2	Expo3	Expo1	Expo2	Expo3
1-Jan-14	786.2	788.224	796.82	833.172	0.26%	1.35%	5.97%
1-Feb-14	784.2	788.224	798.229	840.847	0.51%	1.79%	7.22%
1-Mar-14	782	788.224	799.638	848.801	0.80%	2.26%	8.54%
1-Apr-14	795.1	788.224	801.047	857.033	-0.86%	0.75%	7.79%
1-May-14	794.2	788.224	802.456	865.545	-0.75%	1.04%	8.98%
1-Jun-14	793.3	788.224	803.865	874.335	-0.64%	1.33%	10.21%
1-Jul-14	805.3	788.224	805.275	883.404	-2.12%	0.00%	9.70%
1-Aug-14	806.6	788.224	806.684	892.752	-2.28%	0.01%	10.68%
1-Sep-14	801	788.224	808.093	902.378	-1.60%	0.89%	12.66%
1-Oct-14	810.5	788.224	809.502	912.284	-2.75%	-0.12%	12.56%
1-Nov-14	812.8	788.224	810.911	922.468	-3.02%	-0.23%	13.49%
1-Dec-14	810.1	788.224	812.32	932.931	-2.70%	0.27%	15.16%

- **Recent data are more heavily factored in than the early observations**

Forecasting the Healthcare & Social Assistance Labour Force, using the STEPAR Method

Actual Value		FV	Error
Time	**2014**	**STEPAR**	**STEPAR**
1-Jan-14	786.2	787.862	0.21%
1-Feb-14	784.2	793.434	1.18%
1-Mar-14	782	799.79	2.27%
1-Apr-14	795.1	805.991	1.37%
1-May-14	794.2	809.124	1.88%
1-Jun-14	793.3	812.484	2.42%
1-Jul-14	805.3	817.5	1.51%
1-Aug-14	806.6	823.224	2.06%
1-Sep-14	801	826.446	3.18%
1-Oct-14	810.5	828.112	2.17%
1-Nov-14	812.8	828.562	1.94%
1-Dec-14	810.1	828.273	2.24%

STEPAR (Stepwise Autoregressive)

- **Automatic forecasting method that fit the series with equal weights.**
- **Combine models that are suitable for considering long-term trend *(trend models)* and short-term flux *(autoregressive)***
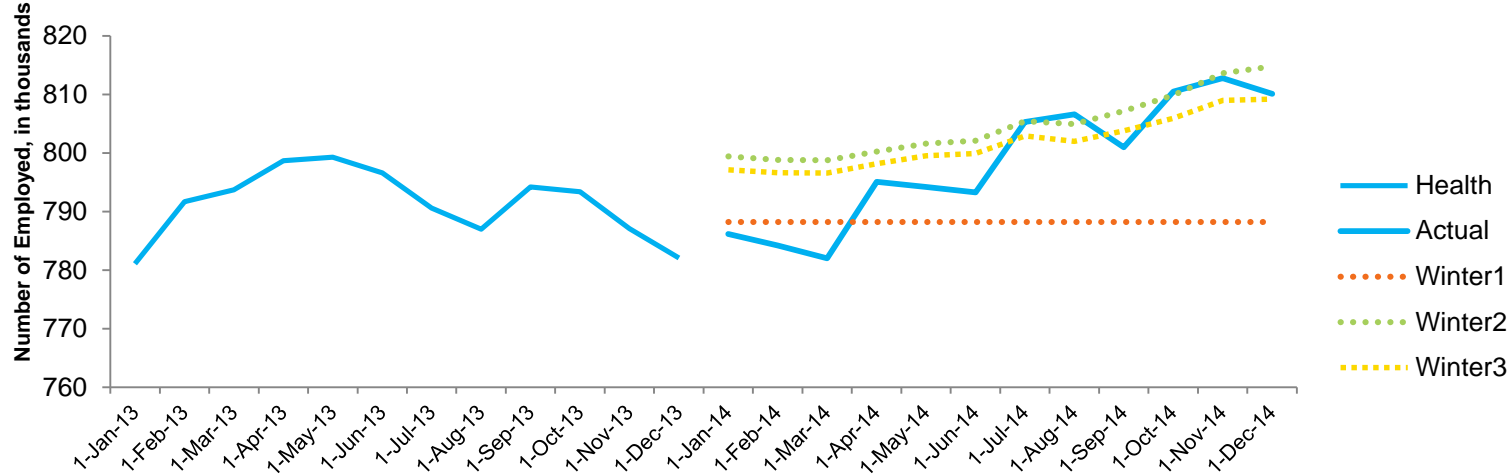
WINTERS Methods

