University of Illinois
Fall 2016

Department of Economics
Roger Koenker

Economics 536
Lecture 9

Introduction to Non-Stationary Time Series

Consider a univariate time series $\{y_t\}_{t=-\infty}^{\infty}$. We say that $\{y_t\}$ is (strictly) stationary if the joint distribution of the vectors $(y_{t_1}, \ldots, y_{t_k})$ and $(y_{t_1+s}, \ldots, y_{t_k+s})$ are the same for any choice of the subscripts $(t_1, t_2, \ldots, t_k, s)$. Thus, in particular, the marginal distributions are identical, so $Ey_t = \mu$ and $Vy_t = \sigma^2$ are independent of t, and furthermore covariances Cov (y_t, y_{t+s}) depend only on s, but not on t. We will say $\{y_t\}$ is weakly stationary, or covariance stationary if only, these mean and covariance conditions hold. In the Gaussian case, i.e., when $\{y_t\}$ is a Gaussian random process weak and strict stationarity are equivalent, but in general this is clearly not true.

In many economic contexts the stationarity assumptions are rather implausible. There are two common models for nonstationarity in economic time series:

(i) deterministic time trends, and cycles,

(ii) unit root processes.

We will begin by contrasting these two cases, from the point of view of forecasting. Before doing so let's introduce a simple way to represent a class of stationary processes

$$y_t = \mu + \psi(L)u_t$$

where $\{u_t\}$ is an iid sequence and $\psi(L)$ is a polynomial in the lag operator L satisfying the conditions:

(i) $\sum_{j=0}^{\infty} |\psi_j| < \infty$,

(ii) The roots of $\psi(z) = 0$ lie *outside* the unit circle.

Condition (i) is needed to assure that variances and covariances are finite. Condition (ii) is essentially an identifiability condition in the Gaussian case, while in non-linear/non-Gaussian cases the situation is rather more complicated. For a detailed discussion of the role of condition (ii) see e.g. Granger Newbold (1986).

Now consider the simplest linear trend model,

$$y_t = \alpha + \delta t + \psi(L)u_t$$

where $\psi(\cdot)$ satisfies the foregoing conditions. Sometimes such models are formulated in logs so in these cases such models may be thought of as exhibiting exponential growth.*

*You can think of this as just a natural approximation to compound interest. If you invest y_0 at r compounded n times per period, then

$$y_1 = y_0(1 + r/n)^n$$

so letting $n \to \infty$, and taking the limit corresponding to continuous compounding, we have $y_1 = y_0 e^r$ and thus $y_t = y_0 e^{rt}$. or $\log y_t = \log y_0 + rt$ (This is the usual economists' *aide memoire* for the well-known identity: $\lim_{n\to\infty}(1 + x/n)^n = e^x$.)

Now consider forecasting y at time $t+s$ given the information at time t, we may write

$$\hat{y}_{t+s|t} = \alpha + \delta(t+s) + \psi_s u_t + \psi_{s+1} u_{t-1} + \dots$$

As $s \to \infty$ we may observe that since the ψ_j are absolutely summable we must have that $\psi_s \to 0$ as $s \to \infty$ and thus as $s \to \infty$

$$E(\hat{y}_{t+s|t} - \alpha - \delta(t+s)) \to 0 \tag{T.1}$$

and

$$V(y_{t+s} - \hat{y}_{t+s|t}) \to \sigma^2(\psi_{s-1}^2 + \psi_{s-2}^2 + \dots + \psi_0^2). \tag{T.2}$$

To see this write

$$\begin{aligned} y_{t+s} - \hat{y}_{t+s|t} &= \alpha + \delta(t+s) + u_{t+s} + \psi_1 u_{t+s-1} + \dots + \psi_{s-1} u_{t+1} + \psi_s u_t + \psi_{s+1} u_{t-1} + \dots) \\ &\qquad -(\alpha + \delta(t+s) + \psi_s u_t + \psi_{s+1} u_{t-1} + \dots) \\ &= u_{t+s} + \psi_1 u_{t+s-1} + \dots + \psi_{s-1} u_{t+1}. \end{aligned}$$

Thus,

$$E(y_{t+s} - \hat{y}_{t+s|t})^2 = \sigma^2(1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{s-1}^2).$$

Note that as $s \to \infty$ this sum converges. (If $\sum |\psi_i|$ converges, then it follows that $\sum \psi_i^2$ converges. Why?)

