Empirical-Finance: Unit-root Revision[1]

Imperial College London
Hormoz Ramian[2]

Background — Economic variables evolve over time. The collection of historical path, present realization and future heading, is the time-series of the economic variable. Values along the history are realized and known to us, in contrast, future values are yet to be realized and are random variables as of today's viewpoint. The time evolution of an economic variable such as income, health status, technological growth, etc. may depend on many external factors as well as its own history. Thus, time-series econometrics studies a wider spectrum that includes cross variable dependencies as well as auto-dependencies, that is the relationship between a variable and its own history. For example, good health status of a typical individual is likely to maintain itself over time, and regardless of other external factors, a good health history is a relevant predictor of good health status in near future. Time-series econometrics provides a foundation to incorporate historical realizations of an economic variable and draw inference about its near future. In particular, such foundations add stricter conditions to classical econometrics to ensure that estimation and inference methods remain valid when working with time-series data. This handout introduces basic concepts, such stationarity and unit-root test that are the building blocks of time series.

Stationarity — Economic time series may have complex features. This translates into a statistical distribution with multiple moments, many of which that may or may not be time-invariant. This complexity motivates the use of weak stationarity as a more simplistic but practical approach that only focuses on the first two moments (covariance stationary) of a time series:

Definition: A time series $\{y_t\}_{t=0}^{\infty}$ is covariance stationary, if:

(i) The mean $\mathbb{E}(y_t) = \mu$ is finite and constant across all time periods t.

(ii) The variance $\mathrm{var}(y_t) = \gamma_0$ is finite and constant across all time periods t.

(iii) Autocovariances $\mathrm{cov}(y_t, y_{t-j}) = \gamma_j$ where $j = \pm 1, \pm 2, \ldots$ describing lags behind or ahead, that are constant for all time periods t.

The first and second properties require the mean and variance of a time series to be constant over time, while the third property requires the covariance between different lags, or autocovariance of the variable with its own lagged values, to be time-invariant. In particular, autocovariance may be different depending on the number of lags apart but still time-invariant. For example, the autocovaraince between a variable observed on Monday and Tuesday this year should be exactly the same as autocovariance between Monday and Tuesday a year from now and so on. But the auto-

[1]This supplementary handout is intended to provide a revision of statistical methods behind unit-root test used in today's class. To cover these methodologies in more details, review any introductory econometrics textbook e.g. C. Kleiber. Applied Econometrics with R, 2008. While students are not examined based on statistical derivations in these notes (optional reading) in this module, basic knowledge of stationarity and its implications to macroeconomics is important.

[2]Email: `h.ramian@imperial.ac.uk`

covariance between Monday and Tuesday (one period apart) this year is not necessarily the same of autocovariance between Monday and Wednesday this year (two periods apart). Autocovariance between Monday and Wednesday, however, must be equal to autocovariance between Wednesday and Friday, since it reflects the same two-day lags distance. Stationarity is a fundamental concept in time series analysis because it implies that the underlying population mean, variance and autocovariances can be estimated from data using the analogous sample measures. Such sample measures may be essentially useless when carrying out inference methods if the underlying process is nonstationary (spurious regression)

The autocovariance at lag j is the covariance between observations y_t and y_{t-j}. Applying the usual definition of a covariance, the autocovariance at lag j gives,

$$\gamma_j = \mathbb{E}[y_t - \mathbb{E}(y_t)][y_{t-j} - \mathbb{E}(y_{t-j})], \tag{1}$$

for $j = \pm 1, \pm 2, \ldots$. For a process with mean that is constant over time, this becomes,

$$\gamma_j = \mathbb{E}[y_t - \mu][y_{t-j} - \mu], \tag{2}$$

where $\mu = \mathbb{E}(y_t)$. The variance is a special case of the autocovariance with $j = 0$. Note that (iii) above in the definition of stationarity requires γ_j to be constant over time. However, this does not imply that the autocovariances γ_1, γ_2, etc are equal to each other. In general, γ_j varies with the lag j; that is, the covariance between y_t and y_{t-j} depends on the time difference between them, namely j. However, for a stationary process, the autocovariance between y_t and y_{t-j} is identical to that between y_s and y_{s-j} for $s \neq j$. That is, the autocovariance γ_j depends only on j (the time distance between the observations) and does not vary with t.

