Chapter 35
Autoregressive Models for Longitudinal Data (120 Mean Monthly Population Records)

General Purpose

Time series are encountered in every field of medicine. Traditional tests are unable to assess trends, seasonality, change points and the effects of multiple predictors like treatment modalities simultaneously. To assess whether autoregressive integrated moving average (ARIMA) methods are able to do all of that.

Specific Scientific Question

Monthly HbA1c levels in patients with diabetes type II are a good estimator for adequate diabetes control, and have been demonstrated to be seasonal with higher levels in the winter. A large patient population was followed for 10 year. The mean values are in the data. This chapter is to assess whether longitudinal summary statistics of a population can be used for the effects of seasons and treatment changes on populations with chronic diseases.

Note:

No conclusion can here be drawn about individual patients. Autoregressive models can also be applied with data sets of individual patients, and with multiple outcome variables like various health outcomes.

Example

The underneath data are from the first year's observation data of the above diabetic patient data. The entire data file is in extras.springer.com, and is entitled "arimafile".

Date	HbA1	nurse	doctor	phone	self	meeting
01/01/1989	11,00	8,00	7,00	3	22	2
02/01/1989	10,00	8,00	9,00	3	27	2
03/01/1989	17,00	8,00	7,00	2	30	3
04/01/1989	7,00	8,00	9,00	2	29	2
05/01/1989	7,00	9,00	7,00	2	23	2
06/01/1989	10,00	8,00	9,00	3	27	2
07/01/1989	9,00	8,00	8,00	3	27	2
08/01/1989	10,00	8,00	7,00	3	30	2
09/01/1989	12,00	8,00	8,00	4	27	2
10/01/1989	13,00	9,00	11,00	3	32	2
11/01/1989	14,00	9,00	7,00	3	29	2
12/01/1989	23,00	10,00	11,00	5	39	3
01/01/1990	12,00	8,00	7,00	4	23	2
02/01/1990	8,00	8,00	6,00	2	25	3

Date = date of observation,
HbA1 = mean HbA1c of diabetes population,
nurse = mean number of diabetes nurse visits,
doctor = mean number of doctor visits,
phone = mean number of phone visits,
self = mean number of self-controls,
meeting = mean number of patient educational meetings

We will first assess the observed values along the time line. The analysis is performed using SPSS statistical software.

Command:

analyze….Forecast….Sequence Charts….Variables: enter HbA1c….Time Axis Labels: enter Date….OK.

Example   213
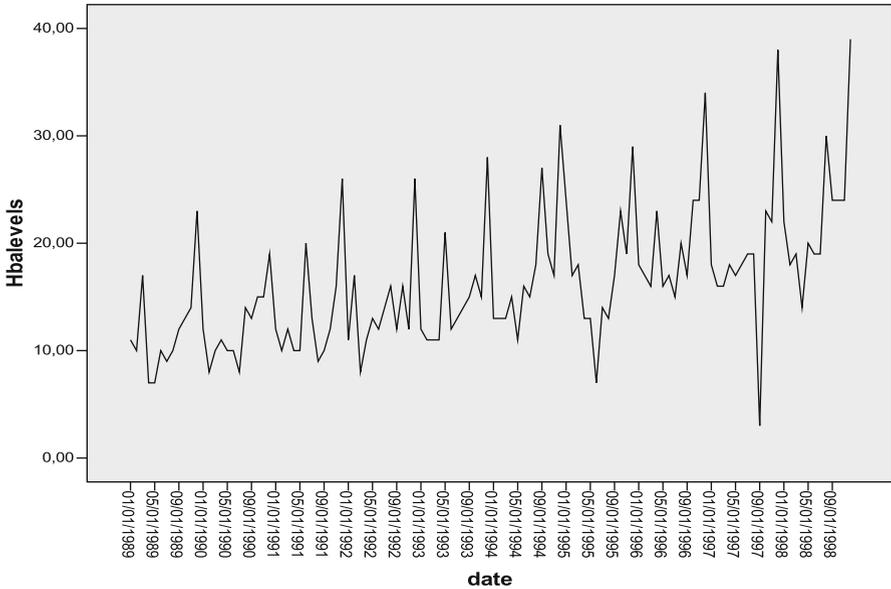

The above output sheets show the observed data. There are (1) numerous peaks, which are (2) approximately equally sized, and (3) there is an upward trend: (2) suggests periodicity which was expected from the seasonal pattern of HbA1c values, (3) is also expected, it suggests increasing HbA1c after several years due to beta-cell failure. Finally (4), there are several peaks that are not part of the seasonal pattern, and could be due to outliers.

