I. Some non-technical definitions
   a. **Stationary**: the mean and variance of a time series do not depend on which segment of the data we use to estimate them.
   b. **Causal**: we can compute (or represent) the $Y_t$, in terms of the current and past error values, $\varepsilon_t$ and $\varepsilon_{t-u}$ for $u>0$ (without needing future ones).
   c. **Invertible**: we can compute the estimate of data, $Y_t$, in terms of the current error, $\varepsilon_t$, and past data values, $Y_{t-u}$ for $u>0$ (without needing future ones). Or, we can estimate each error estimate from just the current data value and past values.

II. Review and Extension to ARMA and ARIMA
   a. **AR(p)**
      i. $Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$
      ii. $\Phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$
      iii. $Y_t \Phi(B) = \varepsilon_t$ or $Y_t = \Phi^{-1}(B) \varepsilon_t$
      iv. Always invertible.
   v. Stationary and causal if the roots of $\Phi(B) = 0$ are all outside the unit circle.
   vi. ACF exponentially declines towards 0 (sinusoidal if the roots are complex)
   vii. PACF zero after lag $p$
   viii. Examples (apply to de-seasoned data only)
b. MA(q)

i. \( y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \) in R or \( y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \) in SAS

ii. \( \Theta(B) = (1+\theta_1 B^1 + \cdots + \theta_q B^q) \) in R or \( \Theta(B) = (1-\theta_1 B^1 - \cdots - \theta_q B^q) \) in SAS

iii. \( y_t = \Theta(B) \varepsilon_t \)

iv. Causal and stationary. Invertible if all roots of \( \Theta(B) = 0 \) are outside the unit circle.

v. ACF zero after lag q (i.e., q+1 peaks on the plot)

vi. PACF declines towards 0

vii. Examples (apply to de-seasoned data only)

Moving Average Models: Data, ACF, PACF

MA(1), Theta=-0.7

MA(2), Theta=-0.6,+0.3

MA(2), Theta=+0.2,-0.9

MA(2), Theta=+1.0,-0.5
III. **ARMA(p,q)**

a. \[ Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad \text{(or flipped \( \theta \) signs in \( R \))} \]

b. \[ \Phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p), \quad \Theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q) \quad \text{(or flipped \( \theta \) signs in \( R \))} \]

c. \[ \Phi(B) Y_t = \Theta(B) \varepsilon_t \quad \text{or} \quad Y_t = [\mu] + \Phi(B)^{-1} \Theta(B) \varepsilon_t \quad \text{or} \quad Y_t = [\mu] + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \]

d. Stationary and causal if the roots of \( \Phi(B) = 0 \) are all outside the unit circle and invertible if the roots of \( \Theta(B) = 0 \) are all outside the unit circle.

e. ACF and PACF have exponential declines

f. Examples

ARMA Models: Data, ACF, PACF
IV. The I in ARIMA: integrated
   a. For the econometrics perspective see
   b. Non-stationary “trending” can be deterministic or stochastic
   c. In a deterministic trend, the long term future is always upwards (or downwards, but not both), the errors (a.k.a. shocks, innovations) have transitory effects, ARMA on the detrended data is all that is needed, and there are no unit roots of Φ(B).
   d. In a stochastic trend, the direction of the long term trend is unpredictable, the errors have permanent effects, ARIMA (i.e., differencing of the time series to achieve a new stationary time series that can be modeled with ARMA) is needed, and there is at least one unit root of Φ(B).
   e. Among other reasons, we need to know which applies because neither the law-of-large-numbers nor the central-limit-theorem hold in the absence of stationarity.

