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 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:

\[ (1 - 0.31005*B^8 - 0.31385*B^{16}) (1 - B)^{45.945} + \epsilon \]

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

\[ DISCH = 112.869 + (.5635*B)^{12} + .67458*B^{24} + \epsilon \]

\[ PAROLE = 83.25 + \epsilon \]

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 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:
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, parolees, 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.

Oklahoma Department of Corrections
Prison Population
1960–1983

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 parole.

Data source: Oklahoma Department of Corrections, Jan 1979-Dec 1982
### Autocorrelations

<table>
<thead>
<tr>
<th>LAG COVARIANCE CORRELATION</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRISPOP</td>
<td>0.93448</td>
<td>0.01447</td>
<td>0.07223</td>
<td>-0.07919</td>
<td>0.00281</td>
<td>0.07424</td>
<td>-0.05669</td>
<td>-0.00929</td>
<td>0.01411</td>
<td>0.01879</td>
<td>0.01364</td>
<td>-0.00332</td>
<td>0.02999</td>
<td>0.0398825</td>
<td>0.408395</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Remarks:**
- Marks two standard errors.

### Inverse Autocorrelations

<table>
<thead>
<tr>
<th>LAG CORRELATION</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRISPOP</td>
<td>0.93448</td>
<td>0.01447</td>
<td>0.07223</td>
<td>-0.07919</td>
<td>0.00281</td>
<td>0.07424</td>
<td>-0.05669</td>
<td>-0.00929</td>
<td>0.01411</td>
<td>0.01879</td>
<td>0.01364</td>
<td>-0.00332</td>
<td>0.02999</td>
<td>0.0398825</td>
<td>0.408395</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Remarks:**
- Marks two standard errors.

### Partial Autocorrelations

<table>
<thead>
<tr>
<th>LAG CORRELATION</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRISPOP</td>
<td>0.93448</td>
<td>0.01447</td>
<td>0.07223</td>
<td>-0.07919</td>
<td>0.00281</td>
<td>0.07424</td>
<td>-0.05669</td>
<td>-0.00929</td>
<td>0.01411</td>
<td>0.01879</td>
<td>0.01364</td>
<td>-0.00332</td>
<td>0.02999</td>
<td>0.0398825</td>
<td>0.408395</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Remarks:**
- Marks two standard errors.

---

**SAS ARIMA Procedure**

**Mean of Working Series:** 5949.69

**Standard Deviation:** 662.335

**Number of Observations:** 60
### UNIVARIATE MODEL

**SAS FORECASTS FOR VARIABLE PRISPOP**

<table>
<thead>
<tr>
<th>OBS</th>
<th>FORECAST</th>
<th>STD ERROR</th>
<th>LOWER 95%</th>
<th>UPPER 95%</th>
<th>ACTUAL</th>
<th>RESIDUAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>5414.2085</td>
<td>86.4155</td>
<td>5280.1166</td>
<td>5548.3091</td>
<td>5455.0000</td>
<td>41.7915</td>
</tr>
<tr>
<td>50</td>
<td>5475.0542</td>
<td>95.5980</td>
<td>5270.8611</td>
<td>5670.2624</td>
<td>5656.2500</td>
<td>131.1958</td>
</tr>
<tr>
<td>51</td>
<td>5504.5602</td>
<td>116.3389</td>
<td>5207.3126</td>
<td>5781.7111</td>
<td>5683.7560</td>
<td>156.3043</td>
</tr>
<tr>
<td>52</td>
<td>5573.1240</td>
<td>170.1089</td>
<td>5227.3885</td>
<td>5914.7500</td>
<td>5994.7560</td>
<td>366.4476</td>
</tr>
<tr>
<td>53</td>
<td>5618.2466</td>
<td>218.2378</td>
<td>5237.0825</td>
<td>6070.2624</td>
<td>6070.2624</td>
<td>399.0034</td>
</tr>
<tr>
<td>54</td>
<td>5642.2662</td>
<td>240.8530</td>
<td>5280.2339</td>
<td>6024.3585</td>
<td>6183.0300</td>
<td>440.7038</td>
</tr>
<tr>
<td>55</td>
<td>5740.7478</td>
<td>267.3005</td>
<td>5277.7116</td>
<td>6074.8116</td>
<td>6074.8116</td>
<td>468.3522</td>
</tr>
<tr>
<td>56</td>
<td>5786.7810</td>
<td>290.2467</td>
<td>5316.7489</td>
<td>6248.8116</td>
<td>6248.8116</td>
<td>518.5769</td>
</tr>
<tr>
<td>57</td>
<td>5865.4784</td>
<td>332.7325</td>
<td>5437.4952</td>
<td>6387.4539</td>
<td>6387.4539</td>
<td>588.0255</td>
</tr>
<tr>
<td>58</td>
<td>5912.2330</td>
<td>352.3527</td>
<td>5469.2585</td>
<td>6438.2219</td>
<td>6438.2219</td>
<td>62.4239</td>
</tr>
<tr>
<td>59</td>
<td>5957.9562</td>
<td>371.6441</td>
<td>5526.3172</td>
<td>6497.6790</td>
<td>6497.6790</td>
<td>140.8303</td>
</tr>
<tr>
<td>60</td>
<td>6017.9845</td>
<td>392.0172</td>
<td>5578.9562</td>
<td>6558.3585</td>
<td>6558.3585</td>
<td>75.3133</td>
</tr>
<tr>
<td>61</td>
<td>6056.3304</td>
<td>413.3254</td>
<td>5609.9853</td>
<td>6610.0219</td>
<td>6610.0219</td>
<td>366.4476</td>
</tr>
<tr>
<td>62</td>
<td>6110.2922</td>
<td>434.6331</td>
<td>5664.6519</td>
<td>6664.6519</td>
<td>6664.6519</td>
<td>75.3133</td>
</tr>
<tr>
<td>63</td>
<td>6164.2466</td>
<td>456.9408</td>
<td>5718.1945</td>
<td>6718.1945</td>
<td>6718.1945</td>
<td>140.8303</td>
</tr>
<tr>
<td>64</td>
<td>6216.8405</td>
<td>479.2486</td>
<td>5772.5933</td>
<td>6772.5933</td>
<td>6772.5933</td>
<td>75.3133</td>
</tr>
<tr>
<td>65</td>
<td>6261.5472</td>
<td>501.5564</td>
<td>5826.1051</td>
<td>6826.1051</td>
<td>6826.1051</td>
<td>140.8303</td>
</tr>
<tr>
<td>66</td>
<td>6305.5363</td>
<td>524.8642</td>
<td>5870.6979</td>
<td>6921.9489</td>
<td>6921.9489</td>
<td>75.3133</td>
</tr>
<tr>
<td>67</td>
<td>6348.3912</td>
<td>549.1721</td>
<td>5914.5732</td>
<td>6973.5276</td>
<td>6973.5276</td>
<td>75.3133</td>
</tr>
<tr>
<td>68</td>
<td>6392.5103</td>
<td>574.4800</td>
<td>5958.3219</td>
<td>7027.6792</td>
<td>7027.6792</td>
<td>75.3133</td>
</tr>
</tbody>
</table>

