Time Series - Australian Iron Production

Austin Ibarra

8/7/2020

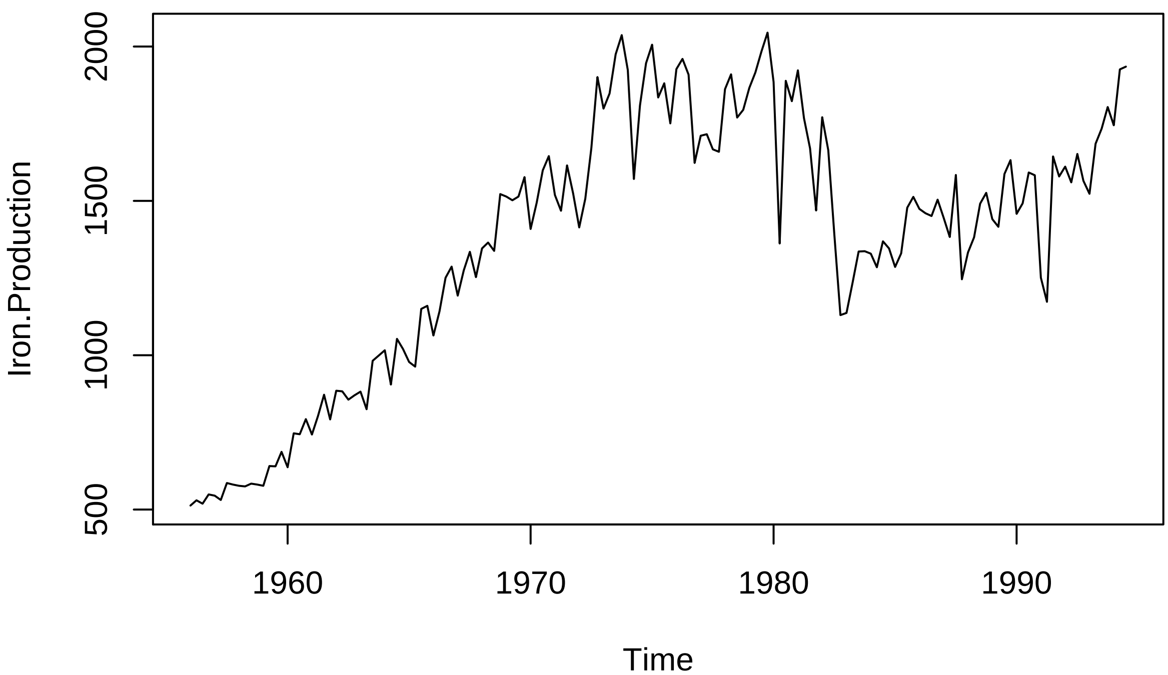
Introduction

This report will be a time series analysis on the quarterly iron production in Australia (in tonnes) from March 1956 to September 1994. Iron ore mining and steel production is a staple in the Australian economy and a global asset. I am interested in the iron production of the country as a function of time and see how the series behaves and if I can forecast it accurately towards the future.

Results & Discussion

I first wanted to see a general graph of the time series of the iron production in Australia.

