# Basic Regression with Time Series Data

ECONOMETRICS (ECON 360)

BEN VAN KAMMEN, PHD

# Introduction

This chapter departs from the cross-sectional data analysis, which has been the focus in the preceding chapters.

Instead of observing many (“n”) elements in a single time period, time series data are generated by observing a single element over many time periods.

The goal of the chapter is broadly to show what can be done with OLS using time series data.

Specifically students will identify similarities in and differences between the two applications and practice methods unique to time series models.

# Outline

The Nature of Time Series Data.

Stationary and Weakly Dependent Time Series.

Asymptotic Properties of OLS.

Using Highly Persistent Time Series in Regression Analysis.

Examples of (Multivariate) Time Series Regression Models.

Trends and Seasonality.

# The nature of time series data

Time series observations have a meaningful order imposed on them, from first to last, in contrast to sorting a cross-section alphabetically or by an arbitrarily assigned ID number.

The values are generated by a stochastic process, about which assumptions can be made, e.g., the mean, variance, covariance, and distribution of the “innovations” (also sometimes called disturbances or shocks) that move the process forward through time.

# The nature of time series data (continued)

So an observation of a time series, e.g.,

$$\{y_t\};\, t \in \{0, \ldots, n\}, \text{ where } n \text{ is the sample size,}$$

can be thought of as a single realization of the stochastic process.

- Were history to be repeated, many other realizations for the path of $y_t$ would be possible.

Owing to the randomness generating the observations of *y*, the properties of OLS that depend on random sampling still hold.

The econometrician's job is to accurately model the stochastic process, both for the purpose of inference as well as <u>prediction</u>.

- Prediction is an application for time series model estimates because knowing the process generating new observations of *y* naturally enables you to estimate a future ("out of sample") value.

# Stationary and weakly dependent time series

Many time series processes can be viewed either as

- regressions on lagged (past) values with additive disturbances or
- as aggregations of a history of innovations.

In order to show this, we have to write down a model and make some assumptions about how present values of *y* ($y_t$) are related to past values (e.g., $y_{t-1}$) and about the variance and covariance structure of the disturbances.

For the sake of clarity, consider a univariate series that does not depend on values of other variables—only on lagged values of itself.

# Stationary and weakly dependent time series (continued)

$$y_t = \rho_1 y_{t-1} + e_t; E(e_t) = 0, E(e_t^2) = \sigma_e^2, E(e_t e_{s\neq t}) = 0,$$

is a simple example of such a model.

- Specifically this is an autoregressive process of order 1—more commonly called “AR(1)” for brevity—because *y* depends on exactly 1 lag of itself.
- In this instance we have also assumed that the disturbances have constant (zero) mean and variance and are not correlated across time periods.

In order to make use of a series in regression analysis, it needs to have an expected value, variance, and auto-covariance (covariance with lagged values of itself), though, and not all series have these (at least not that are finite).

A series will have these properties if it is stationary.

# Stationarity

The property of stationarity implies:

(1) $E(y_t)$ is independent of $t$,

(2) $Var(y_t)$ is a finite positive constant, independent of $t$,

(3) $Cov(y_t, y_{t-s})$ is a finite function of $(t - s)$, but not $t$ or $s$,

(4) The distribution of $y_t$ is not changing over time.

For our purposes the 4$^{th}$ condition is unnecessary, and a process that satisfies the first 3 is still weakly stationary or covariance stationary.

# Stationarity of AR(1) process

The AR(1) process, $y_t$, is covariance stationary under specific conditions.

$$E(y_t) = \rho_1 E(y_{t-1}) + E(e_t); (1) \rightarrow E(y_t) = \rho_1 E(y_t) \Leftrightarrow E(y_t) = 0,$$

$$Var(y_t) = E(y_t^2) = E[\rho_1^2 y_{t-1}^2] + Var(e_t) = \rho_1^2 E(y_t^2) + \sigma_e^2,$$

$$\Leftrightarrow Var(y_t) \equiv \sigma_y^2 = \frac{\sigma_e^2}{1-\rho_1^2}.$$

This is only finite if $\rho_1$ is less than one in absolute value.

