Intermediate Social Statistics Hilary 2009 Lecture 8: Time Series

Raymond Duch

Nuffield College

March 8, 2011

What we'll talk about today

1. Some time-series basics

What we'll talk about today

1. Some time-series basics
2. Memory in time series analysis
   - Memory and autocorrelation
   - One variant of non-stationarity: Deterministic trends
   - Another variant of non-stationarity: Unit roots

What we'll talk about today

1. Some time-series basics
2. Memory in time series analysis
   - Memory and autocorrelation
   - One variant of non-stationarity: Deterministic trends
   - Another variant of non-stationarity: Unit roots
3. The consequences of memory: Trends and spurious regressions
   - A real example with deterministic trends
   - A hypothetical example with stochastic trends

What we'll talk about today

1. Some time-series basics
2. Memory in time series analysis
   - Memory and autocorrelation
   - One variant of non-stationarity: Deterministic trends
   - Another variant of non-stationarity: Unit roots
3. The consequences of memory: Trends and spurious regressions
   - A real example with deterministic trends
   - A hypothetical example with stochastic trends
4. One solution to the spurious regression problem: First-differencing the dependent variable

What we'll talk about today

1. Some time-series basics
2. Memory in time series analysis
   - Memory and autocorrelation
   - One variant of non-stationarity: Deterministic trends
   - Another variant of non-stationarity: Unit roots
3. The consequences of memory: Trends and spurious regressions
   - A real example with deterministic trends
   - A hypothetical example with stochastic trends
4. One solution to the spurious regression problem: First-differencing the dependent variable
5. Another solution to the spurious regression problem: The lagged dependent variable

Outline of the session

1. Some time-series basics
2. Memory in time series analysis
   - Memory and autocorrelation
   - One variant of non-stationarity: Deterministic trends
   - Another variant of non-stationarity: Unit roots
3. The consequences of memory: Trends and spurious regressions
   - A real example with deterministic trends
   - A hypothetical example with stochastic trends
4. One solution to the spurious regression problem: First-differencing the dependent variable
5. Another solution to the spurious regression problem: The lagged dependent variable

The consequences of changing subscripts

Most econometric treatments of regression models use a generic notation in which the subscript "*i*" represents an individual case. In time series notation, individual cases are represented with the subscript "*t*" and the numeric value of "*t*" represents the temporal order in which the cases occurred and this ordering is very likely to matter. Consider the following OLS population model written in the notation that is standard in the literature:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

If the data of interest were time series data, we would rewrite this model as:

$$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

The differences between cross-sections and time series

- Consider the following two scenarios, one dealing with a cross section, the other with time series analysis.

The differences between cross-sections and time series

- Consider the following two scenarios, one dealing with a cross section, the other with time series analysis.
- Both scenarios begin the same way: You're sound asleep one night, and God whispers to you in a dream. He (or She, if you prefer) says: "Ray, I'm going to tell you how I made the social world. The particular dependent variable you're interested in? The one you call Y? It has just three causes, and I'm going to tell you exactly what they are."

The differences between cross-sections and time series

- Consider the following two scenarios, one dealing with a cross section, the other with time series analysis.
- Both scenarios begin the same way: You're sound asleep one night, and God whispers to you in a dream. He (or She, if you prefer) says: "Ray, I'm going to tell you how I made the social world. The particular dependent variable you're interested in? The one you call Y? It has just three causes, and I'm going to tell you exactly what they are."
- Then God tells you the three causes: X_1, X_2, and X_3.

The differences between cross-sections and time series

- Consider the following two scenarios, one dealing with a cross section, the other with time series analysis.
- Both scenarios begin the same way: You're sound asleep one night, and God whispers to you in a dream. He (or She, if you prefer) says: "Ray, I'm going to tell you how I made the social world. The particular dependent variable you're interested in? The one you call Y? It has just three causes, and I'm going to tell you exactly what they are."
- Then God tells you the three causes: X_1, X_2, and X_3.
- This is where the two scenarios diverge.

In the cross-sectional case

- Gather data if necessary. Open data set. Remember the variables God told you.

In the cross-sectional case

- Gather data if necessary. Open data set. Remember the variables God told you.
- Estimate via OLS.

In the cross-sectional case

- Gather data if necessary. Open data set. Remember the variables God told you.
- Estimate via OLS.
- Examine residuals to check that God wasn't lying to you.

In the time-series case

- Gather data if necessary. Open data set. Remember the variables God told you.

In the time-series case

- Gather data if necessary. Open data set. Remember the variables God told you.
- Get very frustrated, because there's a lot you still don't know.

Memory and lags

Aside from changing a subscript from an i to a t, what's so different about time series modeling? Let's focus on one particular feature of time-series analysis that sets it apart from modeling cross-sectional data.
Let's pretend that the time-series scenario above is a model of presidential popularity, and assume that the data are in monthly form:

$$Popularity_t = \alpha + \beta_1 Economy_t + \beta_2 Peace_t + u_t$$

where "Economy" and "Peace" refer to some measures of the health of the national economy and international peace, respectively.

