Econometrics of Seasonality

Based on the book by GHYSELS/OSBORN: *The Econometric Analysis of Seasonal Time Series*

Robert M. Kunst
robert.kunst@univie.ac.at

University of Vienna
and
Institute for Advanced Studies Vienna

March 6, 2012

Outline

Svend Hylleberg's celebrated definition

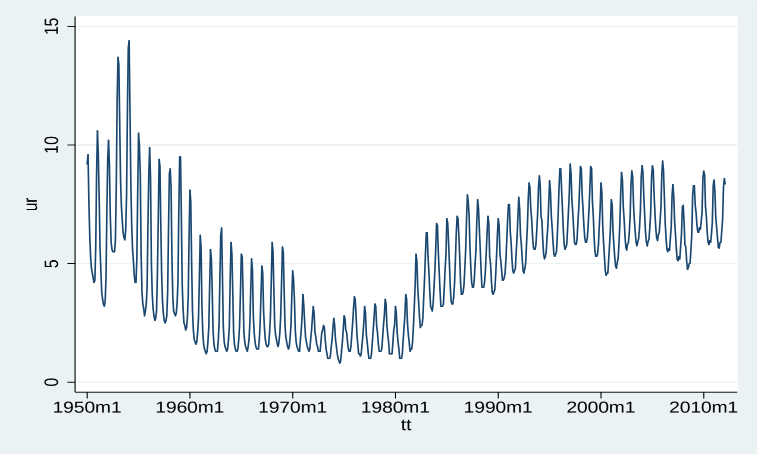
Seasonality is the systematic, although not necessarily regular, intra-year movement caused by the changes of the weather, the calendar, and timing of decisions, directly or indirectly through the production and consumption decisions made by the agents of the economy. These decisions are influenced by endowments, the expectations and preferences of the agents and the production techniques available in the economy. [from *Modelling Seasonality*, HYLLEBERG 1992]

A citation from Thomas J. Sargent

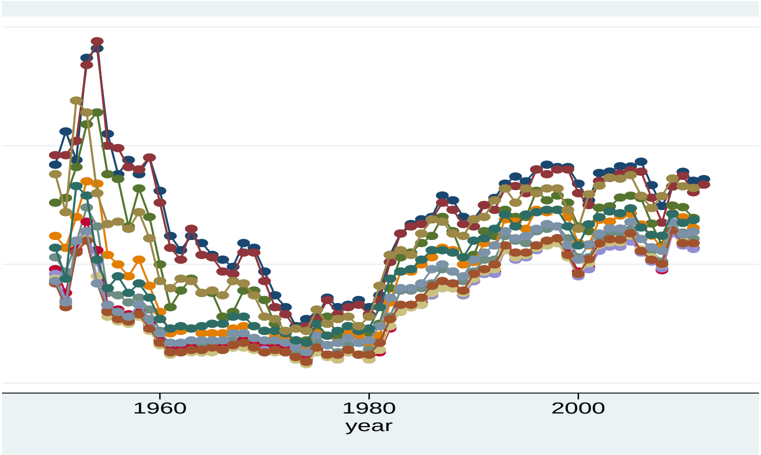
A decision to use seasonally unadjusted data can be justified by a prior suspicion that one's model is least reliable for thinking about seasonal fluctuations. [from the foreword to *The Econometric Analysis of Seasonal Data*, GHYSELS and OSBORN, 2001]

We start with some examples of seasonal series.
Usually, one observes quarters or months. The concept can be extended to weeks or days over a year or days (or business days) within a week, or hours, minutes, seconds within a day etc.

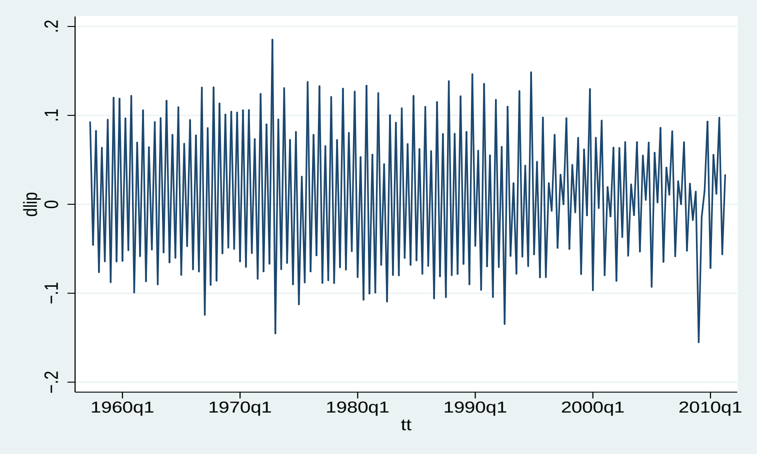
Austrian unemployment (monthly)