- Otherwise the denominator goes to zero and the variance goes to infinity.

$$\sigma_y^2 = \frac{\sigma_e^2}{1-\rho_1^2}; |\rho_1| < 1.$$

# Using highly persistent time series in regression analysis

Even if the weak dependency assumption fails, i.e., $\rho_1 = 1$, an autoregressive process can be analyzed using a (1st difference) transformed OLS model, which makes a non-stationary, strongly dependent process stationary.

- The differences in the following process (called a "random walk") are stationary.

$$y_t = 1 * y_{t-1} + e_t \rightarrow \Delta y_t \equiv y_t - y_{t-1} = e_t,$$

has a finite mean and variance (distribution) that does not depend on $t$.

The Wooldridge book contains more information on testing whether a series has this kind of persistence (see pp. 396-399 and 639-644) and selecting an appropriate transformation of the regression model, but these topics are left to the interested student as optional.

# Stationarity of AR(1) process (continued)

The covariance between two observations that are *h* periods apart is:

$$E(y_t y_{t+h}) = \sigma_y^2 \rho_1^h.$$

This <u>"auto-covariance" does not depend on either of the two places in the time series—only on how far apart they are</u>.

- To derive this, one needs to iteratively substitute for $y_{t+1}$:

$$y_{t+1} = \rho_1(\rho_1 y_{t-1} + e_t) + e_{t+1};\ y_{t+2} = \rho_1[(\rho_1^2 y_{t-1} + \rho_1 e_t) + e_{t+1}] + e_{t+2}.$$

With careful inspection, a pattern emerges as you continue substituting.

$$y_{t+h} = \rho_1^{h+1} y_{t-1} + \sum_{s=0}^{h} \rho_1^s e_{t+h-s}\,.$$

More on this derivation.

# Stationarity of AR(1) process (concluded)

How persistent the series is depends on how close to one $\rho_1$ is in absolute value.

- The closer it is, the more persistent are the values in the series.
- It is also worth noting how the persistence "dies out" when the gap (*h*) between the observations is large.
- This should confirm the intuition that observations with more time intervening between them will be less correlated.

Before moving on, let's summarize a couple more things about the iterative substitution of the AR(1) $\{y_t\}$ process.

# Autocorrelation concluded

The current period's value can be expressed neatly as an infinitely long summation of the past disturbances ("history").

$$y_t = \sum_{s=0}^{\infty} \rho_1^s e_{t-s}\text{, and}$$

the process can accommodate a constant as well, i.e.,

$$y_t = \rho_0 + \rho_1 y_{t-1} + e_t = \sum_{s=0}^{\infty} \rho_1^s (\rho_0 + e_{t-s})\,;E(y_t) = \frac{\rho_0}{1-\rho_1}.$$

Though many variables exhibit no more than 1 order of autocorrelation, it is conceivable to have "p" orders, i.e.,

$$y_t = \rho_0 + \sum_{s=1}^{p} \rho_s y_{t-s} + e_t\text{, is AR(p).}$$

# Asymptotic properties of OLS

The assumptions about autoregressive processes made so far lead to disturbances that are contemporaneously exogenous if the parameters were to be estimated by OLS.

This set (next slide) of assumptions leads to Theorem 11.1, which is that OLS estimation of a time series is consistent.

An AR process, for example, will still be biased in finite samples, however, because it violates the stronger Assumption TS.3 (that all disturbances are uncorrelated with all regressors—not just the contemporary one).

More on the biasedness of OLS.