One specific implication of stationarity is that:

$$\gamma_j = \gamma_{-j} \tag{3}$$

for $j = \pm 1, \pm 2, \ldots$. This follows because:

$$\begin{aligned} \gamma_j &= \mathbb{E}[y_t - \mu][y_{t-j} - \mu] && (4) \\ &= \mathbb{E}[y_{t-j} - \mu][y_t - \mu] && (5) \\ &= \gamma_{-j} && (6) \end{aligned}$$

Correlations are easier to interpret than covariances, and the autocorrelation at lag j is the (theoretical or population) correlation between y_t and y_{t-j}. Again assuming (second-order) stationarity, and hence that $\mathrm{var}(y_t) = \mathrm{var}(y_{t-j})$, this autocorrelation is given by,

$$\rho_j = \frac{\mathrm{cov}(y_t, y_{t-j})}{\mathrm{var}\, y_t} = \frac{\gamma_j}{\gamma_0} \tag{7}$$

Note that by definition $\rho_0 = 1$, since:

$$\rho_0 \;=\; \frac{\gamma_0}{\gamma_0} \;=\; 1 \tag{8}$$

Clearly, since a stationary process has variance and autocovariances that are constant over time, then the autocorrelations ρ_j also do not vary over time. However, they generally vary with the lag j. Also, from (2) and (3), $\rho_j = \rho_{-j}$.

Test Characterization and Intuition — We wish to turn to the issue of testing whether a process is stationary or nonstationary. If a process $\{y_t\}_{t=0}^{+\infty}$ is stationary, we use the notation $y_t \sim I(0)$ where this specifies that the process is stationary after zeroth order of differentiation (alternatively known as integrated of order zero). In time series econometrics, the notation $y_t \sim I(0)$ implies that $\{y_t\}_{t=0}^{+\infty}$ is stationary on level or as it stands without any transformation. Conversely, if a process $\{y_t\}_{t=0}^{+\infty}$ is non-stationary, or more accurately, if a process is non-stationary on level, then there are a few steps that we need to pursue to correctly characterize it:

- In the simplest case, note that nonstationarity may be induced only by the mean of the process (deterministic time-varying mean μ_t), whereas other moments are actually time-invariant. In this case, we can conjecture that the process $\{y_t\}_{t=0}^{+\infty}$ may become stationary around a time trend, which requires us to focus on the de-trended process $\{y_t - \beta t\}_{t=0}^{+\infty}$ where t is called a deterministic trend. A typical economic example of $\{y_t - \beta t\}_{t=0}^{+\infty}$ is a variable that can be predicted around a time trend e.g. population in the next few years is expected to grow at its regular growth rate per year. The deterministic time trend need not be a linear function of time and can be any function $f(t)$, for instance, t^2 which would perhaps be a better time trend for population since its growth rate has been increasing.

- Second, and after examining the de-trended data and establishing that the de-trended process is still nonstationary, we may conjecture that a nonstationary process $\{y_t\}_{t=0}^{+\infty}$ is likely to become stationary after differencing once to form $\{\Delta y_t\}_{t=0}^{+\infty}$. Effectively, in this case we no longer focus on the original variable and turn our focus to its increments per period. For example, stock prices $\{y_t\}_{t=0}^{+\infty}$ wander around and at the very least, have time-varying means, but price differences $\{\Delta y_t\}_{t=0}^{+\infty}$ or roughly speaking stock returns evolve around zero. This suggests that differencing a process once, may yield a stationary process which an econometrician can use for estimation and inference. It is very important to remember that differencing changes the underlying variable and of course, is not a desirable way of proceeding with the data, but it might be the only way. The advantage of using a stationary first-differenced process is that it still carries some information about the original process on level and to some extents is useful. However, one of the drawbacks is that it changes the interpretation of the model parameters.

- Third, and similar to the previous item, we can conjecture that if the first differenced process is still nonstationary, its second difference may become stationary. For example, observed inflation data over time suggests that the series has been increasing at an exponential rate over the past decades and hence differencing it once still yields a process that has a time-varying mean[3]. In this case, we turn our focus to $\{\Delta^2 y_t\}_{t=0}^{+\infty}$ and examine if it satisfies stationarity conditions.

We can now use the previous points to characterize a nonstationary process (note that $y_t \nsim I(0)$). A process $\{y_t\}_{t=0}^{+\infty}$ is nonstationary if it violates the three assumptions (i)-(iii), then it has to be examined for trend-stationarity or the hypothesis "$y_t - \beta t \overset{?}{\sim} I(0)$" or more generally "$y_t - f(t) \overset{?}{\sim} I(0)$". We need to set up a statistical test to examine this hypothesis which considers not only the mean but other features of the process. The conclusion of this hypothesis testing reveals that either the process is trend-stationary and therefore $y_t - f(t) \sim I(0)$, otherwise, we conclude that $y_t - f(t) \nsim I(0)$.