The situation in the unit root model

$$(1-L)y_t = \delta + \psi(L)u_t$$

is quite different. Here since $\Delta y_t = (1-L)y_t$ is stationary we can use standard formula for forecasting,

$$\begin{aligned} \Delta \hat{y}_{t+s|t} &= E(y_{t+s} - y_{t+s-1} | y_t, y_{t-1}, \dots) \\ &= \delta + \psi_s u_t + \psi_{s+1} u_{t-1} + \dots \end{aligned}$$

which looks rather similar to what we had in the trend case, but now

$$\begin{aligned} \hat{y}_{t+s|t} &= \Delta y_{t+s} + \Delta y_{t+s-1} + \dots + \Delta y_{t+1} + y_t \\ &= \delta s + y_t + (\sum_{i=1}^{s} \psi_i) u_t + (\sum_{i=2}^{s+1} \psi_i) u_{t-1} + \dots \end{aligned} \tag{U.1}$$

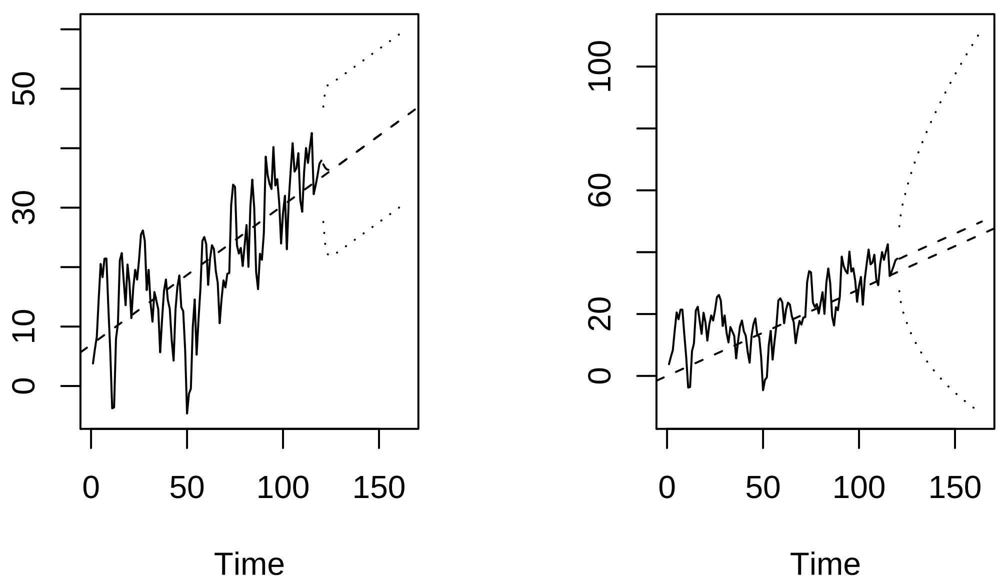
and thus

$$E(y_{t+s} - \hat{y}_{t+s|s})^2 = \sigma^2[1 + (1+\psi_1)^2 + (1+\psi_1+\psi_2)^2 + \dots (1+\psi_1+\dots+\psi_s)^2] \to \infty. \tag{U.2}$$

To summarize the foregoing discussion we can illustrate the comparison of forecasting behavior of the two models in Figure 1.

Note that the point forecast in the trend model reverts to the trend line and the confidence band converges to a constant width. In contrast the unit root model yields a forecast parallel to the trend line (U.1) in which the effect of y_t never disappears. And (U.2) shows that the corresponding confidence band grows even wider as the forecast horizon grows.

Figure 1: Comparison of forecasting behavior of trend stationary and unit root models. The left panel illustrates the trend stationary model, the right panel the unit root model. Note that the forecasts in the former model converge quickly to the estimated trend line, while in the unit root model they converge to a line parallel to the estimated trend shifted by the descrepancy from the trend line prevailing in the last period.



Feller interlude (I.III.5)

In the Stoppard play, Rosencrantz and Guildenstern are Dead, the length of the period in which Rosencrantz is"ahead" is very long. We might think that "luck" would "even out" and there should be only short periods in which either Rosencrantz and Guildenstern would be "ahead." We would be wrong. Let

$$S_n = \sum_{i=1}^{n} u_i$$

with u_i equal $+1$ or -1 with equal probability. A change of sign occurs at n if S_{n-1} and S_{n+1} are of opposite signs.

r	Probability of r sign changes in 100 trials
0	.16
1	.15
2	.14
3	.10

Theorem 1: The probability, $\xi_{r,2n+1}$, that up to epoch $2n+1$ there are exactly r changes of sign satisfies

$$\xi_{r,2n+1} = 2P(S_{2n+1} = 2r+1)$$

Proof: Elementary, but sophisticated logically.

