ØAMET4100 · Spring 2019
Lecture Note 10

Instructor: Fenella Carpena

March 28, 2019

This lecture note provides a review of time series data (Stock & Watson, Chapter 14). This lecture note is not intended to be a comprehensive review of lecture or the textbook, since there is a lot more material than we have time to cover. However, I have tried to focus on the concepts which I believe are necessary to be successful in our class.

1 Review from Last Week

Recall that one of the key assumptions in a time series regression in stationarity, which is equivalent of the "identically distributed" part of the i.i.d. assumption that we have seen before.

Stationary means that the probability distribution does not change over time. Hence, if the random variables $(Y_t, X_{1,t}, \ldots, X_{k,t})$ have a stationary distribution, then the joint distribution when $t = 1$ is the same as when $t = 2$, etc. In words, this means that the distribution of the random variable (as well as the mean, variance, etc.) in the first period is the same as its distribution in the next period, etc. Intuitively, stationarity requires that the future is like the past, in a probabilistic sense.

Why do we care about stationarity? If a series is non-stationary, conventional hypothesis tests, confidence intervals and forecasts can be unreliable. The precise problem caused by non-stationarity and the solution to that problem depends on the specific type of non-stationarity. In this class, we will consider two important types of non-stationarity in time series data: (1) trends, and (2) breaks.

2 Non-stationarity: Trends

A series is said to exhibit a **trend** if it has a persistent long-term movement. There are two types of trends in time series data: **deterministic** and **stochastic** trends. A trend is **deterministic** if it is a nonrandom function of time. For example, suppose that the GDP growth rate increases by 0.75 percentage points per quarter. This trend can be written as $0.75t$, where t is measured in quarters. Here, $0.75t$ is a deterministic trend: the trend is predictable. In contrast, a trend is said to be **stochastic** if it is a random function of time.

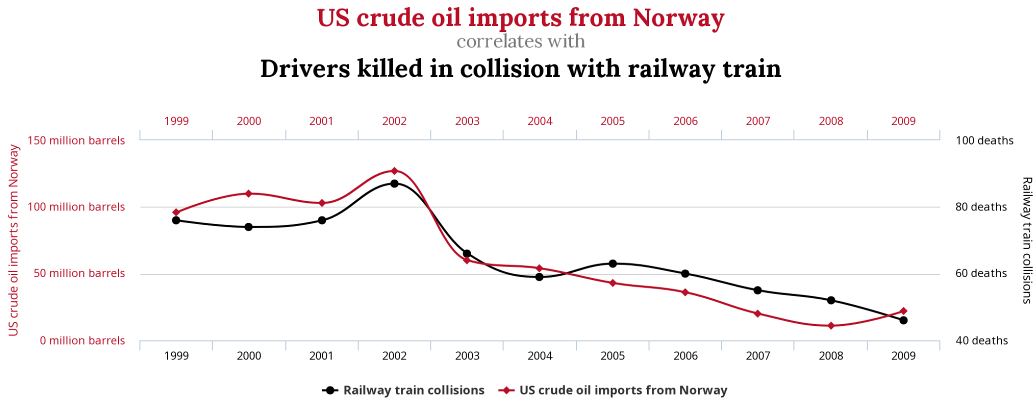
Why do we care about trends? First, many economic time series show a trending behavior, so this needs to be taken account when building forecasting models. The trend in most time series (e.g., GDP, inflation, etc.) is probably best modeled using stochastic (rather than deterministic) trends. This is because the world we live in is very complicated, so it is hard to imagine that the trends of GDP, unemployment, interest rates, etc. are predictable. For this reason, we focus much of our time studying stochastic trends.

Second, OLS estimation of the coefficients on regressors that have a stochastic trend is problematic and raises three important problems.

(a) AR coefficients are biased toward zero: If Y_t is a random walk, the OLS estimate of β_1 is biased toward zero. In particular, $E(\widehat{\beta}_1) = 1 - 5.3/T$. Hence, the bias can be substantial if the sample size (i.e., number of periods in the time series) is small. This estimation bias causes forecasts of Y_t to perform worse than a pure random walk model. The random walk is discussed in more detail in Section 2.1.

(b) Non-normal distribution of t-statistics: If a regressor has a stochastic trend, then its usual OLS t-statistic can have a non-normal distribution under the null hypothesis, even in large samples. This non-normal distribution means that conventional confidence intervals are not valid, and hypothesis tests cannot be conducted as usual.

