

International Airline Passengers

In this paper, we will analyze the international passenger data series (G), as mentioned in the textbook *Time Series: Forecast and Control* by Box, Jenkins and Reinsel (ISBN: 978-0470272848). This textbook was first published in late 1960s, and is considered by many practitioners as the definite foundation textbook on the time series topic.

The international airline passenger series describes monthly totals of the international passengers for the period between Jan 1949 and Dec 1960.

The objective here is to follow the same analysis in the book, and to demonstrate the accuracy of NumXL calculations. Furthermore, **SAS** – a leading statistical software vendor – demonstrate their own analysis for the same data set, so we strongly recommend our users to review their results as well, via this link:







Step 1: Data Transformation

Using the Descriptive Statistic wizard (pictured below), examine the different summary stats of the sample data.

۵ م	Σ III	1	A reprint -			
STA	TS Correlogr	am ARMA AIRI	LINE GARCH	GLM Calibra	Tation Forecast About	
7	2		NumXL		Descriptive Statistics	
	D20	- (**	f_x		2 Input Data set	
	A	В	С	D	Time series Box-Jenkins'!\$8\$2:\$8\$145	
1	Date	Passengers			Ascending	
2	Jan-49	112	1	600 -	ri ricerang	In
3	Feb-49	118			Statistical Testing	ec
4	Mar-49	132		500	Average 🔽 Mean Test	
5	Apr-49	129			STD Deviation	
6	May-49	121				
7	Jun-49	135		400 -		
8	Jul-49	148			I♥ Excess Kurtosis I♥ Kurtosis Test	
9	Aug-49	148		300 -	Median White-noise Test	
10	Sep-49	136			Minimum Vormal Distribution Test	~
11	Oct-49	119		200 -	ARCH Effect Test	Γ
12	Nov-49	104			Significance Level (e.g. 5%)	
13	Dec-49	118		/		
14	Jan-50	115		100 -		
15	Feb-50	126				
16	Mar-50	141		0 +	Output Range \$D\$20	-
17	Apr-50	135		Jan-4	st l	an-5
18	May-50	125		and the second s	Help Cancel OK	
19	Jun-50	149		and the second		
20	Jul-50	170	1			

In the summary statistics table (pictured below), the data series exhibits serial correlation (i.e. failed the white noise test) and fat tails (significant excess kurtosis and ARCH effect).



D	E	F	G	Н	1	J	K	L	М	N	0
Descript	ive Statistics			Significant	e Test		5.00%		Test	p-value	SIG?
				Target	P-Value	SIG?		1% Test White-nois Normal Distributed ARCH Effec		0.00%	FALSE
	AVERAGE:	280		0.000	0.00%	TRUE		% Test White-noi Normal Distribute ARCH Effec		1.15%	FALSE
	STD DEV:	120						A	RCH Effect?	0.00%	TRUE
	SKEW:	0.58		0.000	0.23%	TRUE	1				-
EXCES	S-KURTOSIS:	-0.36		0.000	16.74%	FALSE					
	MEDIAN:	265.5						% Test White-noi: Normal Distributed ARCH Effec			
	MIN:	104									
	MAX:	622									
	Q 1:	180									
	Q 3:	360.5									

The original analysis converts the data series using the natural logarithm function (i.e. LN). Follow the same technique, as shown in the graph below:



This technique should yield the following summary statistics:



E	F	G	H	1	J	K	L	M	N	0	Р
Descript	tive Statistics			Significant	e Test		5.00%		Test	p-value	SIG?
				Target	P-Value	SIG?		v	/hite-noise	0.00%	FALSE
	AVERAGE:	5.54		0.000	0.00%	TRUE		Normal D	istributed?	4.93%	FALSE
	STD DEV:	0.44						A	RCH Effect?	0.00%	TRUE
	SKEW:	-0.12		0.000	27.71%	FALSE					· · · · · · · · · · · · · · · · · · ·
EXCES	S-KURTOSIS:	-0.96		0.000	0.86%	TRUE	1				
	MEDIAN:	5.58									
	MIN:	4.64									
	MAX:	6.43					1				
	Q 1:	5.19					Ī				
	Q 3:	5.89									

Notice that the transformed data series is smoother than the original data and the time trend appears more linear than the original.



 Phone:
 1-888-427-9486

 1-312-324-0367

 Fax:
 1-312-238-9092

 info@spiderfinancial.com

Step 2: Correlogram Analysis

Using the NumXL toolbar, launch the Correlogram wizard.