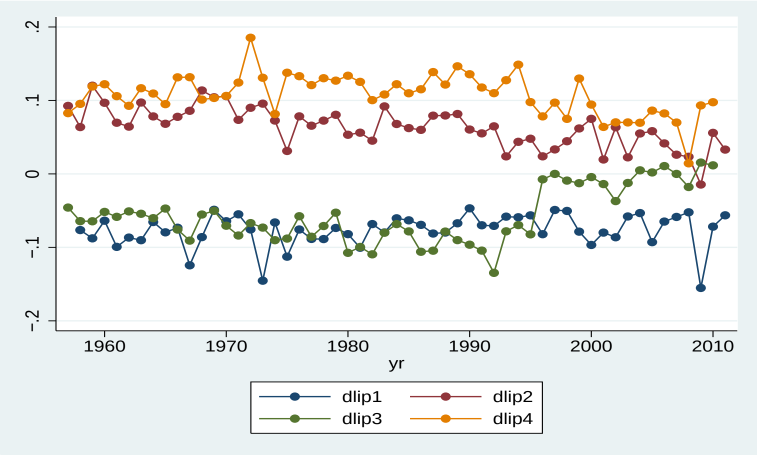
Austrian unemployment by months (Franses graph)



Growth of Austrian industrial production



Growth of Austrian industrial production (seasonal graph)



Stylized facts of many seasonal series

Means of quarters (months) may change over time, hence time-constant seasonal means may be inappropriate. Even the yearly time aggregate may show non-stationary features that are difficult to model by time-series analysis. (example: unemployment rates are notoriously difficult to model)
Modelling typically concentrates on *seasonality in the mean*. There are also models for *seasonality in variance* etc.

Seasonality in the mean: deterministic seasonality

Deterministic seasonality is defined as that part of the seasonal cycle that is known when the "process is started".

Usually, this concept is restricted to time-constant seasonal means or time-constant growth rates that differ across quarters/months.

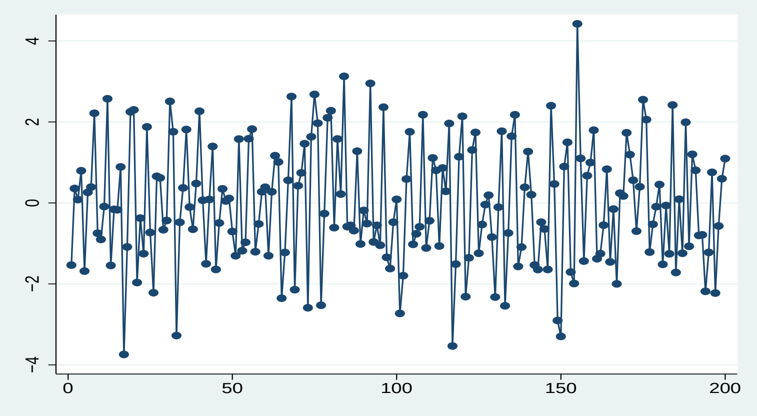
In these cases, deterministic seasonality can be expressed by means of *seasonal dummy variables* that are 1 in specific quarters and 0 otherwise.

