PREDICTING PRISON POPULATION USING THE SAS/ETS™ PRODUCT*

Steven Davis, Oklahoma Department of Corrections

Introduction. Some of the earliest applications of Box-Jenkins time series analyses in the social sciences were concerned with law enforcement or criminal justice data, for example, analysis of data on the Connecticut crackdown on speeding (Glass, 1968). Since that time, many of the reported time series analyses of criminal justice data have continued to center on evaluation of the impact of innovations such as speed limit laws (Johnson, Klein and Levy, 1976) or gun control laws (Deutsch and Alt, 1977). However, events of recent years have increased awareness and application of the forecasting capabilities of the time series analysis. In particular, increasing prison populations, court decisions regarding prisoner housing, and decreased appropriations have increased corrections planners' need for projections of prison population. In Figure 1, the sharp increase since 1979 demonstrates that Oklahoma's prison population has followed the national pattern. A project that was started to use SAS/GRAPH™ programs* to plot this rapid population change for descriptive purposes led to the use of the SAS/ETS product to predict the population's future course. The initial findings of those prediction analyses are presented here.

A Criminal Justice System Model. The procedure for this study loosely followed the guidelines for developing a forecasting system suggested by Jenkins (1979). Part of that process involves developing a conceptual model of the mechanisms influencing the variable of interest, in this case, prison population. In Figure 2 is the model conceived for this study. The arrows show lines of influence and the boxes represent potential data sources in the Oklahoma criminal justice system. Although this model may be amended as links between variables are supported or rejected, it provided the focus for the initial variable selection process.

Data Selection. Initially, three sources of data were identified. Reported offenses and arrests were available from the state Uniform Crime Reporting program. Reports on charges filed by district attorneys and the outcomes of those cases were available from the Arrest Disposition Reporting System. Both programs are part of the Oklahoma State Bureau of Investigation. Prison receptions, discharges and paroles, as well as, total population, are recorded regularly by the Oklahoma Department of Corrections. Because of their immediate availability, and their direct impact on prison population, receptions, paroles and discharges were selected as the first input variables to model. Later, reported offenses, arrests, charges filed, and unemployment were also compiled for inclusion in the analysis. All data were aggregated by month.

Identification of Time Series Models. The input and output variables were first plotted using SAS/GRAPH™ PROC GPLOT (see Figure 3). Next a univariate model was identified for the output series prison population. After examining the plots of the raw data (Figure 3) and sample autocorrelation function (Figure 4), it was clear that there was a trend in the data and a first difference was required. Parameters for several possible models were then estimated for the differenced series and the residual autocorrelations examined. An autoregressive model was found to be the most parsimonious representation of the data:

$$(1 - .310075*B^{4} - .313851*B^{8})\ (1 - B)*PRISPOP = 45.9945 + a_{t}.$$

Univariate models were then identified for each of the input variables to be used in the pre-whitening process:

$$DISCH = 112.869 + (.5635*B)^{12} + .67458*B^{24})*a_{t}$$

$$(1 - B^{12})*RECPT = 18.5288 - .802254*B^{12}*a_{t}$$

$$PAROLE = 83.25 + a_{t}.$$

Even though the PAROLE time series was found to be white noise, it was included in the modelling process as a possible explanatory variable for some of the random behavior in the population series.

Following Box and Jenkins (1976), the output series was next pre-whitened using the univariate model for each input series (the pre-whitening process eliminates spurious cross-correlations between two series which may occur because of autocorrelation in one or both series). The cross-correlation between the pre-whitened output series and each corresponding input series was calculated. Based on those cross-correlation functions, the moving average and autoregressive operators and delay parameter were tentatively identified and estimated for the transfer function between each input series and the output series. Since changes in each input variable were contemporaneous with changes in the output variable, the delay parameter was zero in all cases.