# Conditions under which OLS on time series data is consistent

1. Assumption TS.1' states that the model is linear in parameters (appears in the text in Chapter 10 as TS.1), the process is stationary, and <u>weakly dependent</u> ($Cov(y_t, y_{t+h}) \rightarrow 0$ as h gets large).

2. Assumption TS.2' (same as TS.2) states that the regressors (lagged values) have variation (are not constants) nor are perfectly collinear (functions of other regressors).

3. Assumption TS.3' states that the current period's disturbance is mean independent of the regressors, i.e., the lagged values of $y_t$.

$$E(e_t|\boldsymbol{x_t}) = 0;$$

   $\boldsymbol{x_t}$ is the set of regressors: either lagged values of *y* or other independent variables, as in cross-sectional analysis.

# Asymptotic properties of OLS (concluded)

Under additional Assumptions about the disturbances, inference according to the usual tests is valid:

4. Assumption TS.4’ is the analog of the homoskedasticity assumption:
$$Var(e_t|\boldsymbol{x_t}) = Var(e_t) = \sigma^2,$$
which is called contemporaneous homoskedasticity.

5. And Assumption TS.5’ rules out serial correlation in the disturbances:
$$E(e_t, e_{t-h}|\boldsymbol{x_t}, \boldsymbol{x_{t-h}}) = 0|h \neq 0.$$

# Examples of (multivariate) time series regression models

There are numerous time series applications that involve multiple variables moving together over time that this course will not discuss:

- the interested student should study Chapter 18.

But bringing the discussion of time series data back to familiar realms, consider a simple example in which the dependent variable is a function of contemporaneous and past values of the explanatory variable.

Models that exhibit this trait are called “finite distributed lag” (FDL) models.

# Finite distributed lag models

This type is further differentiated by its order, i.e., how many lags are relevant for predicting $y$.

An FDL of order $q$ is written:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \ldots + \delta_q z_{t-q} + u_t, \text{ or compactly as,}$$

$$y_t = \alpha_0 + \sum_{h=0}^{q} \delta_h z_{t-h} + u_t.$$

Note that this contains the “Static Model,” i.e., in which $\delta_h = 0 | h > 0$, as a special case.

Applications are numerous.

- the *fertility* (responds to tax code incentives to have children) example in the text
- exemplifies short run and long run responses to a market shock.

# Finite distributed lag models (continued)

In a competitive market, a demand increase will raise prices in the short run but invite entry in the long run, along with its price reducing effects.

- Also there is a difference between short run and long run demand elasticity; the latter is more elastic.
- So the effect of a price change on quantity demanded may be modest in the present but significant over a longer period of time.

For example demand for gasoline is quite inelastic in the short run but much more elastic in the long run,

- because consumers can change their vehicles, commuting habits, and locations if given enough time.

These effects could be estimated separately using an FDL model such as:

$$fuel_t = \alpha_0 + \delta_0 price_t + \delta_1 price_{t-1} + \ldots + \delta_q price_{t-q} + u_t.$$

# Finite distributed lag models (concluded)

The familiar prediction from microeconomics is that the static effect—sometimes called the impact multiplier—is quite small or zero while the overall effect—the equilibrium multiplier—of a price increase is large and negative.

Formally these could be stated:

$$\delta_0 \approx 0 \text{ and } \sum_{h=0}^{q} \delta_h < 0.$$

For the long run or "equilibrium" multiplier, one wants to examine the cumulative (sum of the) effects, as a persistently higher price continues to reduce quantity demanded over time periods.

- This is the effect of a permanent increase, as described in the text on page 348.

# Trends and seasonality

A common source of omitted variable bias in a time series regression is time, itself.

If two variables are trending in the same (opposite) direction over time, they will appear related if time is omitted from the regression.

- This is true even when there is no substantive relationship between the two variables.
- Many examples here.