4	Σ	г 🚮	Nr interferes	/ 🖉 🖟 🕗
STA	TS Correlo	gram ARMA A	IRLINE GARCH	GL Correlogram
		2	NumXL	a input Data set
	E65		• fx	3 index bota set
A	А	В	С	Time series
1	Date	Passengers	LOG	Ascending
46	Sep-52	209	5.34	107
47	Oct-52	191	5.25	ACF
48	Nov-52	172	5.15	Calculate ACF IV Calculate PACF
49	Dec-52	194	5.27	Max Lags 24 Max Lags 24
50	Jan-53	196	5.28	4 5
51	Feb-53	196	5.28	✓ Upper & Lower Limits? ✓ Upper & Lower Limits?
52	Mar-53	236	5.46	Fig. Plat Graph2
53	Apr-53	235	5.46	
54	May-53	229	5.43	Significance Level (e.g. 5%)
55	Jun-53	243	5.49	0.05
56	Jul-53	264	5.58	
57	Aug-53	272	5.61	Table Bange
58	Sep-53	237	5.47	seses
59	Oct-53	211	5.35	Graphs Range éméss
60	Nov-53	180	5.19	
61	Dec-53	201	5.30	
62	Jan-54	204	5.32	Help Cancel OK V
63	Feb-54	188	5.24	
64	Mar-54	235	5.46	
65	Apr-54	227	5.42	
66	May-54	234	5.46	1

Highlight the log data and select 24 lags for ACF and PACF. Then, create a correlogram for the data.

	w.spid	erfina	ncial.c	com		AL														F	ax	: sp	ide	l- l- erfi	-31: -31: ina	2-3 2-2 nc	24- 38- ial.	0367 9092 com
Correlogr	am Analysis						100%	-																				$\pm h$
Lag	ACF	UL	ш	PACE	UL	ш	80%									-	ACF									_		- 1
1	95.37%	16.33%	-16.33%	96.35%	16.33%	-16.33%	60%										_	-	1			1						
2	89.89%	16.33%	-16.33%	-21.49%	16.33%	-16.33%	40%					-	-															
3	85.08%	27.42%	-27.42%	15.47%	16.33%	-16.33%	2016																					
4	80.84%	34,40%	-34.40%	8.29%	16.33%	-16.33%	2070																					
5	77.89%	39.62%	-39.62%	32.18%	16.33%	-16.33%	0%	,	1	1 1	1		T		1	-								1				
6	75.64%	43.80%	-43.80%	-1.69%	16.33%	-16.33%	-20%	-																				
7	73.76%	47.35%	-47.35%	7.34%	16.33%	-16.33%	-40%	-			-	_	_															
8	72.71%	50.47%	-50.47%	21.45%	16.33%	-16.33%	-60%	-								-	_	_										
9	73.36%	53.27%	-53.27%	53.92%	16.33%	-16.33%	-80%																	-	_	-	_	- 1
10	74.43%	55.85%	-55.85%	22.89%	16.33%	-16.33%	100%																					
11	75.80%	58.37%	-58.37%	65.09%	16.33%	-16.33%	-100/0	1	2 3	4	5	6 7		9	10	11	12	13	14	15 1	6 1	1	8 19	2	0 21	22	23	24
12	76.19%	60.85%	-60.85%	55.96%	16.33%	-16.33%	100%	1			1				1		-	-									-	
13	71.65%	63.31%	-63.31%	-57.05%	16.33%	-16.33%	000										PAC	F										
14	66.30%	65.71%	-65.71%	-29.03%	16.33%	-16.33%	3076	1																				
15	61.84%	67.77%	-67.77%	11.32%	16.33%	-16.33%	60%								_			-										
16	57.62%	69.48%	-69.48%	-14.28%	16.33%	-16.33%	40%																					
17	54.38%	70.93%	-70.93%	0.17%	16.33%	-16.33%	-									_		_									-	
18	51.95%	72.17%	-72.17%	-1.71%	16.33%	-16.33%	2070	1										_		-			_			1		
19	50.07%	73.25%	-73.25%	14.38%	16.33%	-16.33%	0%	+		-1						-	-	-					-			-		
20	49.04%	74.23%	-74.23%	-6.58%	16.33%	-16.33%	-20%	-			-		-		-	_	-			_		-	_	-	1.0.1	-		
21	49.82%	75.12%	-75.12%	16.06%	16.33%	-16.33%													-									
22	50.62%	75.97%	-75.97%	24.85%	16.33%	-16.33%	-40%	1										_										
23	51.67%	76.84%	-76.84%	25.83%	16.33%	-16.33%	-60%	-					_											_				
24	52.05%	77.72%	-77.72%	8.17%	16.33%	-16.33%	-30%	1																				
								1	2	3 4	5	6	7	8	9 1	0 1	1 12	1	14	15	16	17	18	19	20	21	22	3 24

