

THE USE OF PROC ARIMA
TO TEST DIFFERENT INTERVENTION EFFECTS

Neil S. Fleming, Ph.D., Sr. Vice-President, 2W Systems Co., Inc.
5430 Alpha Rd., Suite 305, Dallas, TX 75240 (972) 733-0588

Eric Gibson, M.D., Dept. of Pediatrics, Thomas Jefferson University

David G. Fleming, M.D., Ph.D., Case-Western Reserve University

INTRODUCTION. In April 1992, the American Academy of Pediatrics (AAP) recommended that health infants be put to sleep on their side or back to reduce the risk of Sudden Infant Death Syndrome (SIDS). We hypothesized two different forms of the intervention to examine the impact as 1) gradually increasing and permanent or 2) abrupt, temporary and gradually declining. These were all tested using the Box and Tiao time-series intervention methodology based on the Box-Jenkins autoregressive, integrated, moving average (ARIMA) model that considers the time-dependence in the data. These hypotheses were tested on SIDS rates for white and non-white populations in two major metropolitan areas.

This paper presents the SAS programming code for both the Data Step and statistical PROC along with the models estimated, including output statistics from SAS' Proc ARIMA to test the hypotheses. Rates are normalized by the annual birth rate for the non-white population in one of the cities to mitigate changes due to births. The Data Step is presented to prepare 32 quarters of SIDS data to compare rates prior to the first half of 1992 versus subsequent quarters through 1994.

The models are examined using PROC ARIMA to test the form of the intervention effect subsequent to the AAP recommendations. The models also include the seasonality effect, i.e., the increase during the winter quarter for the first three months of the year; and the autoregressive/moving average components of the error structure.

Discussion includes the theoretical ARIMA models pertaining to the two hypotheses, including the intervention specifications (step or pulse functions). Also presented are the results with interpretation and testing of the model parameters in relation to their standard errors. Critical to the testing of the two hypotheses is the estimation of the denominator term that influences the nature of the intervention effect (step or pulse).

Since the model is analyzed in terms of the natural logarithm of the SIDS rates, results are interpreted through percentage changes caused by the intervention and seasonality. This examination demonstrates how time-series intervention analysis can be used to quantify the impact of public policy recommendations in quasi-experimental designs.

STATISTICAL MODEL. The ARIMA model used for this analysis in general and for the abrupt, permanent hypothesis is:

$$Z_t = \mu + w_{is}S_t + w_sS_t + (\theta_q(\beta)/\phi_p(\beta)) a_t, \text{ wherein:}$$

Z_t is the natural log of SIDS deaths per 10,000 annual live births; e^μ is the average ratio of deaths/10,000 annual live births in non-winter quarters prior to the intervention; w_{is} is an abrupt, permanent effect of the intervention upon the SIDS rate after the intervention, representing a step function. IS_t is a dummy code with a value of 0 for all 22 quarters prior to the intervention, and 1 for 10 subsequent quarters; W_s is the shift from winter to non-winter quarters; S_t is the dummy code for seasonality, winter (1), versus non-winter(0).

The final term in the equation represents the ARIMA noise process: θ_q is the coefficient of the noise model moving average; ϕ_p is the coefficient of the noise model autoregressive term; β is the back shift operator; and a_t is the stationary white noise process.

To test a gradually increasing, permanent intervention, a denominator is added to the $w_{is}I_t$ term $(1-\delta B)$. This denominator represents the rate of increase in the decline in SIDS deaths. Under a stable, stationary process, δ is constrained between 0 and 1. As δ increases, the intervention effect is more gradual rather than abrupt. As δ approaches zero, the intervention is identical to an abrupt, permanent effect.

To test a abrupt, temporary and gradually declining intervention, a denominator is added to the intervention term $(1-\delta B)$. However, the intervention term is respecified (IP_t) to represent a pulse (one-time only) effect. The estimated coefficient w_{ip} represents the abrupt effect in the first quarter after the intervention representing a pulse function; IP_t is the dummy code with 0 for all quarters except 1 for the first quarter after the intervention.