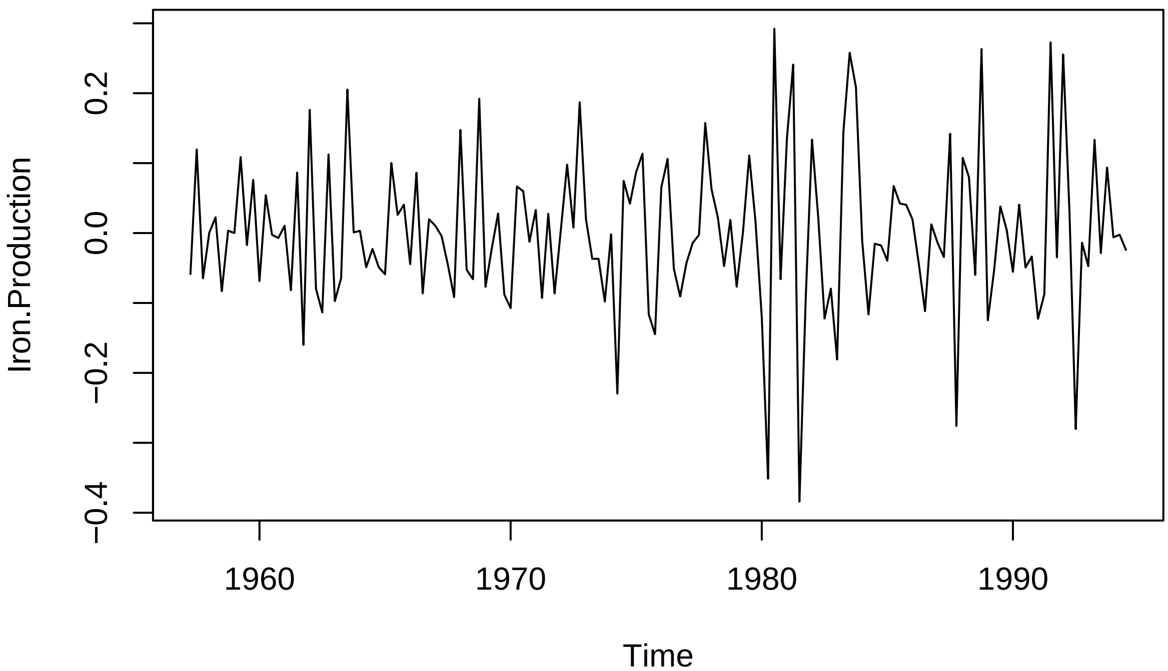
```
iron <- read.csv("basic-quarterly-iron-production(1).csv")
tsiron <- ts(iron, start = 1956, frequency = 4)
plot(tsiron)
```



From figure 1, we see a general time series plot of the iron production plotted against time, from 1956 to 1994. We observe a steady increasing trend in production from 1954 to around 1972 where we see a jump in production that generally keeps a stagnant trend with cycling oscillations. This trend continues until a drop-in production begins in 1980 to about 1983, most likely due to the global steel crisis of 1980. The trend begins to pick back up from 1984 onwards with a steady increase with seasonal oscillations.

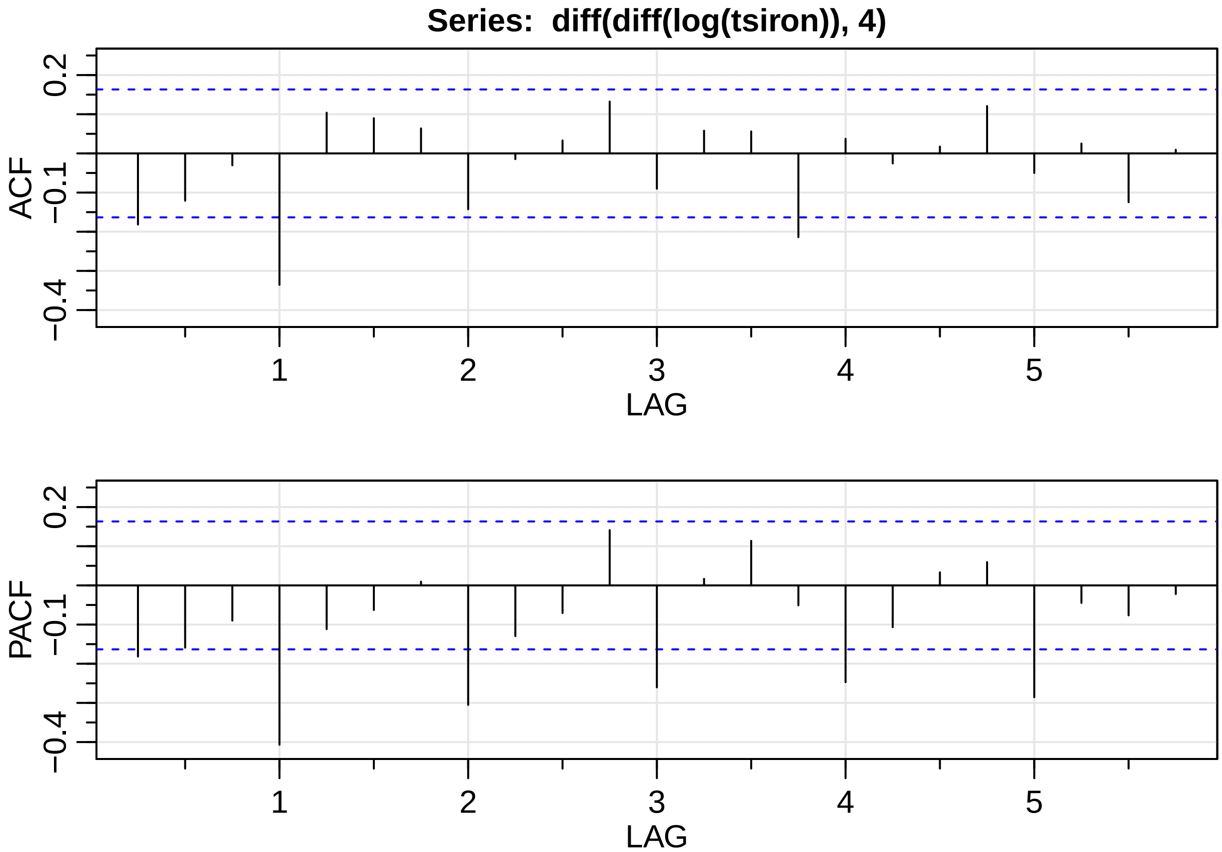
In order to begin modeling and analyzing the periodic components that make up the series, I begin transforming the data to make it a stationary time series. I control for the rising trend of the series through differencing and take an additional seasonal lag difference to account for the quarterly aspect of the data. I also take the log of the data to smooth out the large oscillations, most notably the steel crisis of 1980. The figure below depicts the stationary version of the series.

```
plot(diff(diff(log(tsiron)), 4))
```