ARIMA (autoregressive integrated moving average methodology) is used for modeling this complex data pattern. It uses the Export Modeler for outlier detection, and produces for the purpose XML (eXtended Markup Language) files for prediction modeling of future data.

Command:

Analyze….Forecast….Time Series Modeler….Dependent Variables: enter HbA1c….Independent Variables: enter nurse, doctor, phone, self control, and patient meeting….click Methods: Expert Modeler….click Criteria….Click Outlier Table….Select automatically ….Click Statistics Table….Select Parameter Estimates….mark Display forecasts….click Plots table….click Series, Observed values, Fit values….click Save….Predicted Values: mark Save….Export XML File: click Browse….various folders in your PC come up….in "File Name"of the appropriate folder enter "exportarima"….click Save….click Continue….click OK.

Date
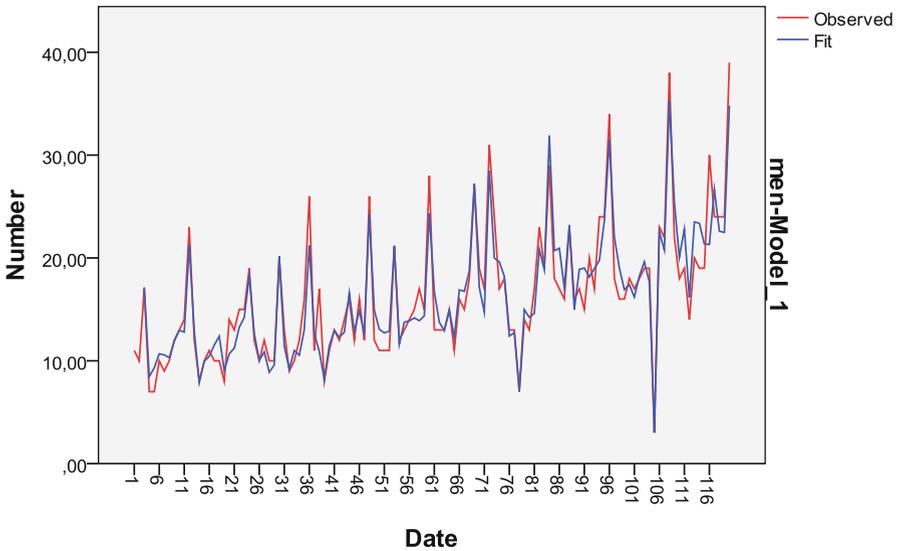
The above graph shows that a good fit of the observed data is given by the ARIMA model, and that an adequate predictive model is provided. The upward trend is in agreement with beta-cell failure after several years.

The underneath table shows that 3 significant predictors have been identified. Also the goodness of fit of the ARIMA (p, d, q) model is given, where p=number of lags, d=the trend (one upward trend means d=1), and q=number of moving averages (=0 here). Both Stationary R square, and Ljung-Box tests are insignificant. A significant test would have meant poor fit. In our example, there is an adequate fit, but the model has identified no less than 7 outliers. Phone visits, nurse visits, and doctor visits were significant predictors at p<0.0001, while self control and educational patient meetings were not so. All of the outliers are significantly more distant from the ARIMA model than could happen by chance. All of the p-values were very significant with p<0.001 and<0.0001.

Model statistics

Model	Number of predictors	Model Fit statistics	Ljung-BoxQ(18)			Number of outliers
		Stationary R-squared	Statistics	DF	Sig.	
men-Model_1	3	,898	17,761	18	,471	7

ARIMA model parameters

					Estimate	SE	t	Sig.
men-Model_1	men	Natural log	Constant		−2,828	,456	−6,207	,000
	phone	Natural log	Numerator	Lag 0	,569	,064	8,909	,000
	nurse	Natural log	Numerator	Lag 0	1,244	,118	10,585	,000
	doctor	Natural log	Numerator	Lag 0	,310	,077	4,046	,000
				Lag 1	−,257	,116	−2,210	,029
				Lag 2	−,196	,121	−1,616	,109
			Denominator	Lag 1	,190	,304	,623	,535

Example   215

Outliers

			Estimate	SE	t	Sig.
men-Model_1	3	Additive	,769	,137	5,620	,000
	30	Additive	,578	,138	4,198	,000
	53	Additive	,439	,135	3,266	,001
	69	Additive	,463	,135	3,439	,001
	78	Additive	−,799	,138	−5,782	,000
	88	Additive	,591	,134	4,409	,000
	105	Additive	−1,771	,134	−13,190	,000

When returning to the data view screen, we will observe that SPSS has added HbA1 values (except for the first two dates due to lack of information) as a novel variable. The predicted values are pretty similar to the measured values, supporting the adequacy of the model.

