When working with time series data, issues arise that can’t be properly addressed with
the techniques usually used in analyzing cross-sectional data. In these situations, it is
more appropriate to use specialized time series models.

One of the most important aspects of a time series model is distributed lags, the idea that
something happening in one period may have impacts for a number of periods afterward.
As a result, time series models often include lagged values of variables.

Distributed Lag Models

A simple version of a model with simple lags is given in this example where the
dependent variable, Y, depends on two explanatory variables. The first, X1, impacts Y
after a lag of one period while the second, X2, impacts Y immediately. The model of Y
would be:

\[ Y_t = \beta_0 + \beta_1 X_{1,t-1} + \beta_2 X_{2,t} + \epsilon_t \]

A more complicated version occurs when the impact of a change in an explanatory
variable occurs over a number of time periods.

For example, the effect of a change in the money supply might have an impact on GDP
over a period of time rather than immediately because it takes the economy time to react.

In such a case, an appropriate model might be:

\[ Y_t = a_0 + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_p X_{t-p} + \epsilon_t \]

This sort of model is a distributed lag model, in which the effect of changes in X are
distributed over a number of time periods. In the case shown above, the effect of a
change in X lasts for p time periods.

If effects deteriorate over time, then \( \beta_t \) should decrease in magnitude as t increases. That
is, things that happened further in the past should have less of an impact.
So, for example, the change in the growth rate of GDP might be regressed on the change in the growth rate of the money supply for the past eight quarters to see what the relationship seems to be. We might expect that the effect of recent changes in the money supply will be greater than the effect of earlier changes.

There are several problems with estimating an equation such as this with ordinary least squares (OLS).

1. Multicollinearity in the lagged values of \( X \). This will make the estimated coefficients nonrobust and will make it tricky to determine exactly what the impact of certain explanatory variables is.

2. Because of the multicollinearity and lack of precise and robust estimates, the estimated coefficients might not decline as the length of the lag increases. This can lead to some odd and difficult-to-explain coefficient estimates.

3. The degrees of freedom will decrease as the number of lags included increases, which can be a real problem with time series data because the time series might be a bit short and you can’t easily expand the data set overnight.

**Koyck Lags**

Koyck lags are an attempt to deal with some of the problems with distributed lag models.

A Koyck lag model uses the following form for coefficients on lagged values:

\[
\beta_i = \beta_0 \lambda^i \quad \text{for} \quad 0 < \lambda < 1, \ i \geq 0
\]

Where \( i \) is the length of the lag and \( \lambda \) is a measure of how quickly or slowly the effect deteriorates.

Plugging this into the lagged model shown above and letting the lags go on for an infinite amount of time gives us

\[
Y_t = a_0 + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \varepsilon_t
\]

gives us:

\[
Y_t = a_0 + \beta_0 (X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \ldots) + \varepsilon_t
\]

So now, the only parameters that need to be estimated are \( a_0, \beta_0 \), and \( \lambda \).
The problem is that the equation is non-linear in these coefficients, what with the higher order terms of $\lambda$ and $\beta_0$ and $\lambda$ being multiplied together and all that, so this presents a bit of a problem.

However, a little clever algebra comes in handy here. Consider that

$$\lambda Y_{t-1} = \lambda a_0 + \beta_0(\lambda X_{t-1} + \lambda^2 X_{t-2} + \lambda^3 X_{t-3} + \ldots) + \epsilon_{t-1}$$

Now, subtracting the two equations (which is always a fun trick) gives us

$$Y_t = a_0 + \beta_0(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \ldots) + \epsilon_t$$

$$\lambda Y_{t-1} = \lambda a_0 + \beta_0(\lambda X_{t-1} + \lambda^2 X_{t-2} + \lambda^3 X_{t-3} + \ldots) + \lambda \epsilon_{t-1}$$

$$Y_t - \lambda Y_{t-1} = a_0 - \lambda a_0 + \beta_0 X_t + \epsilon_t - \lambda \epsilon_{t-1}$$

Which can be rewritten as

$$Y_t = a_0 + \beta_0 X_t + \lambda Y_{t-1} + u_t$$

Where

$$a_0 = a_0 - \lambda a_0$$

$$u_t = \epsilon_t - \lambda \epsilon_{t-1}$$

The equation

$$Y_t = a_0 + \beta_0 X_t + \lambda Y_{t-1} + u_t$$

is called an *autoregressive* equation because the variable $Y$ is regressed on a lagged value of itself.