(c) Spurious regression: Two series that are independent and each have stochastic trends might appear to be related. This is what econometricians call a **spurious** relationship. The following figure illustrates a spurious relationship.



The time series "US crude oil imports from Norway" (red) is almost perfectly correlated with the time series "Drivers killed in collision with railway train" (black). This high correlation can be explained by the fact that both series have stochastic trends that are similar, likely by chance. Since there is no compelling reason to think that both trends are related, the correlation may be spurious.

2.1 Random Walk Model of a Trend

The simplest way to model a time series Y_t that has a stochastic trend is a random walk

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

where the u_t are i.i.d. errors with $E(u_t|Y_{t-1}, Y_{t-2}, \ldots) = 0$. There are several key points to understand about the random walk model.

1. Note that $E(Y_t|Y_{t-1}, Y_{t-2}, \ldots) = E(Y_{t-1}|Y_{t-1}, Y_{t-2}, \ldots) + E(u_t|Y_{t-1}, Y_{t-2}, \ldots) = Y_t$. This means that in a random walk model, the best forecast for Y_t is the previous period's observation, Y_{t-1}. Hence, the difference between Y_t and Y_{t-1} is unpredictable. The path followed by Y_t consists of random steps u_t, hence, the equation is called a random walk.

2. We could extend equation (1) by adding a constant:

$$Y_t = \beta_0 + Y_{t-1} + u_t.$$

   This model is called a **random walk model with drift**. The "drift" allows us to model the tendency of a series to move in one direction. If $\beta_0 > 0$, then the series drifts upwards, and if $\beta_0 < 0$, then the series drifts downwards.

3. A random walk is non-stationary. To see this, note that $var(Y_t) = var(Y_{t-1}) + var(u_t)$. If Y_t were stationary, we need to have $var(Y_t) = var(Y_{t-1})$, and this would require that $var(u_t) = 0$ (i.e., that u_t is constant for all t, which would be a non-stochastic trend). Thus, it is impossible to have a random walk that is stationary. Another way to see that the random walk is non-stationary is the following. Suppose that Y_0, the starting value of the random walk, is zero. We can then write the time series as

$$\begin{aligned}
Y_0 &= 0 \\
Y_1 &= 0 + u_1 \\
Y_2 &= 0 + u_1 + u_2 \\
&\vdots \\
Y_t &= \sum_{i=1}^{t} u_i.
\end{aligned}$$

Therefore, we have $var(Y_t) = var(u_1 + u_2 + \ldots + u_t) = t\sigma_u^2$. Thus, the variance of a random walk depends on t, so the random walk is not stationary.

4. A random walk is a special case of the AR(1) model, with $\beta = 1$. What is this important? Since a random walk is non-stationary, we can see that if $\beta = 1$ in an AR(1) model, then Y_t is non-stationary. Consequently, if $|\beta| < 1$ and u_t are stationary, then Y_t is stationary.

More generally, **when is AR(p) stationary?** Here, stationarity is linked to the roots of the polynomial

$$1 - \beta_1 z - \beta_2 z^2 - \beta_3 z^3 - \ldots - \beta_p z^p. \tag{2}$$

The roots of this polynomials are the values of z that set the above equation equal to zero (e.g., the roots of $3z^2 - 27$ are $+3$ and -3, where we have two roots because we have a second-order polynomial).

For an AR(p) to be stationary, the roots of equation (2) must all be greater than 1 in absolute value. If at least one root of the AR(p) equals one, then the series is said to have a **unit root**, and if Y_t has a unit root, then it contains a stochastic trend and it is non-stationary. If Y_t is stationary (and thus does not have a unit root), it does not contain a stochastic trend. Hence, we use the terms **stochastic trend** and **unit root** interchangeably.

As an example, let's return to an AR(1) model. **What is the root of AR(1)?** The root is the value of z that solves $1 - \beta_1 z = 0 \implies z = 1/\beta_1$. **When is AR(1) stationary?** For AR(1) to be stationary, the root must be greater than 1 in absolute value. Thus, stationarity occurs when $|1/\beta_1| > 1 \implies 1 > |\beta_1|$, which is exactly the condition that was mentioned in point 4 above.

2.2 Dickey-Fuller Test

How can we detect stochastic trends? We can do so by testing for unit root. Specifically, the test we implement is the Dickey-Fuller test, and it tells us whether there is evidence of stochastic trends in the data. Here, our null hypothesis is that the time series has a stochastic trend (i.e., had a unit root), and the alternative hypothesis is that the time series is stationary.