To model a time trend in $y$ that increases it by a constant amount ($\alpha_1$) each period.

$$y_t = \alpha_0 + \alpha_1 t + e_t \rightarrow \Delta y_t = \alpha_0 + \alpha_1 t + e_t - (\alpha_0 + \alpha_1 (t-1) + e_{t-1}),$$

$$\Leftrightarrow \Delta y_t = \alpha_1 (t - t + 1) + e_t - e_{t-1} = \alpha_1 + \Delta e_t.$$

# Trends and seasonality (continued)

The difference in consecutive errors has an expectation of zero, so $\alpha_1$ is the expected change per period.

Were $y_t$ to grow at a constant *rate* instead of by a constant amount each period, the semi-log specification would more accurately capture the time trend, i.e.,

$$\ln(y_t) = \alpha_0 + \alpha_1 t + e_t \rightarrow \Delta\ln(y_t) = \alpha_1 \approx \frac{\Delta y_t}{y_{t-1}}.$$

This is the (expected) <u>growth rate</u> of *y* per period: *y* grows at $100 * \alpha_1\%$ per period.

To contrast the two varieties of time trends, observe the following two figures.

# Time trends illustrated (1)

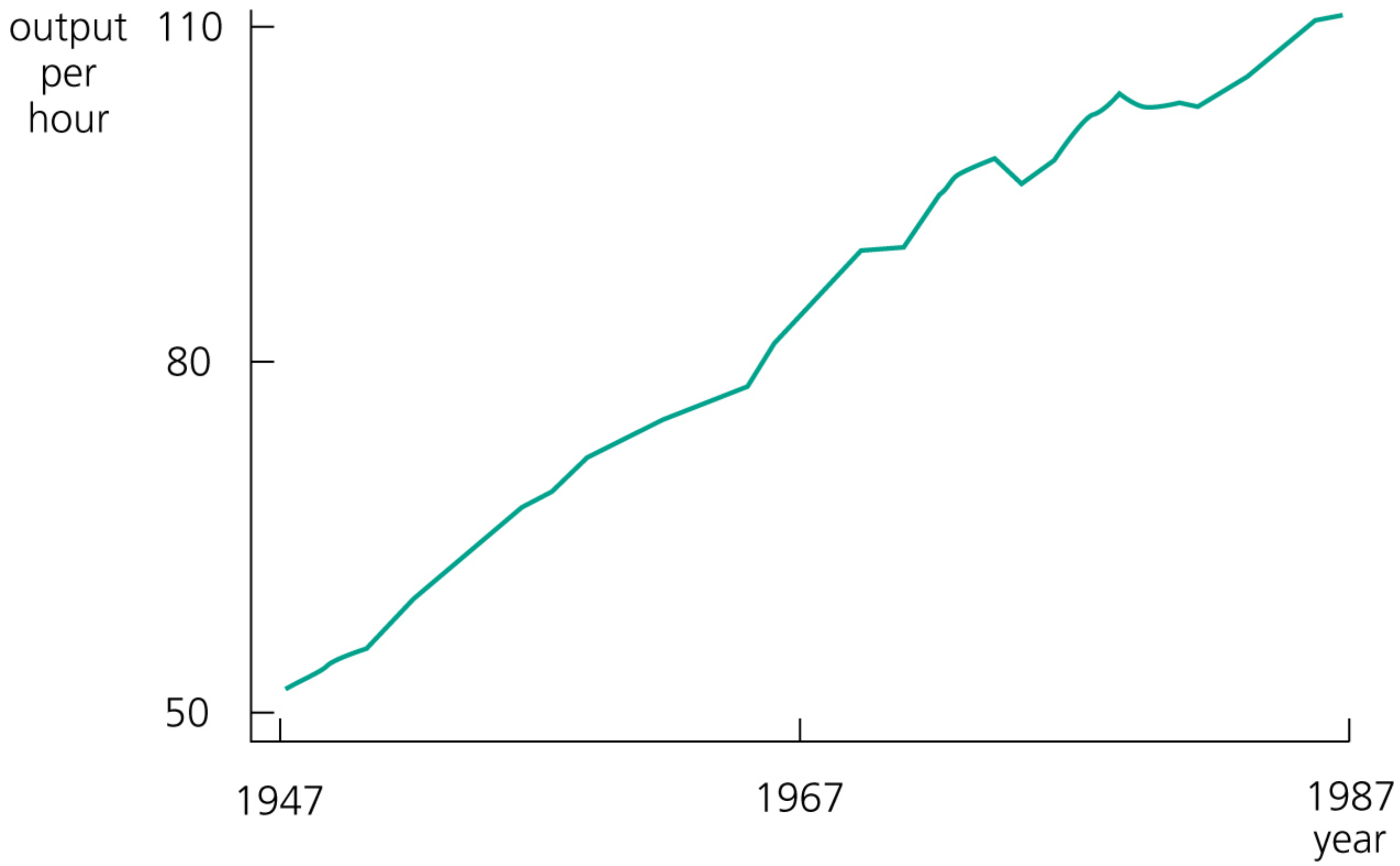


*Figure 1: Linear Trend*

# Time trend illustrated (2)

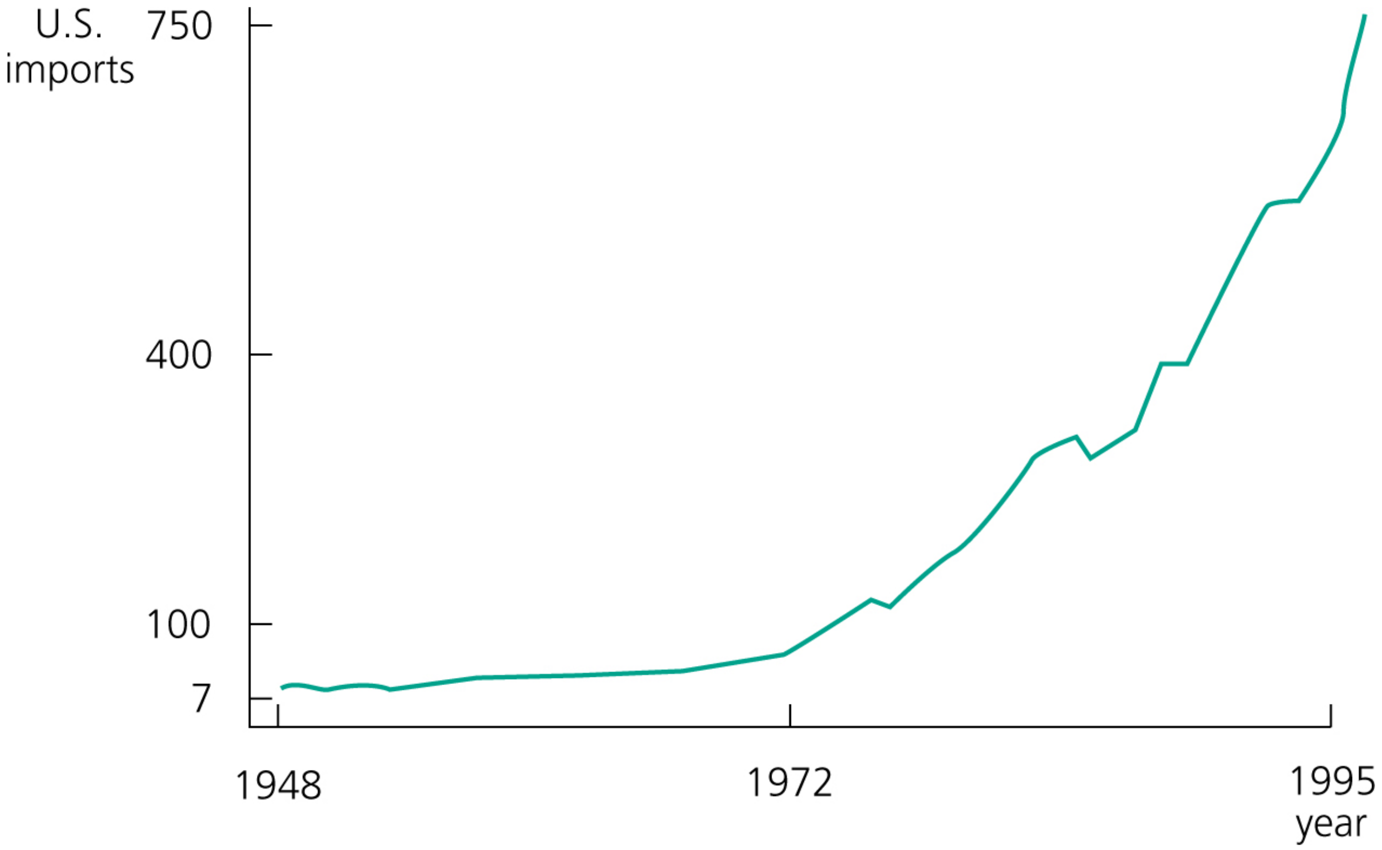


*Figure 1: Exponential Trend*

# Time trends and seasonality (continued)

Accounting for the time trend when regressing two time series variables avoids the omitted variable problem that would result from estimating the model,

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 t + u_t,$$

using only the *x* variables, thus generating biased estimates of their coefficients.

# A detrending interpretation of regressions with a time trend

Were the researcher to “de-trend” all of the variables prior to performing the above regression, however, the estimates would <u>not</u> be biased.