The transfer functions were combined into a single model and parameters estimated. The residuals were examined to help identify the noise portion of the model and then the parameters were re-estimated. No significant autocorrelations were found among the residuals and the following model was accepted:

*SAS/ETS and SAS/GRAPH are trademarks of SAS Institute, Inc.

$$PRISPOP = 4151.57 + (-.0478 - .4441*B)*PAROLE + (.3288 + .7231*B)*RECPT + (-.5183 - .6151*B - .2595*B^2)*DISCH + (1 - 1.0549*B)^{-1}*a_t.$$

Using the SAS/ETS™ ARIMA procedure, forecasts were calculated through the 12 months of 1983 based on the univariate and transfer function models (Tables 1 and 2, respectively). When these forecasts were compared to the actual prison population data available for the first four months of 1983, the transfer function model provided the best fit (see Figure 6).

Conclusions. A transfer function model of a prison population time series with receptions, paroles, and discharges as input series was found to produce more accurate short-term forecasts of prison population than the univariate model relying on the population history alone. Compilation of population data is continuing and the longer-term accuracy of predictions by both models will be monitored. Time series data for unemployment, offenses and arrests, and charges filed have now been compiled and models are being identified. As adequate univariate time series models are developed for these variables, their contributions to the larger transfer function model will also be investigated.

REFERENCES

Box, G.E.P. and Jenkins, G.M. Time Series Analysis: Forecasting and Control, Revised Edition. San Francisco: Holden-Day, 1976.

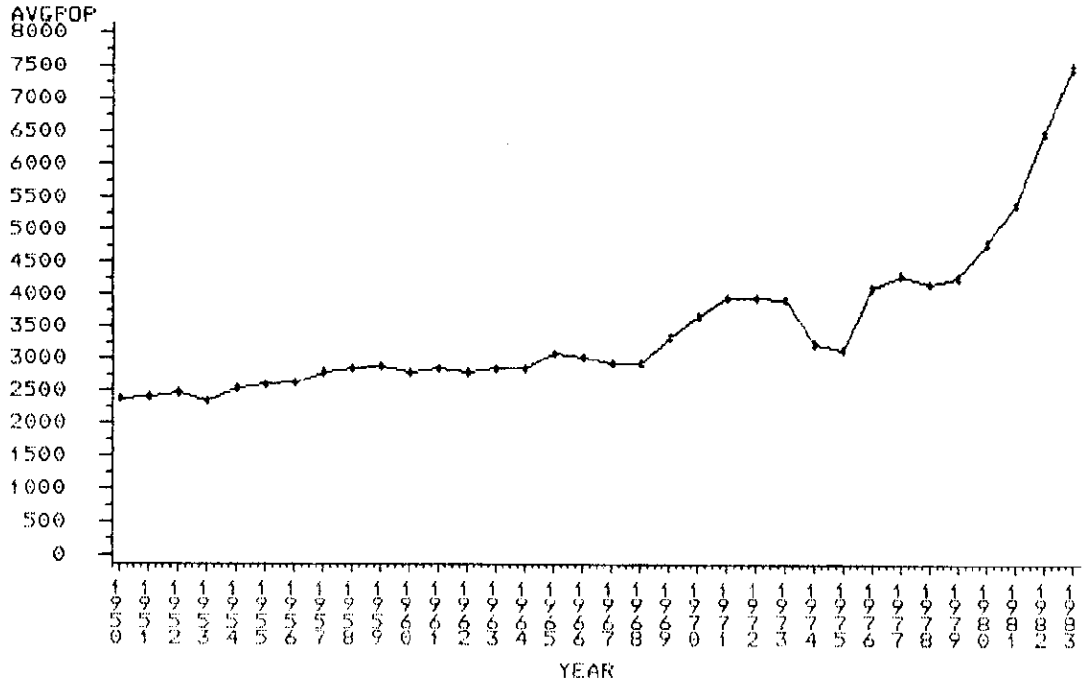
Deutsch, S.J. and Alt, F.B. The effect of Massachusetts' gun control law on gun-related crimes in the city of Boston. Evaluation Quarterly, 1, 543-568, 1977.

Glass, G.V. Analysis of data on the Connecticut speeding crackdown as a time series quasi-experiment. Law and Society Review, 3, 55-76, 1968.

Jenkins, G.M. Practical Experiences with Modelling and Forecasting Time Series. St. Helier, UK: Gwilym Jenkins and Partners, 1979.