More ...

- In the model from the previous slide, a president's popularity in any given month t is a function of that month's economy and that month's level of international peace, plus some random error term, and nothing else, at any points in time.

More ...

- In the model from the previous slide, a president's popularity in any given month t is a function of that month's economy and that month's level of international peace, plus some random error term, and nothing else, at any points in time.
- What about last month's economic shocks, or the war that ended three months ago? They are nowhere to be found in this equation, which means quite literally that they can have no effect on a president's popularity ratings in this month.

More ...

- Every month—according to this model—the public starts from scratch evaluating the president, as if to say, on the first of the month: "Okay, let's just forget about last month.

More ...

- Every month—according to this model—the public starts from scratch evaluating the president, as if to say, on the first of the month: "Okay, let's just forget about last month.
- Instead, let's check this month's economic data, and also this month's international conflicts, and render a verdict on whether the president is doing a good job or not."

More ...

- Every month—according to this model—the public starts from scratch evaluating the president, as if to say, on the first of the month: "Okay, let's just forget about last month.

- Instead, let's check this month's economic data, and also this month's international conflicts, and render a verdict on whether the president is doing a good job or not."

- There is no memory from month to month whatsoever. Every independent variable has an immediate impact, and that impact lasts exactly one month, after which the effect immediately dies out entirely. (This is preposterous, of course!)

Is this a problem?

- If we are convinced that at least some past values of the economy still have effects today, and if at least some past values of international peace still have effects today, but we instead only estimate the contemporary effects (from period t), then we have committed omitted variables bias—which, of course, is one of the most serious mistakes a social scientist can make.

Is this a problem?

- If we are convinced that at least some past values of the economy still have effects today, and if at least some past values of international peace still have effects today, but we instead only estimate the contemporary effects (from period t), then we have committed omitted variables bias—which, of course, is one of the most serious mistakes a social scientist can make.
- Failing to account for how past values of our independent variables might affect current values of our dependent variable is a serious issue in time series observational studies, and nothing quite like this issue exists in the cross-sectional world. In time series analysis, even if we know that Y is caused by X_1 and X_2, we still have to worry about how many past lags of X_1 and X_2 might affect Y.

Oh, give me lags, lots of lags...

So maybe we should just do this:

$$Popularity_t = \alpha + \beta_1 Economy_t + \beta_2 Economy_{t-1} + \beta_3 Economy_{t-2} +$$

$$\beta_4 Economy_{t-3} + \beta_5 Peace_t \beta_6 Peace_{t-1} \beta_7 Peace_{t-2} + \beta_8 Peace_{t-3} + u_t$$

This is, indeed, one possible solution to the question of how to incorporate the lingering effects of the past on the present. But the model is getting a bit unwieldy, with lots of parameters to estimate. And there are other problems, too, like the issue of how many lags to specify. Once you get a Y with several different Xs and not too many data points (say, with annual data), this becomes very unwieldy indeed.

Outline of the session

Autocorrelation: How memory is embodied in time series

Recall the assumptions about OLS, and specifically about autocorrelation. For the model:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

We assume that:

$$Cov[u_t, u_{t+k}] = 0 \forall k \neq 0$$

But with autocorrelation, this assumption is violated. The most common form of autocorrelation is what is known as a first-order autoregressive (AR(1)) process, where:

$$u_t = \rho u_{t-1} + e_t$$

where e_t is white noise and the parameter $\rho < |1|$. It (ρ) embodies the "memory" of a series.

Autocorrelation and stationarity

In this model, what is the influence of time t-2 on time t? We can substitute and get to:

$$\varepsilon_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \cdots$$

A "stationary" process is one where ρ is strictly less than 1. What happens if it were equal to 1, or greater than 1?

Correcting for autocorrelation versus accounting for memory in time series

- In the standard econometric literature, detecting autocorrelation is easy—which is true—and the corrections for it are similarly easy.

Correcting for autocorrelation versus accounting for memory in time series

- In the standard econometric literature, detecting autocorrelation is easy—which is true—and the corrections for it are similarly easy.
- ", robust" econometrics?

Correcting for autocorrelation versus accounting for memory in time series

- In the standard econometric literature, detecting autocorrelation is easy—which is true—and the corrections for it are similarly easy.
- ", robust" econometrics?
- But this ignores the three causes of autocorrelation:

Correcting for autocorrelation versus accounting for memory in time series

- In the standard econometric literature, detecting autocorrelation is easy—which is true—and the corrections for it are similarly easy.
- “, robust” econometrics?
- But this ignores the three causes of autocorrelation:
    1. Poor specification (of lag lengths).

Correcting for autocorrelation versus accounting for memory in time series

- In the standard econometric literature, detecting autocorrelation is easy—which is true—and the corrections for it are similarly easy.
- ", robust" econometrics?
- But this ignores the three causes of autocorrelation:
  1. Poor specification (of lag lengths).
  2. Omitted-variables (a different kind of specification error). Most often, the variables that we omit have memories—i.e., X_t and X_{t-1} are correlated.