Next, we hypothesize "$\Delta y_t \overset{?}{\sim} I(0)$". In this case, the test conclusion either indicates that $\Delta y_t \sim I(0)$, meaning that process becomes stationary after differencing once, or alternatively, when using the original process on level $y_t \sim I(1)$. This is an important notation since it specifies that the first-differenced transformation is stationary or that the level is non-stationary but becomes stationary after differenting once. However, the test conclusion may reveal that $\Delta y_t \nsim I(0)$ which motivates the use of second-differenced transformation and the hypothesis $\Delta^2 y_t \overset{?}{\sim} I(0)$. Similarly, if we conclude that $\{\Delta^2 y_t\}_{t=0}^{+\infty}$ satisfies stationarity conditions, then $\Delta^2 y_t \sim I(0)$, or alternatively $\Delta y_t \sim I(1)$ and also that $y_t \sim I(2)$.

Hypothesis Testing — We now turn to the issue of testing whether a process is $I(1)$ or $I(0)$. The null and alternate hypotheses concerning a unit root can be either:

$$H_0 \; : \; y_t \sim I(1) \tag{9}$$
$$H_A \; : \; y_t \sim I(0) \tag{10}$$

or

$$H_0 \; : \; y_t \sim I(0) \tag{11}$$
$$H_A \; : \; y_t \sim I(1) \tag{12}$$

The first of these is more common, because it is straightforward to express the $I(1)$ null hypothesis as an equality on a single coefficient. For example, consider the AR(1) process:

$$y_t = \delta + \phi y_{t-1} + \epsilon_t \tag{13}$$

The unit root case is $y_t \sim I(1)$, this applies if and only if $\phi = 1$, while $y_t \sim I(0)$ implies the inequality $\phi < 1$. Thus equivalent to (10) case, the null and alternative hypotheses for the AR(1)

[3]In fact, we should examine the variances and autocovariances as well, but it is easier to run the first check on the mean or just the time plot of series to visually inspect if a process has an apparent tendency to violate the time-invariant mean assumption.

case can be specified as:

$$H_0 \quad : \quad \phi = 1 \tag{14}$$

$$H_A \quad : \quad \phi < 1 \tag{15}$$

As usual with any hypothesis test, we want to specify the null hypothesis in terms of an equality on a parameter to enable the distribution of the test statistic to be examined under the null hypothesis. Also note that the representation of the alternative hypothesis in (12) makes it clear that this is a one-sided test. Although our discussion focuses on testing the unit root null hypothesis of (10), tests of the stationarity null hypothesis (12) exist. These latter tests are, however, much less widely applied than the unit root tests we consider.

Simple Dickey-Fuller Regression — Consider again the AR(1) case of (13). By subtracting y_{t-1} from each side, this can be equivalently written as:

$$\Delta y_t \quad = \quad \delta + \alpha y_{t-1} + \epsilon_t \tag{16}$$

where $\alpha = \phi - 1$ and the null and alternative hypotheses for ϕ as specified in (15) are equivalent to:

$$H_0 \quad : \quad \alpha = 0 \tag{17}$$

$$H_A \quad : \quad \alpha < 0 \tag{18}$$

Since the test is one-sided, H_0 will be rejected only if $\widehat{\alpha}$ takes a sufficiently extreme negative value: a positive $\widehat{\alpha}$ cannot constitute evidence against H_0 in the direction of H_A. Put differently, $\alpha = \phi - 1$ implies that $\widehat{\alpha} > 0$ is equivalent to $\widehat{\phi} > 1$, which is not compatible with the conclusion that the process is stationary.

Dickey and Fuller were the first authors to develop the statistical theory associated with the test of (17) against (18), and hence (16) is often referred to as a Dickey-Fuller regression. At this point, note that there is an important practical advantage of using the test regression in the form of (16) rather than (13). That is, the test can be performed using the "t-ratio" $t(\widehat{\alpha})$ for the null hypothesis of a zero coefficient, as computed by virtually all regression programs for every coefficient in a regression model. Were the test to be performed using (18), an additional calculation would have to be performed in order to calculate the "t-ratio" for a test of $\phi = 1$ as $(\widehat{\phi} - 1)/s.e.(\widehat{\phi})$. Nevertheless, the advantage is purely a practical one, since the two values of the "t-ratio", namely for $\phi = 1$ in (13) and for $\alpha = 0$ in (16) will necessarily be numerically identical.