



Forecasting made easy with SPSS Statistics

Jarlath Quinn

www.sv-europe.com

FAQ's

- Is this session being recorded? Yes
- Can I get a copy of the slides? Yes, we'll email a PDF copy to you after the session has ended.
- Can we arrange a re-run for colleagues? Yes, just ask us.
- How can I ask questions? All lines are muted so please use the chat facility – if we run out of time we will follow up with you.









- Premier accredited partner to IBM and Predictive
 Solutions specialising in advanced analytics & big
 data technologies
- Work with open source technologies (R, Python, Spark etc.)
- Team each has 15 to 30 years of experience working in the advanced and predictive analytics industry

- Deep experience of applied advanced analytics applications across sectors
 - Retail
 - Gaming
 - Utilities
 - Insurance
 - Telecommunications
 - Media
 - FMCG





 If you can see sub-menu, then it is installed on your copy of SPSS Statistics



Edit View Data Transform Analyze Graphs Utilities Extensions Window Help File Reports 0 Descriptive Statistics **Bayesian Statistics** Ta<u>b</u>les var var var var var va Compare Means 1 General Linear Model 2 Generalized Linear Models Mixed Models 3 Correlate 4 Regression 5 Loglinear Neural Networks 6 Classify 7 Dimension Reduction Scale 8 Nonparametric Tests 9 Forecasting . 🔀 Create Temporal Causal Models... 10 Survival Create Traditional Models... Multiple Response 🕅 Apply Temporal Causal Models... 11 💋 Missing Value Analysis... Apply Traditional Models... 12 Multiple Imputation ъ Seasonal Decomposition... 13 Complex Samples 🕎 Spectral Analysis... Bimulation... 14 🚰 Sequence Charts... Quality Control 15 Autocorrelations... Spatial and Temporal Modeling. E Cross-Correlations... Direct Marketing 16 ъ 17

🔚 Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor

Agenda

- The principles of Time Series forecasting
- Visualising time series
- Smoothing techniques
- Exponential smoothing methods
- Interpreting output and model fit
- Using predictor fields to improve accuracy
- Generating forecasts







The Principles of Time Series Forecasting

www.sv-europe.com

What is Time Series?

- A 'Time Series' is simply a series of values of a quantity collected over a specific time period, often with equal intervals between them
- Examples of time series include:
 - Airline passenger numbers for a particular country over the last 40 years
 - Daily website hits during a three-month period
 - Hourly traffic volumes over the course of a week



Time Series Forecasting

"'Things' that are observed repeatedly over time, with past values and

other factors being used to predict future values"



What is Time Series?

- Time series analysis is based on the principle that the past provides a model for the future
- Time series forecasting models often don't require predictor/independent variables
- The goal of time series analysis is to separate the random variability ('noise') from the variability that can be explained
- A single time series may have several elements that enable effective forecasting



What's in a 'Time Series'?

Time Series







Considerations

- Are the data points regularly spaced?
- How far can I forecast into the future?
- What is the periodicity?
- Should I exclude data?
- Do I have other predictor fields?
- Are there special events that need to be marked?
- Do I need to forecast more than one series?







www.sv-europe.com

- Using the dataset Ticket Sales Missing.sav
- We can visualise the data series with **Sequence Charts**

ta Sequence Charts		×
Qne chart per variable	Variables: Tickets Time <u>A</u> xis Labels: Transform Natural log transform Difference: 1 Seasonally difference: 1 Current Periodicity: None	<u>Time Lines</u> <u>F</u> ormat
ОК	Paste Reset Cancel Help	

	🛷 Tickets	var	V
1			
2	15397		
3	18270		
4	20384		
5	18993		
6	15059		
7	14125		
8			
9	19760		
10	19435		
11	22703		
12	17771		
13	13091		
14	17039		
15	20224		
16	19862		
17	20723		
18	17754		
19	12072		
20	15986		
21	17767		
22	19823		
	10452		



- A few things to note:
 - You don't need a variable for the time axis
 - There are gaps in the series i.e. missing data





- One way to deal with missing values in a sequence
- Using the method 'Mean of nearby points'

Europe







Assigning Periodicity

www.sv-europe.com

Assigning Periodicity

- SPSS Statistics can only see a sequence of numbers
- It doesn't know what separates each row of data
- They could be separated by minutes, weeks or years
- To address this, we explicitly assign periodicity to the dataset

Define Dates	×
Cases Are:	First Case Is: Periodicity at higher level Year: 2008 Month: 12
Year(?)Month(?;12)	Cancel Help



Assigning Periodicity

 SPSS Statistics now creates a series of fields that it can use in time series analysis to identify the periodic (and seasonal) separation between the sequence values

✓ YEAR_	✓ MONTH_	💑 DATE_
2008	1	JAN 2008
2008	2	FEB 2008
2000	2	MAD 2000









www.sv-europe.com

- Creating a sequence plot of the file **Batteries sales.sav**
- Note the periodicity is already defined





- We can use SPSS to create a **smoothed** version of this series
- Smoothing is usually done to reveal a clearer picture of the series by simplifying it and removing some of the randomness.
- Smoothing a series is similar to the methods that pure Time Series techniques employ to create a model for forecasting
- The are *many different* ways that smoothing techniques can be applied.