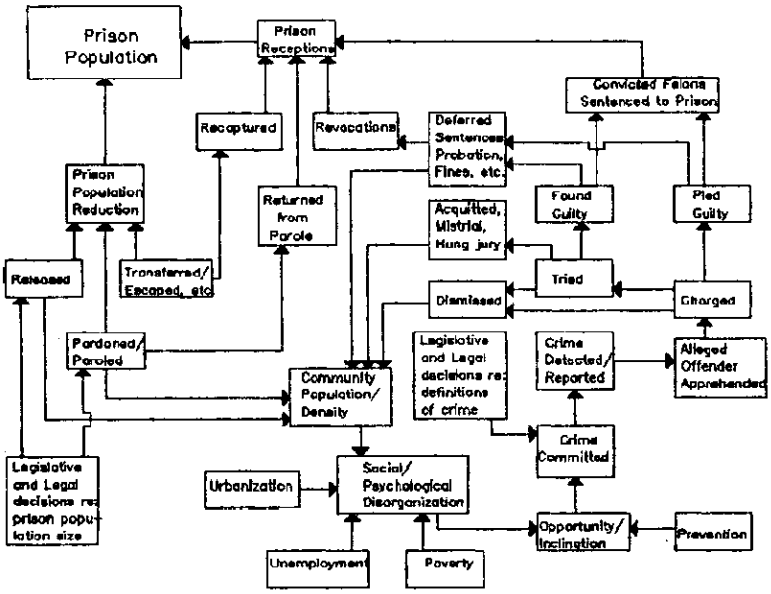
Johnson, D.M., Klein, T.M., and Levy, P.S. Effect of the 55 MPH speed limit law on fatal crashes in Texas. NHTSA Technical Report, U.S. Department of Transportation, July 1976.

Oklahoma Department of Corrections
Prison Population
1950–1983



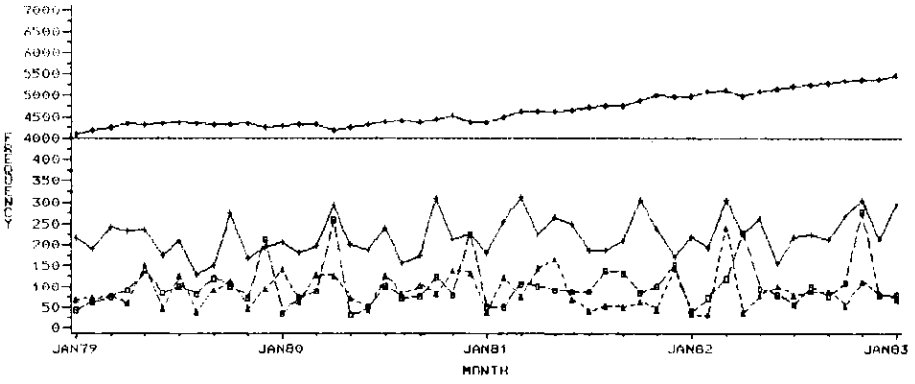
SOURCE: Oklahoma Department of Corrections Planning and Research

Prison Population Model



Available data sources: ADRS, DOC, UCR

Oklahoma Department of Corrections
Prison Population Variables



Top line is total population. Dashed line with triangle is receptions.
Dashed line with square is discharges. Solid line with plus is parolees.

Data source: Oklahoma Department of Corrections, Jan 1979–Dec 1982

SAS
ARIMA PROCEDURE

NAME OF VARIABLE = PRISPOP

MEAN OF WORKING SERIES= 4904.69
STANDARD DEVIATION = 662.335
NUMBER OF OBSERVATIONS= 60

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	STD
0	438688	1.00000	|********************|	0
1	409944	0.93448	|******************* . |	0.129099
2	378680	0.86321	|*****************. |	0.213951
3	349582	0.79688	|****************. |	0.26573
4	325159	0.74121	|***************. |	0.302952
5	299750	0.68329	|**************. |	0.331803
6	275327	0.62761	|*************. |	0.35448
7	252186	0.57486	|***********. |	0.37254
8	231233	0.52710	|**********. |	0.387042
9	211166	0.48136	|*********. |	0.398826
10	193018	0.43999	|*********. |	0.408395