Correcting for autocorrelation versus accounting for memory in time series

- In the standard econometric literature, detecting autocorrelation is easy—which is true—and the corrections for it are similarly easy.
- ", robust" econometrics?
- But this ignores the three causes of autocorrelation:
  1. Poor specification (of lag lengths).
  2. Omitted-variables (a different kind of specification error). Most often, the variables that we omit have memories—i.e., X_t and X_{t-1} are correlated.
  3. Non-stationarity.

Stationarity and non-stationarity

- A stationary time series is one whose statistical properties—typically its mean and variance—do not depend on time.

Stationarity and non-stationarity

- A stationary time series is one whose statistical properties—typically its mean and variance—do not depend on time.

- More casually, some refer to a stationary series as "mean reverting"—as in, a series is stationary if, over the long run, it reverts to a stable, long-run mean. A different way of saying this is that the long-run forecast of a series into $t + k$ time is its mean.

Stationarity and non-stationarity

- A stationary time series is one whose statistical properties—typically its mean and variance—do not depend on time.

- More casually, some refer to a stationary series as "mean reverting"—as in, a series is stationary if, over the long run, it reverts to a stable, long-run mean. A different way of saying this is that the long-run forecast of a series into $t + k$ time is its mean.

- The notion of the variance being dependent on time receives less attention, but should not be ignored. One implication of the requirement that a stationary series have a stable, long-run variance is that all series that are bounded at both low and high ends should be stationary (because they cannot have ever-growing variances).

Stationarity and non-stationarity

- A stationary time series is one whose statistical properties—typically its mean and variance—do not depend on time.
- More casually, some refer to a stationary series as "mean reverting"—as in, a series is stationary if, over the long run, it reverts to a stable, long-run mean. A different way of saying this is that the long-run forecast of a series into $t + k$ time is its mean.
- The notion of the variance being dependent on time receives less attention, but should not be ignored. One implication of the requirement that a stationary series have a stable, long-run variance is that all series that are bounded at both low and high ends should be stationary (because they cannot have ever-growing variances).
- We often refer to non-stationary time series as having "trends." There are two types of such series—one that's pretty easy to spot, the other that sometimes is not.

Deterministic trends

- A time series with a deterministic trend—sometimes called a "trend stationary" series—simply increases (or decreases) ad infinitum.

Deterministic trends

- A time series with a deterministic trend—sometimes called a "trend stationary" series—simply increases (or decreases) ad infinitum.
- They have neither a stable long-run mean nor a stable long-run variance. And there is no meaningful long-run "forecast" of such a series.

Deterministic trends

- A time series with a deterministic trend—sometimes called a "trend stationary" series—simply increases (or decreases) ad infinitum.
- They have neither a stable long-run mean nor a stable long-run variance. And there is no meaningful long-run "forecast" of such a series.
- Eg., GDP. What's the "mean" of GDP? That depends entirely on the sample of data you're looking at. The same is true of the variance.

Deterministic trends

- A time series with a deterministic trend—sometimes called a "trend stationary" series—simply increases (or decreases) ad infinitum.
- They have neither a stable long-run mean nor a stable long-run variance. And there is no meaningful long-run "forecast" of such a series.
- Eg., GDP. What's the "mean" of GDP? That depends entirely on the sample of data you're looking at. The same is true of the variance.
- A trend-stationary series can be made stationary by removing its trend (a linear trend or a polynomial).

Deterministic trends

- A time series with a deterministic trend—sometimes called a "trend stationary" series—simply increases (or decreases) ad infinitum.
- They have neither a stable long-run mean nor a stable long-run variance. And there is no meaningful long-run "forecast" of such a series.
- Eg., GDP. What's the "mean" of GDP? That depends entirely on the sample of data you're looking at. The same is true of the variance.
- A trend-stationary series can be made stationary by removing its trend (a linear trend or a polynomial).
- Detrending such time series - important and not always straightforward.

Deterministic trends

- A time series with a deterministic trend—sometimes called a "trend stationary" series—simply increases (or decreases) ad infinitum.
- They have neither a stable long-run mean nor a stable long-run variance. And there is no meaningful long-run "forecast" of such a series.
- Eg., GDP. What's the "mean" of GDP? That depends entirely on the sample of data you're looking at. The same is true of the variance.
- A trend-stationary series can be made stationary by removing its trend (a linear trend or a polynomial).
- Detrending such time series - important and not always straightforward.
- We tend not to be interested in GDP level—but instead in things like GDP growth. It's important to note that GDP growth has been detrended.

Stochastic trends (or unit roots)

- A time series with a unit root—sometimes called a "stochastic trend"—is considerably more complicated, in part because they aren't obviously identifiable by simple visual inspection.

