AFRE 835: Introductory Econometrics

Chapter 10: Basic Time Series Regression Analysis

Spring 2017

Introduction

- This chapter provides a basic introduction to regression analysis with time series data.
- The key distinction here is that there is now a temporal ordering to the data.
- In the cross-sectional setting, the order in which the data were listed was irrelevant, as long as they represented a random sample from the population of interest.
- The notion of randomness is more complex.
- We now have a sequence of random variables, known as a **stochastic process** (or a **time series process**).
- In many ways, what we observe is a single *thread* or realization of that process.

TABLE 10.1 Partial Listing of Data on U.S. Inflation and Unemployment Rates, 1948–2003

Year	Inflation	Unemployment
1948	8.1	3.8
1949	−1.2	5.9
1950	1.3	5.3
1951	7.9	3.3
.	.	.
.	.	.
.	.	.
1998	1.6	4.5
1999	2.2	4.2
2000	3.4	4.0
2001	2.8	4.7
2002	1.6	5.8
2003	2.3	6.0

Outline

1. Examples of Time Series Regression Models
2. Finite Sample Properties of OLS under Classical Assumptions
3. Functional Form and Dummy Variables
4. Trends and Seasonality

The Static Model

- The simplest of time series models is the **static model**.
- As the name suggests, the model is little different from the cross-sectional models we've considered thusfar;
  - The model does involve time series data, but makes no real use of it.
  - The relationship between our dependent variable and the explanatory variables is a *contemporaneous* one only.
  - Specifically, with a single independent variable, we have a population model
$$y_t = \beta_0 + \beta_1 z_t + u_t \quad t = 1, \ldots, n \tag{1}$$
  - The interpretation of the coefficients follows the same pattern as in the cross-sectional setting.
- The static model is not frequently used or applicable.

Finite Distributed Lag Models

- A more common specification is to allow one or more of our regressors to impact y with a lag.
- For example, a **finite distributed lag (FDL) model** of order two, would specify
$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t \quad t = 1, \ldots, n \tag{2}$$
- Wooldridge gives the example of how a regressor pe_t (denoting the size of the personal exemption in federal taxes) takes time to impact gfr_t denoting the general fertility rate.
- One might also expect time delays in how
  - faculty hires impact the ranking of an economics department;
  - investments in expenditures per student impact high school graduation rates;
  - voting rights laws impact diversity in elected officials.

Finite Distributed Lag Models - Temporary Changes

- Interpreting finite distributed lag models is more difficult. The effect of a change at one point in time ripples through several time periods.
- With a temporary change in a regressor at time t, say from $z = c$ to $z = c + 1$, the impact will ripple through time with an effect dictated by the δ_j's.

$$\begin{aligned}
y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\
y_t &= \alpha_0 + \delta_0 (c+1) + \delta_1 c + \delta_2 c \\
y_{t+1} &= \alpha_0 + \delta_0 c + \delta_1 (c+1) + \delta_2 c \\
y_{t+2} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 (c+1) \\
y_{t+3} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c
\end{aligned}$$

  where we've set $u_t = 0$ to simplify the comparisons.
- The immediate impact of the change is measured by δ_0, sometimes referred to as the **impact propensity** or **impact multiplier**.

Finite Distributed Lag Models - Permanent Changes

- A permanent shift in a regressor to a new level takes time to accumulate, with

$$\begin{aligned}
y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\
y_t &= \alpha_0 + \delta_0 (c+1) + \delta_1 c + \delta_2 c \\
y_{t+1} &= \alpha_0 + \delta_0 (c+1) + \delta_1 (c+1) + \delta_2 c \\
y_{t+2} &= \alpha_0 + \delta_0 (c+1) + \delta_1 (c+1) + \delta_2 (c+1)
\end{aligned}$$

- The implications of the permanent change (holding u unchanged) is given by

$$\frac{\Delta y_s}{\Delta z_t} = \begin{cases} \delta_0 & s = t \\ \delta_0 + \delta_1 & s = t+1 \\ \delta_0 + \delta_1 + \delta_2 & s \geq t+2 \end{cases} \tag{3}$$

- More generally, in a FDL model of order q, the **long-run propensity (LRP)** or **long-run multiplier** is given by $\delta_0 + \delta_1 + \cdots + \delta_q$.