	CONSTANT	LINEAR	QUADRATIC
Description	• Presume the values of the time series' variable are positive. • Seasonal parameters are multiplied with the trend.		
Features	• Predicts a constant forecast value for all future time intervals.	• If *trend=n* is not specified in the code, this is usually the default option.	• Unstable and not advised to use.
Code	proc forecast data=health.five lead=12 interval=month out=health.forecast_winter1 method=winters trend=1; id time; var &v1; run;	proc forecast data=health.five lead=12 interval=month out=health.forecast_winter2 method=winters trend=2; id time; var &v1; run;	proc forecast data=health.five lead=12 interval=month out=health.forecast_winter3 method=winters trend=3; id time; var &v1; run;
Options	• The consideration of seasonality makes use of the INTERVAL and SEASONS options		

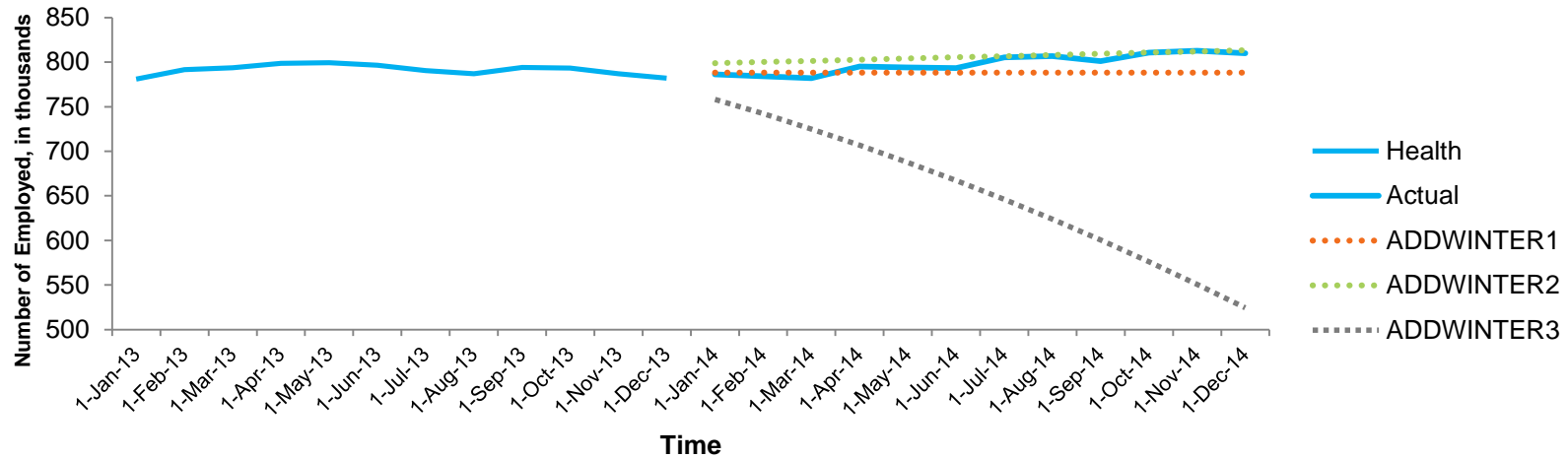# Forecasting the Healthcare & Social Assistance Labour Force, using the Winter Method

Actual Value		Forecasted Values			Error Percentage		
Time	2014	Winter1	Winter2	Winter3	Winter1	Winter2	Winter3
1-Jan-14	786.2	788.224	799.421	797.112	0.26%	1.68%	1.39%
1-Feb-14	784.2	788.224	798.853	796.633	0.51%	1.87%	1.59%
1-Mar-14	782	788.224	798.762	796.613	0.80%	2.14%	1.87%
1-Apr-14	795.1	788.224	800.274	798.179	-0.86%	0.65%	0.39%
1-May-14	794.2	788.224	801.63	799.547	-0.75%	0.94%	0.67%
1-Jun-14	793.3	788.224	802.128	799.948	-0.64%	1.11%	0.84%
1-Jul-14	805.3	788.224	805.424	802.901	-2.12%	0.02%	-0.30%
1-Aug-14	806.6	788.224	804.931	801.992	-2.28%	-0.21%	-0.57%
1-Sep-14	801	788.224	807.177	803.807	-1.60%	0.77%	0.35%
1-Oct-14	810.5	788.224	809.884	805.947	-2.75%	-0.08%	-0.56%
1-Nov-14	812.8	788.224	813.664	808.968	-3.02%	0.11%	-0.47%
1-Dec-14	810.1	788.224	814.761	809.233	-2.70%	0.58%	-0.11%