Stochastic trends (or unit roots)

- A time series with a unit root—sometimes called a “stochastic trend”—is considerably more complicated, in part because they aren't obviously identifiable by simple visual inspection.

- Though it's not always obvious visually, they neither have a stable long-run mean or variance. Neither is its long-run forecast particularly meaningful.

Stochastic trends (or unit roots)

- A time series with a unit root—sometimes called a "stochastic trend"—is considerably more complicated, in part because they aren't obviously identifiable by simple visual inspection.
- Though it's not always obvious visually, they neither have a stable long-run mean or variance. Neither is its long-run forecast particularly meaningful.
- A series with a unit root can be made stationary by taking first differences. That is, if Y_t has a unit root, it can be made stationary by taking $\Delta Y_t = Y_t - Y_{t-1}$.

Stochastic trends (or unit roots)

- A time series with a unit root—sometimes called a "stochastic trend"—is considerably more complicated, in part because they aren't obviously identifiable by simple visual inspection.
- Though it's not always obvious visually, they neither have a stable long-run mean or variance. Neither is its long-run forecast particularly meaningful.
- A series with a unit root can be made stationary by taking first differences. That is, if Y_t has a unit root, it can be made stationary by taking $\Delta Y_t = Y_t - Y_{t-1}$.
- The most prominent examples of series with unit roots are stock prices. Can you see why?

Some specifics about unit roots

So in the aforementioned setup, if $\rho = 1$, we have a unit root process, which simplifies to:

$$Y_t = Y_{t-1} + e_t$$

In fact, just backshift the subscripts, and you'll eventually see that:

$$Y_t = Y_0 + \Sigma e_t$$

So there is no decay whatsoever here, no mean-reversion, and no meaningful long-run forecast. Today's Y is simply the sum of (never-decaying) mean-zero shocks.

Outline of the session

Do deterministic or stochastic trends pose problems for inference?

- Some series have memories of their pasts that are sufficiently long to induce statistical problems. In particular, we'll mention one called the **spurious regression problem**.

Do deterministic or stochastic trends pose problems for inference?

- Some series have memories of their pasts that are sufficiently long to induce statistical problems. In particular, we'll mention one called the **spurious regression problem**.
- It affects both deterministic and stochastic trends, but a bit differently. We'll show the problem with deterministic trends using a real (but entirely playful) example. We'll show the problem with stochastic trends by using hypothetical data.

Deterministic Trends: Does golf cause divorce?

- Consider the following facts: In post-World War II America, golf became an increasingly popular sport. As its popularity grew, perhaps predictably the number of golf courses in America grew to accommodate the demand for places to play. That growth continued steadily into the early 21st century. We can think of the number of golf courses in America as a time series, of course, presumably one on an annual metric.

Deterministic Trends: Does golf cause divorce?

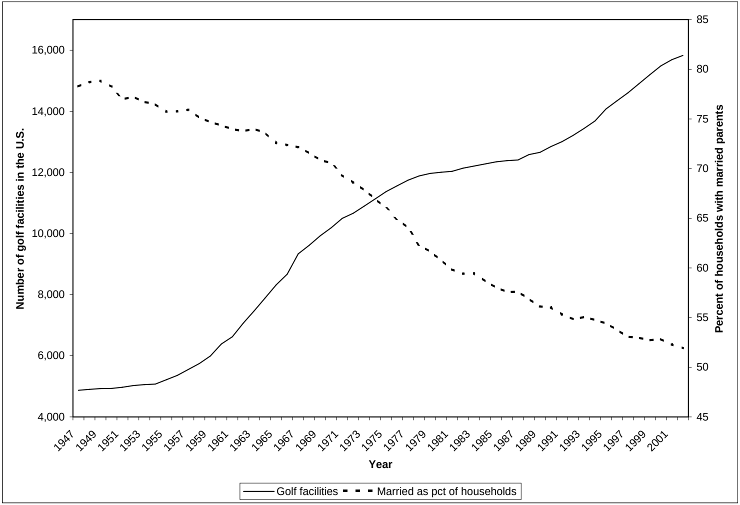
- Consider the following facts: In post-World War II America, golf became an increasingly popular sport. As its popularity grew, perhaps predictably the number of golf courses in America grew to accommodate the demand for places to play. That growth continued steadily into the early 21st century. We can think of the number of golf courses in America as a time series, of course, presumably one on an annual metric.

- Over the same period of time, divorce rates in America grew and grew. Where divorce was formerly an uncommon practice, today it is commonplace in American society. We can think of family structure as a time series, too—in this case, the percentage of households in which a married couple is present.

Both of these time series have deterministic trends

Both of these time series—likely for different reasons—have long memories. In the case of golf courses, the number of courses in year t obviously depends heavily on the number of courses the previous year. In the case of divorce rates, the dependence on the past presumably stems from the lingering, multi-period influence of the social forces that lead to divorce in the first place.

Golf courses and the demise of the family, 1947 - 2002



Is there a problem, officer?