Moving Average Smoothing Example

- One of the most common forms of smoothing is using a moving average
- The graphic shows how a **moving average** is calculated with a span of 3 cases





• The **Create Time Series** procedure allows us to create smoothed series



		V <u>a</u> riable-> New name
Batteries YEAR, not periodic [MONTH, period 12 [*	Battery_MA3=MA(Batteries 3)
		Name: Battery_MA3 Change
		Eunction:
		Centered moving average
		Order: 1 Span: 3
		Current Periodicity: 12



Europe

• The sequence chart clearly shows that the smoothed series moving average is a less noisy, albeit simplified version of the Batteries sales variable







Exponential Smoothing Methods

www.sv-europe.com

Exponential Smoothing

- Using the simple moving average smoothing techniques that we saw earlier, each time point in the calculation has equal weight. For example, by using a span of say 5, the cases that are two time points away are treated as equally as those that are one time point away when the moving average is calculated.
- In *Exponential* smoothing, values that are *closer* (in time) are *given* greater weight than those that are further away.
- This approach can be employed in forecasting where the values that are *more recent* have a greater influence on estimating the future than those that are less recent.



Additive vs Multiplicative Series

Additive Series

Multiplicative Series



Seasonal Variation Constant



Seasonal Variation Changing

Basic Exponential Smoothing Methods

	Non Seasonal	Additive Seasonal	Multiplicative Seasonal
Constant Level		$\sim \sim \sim$	$\sim \sim \sim$
Linear Trend			PAPA
Damped Trend			AAA
Exponential Trend			NAN



SPSS Statistics Exponential Smoothing Models

- SPSS Statistics has 7 standard <u>Exponential smoothing</u> models
- The standard models are divided into:
 - 4 non-seasonal methods
 - 3 seasonal methods
- However using the Expert Modeler method means that SPSS automatically chooses a model type for the series

🝓 Time Series Modeler: Exponential Smoothing Criteria	×
Model Type Dependent Nonseasonal: Simple Holt's linear trend Brown's linear trend Damped trend Naţu Naţu Simple seasonal Winters' additive Winters' multiplicative Winters' nultiplicative Current periodicity: 12 Dependent Sigu: Non Sigu: Naţu Naţu Naţu Minters' additive Minters' multiplicative Minters' nultiplicative Minters' nultiplicative Națu Națu Minters' nultiplicative Minters' nultiplicative 	Variable Transformation — re root al log
Cancel Help	



*clothing sales simple.sav [DataSet1] - IBM SPSS Statistics Data Editor

- To create a forecast using **Expert Modeler** from the main menu click:
 - Analyze
 - Forecasting
 - Create Traditional Models

e <u>E</u> di	<u>V</u> iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	E <u>x</u> tensio	ns	<u>W</u> indow	<u>H</u> elp	
<u>)</u>				Reports Descriptive Statistics Bayesian Statistics							
		date	🥠 (, Ta <u>b</u> le	S		•		🖋 MON	NTH_	🖧 DATE_
1	01/0	01/19	89	Co <u>m</u> r Cono	oare Means) Jodel	۲ ۲			1	JAN 1989
2	02/0	01/19	89	Gene	ralized Line	ear Models	, ,			2	FEB 1989
3	03/0	01/19	89	Mi <u>x</u> ed	Models		•			3	MAR 1989
4	04/0	01/19	89	<u>C</u> orre	late		•			4	APR 1989
5	05/0	01/19	89	L <u>og</u> lir	near		•			5	MAY 1989
6	06/0)1/19	89	Neura	al Net <u>w</u> orks	3	•			6	JUN 1989
7	07/0)1/19	89	Class	si <u>f</u> y nsion Red	uction	4			7	JUL 1989
8	08/0)1/19	89	Sc <u>a</u> le			•			8	AUG 1989
9	09/0)1/19	89	Nonp	arametric 1	Fests	•				
10	10/0)1/19	89	Forec Surviv	asting /al				Create To	emporal Ca	ausal Models
11	11/0)1/19	89	M <u>u</u> ltip	le Respon	se	•		Apply Ter	mporal Cau	Isal Models
12	12/0)1/19	89	🔣 Missir	ng Value Ar	nal <u>v</u> sis			Apply Tra	ditional Mo	dels
13	01/0)1/19	90	Com	ole imputati olex Sampl	on es			🗾 <u>S</u> easona	l Decompo	sition
14	02/0	1/10	90	🐺 S <u>i</u> mul	ation				Spectral .	Analysis	
15	02/0	1/10	20	<u>Q</u> uali	ty Control		•		Autocorre	e Charts	
16	03/0	1/10	20	Spati: Direct	al and Tem t Marketing	iporal Mode	eling ►		Cross-Co	orrelations.	
17	04/0	1/19	90		21700	15	100	•			MAX 1000
17	05/0	1/19	90	•	51790	. 15	199	U		5	WAT 1990



