These notes will cover some of the general discussion. Much of the other effort will be on the handouts, one related to Lowe’s Home Goods Store and other using gasoline price data.

Let’s think first about a time series $y_1, y_2, y_3, \ldots, y_n$. The notation varies. Sometimes there is a start value $y_0$. Often the index is $t$ (rather than $i$) and the series length is noted as $T$ (rather than $n$).

There are no independent variables. The task will be just to explain the time series phenomenon. The CPI kept over many years (and analyzed by itself) would be a time series. The pound-to-dollar exchange rate kept monthly would be a time series.

As another example, we have the Canadian lynx captures from 1821-1934. There are many caveats and side notes. These are from the site http://eom.springer.de/c/c110040.htm.

Here is what these look like:

![Scatterplot of Lynx vs Year](image)

The time-based phenomenon is obvious. But how should we describe these? What descriptions are we seeking? What would be helpful?

Time series work divides into two broad varieties.

The first is called time domain. These attempt to explain $y_t$ in terms of previous $y$’s and/or linear combinations of independent, but overlapping, noise terms. An example of such a model might be
\[ y_t = \mu + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t \]

In this form, the \( \varepsilon \)'s are independent noise terms. This particular example is called AR2, standing for “autoregressive, order 2.”

Another example starts with the independent noise terms and then is set up as

\[ y_t = \mu + \alpha_1 \varepsilon_t + \alpha_2 \varepsilon_{t-1} + \alpha_3 \varepsilon_{t-3} + \zeta_t \]

Here the \( \zeta \)'s are independent noise terms as well. Nearby (in time) \( y \) values are correlated because they share some random components. This particular example is called MA3, standing for “moving average, order 3.”

The AR and MA can be combined. These would be called ARMA models. More generally, we have ARIMA, with the I standing for “integrated.” If the I parameter is 1, then the series will be differenced before analysis. If the I parameter is 2, then it will be differenced twice.

Models do not come with official labels that indicate their varieties. Thus, the first very difficult part of time domain analysis is identifying exactly what kind of ARIMA model you’ve got.

The second variety is frequency domain. This works from the construction of the data as the sum of cosine waves of various frequencies, combined with random noise. A very trivial example would be this:

\[ y_t = \mu + \sum_{j=1}^{5} A_j \cos(\omega_j t + \phi_j) + \varepsilon_t \]

This has five cosine waves.

The \( A_j \) is the amplitude of the \( j^{th} \) wave; if it’s big, then you’ll see that wave more than the others.

The \( \omega_j \) is the frequency of the \( j^{th} \) wave; if \( \omega_j \) is big, then your data series will go through many cycles of this wave. Most of the time, we are interested in low frequency waves.

The \( \phi_j \) is the phase shift of wave \( j \). It’s of little interest by itself, but the fact that the different waves have different phase shifts makes them combine in weird ways.
Frequency domain analysis is critical when the essence of the phenomenon is its frequency behavior. Engineering problems are common here, because we think about the vibrational frequencies of building materials. In music, the frequencies are exactly what we hear.

In thinking about time series, a critical property is *stationarity*. This means that the distribution of $X_t$ (considered in isolation) is exactly the same for every $t$. An immediate consequence is that the mean of $X_t$ and the standard deviation of $X_t$ does not change over time. Moreover any set of same-spaced $k$-tuples has the same distribution for all time.

What does that mean? Pick any positive integer $k$ and then pick any positive integers $n_1, n_2, \ldots, n_k$ with $n_1 < n_2 < \ldots < n_k$. Then $(X_{t+n_1}, X_{t+n_2}, \ldots, X_{t+n_k})$ has the same distribution for every $t$.

The most common (and probably the most believable) time domain model is the AR1, autoregressive of order 1. This is described in the time series pamphlet. The model can be written in several forms.

The model starts with a nonrandom value $X_0$. Thereafter,

$$X_t = \rho X_{t-1} + \varepsilon_t \quad [4a]$$

This says that the value obtained at time $t$ is a multiple of the value at time $t-1$, plus an added random noise term. The set of noise terms $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$ is assumed to be white noise, with a mean of zero. In addition, it is assumed that $\varepsilon_t$ is also independent of $X_0, X_1, \ldots, X_{t-1}$. This model is only useful in the case of stationarity. For reasons of stationarity, as will be made clear below, it is necessary to assume that $-1 < \rho < 1$. Here $\rho$ is called the autoregressive parameter.

It looks a little more believable with a mean term:

$$X_t - \mu = \rho (X_{t-1} - \mu) + \varepsilon_t \quad [4b]$$

In this form, $E X_t = \mu$ at every time point.