### TRANSFER FUNCTION MODEL

**SAS FORECASTS FOR VARIABLE PRISPOP**

<table>
<thead>
<tr>
<th>OBS</th>
<th>FORECAST</th>
<th>STD ERROR</th>
<th>LOWER 95%</th>
<th>UPPER 95%</th>
<th>ACTUAL</th>
<th>RESIDUAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>5117.6740</td>
<td>20.5768</td>
<td>5035.8012</td>
<td>5200.0000</td>
<td>40.7780</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5265.1075</td>
<td>81.8289</td>
<td>5204.7262</td>
<td>5325.4889</td>
<td>5576.2500</td>
<td>163.8575</td>
</tr>
<tr>
<td>51</td>
<td>5390.0173</td>
<td>93.6161</td>
<td>5330.4706</td>
<td>5450.5572</td>
<td>5606.2500</td>
<td>70.0700</td>
</tr>
<tr>
<td>52</td>
<td>5525.2306</td>
<td>114.6757</td>
<td>5465.4706</td>
<td>5605.5572</td>
<td>5761.7500</td>
<td>82.4239</td>
</tr>
<tr>
<td>53</td>
<td>5655.2945</td>
<td>135.3731</td>
<td>5595.4706</td>
<td>5805.5572</td>
<td>5861.7500</td>
<td>95.7280</td>
</tr>
<tr>
<td>54</td>
<td>5795.3584</td>
<td>156.2831</td>
<td>5735.4706</td>
<td>5905.5572</td>
<td>5961.7500</td>
<td>109.0280</td>
</tr>
<tr>
<td>55</td>
<td>5935.4223</td>
<td>177.2931</td>
<td>5875.4706</td>
<td>6085.5572</td>
<td>6061.7500</td>
<td>122.3280</td>
</tr>
<tr>
<td>56</td>
<td>6075.4862</td>
<td>198.3031</td>
<td>5915.4706</td>
<td>6205.5572</td>
<td>6161.7500</td>
<td>135.6280</td>
</tr>
<tr>
<td>57</td>
<td>6215.5501</td>
<td>219.3131</td>
<td>6055.4706</td>
<td>6395.5572</td>
<td>6261.7500</td>
<td>148.9280</td>
</tr>
<tr>
<td>58</td>
<td>6355.6140</td>
<td>240.3231</td>
<td>6195.4706</td>
<td>6535.5572</td>
<td>6361.7500</td>
<td>162.2280</td>
</tr>
<tr>
<td>59</td>
<td>6495.6779</td>
<td>261.3331</td>
<td>6335.4706</td>
<td>6675.5572</td>
<td>6461.7500</td>
<td>175.5280</td>
</tr>
<tr>
<td>60</td>
<td>6635.7418</td>
<td>282.3431</td>
<td>6475.4706</td>
<td>6815.5572</td>
<td>6561.7500</td>
<td>188.8280</td>
</tr>
<tr>
<td>61</td>
<td>6775.8057</td>
<td>303.3531</td>
<td>6615.4706</td>
<td>7055.5572</td>
<td>6661.7500</td>
<td>202.1280</td>
</tr>
<tr>
<td>62</td>
<td>6915.8696</td>
<td>324.3631</td>
<td>6755.4706</td>
<td>7295.5572</td>
<td>6761.7500</td>
<td>215.4280</td>
</tr>
<tr>
<td>63</td>
<td>7055.9335</td>
<td>345.3731</td>
<td>6895.4706</td>
<td>7535.5572</td>
<td>6861.7500</td>
<td>228.7280</td>
</tr>
<tr>
<td>64</td>
<td>7196.0674</td>
<td>366.3831</td>
<td>6935.4706</td>
<td>7775.5572</td>
<td>6961.7500</td>
<td>242.0280</td>
</tr>
</tbody>
</table>

---

60
Oklahoma Department of Corrections
Prison Population, Projection, and Confidence Limits

Model for: $4151.57 + (-.0478 \times 44418$) Poisn $+ (.3259 \times 72518) Propt \times (-.5185 \times 1.518 \times 22558^t)$ Estim $+ (1 \times 10449)^{-3}$

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