- Specify **Sales of Women's' Clothing** as the dependent variable
- Note that the default **Method** for model fitting is **Expert Modeler**
- Based on a measure of model fit, this method will try a number of model types and select the one with the best fit over the series

🙀 Time Series Modeler	×
Variables Statistics Plots Output Filter Save Op	tions
Variables:	Dependent Variables:
Date [date]	Sales of Women's Clothing [clothing_sales]
WONTH period 12 MONTH 1	
	Independent Variables:
Method: Expert M	odeler 🔹 Criteria
Model Typ	e: All models
Estimation Period	Forecast Period
Start: First case	Start. First case after end of estimation period
End: Last case	End: Last case in active dataset
OK Paste	Reset Cancel Help



• We can request a plot to enable us to see how the model as been fitted to actual series

ta Time Series Modeler	×
Variables Statistics Plots Output Filter Save Options	
Plots for Comparing Models	
Stationary R square Maximum absolute percentage error	
R square Maximum absolute error	
Root mean sguare error Normalized BIC	
Mean absolute percentage error Residual autocorrelation function (ACF)	
Mean absolute error Residual partial autocorrelation function (PACF)	
┌ Plots for Individual Models	
Series Residual autocorrelation function (ACF)	
Each Plot Displays Residual partial autocorrelation function (PACF)	
Observed values	
✓ Forecasts	
✓ Fit values	
Confidence intervals for forecasts	
Confidence intervals for fit values	
OK Paste Cancel Help	
A SFLECT INTERNATIONAL COM	PANY



 We can specify how far we wish to forecast into the future



	Ontions	
Tables Statistics Plots Output Filte	Save Options	
orecast Period		
) <u>F</u> irst case after end of estimation pe	eriod through last case in active dataset	
Date:	eriod through a specified date	
Year Month		
1999 12		
ser-Missing Values	Confidence Interval Width (%):	05
ser-Missing Values	Confidence Interval <u>W</u> idth (%):	95
Iser-Missing Values Treat as invalid Treat as valid	Confidence Interval <u>W</u> idth (%): <u>P</u> refix for Model Identifiers in Output:	95 Model
Iser-Missing Values 	Confidence Interval <u>W</u> idth (%): <u>P</u> refix for Model Identifiers in Output: Ma <u>x</u> imum Number of Lags Shown in ACF	95 Model 24
Iser-Missing Values) <u>T</u> reat as invalid) Treat as valid	Confidence Interval <u>W</u> idth (%): <u>P</u> refix for Model Identifiers in Output: Ma <u>x</u> imum Number of Lags Shown in ACF and PACF Output:	95 Model 24
ser-Missing Values) Treat as invalid) Treat as valid	Confidence Interval <u>W</u> idth (%): <u>P</u> refix for Model Identifiers in Output: Ma <u>x</u> imum Number of Lags Shown in ACF and PACF Output:	95 Model 24
ser-Missing Values preat as invalid Treat as valid	Confidence Interval <u>W</u> idth (%): <u>P</u> refix for Model Identifiers in Output: Ma <u>x</u> imum Number of Lags Shown in ACF and PACF Output:	95 Model 24
ser-Missing Values) <u>T</u> reat as invalid) Treat as valid	Confidence Interval <u>W</u> idth (%): <u>P</u> refix for Model Identifiers in Output: Ma <u>x</u> imum Number of Lags Shown in ACF and PACF Output:	95 Model 24





Interpreting Output and Model Fit

www.sv-europe.com

 Note that expert modeler chooses Winters' Additive as the best model type – this model is used with series showing a trend and seasonal effects



Model Summary

								Percentile			
Fit Statistic	Mean	SE	Minimum	Maximum				50		90	
Stationary R-squared	.735		.735	.735	.735	.735	.735	.735	.735	.735	.735
R-squared	.815	.V	Ve can	ignore	these	statisti	cal valu	les as t	they⊫ar	e used	to .815
RMSE	5296.069		5296.069	5296 069	5296.069	5296.069	-f ^{296,069}	5296.069	5296 p69	5296.069	, 5296.069
MAPE	9.412		compa			nance	9.412		oue _{9.412}		9.412
MaxAPE	83.036		83.036	83.036	83.036	releva	nt here	83.036	83.036	83.036	83.036
MAE	3624.699		3624.699	3624.699	3624.699	3624.699	3624.699	3624.699	3624.699	3624.699	3624.699
MaxAE	19302.778		19302.778	19302.778	19302.778	19302.778	19302.778	19302.778	19302.778	19302.778	19302.778
Normalized BIC	17.269		17.269	17.269	17.269	17.269	17.269	17.269	17.269	17.269	17.269