- What's the problem here? Any time one time series with a long memory is placed in a regression model with another series which also has a long memory, it can lead to falsely finding evidence of a causal connection between the two variables. This is known as the "spurious regression problem."

Is there a problem, officer?

- What's the problem here? Any time one time series with a long memory is placed in a regression model with another series which also has a long memory, it can lead to falsely finding evidence of a causal connection between the two variables. This is known as the "spurious regression problem."
- If we take the demise of marriage as our dependent variable and use golf facilities as our independent variable, we would surely see that these two variables are related, statistically. In substantive terms, we might be tempted to jump to the conclusion that the growth of golf in America has led to the breakdown of the nuclear family.

Is this relationship causal?

Variable	coefficient (Std. Err.)
Golf facilities	-2.53*
	(0.09)
Constant	91.36*
	(1.00)
N	56
R^2	0.93

* indicates $p < .05$

Interpretations and objections

- Some of you—presumably, non-golfers—are nodding your heads and thinking, "But maybe golf does cause divorce rates to rise! Does the phrase "golf widow" ring a bell?"

Interpretations and objections

- Some of you—presumably, non-golfers—are nodding your heads and thinking, "But maybe golf does cause divorce rates to rise! Does the phrase "golf widow" ring a bell?"

- But here's the problem with variables with deterministic trends, and why it's such a potentially nasty problem in the social sciences. We could substitute any variable with a trend in it and come to the same "conclusion."

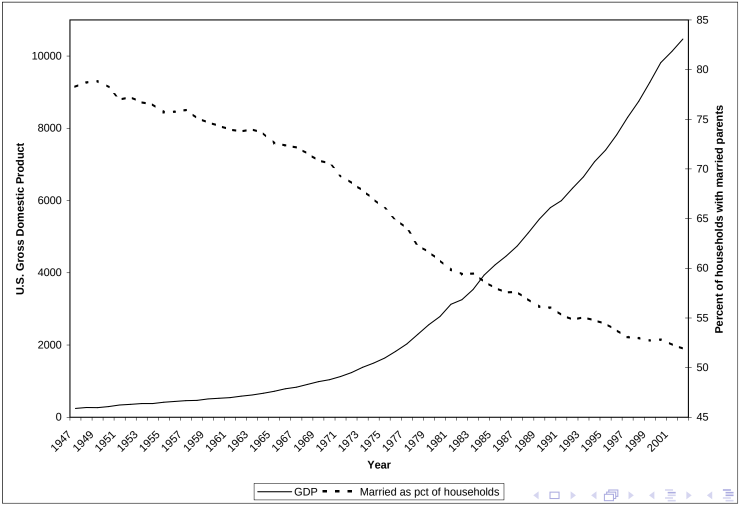
Interpretations and objections

- Some of you—presumably, non-golfers—are nodding your heads and thinking, "But maybe golf does cause divorce rates to rise! Does the phrase "golf widow" ring a bell?"

- But here's the problem with variables with deterministic trends, and why it's such a potentially nasty problem in the social sciences. We could substitute any variable with a trend in it and come to the same "conclusion."

- To prove the point, let's take another example. Instead of examining the growth of golf, let's look a different kind of growth—economic growth. In post-war America, Gross Domestic Product (GDP) as grown steadily, with few interruptions in its upward trajectory. Obviously, GDP has a deterministic trend, where current values of the series depend extremely heavily on past values.

The growth of the U.S. economy and the decline of the family, 1947 - 2002



GDP and the demise of the family, 1947 - 2002

Variable	coefficient (Std. Err.)
GDP (in trillions)	-2.71*
	(0.16)
Constant	74.00*
	(0.69)
N	56
R^2	0.84

* indicates $p < .05$

It's not the golf, it's not the economy, it's the trend

Using divorce as our dependent variable and GDP as our independent variable, the regression results show a strong, negative, and statistically significant relationship between the two. This is not occurring because higher rates of economic output have led to the destruction of the American family. It is occurring because both variables have trends in them, and a regression involving two variables with trends—even if they are not truly associated—will produce spurious evidence of a relationship.

What we're about to show you

Remember that a unit-root series looks like the following:

$$Y_t = Y_{t-1} + e_t$$

And then we can generate another series, which we'll call X, using an identical procedure:

$$X_t = X_{t-1} + v_t$$

We can generate data like these very easily in a spreadsheet. As we do, keep in mind that both e and v are entirely independent of one another (by construction). They will both be random draws from a mean zero, normal distribution. And Y and X will merely be the infinite sum of past shocks.

What to expect

- Will e and v be correlated?

What to expect

- Will e and v be correlated?
- What will Y and X look like, visually?

What to expect

- Will e and v be correlated?
- What will Y and X look like, visually?
- Will Y and X be correlated?

What to expect

- Will e and v be correlated?
- What will Y and X look like, visually?
- Will Y and X be correlated?
- Let's have a look

What we saw

- Indeed, e and v are uncorrelated. (That's a relief!)

