

Jason & Glenwick, eds.

Methodological Approaches to Community-Based Research

Chapter 8/Web Appendix guide

Web Appendix Guide for Chapter 8/Time Series Analysis

The Web Appendix materials for the chapter on time series analysis (TSA) include data sets, syntax files, and annotated SAS output (included in this guide). Only one data set and one syntax file are necessary to conduct the analyses described in the chapter. The other data set and corresponding syntax file are provided to showcase some of the necessary data management steps a researcher may need to perform in order to transform the traditional participant-level data to a clinic-level biweekly data set. Both syntax files include notes to explain the purpose of each syntax step.

- **Datasets (in SAS and SPSS)**
 - Participant-level data
 - TSA_Participants.sav
 - TSA_Participants.sas7bdat
 - Clinic-level data
 - TSA_Clinic.sav
 - TSA_Clinic.sas7bdat
- **SAS syntax files**
 - TSA Chapter – Participant to Clinic Scores.sas

This syntax:

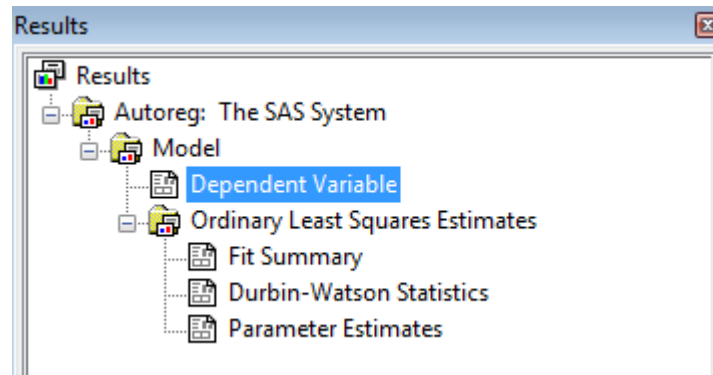
 - Imports the participant-level data from the SPSS file (there are sometimes problems with format labels in SAS when the data were saved using a 64-bit installation and then opened with a 32-bit installation, so importing SPSS format data is a safer way)
 - Calculates clinic-level average biweekly scores
 - Accounts for biweekly intervals without any interviews
 - Creates dummy variables for testing intervention effects
 - Labels all variables
 - Saves final dataset as a SAS (i.e., TSA_Clinic.sas7bdat) or SPSS (i.e., TSA_Clinic.sav) data set
 - TSA Chapter – Autoregressive Analyses.sas

This file provides the SAS syntax necessary to fit the autoregressive model discussed in the chapter, using the clinic-level data.
- **Annotated SAS Output (below)**

This document provides screenshots of the SAS output generated by the syntax file “TSA Chapter – Autoregressive Analyses.sas” and provides explanations of how the output is organized and how the output would be interpreted.

Annotated SAS Output

In step 1 of the autoregressive regression analysis, we are fitting a basic linear regression model while testing up to five lags for statistically significant autocorrelations using the Durbin-Watson Statistic. SAS reminds us that up to this point we have not actually modeled any autoregressive terms by grouping all of our output under the “Ordinary Least Square Estimation” header in the SAS output window:



Below is the output we obtain:

The AUTOREG Procedure

Dependent Variable kqscore
KQ Score

Ordinary Least Squares Estimates

SSE	24.6649745	DFE	36
MSE	0.68514	Root MSE	0.82773
SBC	98.6908264	AIC	95.415654
MAE	0.60171082	AICC	95.7585112
MAPE	7.90800146	Regress R-Square	0.1063
		Total R-Square	0.1063