Deterministic seasonality: simple model

$$y_t = \sum_{s=1}^{S} \delta_{st} m_s + z_t$$

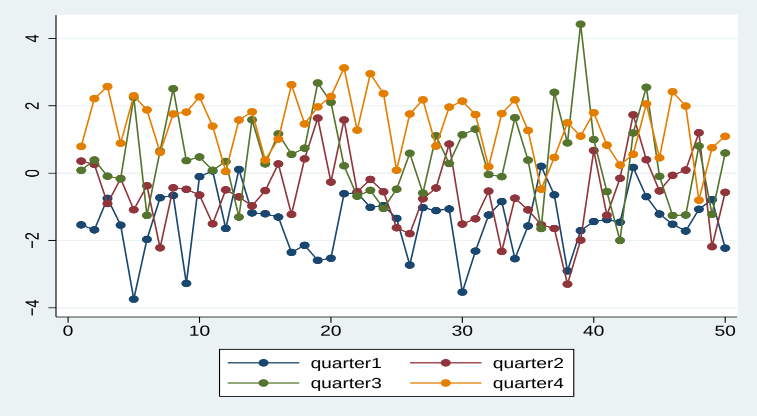
Here, $\delta = 1$ if t falls to season s, and $\delta = 0$ otherwise. m_s is the mean for season s, S is the number of seasons (usually 4 or 12), and z_t is zero-mean stationary.

Simulating the simple model with white-noise z



Quarter means are $-1.5, -0.5, 0.5, 1.5$; Gaussian white noise with variance 1.

Simulated data: graph by quarters



Even for this model with time-constant seasonal means, changes in ranking among quarters occur frequently.

Representation by trigonometric functions

$$y_t = \sum_{s=1}^{S} \delta_{st} m_s + z_t$$

This representation is often inconvenient, as we do not see the overall average mean and the components of the seasonal cycle. This suggests using instead the equivalent model

$$y_t = \alpha_0 + \sum_{k=1}^{S/2} \left\{ \alpha_k \cos\left(\frac{2\pi kt}{S}\right) + \beta_k \sin\left(\frac{2\pi kt}{S}\right) \right\} + z_t$$

Trigonometric representation for $S = 4$

For $S = 4$, the trigonometric components are quite simple. We have

$$\begin{aligned}
\cos\left(\frac{2\pi t}{4}\right) &= \cos\left(\frac{\pi t}{2}\right) = 0, -1, 0, 1, \ldots \\
\cos\left(\frac{4\pi t}{4}\right) &= \cos(\pi t) = -1, +1, -1, \ldots \\
\sin\left(\frac{2\pi t}{4}\right) &= \sin\left(\frac{\pi t}{2}\right) = 1, 0, -1, 0, \ldots \\
\sin\left(\frac{4\pi t}{4}\right) &= 0
\end{aligned}$$

α_0 is the average mean, α_1 and β_1 denote the annual wave, and α_2 gives the half-year component.

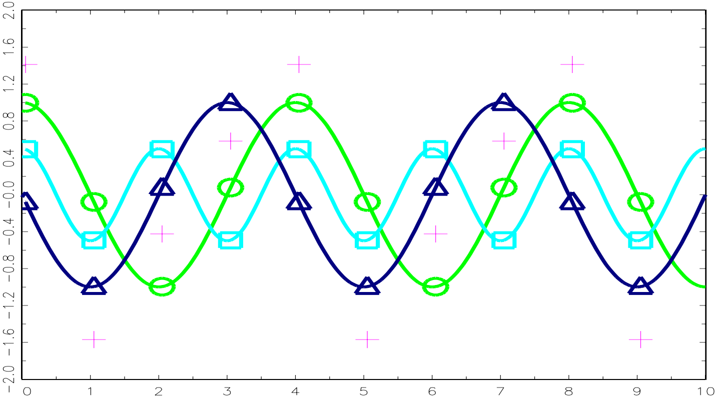
Example: trigonometric representation for the textbook example

Quarter means are $(-1.5, -0.5, 0.5, 1.5)$
$m_1 = -1.5$; $m_2 = -0.5$; $m_3 = 0.5$; $m_4 = 1.5$
$\alpha_0 = 0$; $\alpha_1 = ?$; $\alpha_2 = ?$; $\beta_1 = ?$
Apply formulas (book (1.3)-(1.5) are in error!):

$$\begin{aligned}
\alpha_1 &= \frac{1}{2}\sum_{s=1}^{4} m_s \cos\left(\frac{s\pi}{2}\right) = \frac{1}{2}(-m_2 + m_4) = 1 \\
\alpha_2 &= \frac{1}{4}\sum_{s=1}^{4} m_s \cos(s\pi) = \frac{1}{4}(-m_1 + m_2 - m_3 + m_4) = \frac{1}{2} \\
\beta_1 &= \frac{1}{2}\sum_{s=1}^{4} m_s \sin\left(\frac{s\pi}{2}\right) = \frac{1}{2}(m_1 - m_3) = -1
\end{aligned}$$

"yearly cycle at $\pi/2$ dominates"

Ghysels/Osborn example: decomposition into trigonometric components



Plus signs mark the seasonal deterministic cycle.

