Time Series Analysis

Models with Trends

Jude C. Hays

February 16, 2021

Outline

Deterministic and Stochastic Trends

- Time series processes can be decomposed into three parts: the trend, the stationary component, and noise.
- The trend component accounts for changes in the nature of the time series over time.
- Time series processes with trends are non-stationary. The mean, variance, or both are a function of time.
- We need to properly account for trends in dynamic processes in order to test hypotheses with time series data.

Deterministic Trends

- A deterministic trend is one where realizations of the time series process are a fixed function of time, such as a high-order polynomial

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$$

- Clearly, in this case, $E(y_t)$ depends on t.
- If we add a stationary component to the trend, for example,

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + A(L)\varepsilon_t$$

  The process is said to be **trend-stationary**. Long-run forecasts will converge to the trend.
- In the simplest case, we have $y_t = \beta_0 + \beta_1 t$, a linear trend, which can be expressed as

$$\Delta y_t = \beta_0, \text{ or } \Delta y_t = \beta_0 + \varepsilon_t \text{ (with noise)}$$

Stochastic Trends

- A stochastic trend is one where realizations of a random process have permanent effects on the nature of a time series.
- In the simplest case, we have a random walk process

$$y_t = y_{t-1} + \varepsilon_t \text{ or } \Delta y_t = \varepsilon_t$$

- The solution to this first-order difference equation is

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

- Hence, the effects of stochastic shocks do not decay over time.
- This means that the variance of a random walk is time dependent

$$\mathrm{var}(y_t) = \mathrm{var}(\varepsilon_t + \varepsilon_{t-1} + \ldots + \varepsilon_1) = t\sigma^2$$

Mixed Trends

- A random walk with drift

$$y_t = a_0 + y_{t-1} + \varepsilon_t \text{ or } \Delta y_t = a_0 + \varepsilon_t$$

  has a trend that is partially deterministic and partially stochastic.
- The solution to this model is $y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i$
- Generalizations include the *trend plus noise* and *trend plus irregular* models

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i + \eta_t$$

and

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i + A(L)\eta_t$$

respectively, where η_t is white noise.

Detrending and Differencing

- In order to do time series analysis we need to distinguish the trend and stationary components, and the appropriate method depends on whether the trend is deterministic or stochastic.
- If a trend is stochastic, we difference the data to isolate the stationary component. The process is **difference-stationary**.
- In the case of a random walk with drift, we have

$$E(\Delta y_t) = E(a_0 + \varepsilon_t) = a_0$$
$$\mathrm{var}(\Delta y_t) = E(\Delta y_t - a_0)^2 = E(\varepsilon_t)^2 = \sigma^2$$
$$\mathrm{cov}(\Delta y_t, \Delta y_{t-s}) = E(\varepsilon_t, \varepsilon_{t-s}) = 0$$

- If the trend is deterministic, to isolate the stationary component, we detrend the data by regressing $\{y_t\}$ on a high-order polynomial function of time.
- The order of the polynomial can be determined by t-tests and F-tests as well AIC and SBC measures of fit.

Spurious Regressions

- It should be clear that two times series with deterministic trends will correlate spuriously.
- If the true data generating process is

$$y_t = f_y(t) + \varepsilon_{yt}$$
$$z_t = f_z(t) + \varepsilon_{zt}$$

and we estimate the regression

$$y_t = \beta z_t + e_{yt},$$

then e_{yt} will contain $f_y(t)$, which will correlate with z_t through $f_z(t)$. The estimate of β will suffer from omitted variable bias. Moreover, because the $\{e_{yt}\}$ are not independent, our standard error estimates will be biased as well.

Spurious Regressions

- What if two times series have stochastic trends? Will they also correlate spuriously?
- The answer in this case is not as obvious. However, Granger and Newbold (1974) showed that if the true data generating process is

$$y_t = y_{t-1} + \varepsilon_{yt}$$
$$z_t = z_{t-1} + \varepsilon_{zt}$$

and we estimate the regression

$$y_t = \beta_0 + \beta_1 z_t + e_{yt}$$

we will reject the null hypothesis more often than suggested by our p-values. In their Monte Carlo simulations, *t*-tests rejected the null hypothesis $\beta = 0$ about **seventy-five percent** of the time.