What we saw

- Indeed, e and v are uncorrelated. (That's a relief!)
- In our sample of 200 observations, sometimes Y looked like it had a stable, long-run mean, but sometimes it did not.

What we saw

- Indeed, e and v are uncorrelated. (That's a relief!)
- In our sample of 200 observations, sometimes Y looked like it had a stable, long-run mean, but sometimes it did not.
- Even though e and v are uncorrelated, Y and X are correlated. Randomness is uncorrelated with (independent) randomness, but the sum of randomness is correlated with the sum of (independent) randomness.

Skepticism about detecting unit roots empirically

- There are a variety of ways to test for unit roots in a time series—that is, where the null hypothesis is that the series contains a unit root—the most popular of which is the Augmented Dickey-Fuller test. These are easy to implement and interpret.

Skepticism about detecting unit roots empirically

- There are a variety of ways to test for unit roots in a time series—that is, where the null hypothesis is that the series contains a unit root—the most popular of which is the Augmented Dickey-Fuller test. These are easy to implement and interpret.
- There are also a couple of ways to test for both trend stationarity and level stationarity—that is, where the null hypothesis is that a series is either level stationary or trend stationary—which are easy to implement and only a little difficult to interpret. The KPSS test is the most popular.

Skepticism about detecting unit roots empirically

- There are a variety of ways to test for unit roots in a time series—that is, where the null hypothesis is that the series contains a unit root—the most popular of which is the Augmented Dickey-Fuller test. These are easy to implement and interpret.
- There are also a couple of ways to test for both trend stationarity and level stationarity—that is, where the null hypothesis is that a series is either level stationary or trend stationary—which are easy to implement and only a little difficult to interpret. The KPSS test is the most popular.
- The problem: These tests only work asymptotically. In time series data, we're very lucky to have 200 data points. I've done work with 28 data points, which is surely no

Skepticism about detecting unit roots empirically

- There are a variety of ways to test for unit roots in a time series—that is, where the null hypothesis is that the series contains a unit root—the most popular of which is the Augmented Dickey-Fuller test. These are easy to implement and interpret.
- There are also a couple of ways to test for both trend stationarity and level stationarity—that is, where the null hypothesis is that a series is either level stationary or trend stationary—which are easy to implement and only a little difficult to interpret. The KPSS test is the most popular.
- The problem: These tests only work asymptotically. In time series data, we're very lucky to have 200 data points. I've done work with 28 data points, which is surely no
- What to do? I suggest relying on theory instead of these tests, and (later) will recommend a strategy that is invariant to whether or not there's a unit root (but not a deterministic trend) in your data.

Outline of the session

1. Some time-series basics
2. Memory in time series analysis
   - Memory and autocorrelation
   - One variant of non-stationarity: Deterministic trends
   - Another variant of non-stationarity: Unit roots
3. The consequences of memory: Trends and spurious regressions
   - A real example with deterministic trends
   - A hypothetical example with stochastic trends
4. One solution to the spurious regression problem: First-differencing the dependent variable
5. Another solution to the spurious regression problem: The lagged dependent variable

"Levels" versus "changes"

- One way to avoid the problems of spurious regressions is to use a **differenced dependent variable**.

"Levels" versus "changes"

- One way to avoid the problems of spurious regressions is to use a **differenced dependent variable**.
- A differenced (or, equivalently, "first differenced") variable is calculated by subtracting the first lag of the variable (Y_{t-1}) from the current value Y_t.

"Levels" versus "changes"

- One way to avoid the problems of spurious regressions is to use a **differenced dependent variable**.
- A differenced (or, equivalently, "first differenced") variable is calculated by subtracting the first lag of the variable (Y_{t-1}) from the current value Y_t.
- The resulting time series is typically represented as $\Delta Y_t = Y_t - Y_{t-1}$.

"Levels" versus "changes"

- In fact, when time series have either deterministic or stochastic trends, taking first differences of both independent and dependent variables can be done.

"Levels" versus "changes"

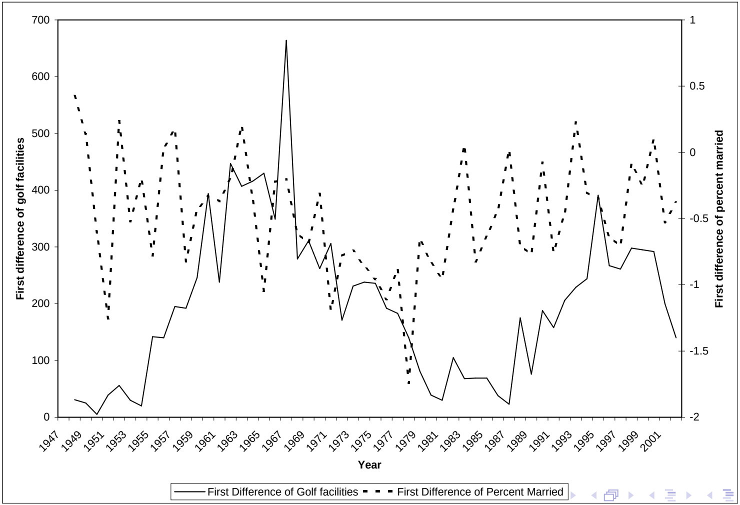
- In fact, when time series have either deterministic or stochastic trends, taking first differences of both independent and dependent variables can be done.