This leads to the following normal approximation.

Theorem 2: As $n \to \infty$, the probability of fewer than $x\sqrt{n}$ sign changes before n tends to $2\Phi(2x) - 1$.

Proof: Let X_n denote the number of sign changes up to epoch n, for fixed x and large n,

$$\begin{aligned} P(X_n < x\sqrt{n}) &= \sum_{i=1}^{\lfloor x\sqrt{n} \rfloor} \xi_{i,n} \\ &\approx 2P(0 < S_n < 2x\sqrt{n}) \\ &\to 2\Phi(2x) - 1 \qquad \square \end{aligned}$$

The last limit follows from the fact that

$$\begin{aligned} &P(S_n > x\sqrt{n}) = 1 - \Phi(x) \\ &S_n = \sum u_i \qquad u_i = \begin{cases} +1 & 1/2 \\ -1 & 1/2 \end{cases} \\ &Eu_i = 0 \qquad Vu_i = E(u_i^2) = 1 \end{aligned}$$

so

$$\frac{S_n}{\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$$

This can be used to do testing since for stationary series X_n grows like n but for random walk X_n grows like $\sqrt{n}$. It is a worthwhile exercise to compare sample paths of various AR(1) models with this coin tossing random walk model to get a feeling for the difference in behavior. In R this is relatively simple: `plot(cumsum(sample(c(-1,1),1000,replace = TRUE)),type ='l')` vs. `plot(filter(rnorm(1000),.9, method = 'recursive'),type = 'l')`

It also follows from the foregoing distributional convergence that the median of the X_n is approximately $\frac{1}{2}\Phi^{-1}(3/4) \approx .337\sqrt{n}$ and the 10th percentile is roughly $Q_{X_n}(.10) \approx 0.0628\sqrt{n}$, that is, fewer than $.0628\sqrt{n}$ sign changes has probability .10. So in 1000 tosses there is a 10 percent chance that there will be 2 or less sign changes.

This seems quite surprising to most people. Even those quite sophisticated about probability may have difficulty reconciling these results with "common sense." Why should there be a knife edge in the simple AR(1) model

$$y_t = \rho y_{t-1} + u_t,$$

when $\rho = 1$ the number of sign changes grows like $\sqrt{n}$, whereas for any $\rho < 1$ this number grows like n.

Motivating Testing for unit roots

There are several motivations for the vast amount of attention lavished on the problem of testing for unit roots in the recent literature of econometrics. One of the more compelling is the work of Newbold and Granger (1974) on "spurious regression." This paper revived an observation made in Yule (1926) and focused attention on the unit root model throughout econometrics. They consider the following situation. The investigator has a simple bivariate model

$$(*) \qquad y_t = \beta_0 + \beta_1 x_t + e_t$$

but in fact,

$$\begin{aligned} y_t &= y_{t-1} + u_t \\ x_t &= x_{t-1} + v_t. \end{aligned}$$

and $\{u_t\}, \{v_t\}$ are iid. Now, one would hope that the usual theory of regression would apply and that a test of $H_0 : \beta_1 = 0$ would reveal (eventually, of course) that the model (*) was bogus. Surprisingly, this isn't the case and the usual theory doesn't apply here and if used naively can be badly misleading.

Out 100 replications the hypothesis $H_0 : \beta_1 = 0$ is rejected 77 times, at the $\alpha = .05$ level. If we extend the model to include more I(1) x's, the situation is even more disturbing as you can see from the Table below.

Spurious Regressions of I(1) Variables

Number of Regressors	Percentage of F Rejections	Mean DW-value	Mean R^2
1	76	.32	.26
2	78	.46	.34
3	93	.55	.46
4	95	.74	.55
5	96	.88	.59

Source: Granger and Newbold (1974)

There are several points which are important to make about this table. First, since the dependent variable in these models is generated as a random walk, we have, in effect, omitted y_{t-1} which should have appeared with coefficient one, and at the same time we have included extraneous variables $(x_{1t}, \ldots, x_{pt})$ which are independent of y_t. We have seen that I(1) variables behave in some respects like trended variables and thus it is not surprising that one or more of the extraneous x's behaves sufficiently similarly to the omitted y_{t-1} that we mistake their estimated coefficients as significant.