Since $[4b]$ can be written as $X_t = \mu (1 - \rho) + \rho X_{t-1} + \varepsilon_t$, you may also see this model in form

$$X_t = \nu + \rho X_{t-1} + \varepsilon_t \quad [4c]$$
Let’s use [4c] to investigate the variance. We will assume that \( \sigma^2 = \text{Var}(\varepsilon_t) \) for every time point \( t \). Assume also that \( \text{Var}(X_t) = \text{Var}(X_{t-1}) = \tau^2 \). Stationarity implies that the variance will be the same at all points in time.

\[
\text{Var}(X_t) = \text{Var}(\nu + \rho X_{t-1} + \varepsilon_t)
\]

\[
= \text{Var}(\rho X_{t-1} + \varepsilon_t) \quad \text{since } \nu \text{ is not random}
\]

\[
= \text{Var}(\rho X_{t-1}) + \text{Var}(\varepsilon_t) \quad \text{since } \varepsilon_t \text{ is independent of } X_{t-1}
\]

\[
= \rho^2 \text{Var}(X_{t-1}) + \sigma^2
\]

With \(-1 < \rho < 1\), we can have \( \text{Var}(X_t) \) the same for every value of \( t \). Then

\[
\tau^2 = \rho^2 \tau^2 + \sigma^2
\]

and the variance of the \( X_t \)'s is related to the noise variance through

\[
\tau^2 = \frac{\sigma^2}{1 - \rho^2}
\]

If we have \( \rho > 1 \) or \( \rho < -1 \), then certainly \( \rho^2 > 1 \). This would have \( \text{Var}(X_t) \) growing to infinity at an exponential rate. This is almost certainly *not* a property that we want a model to have.

**NOTE:** With \( \rho = 1 \), this is a random walk. In any of [4a] or [4b] or [4c] with \( \rho = 1 \), the model is \( X_t = X_{t-1} + \varepsilon_t \). Then

\[
X_1 = X_0 + \varepsilon_1
\]

\[
X_2 = X_1 + \varepsilon_2 = X_0 + \varepsilon_1 + \varepsilon_2
\]

\[
X_3 = X_2 + \varepsilon_3 = X_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3
\]

\[
X_4 = X_3 + \varepsilon_4 = X_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4
\]

and so on

**NOTE:** Why can we not have \( \rho > 1 \)? In terms of just modeling, we do have the freedom to create any model we desire, but \( \rho > 1 \) creates some consequences that we might wish to avoid.

We can examine the sunspot data, too. This shows clear periodic behavior. We are quite sure that this is a set of data for which humans had absolutely no control.
Now for the Lowe’s data set.

Here’s a graph of quarterly sales:

This will lead to a reasonably orderly solution.
We will use GasolineMarket.mpj.

The P variables are “price of.”

Let’s try to regress GasPrice on P_NewCars, P_UsedCars, Population.

We will start with time plots for all variables. These are not shown here, but they all go up with time.

Then we will ask for VIF numbers, the Durbin-Watson statistic, and the four-in-one plot set for the residuals.

Regression Analysis: GasPrice versus P_NewCars, P_UsedCars, Population

The regression equation is

\[
\text{GasPrice} = -80.2 + 1.04 \text{P_NewCars} - 0.421 \text{P_UsedCars} + 0.000326 \text{Population}
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-80.19</td>
<td>20.88</td>
<td>-3.84</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>P_NewCars</td>
<td>1.0364</td>
<td>0.3611</td>
<td>2.87</td>
<td>0.006</td>
<td>87.684</td>
</tr>
<tr>
<td>P_UsedCars</td>
<td>-0.4213</td>
<td>0.2540</td>
<td>-1.66</td>
<td>0.104</td>
<td>82.269</td>
</tr>
<tr>
<td>Population</td>
<td>0.0003263</td>
<td>0.0001203</td>
<td>2.71</td>
<td>0.009</td>
<td>10.253</td>
</tr>
</tbody>
</table>

\[ S = 10.2123 \quad \text{R-Sq} = 89.7\% \quad \text{R-Sq(adj) = 89.0}\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>3</td>
<td>43461</td>
<td>14487</td>
<td>138.91</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>48</td>
<td>5006</td>
<td>104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>48467</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source  DF  Seq SS
P_NewCars 1  42466
P_UsedCars 1  226
Population 1  768