Finite Sample Properties of OLS under Classical Assumptions

- This section describes the finite sample properties of OLS in a time series setting, much like we considered for the cross section setting.
- As we'll see, we need particularly strong assumptions in this setting for OLS to remain unbiased.
- The assumptions parallel many of those developed in the cross-section setting.

Assumptions TS.1 and TS.2

- Assumption TS.1 (Linear in Parameters): The stochastic process $\boldsymbol{x}_t, y_t) : t = 1, \ldots, n$ follows the linear model:

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t \tag{4}$$

  where $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tk})$ is a row vector of the regressors for time period t and $u_t : t = 1, \ldots, n$ denotes a sequence of errors.

  ... This parallels assumptions MLR.1.
- Assumption TS.2 (No Perfect Collinearity): In the sample (and therefore the underlying time series process), no independent variable is constant nor a perfect linear combination of others.

  ... This parallels assumptions MLR.3.

Zero Conditional Mean

- Assumption TS.3 (Zero Conditional Mean): For each t, the expected value of the error u_t, given the explanatory variables for *all* time periods, is zero; i.e.,

$$E(u_t|\boldsymbol{X}) = 0 \quad t = 1, \dots, n. \tag{5}$$

  where $\boldsymbol{X} = (\boldsymbol{x}_1', \dots, \boldsymbol{x}_n')'$ is $n \times k$ matrix containing the regressors for all time periods.
- Assumption TS.3 is what is known as **strict exogeneity**.
- It implies that u_t is uncorrelated with *any* x_{sj} in *any* time periods.
- This is a much stronger assumption than what is known as **contemporaneous exogeneity**: $E(u_t|\boldsymbol{x}_t) = 0$.
- The distinction was not an issue in the cross-section setting due to the random sample assumption (MLR.2), which precluded correlations across observations.

Understanding Strict Exogeneity

- Strict exogeneity does not restrict correlation in the error terms over time.
- Nor does it restrict correlation patterns among the regressors.
- It does, however, preclude any correlation between *any* u_t and *any* x_{sj}.
- As in the cross-sectional setting, TS.3 could be violated if we have omitted variables or measurement errors.
- TS.3 could be violated if there any temporal feedback.

Unbiasedness of OLS

- **Theorem 10.1** (Unbiasedness of OLS): Under Assumptions TS.1, TS.2, and TS.3, the OLS estimators are unbiased conditional on $\boldsymbol{X}$ and therefore unconditionally as well.

Additional Assumptions and Theorem for Time Series Data

- Assumption TS.4 (Homoskedasticity): Conditional on $\boldsymbol{X}$, $Var(u_t|\boldsymbol{X}) = Var(u_t) = \sigma^2$.
- Assumption TS.5 (No Serial Correlation): Conditional on $\boldsymbol{X}$, the errors across time are uncorrelated; i.e., $Corr(u_t, u_s|\boldsymbol{X}) = 0$.

  ...Note: This was not an issue in the cross-sectional setting due to the random sampling assumption.
- **Theorem 10.2** (OLS Sampling Variances): Under the Gauss-Markov Assumptions TS.1 through TS.5: $Var(\hat{\beta}|\boldsymbol{X}) = \sigma^2 / \left[SST_j(1 - R_j^2) \right]$, where SST_j denotes the total sum of squares for x_{tj} and R_j^2 is the R-squared from regressing x_j on all the other regressors.
- **Theorem 10.3** (Unbiased Estimation of σ^2): Under assumptions TS.1 through TS.5, the estimator $\hat{\sigma}^2 = SSR/(n - k - 1)$ is an unbiased estimator of σ^2.

The Gauss-Markov Theorem for Time Series Data

- **Theorem 10.4** (Gauss-Markov Theorem): Under assumptions TS.1 through TS.5, the OLS estimators are best linear unbiased estimators conditional on $\boldsymbol{X}$.
- Assumption TS.6 (Normality): The errors u_t are independent of $\boldsymbol{X}$ and are independent and identically distributed as $\mathcal{N}(0, \sigma^2)$.
- **Theorem 10.5** (Normal Sampling Distributions): Under assumptions TS.1 through TS.6, the CLM assumptions for time series, the OLS estimators are normally distributed conditional on $\boldsymbol{X}$. Further, under the null hypothesis, each t-statistic has a t distributions, and each F-statistic has an F distribution. The usual construction of confidence intervals is valid.