Dickey-Fuller Tests

- Detecting purely deterministic trends is relatively easy. F and t-tests will work.
- Detecting purely stochastic and mixed trends is more complicated.
- We cannot rely on ACF plots. In finite samples, the ACF of an integrated process will look like a stationary near-integrated process.
- Dickey and Fuller developed their tests around three equations:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$
$$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$$
$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$$

Dickey-Fuller Tests

- Dickey and Fuller developed their tests around three equations:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$
$$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$$
$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$$

- Under the null hypothesis $\gamma = 0$, the first equation is a pure random walk. The second equation adds a drift term, and the third equation adds a drift and linear time trend.
- Note that under the null hypothesis. The process is non-stationary, the variance of y_t is a function of time and becomes infinitely large as t increases.
- Thus, standard t and F-tests, which assume constant and finite variance, are not appropriate.

Dickey-Fuller Tests

- A Dickey-Fuller test is conducted by estimating one of the three regression, computing the relevant t and F-statistics and comparing these against the empirical critical values identified via simulation under the null hypothesis.

Table 4.2 Summary of the Dickey-Fuller Tests

Model	Hypothesis	Test Statistic	Critical Values for 95% and 99% Confidence Intervals
$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$	$\gamma = 0$	τ_τ	−3.45 and −4.04
	$\gamma = a_2 = 0$	ϕ_3	6.49 and 8.73
	$a_0 = \gamma = a_2 = 0$	ϕ_2	4.88 and 6.50
$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$	$\gamma = 0$	τ_μ	−2.89 and −3.51
	$a_0 = \gamma = 0$	ϕ_1	4.71 and 6.70
$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$	$\gamma = 0$	τ	−1.95 and −2.60

Note: Critical values are for a sample size of 100.

Extensions: Augmented Dickey-Fuller Test

- What if the time series process is a higher order autoregressive process?

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t$$

- The *augmented* Dickey-Fuller test uses the regression

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t$$

where $\gamma = -\left(1 - \sum_{i=1}^{p} a_i\right)$ and $\beta_i = -\sum_{j=i}^{p} a_j$.

Extensions: Perron's Test

- What if there's a structural break in the time series?
- Perron's test allows us to distinguish a random walk with a pulse from a trend-stationary process with a structural break in the intercept.

$$H_1 : y_t = a_0 + y_{t-1} + \mu_1 D_p + \varepsilon_t$$
$$A_1 : y_t = a_0 + a_2 t + \mu_2 D_L + \varepsilon_t$$

- The test proceeds as follows:
  1. Detrend the data using the alternative model with residuals $\hat{y}_t$.
  2. Estimate the regression $\hat{y}_t = a_1 \hat{y}_{t-1} + \varepsilon_t$.
  3. Check for serial correlation and estimate $\hat{y}_t = a_1 \hat{y}_{t-1} + \sum_{i=1}^{k} \beta_i \Delta \hat{y}_{t-i} + \varepsilon_t$, if needed.
  4. Calculate the t-statistic for the null hypothesis $a_1 = 1$ and compare against Perron's table of critical values.

Beveridge and Nelson Decomposition

- If a time series has a stochastic trend plus either noise or an irregular part, it may be interesting to decompose the series into its permanent (trend) and temporary (stationary) components.
- One method for doing this is the Beveridge and Nelson decomposition, which can be used for any ARIMA (p,1,q) model.
- Start by considering an ARIMA (0,1,2) process

$$y_t = a_0 + y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}$$

- The solution for y_t is

$$y_t = a_0 t + y_0 + \sum_{i=1}^{t} e_i$$

where $e_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}$.