2.2.1 Dickey-Fuller Test: AR(1) model

As discussed above, a time series that follows an AR(1) model with $\beta = 1$ has a stochastic trend. Thus, testing for a unit root is equivalent to the hypothesis test

$$H_0 : \beta_1 = 1, H_1 : |\beta_1| < 1$$

in the regression model

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t.$$

The null hypothesis is that the AR(1) model has a unit root and the alternative hypothesis is that it is stationary. In practice, we rewrite this hypothesis test by subtracting Y_{t-1} from both sides of the AR(1) model to obtain

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$$

where $\delta = \beta_1 - 1$. The hypothesis test then becomes

$$H_0 : \delta = 0, H_1 : \delta < 0.$$

The OLS t-statistic testing $\delta = 0$ is called the **Dicky-Fuller Statistic**, which I write as DF-statistic or DF-stat for short. The DF-stat should be computed **using homoskedastic** (i.e., non-robust) standard errors because under the null hypothesis of a unit root, the usual "nonrobust" standard errors produce a t-statistic that is, in fact, robust to heterskedasticity.

Note that the formulation above also makes it clear that if Y_t follows a random walk (and is stochastic with $\delta = 0$), then we get $\Delta Y_t = \beta_0 + u_t$. This process is stationary. Thus, one way to handle a stochastic trend in a series is to transform the series to first differences, so that it does not have the trend.

2.2.2 Dickey-Fuller Test: AR(p) model

The Dicky-Fuller test can also be applied to an AR(p) model, and is called the **augmented Dickey-Fuller test (ADF)**. The ADF tests the null hypothesis $H_0 : \delta = 0$ vs. $H_1 : \delta < 0$ in the model

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \ldots + \gamma_p \Delta Y_{t-p} + u_t.$$

Under the null hypothesis, Y_t has a stochastic trend. Under the alternative hypothesis, Y_t is stationary. the ADF statistic is the OLS t-statistic testing $\delta = 0$ in the above regression.

If instead, the alternative hypothesis is that Y_t is stationary around a deterministic linear time trend, then this trend must be added as an additional regressor, and the Dickey-Fuller regression becomes

$$\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \ldots + \gamma_p \Delta Y_{t-p} + u_t$$

where α is an unknown coefficient. The ADF-stat is the OLS t-stat testing $\delta = 0$ in the above equation.

How do we choose the lag length for the ADF regression? Generally, we use the AIC or BIC. Some studies suggest it is better to have too many lags than too few lags, so some might say it is better to use the AIC than the BIC.

What is the critical value that we should compare the ADF statistic to? Under the null hypothesis of unit root, the ADF statistic does **NOT** have a normal distribution, even in large samples. The critical values depend on whether the test is based on a regression with or without a linear time trend. The critical values are given in Table 14.4 of the textbook (reproduced below). Because we have a one-sided test where the alternative hypothesis is $\delta < 0$, we reject the hypothesis of unit root when the ADF stat is less than the critical value.

TABLE 14.4 Large-Sample Critical Values of the Augmented Dickey–Fuller Statistic

Deterministic Regressors	10%	5%	1%
Intercept only	−2.57	−2.86	−3.43
Intercept and time trend	−3.12	−3.41	−3.96

Finally, as before, we must be careful with the conclusions we make in a hypothesis test. Failure to reject the null hypothesis does not mean that the null hypothesis true. It just means that there is no evidence in the data to conclude that the null is false. Hence, even if we fail to reject the null of unit root, it does not mean that the series actually *has* a unit root.

3 Non-stationarity: Breaks

A second type of non-stationarity occurs when the population regression function changes over a given period. These changes are called **breaks**, and they may occur because of changes in economic policy, changes in an industry because of new inventions, etc.

If breaks are not accounted for in the regression model, OLS estimates will reflect the average relationship. Since these estimates might be strongly misleading and result in poor forecast quality, we are interested in testing for breaks. One distinguishes between testing for a break when the date is known and testing for a break with an unknown break date.

3.1 Testing for a break at a known break date

Let τ denote a known break date and let $D_t(\tau)$ be a binary variable indicating time periods before and after the break. Incorporating the break in an ADL(1,1) regression model yields

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1} + \gamma_0 D_t(\tau) + \gamma_1 \left[D_t(\tau) \cdot Y_{t-1}\right] + \gamma_2 \left[D_t(\tau) \cdot X_{t-1}\right] + u_t, \quad (3)$$

where we allow for discrete changes in β_0, β_1 and δ_1 at the break date τ. The null hypothesis is that there is no break, which is given by

$$H_0 : \gamma_0 = \gamma_1 = \gamma_2 = 0.$$

This hypothesis can be tested against the alternative that at least one of the γ's is not zero using an F-test. This idea is called a Chow test, named after Gregory Chow (1960).