SAS PROGRAM CODE. NOTE THAT DATA HAVE BEEN REPLACED BY 'X' BECAUSE PUBLICATION IS PENDING FOR RESEARCH. The Data Step was the following:

```

LIBNAME RESID 'C:\SCSUG\RESID';
DATA RESID.SIDS2;
INPUT DATE WHITE;
INFORMAT DATE YYQ4.;
IF DATE GT '30JUN92'D THEN INTER=1; ELSE INTER=0;
IF DATE GT '30JUN92'D AND DATE LT '01OCT92'D THEN PULSE=1; ELSE PULSE=0;
IF WHITE=0 THEN LWHITE=0; ELSE LWHITE=LOG(WHITE);
IF QTR(DATE)=1 THEN SEASON=1; ELSE SEASON=0;
IF YEAR(DATE)=1987 THEN WBIRTH=12589;
ELSE IF YEAR(DATE)=1988 THEN WBIRTH=X;
ELSE IF YEAR(DATE)=1989 THEN WBIRTH=X;
ELSE IF YEAR(DATE)=1990 THEN WBIRTH=X;
ELSE IF YEAR(DATE)=1991 THEN WBIRTH=X;
ELSE IF YEAR(DATE)=1992 THEN WBIRTH=X;
ELSE IF YEAR(DATE)=1993 THEN WBIRTH=X;
ELSE IF YEAR(DATE)=1994 THEN WBIRTH=X;
IF WHITE=0 THEN LWRATIO=LOG(1/WBIRTH*10000);
ELSE LWRATIO=LOG(WHITE/WBIRTH*10000);
CARDS;
      87Q1   X
      87Q2   X
      87Q3   X
      87Q4   X
      88Q1   X
      .
      .
      .
      94Q1   X
      94Q2   X
      94Q3   X
      94Q4   X
RUN;

```

The following was the programming code to test the hypothesis of a gradually increasing and permanent effect . The was also presented without and with the ARIMA noise terms in the model.

```

OPTIONS PAGENO=1;
%LET DEPVAR=LWRATIO;
PROC ARIMA DATA=RESID.SIDS2;
IDENTIFY VAR=&DEPVAR(0) CROSSCOR=(INTER(0) SEASON(0)) NLAG=4 NOPRINT;
ESTIMATE INPUT=/(1)INTER SEASON) ML;
TITLE1 "ANALYSIS OF PHILLY LWRATIO FOR LOG OF WHITE DEATHS PER 10,000";
TITLE2 "ANALYSIS OF INCREASING IMPACT WITHOUT ANY AR/MA TERMS";

```

RUN;

```
PROC ARIMA DATA=RESID.SIDS2;  
IDENTIFY VAR=&DEPVAR(0) CROSSCOR=(INTER(0) SEASON(0)) NLAG=4 NOPRINT;  
ESTIMATE P=(7) Q=(3) INPUT=/(1)INTER SEASON) ML;  
TITLE2 "ANALYSIS OF INCREASING IMPACT WITH ANY AR/MA TERMS";  
RUN;
```

The following was the programming code to test the hypothesis of an abrupt, temporary and gradually declining effect . The was also presented without and with the ARIMA noise terms in the model.

```
OPTIONS PAGENO=1;  
%LET DEPVAR=LWRATIO;  
PROC ARIMA DATA=RESID.SIDS2;  
IDENTIFY VAR=&DEPVAR(0) CROSSCOR=(PULSE(0) SEASON(0)) NLAG=4 NOPRINT;  
ESTIMATE INPUT=/(1)PULSE SEASON) ML;  
TITLE1 "ANALYSIS OF PHILLY LWRATIO FOR LOG OF WHITE DEATHS PER 10,000";  
TITLE2 "ANALYSIS OF DECLINING IMPACT WITHOUT ANY AR/MA TERMS";  
RUN;
```

```
PROC ARIMA DATA=RESID.SIDS2;  
IDENTIFY VAR=&DEPVAR(0) CROSSCOR=(PULSE(0) SEASON(0)) NLAG=4 NOPRINT;  
ESTIMATE P=(3)(7) INPUT=/(1)PULSE SEASON) ML;  
TITLE2 "ANALYSIS OF DECLINING IMPACT WITH ANY AR/MA TERMS";  
RUN;
```