'.' MARKS TWO STANDARD ERRORS

INVERSE AUTOCORRELATIONS

LAG	CORRELATION	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	-0.53743	|**********. |
2	0.01417	| . |
3	0.07223	| . |* . |
4	-0.06933	| . *| . |
5	0.02002	| . | . |
6	-0.00398	| . | . |
7	0.01411	| . | . |
8	-0.01879	| . | . |
9	0.01364	| . | . |
10	-0.00332	| . | . |

PARTIAL AUTOCORRELATIONS

LAG	CORRELATION	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	0.93448	| . |****************** |
2	-0.07919	| . **| . |
3	0.00281	| . | . |
4	0.04424	| . |* . |
5	-0.05609	| . *| . |
6	-0.00929	| . | . |
7	-0.00805	| . | . |
8	0.00175	| . | . |
9	-0.01417	| . | . |
10	0.00677	| . | . |

SAS
ARIMA PROCEDURE

NAME OF VARIABLE = PRISPOP
PERIODS OF DIFFERENCING= 1.

MEAN OF WORKING SERIES= 39.8525
STANDARD DEVIATION = 74.1721
NUMBER OF OBSERVATIONS= 59

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	STD
0	5501.5	1.00000	|********************|	0
1	421.143	0.07655	| . |** . |	0.130189
2	-1230.12	-0.22360	| .****| . |	0.13095
3	804.121	0.14616	| . |*** . |	0.137268
4	2086.26	0.37922	| . |******** |	0.139881
5	114.332	0.02078	| . | . |	0.156338
6	-694.464	-0.12623	| . ***| . |	0.156385
7	-140.562	-0.02555	| . *| . |	0.158102
8	1836.22	0.33377	| . |******* |	0.158172
9	444.498	0.08080	| . |** . |	0.16969
10	-1703.44	-0.30963	| .******| . |	0.170341
11	214.199	0.03893	| . |* . |	0.179627
12	1208.7	0.21970	| . |**** . |	0.17977
13	-834.054	-0.15160	| . ***| . |	0.184265
14	-1051.01	-0.19104	| . ****| . |	0.186367
15	871.834	0.15847	| . |*** . |	0.189657
16	1180.35	0.21455	| . |**** . |	0.191888
17	-495.672	-0.09010	| . **| . |	0.195912
18	-849.608	-0.15443	| . ***| . |	0.196613
19	145.017	0.02636	| . |* . |	0.198658
20	703.848	0.12794	| . |*** . |	0.198718
21	-219.761	-0.03995	| . *| . |	0.200109
22	-1024.1	-0.18615	| . ****| . |	0.200244
23	577.603	0.10499	| . |** . |	0.203156
24	1247.07	0.22668	| . |***** . |	0.204073

'.' MARKS TWO STANDARD ERRORS

UNIVARIATE MODEL

SAS
FORECASTS FOR VARIABLE PRISPOP

OBS	FORECAST	STD ERROR	LOWER 95%	UPPER 95%	ACTUAL	RESIDUAL
--------FORECAST BEGINS--------						
49	5414.2085	68.4155	5280.1168	5548.3001	5456.0000	41.7915
50	5475.0542	99.5980	5279.8461	5670.2624	5606.2500	131.1958
51	5502.6022	116.3389	5274.5825	5730.6218	5550.5000	47.8978
52	5524.5417	136.3443	5257.3124	5791.7711	5680.7500	156.2083
53	5573.1240	170.1034	5239.7280	5906.5199	5850.0000	276.8760
54	5625.1024	197.7407	5237.5385	6012.6663	5991.7500	366.6476
55	5658.2466	218.2378	5230.5092	6085.9841	6057.2500	399.0034
56	5692.2962	240.8530	5220.2339	6164.3585	6133.0000	440.7038
57	5740.9745	267.3005	5217.0762	6264.8728	6229.0000	488.0255
58	5788.2610	290.2467	5219.3892	6357.1329	6453.8000	665.5390
59	5825.5224	309.9560	5218.0211	6433.0238	6514.0000	688.4776
60	5865.4674	330.7227	5217.2642	6513.6706	6445.5000	580.0326
61	5912.9230	352.3527	5222.3258	6603.5202		
62	5957.9562	371.8441	5229.1567	6686.7557		
63	5998.0055	389.8152	5233.9833	6762.0276		
64	6040.6300	408.1972	5240.5798	6840.6802		
65	6086.8405	426.5111	5250.8958	6922.7851		
66	6130.9282	443.5126	5261.6612	7000.1952		
67	6172.7369	459.7418	5271.6614	7073.8124		
68	6216.4685	476.0398	5283.4495	7149.4875		
69	6261.7588	492.0175	5297.4243	7226.0933		
70	6305.5368	507.1782	5311.4877	7299.5858		
71	6348.3912	521.8964	5325.4952	7371.2872		
72	6392.5103	536.5219	5340.9489	7444.0718		