Note that in equation (3), we are testing for a break in both regressors Y and X, but it is also possible to test for a break only in X. In this case, the regression would not include the term $\gamma_1 [D_t(\tau) \cdot Y_{t-1}]$.

3.2 Testing for a break at an unknown break date

When the break date is unknown the **Quandt likelihood ratio (QLR)** may be used. It is a modified version of the Chow test which uses the largest of all F-statistics obtained when applying the Chow test for all possible break dates in a predetermined range $[\tau_0, \tau_1]$.

Specifically, the QLR test statistic is the largest Chow $F(\tau)$ statistic computed over a range of eligible break dates $\tau_0 \leq \tau \leq \tau_1$:

$$QLR = \max [F(\tau_0), F(\tau_0 + 1), \ldots, F(\tau_1)]. \tag{4}$$

The QLR test has several important properties, as follows.

1. The large-sample distribution of QLR depends on q, the number of restrictions being tested, and the ratios of end points to the sample size, $\tau_0/T, \tau_1/T$. A common approach is to use 15% trimming, i.e., $\tau_0 = 0.15T$ and $\tau_1 = 0.85T$. Similar to the ADF test, the large-sample distribution of QLR is nonstandard. Critical values are presented in Table 14.5 of the textbook (reproduced below).
2. The QLF test can detect a single discrete break, multiple discrete breaks, or a slow evolution of the regression function.
3. When there is a single discrete break in the population regression function that lies a date within the range tested, the time period $\widehat{\tau}$ corresponding to the QLR test statistic is a consistent estimator of the break date.
4. Like the Chow test F-stat, the QLR statistic can be used to test for a break in all or just some of the regression coefficients.

TABLE 14.5 Critical Values of the QLR Statistic with 15% Trimming

Number of Restrictions (q)	10%	5%	1%
1	7.12	8.68	12.16
2	5.00	5.86	7.78
3	4.09	4.71	6.02
4	3.59	4.09	5.12
5	3.26	3.66	4.53
6	3.02	3.37	4.12
7	2.84	3.15	3.82
8	2.69	2.98	3.57
9	2.58	2.84	3.38
10	2.48	2.71	3.23
11	2.40	2.62	3.09
12	2.33	2.54	2.97
13	2.27	2.46	2.87
14	2.21	2.40	2.78
15	2.16	2.34	2.71
16	2.12	2.29	2.64
17	2.08	2.25	2.58
18	2.05	2.20	2.53
19	2.01	2.17	2.48
20	1.99	2.13	2.43

Note: These critical values apply when $\tau_0 = 0.15T$ and $\tau_1 = 0.85T$ (rounded to the nearest integer), so the F-statistic is computed for all potential break dates in the central 70% of the sample. The number of restrictions q is the number of restrictions tested by each individual F-statistic. Critical values for other trimming percentages are given in Andrews (2003).

4 Pseudo Out-of-Sample Forecasts

Pseudo out-of-sample forecasts are used to simulate the out-of-sample performance (i.e., the real time forecast performance) of a time series regression model. The idea is simple: pick a date near the end of the end of the sample, estimate your forecasting model using data up to that date, and then use that estimated model to make a (pseudo) forecast.

How exactly do we implement pseudo out-of-sample forecasts?

Step 1 Divide the sample data into $s = T - P$ and P subsequent observations. The P observations are used as pseudo-out-of-sample observations.

Step 2 Estimate the model using the first s observations.

Step 3 Compute the pseudo-forecast $\widetilde{Y}_{s+1|s}$.

Step 4 Compute the pseudo-forecast-error $\widetilde{u}_{s+1} = Y_{s+1} - \widetilde{Y}_{s+1|s}$.

Step 5 Repeat steps 2 trough 4 for the remaining dates, $s = T - P + 1$ to $T - 1$ (re-estimate the regression at each date).

Why do we care about pseudo out-of-sample forecasts? There are several reasons. First, as mentioned above, it gives us a sense of the performance of our model. Second, we can use the pseudo out-of-sample forecasts allow to estimate the RMSFE of the model. In particular, the sample standard deviation of the forecast errors in an estimator of the RMSFE, that incorporates the uncertainty coming from both the error term and the estimation of the coefficients. Third, we can use pseudo out-of-sample forecasts to compare two ore more candidate forecasting models, to check their potential for providing reliable forecasts.