Functional Form

- The lessons regarding function form (e.g., the use and interpretation of logged variables and quadratic terms) apply as well to time series data.
- Consider a static log-log model of consumption as a function of income (problem C7 using consump.dta); i.e.,

$$log(c_t) = \beta_0 + \delta_0 log(y_t) + u_t \tag{6}$$

- The parameter δ_0 measures the elasticity of per capita real consumption (c_t) with respect to per capita real income (y_t).
- A Finite Distributed Lag model (of order 1) would allow for a delay in how consumption responds to changes in income, with

$$log(c_t) = \beta_0 + \delta_0 log(y_t) + \delta_1 log(y_{t-1}) + u_t \tag{7}$$

- In this case, the short-run income elasticity of consumption would be δ_0, whereas the long-run income elasticity would be $\delta_0 + \delta_1$.

$log(c_t)$			
	Static	FDL 1	FDL1 with real int.
$log(y_t)$	0.944	0.637	0.5588
	(0.009)**	(0.116)**	(0.1220)**
$log(y_{t-1})$		0.303	0.3767
		(0.113)*	(0.1185)**
$realint_t$			0.0017
			(0.0010)
_cons	0.329	0.377	0.4146
	(0.082)**	(0.085)**	(0.0851)**
R^2	1.00	1.00	1.00
N	37	36	36

* $p < 0.05$; ** $p < 0.01$

The last column adds the real interest rate as a regressor.

Computing the Long-Run Propensity

- Recall that in a FDL of order q

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \cdots + \delta_q z_{t-q} + u_t \quad t = 1, \ldots, n \tag{8}$$

  the long-run propensity (LRP) is equal to:

$$\theta_0 = \delta_0 + \delta_1 + \cdots + \delta_q \tag{9}$$

- The standard error associated with the *LRP* can be obtained by:
  - Using Stata's *lincom* command.
  - Using $\delta_0 = \theta_0 - \delta_1 - \cdots - \delta_k$ to rewrite the model as

$$y_t = \alpha_0 + \theta_0 z_t + \delta_1(z_{t-1} - z_t) + \cdots + \delta_q(z_{t-q} - z_t) + u_t \quad t = 1, \ldots, n \tag{10}$$

  - Using bootstrapped draws from the joint distribution of the $\hat{\delta}_j$'s to simulate draws from $\hat{\theta}$.

Example Code: Log(Consumption)

```
reg      lc ly;
outreg   using "`TableA'", bdec(3) se tex title(log(consumption))
         ctitle("", static) replace;
tsset    year;
reg      lc ly l.ly;
outreg   using "`TableA'", bdec(3) se tex title(log(wages))
         ctitle("", FDL1) merge;
reg      lc ly l.ly r3;
outreg   using "`TableA'", bdec(4) se tex title(log(wages))
         ctitle("", FDL1 w/r) merge;
lincom   ly+l.ly ;
matrix   M=e(b);
matrix   V=e(V);
drawnorm d0 d1 b1 b0, n(1000) cov(V) means(M);
sum      b0 d0 d1 b1;
gen      theta=d0+d1;
sum      theta;
```


Example Results: Log(Consumption)

- In our FDL1 model of log-consumption, using the lincom command, we get $\widehat{se}(\hat{\theta}) = 0.0088696$
- Using 1000 bootstraps, the simulation-based estimate of the standard error yields $\widehat{se}(\hat{\theta}) = 0.0090643$.
- Using 100000 bootstraps, the simulation-based estimate of the standard error yields $\widehat{se}(\hat{\theta}) = 0.008875$.