Winters
- **An exponentially smoothed trend, which factors in seasonality**

Forecasting the Healthcare & Social Assistance Labour Force, using the ADDWINTER Method

Actual Value		Forecasted Values			Error Percentage		
Time	**2014**	**AW1**	**AW2**	**AW3**	**AW1**	**AW2**	**AW3**
1-Jan-14	786.2	788.224	798.769	758.256	0.26%	1.60%	-3.55%
1-Feb-14	784.2	788.224	800.093	742.123	0.51%	2.03%	-5.37%
1-Mar-14	782	788.224	801.417	724.964	0.80%	2.48%	-7.29%
1-Apr-14	795.1	788.224	802.742	706.778	-0.86%	0.96%	-11.11%
1-May-14	794.2	788.224	804.066	687.565	-0.75%	1.24%	-13.43%
1-Jun-14	793.3	788.224	805.39	667.326	-0.64%	1.52%	-15.88%
1-Jul-14	805.3	788.224	806.714	646.06	-2.12%	0.18%	-19.77%
1-Aug-14	806.6	788.224	808.039	623.768	-2.28%	0.18%	-22.67%
1-Sep-14	801	788.224	809.363	600.449	-1.60%	1.04%	-25.04%
1-Oct-14	810.5	788.224	810.687	576.104	-2.75%	0.02%	-28.92%
1-Nov-14	812.8	788.224	812.011	550.732	-3.02%	-0.10%	-32.24%
1-Dec-14	810.1	788.224	813.336	524.333	-2.70%	0.40%	-35.28%

Additive Winters
- **Similar to the Winters Method, except the seasonal parameters are added to the trend**

Quirks

- **The interval variable must be formatted into a SAS date**.

- **Since we are forecasting with a time series, make sure the observations are sorted by time from past to present**.

- **Always try a variety conditions such as the time period, smoothing weight, methods, etc on a test run since every time series have different characteristics**.

Univariate Forecasting Models

	PROC FORECAST	PROC ARIMA (Auto Regression Integrated Moving Average)
Features	• Automatic trend extrapolation. • No need for complex statistical analysis.	• Outputs detail analysis of a series. • Selects the best fit with the least mean squared forecast residuals. • Generates forecast values as a linear combination of the series' historical values and errors.
Best For	• Forecasting numerous of time series at once.	• Forecasting one series via the identification, estimation, and forecast stages.

Multivariate Forecasting Models

	PROC VARMAX	PROC STATESPACE
Features	• Predict forecasts from the regression model. • Provide tests to find constant error terms, etc.	• Forecast numerous of relevant variables based on the autocorrelations.
General	• Forecasts the response variable based on the influence of the predictor variables $(X_1, X_2, X_3, .., X_n)$. • Able to project forecasts, based on scenarios (Example, oil price falls, the forecast of GDP is expect to fall in Venezuela).	

TransUnion

Contact

Linda Lieu

Advanced Analytics Intern, **TransUnion Canada**, January 2015 - August 2015

3rd Year Student, Actuarial and Financial Mathematics, **McMaster University**

Expected Graduation: April 2017

llieu@transunion.com

lieult@mcmaster.ca

A special thank you to Peter Loparco and Justin Jia for their feedbacks and guidance.

Appendix: Ontario Healthcare Statistics

- By 2035, the senior population will consist of 23.8% of the Ontario's population (grow to 4.1 millions).

- More people are beginning their career past the age of 18 due to post-secondary education.

- Health care professionals tend to be older.

- Factors that affect skills shortage: age structure of the labour sector, training length, mobility, and workplace conditions (long hours, inconsistent shifts, and difficult patients).

TransUnion

References

Derby, Nate. "Univariate Forecasting Methods." *Time-Series Forecasting*(2000): n. pag. *Calgary SAS Users Group.*
11 Dec. 2009. Web. 28 May 2015.
http://www.sas.com/content/dam/SAS/en_ca/User%20Group%20Presentations/Calgary-User-Group/Derby-
TimeSeriesForecasting-Nov2009.pdf

SAS Institute Inc. 2013. SAS/ETS® 12.3 User's Guide. Cary, NC: SAS Institute Inc.
http://support.sas.com/documentation/cdl/en/etsug/66100/PDF/default/etsug.pdf

Weber, Richard. "Time-Series." *SpringerReference* (2011): n. pag. *Time Series.* Web. 28 May 2015.
http://www.statslab.cam.ac.uk/~rrw1/timeseries/t.pdf

TransUnion