We will now apply the XML file and the Apply Models modus for making predictions about HbA1 values in the next 6 months, assuming that the significant variables nurse, doctor, phone are kept constant at their overall means.

First add the underneath data to the original data file and rename the file, e.g., "arimafile2", and store it at an appropriate folder in your computer.

Date	HbA1	nurse	doctor	phone	self	meeting
01/01/1999	10,00	8,00	4,00			
01/02/1999	10,00	8,00	4,00			
01/03/1999	10,00	8,00	4,00			
01/04/1999	10,00	8,00	4,00			
01/05/1999	10,00	8,00	4,00			
01/06/1999	10,00	8,00	4,00			

**Then open "arimafile2.sav" and command:**

Analyze….click Apply Models….click Reestimate from data….click First case after end of estimation period through a specified date….Observation: enter 01/06/1999….click Statistics: click Display Forecasts….click Save: Predicted Values mark Save….click OK.

The underneath table shows the predicted HbA1 values for the next 6 months and their upper and lower confidence limits (UCL and LCL).

Forecast

Model		121	122	123	124	125	126
HbA1-Model_1	Forecast	17,69	17,30	16,49	16,34	16,31	16,30
	UCL	22,79	22,28	21,24	21,05	21,01	21,00
	LCL	13,49	13,19	12,58	12,46	12,44	12,44

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier
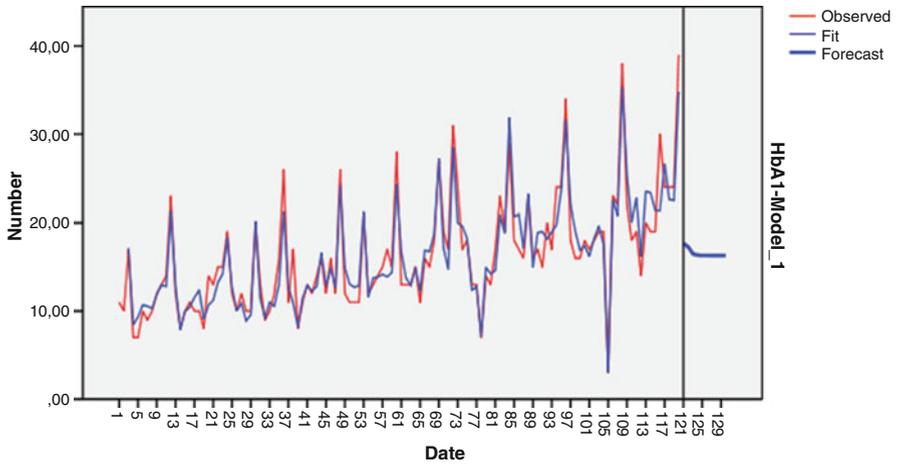
Also a graph of the HbA1 pattern after the estimation period is given as shown in the above graph. When returning to the data view of the arimafile2, we will observe that SPSS has added the predicted values as a novel variable.

Date	HbA1	nurse	doctor	phone	self	meeting	modeled HbA1	predicted HbA1
07/01/1998	19,00	11,00	8,00	5,00	28,00	4,00	21,35	21,35
08/01/1998	30,00	12,00	9,00	4,00	27,00	5,00	21,31	21,31
09/01/1998	24,00	13,00	8,00	5,00	30,00	5,00	26,65	26,65
10/01/1998	24,00	12,00	10,00	4,00	28,00	6,00	22,59	22,59
11/01/1998	24,00	11,00	8,00	5,00	26,00	5,00	22,49	22,49
12/01/1998	39,00	15,00	10,00	5,00	37,00	7,00	34,81	34,81
01/01/1999		10,00	8,00	4,00				17,69
01/02/1999		10,00	8,00	4,00				17,30
01/03/1999		10,00	8,00	4,00				16,49
01/04/1999		10,00	8,00	4,00				16,34
01/05/1999		10,00	8,00	4,00				16,31
01/06/1999		10,00	8,00	4,00				16,30

modeled HbA1 = calculated HbA1 values from the above arima model
Predicted HbA1 = the predicted HbA1 values using the XML file for future dates.

Conclusion

Autoregressive integrated moving average methods are appropriate for assessing trends, seasonality, and change points in a time series. In the example given no conclusion can be drawn about individual patients. Autoregressive models can,

however, also be applied for data sets of individual patients. Also as a multivariate methodology it is appropriate for multiple instead of a single outcome variable like various health outcomes.

Note

More background theoretical and mathematical information of autoregressive models for longitudinal data is in Machine learning in medicine part two, Multivariate analysis of time series, pp 139–154, Springer Heidelberg Germany, 2013, from the same authors.