By “de-trending” what is meant is to regress each of them on time and subtract the fitted values—so store the residuals.

- The following example will illustrate this.

Suppose one regressed a time series of *employment* in a local area (San Francisco, CA) observations on a time series of the *minimum wage*.

# Detrending (continued)

The results would look like this.

```
. reg l_emp minimum_wage

      Source |       SS       df       MS              Number of obs =      84
-------------+------------------------------           F(  1,    82) =   20.42
       Model |  .020840385     1  .020840385           Prob > F      =  0.0000
    Residual |  .083671828    82  .001020388           R-squared     =  0.1994
-------------+------------------------------           Adj R-squared =  0.1896
       Total |  .104512213    83  .001259183           Root MSE      =  .03194

------------------------------------------------------------------------------
       l_emp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
minimum_wage |   .0174864   .0038693     4.52   0.000     .0097892    .0251836
       _cons |   12.86403   .0338867   379.62   0.000     12.79662    12.93144
------------------------------------------------------------------------------
```

The positive coefficient estimate seems to contradict the usual theoretical prediction that a wage floor decreases equilibrium employment.

However both variables trend upward over time, as the following regressions demonstrates.

# Detrending (continued)

```
. reg l_emp time

      Source |       SS       df       MS              Number of obs =      84
-------------+------------------------------           F(  1,    82) =   79.89
       Model |  .051574366     1  .051574366           Prob > F      =  0.0000
    Residual |  .052937847    82  .000645584           R-squared     =  0.4935
-------------+------------------------------           Adj R-squared =  0.4873
       Total |  .104512213    83  .001259183           Root MSE      =  .02541

------------------------------------------------------------------------------
       l_emp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0010219   .0001143     8.94   0.000     .0007945    .0012494
       _cons |   12.44664   .0638019   195.08   0.000     12.31972    12.57356
------------------------------------------------------------------------------

. reg minimum_wage time

      Source |       SS       df       MS              Number of obs =      84
-------------+------------------------------           F(  1,    82) =  318.05
       Model |  54.1859371     1  54.1859371           Prob > F      =  0.0000
    Residual |  13.9702914    82  .170369408           R-squared     =  0.7950
-------------+------------------------------           Adj R-squared =  0.7925
       Total |  68.1562286    83   .82115938           Root MSE      =  .41276

------------------------------------------------------------------------------
minimum_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0331242   .0018574    17.83   0.000     .0294293    .0368191
       _cons |  -9.755329   1.036462    -9.41   0.000    -11.81718   -7.693475
------------------------------------------------------------------------------
```