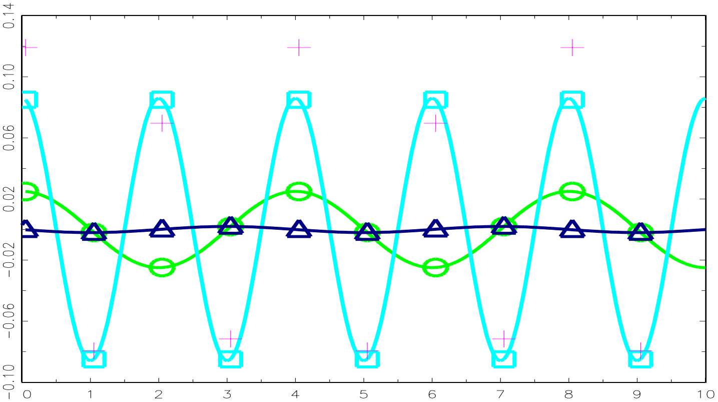
Example: trigonometric representation for growth of Austrian industrial production

Sample quarter averages are: $-0.079, 0.071, -0.075, 0.121$
$m_1 = -0.079$; $m_2 = 0.071$; $m_3 = -0.075$; $m_4 = 0.121$
$\alpha_0 = 0.0094$; $\alpha_1 =?$; $\alpha_2 =?$; $\beta_1 =?$

$$\begin{aligned}
\alpha_1 &= \frac{1}{2}\sum_{s=1}^{4} m_s \cos\left(\frac{s\pi}{2}\right) = \frac{1}{2}(-m_2 + m_4) = 0.025 \\
\alpha_2 &= \frac{1}{4}\sum_{s=1}^{4} m_s \cos(s\pi) = \frac{1}{4}(-m_1 + m_2 - m_3 + m_4) = 0.086 \\
\beta_1 &= \frac{1}{2}\sum_{s=1}^{4} m_s \sin\left(\frac{s\pi}{2}\right) = \frac{1}{2}(m_1 - m_3) = -0.002
\end{aligned}$$

Semiannual cycle dominates (π).

Industrial production growth: decomposition into trigonometric components



Plus signs mark the seasonal deterministic cycle.

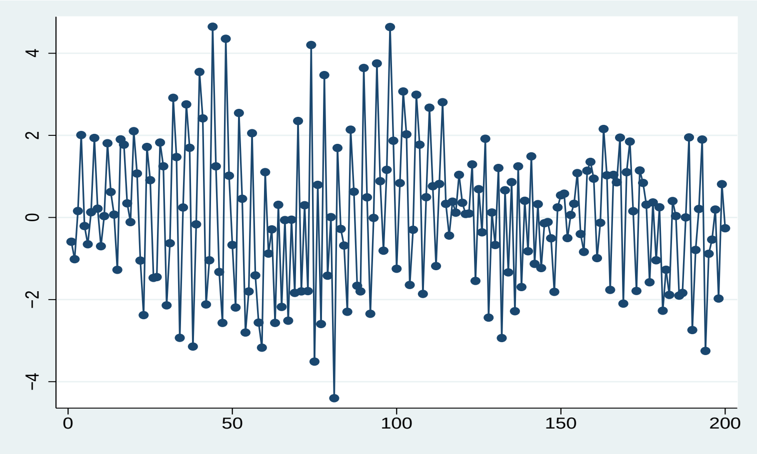
Another model class: linear stationary seasonal models

For example:

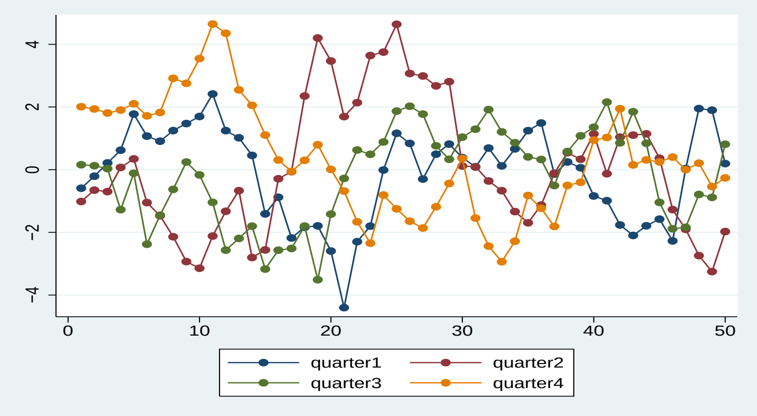
$$y_t = \phi_S y_{t-S} + \varepsilon_t, \quad |\phi_S| < 1$$