V. ARIMA(p,d,q) where I is “integrated”
   a. Φ(B) \(\nabla^{d}Y_t = \Theta(B) \varepsilon_t\) where \(\nabla\) = \((1 - B)\) is the difference operator (called “nabla”):
      \(\nabla^{1}Y_t = \nabla Y_t = (1-B)Y_t = Y_{t}-Y_{t-1}\)
      \(\nabla^{2}Y_t = (1-B)\nabla Y_t = (1-B)(1-B)Y_t = Y_{t} - 2Y_{t-1} + Y_{t-2}\)
      Or \(\nabla^{2}Y_t = (1-B)^2Y_t = (1-B)[(1-B)Y_t] = (1-B)(Y_{t-1} - Y_{t-2}) = Y_{t} - 2Y_{t-1} + Y_{t-2}\)
      Also, we define the seasonal version as \(\nabla^{12}Y_t = (1-B^{12})Y_t = Y_{t} - Y_{t-12}\)
   b. Think of this as \(W_t = \nabla^{d}Y_t\) and \(W_t\) follows ARMA(p,q) model.
   c. The differenced time series is \(d\) shorter than the original.
   d. This model is used for non-stationary time series.
   e. With a constant and \(d=1\) an underlying linear trend is included: \(Y_t - Y_{t-1} = \mu + \text{error}\).
   f. With a constant and \(d=2\) and underlying deterministic quadratic trend is included.
   g. Most commonly no constant is used, and no long-term direction is implied.
   h. Forecast intervals get ever wider for longer “leads”.
   i. Special cases:
      i. ARIMA(0,0,0), is white noise
      ii. ARIMA(0,1,0) is a random walk.
      iii. ARIMA(0,1,1) is ETS(A,N,N) with \(\alpha = \theta_1 + 1\)
      iv. ARIMA(0,2,2) is ETS(A,A,N) with \(\alpha = 1-\theta_2, \beta = \theta_1-\theta_2+1\)
      v. ARIMA(1,1,2) is ETS(A,Ad,N) with \(\phi = \phi_2, \alpha = 1-\phi^2/\phi_2, \beta = \phi_1-\phi_2/\phi_1^2-1\)
   j. Not all ETS models have an ARIMA equivalent. Not all ARIMA models have an ETS equivalent.
   k. Unit root tests are useful for detecting the need for \(d>0\). A linear decline in the ACF suggests the need for differencing. A linear decline in the IACF is a good indicator of over-differencing. Available unit-root tests include Dickey-Fuller, Augmented Dickey-Fuller (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and some others.
VI. Principles of SAS PROC ARIMA modeling
a. Identification of the appropriate d, p, and q is an art.
b. George Box said “All models are wrong, but some are useful”. In this context, “nearby” ARIMA models tend to have similar predictions. Also, multiple representations of the same model are sometimes possible.
c. Once a model is chosen, the software can make appropriate predictions. But most software gives 95% intervals that hold less than 95% of the future even if the model is correct and the underlying process remains unchanged (i.e., no drift in the parameters and no drastic unexpected events) because they do not include uncertainty in the estimation of the parameters. (Do you know an approach that, in general, automatically includes parameter uncertainty?)
d. Predictions may have a level mean path or have a trending mean path. Think carefully (and talk to your client) about whether you are using the appropriate one before making decisions based on the computer predictions. Ignoring unit roots or deterministic trends is a major cause of bad predictions.
e. Ljung and Box’s “autocorrelation check for white noise” performed on residuals can be used to check the adequacy of choice of p, d, and q. (Also ACF, PACF, IACF.) If the residuals are not nearly Gaussian white noise, the predictions will be incorrect!
f. Special tests (ESACF, MINIC, SCAN) are sometimes helpful to complement finding p, q, and d using EDA plots, subject matter theory, and ACF/PACF/IACF.
g. Practical tip: Some over-fitting of one side of an ARMA is usually no problem, and large p-values for higher order terms indicate they should be dropped. Simultaneously over-fitting p and q will result in unstable estimates and/or convergence problems.
h. METHOD=ML is better for the ESTIMATE statement than the default CLS. With d>0, NOCONSTANT is usually appropriate for the ESTIMATE statement (0 mean difference).
i. Forecasting the end of an observed series is one good check of model adequacy (e.g., FORECAST BACK=6 LEAD=6).
j. (PROC MODEL can be used to forecast non-linear trends with ARMA error.)