Examining the ACF plot, the data appear to be integrated at lag one(1) and at lag twelve (12). Difference the data for both lags (i.e. $(1-L)(1-L^{12})LN(X)$), as shown in the graph below:



The differenced data set should generate the following correlogram:

SPIDERFINANCIAL

Phone: 1-888-427-9486

5	חוק	FRI	FIN	ΛN		A I													Ph	on	e:	1-	888	3-4	27-	-94	186
91						AL																1-	31	2-3	24	-03	36/
ww	w.spid	errina	ncial.	com															Fa	X :		1-	31	2-2	38-	-90)92
																		in	fo	@S	bid	erfi	ina	nc	ial	.co	om
Correlogr	am Analysis						30%																	-			
Lag	ACF	UL	u	PACE	UL	u									A	CF					_			_		_	_
1	-34.11%	17.12%	-17.12%	-34.12%	17.12%	-17.12%	20%	-	_	-	-			_													
2	10.50%	17.12%	-17.12%	-1.44%	17.12%	-17.12%	10%																				
3	-20.21%	19.01%	-19.01%	-19.40%	17.12%	-17.12%					_							-									
4	2.14%	19.18%	-19.18%	-12.64%	17.12%	-17.12%	0% -				1						1		-		-	-	1			1	
5	5.57%	19.80%	-19.80%	3.34%	17.12%	-17.12%	-10%					-															
6	3.08%	19.80%	-19.80%	3.41%	17.12%	-17.12%		-																			
7	-5.56%	19.85%	-19.85%	-6.72%	17.12%	-17.12%	-20% -		-								-		_	-	_	-	_	_		_	_
8	-0.08%	19.86%	-19.86%	-2.62%	17.12%	-17.12%	-30% -																				
9	17.64%	19.91%	-19.91%	24.36%	17.12%	-17.12%		-																			
10	-7.64%	19.91%	-19.91%	4.21%	17.12%	-17.12%	-40% -																				
11	6.44%	20.36%	-20.36%	4.89%	17.12%	-17.12%	-50%																				
12	-38.66%	20.45%	-20.45%	-38.34%	17.12%	-17.12%		1 2	3	4 5	6	7 4	9	10	11	12 1	3	14 1	5 1	15 1	7 1	8 19	20	21	22	23	24
13	15.16%	20.51%	-20.51%	-13.86%	17.12%	-17.12%	30%																				
14	-5.76%	22.54%	-22.54%	-7.94%	17.12%	-17.12%									P	ACI	-										
15	14.96%	22.84%	-22.84%	0.22%	17.12%	-17.12%	20%	1	_	-	-		_	-			-			-				<u> </u>	_	-	-
16	-13.89%	22.88%	·22.88%	-16.92%	17.12%	-17.12%	10%	1							-												
17	7.05%	23.17%	-23.17%	1.75%	17.12%	-17.12%	0%	to be					-	-	-	-			-	h r	-		-	and the second	-	-	
18	1.56%	23.41%	-23.41%	12.48%	17.12%	-17.12%	-10%					-						-					_				
19	-1.06%	23.47%	·23.47%	-1.95%	17.12%	-17.12%	-20%		-	-	-		-				-	_	_		_	-	_		_	-	_
20	-11.67%	23.47%	-23.47%	-20.47%	17.12%	-17.12%	2000																				
21	3.86%	23.48%	-23.48%	14.18%	17.12%	-17.12%	-30%																				
22	-9.14%	23.65%	·23.65%	-13.66%	17.12%	-17.12%	-40%	1																			
23	22.33%	23.66%	-23.66%	18.40%	17.12%	-17.12%	-50%	1																			
24	-1.84%	23.77%	-23.77%	-9.67%	17.12%	-17.12%		1 2	2 3	4 5	6	7	8	9 10	11	12	13	14	15	16	17	18	19 2	10 2	1 22	2 2	3 24

Also, notice that the ACF plot of the differenced data series shows a significant autocorrelation at lag one (1) and lag twelve (12).