Model Fit


Model Fit Statistics

- Stationary R square A stationary model is effectively one with the trend removed so that the values have the same variance and mean over time. Larger values indicate better fit.
- **Root mean square error (RSME)** Squared errors are based on the square of the differences between *the fitted values and the observed values*. It's similar to a standard deviation value. *Smaller values indicate better fit.*
- Mean absolute percentage error (MAPE) The average error values in percentage terms. Values such as 0.15 equate to an average of 15% error. *Smaller values indicate better fit.*
- **Mean absolute error** Mean of the absolute values of the forecast errors. MAE is in the same units as the dependent series. MAE is appropriate when the cost of the forecast errors is proportional to the absolute size of the forecast error. *Smaller values indicate better fit.*
- **Maximum absolute percentage error** The largest forecast error, expressed as a percentage. This measure gives a worst case scenario indication of model performance. It works best if there are no extremes to the data. *Smaller values indicate better fit.*
- **Maximum absolute error** -The largest forecast error. Expressed in the same units as the dependent series. This measure gives a worst case scenario indication of model performance. *Smaller values indicate better fit.*
- **Normalized BIC** The Normalized *Bayesian Information Criterion* (BIC) fit measure enables you to compare different models for the same series. Normalized BIC "rewards" simpler models that fit better, while it "penalizes" models that use more parameters. It is based on a mean squared **This is the fit measure that Expert Modeler uses when comparing candidate models**. *Smaller values indicate better fit*.



• Ljung-Box Q: a lack of fit test to check that the model is correctly specified

		Model Fit statistics	Ljung-Box Q(18)		8)		
Model	Number of Predictors	Stationary R-squared	Statistics	DF	Sig.	Number of Outliers	
Sales of Women's Clothing-Model_1	0	.735	23.635	15	.072	0	

Model Statistics

- Time series models are often evaluated by focussing on the *errors* in the model.
- These are also known as residuals and they represent the difference between the fit values and the actual values.
- Ideally when the model has been fitted, the correlation between the residuals in the sequence should be random (white noise).
- Values in the Sig. column above 0.05 indicate that the model doesn't leave any significant correlations after the model has been specified. So we *might* assume that it's doing a reasonable job (overall) of fitting the series.







 Lets look at the effect of changing the *range of time* that we submit to the Expert Modeler algorithm.

Date [date] Sales of Women's Clothing [clothing_sales] YEAR, not periodic [YEAR_] MONTH, period 12 [MONTH_]	Select ^A All cases ^A f condition is satisfied ^I f ^A Rangom sample of cases ^{Sample} ^B Based on time or case ranter and the set response of the set of the
Current Status:	Paste Reset Cancel Help



• Although the chosen Model Type is still Winter's Additive. The fit statistics are slightly different when using only the last 5 years of data

Fit Statistic	Mean
Stationary R-squared	.735
R-squared	.815
RMSE	5296.069
MAPE	9.412
MaxAPE	83.036
MAE	3624.699
MaxAE	19302.778
Normalized BIC	17.269

Model 1: 1989 - 1998

Fit Statistic	Mean
Stationary R-squared	.822
R-squared	.815
RMSE	5552.480
MAPE	9.634
MaxAPE	68.892
MAE	3897.216
MaxAE	19307.360
Normalized BIC	17.449

Model 2: 1994 - 1998







www.sv-europe.com

- As we have seen, the Expert Modeler in SPSS Forecasting attempts to automatically find the best-fitting model for each dependent series. So far we have only been working with Exponential Smoothing models However, by default the Expert Modeler considers two types of Time Series model: Exponential Smoothing and **ARIMA**.
- If we choose to work with any **independent** (predictor) variables, then the Expert Modeler will select an ARIMA model if any of these independent variables have a statistically significant relationship with the dependent series.
- The acronym ARIMA refers to the **three components** of this modelling approach.
- It stands for Autoregressive (**AR**) Integrated (**I**) Moving Average (**MA**)
- This is the <u>structure</u> of an ARIMA model. But not all ARIMA models use <u>all</u> these elements. In fact, most don't.



- The three components within the ARIMA model are usually shortened to *p,d,q*
 - **p** refers to the **autoregression** component
 - d refers to the integration or differencing
 - **q** refers to the **moving average** component



Autoregression (p) refers to the correlation between a value in a series and the previous value(s). If I want to know the temperature today, is it useful to know the temperature yesterday? If so, then the autoregression p component is equal to 1 as this represents a lag of 1 day (a *first order* autoregression). If however the *day before yesterday* is a better component, then a value of 2 should be used to signify this (a *second order* autoregression).