VII. SAS Code for EDA and Basic Identification of d, p, and q
a. Iron and Steel Example

```sas
TITLE '(1937-1980) Iron and Steel Exports (millions of tons)';
DATA steel;
  INPUT export @@;
  year = INTNX("YEAR", "01JAN1936"D, _N_);
  FORMAT year YEAR4.;
DATALINES;
  3.89 2.41 2.8  8.72 7.12 7.24 7.15 6.05 5.21 5.03 6.88 4.7  5.06 3.16
  3.62 4.55 2.43 3.16 4.55 5.17 6.95 3.46 2.13 3.47 2.79 2.52 2.8  4.04
  3.08 2.28 2.17 2.78 5.94 8.14 3.55 3.61 5.06 7.13 4.15 3.86 3.22
  3.5  3.76 5.11
RUN;

PROC SGPLOT DATA=steel;
  SERIES X=year Y=export;
RUN;
```

PROC ARIMA DATA=steel PLOTS=FORECAST(ALL);
  IDENTIFY VAR=export STATIONARITY=(ADF=3) NLAG=10;
RUN;

Name of Variable = export

Mean of Working Series    4.418182
Standard Deviation         1.73354
Number of Observations          44

Autocorrelation Check for White Noise

To    Chi-    Pr >
  Lag  Square  DF  Chisq  --------------------------Autocorrelations--------------------------
       6        0.0586

Augmented Dickey-Fuller Unit Root Tests

<table>
<thead>
<tr>
<th>Type</th>
<th>Lags</th>
<th>Rho</th>
<th>Pr &lt; Rho</th>
<th>Tau</th>
<th>Pr &lt; Tau</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero Mean</td>
<td>0</td>
<td>-2.8498</td>
<td>0.2416</td>
<td>-1.15</td>
<td>0.2238</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-1.9403</td>
<td>0.3337</td>
<td>-0.87</td>
<td>0.3330</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.9228</td>
<td>0.4814</td>
<td>-0.57</td>
<td>0.4647</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-1.0378</td>
<td>0.4613</td>
<td>-0.89</td>
<td>0.3241</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Mean</td>
<td>0</td>
<td>-22.6286</td>
<td>0.0022</td>
<td>-3.82</td>
<td>0.0052</td>
<td>7.31</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-30.4238</td>
<td>0.0003</td>
<td>-3.84</td>
<td>0.0050</td>
<td>7.40</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-24.5872</td>
<td>0.0010</td>
<td>-3.01</td>
<td>0.0421</td>
<td>4.55</td>
<td>0.0667</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-15.1798</td>
<td>0.0243</td>
<td>-2.34</td>
<td>0.1657</td>
<td>2.79</td>
<td>0.3803</td>
</tr>
<tr>
<td>Trend</td>
<td>0</td>
<td>-24.2698</td>
<td>0.0012</td>
<td>-3.96</td>
<td>0.0179</td>
<td>7.83</td>
<td>0.0247</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-36.0520</td>
<td>0.0002</td>
<td>-4.21</td>
<td>0.0095</td>
<td>8.91</td>
<td>0.0074</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-34.7977</td>
<td>0.0002</td>
<td>-3.54</td>
<td>0.0484</td>
<td>6.33</td>
<td>0.0680</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-18.5709</td>
<td>0.0532</td>
<td>-2.34</td>
<td>0.4048</td>
<td>2.88</td>
<td>0.6144</td>
</tr>
</tbody>
</table>

For these data (mean not 0)
ESTIMATE P=1 METHOD=ML; /* ADF rejects unit root, so this analysis is OK */
FORECAST BACK=5 LEAD=5;
ESTIMATE Q=1 METHOD=ML;
FORECAST BACK=5 LEAD=5;
RUN;

Maximum Likelihood Estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>MU</td>
<td>4.42129</td>
<td>0.43002</td>
<td>10.28</td>
<td>&lt;.0001</td>
<td>0</td>
</tr>
<tr>
<td>AR1,1</td>
<td>0.46415</td>
<td>0.13579</td>
<td>3.42</td>
<td>0.0006</td>
<td>1</td>
</tr>
</tbody>
</table>