- In effect, instead of Y_t representing the levels of a variable, ΔY_t represents the period-to-period changes in the level of the variable.

“Levels” versus “changes”

- In fact, when time series have either deterministic or stochastic trends, taking first differences of both independent and dependent variables can be done.
- In effect, instead of Y_t representing the levels of a variable, ΔY_t represents the period-to-period changes in the level of the variable.
- For many (but not all) variables with such long memories, taking first differences will eliminate the visual pattern of a variable that just seems to keep going up.

First differences of the number of golf courses and percentage of married families, 1947 - 2002



Some cautions

- Because, in these cases, taking first differences of the series removes the long memories from the series, these transformed time series will not be subject to the spurious regression problem. But I'd like to caution against thoughtless differencing of time series (which is disturbingly common!). In particular, taking first differences of time series can eliminate some (true) evidence of an association between time series in certain circumstances.

Some cautions

- Because, in these cases, taking first differences of the series removes the long memories from the series, these transformed time series will not be subject to the spurious regression problem. But I'd like to caution against thoughtless differencing of time series (which is disturbingly common!). In particular, taking first differences of time series can eliminate some (true) evidence of an association between time series in certain circumstances.
- I'd recommend that, wherever possible, you use theoretical reasons to either detrend a time series, or to analyze it in levels. In effect, you should ask yourself if your theory about a causal connection between X and Y makes more sense in levels or first-differences (or something else). For example, if you are analyzing budgetary data from a government agency, does your theory specify particular things about the sheer amount of agency spending (in which case, you would analyze the data in levels), or does it specify particular things about what causes budgets to shift from year to year (in which case, you would analyzed the data in first differences)?

Outline of the session

1. Some time-series basics
2. Memory in time series analysis
   - Memory and autocorrelation
   - One variant of non-stationarity: Deterministic trends
   - Another variant of non-stationarity: Unit roots
3. The consequences of memory: Trends and spurious regressions
   - A real example with deterministic trends
   - A hypothetical example with stochastic trends
4. One solution to the spurious regression problem: First-differencing the dependent variable
5. Another solution to the spurious regression problem: The lagged dependent variable

Multiple lags of our independent variables

Consider a simple two-variable system with our familiar variables Y and X, except where, to allow for the possibility that previous lags of X might affect current levels of Y, we include a large number of lags of X in our model.

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_1 + \cdots + \beta_k X_{t-k} + u_t$$

This model is known as a **distributed lag model**. Notice the slight shift in notation here, where we are subscripting our β coefficients by the number of periods that that variable is lagged from the current value; hence, the β for X_t is β_0 (because $t - 0 = 0$). Under such a setup, the **cumulative impact** β of X on Y is equal to:

$$\beta = \beta_0 + \beta_1 + \beta_2 + \cdots + \beta_k = \sum_{i=0}^{k} \beta_i$$

The cumulative impact of X on Y

$$\beta = \beta_0 + \beta_1 + \beta_2 + \cdots + \beta_k = \sum_{i=0}^{k} \beta_i$$

It is worth emphasizing that we are interested in that cumulative impact of X on Y, not merely the **instantaneous effect** of X_t on Y_t represented by the coefficient β_0.

But how can we capture the effects of X on Y without estimating such a cumbersome model like the one above?

Assume a thing or two

Assume that the effects of a change in X on Y eventually decay over time, such that:

$$\beta_k = \beta_0 \lambda^k$$

$$k = 0, 1, 2, \ldots$$

$$0 < \lambda < 1$$

That means that every β is less than the β from the preceding time period. This is often, but maybe not always, a reasonable assumption.

Now we can write the distributed-lag model as:

$$Y_t = \alpha + \beta_0 X_t + \beta_0 \lambda X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \cdots + u_t$$

Notice that, unlike the generic distributed-lag model, in this case all of the β coefficients are labeled β_0. For the moment, this doesnt look any easier. But lag everything by one period (which is true by definition) and multiply every term through by λ to get:

$$\lambda Y_{t-1} = \lambda\alpha + \lambda\beta_0 X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \beta_0 \lambda^3 X_{t-3} + \cdots + \lambda u_{t-1}$$

Subtract and rearrange terms to get

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t$$

where

$$v_t = (u_t - \lambda u_{t-1})$$

This is known as the **Koyck transformation**, and is commonly referred to as the **lagged dependent variable** model.

The mechanics of the Koyck model

Compare the Koyck transformation to the equivalent distributed lag model above. Both have the same dependent variable, Y_t. Both have a variable representing the immediate impact of X_t on Y_t. But where the distributed lag model also has a slew of coefficients for variables representing all of the lags of 1 through k of X on Y_t, the lagged dependent variable model instead contains a single variable and coefficient, λY_{t-1}.