To determine the order for the model of the series, I examine the ACF and PACF of the newly transformed series.

```
acf2(diff(diff(log(tsiron)), 4))
```

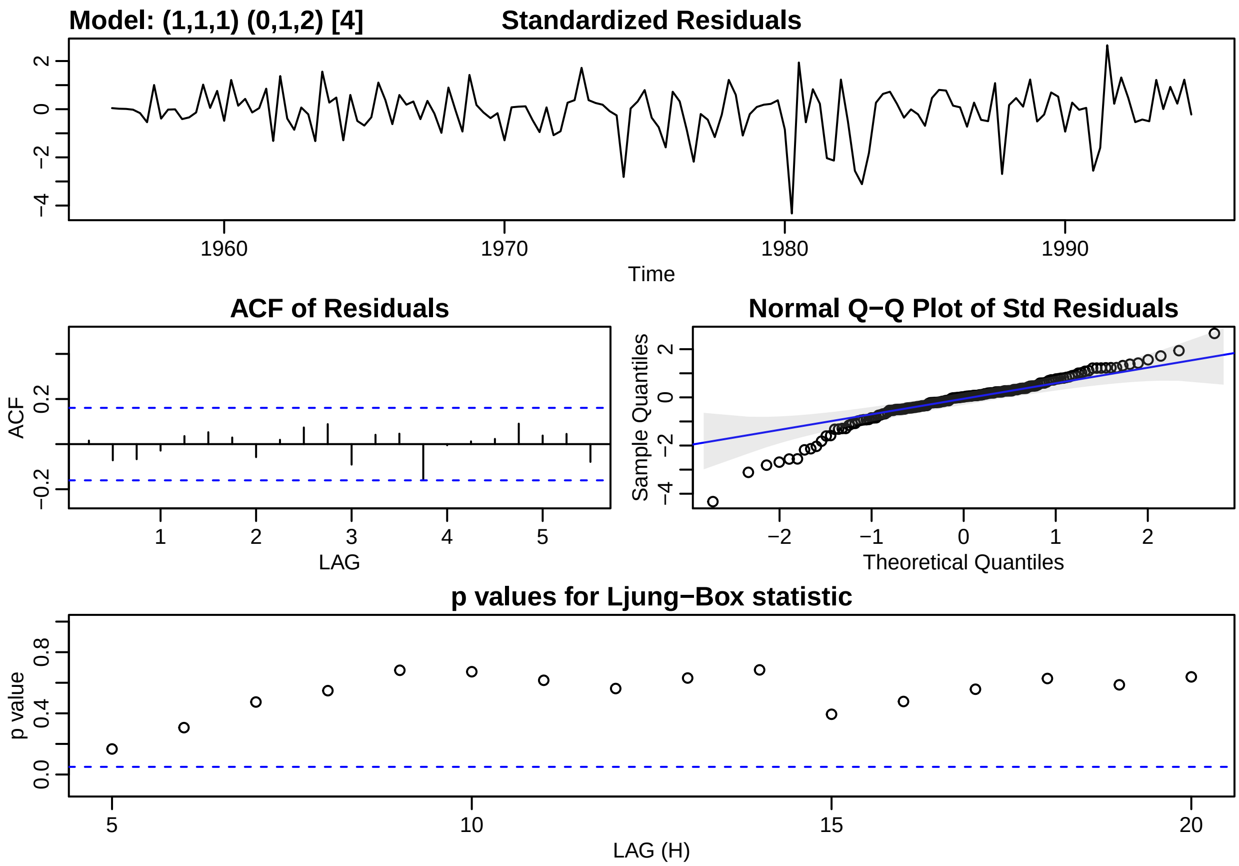



```
##         ACF  PACF
##  [1,] -0.18 -0.18
##  [2,] -0.12 -0.16
##  [3,] -0.03 -0.09
##  [4,] -0.34 -0.41
##  [5,]  0.10 -0.11
##  [6,]  0.09 -0.06
##  [7,]  0.06  0.01
##  [8,] -0.14 -0.30
##  [9,] -0.01 -0.13
## [10,]  0.03 -0.07
## [11,]  0.13  0.14
## [12,] -0.09 -0.26
## [13,]  0.06  0.02
## [14,]  0.06  0.11
## [15,] -0.21 -0.05
## [16,]  0.04 -0.25
## [17,] -0.03 -0.11
## [18,]  0.02  0.03
## [19,]  0.12  0.06
## [20,] -0.05 -0.29
## [21,]  0.03 -0.04
## [22,] -0.12 -0.08
## [23,]  0.01 -0.02
```