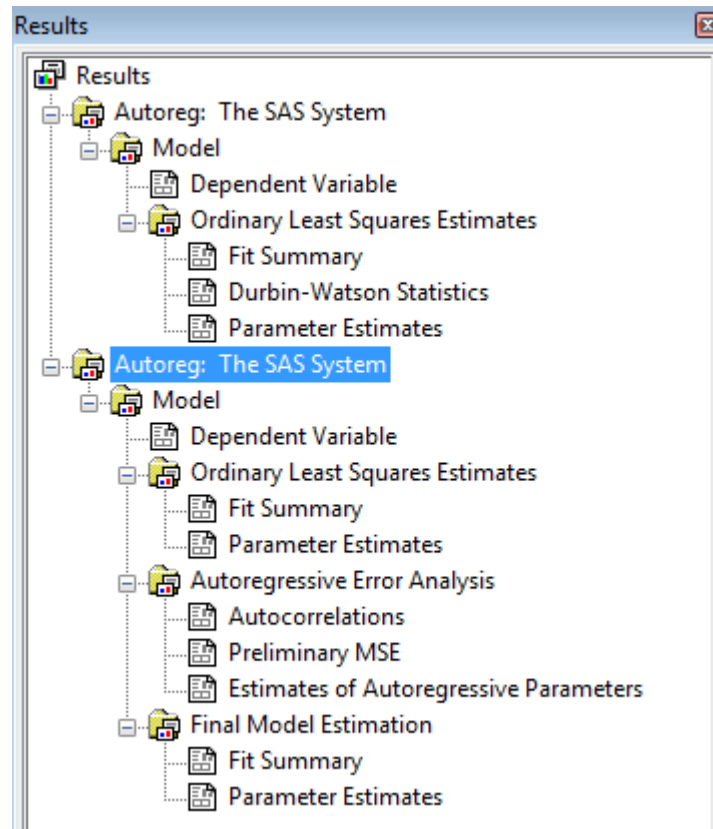
Durbin-Watson Statistics

Order	DW	Pr < DW	Pr > DW
1	2.3254	0.8043	0.1957
2	1.2888	0.0120	0.9880
3	2.2803	0.8630	0.1370
4	2.0656	0.7291	0.2709
5	2.0101	0.7350	0.2650

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t	Variable Label
Intercept	1	7.2606	0.2926	24.81	<.0001	
intv	1	0.6814	0.3294	2.07	0.0458	intv

We find that lag 2 is potentially positively autocorrelated (highlighted here by a rectangular box). Thus, in Step 2, we are fitting a model that includes autoregressive terms up to lag 2. Note that this time, the AUTOREG procedure in SAS produces three different output headers:



As before, the ordinary least squares model is fit. Additionally, the new model is fit, which includes both the parameters of primary interest, the intervention effect, as shown under “Final Model Estimation” and the requested autoregressive parameters, lags 1 and 2, as shown under “Autoregressive Error Analysis”. First, we will take a look at parameter estimates of the autoregressive terms.

Estimates of Autocorrelations																								
Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
0	0.6491	1.000000												*****										
1	-0.1149	-0.177043										****		*****										
2	0.1189	0.183189												****										
Preliminary MSE									0.6133															
Estimates of Autoregressive Parameters																								
Lag	Coefficient	Standard Error	t Value																					
1	0.149290	0.169378	0.88																					
2	-0.156758	0.169378	-0.93																					
Yule-Walker Estimates																								
SSE	23.340395	DFE	34																					
MSE	0.68648	Root MSE	0.82854																					
SBC	104.520548	AIC	97.9702029																					
MAE	0.58537607	AICC	99.1823242																					
MAPE	7.72778938	Regress R-Square	0.1148																					
Durbin-Watson	1.8769	Total R-Square	0.1543																					

Here we note that indeed parameter estimates for lag 1 and 2 were fit. We also note, however, that after accounting for the small lag 1 autocorrelation (-0.177), the lag 2 autocorrelation (0.183) is no longer statistically significant. Thus, we do not actually need to include these terms in our regression model. If we did want to include them, the final parameter estimates, while including the lag 1 and 2 parameter estimates are shown under the header “Final Model Estimation”:

The AUTOREG Procedure						
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t	Variable Label
Intercept	1	7.2841	0.2929	24.87	<.0001	
intv	1	0.6895	0.3283	2.10	0.0432	intv

Note that these parameter estimates are very similar to the previously obtained ordinary least squares solution. Thus, the decision as to whether or not to model the underlying autocorrelations does not meaningfully impact the overall finding that clinic-wide Hepatitis C Knowledge scores were on average higher after the interventions started compared to before the intervention started.