- Integration or *differencing (d)*. Many analysts prefer the term 'differencing' rather than 'integration' for the **d** component of ARIMA.
- Differencing is a technique to used to make a time series stationary. A stationary series is one where the mean, the variance and autocorrelation values are constant over time.
- You may recall that Time Series analysis attempts to break a series up into different components by isolating elements like the trend component. ARIMA models require a series to be stationary in order to estimate correctly and the differencing component deals with this.
- Differencing (d) simply refers to the difference between one value in a sequence and another. If we were looking at daily temperatures, A first order differencing would result in a sequence of values showing the difference between one day and the previous day. These values could of course be positive or negative.



- Moving Average (q). We shouldn't confuse the term moving average here with the kind of moving average *smoothing* exercise that we undertook earlier. In fact the **q** component is another kind of autoregression. Except that this time it focusses on the *errors* (sometimes called 'shocks') in the forecasting model.
- A first order **q** component (denoted as '1') in an ARIMA model indicates that the model's error in the immediate previous period is related to what the dependent variable will be now. For example, knowing that the model overestimated or underestimated by 20% in the previous period helps to predict the current value.



- Normally analysts thoroughly explore their time series data before attempting to fit a model
- In doing so they tend to pay a lot of attention to special charts known as correlograms
- These important charts have two forms:
 - **1.** ACF Autocorrelation Function
 - 2. PACF Partial Autocorrelation Function
- These charts are especially important if one needs to *manually* specify an ARIMA model as they can be used to give a clues to the what the **p**, **d**, **q** values should be.



• ACF – Autocorrelation Function

- This displays how well the present value of the series is related with its past values
- Each bar shows the correlation with the present value at a given lag number
- In the batteries data we can see strong correlations at lag = 1 (i.e. the previous month's sales) and at lag = 12 (i.e. the same month in the previous



• PACF – Partial Autocorrelation Function

• Unlike the ACF it looks at the correlation of the residuals (the error remaining after removing the variation already explained by the earlier lags) with the current time point. It is a pure measure of correlation as it controls for all correlation values up to that time point.



 Before building an ARIMA model, we can investigate the autocorrelation in the series.





Demo 4: Using predictor fields with ARIMA

- Both the ACF and PACF plots show strong seasonal correlations at the 12 and 24 month lag points.
- A weaker correlation is also shown at the non-seasonal 1st order (i.e. lag 1)
- Based on this, the ARIMA model should probably have at least this structure: (0,0,0) (1,0,0)





- In this section we can use the clothing sales file but with two additional independent variables: **mail** and **strike**
- The variable **strike** has only two values: 1 and 0. This is called an **event** variable.
- Remember this is what the series looks like:





Demo 4: Using predictor fields with ARIMA

- As an experiment, let's return to the Time Series Modeler dialog an **manually** specify and ARIMA model.
- To override Expert Modeler and specify our own ARIMA model, click the drop-down button marked Method and change it to ARIMA
- To specify the **p**,**d**,**q** values we click:
 - Criteria

ime Series Modeler	×
Variables Statistics Plots Output Filter Save Opt	tions
Variables:	Dependent Variables:
Date [date] Number of Catalogs Mailed [mail] strike YEAR, not periodic [YEAR_] MONTH, period 12 [MONTH_] DIFF(clothing_sales,1) [DIFF1]	Sales of Women's Clothing [clothing_sales]
	Independent Variables:
	•
Method: ARIMA	Criteria
Model Type	e: ARIMA(0, 0, 0)(0, 0, 0)
Estimation Period	Forecast Period
Start: First case	Start: First case after end of estimation period
End: Last case	End: Last case in active dataset
OK Paste	Reset Cancel Help



Demo 4: Using predictor fields with ARIMA

- Further exploratory analyses indicate that not only is there strong seasonal correlation as shown in the ACF and PACF charts but that the data is not stationary so we should add a differencing term
- With this in mind let's specify the P,D,Q values in the seasonal column as (1,1,1).
- This means our ARIMA model is (0,0,0), (1,1,1)
- To see how the model performs click:
 - Continue
 - OK

🚰 Time Series Modeler: ARIMA Criteria	a	×
ARIMA Orders		
	Nonseasonal	Seasonal
Autoregressive (p)	0	1
Difference (d)	0	1
Moving Average (g)	0	1
Transformation	Current periodicity: 12	
One		
◎ S <u>q</u> uare root		
◎ Naṯural log		
Include constant in model		
	Cancel Help	



Winter's Additive Exponential Smoothing

Fit Statistic	Mean
Stationary R-squared	.735
R-squared	
RMSE	5296.069
MAPE	9.412
MaxAPE	83.036
MAE	3624.699
MaxAE	19302.778
Normalized BIC	17.269

ARIMA (0,0,0) (1,1,1)

Fit Statistic	Mean
Stationary R-squared	.306
R-squared	.753
RMSE	6084.494
MAPE	11.374
MaxAPE	84.309
MAE	4409.612
MaxAE	21318.207
Normalized BIC	17.557

• The first thing we notice is that the Stationary R-Squared value is much smaller. This is the effect of adding the **D** (differencing component to the model). But apart from that there's not a **huge** difference between the fit values.