The ACF and PACF of the series both die after lag 1, which shows signs of an AR(1) and MA(1) process. Examining the seasonal lags, we see that the ACF dies after the first seasonal lag while the PACF tails off which depicts a SMA(1) or even (2) process. The cutoff points for the lags seem very close so we will try multiple models and select the one with the lowest AIC, significant p-values for the parameters, and normal residuals.

```
sarima(log(tsiron), 1,1,1,0,1,2,4)
```

```
## initial  value -2.202444
## iter   2 value -2.347667
## iter   3 value -2.430646
## iter   4 value -2.436151
## iter   5 value -2.444854
## iter   6 value -2.457144
## iter   7 value -2.457838
## iter   8 value -2.460063
## iter   9 value -2.463990
## iter  10 value -2.467391
## iter  11 value -2.468337
## iter  12 value -2.470067
## iter  13 value -2.474015
## iter  14 value -2.474355
## iter  15 value -2.474592
## iter  16 value -2.474615
## iter  17 value -2.474617
## iter  18 value -2.474617
## iter  18 value -2.474617
## iter  18 value -2.474617
## final  value -2.474617
## converged
## initial  value -2.482465
## iter   2 value -2.484708
## iter   3 value -2.484945
## iter   4 value -2.486521
## iter   5 value -2.486835
## iter   6 value -2.487049
## iter   7 value -2.487144
## iter   8 value -2.487171
## iter   9 value -2.487172
## iter  10 value -2.487173
## iter  10 value -2.487173
## final  value -2.487173
## converged
```