Constant Estimate 2.36913
Variance Estimate 2.443378
Std Error Estimate 1.563131
AIC 166.3711
SBC 169.9395
Number of Residuals 44

Correlations of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MU</th>
<th>AR1,1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MU</td>
<td>1.000</td>
<td>0.016</td>
</tr>
<tr>
<td>AR1,1</td>
<td>0.016</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Autocorrelation Check of Residuals

| To | Chi- Square | DF | ChiSq | Pr > | ----------------- | Autocorrelations ----------------- |}
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2.17</td>
<td>5</td>
<td>0.8258</td>
<td>-0.146</td>
<td>-0.053</td>
<td>0.073</td>
</tr>
<tr>
<td>12</td>
<td>4.29</td>
<td>11</td>
<td>0.9606</td>
<td>-0.020</td>
<td>-0.072</td>
<td>-0.019</td>
</tr>
<tr>
<td>18</td>
<td>7.23</td>
<td>17</td>
<td>0.9804</td>
<td>0.095</td>
<td>0.014</td>
<td>0.007</td>
</tr>
<tr>
<td>24</td>
<td>13.03</td>
<td>23</td>
<td>0.9512</td>
<td>-0.217</td>
<td>-0.095</td>
<td>-0.084</td>
</tr>
</tbody>
</table>

7
Model for variable export

Estimated Mean  4.421293

Autoregressive Factors
Factor 1:  1 - 0.46415 B**(1)

(then forecasts as text)
Autocorrelation Check of Residuals

To     Chi-    Pr >
Lag  Square  DF  ChiSq------------------Autocorrelations---------------
  6  1.31     5  0.9334  0.067  -0.024  0.084  0.078  -0.019
 12  3.25    11  0.9507  0.006  -0.079  0.052  -0.013  0.146  0.040
 18  6.66    17  0.9876  0.063  0.000  0.044  -0.092  0.095  -0.149
 24 14.10   23  0.9239 -0.208  -0.135  0.116  -0.085  0.013  -0.073

Estimated Mean  4.424887

Moving Average Factors
Factor 1:  1 + 0.49072 B**(1)

TITLE2 "MA(1) Model"; /* (Unfortunately, graph not affected) */
FORECAST LEAD=20; /* Based on Q=1, from last ESTIMATE statement */
QUIT;
VIII. R code: see LNTS7.R

# (1937-1980) Iron and Steel Exports (millions of tons)
steel = c(3.89, 2.41, 2.8, 8.72, 7.12, 7.24, 7.15, 6.05, 5.21, 5.03,
         6.88, 4.7, 5.06, 3.16, 3.62, 4.55, 2.43, 3.16, 4.55, 5.17,
         6.95, 3.46, 2.13, 3.47, 2.79, 2.52, 2.8, 4.04, 3.08, 2.28,
         2.17, 2.78, 5.94, 8.14, 3.55, 3.61, 5.06, 7.13, 4.15, 3.86,
         3.22, 3.5, 3.76, 5.1)  
steel = ts(steel, start=1937, end=1980)
library(forecast)
tsdisplay(steel, main="LNTS7")

LNTS7

Lag ACF
2 4 6 8
-0.4 0.0 0.4
Lag PACF
2 4 6 8 10 12 14
-0.4 0.0 0.4
Start of multi-step forecasts
# Diagnostics
oldpar = par(no.readonly=TRUE)
par(oma=c(0,0,3,0))
ar1 = Arima(steel, order=c(1,0,0), method="ML")
tsdia(ar1)
mtext("Residuals from AR1 model applied to steel data", outer=TRUE)

ma1 = Arima(steel, order=c(0,0,1), method="ML")
tdia(ma1)
mtext("Residuals from MA1 model applied to steel data", outer=TRUE)

arma11 = Arima(steel, order=c(1,0,1), method="ML")
tsdia(arma11)
mtext("Residuals from ARMA11 model applied to steel data", outer=TRUE)
par(oldpar)