Estimation of this equation carries with it a number of problems. In particular, because the variable $X$ influences $Y$ in one time period, it indirectly affects all of the subsequent $Y$’s as well.

To state this more clearly, $X_{t-1}$ affects the value of $Y_{t-1}$, but because the value of $Y_{t-1}$ affects the value of $Y_t$, the $X_{t-1}$ also affects $Y_t$ according to the parameter $\lambda$. This may not be something you want to have reflected in your model.
**Koyck Example**

Imagine that household consumption in one year is dependent on disposable income in that year as well as on disposable income in some preceding years:

\[ C_t = f(YD_t, YD_{t-1}, YD_{t-2}, \ldots) \]

Which might be written as

\[ C_t = \gamma_0 + \gamma_1 YD_t + \gamma_2 YD_{t-1} + \gamma_3 YD_{t-2} + \ldots + \xi_t \]

If the coefficients on YD decrease over time according to a geometric pattern, this could be rewritten as

\[ C_t = \alpha_0 + \beta_0 YD_t + \lambda C_{t-1} + u_t \]

So that consumption is modeled in an autoregressive fashion. The results given in the book for the U.S. economy from 1964 to 1994 are:

\[ \hat{C}_t = -38.11 + 0.52YD_t + 0.46C_{t-1} \quad (\alpha_0=-38.11, \beta_0=0.52, \lambda=0.46) \]

To thing about what this means in the context of the original equation

\[ C_t = \gamma_0 + \gamma_1 YD_t + \gamma_2 YD_{t-1} + \gamma_3 YD_{t-2} + \ldots + \xi_t \]

We need to realize that the constant term

\[ \gamma_0 = \frac{\alpha_0}{1-\lambda} \] (this is an infinite series sort of thing)

and that
\[ \gamma_1 = \beta_0 \]
\[ \gamma_2 = \beta_0 \lambda^1 \]
\[ \gamma_3 = \beta_0 \lambda^2 \]
\[ \gamma_4 = \beta_0 \lambda^3 \]

and so on so that

\[ \hat{C}_t = -70.57 + 0.52YD_t + 0.24YD_{t-1} + 0.11YD_{t-2} + 0.05YD_{t-3} + ... \]

So the coefficients decline smoothly as would be expected.

To see where these coefficients come from, consider the following. We start out with:

\[ C_t = \alpha_0 + \beta_0 YD_t + \lambda C_{t-1} + u_t \]

If we lag this one period we get

\[ C_{t-1} = \alpha_0 + \beta_0 YD_{t-1} + \lambda C_{t-2} + u_{t-1} \]

Substituting this into the above equation we get

\[ C_t = \alpha_0 + \beta_0 YD_t + \lambda (\alpha_0 + \beta_0 YD_{t-1} + \lambda C_{t-2} + u_{t-1}) + u_t \]

We then note that

\[ C_{t-2} = \alpha_0 + \beta_0 YD_{t-2} + \lambda C_{t-3} + u_{t-2} \]

And we substitute this to get

\[ C_t = \alpha_0 + \beta_0 YD_t + \lambda (\alpha_0 + \beta_0 YD_{t-1} + \lambda (\alpha_0 + \beta_0 YD_{t-2} + \lambda C_{t-3} + u_{t-2}) + u_{t-1}) + u_t \]

If we were to continue substituting for increasingly lagged values of \( C_{t-j} \) we would wind up with:

\[ C_t = \alpha_0 (1 + \lambda + \lambda^2 + \lambda^3 + ... ... ) + \beta_0 (YD_t + \lambda YD_{t-1} + \lambda^2 YD_{t-2} + \lambda^3 YD_{t-3} + ... ... ) + \text{error terms} \]
One infinite sum relationship is that, for $0 < \lambda < 1$,

$$\sum_{i=0}^{\infty} \lambda^i = \frac{1}{1-\lambda}.$$ 

For example, if we have $\lambda = 0.5$, this would be $1 + 0.5 + 0.25 + 0.125 + \ldots = 2$

Applying this to the above equation, we get

$$C_t = \alpha_0 \frac{1}{1-\lambda} + \beta_0(YD_t + \lambda YD_{t-1} + \lambda^2 YD_{t-2} + \lambda^3 YD_{t-3} + \ldots) + \text{error terms}$$

$$C_t = \alpha_0 \frac{1}{1-\lambda} + \beta_0 YD_t + \beta_0 \lambda YD_{t-1} + \beta_0 \lambda^2 YD_{t-2} + \beta_0 \lambda^3 YD_{t-3} + \ldots + \text{error terms}$$

and this gives us

$$\hat{C}_t = -70.57 + 0.52YD_t + 0.24YD_{t-1} + 0.11YD_{t-2} + 0.05YD_{t-3} + \ldots$$

Now, what if this relationship was to be estimated with a normal OLS model?