- The sequence charts for the Exponential Smoothing model vs our manually specified ARIMA model show that they look very similar.
- Note that the ARIMA model doesn't include a model fit line for the first 12 months of the data (as it would need the previous 12 months to do so)





- Returning to Time Series Modeler dialog we can use the Expert Modeler function to automatically specify a model for us
- Obviously if we just ran this procedure as we did at the start, it will select a Winters Additive Exponential Smoothing model again
- So we will force it to *only consider ARIMA models*

Time Series Modeler	ots Output Filter Save C	ptions		×
(ariables:	Mailed [mail] /EAR_] JONTH_]	•	Dependent Variables: Papendent Variables: Papendent Variables:	C Time Series Modeler: Expert Modeler Criteria C Model Outliers Model Type ○ All models ○ Exponential smoothing models only ⓒ ABIMA models only ⓒ ABIMA models only ⓒ Expert Modeler considers seasonal models Current periodicity: 12 Events Independent Variables:
	Method: Expert	Modeler	Criteria	Event Type Variable
Estimation Period Start: First case End: Last case	Model I	Forecas Start: F End: L Reset	models only st Period irst case after end of estim ast case in active dataset Cancel Help	mi Event variables are special independent variables that are used to model effects of external occurrences such as a flood, strike, or introduction of a new product line. Check all variables you want to treat as event variables. Each should be coded such that i unicates a time point where an event is thought to have had an effect.
1990	9 SEP 19	90		Continue Cancel Help
1990	11 NOV 19	90		



Model Description

			Model Type
Model ID	Sales of Women's Clothing	Model_1	ARIMA(0,0,0) (0,1,1)

- The results show that Expert Modeler has *dropped* the seasonal autoregression term (P). Remember, this method looked at the previous relevant time point and uses it to estimate the next timepoint based on a simple regression formula
- It has however kept the differencing function (D). This method is used to make the series stationary so the variance doesn't change over time.
- It has also kept the moving average term (Q). This method fits a model by starting from an average value and constantly takes into consideration how wrong it was in the previous time point every time it makes a forecast one time point into the future.



Winter's Additive Exponential Smoothing

Fit Statistic	Mean
Stationary R-squared	.735
R-squared	.815
RMSE	5296.069
MAPE	9.412
MaxAPE	83.036
MAE	3624.699
MaxAE	19302.778
Normalized BIC	17.269

ARIMA (0,0,0) (1,1,1)

Fit Statistic	Mean
Stationary R-squared	.306
R-squared	.753
RMSE	6084.494
MAPE	11.374
MaxAPE	84.309
MAE	4409.612
MaxAE	21318.207
Normalized BIC	17.557

ARIMA (0,0,0) (0,1,1)

Fit Statistic	Mean
Stationary R-squared	.304
R-squared	.752
RMSE	6062.496
MAPE	11.364
MaxAPE	85.067
MAE	4400.095
MaxAE	21802.869
Normalized BIC	17.506

• Remember that the Expert Modeler attempts to find models with the smallest Normalised BIC value. That's why it chose Winter's Additive in our first example and the ARIMA (0,0,0) (0,1,1) when we forced it to only consider ARIMA models. Nevertheless the improvement in fit over our manually specified model is very small.



- The dataset also contains two predictor fields which can be specified as independent variables. In this context Exponential Smoothing algorithms don't make use of predictor fields, but ARIMA can.
- The variable **mail** refers to the number of catalogues that were posted to customers. Note that if we wish to use this to forecast into the future we would need to specify the future values for each monthly mailing volume. You can see these anticipated future values at the bottom of the data file.
- The variable **strike** indicates whether or not a postal strike *event* had occurred that month.