Beveridge and Nelson Decomposition

- The solution for y_{t+s} is $a_0(t+s) + y_0 + \sum_{i=1}^{t+s} e_i$
- Substitute for y_0 using the solution for y_t

$$\begin{aligned} y_{t+s} &= a_0(t+s) + \left[y_t - a_0 t - \sum_{i=1}^{t} e_i \right] + \sum_{i=1}^{t+s} e_i \\ &= a_0 s + y_t + \sum_{i=1}^{s} e_{t+i} \end{aligned}$$

- Rewriting in terms of ε_t, we have

$$y_{t+s} = a_0 s + y_t + \sum_{i=1}^{s} \varepsilon_{t+i} + \beta_1 \sum_{i=1}^{s} \varepsilon_{t-1+i} + \beta_2 \sum_{i=1}^{s} \varepsilon_{t-2+i}$$

- Through recursive substitution, it can be shown that

$$E_t y_{t+s} = a_0 s + y_t + (\beta_1 + \beta_2)\varepsilon_t + \beta_2 \varepsilon_{t-1}$$

Beveridge and Nelson Decomposition

- The time-t forecast of y_{t+s} is the current level of the stochastic trend plus the forecast of the deterministic trend: $E_t y_{t+s} = \mu_t + a_0 s$. Solving for μ_t gives

$$\mu_t + a_0 s = y_t + a_0 s + (\beta_1 + \beta_2)\varepsilon_t + \beta_2 \varepsilon_{t-1}$$

or

$$\mu_t = y_t + (\beta_1 + \beta_2)\varepsilon_t + \beta_2 \varepsilon_{t-1}$$

- Since the temporary component of y_t is $y_t - \mu_t - a_0 t$, we have

$$\begin{aligned} \text{Temporary} &= y_t - [y_t - a_0 t + (\beta_1 + \beta_2)\varepsilon_t + \beta_2 \varepsilon_{t-1}] \\ &= -(\beta_1 + \beta_2)\varepsilon_t - \beta_2 \varepsilon_{t-1} \end{aligned}$$

- Hence, the difference between the observed $\{y_t\}$ and the stochastic trend is perfectly negatively correlated with the temporary part of $\{y_t\}$. This is the identifying assumption that allows us to perform the decomposition.
- The decomposition proceeds in five steps.

Beveridge and Nelson Decomposition

1. Difference y_t and estimate a_0, β_1, and β_2 as the parameters of a stationary ARMA(0,2) model.
2. Use the ARMA model to make in-sample forecasts of each y_t and y_{t-1}.
3. Set the forecast errors equal to ε_t and ε_{t-1} respectively.

$$\begin{aligned}
\varepsilon_0 &= y_0 - a_0 \\
\varepsilon_1 &= y_1 - y_0 - a_0 - \beta_1\varepsilon_0 \\
\varepsilon_2 &= y_2 - y_1 - a_0 - \beta_1\varepsilon_1 - \beta_2\varepsilon_0 \\
&\vdots
\end{aligned}$$

4. Solve for the irregular component $(\beta_1 + \beta_2)\varepsilon_t + \beta_2\varepsilon_{t-1}$
5. The trend is $\mu_t = (\beta_1 + \beta_2)\,\varepsilon_t + \beta_2\varepsilon_{t-1} + y_t$ and the transitory component is $y_t - \mu_t$.

Hodrick and Prescott Decomposition

- The Hodrick-Prescott filter chooses μ_t to minimize the following sum of squares

$$\frac{1}{T}\sum_{t=1}^{T}(y_t-\mu_t)^2+\frac{\lambda}{T}\sum_{t=2}^{T-1}[(\mu_{t+1}-\mu_t)-(\mu_t-\mu_{t-1})]^2$$

- Note that when λ is zero, the solution is $\mu_t = y_t$, and when $\lambda \to \infty$, the result is a linear time trend.
- Hence, high values of λ smooth the time series.
- There is also an *unobserved components* decomposition, which we will cover when we get to state-space models.