# Detrending (continued)

Controlling for the time trend yields support for the more familiar theoretical prediction—that wage floors decrease equilibrium unemployment.

```
. reg l_emp minimum_wage time

      Source |       SS       df       MS              Number of obs =      84
-------------+------------------------------           F(  2,    81) =   75.62
       Model |  .06805971     2  .034029855           Prob > F      =  0.0000
    Residual | .036452504    81  .000450031           R-squared     =  0.6512
-------------+------------------------------           Adj R-squared =  0.6426
       Total | .104512213    83  .001259183           Root MSE      =  .02121

------------------------------------------------------------------------------
       l_emp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
minimum_wage |  -.0343515   .0056757    -6.05   0.000    -.0456444   -.0230587
        time |   .0021598   .0002109    10.24   0.000     .0017403    .0025793
       _cons |   12.11153   .0768328   157.63   0.000     11.95865     12.2644
------------------------------------------------------------------------------
```

# Detrending (concluded)

The following Stata code would enable you to obtain the same results using *de-trended* variables.

**reg l_emp time**

**predict l_emp_detr, residuals**

**reg minimum_wage time**

**predict mw_detr, residuals**

Then you could regress the residuals on one another (without the time trend) to obtain the estimates without the bias of the omitted time trend.

```
. reg l_emp_detr mw_detr

      Source |       SS           df       MS      Number of obs   =        84
-------------+----------------------------------   F(1, 82)        =     37.08
       Model |  .016485344         1  .016485344   Prob > F        =    0.0000
    Residual |  .036452504        82  .000444543   R-squared       =    0.3114
-------------+----------------------------------   Adj R-squared   =    0.3030
       Total |  .052937848        83  .000637805   Root MSE        =    .02108

------------------------------------------------------------------------------
  l_emp_detr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     mw_detr |  -.0343515    .005641    -6.09   0.000    -.0455732   -.0231299
       _cons |   9.42e-11   .0023005     0.00   1.000    -.0045764    .0045764
------------------------------------------------------------------------------
```

# Seasonality

The same statements about time trends can be made about seasonal (calendar period-specific) effects:

- variables that move in predictable ways through the calendar year.
- These can be accounted for using seasonal indicator variables, e.g.,

$$march_t = \begin{cases} 1, \text{month of observation is } March \\ 0, \text{otherwise.} \end{cases}$$

Including a set of monthly (quarterly) indicators in a regression accomplishes the same thing as including a time trend:

- controlling for a spurious variable that would otherwise bias the estimates.

The same process that works with trends works with seasonality: "de-seasonalizing" of variables.

# Conclusion (1)

This lesson has shown a representative sample of basic time series regression methods.

Time series analysis has been generalized to univariate processes that exhibit autoregressive and moving average ("ARMA" models) properties and are most concisely represented by the Wold Decomposition:

$$y_t = E^*(y_t|y_{t-1}..., y_{t-p}) + \sum_{i=0}^{\infty} \pi_i e_{t-1} ,$$

i.e., there is a deterministic component (the optimal linear prediction of the contemporary value, conditional on past values) and the linearly indeterministic component (the weighted influence of past disturbances).

# Conclusion (2)

Other directions for more thorough examination by the interested student include:

1. Descriptive measures of autocorrelation, e.g., the autocovariance function and partial autocorrelation coefficients, and the correlogram.
2. Tests for autocorrelation, e.g., the Durbin-Watson Test, Godfrey-Breusch Test, and Box-Pierce Test.
3. Heteroskedasticity robust inference about time series estimates.
4. Tests for the order of integration of a process.
5. The use of time series models for out of sample prediction ("forecasting").