Phone: 1-888-427-9486



 Phone:
 1-888-427-9486

 1-312-324-0367

 Fax:
 1-312-238-9092

 info@spiderfinancial.com

Step 3: Airline Modeling

The proposed model for the log passenger data series is an airline model with a season length of 12 months.

$$(1-L)(1-L^{12})X_{t} = \mu + (1+\theta L)(1+\Theta L^{12})a_{t}$$

Where:

L = the backshift operator (aka B)

 a_t = the error term, shock, innovation, or simply the model residual at time t

 μ = the mean of the seasonal differenced time series

Using the NumXL toolbar, click on the Airline icon to launch the Airline model wizard.

^	Σ it		2~ mp	• 1	AIRLINE Model
ST/	ATS Correlogra	m ARMA	AIRLINE GARC	H GLM C	Input Data set
		2	Num)	(L	Time series Modeling!\$C\$2:\$C\$145
		2	, ,	AIRLIN	Ascending
	В	С	D	E	
1	Passengers	LOG	AIRLINE		Model Specification Options
11	119	4.78	#N/A		Length of Seasonality 12 Guess parameters
12	104	4.64	#N/A		Goodness of Fit
13	118	4.77	#N/A	1	▲ ✓ Residuals Diagnosis
14	115	4.74	#N/A	1	Significance Level (e.g. 5%)
15	126	4.84	122,**	1	
16	141	4.95	139		0.05
17	135	4.91	137		
18	125	4.83	127		Output Range \$F\$23
19	149	5.00	141		
20	170	5.14	160		Help Cancel OK
21	170	5.14	166		5
22	158	5.06	155		
23	133	4.89	137		AIRLINE(12) Goodness-of-fit
24	114	4.74	117	1	Param Value LLF AIC

F	G	Н	1	J	К	L	М	N	0	P	Q	R	S	Т	U
AIRLINE(12)			Goodness	-of-fit		_		Residuals	(standardi	zed) Anal	ysis			
	Param	Value		LLF	AIC	CHECK			AVG	STDEV	SKEW	KURTOSIS	Noise?	Normal?	ARCH?
	μ	0.00029		218.42	-428.52	1.			0.00	1.00	0.04	1.24	FALSE	FALSE	TRUE
	θ	0						Target	0.00	1.00	0.00	0.00			11.1
	Θ	0						SIG?	FALSE	FALSE	FALSE	TRUE			
	σ	0.04567										9			
	5	12										1			



Step 4: Calibration

Select the cell at the top of the airline model table (i.e. "AIRLINE(12)") and click on the Calibration icon in the toolbar.

4	Σ 4	44.5	1	NH			V	Solver Parameters				23
STA	TS Corre	logram /	ARMA AIF	RLINE GARC		Calibeat	For	Se <u>t</u> Objective:	\$3\$25			
			(j	f _x Z	90879	37839:	To: O Max	© Mi <u>n</u> ─ <u>V</u> alu	e Of: 0		
	E	F	G	н	1	1	к	By Changing Variable Cel	s:			
20							_	\$H\$25:\$H\$28				1
21							-	Subject to the Constraint	s:			
23		AIRUNE	(2)			Goodness	of-fit	t +25 > = 0.00000	2.		1	
24			Param	Value		LLF	AIC	\$6\$25 >= 0.99999		^	Add	
25		_7 _	μ	0.00029		218.42	-428					5
26	- C .	5	θ	0							Change	
27	1		Θ	0								5
28			σ	0.04567			_				Delete	
29		_	s	12			_					_
30							_				Denot All	
31							_				Keset All	
32							_				Land/Dave	
33							_			*	Load/Save	
34							_	Make Unconstrained	/ariables Non-Negative			
35		Forecast					_	Coloritor Colorization de	an a u - t		1	_
36		Step	Mean	STD	UL	ш	_	Select a Solving Method:	GRG Nonlinear		Options	
37	Jan-61	1	6.10	0.05	6.19	6.01	0.18					
38	Feb-61	2	6.03	0.06	6.16	5.91	0.25	Solving Method				
39	Mar-61	3	6.10	0.08	6.26	5.95	0.31	Select the GRG Nonlines	r engine for Solver Probler	ns that are smooth nonli	near. Select the LP Simp	lex
40	Apr-61	4	6.20	0.09	6.38	6.02	0.36	engine for linear Solver	Problems, and select the E	volutionary engine for S	olver problems that are	
41	May-61	5	6.22	0.10	6.42	6.02	0.40	non-smooth.				
42	Jun-61	0	6.55	0.11	6.3/	6.15	0.44		4	_		
43	JUI-61	,	6.50	0.12	6.74	6.20	0.4/		4		<u></u>	
44	Aug-01 Sen-61	0	6.30	0.15	6.57	6.03	0.51	Help		3 Solver	Close	:
45	Oct-61	10	6.30	0.14	6.48	5.92	0.54			1	N .	

