Chapter 10 and Chapter 11: Basic Time Series Regression

1. Time series data have temporal ordering. Past can affect future, not vice versa. So we cannot switch (change the ordering of) the observations. That is the biggest difference from the cross sectional data.

2. In time series data there is a variable that serves as index for the time periods. That variable also indicates the frequency of observations. For instance, we have yearly data if one observation becomes available once a year. The stata command `tsset` declares the data are time series.

3. The regression that uses time series can have (dynamic) causal interpretation, provided that the regressor is uncorrelated with the error term. More often, time series are used for forecasting purpose.

4. Consider the causal study first.

   (a) For simplicity a simple regression is considered

   $$Y_t = \beta_0 + \beta_1 X_t + u_t \tag{1}$$

   where subscript t is used to emphasize that data are time series.

   (b) Model (1) is a static model because it only captures the immediate or contemporaneous effect of X on Y. When X_t changes by one unit, it only has effect on Y_t (in the same period). In other words, Y_{t+1}, Y_{t+2}, and so on are unaffected. In light of this, model (1) is static.

   (c) It is easy to account for lag effect by using lag value. There are first lag, second lag, and so on. The table below is one example

t	Y_t	Y_{t-1}	Y_{t-2}	ΔY_t
1	2.2	na	na	na
2	3	2.2	na	0.8
3	-2	3	2.2	-5

   i. t is the index series. $t = 1$ corresponds to the first (or earliest) observation. By default, the first observation should be the earliest.

   ii. Y_{t-1} is the first lag of Y_t. When $t = 2$, $Y_{t-1} = Y_{2-1} = Y_1 = 2.2$. So the second observation of Y_{t-1} is the first observation of Y_t. The first observation of Y_{t-1}

is missing value (denoted by na) because there is no data for Y_0. Essentially you get the first lag by pushing the whole series one period down.

iii. Y_{t-2} is the second lag of Y_t. You get the second lag by pushing the whole series two periods down (and two missing values are generated).

iv. ΔY_t is the first difference. By definition,

$$\Delta Y_t \equiv Y_t - Y_{t-1} \tag{2}$$

for example, when $t = 2$, $\Delta Y_2 \equiv Y_2 - Y_{2-1} = 3 - 2.2 = 0.8$. ΔY_1 is missing since Y_0 is missing

v. $\boxed{\text{Exercise}}$: Use the above table and find ΔY_{t-1} for $t = 1, 2, 3$

vi. The stata commands to generate the first lag and second lag are

```
gen ylag1 = y[_n-1]
gen ylag2 = y[_n-2]
```

vii. Alternatively you can refer to the first and second lags by using the lag operator `L` (after declaring time series using `tsset`)

`L.y, L2.y`

(d) The distributed lag (DL) model uses lag value of X to account for the lag effect. A DL model with two lags are

$$Y_t = \beta_0 + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + u_t \tag{3}$$

where the parameter δ is called (short-run) multiplier.

i. It follows that a change in X_t has effect on Y_t, Y_{t+1} and Y_{t+2}. Then the effect of X_t vanishes.

ii. Mathematically, we can show

$$
\begin{aligned}
\frac{dY_t}{dX_t} &= \delta_0 && (4)\\
\frac{dY_{t+1}}{dX_t} &= \delta_1 && (5)\\
\frac{dY_{t+2}}{dX_t} &= \delta_2 && (6)\\
\frac{dY_{t+j}}{dX_t} &= 0, \quad (j > 2) && (7)
\end{aligned}
$$

where $\frac{dY_{t+1}}{dX_t}$ can be obtained from the regression where Y_{t+1} is the dependent variable.

iii. When we graph δ_j as a function of j, we obtain the lag distribution.

iv. The cumulative effect of X_t on Y_t, Y_{t+1} and Y_{t+2} is called long run propensity (LRP) or long run multiplier.

$$LRP \equiv \delta_0 + \delta_1 + \delta_2 \tag{8}$$

There are two steps to obtain the estimate and standard error for LRP. First, write $\theta = \delta_0 + \delta_1 + \delta_2$, which implies that $\delta_0 = \theta - \delta_1 - \delta_2$. Then equation (3) becomes

$$
\begin{aligned}
Y_t &= \beta_0 + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + u_t && (9)\\
&= \beta_0 + (\theta - \delta_1 - \delta_2) X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + u_t && (10)\\
&= \beta_0 + \theta X_t + \delta_1 (X_{t-1} - X_t) + \delta_2 (X_{t-2} - X_t) + u_t && (11)
\end{aligned}
$$

The last equation suggests running the transformed regression of Y_t onto X_t, $(X_{t-1} - X_t)$ and $(X_{t-2} - X_t)$. Then $\hat{\theta} = \widehat{LRP}$, $\texttt{se}(\hat{\theta}) = \texttt{se}(\widehat{LRP})$ and the confidence interval for θ is the confidence interval for LRP.

v. Due to multicollinearity, it can be difficult to obtain precise estimates of the short run multiplier δ_j. However we can often get good estimate of LRP.

vi. To mitigate multicollinearity (which is more common for time series data), we can impose some restriction on the lag distribution. For example, if we assume the lag distribution is flat, i.e., $\delta_0 = \delta_1 = \delta_2 = \delta$ then model (3) reduces to

$$Y_t = \beta_0 + \delta Z + u_t, \quad (Z \equiv X_t + X_{t-1} + X_{t-2}) \tag{12}$$