The following are the the results for the analysis:

ARIMA Procedure

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	0.71060	0.14089	5.04	0	LWRATIO	0
NUM1	-0.60158	0.95273	-1.71	0	INTER	0
DEN1,1	-0.85104	0.27987	-3.04	1	INTER	0
NUM2	0.41443	0.26405	1.57	0	SEASON	0

Constant Estimate = 0.71060404

Variance Estimate = 0.34482716

Std Error Estimate = 0.58721986

AIC = 58.6854575

SBC = 64.4214063

Number of Residuals= 31

Correlations of the Estimates

Variable	Parameter	LWRATIO MU	INTER NUM1	INTER DEN1,1	SEASON NUM2
LWRATIO	MU	1.000	-0.458	0.323	-0.446
INTER	NUM1	-0.458	1.000	-0.152	-0.034
INTER	DEN1,1	0.323	-0.152	1.000	-0.285
SEASON	NUM2	-0.446	-0.034	-0.285	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	4.06	6	0.669	-0.221	-0.124	-0.146	0.155	0.044	0.047
12	8.48	12	0.747	-0.283	0.044	-0.109	-0.035	0.069	0.035
18	16.72	18	0.543	-0.216	0.285	0.022	-0.081	-0.045	0.029
24	23.61	24	0.484	0.192	-0.028	-0.152	-0.025	0.092	-0.034

Model for variable LWRATIO

Estimated Intercept = 0.71060404

Input Number 1 is INTER.

Overall Regression Factor = -0.60158

The Denominator Factors are

Factor 1: $1 + 0.85104 B^{**}(1)$

Input Number 2 is SEASON.

Overall Regression Factor = 0.41443

ARIMA Procedure

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	0.61396	0.06440	9.53	0	LWRATIO	0
MA1,1	0.59393	0.19037	3.12	3	LWRATIO	0
AR1,1	-0.71019	0.12982	-5.47	7	LWRATIO	0
NUM1	-0.62114	0.21017	-2.96	0	INTER	0
DEN1,1	-0.60137	0.50584	-1.19	1	INTER	0
NUM2	0.95304	0.21045	4.53	0	SEASON	0

Constant Estimate = 1.04998902

Variance Estimate = 0.20618424

Std Error Estimate = 0.45407515

AIC = 50.5942868

SBC = 59.19821

Number of Residuals= 31

Correlations of the Estimates

Variable	Parameter	LWRATIO MU	LWRATIO MA1,1	LWRATIO AR1,1	INTER NUM1	INTER DEN1,1	SEASON NUM2
LWRATIO	MU	1.000	-0.211	0.026	-0.232	0.068	-0.852
LWRATIO	MA1,1	-0.211	1.000	-0.414	-0.003	-0.003	0.233
LWRATIO	AR1,1	0.026	-0.414	1.000	0.080	-0.040	-0.083
INTER	NUM1	-0.232	-0.003	0.080	1.000	0.726	0.020
INTER	DEN1,1	0.068	-0.003	-0.040	0.726	1.000	-0.022
SEASON	NUM2	-0.852	0.233	-0.083	0.020	-0.022	1.000

Autocorrelation Check of Residuals

To	Chi	Autocorrelations							
Lag	Square	DF	Prob						
6	6.55	4	0.162	-0.380	-0.058	-0.010	0.085	0.041	-0.169
12	9.13	10	0.519	0.194	-0.059	-0.069	-0.075	0.032	0.072
18	20.39	16	0.203	-0.255	0.331	-0.020	-0.106	-0.017	0.047
24	28.55	22	0.158	0.158	-0.216	0.100	-0.011	0.010	-0.067

Model for variable LWRATIO

Estimated Intercept = 0.61395887

Autoregressive Factors

Factor 1: 1 + 0.71019 B**(7)

Moving Average Factors

Factor 1: 1 - 0.59393 B**(3)

Input Number 1 is INTER.

