Chapter 8: Regression with Lagged Explanatory Variables

- **Time series data: Y_t for t=1,..,T**

- **End goal: Regression model relating a dependent variable to explanatory variables.**

With time series new issues arise:

1. **One variable can influence another with a time lag.**

2. **If the data are <u>nonstationary</u>, a problem known as <u>spurious regression</u> may arise.**

- **You will not understand 2. at this stage.**

- **In this chapter, we focus on 1.**

- **Assume data are stationary (explain later what this means).**

The Regression Model with Lagged Explanatory Variables

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \ldots + \beta_q X_{t-q} + e_t$$

- **Multiple regression model with current and past values (*lags*) of X used as explanatory variables.**
- **q = *lag length = lag order***
- **OLS estimation can be carried out as in Chapters 4-6.**
- **Statistical methods same as in Chapters 4-6.**
- **Verbal interpretation same as in Chapter 6.**

Ex. "β_2 measures the effect of the explanatory variable *2 periods ago* on the dependent variable, *ceteris paribus*".

Aside on Lagged Variables

- **X_t is the value of the variable in period t.**

- **X_{t-1} is the value of the variable in period t-1 or "lagged one period" or "lagged X".**

<u>Defining X and lagged X in a spreadsheet</u>

"X"	"lagged X"
X_2	X_1
X_3	X_2
X_4	X_3
.	.
.	.
.	.
.	.
.	.
.	.
X_T	X_{T-1}

- **Each column will have T-1 observations.**

- **In general, when creating "X lagged q periods" you will have T-q observations.**

Example: Lagged Variables

T = 10

$$Y_t = \alpha + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \beta_4 X_{t-3} + e_t \, .$$

	Col. A	Col. B	Col. C	Col. D	Col. E
	Y	X	X lagged 1 period	X lagged 2 periods	X lagged 3 periods
Row 1	Y_4	X_4	X_3	X_2	X_1
Row 2	Y_5	X_5	X_4	X_3	X_2
Row 3	Y_6	X_6	X_5	X_4	X_3
Row 4	Y_7	X_7	X_6	X_5	X_4
Row 5	Y_8	X_8	X_7	X_6	X_5
Row 6	Y_9	X_9	X_8	X_7	X_6
Row 7	Y_{10}	X_{10}	X_9	X_8	X_7

Example: Long Run Prediction of a Stock Market Price Index

The issue of whether stock market returns are predictable is a very important (but difficult) one in finance.

This is not a book on financial theory and, hence, we will not describe the theoretical model which motivates this example.

Variables: stock prices, dividends and returns.

The basic equation relating these three concepts is:

$$\text{Return} = R_t = \frac{(P_t - P_{t-1} + D_t)}{P_{t-1}} \times 100,$$

where R_t is the return on holding a share from period t-1 through t,

P_t is the price of the stock at the end of period t

D_t is the dividend earned between period t-1 and t.

This relationship, along with assumptions about how these variables evolve in the future, can be used to develop various theoretical financial models.

One example: the ratio of dividends to stock price should have some predictive power for future returns, particularly at long horizons.

How does such a theory relate to our regression model with lagged explanatory variables?

Dependent variable (Y) is the total return on the stock market index over a future period but the explanatory variable (X) is the current dividend-price ratio.

$$Y_{t+h} = \alpha + \beta X_t + e_{t+h},$$

h is forecast horizon

Y_{t+h} is calculated using the returns R_{t+1}, R_{t+2},.., R_{t+h}.

Equivalently:

$$Y_t = \alpha + \beta X_{t-h} + e_t .$$

This is a specialized version of the regression model with lagged explanatory variables.

Financial theory suggests that the explanatory power for this regression should be poor at short horizons (e.g. h=1 or 2) but improve at longer horizons.

Our data (monthly)
Y = twelve month returns (i.e. h=12) from a stock market

X = dividend-price ratio (twelve months ago).

	Coeff	t Stat	P-value	Lower 95%	Upper 95%
Inter.	-0.003	-0.662	0.508	-0.013	0.006
X_{t-12}	0.022	4.833	1.5E-6	0.013	0.032

Dividend-price ratio does have significant explanatory power for twelve month returns (since P-value less than .05).

Theory that dividend-price ratio has some predictive power for long run returns is supported.

However, R^2=0.019 indicating that this predictive power is weak.

Only 1.9% of the variation in twelve month returns can be explained by the dividend-price ratio.

Example: The Effect of Bad News on Market Capitalization

Motivation: Share price of a company can be sensitive to bad news.

E.g. Company B is in an industry which is particularly sensitive to the price of oil.

If the price of oil goes up, then the profits of Company B will tend to go down and some investors, anticipating this, will sell their shares in Company B driving its price (and market capitalization) down.

However, this effect might not happen immediately so lagged explanatory variables might be appropriate.