Collectively these topics would consume far more time that we have in this course.  The interested student is advised to indulge any interest in a course such as ECON 573 (Financial Econometrics).

# Optional: autocovariance of AR(1) process

$$y_t y_{t+h} = \left[\rho_1^{h+1} y_{t-1} + \sum_{s=0}^{h} \rho_1^s e_{t+h-s}\right] (\rho_1 y_{t-1} + e_t)$$

$$= \rho_1^{h+2} y_{t-1}^2 + y_{t-1} \sum_{s=0}^{h} \rho_1^{s+1} e_{t+h-s} + \rho_1^{h+1} y_{t-1} e_t + e_t \sum_{s=0}^{h} \rho_1^s e_{t+h-s} \,.$$

Take the expectation of this to find the covariance.

# Optional: autocovariance of AR(1) process (continued)

$$Cov(y_t, y_{t+h}) = E(y_t y_{t+h})$$

$$= \rho_1^{h+2} E(y_{t-1}^2) + E(y_{t-1}) \sum_{s=0}^{h} \rho_1^{s+1} e_{t+h-s} + \rho_1^{h+1} E(y_{t-1} e_t) + E\left[e_t \sum_{s=0}^{h} \rho_1^s e_{t+h-s}\right],$$

in which (fortunately) most of the terms are uncorrelated:

$$E(y_t y_{t+h}) = \rho_1^{h+2} \sigma_y^2 + 0 + 0 + E\left[e_t \sum_{s=0}^{h} \rho_1^s e_{t+h-s}\right] = \rho_1^{h+2} \sigma_y^2 + \rho_1^h \sigma_e^2,$$

where the 2nd equality uses the assumption that the disturbances are uncorrelated across time periods.

$$E(e_t e_{t+h-s}) = \begin{cases} \sigma_e^2, & s = h \\ 0, & s \neq h. \end{cases}$$

# Optional: autocovariance of AR(1) process (concluded)

Now the whole expression can be written in terms of the variance of *y*.

$$E(y_t y_{t+h}) = \rho_1^{h+2}\sigma_y^2 + \rho_1^h\sigma_y^2(1-\rho_1^2) = \sigma_y^2\rho_1^h(\rho_1^2 + 1 - \rho_1^2) = \sigma_y^2\rho_1^h.$$

With this result, it is possible to make some sense of how persistent *y* is.

Back.

# Optional: biasedness of OLS

The proof of this is fairly complex, but intuitively the bias comes from the fact that the present period disturbance $e_t$ can be expressed as a function of lagged values of the dependent variable;

- these are correlated with the regressor, $y_{t-1}$, resulting in the bias in finite samples.

But the OLS bias with lagged dependent variables as regressor does disappear with large samples,

- so at least OLS is consistent in these circumstances.

# Optional: biasedness of OLS (sketch of "proof")

The estimator of $\rho$ in an AR(1) regression is:

$$\hat{\rho}_1 = \rho_1 + \frac{\sum_{t=1}^{n} y_{t-1} e_t}{\sum_{t=1}^{n} y_{t-1}^2}.$$

$e_t$ can be written using a lag operator, $(1 - \rho_1 L) y_t = y_t - \rho_1 y_{t-1}$.

# Optional: biasedness of OLS (sketch of "proof")

So,

$$\hat{\rho}_1 = \rho_1 + \frac{\sum_{t=1}^{n} y_{t-1}(1-\rho_1 L)y_t}{\sum_{t=1}^{n} y_{t-1}^2}\text{, which is where the bias comes from.}$$

I.e., $E(y_{t-1}y_t) \neq 0$.

But when the sample size *n* gets large, the term in parentheses goes to zero, rendering the estimator consistent.

This explanation should be taken with a grain of salt because it is intended merely to illustrate the difference between Assumptions TS.3 and TS.3'.

Back.