The book gives the results for 1964-94 of

$$\hat{C}_t = -142.64 + 0.93YD_t + 0.11YD_{t-1} - 0.05YD_{t-2} - 0.03YD_{t-3} + \ldots$$

These coefficients become a bit weird after a while and so most people choose to use Koyck lags to model time series data because the results are smoother and can be more easily explained.

*(In fact, the numbers reported above are from the third edition of Studenmund and are a bit different from those reported in the fourth edition. One can only wonder at the reason for the difference between the two editions of this excellent text.)*

**A Bit of a Math Exercise**

According to the text, an estimate of the long run multiplier, or the long run impact of a one time increase in income on consumption after all the lagged effects have been felt, can be calculated based on the equation from above:
\[ C_t = \alpha_0 + \beta_0 YD_t + \lambda C_{t-1} + u_t \]

which is estimated to be
\[ \hat{C}_t = -38.11 + 0.52 YD_t + 0.46 C_{t-1} \]

The book claims that the calculation of the long run multiplier is:
\[ \hat{M}_{LR} = \hat{\beta}_0 \left[ \frac{1}{1 - \hat{\lambda}_0} \right] = 0.52 \left[ \frac{1}{1 - 0.46} \right] = 0.96, \]

which suggests that an increase in income of $1 will increase consumption by $0.96 over all of the future.

Can you explain where this equation comes from?

Well, the immediate impact of a $1 increase in YD on C is $0.52, from the equation above. The impact of this $0.52 increase in this period’s consumption on next period’s consumption is $0.52 \times 0.46 = $0.2392 and this process continues indefinitely.

In terms of the equation \( C_t = \alpha_0 + \beta_0 YD_t + \lambda C_{t-1} + u_t \), this accumulated impact is
\[ \beta_0 + \lambda \beta_0 + \lambda^2 \beta_0 + \lambda^3 \beta_0 + \lambda^4 \beta_0 + \ldots = \frac{\beta_0}{1 - \lambda}. \]

**Serial Correlation and Koyck Distributed Lags**

Koyck lags are pretty, but they suffer from a high probability of serial correlation. The reason for this goes back to the equations:
\[ Y_t = \alpha_0 + \beta_0 \left( X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \lambda^3 X_{t-3} + \ldots \right) + \varepsilon_t \quad (12.4 \text{ in } 4^{th} \text{ edition}) \]
\[ \lambda Y_{t-1} = \lambda \alpha_0 + \beta_0 \left( \lambda X_{t-1} + \lambda^2 X_{t-2} + \lambda^3 X_{t-3} + \ldots \right) + \lambda \varepsilon_{t-1} \quad (12.6 \text{ in } 4^{th} \text{ edition}) \]
and if you subtract the second from the first you get:

\[ Y_t - \lambda Y_{t-1} = a_0 - \lambda a_0 + \beta_0 X_t + \varepsilon_t - \lambda \varepsilon_{t-1} \]

which the book rewrites as:

\[ Y_t = a_0 - \lambda a_0 + \beta_0 X_t + \lambda Y_{t-1} + \varepsilon_t - \lambda \varepsilon_{t-1} \]

\[ Y_t = \alpha_0 + \beta_0 X_t + \lambda Y_{t-1} + u_t \]

But this error term, \( u_t \), is equal to:

\[ u_t = \varepsilon_t - \lambda \varepsilon_{t-1} \]

And the error term from the previous period, \( u_{t-1} \), is equal to:

\[ u_{t-1} = \varepsilon_{t-1} - \lambda \varepsilon_{t-2} \]

So, the two error terms, \( u_t \) and \( u_{t-1} \), both depend on the same term, \( \varepsilon_{t-1} \), and are almost certain to be correlated. Correlated error terms are a violation of one of the classical assumptions and are a potential problem.

Serial correlation in the Koyck distributed lag model is even worse than usual because the Koyck model includes a lagged dependent variable (\( Y_{t-1} \)) as an explanatory variable. Specifically:
1. The usual corrections for serial correlation won’t work.
2. The estimated coefficients will be biased.
3. Estimates of standard errors of coefficients are biased so that hypothesis testing is invalid.
4. Estimates of standard errors of residuals are biased so that the Durbin-Watson \( d \) test is potentially invalid.