Example Output: Log(Consumption)

```
. lincom  ly+l.ly ;

 ( 1)  ly + L.ly = 0

------------------------------------------------------------------------------
          lc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .9354912   .0088696   105.47   0.000     .9174243    .9535581
------------------------------------------------------------------------------

. matrix  M=e(b);

. matrix  V=e(V);

. drawnorm d0 d1 b1 b0, n(100000) cov(V) means(M);
(obs 100,000)

. sum     b0 d0 d1 b1;

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          b0 |    100,000    .4145382    .0851747   .0463891   .7508087
          d0 |    100,000     .558572    .1221505   .0192086   1.108961
          d1 |    100,000    .3769279    .1186137  -.1580337    .882032
          b1 |    100,000    .0016985    .0010001  -.0024806   .0059936

. gen     theta=d0+d1;

. sum     theta;

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       theta |    100,000    .9354999     .008875   .9004927   .9738178
```


Binary (Dummy) Variables

- Binary variables are often used in **event studies**, to isolate the impact of a particular event on outcomes of interest.
- Examples include
  - Financial events such as the Great Depression or Great Recession;
  - Political events such as elections or scandals;
  - Historical events, such as WW1 or WW2;
  - Regulatory events, such as passage of the Clean Air Act;
- Binary variables can be used to isolate subperiods before, after and during an event.

Trending in Time Series

- Many economic time series follow a general upward trend.
- These trends may be due to a variety of factors shared by both independent and dependent variables of interest, such as
  - Population growth;
  - Technological change;
  - Cyclical or trending weather patterns.
- Failing to control for these trends amounts to a form of omitted variables bias and will lead to spurious correlation between two variables.

Linear and Exponential Time Trends

- One popular way to account for underlying trends is to introduce a **linear time trend**:

$$y_t = \alpha_0 + \alpha_1 t + e_t \quad t = 1, \ldots, n. \tag{11}$$

  where e_t is independent and identically distributed ($i.i.d.$).
- If $\alpha_1 > 0$, our dependent variable is growing over time.
- Many economic time series variables are better approximated by an **exponential** trend, with

$$log(y_t) = \beta_0 + \beta_1 t + e_t \quad t = 1, \ldots, n. \tag{12}$$

  ... This is just a log-level model with time as the independent variable, so that (holding $\Delta e_t = 0$)

$$100 \cdot \beta_1 = 100 \cdot \frac{\partial E(y_t|t)}{\partial t} \approx \frac{\%\Delta E(y_t|t)}{\Delta t} \tag{13}$$

Quadratic Trends

- One can, of course, also consider nonlinear trends.
- It is not uncommon to allow for a quadratic trend, with

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + e_t \quad t = 1, \ldots, n. \tag{14}$$

Example: Housing Investment as a Function of Price

	$log(invpc_t)$		
	static	FDL1	FDL1
$log(price_t)$	1.241	-0.381	3.260
	(0.382)**	(0.679)	(0.960)**
t		0.010	0.013
		(0.004)**	(0.003)**
$log(price_{t-1})$			-4.487
			(0.959)**
_cons	-0.550	-0.913	-1.086
	(0.043)**	(0.136)**	(0.115)**
R^2	0.21	0.34	0.56
N	42	42	41

* $p < 0.05$; ** $p < 0.01$

Some Notes on Time Series Models and Trending

- The OLS estimator in the case of a linear trend can be obtained in two steps:
  1. Forming a *detrended* version of both the dependent and independent variable; e.g., $\ddot{y}_t = y_t - \hat{y}_t$ where $\hat{y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 t$.
  2. Regressing $\ddot{y}_t$ on $\ddot{x}_{t1}, \ddot{x}_{t2}, \ldots, \ddot{x}_{tk}$.

  ... Thus, the inclusion of a linear trend is (mathematically) equivalent to detrending all of the variables prior to OLS.
- The R^2's from time series models will tend to be large due to underlying trends. Looking at the R^2's from the second step above is more informative in terms of fit.

Controlling for Seasonal, Weekly and Hourly Patterns

- In time series data, the underlying data may exhibit patterns that repeat over seasons, months, days of the week, and hours of the day.
- As we are often interested in departures from these routine patterns, it is often appropriate to include binary variables to reflect seasonality (e.g., month dummy variables, weekday dummy variables, etc.).
- Just like the linear time trend can be viewed as *detrending* the data, the inclusion of seasonal dummies effectively creates *seasonally adjusted variables*.