Unusual Observations

<table>
<thead>
<tr>
<th>Obs</th>
<th>P_NewCars</th>
<th>GasPrice</th>
<th>Fit</th>
<th>SE Fit</th>
<th>Residual</th>
<th>St Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>94</td>
<td>84.02</td>
<td>59.59</td>
<td>2.76</td>
<td>24.43</td>
<td>2.48R</td>
</tr>
<tr>
<td>30</td>
<td>97</td>
<td>79.77</td>
<td>59.13</td>
<td>1.62</td>
<td>20.64</td>
<td>2.05R</td>
</tr>
<tr>
<td>32</td>
<td>103</td>
<td>76.00</td>
<td>56.11</td>
<td>3.89</td>
<td>19.90</td>
<td>2.31R</td>
</tr>
<tr>
<td>46</td>
<td>141</td>
<td>71.87</td>
<td>92.31</td>
<td>2.49</td>
<td>-20.44</td>
<td>-2.06R</td>
</tr>
<tr>
<td>52</td>
<td>134</td>
<td>123.90</td>
<td>98.36</td>
<td>4.24</td>
<td>25.54</td>
<td>2.75R</td>
</tr>
</tbody>
</table>

R denotes an observation with a large standardized residual.

Durbin-Watson statistic = 0.453959

There are many things wrong here!
The fit is good, but the VIF numbers are terrible. In addition, the residuals seem to violate assumptions.

The Durbin-Watson statistic, as we’ll see, is a prime indicator that there are “issues.”

Let’s try looking at the residuals, after adjusting out for time. It’s a bit of work to set this up, but here’s what we get:

**Regression Analysis: GasPrice|Year versus P_NewCars|Year, P_UsedCars|Year, ...**

The regression equation is

\[
\text{GasPrice|Year} = -0.00 + 0.937 \text{ P_NewCars|Year} - 0.381 \text{ P_UsedCars|Year} - 0.000070 \text{ Popn|Year}
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.00</td>
<td>1.411</td>
<td>-0.00</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>P_NewCars</td>
<td>Year</td>
<td>0.9368</td>
<td>0.3946</td>
<td>2.37</td>
<td>0.022</td>
</tr>
<tr>
<td>P_UsedCars</td>
<td>Year</td>
<td>-0.3813</td>
<td>0.2613</td>
<td>-1.46</td>
<td>0.151</td>
</tr>
<tr>
<td>Popn</td>
<td>Year</td>
<td>-0.0000699</td>
<td>0.0006568</td>
<td>-0.11</td>
<td>0.916</td>
</tr>
</tbody>
</table>

\[ S = 10.1725 \quad \text{R-Sq} = 20.6\% \quad \text{R-Sq(adj)} = 15.6\% \]

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>3</td>
<td>1289.1</td>
<td>429.7</td>
<td>4.15</td>
<td>0.011</td>
</tr>
<tr>
<td>Residual Error</td>
<td>48</td>
<td>4967.0</td>
<td>103.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>6256.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Seq SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_NewCars</td>
<td>Year</td>
<td>1</td>
</tr>
<tr>
<td>P_UsedCars</td>
<td>Year</td>
<td>1</td>
</tr>
<tr>
<td>P_popn</td>
<td>Year</td>
<td>1</td>
</tr>
</tbody>
</table>
Unusual Observations
<p>|</p>
<table>
<thead>
<tr>
<th>Obs</th>
<th>P_NewCars</th>
<th>Year</th>
<th>GasPrice</th>
<th>Year</th>
<th>Fit</th>
<th>SE Fit</th>
<th>Residual</th>
<th>St Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>0.3</td>
<td>27.93</td>
<td>3.77</td>
<td>2.77</td>
<td>24.16</td>
<td>2.47</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.6</td>
<td>21.78</td>
<td>1.75</td>
<td>1.86</td>
<td>20.03</td>
<td>2.00</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>2.3</td>
<td>14.22</td>
<td>-4.11</td>
<td>4.61</td>
<td>18.33</td>
<td>2.02</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>-13.4</td>
<td>24.15</td>
<td>-2.87</td>
<td>4.22</td>
<td>27.02</td>
<td>2.92</td>
<td>R</td>
<td></td>
</tr>
</tbody>
</table>

R denotes an observation with a large standardized residual.

Durbin-Watson statistic = 0.455668

The Durbin-Watson statistic says that we have not succeeded. The $R^2$ took a big drop, but we are not concerned; we are still searching for a valid model. The four-in-one picture also indicates that we are still searching:

There are a number of corrections that can be tried.
On the Durbin-Watson statistic, the tables are given as lower cutoffs only. We are nearly always interested in testing for positive autocorrelation, and we reject only for small values of Durbin-Watson. The table has three items of entry.

* $n$, the sample size
* $k$, the number of predictors (perhaps as $\Lambda = k + 1$)
* $\alpha$, the significance level

The table entries can be described as $Q_L$ and $Q_U$.

If Durbin-Watson $< Q_L$, then reject $\rho = 0$.
If $Q_L \leq$ Durbin-Watson $\leq Q_U$, the test is undecided.
If Durbin-Watson $> Q_U$, then accept $\rho = 0$. 