The second of these reasons, that the estimated coefficients will be biased, is worth an explanation. The basic Koyck model is:

\[ Y_t = \alpha_0 + \beta_0 X_t + \lambda Y_{t-1} + u_t \]

\[ u_t = \varepsilon_t - \lambda \varepsilon_{t-1} \]

Consider the following two equations:

\[ Y_t = \alpha_0 + \beta_0 X_t + \lambda Y_{t-1} + \varepsilon_t - \lambda \varepsilon_{t-1} \]
\[ Y_{t-1} = \alpha_0 + \beta_0 X_{t-1} + \lambda Y_{t-2} + \varepsilon_{t-1} - \lambda \varepsilon_{t-2} \]
From the second of these equations, \( Y_{t-1} \) depends on the error term \( \varepsilon_{t-1} \). For example, if \( \varepsilon_{t-1} \) is positive, then \( Y_{t-1} \) will be larger than it would have been otherwise.

Looking at the first of the equations, this means that one of the explanatory variables from the first of these two equations, \( Y_{t-1} \), is correlated with the error term, which contains \( \varepsilon_{t-1} \). This is a violation of one of the classical assumptions and will bias the estimate of \( \lambda \).

**A Modified Durbin-Watson Test for Serial Correlation in Koyck Models**

The Durbin-Watson test for serial correlation is potentially invalid for equations containing lagged dependent variables as explanatory variables. This means that you can’t use the usual D-W test for Koyck models.

Of course, some clever person has come up with a modified version that is valid for large sample Koyck models. The test statistic, which is normally distributed, is:

\[
\begin{align*}
  h &= (1 - 0.5d) \left( \frac{n}{1 - nS_\lambda^2} \right) \\
  \text{Where} & \\
  d & \text{is the Durbin-Watson statistic} \\
  n & \text{is the sample size} \\
  S_\lambda^2 & \text{is the square of the estimated standard error of } \lambda \text{ from the Koyck model}
\end{align*}
\]

The null hypothesis is that there is no serial correlation. The alternative is that there is. So, the bigger the absolute value of \( h \) is, the more likely it is that there is serial correlation. Specifically, if \( |h| > 1.96 \) then the null hypothesis of no serial correlation would be rejected.

You might recall that when the D-W statistic \( d \) is close to 2, the implication is that there is no serial correlation.

When \( d = 2 \), \( h = 0 \). Don’t reject the null hypothesis of no serial correlation.

I have no sense of the other variables and what role they play here.

This test has two problems.
1. It doesn’t work when there’s more than one lagged dependent variable.
2. It is undefined when \( 1 - nS_\lambda^2 < 0 \).
A Lagrange Multiplier Test for Serial Correlation in Koyck Models

This is an alternative test that looks at the extent to which the explanatory variables and lagged residuals explain the residuals in the model. The process is rather like the Park test for heteroskedasticity in that you estimate a primary model, save the residuals, and then estimate a second model involving the residuals.

To do this, do a regression of the primary model and save the residuals.

\[ Y_t = \alpha_0 + \beta_0 X_t + \lambda Y_{t-1} + \varepsilon_t - \lambda e_{t-1} \]

Then regress the residuals on the explanatory variables and the lagged residuals.

\[ e_t = a_0 + a_1 X_t + a_2 Y_{t-1} + a_3 e_{t-1} + u_t \]

The important question is whether the estimated value of a_3 is significantly different from zero. If it is zero, there is no serial correlation, if it is different from zero then there is correlation between the residual in period t and the residual in period t-1 and there is serial correlation. To determine whether the correlation is significant, calculate the test statistic

\[ LM = nR^2 \]

where

- n is the number of observations
- \( R^2 \) is the \( R^2 \) from the second regression

Under the null hypothesis of no serial correlation, this test statistic follows a chi-square distribution with one degree of freedom and the value of this test statistic should be small. A big value for LM would suggest that there is serial correlation.

Solutions to Serial Correlation in Koyck Distributed Lag Models?
They’re tricky and beyond the scope of this course.

Granger Causality is Cool

Granger causality allows a researcher to test an assumption about causality in time series data.

For example, you could test the causal relationship between monetary growth and GDP. That is, does GDP growth lead to increases in the money supply or does growth of the money supply cause GDP growth?
There’s a better example…

The test itself is for a specific type of causality called Granger causality. We say that one factor “Granger causes” another factor. Basically it’s based on changes in one consistently leading changes in the other.