One indication of the specification problem is the highly significant Durbin Watson statistic in most realizations. Indeed, Paul Newbold's frequent comment regarding this phenomenon was, "expect nonsense when $DW \approx R^2$."

Testing for unit roots

Much of the early history of econometrics was preoccupied with testing for iid errors in time-series. Much of recent time series-econometrics has been preoccupied by the problem of testing for unit roots. One can place this in the context of Box-Jenkins theory by considering their class of ARIMA(p, d, q) processes where we write as,

$$\phi(L)(1-L)^d y_t = \theta(L)u_t$$

with iid u_t. We say such a model is "integrated of order d" since exactly d roots of the AR component lie on the unit circle and we presume that after applying $(1-L)^d$ to y_t the model is stationary.

Why is unit root testing different?

Consider the simplest random walk model

$$y_t = \rho y_{t-1} + u_t$$

where under the null we suppose

$$H_0 : \rho = 1$$

with u_t iid $\mathcal{N}(0, \sigma^2)$. We might imagine based on naive regression analogies that we could estimate the model and use the usual t-test. Why not? Consider the OLS estimator of $|\rho| < 1$,

$$\hat{\rho}_T = \frac{\sum y_{t-1} y_t}{\sum y_{t-1}^2}$$

we have from general principles,

$$\sqrt{n}(\hat{\rho}_T - \rho) \rightsquigarrow \mathcal{N}(0, \sigma^2 (X'X/n)^{-1})$$

what is $(X'X/n)^{-1}$ here?

$$X'X = \sum y_{t-1}^2$$

so

$$n^{-1}X'X = n^{-1}\sum y_{t-1}^2 \rightsquigarrow \sigma^2(1-\rho^2)$$

Since $E(y_t-\mu)^2 = E(u_t+\rho u_{t-1}+\rho^2 u_{t-2}+\ldots)^2 = \sigma^2(1+\rho^2+\rho^4+\ldots) = \sigma^2/(1-\rho^2)$.
But already we see that we are in trouble since for $\rho = 1$ we get the conclusion

$$\sqrt{n}(\hat{\rho}_T - \rho) \rightsquigarrow \mathcal{N}(0, 1-\rho^2)$$

i.e., we see that $\hat{\rho}_T$ seems to converge to 1 in the $\rho = 1$ case *faster* than the "usual" rate $1/\sqrt{n}$. Note also the cute way that the σ^2 cancels.
What to do? To take a closer look at this phenomena consider,

$$\hat{\rho}_T - 1 = \frac{\sum y_{t-1}u_t}{\sum y_{t-1}^2}$$

Recall that $y_t = y_0 + \sum_{s=1}^t u_s$ and for convenience assume that $y_0 = 0$, then

$$y_t^2 = (y_{t-1}+u_t)^2 = y_{t-1}^2 + 2y_{t-1}u_t + u_t^2$$

so,

$$y_{t-1}u_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - u_t^2)$$

Summing over $t = 1, 2, \ldots T$ we have,

$$\sum y_{t-1}u_t = \frac{1}{2}(y_T^2 - y_0) - \frac{1}{2}\sum_{t=1}^{T} u_t^2.$$

Now recall that, using $y_0 = 0$,

$$y_T \sim \mathcal{N}(0, \sigma^2 T)$$

so

$$\frac{y_T^2}{(\sigma^2 T)} \sim \chi_1^2.$$

and

$$\sigma^{-2}T^{-1}\sum u_t^2 \to 1$$

so

$$\frac{1}{\sigma^2 T}\sum y_{t-1}u_t \rightsquigarrow \frac{1}{2}(X-1)$$

where $X \sim \chi_1^2$. Next consider $\sum y_{t-1}^2$ but $y_{t-1} \sim \mathcal{N}(0, \sigma^2(t-1))$, so $Ey_{t-1}^2 = \sigma^2(t-1)$, so

$$E\sum y_{t-1}^2 = \sigma^2\sum_{t=1}^{T}(t-1) = \sigma^2(T-1)T/2$$

thus $\sum y_{t-1}^2 = \mathcal{O}(T^2)$. This means that in order to get a stable limiting form for $\hat{\rho}_T - 1$ we must rescale by T rather than $\sqrt{T}$. We can write

$$T(\hat{\rho}_T - \rho) \sim \frac{T^{-1}\sum y_{t-1}u_t}{T^{-2}\sum y_{t-1}^2} \sim \text{ rescaled and recentered}\chi_1^2$$

Further, one can look carefully at the usual t-statistic for this case

$$t_\rho = \frac{\hat{\rho}_T - 1}{(\hat{\sigma}_T^2/\sum y_{t-1}^2)^{1/2}}$$

Two things are reasonably clear about this test statistic: (i) it is *not* asymptotically Normal and (ii) It *does* converge in Law. This is the leading example of what is usually referred to as the Dickey Fuller distribution.