	🖋 mail	🗞 strike	I YEA
1	11288	0	19
7	11096	0	19
6	11224	0	19
6	11483	0	19
5	11643	0	19
6	10893	0	19
7	11147	1	19
7	12260	0	19
4	11168	0	19
7	14370	0	19
8	11890	0	19
8	11722	0	19
5	11589	0	19
7	11633	0	19
5	11951	0	19
6	11706	0	1 9
0	11/60	0	10



Winter's Additive Exponential Smoothing

Fit Statistic	Mean
Stationary R-squared	.735
R-squared	.815
RMSE	5296.069
MAPE	9.412
MaxAPE	83.036
MAE	3624.699
MaxAE	19302.778
Normalized BIC	17.269

ARIMA (0,0,0) (0,1,1) <u>without</u> independent variables

Fit Statistic	Mean
Stationary R-squared	.304
R-squared	.752
RMSE	6062.496
MAPE	11.364
MaxAPE	85.067
MAE	4400.095
MaxAE	21802.869
Normalized BIC	17.506

ARIMA (0,0,0) (0,1,1) <u>with</u> independent variables

Fit Statistic	Mean
Stationary R-squared	.492
R-squared	.803
RMSE	5349.490
MAPE	9.810
MaxAPE	46.514
MAE	4045.744
MaxAE	16525.282
Normalized BIC	17.309

 The model fit seems to have improved slightly on the previous ARIMA model that did not include independent variables. Many of the statistics in the Exponential Smoothing model still seem to indicate a slightly better overall fit. But the biggest change is in the MaxAPE and MaxAE where the maximum absolute percentage error and maximum absolute error values are both quite smaller. This is almost certainly due to the inclusion of the **strike** variable.



The immediate effect of adding an event variable can be seen in the sequence chart for the model. Postal strikes occurred in June 1996 and September 1997. As such, the event variable contains the value 1 at each of these time points. The model immediately picks this up so the fit line correctly reflects the downturn in revenue on these two occasions. Thus an event variable can be used to help explain anomalous values caused by external effects with the aim of providing more accurate estimates.







www.sv-europe.com

- One of the simplest ways to create new data *showing forecasted values* is to simply add the forecasts to the end of the data file itself.
- In our current example file, it's necessary to add values for the two independent variables mail and strike because they are part of the model and will therefore be needed to generate these new values.

Eile <u>E</u> dit	<u>View Data Tr</u>	ansform <u>Analyze G</u> raphs <u>U</u> tili	ties Extensions	<u>W</u> indow <u>H</u> elp			
🚖 🔚		r 🤉 🖹 🛓 重	P # 9	A 			
132 : strike	0						
	🔏 date	clothing_sales	🖋 mail	💰 strike	✓ YEAR_	✓ MONTH_	💑 DATE_
113	05/01/1998	41038.75	11951	0	1998	5	MAY 1998
114	06/01/1998	48329.36	11706	0	1998	6	JUN 1998
115	07/01/1998	44961.98	11460	0	1998	7	JUL 1998
116	08/01/1998	58660.76	11808	0	1998	8	AUG 1998
117	09/01/1998	57791.14	12781	0	1998	9	SEP 1998
118	10/01/1998	56329.40	11690	0	1998	10	OCT 1998
119	11/01/1998	54617.35	11393	0	1998	11	NOV 1998
120	12/01/1998	80245.97	15263	0	1998	12	DEC 1998
121			14370	0	1999	1	JAN 1999
122] .		11890	0	1999	2	FEB 1999
123] .		11722	0	1999	3	MAR 1999
124] .		11589	0	1999	4	APR 1999
125			11633	0	1999	5	MAY 1999
126			11951	0	1999	6	JUN 1999
127			11706	0	1999	7	JUL 1999
128			11460	0	1999	8	AUG 1999
129] .		11808	0	1999	9	SEP 1999
130] .		12781	0	1999	10	OCT 1999
131			11690	0	1999	11	NOV 1999
132			11393	0	1999	12	DEC 1999



- Simply return to the Time Series dialog and click the tab marked:
 - Save
- Then check the appropriate options to request the predicted values

Variables:	15		
	Description	Save	Variable Name Prefix
Predicted V	alues	V	Predicted
Lower Cont	idence Limits		LCL
Opper Con	idence Limits		UCL
Noise Resi	duals		NResidual
For each ite	m you select, one variable File	is saved per dependent varia	ble.
For each iter xport Model KML <u>F</u> ile:	m you select, one variable File	is saved per dependent varia	ble. Browse ations.
For each iter Export Model KML <u>F</u> ile:	m you select, one variable File XML files are only	is saved per dependent varia	ble. Browse ations. Browse



• The forecasted clothing sales for the year 1999 are now shown in the dataset.