(e) In theory if the effect of X never dies out, we need to include infinite lag value of X, which is infeasible. Instead we may consider an autoregressive distributed lag model (ADL) given by

$$Y_t = \beta_0 + \rho Y_{t-1} + \delta X_t + u_t \tag{13}$$

By recursive substitution

$$\begin{aligned} Y_{t+1} &= \beta_0 + \rho Y_t + \delta X_{t+1} + u_{t+1} & (14) \\ &= \beta_0 + \rho(\beta_0 + \rho Y_{t-1} + \delta X_t + u_t) + \delta X_{t+1} + u_{t+1} & (15) \\ &= \ldots + \rho\delta X_t + \ldots & (16) \end{aligned}$$

we can show the effect of X on Y is

$$\frac{dY_t}{dX_t} = \delta, \quad \frac{dY_{t+1}}{dX_t} = \rho\delta, \quad \ldots, \quad \frac{dY_{t+j}}{dX_t} = \rho^j \delta \tag{17}$$

Alternatively you use apply the chain rule of calculus and get

$$\begin{aligned} \frac{dY_{t+1}}{dX_t} &= \frac{dY_{t+1}}{dY_t}\frac{dY_t}{dX_t} = \rho\delta & (18) \\ \frac{dY_{t+2}}{dX_t} &= \frac{dY_{t+2}}{dY_{t+1}}\frac{dY_{t+1}}{dX_t} = \rho(\rho\delta) = \rho^2\delta & (19) \end{aligned}$$

The effect of X decreases exponentially if $|\rho| < 1$, but never becomes zero no matter how large j is.

5. We get the autoregressive (AR) model (or autoregression) if X is excluded from the ADL model. More explicitly, a first order autoregressive model is

$$Y_t = \beta_0 + \rho Y_{t-1} + u_t \tag{20}$$

(a) The AR model is more suitable for forecasting than ADL and DL models since we do not need to forecast X before forecasting Y. This is because in AR model the regressor is just the lagged dependent variable.

(b) The AR model can be used to test serial correlation. A series is serially correlated if $\hat{\rho}$ in (20) is significantly different from 0 (when its p-value is less than 0.05).

(c) (Optional) The time series has a unit root (and becomes non-stationary) if it is extremely serially correlated (or highly persistent). In theory, a series has unit root if $\rho = 1$ in (20). In practice $\hat{\rho} \approx 1$ is suggestive of unit root. We can show the variance of a unit root series goes to infinity as t rises. The conventional law of large number and central limit theorem both require finite variance, so neither apply to a unit root series. That means in general we need to difference a unit root series (to make it stationary) before using it in regression. One exception is, you do not need to difference unit root series if they are cointegrated. The unit root series is also called random walk.

(d) The AR model with one lag can be fitted by stata command

```
reg y L.y
```

(e) Then the one-period ahead forecast is computed as

$$\hat{Y}_{t+1} = \hat{\beta}_0 + \hat{\rho} Y_t \tag{21}$$

and the two-period ahead forecast is

$$\hat{Y}_{t+2} = \hat{\beta}_0 + \hat{\rho} \hat{Y}_{t+1} \tag{22}$$

In general the h-period ahead forecast is

$$\hat{Y}_{t+h} = \hat{\beta}_0 + \hat{\rho} \hat{Y}_{t+h-1} \quad (h = 2, 3, \ldots) \tag{23}$$

In short all forecasts can be computed in an iterative (recursive) fashion.

(f) The AR model with two lags can be fitted by stata command

```
reg y L.y L2.y
```

(g) We can add third lag, fourth lag, and so on until the new lag has insignificant coefficient.

6. If the time series is trending, we can add a linear trend (and quadratic trend if neces-

sary) into the AR model and run below regression:

$$Y_t = \beta_0 + \rho Y_{t-1} + ct + u_t \tag{24}$$

where t is the linear trend. According to the Frisch-Waugh Theorem, the estimated coefficient $\hat{\rho}$ in (24) can be obtained by regressing $Y_t^{\texttt{detrended}}$ onto $Y_{t-1}^{\texttt{detrended}}$, where $Y_t^{\texttt{detrended}}$ denotes the detrended series, and it is the residual of regressing Y_t on the trend. In short, $\hat{\rho}$ in (24) measures the effect on the deviation of the series around its trend.

7. The trend term can also be used in DL model as proxy for unobserved trending factors.

8. If we have monthly or quaterly data, there may exist regular up and down pattern at specific interval. This phenomenon is called seasonality. For instance, the sale data typically go up in November and December. It is easy to account for seasonality by using seasonal dummy that equals one during a particular season.

9. A structural break may happen at a particular date t_{br}. For example, the expected value of Y_t may increase a lot after t_{br}. In the time series plot, there is a jump at t_{br}. We can use dummy variable to capture that break. Moreover, we can apply the Chow test for the null hypothesis that there is no break (eg., there is no change in the expected value before and after t_{br}). In this context, the Chow test is called the test for structural break.

10. To summarize, a time series can be characterized by (1) whether there is a trend; (2) whether there is seasonality (regular up-and-down movement); (3) how strong the current is related to the past (or how strong is the serial correlation); and (4) whether there is structural break. It is a good idea to draw the time-series plot as it can inform us about what kind of regressors are needed. In general time series data are not iid, because current is correlated with the past (so independence is violated).