# Forecasts
plot(forecast(ar1, h=10), xlab="year", ylab="Steel Production", main="AR1")

mal = Arima(steel, order=c(0,0,1), method="ML")
tdia(mal)
mtext("Residuals from MA1 model applied to steel data", outer=TRUE)

arma11 = Arima(steel, order=c(1,0,1), method="ML")
tdia(arma11)
mtext("Residuals from ARMA11 model applied to steel data", outer=TRUE)
par(oldpar)

# Forecasts
plot(forecast(ar1, h=10), xlab="year", ylab="Steel Production", main="AR1")
plot(forecast(ma1, h=10), xlab="year", ylab="Steel Production", main="MA1")
plot(forecast(arma11, h=10), xlab="year", ylab="Steel Production", main="ARMA11")

# SAS type predict plotter
# Note than back and lead are in time points (e.g., months, not years)
# tsPredPlot = function(y, model, main=NULL, ylab="y", back=0, lead=1) {
  if (!require(forecast)) stop("need to install package 'forecast'.")
  n = length(y)

  # compute main
  if (is.null(main)) {
    main = paste("ARIMA(" , length(model$model$phi),", ", length(model$model$Delta), ", ", length(model$model$theta), ")", sep="")
  }

  # refit if back>0
  if (back>0) {
    lead = back
    tsp = tsp(y)
    save = ts(y[(n-back+1):n], start=tsp[2] - (back-1)/tsp[3], end=tsp[2],
      frequency=tsp[3])
    y = ts(y[1:(n-back)], start=tsp[1], end=tsp[2]-back/tsp[3],
      frequency=tsp[3])
    model = Arima(y, order=c(length(model$model$phi), 0, length(model$model$theta)))
  }

  pred = as.data.frame(forecast(model, h=lead))
  tsp = tsp(y)
  P = ts(pred[,"Point Forecast"], start=tsp[2]+1/tsp[3],
    end=tsp[2]+lead/tsp[3], freq=tsp[3])
    freq=tsp[3])
    freq=tsp[3])
  plot(y, xlim=c(tsp[1], tsp[2]+lead/tsp[3]), main=main, ylab=ylab,
    ylim=range(y, P, L, U))
  if (lead>1) {
    lines(P, col=2)
    lines(L, col=2, lty=3)
    lines(U, col=2, lty=3)
  } else {
    lines(c(tsp[2],tsp[2]+1/tsp[3]), rep(as.numeric(P),2), col=2)
    lines(c(tsp[2],tsp[2]+1/tsp[3]), rep(as.numeric(L),2), col=2, lty=3)
    lines(c(tsp[2],tsp[2]+1/tsp[3]), rep(as.numeric(U),2), col=2, lty=3)
  }
  if (back>0) points(save)
  invisible(pred)
}
# Forecast like SAS

tsPredPlot(steel, ar1, back=5, lead=5)

![ARIMA(1, 0, 0)](image)

```r
Box.test(steel, 6, type="Ljung-Box")
```

```
# Box-Ljung test
# data: steel
# X-squared = 12.1546, df = 6, p-value = 0.05861
```

```
Box.test(resid(ar1), 6, type="Ljung-Box")
```

```
# data: resid(ar1)
# X-squared = 2.1702, df = 6, p-value = 0.903
```

tsPredPlot(steel, armal1, back=5, lead=5)

![ARIMA(1, 0, 1)](image)

```
tsPredPlot(steel, armal1, lead=20)
```

```
# Box & Leung test for white noise across, e.g., 6 lags
Box.test(steel, 6, type="Ljung-Box")
```

```
# data: steel
# X-squared = 12.1546, df = 6, p-value = 0.05861
```

```
Box.test(resid(armal1), 6, type="Ljung-Box")
```

```
# data: resid(armal1)
# X-squared = 2.1702, df = 6, p-value = 0.903
```