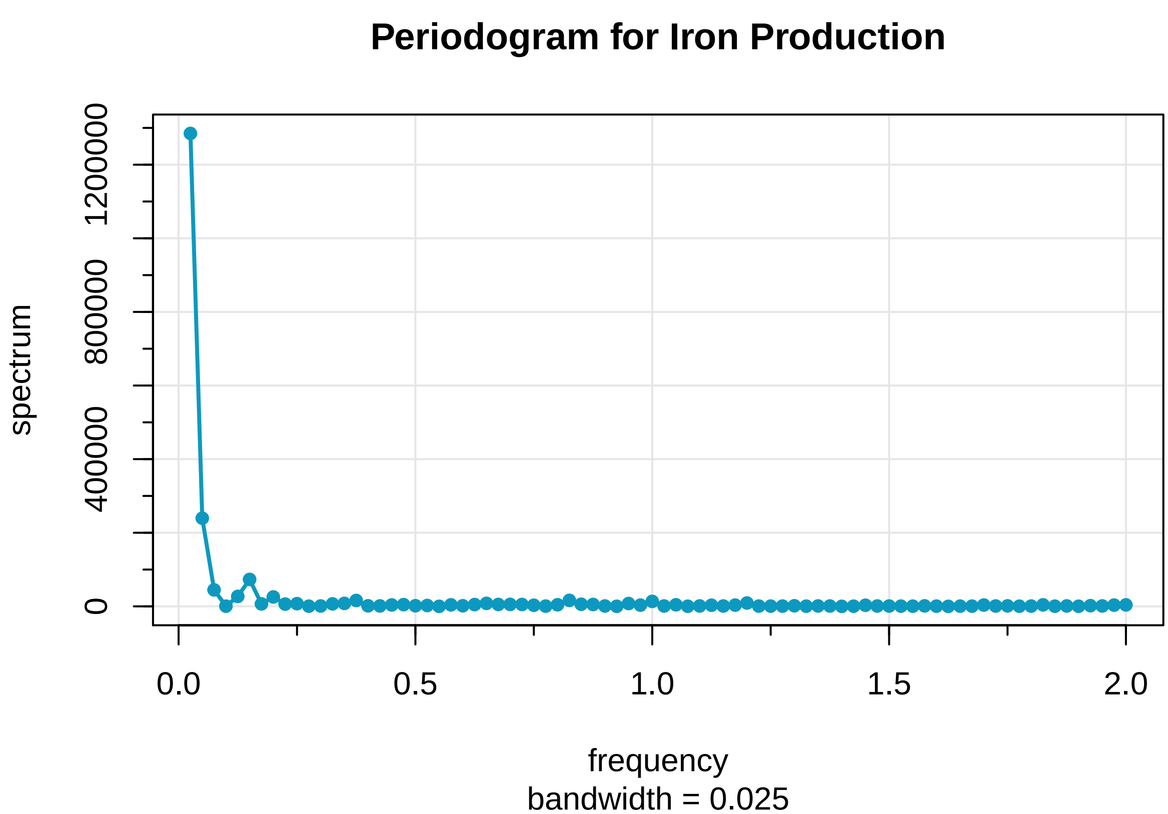

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), include.mean = !no.constant, transform.pars = trans, fixed = fixed,
##     optim.control = list(trace = trc, REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1      ma1     sma1     sma2
##       0.3843  -0.7128  -0.7371  -0.2028
## s.e.  0.1385   0.1093   0.1032   0.0888
##
## sigma^2 estimated as 0.006526:  log likelihood = 160.24,  aic = -310.47
##
## $degrees_of_freedom
## [1] 146
##
## $ttable
##      Estimate     SE t.value p.value
## ar1    0.3843 0.1385  2.7737  0.0063
## ma1   -0.7128 0.1093 -6.5203  0.0000
## sma1  -0.7371 0.1032 -7.1415  0.0000
## sma2  -0.2028 0.0888 -2.2841  0.0238
##
## $AIC
## [1] -2.029218
```