Such models generate cycles of length S. Unfortunately, these cycles are too transitory to be a useful representation for seasonal cycles.

Simulating the simple model with $S = 4$, $\phi = 0.9$, Gaussian white-noise ε, $T = 1000 + 200$



Stationary model: simulated data by season



Too many crossings, not a realistic representation

More general stationary models

For example,

$$\phi(B)\left(1-\varphi B^S\right)y_t=\theta(B)\varepsilon_t,$$

or, in the Box-Jenkins tradition

$$\phi_S(B^S)y_t=\theta_S(B^S)\varepsilon_t,$$

Such variants tend to be a bit more realistic.

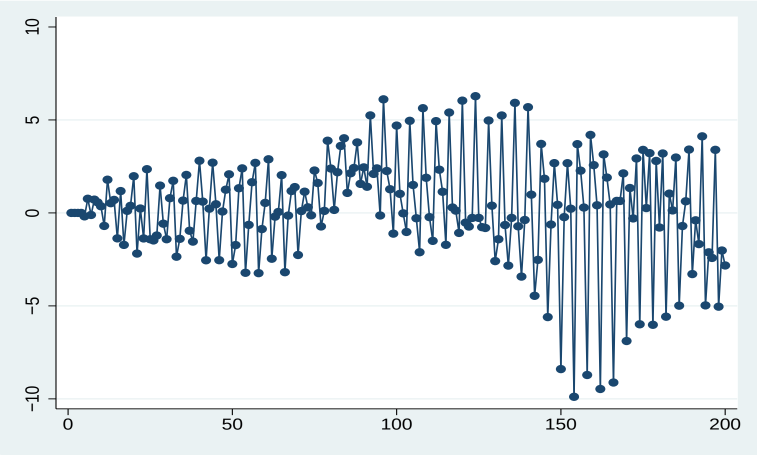
Important model class: seasonal unit-root non-stationary model

The basic process of this class is the *seasonal random walk*

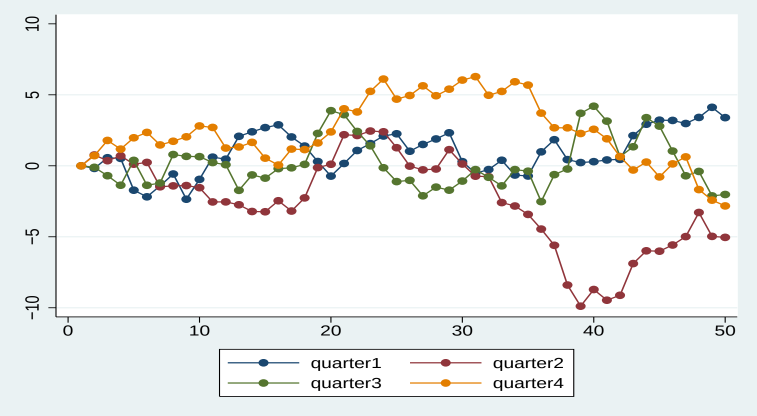
$$y_t = y_{t-S} + \varepsilon_t, \Delta_S y_t = \varepsilon_t,$$

using the notation $\Delta_S = 1 - B^S$. The seasonal random walk consists of S independent random walks that alternate. Observation t depends on observations at $t - S$, $t - 2S$, ... only. Seasonal cycle is very persistent but it also changes persistently. Variance is increasing and therefore the process is not stationary.

Simulating the seasonal random walk with $S = 4$



Simulated SRW by season



Few crossings but: some may feel uneasy at the divergence between seasons.