Overall Regression Factor = -0.62114

The Denominator Factors are

Factor 1: 1 + 0.60137 B**(1)

ARIMA Procedure

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	0.62146	0.12620	4.92	0	LWRATIO	0
NUM1	-0.74584	0.58371	-1.28	0	PULSE	0
DEN1,1	-0.51911	0.53473	-0.97	1	PULSE	0
NUM2	0.39789	0.27268	1.46	0	SEASON	0

Constant Estimate = 0.62145832

Variance Estimate = 0.35502907

Std Error Estimate = 0.59584316

AIC = 59.5893048

SBC = 65.3252536

Number of Residuals= 31

Correlations of the Estimates

Variable	Parameter	LWRATIO MU	PULSE NUM1	PULSE DEN1,1	SEASON NUM2
LWRATIO	MU	1.000	-0.179	0.259	-0.515
PULSE	NUM1	-0.179	1.000	-0.478	0.120
PULSE	DEN1,1	0.259	-0.478	1.000	-0.341
SEASON	NUM2	-0.515	0.120	-0.341	1.000

Autocorrelation Check of Residuals

To	Chi	Autocorrelations							
Lag	Square	DF	Prob						
6	4.07	6	0.667	-0.150	-0.127	-0.084	0.127	0.141	0.163
12	8.62	12	0.735	-0.278	0.085	-0.117	-0.057	0.051	-0.005
18	16.29	18	0.572	-0.249	0.203	-0.054	-0.127	-0.072	-0.012
24	23.00	24	0.520	0.165	-0.007	-0.178	-0.012	0.094	-0.007

Model for variable LWRATIO

Estimated Intercept = 0.62145832

Input Number 1 is PULSE.

Overall Regression Factor = -0.74584

The Denominator Factors are

Factor 1: $1 + 0.51911 B^{**}(1)$

Input Number 2 is SEASON.

Overall Regression Factor = 0.397889

ARIMA Procedure

Maximum Likelihood Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag	Variable	Shift
MU	0.62985	0.07811	8.06	0	LWRATIO	0
AR1,1	-0.38530	0.17722	-2.17	3	LWRATIO	0
AR2,1	-0.58700	0.15630	-3.76	7	LWRATIO	0
NUM1	-0.97564	0.40610	-2.40	0	PULSE	0
DEN1,1	0.71996	0.16252	4.43	1	PULSE	0
NUM2	0.88629	0.21070	4.21	0	SEASON	0

Constant Estimate = 1.38470651

Variance Estimate = 0.23840822

Std Error Estimate = 0.48827065

AIC = 52.2986937

SBC = 60.9026169

Number of Residuals= 31

Correlations of the Estimates

Variable	Parameter	LWRATIO MU	LWRATIO AR1,1	LWRATIO AR2,1	PULSE NUM1	PULSE DEN1,1	SEASON NUM2
LWRATIO	MU	1.000	0.108	0.050	-0.361	0.029	-0.703
LWRATIO	AR1,1	0.108	1.000	-0.001	0.009	0.073	-0.139
LWRATIO	AR2,1	0.050	-0.001	1.000	-0.021	-0.029	-0.089
PULSE	NUM1	-0.361	0.009	-0.021	1.000	0.689	0.249
PULSE	DEN1,1	0.029	0.073	-0.029	0.689	1.000	0.203
SEASON	NUM2	-0.703	-0.139	-0.089	0.249	0.203	1.000

Autocorrelation Check of Residuals

To	Chi	Autocorrelations							
Lag	Square	DF	Prob						
6	5.34	4	0.254	-0.306	-0.062	-0.118	0.101	0.084	-0.143
12	6.61	10	0.762	0.147	-0.074	-0.017	-0.026	0.032	-0.013
18	17.38	16	0.361	-0.237	0.317	-0.091	-0.110	-0.040	0.030
24	23.34	22	0.383	0.107	-0.193	0.096	-0.012	0.047	-0.045

Model for variable LWRATIO

Estimated Intercept = 0.62984917

Autoregressive Factors

Factor 1: 1 + 0.3853 B**(3)

Factor 2: 1 + 0.587 B**(7)

Input Number 1 is PULSE.

Overall Regression Factor = -0.97564

The Denominator Factors are

Factor 1: 1 - 0.71996 B**(1)

Input Number 2 is SEASON.