🔏 date	clothing_sales	🛷 mail	\delta strike	✓ YEAR_	MONTH_	💑 DATE_	Predicted_clothin
							<pre>g_sales_Model_1</pre>
05/01/1998	41038.75	11951	0	1998	5	MAY 1998	47359.96
06/01/1998	48329.36	11706	0	1998	6	JUN 1998	47391.22
07/01/1998	44961.98	11460	0	1998	7	JUL 1998	49934.48
08/01/1998	58660.76	11808	0	1998	8	AUG 1998	45590.14
09/01/1998	57791.14	12781	0	1998	9	SEP 1998	54405.31
10/01/1998	56329.40	11690	0	1998	10	OCT 1998	56927.51
11/01/1998	54617.35	11393	0	1998	11	NOV 1998	53767.85
12/01/1998	80245.97	15263	0	1998	12	DEC 1998	73884.71
		14370	0	1999	1	JAN 1999	34621.09
		11890	0	1999	2	FEB 1999	33784.23
		11722	0	1999	3	MAR 1999	48400.32
		11589	0	1999	4	APR 1999	49167.51
		11633	0	1999	5	MAY 1999	49522.23
		11951	0	1999	6	JUN 1999	55169.71
		11706	0	1999	7	JUL 1999	45732.82
		11460	0	1999	8	AUG 1999	50807.40
		11808	0	1999	9	SEP 1999	59576.83
		12781	0	1999	10	OCT 1999	68085.74
		11690	0	1999	11	NOV 1999	54808.51
		11393	0	1999	12	DEC 1999	76528.76



 An alternative approach is to save the model as an XML file.

Time Series Modeler			×					
Variables Statistics Plots Output Filter Save Options								
Save Variables								
Variables								
Valiables.								
Predicted Values	Save	Predicted						
Lower Confidence Limits		LCL						
Upper Confidence Limits		UCL						
Noise Residuals		NResidual						
For each item you select, one variable is saved per depen	dent variable.							
Export Model File								
XML File: D:\SE Demos\SmartVision Seminars and W	ebinars\Training Webinars\Introduction to Time Se	ries Forecasting\DataARIMA Model 1.xml	Browse					
() XML files are only compatible with SP	SS applications.							
J								
PMML File:			Browse					
	energiant configurations, including ODOO							
PMML lifes are compatible with PMML	-compliant applications, including SPSS.							
		_						
	OK Paste Reset Cancel Hel	p						



🔚 *clothing sales with independent variables.sav (DataSet5) - IBM SPSS Statistics Data Editor

 We can apply the saved xml model file to a new dataset by using the SPSS
 Scoring Wizard

ile	<u>E</u> dit	<u>V</u> iew <u>D</u> ata <u>T</u> ra	ansform <u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	E <u>x</u> tensions	<u>W</u> indow	<u>H</u> elp	
2				- 	📕 <u>V</u> ariab	les			0
					<u>ವಿ о</u> мз с	Control Panel		14	
:		<u> </u>			📇 OMS [dentifiers			
		💰 date	🖉 🖉 clothing	g_sales	🚑 Scorin	g <u>W</u> izard		trike	🖋 YEA
	1	01/01/1989		16578.	🙀 Merge	Model <u>X</u> ML		0	19
2	2	02/01/1989		18236.	E Calcul	ate with Pivot	Table	0	19
(3	03/01/1989		43393.	Data F	ile <u>C</u> omments Variable Macr	·	0	19
4	4	04/01/1989		30908.	Z Define	Variable Sets		0	19
į	5	05/01/1989		28701.	🛨 Censo	or Table		0	19
(6	06/01/1989		29647.	🌀 <u>U</u> se V	ariable Sets		0	19
-	7	07/01/1989		31141.	Show	All Variables		0	19
8	В	08/01/1989		31177.	₩ <u>S</u> pellir	ng		0	19
(9	09/01/1989		30672.	+ Proces	ss Data Files		0	19
1	0	10/01/1989		37633.	b <u>R</u> un S	cript		0	19
1	1	11/01/1989		33890.	Produ	ction Facility		0	19
1	2	12/01/1090		51279	<u>M</u> ap C	10200	ty	0	10



For a deeper dive, try our online course



Introduction to Time Series Forecasting with IBM SPSS Statistics

Home / Online courses / Introduction to Time Series Forecasting with IBM SPSS Statistics

176 students

Overview Curriculum Instructor

Introduction to Time Series Forecasting

Course materials

£75.00

Buy this course



Download our e-book for free



The insider's guide to predictive analytics

£0.00



Category: books


Working with Smart Vision Europe Ltd.

• Sourcing Software

- You can buy your analytical software from us often with discounts
- Assist with selection, pilot, implementation & support of analytical tools
- <u>http://www.sv-europe.com/buy-spss-online/</u>
- Training and Consulting Services
 - Guided consulting & training to develop in house skills
 - Delivery of classroom training courses / side by side training support
 - Identification & recruitment of analytical skills into your organisation
- Advice and Support

urope

offer 'no strings attached' technical and business advice relating to analytical activities

Technical support services around SPSS



Contact us:

+44 (0)207 786 3568 info@sv-europe.com Twitter: @sveurope Follow us on Linked In Sign up for our Newsletter

Thank you

www.sv-europe.com

A SELECT INTERNATIONAL COMPANY