Because the two setups are equivalent, then this means that the lagged dependent variable does not represent how Y_{t-1} somehow causes Y_t, but instead Y_{t-1} is a stand-in for the cumulative effects of all past lags of X (that is, lags 1 through k) on Y_t. All of that through estimating a single coefficient instead of a very large number of them.

More on the Koyck mechanics

The coefficient λ, then, represents the ways in which past values of X affect current values of Y, which nicely solves the problem outlined at the start of this section. Normally, the values of λ will range between 0 and 1. You can readily see that if $\lambda = 0$ then there is literally no effect of past values of X on Y_t. Such values are uncommon in practice. As λ gets larger, that indicates that the effects of past lags of X on Y_t persist longer and longer into the future.

The λ coefficient

In these models, the cumulative effect of X on Y is conveniently described as:

$$\beta = \frac{\beta_0}{1 - \lambda}$$

Examining the formula, it is easy to see that when $\lambda = 0$, the denominator is equal to 1, and the cumulative impact is exactly equal to the instantaneous impact. There is no lagged effect at all. When $\lambda = 1$, however, we run into problems; the denominator equals zero, so the quotient is undefined. But as λ approaches 1, you can see that the cumulative effect grows. Thus, as the values of the coefficient on the lagged dependent variable move from zero toward one, the cumulative impact of changes in X on Y grows.

From distributed lags to Error Correction Models

The ARDL(1,1) model:

$$y_t = \mu + \gamma_1 Y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

can be rearranged as:

$$\Delta y_t = \mu + \beta_0 \Delta x_t + (\gamma_1 - 1)(y_{t-1} - \theta x_{t-1}) + \varepsilon_t$$

Advantages of ECMs

- They were originally conceived of as being used when the series have unit roots (and are what is called "cointegrated"). That is no longer considered a necessary precondition for using ECMs.

Advantages of ECMs

- They were originally conceived of as being used when the series have unit roots (and are what is called "cointegrated"). That is no longer considered a necessary precondition for using ECMs.
- As you can see from the setup, they allow the analyst to capture both short-run effects (the Δ) as well as the long-run effects (in the form of the lagged levels, which is less obvious).

Advantages of ECMs

- They were originally conceived of as being used when the series have unit roots (and are what is called "cointegrated"). That is no longer considered a necessary precondition for using ECMs.
- As you can see from the setup, they allow the analyst to capture both short-run effects (the Δ) as well as the long-run effects (in the form of the lagged levels, which is less obvious).
- If there are long-run effects, this can be interpreted in the following way: When Y and X are out of long-run equilibrium with one another, that disequilibrium will be corrected in subsequent time periods at the rate of the coefficient for the lagged levels of Y. So disequilibrium can be corrected either quickly, slowly, or (as the case may be) not at all.

Advantages of ECMs

- They were originally conceived of as being used when the series have unit roots (and are what is called "cointegrated"). That is no longer considered a necessary precondition for using ECMs.
- As you can see from the setup, they allow the analyst to capture both short-run effects (the Δ) as well as the long-run effects (in the form of the lagged levels, which is less obvious).
- If there are long-run effects, this can be interpreted in the following way: When Y and X are out of long-run equilibrium with one another, that disequilibrium will be corrected in subsequent time periods at the rate of the coefficient for the lagged levels of Y. So disequilibrium can be corrected either quickly, slowly, or (as the case may be) not at all.
- Two substantive illustrations:

Advantages of ECMs

- They were originally conceived of as being used when the series have unit roots (and are what is called "cointegrated"). That is no longer considered a necessary precondition for using ECMs.
- As you can see from the setup, they allow the analyst to capture both short-run effects (the Δ) as well as the long-run effects (in the form of the lagged levels, which is less obvious).
- If there are long-run effects, this can be interpreted in the following way: When Y and X are out of long-run equilibrium with one another, that disequilibrium will be corrected in subsequent time periods at the rate of the coefficient for the lagged levels of Y. So disequilibrium can be corrected either quickly, slowly, or (as the case may be) not at all.
- Two substantive illustrations:
    1. Long-run and short-run interest rates

Advantages of ECMs

- They were originally conceived of as being used when the series have unit roots (and are what is called "cointegrated"). That is no longer considered a necessary precondition for using ECMs.
- As you can see from the setup, they allow the analyst to capture both short-run effects (the Δ) as well as the long-run effects (in the form of the lagged levels, which is less obvious).
- If there are long-run effects, this can be interpreted in the following way: When Y and X are out of long-run equilibrium with one another, that disequilibrium will be corrected in subsequent time periods at the rate of the coefficient for the lagged levels of Y. So disequilibrium can be corrected either quickly, slowly, or (as the case may be) not at all.
- Two substantive illustrations:
  1. Long-run and short-run interest rates
  2. Public policy and public opinion