Monthly data for 5 years (i.e. 60 months) on the following variables:

- **Y = market capitalization of company ($000)**
- **X = price of oil (dollars per barrel)**

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + e_t .$$

Example: The Effect of Bad News on Market Capitalization (cont.)

Results:

	Coeff.	St. Err.	t Stat	P-val	Lower 95%	Upper 95%
Inter.	92001.5	2001.7	45.96	6.E-42	87979	96024.
X_t	-145.0	47.6	-3.04	.0037	-240.7	-49.3
X_{t-1}	-462.1	47.7	-9.70	6E-13	-557.9	-366.4
X_{t-2}	-424.5	46.2	-9.19	3.E-12	-517.3	-331.6
X_{t-3}	-199.6	47.8	-4.18	.0001	-295.5	-103.6
X_{t-4}	-36.9	47.5	-.78	.44	-132.3	58.5

Example: The Effect of Bad News on Market Capitalization (cont.)

What can the company conclude about the effect of the price of oil on its market capitalization?

Increasing the price of oil by $1 per barrel in a given month is associated with:

- **An immediate reduction in market capitalization of $145,000, *ceteris paribus*.**
- **A reduction in market capitalization of $462,140 one month later, *ceteris paribus*.**
- **A reduction in market capitalization of $424,470 two months later, *ceteris paribus.***
- **A reduction in market capitalization of $199,550 three months later, *ceteris paribus*.**
- **A reduction in market capitalization of $36,900 four months later, *ceteris paribus*.**

Example: The Effect of Bad News on Market Capitalization (cont.)

Intuition about what the *ceteris paribus* condition implies:

"Increasing the oil price by one dollar in a given month will tend to reduce market capitalization in the following month by $462,120, *assuming that no other change in the oil price occur.*"

Total effect = $145,000 + $462,140 + $424,470 + $199,550 + $36,900 = $1,268,060

"After four months, the effect of adding one dollar to the price of oil is to decrease market capitalization by $1,268,060".

Selection of Lag Order

How to choose q (lag length)?

One way: Use t-tests discussed in Chapter 5 sequentially (another way is to use information criteria which we will discuss later).

<u>Step 1</u>

Choose the maximum possible lag length, qmax, that seems reasonable to you.

<u>Step 2</u>

Estimate the model:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \ldots + \beta_{q\max} X_{t-q\max} + e_t .$$

If the P-value for testing $\beta_{qmax}=0$ is less than the significance level you choose (e.g. .05) then go no further. Use qmax as lag length. Otherwise go on to the next step.

Selection of Lag Order (cont.)

Step 3

Estimate the model:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \ldots + \beta_{q\max-1} X_{t-q\max+1} + e_t\,.$$

If the P-value for testing β_{qmax-1}=0 is less than the significance level you choose (e.g. .05) then go no further. Use qmax-1 as lag length. Otherwise go on to the next step.

Step 4

Estimate the model:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \ldots + \beta_{q\max-2} X_{t-q\max+2} + e_t\,.$$

If the P-value for testing β_{qmax-2}=0 is less than the significance level you choose (e.g. .05) then go no further. Use qmax-2 as lag length. Otherwise go on to the next step, etc.

Aside: Lag Length

- **The number of observations used in a model with lagged explanatory variables is equal to the original number of observations, T, minus the maximum lag length.**

- **In Step 2 you have T-qmax observations**

- **In Step 3, T-qmax+1 observations, etc.**

Example: The Effect of Bad News on Market Capitalization (cont.)

- **Suppose qmax=4**

- **P-value for X_{t-4} = .44>.05 (see previous table)**

- **Drop X_{t-4} and re-estimate with q = 3.**

	Coeff.	St. Err.	t Stat	P-value	Lower 95%	Upper 95%
Inter.	90402.2	1643.18	55.02	9.E-48	87104.9	93699.5
X_t	-125.90	46.24	-2.72	.0088	-218.69	-33.11
X_{t-1}	-443.49	45.88	-9.67	3.E-13	-535.56	-351.42
X_{t-2}	-417.61	45.73	-9.13	2.E-12	-509.38	-325.84
X_{t-3}	-179.90	46.25	-3.89	.0003	-272.72	-87.09

- **P-value for X_{t-3} is .0003 < .05.**

- **Select q=3 and present this table in a report.**

Chapter Summary

1. Regressions with time series variables involve two issues we have not dealt with in the past. First, one variable can influence another with a time lag. Second, if the variables are non-stationary, the spurious regressions problem can result. The latter issue will be dealt with later on.

2. Distributed lag models have the dependent variable depending on an explanatory variable and lags of the explanatory variable.

3. If the variables in the distributed lag model are stationary, then OLS estimates are reliable and the statistical techniques of multiple regression (e.g. looking at P-values or confidence intervals) can be used in a straightforward manner.

4. The lag length in a distributed lag model can be selected by sequentially using t-tests beginning with a reasonably large lag length.