```
##
## $AICc
## [1] -2.027451
##
## $BIC
## [1] -1.930831
```

After running several models, I came up with the SARIMA (1,1,1)x(0,1,2)4 model that I believe does an acceptable job of modeling the data. The ACF of residuals seem properly bounded, the p-values for the Ljung-Box statistic indicate normality, the p-values for the estimates are all significant, and the AIC was the lowest amongst all other models I ran. The qqplot doesn't properly bound all the residuals within the standard error, though this was the best model that had the most accuracy compared with the rest. Overall, the model is the most accurate while keeping the parsimonious principle in mind to keep from overfitting the data.

I also want to examine a periodogram of the series to determine which frequencies dominate the periods that make up the series.

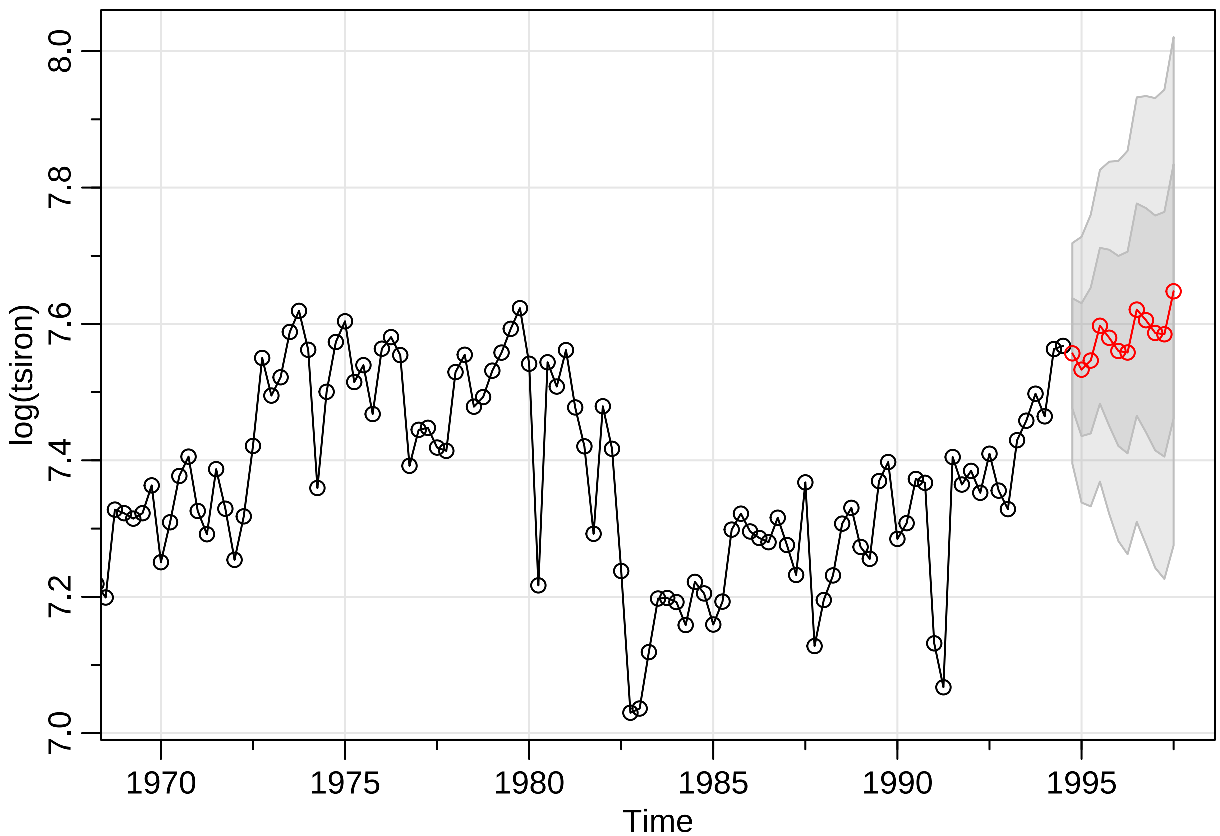
```
mvspec(tsiron, main="Periodogram for Iron Production", col=rgb(.05,.6,.75), lwd=2, type="o", pch=20)
```



The figure depicts a large spike to the left of the series. This shows that the series is dominated by slower oscillations, indicating that the phis that show the correlation between each consecutive point are highly correlated in the periods with little variation. This confirms our initial analysis of the first figure how the series trend is steadily increasing until 1980, then drops, then steadily increases again until 1994.

Next, we will use our model to forecast the next 12 quarters.

```
pred.tsiron <- sarima.for(log(tsiron), 12,1,1,1,0,1,2,4)
```




```
predicted <- exp(pred.tsiron$pred)
L <- predicted - 1.96*exp(pred.tsiron$se)
U <- predicted + 1.96*exp(pred.tsiron$se)
predicted
```

```
##          Qtr1     Qtr2     Qtr3     Qtr4
## 1994                            1913.880
## 1995 1868.661 1893.871 1993.012 1958.253
## 1996 1920.600 1916.283 2040.733 2009.155
## 1997 1972.037 1968.186 2096.245
```

From our forecast, we see that the model does a reasonably good job in predicting the next 12 quarters with an increasingly upward trend. The standard error grows substantially larger showing how the uncertainty of the model as it goes farther into the future.

Conclusion Overall, I reached the goals set in the introduction through analysis of the time series regarding iron production in Australia, examining the periodic elements that dominate the series, SARIMA modeling of the data, and finally using the model to forecast 12 quarters into the future with moderate accuracy.