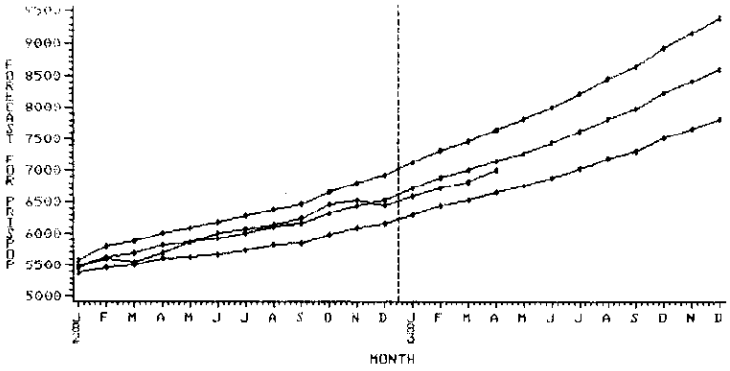
TRANSFER FUNCTION MODEL

SAS
FORECASTS FOR VARIABLE PRISPOP

OBS	FORECAST	STD ERROR	LOWER 95%	UPPER 95%	ACTUAL	RESIDUAL
--------FORECAST BEGINS--------						
49	5476.7740	50.5251	5377.7468	5575.8012	5456.0000	-20.7740
50	5625.1075	81.8289	5464.7262	5785.4889	5606.2500	-18.8575
51	5690.1703	93.6561	5506.6081	5873.7325	5550.5000	-139.6703
52	5802.1214	104.1350	5598.0210	6006.2219	5680.7500	-121.3714
53	5855.2304	114.6757	5630.4706	6079.9901	5850.0000	-5.2304
54	5915.2415	125.3677	5669.5259	6160.9572	5991.7500	76.5085
55	5994.8261	136.2831	5727.7167	6261.9355	6057.2500	62.4239
56	6095.3931	147.4830	5806.3323	6384.4539	6133.0000	37.6069
57	6149.3739	159.0216	5837.6979	6461.0498	6229.0000	79.6261
58	6312.9697	170.9485	5977.9175	6648.0219	6453.8000	140.8303
59	6438.6865	183.3107	6079.4049	6797.9681	6514.0000	75.3135
60	6522.2261	196.1534	6137.7732	6906.6790	6445.5000	-76.7261
61	6703.8951	210.1825	6291.9459	7115.8443		
62	6871.1948	225.0072	6430.1896	7312.2000		
63	6999.0043	239.6064	6529.3853	7468.6233		
64	7151.4041	254.7234	6652.1564	7650.6517		
65	7274.4254	270.5543	6744.1498	7804.7011		
66	7441.5299	287.1467	6878.7340	8004.3259		
67	7618.5857	304.5496	7021.6807	8215.4908		
68	7821.1103	322.8144	7188.4069	8453.8137		
69	7983.8230	341.9946	7313.5273	8654.1187		
70	8222.4913	362.1459	7512.6999	8932.2827		
71	8425.1666	383.3267	7673.8616	9176.4717		
72	8597.3618	405.5984	7802.4052	9392.3184		

Oklahoma Department of Corrections
Prison Population, Projection, and Confidence Limits

Model: Pop = 4151.57 + (-.0478 - .4441B) Parole + (.3288 + .7231B) Recpt + (-.5183 - .8151B - .2595B^{4}) Disch + (1 - 1.0549B)$^{-1}$ a



Data source: Oklahoma Department of Corrections, Jan 1978—Dec 1983