Some generalization to the case where our original model has a.) an intercept b.) a time trend, are needed and result in alterations of the critical values as indicated in the distributed tables. Note that even for the relatively simple case of the pure random walk the critical values are considerably larger than the ones we are used to from the t-table.

What to do if we have more complicated error process? For example, suppose $u_t \sim ARMA(1,1)$ so

$$(1 - \phi_1 L)u_t = (1 - \theta_1 L)\varepsilon_t$$

with $\varepsilon \sim iid$. Then

$$\varepsilon_t = \sum_{j=0}^{\infty} \theta_1^j (u_{t-j} - \phi_1 u_{t-j-1})$$

so

$$\begin{aligned}
\Delta y_t &= (\rho - 1)y_{t-1} + u_t \\
&= (\rho - 1)y_{t-1} + \phi_1 u_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\
&= (\rho - 1)y_{t-1} + \phi_1 u_{t-1} + \varepsilon_t - \theta_1 \sum_{j=1}^{\infty} \theta_1^{j-1}(u_{t-j} - \phi_1 u_{t-j-1}) \\
&= (\rho - 1)y_{t-1} + (\phi_1 - \theta_1)\sum u_{t-i}\theta_1^{i-1} + \varepsilon_t \\
&= (\rho - 1)y_{t-1} + \sum_{i=1}^{q} \delta_i \Delta y_{t-i} + \varepsilon_t
\end{aligned}$$

This is called the augmented Dickey-Fuller(ADF) version of the test and rather remarkably the t-test statistic in this regression has the same asymptotic distribution as in the simple case.

Granger Causation

Lets begin by recalling some definitions from 507.

Def The random variables X, Y are stochastically independent, $X \perp\!\!\!\perp Y$, if $F_{Y|X}(y|x) = F_Y(y)$.

Def. The random variables X, Y are mean independent, $X \perp Y$, if $E(Y|X) = EY$.

The former definition is obviously much stronger than the latter, i.e.,

$$X \perp\!\!\!\perp Y \Rightarrow X \perp Y,$$

and can with some effort be shown to imply

$$X \perp\!\!\!\perp Y \Rightarrow h(X) \perp g(Y)$$

for any nice functions h, g. Note mean independent is also often termed uncorrelatedness.

We can obviously regard X as a vector of $r.v$'s in the foregoing definitions and it may be convenient to consider groups of conditioning variables which include the entire historical past. For example, let

$$\Omega_t = \{X_{t-1}, X_{t-2}, \dots, Y_{t-1}, Y_{t-2}, \dots\}$$

Granger (1969) suggested the following definition of causal ordering among time series.

Def. We will say that Y_t does not Granger cause X_t iff

$$E(X_t|\Omega_t) = E(X_t|X_{t-1}, X_{t-2}, \dots)$$

In other words, Y_t does not help to predict the mean of X_t. For some purposes, although this is rarely done, one might want to strengthen this mean independence notion of Granger causality to require

$$F_{X_t|\Omega_t} = F_{X_t|X_{t-1},X_{t-2},\dots}$$

Note that since Granger causation is purely a definition based on first moments of the series; we may return to this idea briefly when we encounter quantile regression.

An interesting application of Granger causation is the note by Thurman and Fisher (1988), who show that – at least in the U.S. – eggs Granger cause chickens, but chickens do not Granger cause eggs, thus, resolving a long standing open problem in domestic agriculture. See Harvey for a more serious elaboration of the issues here.

References

Feller, W. (1968) Introduction to Probability Theory and its Applications, Wiley, 3rd ed.

Granger, C. and P. Newbold (1974) Spurious Regression in Econometrics, *J. of Econometrics*, 2, 111-120.

Granger, C. (1969) Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, 37, 424-438.

Granger, C.W.J. and P Newbold, (1986) *Forecasting Eonomic Time Series*, Academic Press.

Thurman, W.N. and M.E. Fisher (1988) Chickens, Eggs, and Causality, or which came first? *Am J. Ag. Econ*, 237-8.

Yule, G.U. (1926) Why do we sometimes get nonsense correlations in time series? *JRSS(B)*, 89, 1-64.