The Excel solver will try to determine the optimal values for the airline model's parameters (i.e. θ, Θ).

<u>Keep Solver Solution</u>	Answer Sensitivity Limits
○ <u>R</u> estore Original Values □ Return to Solver Parameters Dialog	Outline Reports
<u>Cancel</u>	Save Scenario
olver found a solution. All Constraints and	optimality conditions are



The new optimal values for the model parameters are shown below:

F	G	Н	1	1	ĸ	L	М	N	0	Р	Q	R	S	Т	U
AIRUNE	(12)			Graness	-of-fit				Residuals	(standard	ized) Ana	lysis			
	Param	Value	6	LLF	AIC	ECK		0	AVG	STDEV	SKEW	KURTOSIS	Noise?	Normal?	ARCH?
1	μ	0.00029		240.53	-472.75	1.		1	0.00	0.98	0.09	0.63	TRUE	TRUE	FALSE
	θ	0.4171						Target	0.00	1.00	0.00	0.00	1	1	A
	0	0.44248				<i>v</i>		SIG?	FALSE	FALSE	FALSE	FALSE	\sim	\checkmark	\checkmark
	σ	0.03825		-					A	A	A	A	~	40	
	5	12			-				\checkmark	\checkmark	\checkmark	\checkmark			
										-					

Examining the residuals analysis table, the calibrated values satisfy all assumptions of the underlying model (i.e. Gaussian distributed residuals).

The parameter values of the calibrated model on the SAS website are slightly different from the ones we calculated earlier:

	Maximu	m Likelihood Es	timation		
				Арргох	
Parameter	Estimate	Standard Error	t Value	Pr > t	Lag
MA1,1	0.40194	0.07988	5.03	<.0001	1
MA2,1	0.55686	0.08403	6.63	<.0001	12

However, our values are within the error tolerance limits (i.e. $\pm 1.96 \times \sigma$) and our Akaike's information criterion (AIC) is better.

The core difference between NUMXL and SAS values can be explained by noting that we did not set the intercept (μ) value to zero.



Step 5: Forecast

The residuals of the calibrated model satisfy the assumptions of an airline model. Now, we are ready to conduct a 24-month forecast for the monthly international airline passenger totals.

The forecast will follow two stages:

- Forecasting for the log of the monthly totals
- Transforming the forecast back into regular monthly totals

Select the cell labeled "AIRLINE(12)" and click on the Forecast icon in the toolbar.



Please note that for forecast purposes, the input time series here refers to the latest 13 months, or the observations between Nov 1959 and Dec 1960. The output table is shown below:

SPI	DEF	RFI	NA	N	C	Α	L
www.	spiderfi	nanci	al.con	n			

Step	Mean	STD	UL	LL
1	6.11	0.04	6.18	6.03
2	6.05	0.04	6.13	5.96
3	6.11	0.05	6.21	6.02
4	6.21	0.05	6.32	6.10
5	6.23	0.06	6.35	6.12
6	6.36	0.06	6.48	6.24
7	6.51	0.07	6.64	6.38
8	6.49	0.07	6.62	6.35
9	6.31	0.07	6.45	6.16
10	6.20	0.08	6.36	6.05
11	F 106	0,01	5.21	<u> </u>



To convert back to regular monthly totals, use these equations:

$$UL = e^{UL_{\log}}$$
$$LL = e^{LL_{\log}}$$
$$\mu = e^{\mu_{\log} + \frac{\sigma_{\log}^2}{2}}$$



 Phone:
 1-888-427-9486

 1-312-324-0367

 Fax:
 1-312-238-9092

 info@spiderfinancial.com

Step	Mean	STD	UL	LL
1	450		485	418
2	423		460	387
3	453		499	411
4	499		554	448
5	511		572	455
6	579		654	511
7	674		766	590
8	657		752	571
9	551		635	476
10	496		575	425
11	428		499	365
12	476		559	403
13	695		▲592▲	A10