You’ve still got to be careful about inferring actual causality. Umbrellas might Granger cause rain, but they don’t actually cause rain. Also, rain might Granger cause umbrellas to open, but it doesn’t actually cause umbrellas to open.

To put this slightly differently, image the old story of man looking into a cow pasture through a crack in a fence. Everyone once in a while a cow walks by and the man sees, in order, the head, the body and the tail. After thinking for a while he exclaims, “It’s obvious! The head causes the tail!” Actually, the head may Granger cause the tail but, as always, it’s important to understand the underlying process.

Anyway, here’s the deal with Granger causality:

If you have two variables, Y and B, and you want to see if B Granger causes Y, you would do a regression of Y on lagged values of Y and lagged values of B and then test whether the coefficients on the lagged B values are jointly equal to zero. If you reject this null hypothesis, then the conclusion is that B Granger causes Y.

\[ Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 B_{t-1} + \epsilon_t \]

If the coefficient on B_{t-1} is significantly different from zero, the implication is that B Granger causes Y. The model might also include additional lagged terms.

You can also reverse this to test whether Y Granger causes B.

\[ B_t = \beta_0 + \beta_1 B_{t-1} + \beta_2 Y_{t-1} + \epsilon_t \]

If the coefficient on Y_{t-1} is significantly different from zero, the implication is that Y Granger causes B.

In case you’re wondering, it can be the case that Y Granger causes B and that B also Granger causes Y.

The funnest Granger causality example is from:

Author(s): Thurman, Walter N.; Fisher, Mark E.
Title: Chickens, Eggs, and Causality, or Which Came First?
Spurious Correlation and Nonstationarity

“One problem with time-series data is that independent variables can appear to be more significant than they actually are if they have the same underlying trend as the dependent variable. In a country with rampant inflation, for example, almost any nominal variable will appear to be highly correlated with all other nominal variables…

“Such a problem is an example of spurious correlation, a strong relationship between two or more variables that is caused by a statistical fluke or by the nature of the specification of the variables, not by a real underlying causal relationship.”

The focus of this section is on spurious correlations resulting from non-stationary time series.

Definitions of Stationarity and Non-Stationarity

A stationary time series is one whose basic properties (especially the mean value and the standard deviation) don’t change over time.

A non-stationary time series is one whose basic properties exhibit some consistent upward or downward trend.

For example, if you look at price levels in a country that has experienced consistent inflation over a long period of time, price levels would appear to be non-stationary. If, however, you look at hours worked per week or interest rates, these would likely appear to be stationary.

Formalities? You want Formalities?
OK, to be formal about this, a time series variable $X_t$ is stationary if:
1. the mean of $X_t$ is constant over time
2. the variance of $X_t$ is constant over time
3. the correlation between $X_t$ and $X_{t-k}$ (the autocorrelation function) depends on the length of the lag, $k$, but not on anything else

Implications of Non-Stationarity

The main implications of analysis of non-stationary time series is that the $R^2$ values will be inflated, as will the t-stats.

Basically, things will seem to explain each other much better than they really do simply because they’re all growing together as a result of the same process.
Testing for Non-Stationarity

1. Visual inspection
Does a graph of the data seem to show that it is heteroskedastic or growing out of control? Then it may be non-stationary.

2. Autocorrelation functions
Do the autocorrelation functions (the correlation between $Y_t$ and $Y_{t+k}$) tend to zero as the number of lags, $k$, increases? If they do then the time series is probably stationary. If not, then it is nonstationary.

3. The Dickey-Fuller test
To do this test, you need to first difference the variable of interest. That is, you need to calculate its value less its lagged value:

$$\Delta Y_t = (Y_t - Y_{t-1})$$

And then regress this on the previous value of the variable:

$$\Delta Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$$

The Dickey-Fuller test is about the value of $\beta_1$.

The null hypothesis is that the series is non-stationary, basically meaning that $\beta_1$ is equal to or greater than zero. The alternative hypothesis is that the series is stationary, or that $\beta_1$ is less than zero.

$$H_0: \beta_1 = 0$$
$$H_A: \beta_1 < 0$$

You need to be careful about critical values for this test as they don’t follow the standard values for a t-test.

Adjusting for Non-Stationarity
The usual way of dealing with non-stationarity is to analyze first differences rather than to do analysis of the time series itself.

The problems with this are:
1. It changes the meaning of the estimated relationship
2. It ignores information about the long run trend of the variable