More general seasonal unit-root models

The model may be generalized by assuming seasonal differences to be autocorrelated. For example, consider:

$$\phi(B)\Delta_S y_t = \sum_{s=1}^{S} \delta_{st} m_s + \theta(B)\varepsilon_t,$$

assuming $\phi(B)$ and $\theta(B)$ to be well-behaved (all roots outside unit circle). These are, according to Box-Jenkins, members of the SARIMA class. y is non-stationary and seasonal cycles are persistent and also persistently changing.

Summary: important seasonal models

- I Deterministic seasonality: time-constant seasonal means: non-stationary but sub-series for seasons are stationary
- II Stochastic seasonality
  - a stationary stochastic seasonality: time-constant means (not so good models)
  - b unit-root seasonality: non-stationary, seasonal means evolve over time

Models from class II may contain elements from models I.
Stochastic models usually also contain deterministic parts.
I: summer remains summer
II: summer may become winter

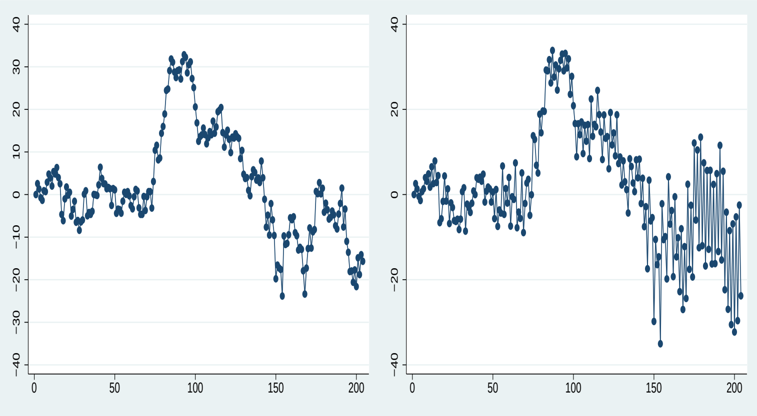
Unobserved components

Ideology: Box-Jenkins type time-series analysis looks for possible unit roots in ARMA representations. Unobserved-components models assume all unit-root components to be there and may test for their "significance". For example (ns=non-seasonal, s=seasonal part):

$$
\begin{aligned}
y_t &= y_t^{ns} + y_t^{s}, \\
\Delta y_t^{ns} &= \varepsilon_t^{ns}, \\
\left(1 + B + B^2 + \ldots + B^{s-1}\right) y_t^{s} &= \varepsilon_t^{s}
\end{aligned}
$$

Here, a crucial parameter is the variance ratio between errors.

Unobserved components: simulated data



Left graph with $\sigma^{ns}/\sigma^{s}=3$, right graph with $\sigma^{ns}/\sigma^{s}=1/3$. Both cases are really seasonal unit-root models.

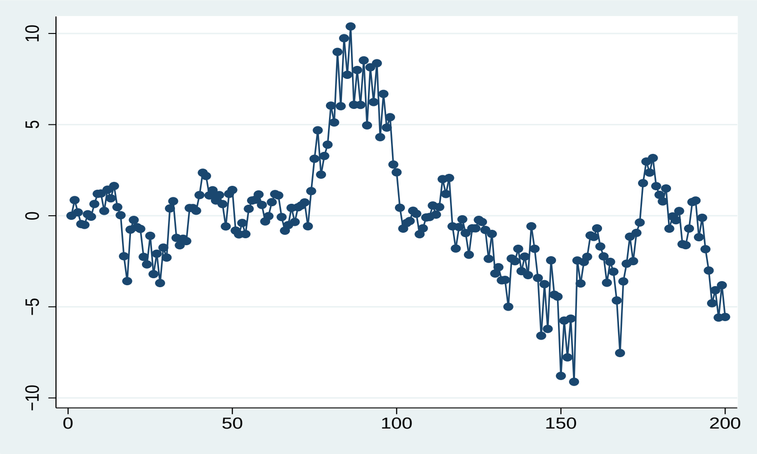
More models: periodic processes

Different coefficients for different seasons. A simple example:

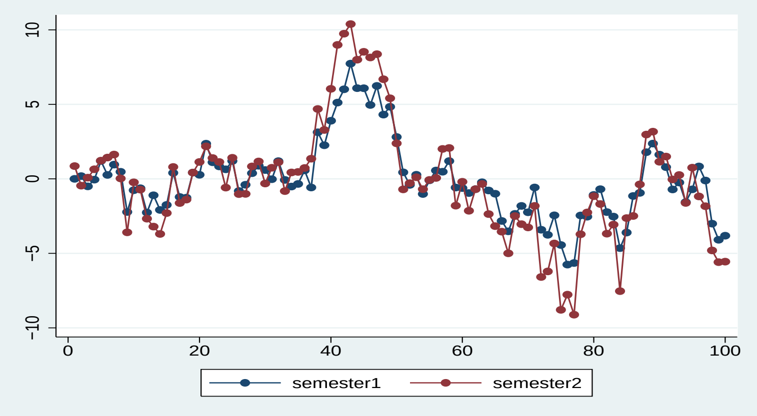
$$y_t = \left(\sum_{s=1}^{S} \delta_{st} \phi_s^P \right) y_{t-1} + \varepsilon_t$$

This is a *first-order* autoregression with periodically varying coefficients. Stability conditions are relatively complex. Here, absolute value of product of all coefficients should be less than one.

Simulating a periodic model: $S = 2$, $\phi_1^P = 0.7$, $\phi_2^P = 1.35$



Simulated periodic process by semester



Seasons do not wander apart as in the SRW but model class is complex. There are many parameters to be estimated.

Seasonality in higher moments I

Stochastic seasonal unit roots are really random-coefficients models, for example:

$$\begin{aligned} y_t &= (1-\phi_t)\, y_{t-S} + \varepsilon_t, \\ \phi_t &= \rho\phi_{t-1} + \xi_t, \\ |\rho| &< 1, \end{aligned}$$

with white-noise ε_t and ξ_t. Note that $\mathrm{E}\phi_t = 0$. This model class can be similar to ARCH models.

Seasonality in higher moments II

Seasonal GARCH models may look like

$$\begin{aligned} y_t &= \varepsilon_t, \\ \varepsilon_t &= \nu_t \sqrt{h_t}, \\ h_t &= \mu + \phi \varepsilon_{t-S}^2 + \theta h_{t-S} \end{aligned}$$

where ν_t is iid. Similarly, one may consider periodic GARCH. Note error in book: (1.20)-(1.21) yield imaginary data!

More on the deterministic seasonal model

Now, the model has an average-mean constant μ, and the sum of all m_s is assumed as 0:

$$y_t = \mu + \sum_{s=1}^{S} m_s \delta_{st} + z_t,$$

where z is a mean-zero stationary process. The model can be generalized to contain seasonal trends, but this is usually implausible (diverging seasons). Also note the sample-size convention:
There are T observations for $T/S = T_\tau$ years.

Transformation to trigonometric representation

For quarterly data (S=4), it is relatively straight forward to switch between a variant of the dummy representation (sum of γ_s is non-zero)

$$y_t = \sum_{s=1}^{4} \gamma_s \delta_{st} + z_t$$

and the trigonometric form ($\beta_2 = 0$)

$$y_t = \mu + \sum_{k=1}^{2} \left\{ \alpha_k \cos\left(\frac{\pi k t}{2}\right) + \beta_k \sin\left(\frac{\pi k t}{2}\right) \right\} + z_t .$$

Using simple algebra, one can show that

$$\Gamma = RA,$$

where $\Gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)'$ and $A = (\mu, \alpha_1, \beta_1, \alpha_2)'$ are related via

$$R = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 1 & -1 & 0 & 1 \\ 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

R is non-singular and it also has a simple inverse.

The inverse of R

$$R^{-1} = \frac{1}{4} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 0 & 2 \\ 2 & 0 & -2 & 0 \\ -1 & 1 & -1 & 1 \end{pmatrix}$$

can be used for the transformation $A = R^{-1}\Gamma$.

Re-consider the example

In that example, it was given that:
$m_1 = -1.5$; $m_2 = -0.5$; $m_3 = 0.5$; $m_4 = 1.5$
The sum is 0, hence γ_s and m_s coincide. Evaluation by insertion yields

$$R^{-1}\Gamma = A = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0.5 \end{pmatrix}$$

Note that here the semi-annual 'fast' seasonal α_2 is in the last position.

Re-consider Austrian production example

In that example, we estimated (new γ notation):
$\gamma_1 = -0.079$; $\gamma_2 = 0.071$; $\gamma_3 = -0.075$; $\gamma_4 = 0.121$
Note that the sum is not 0. We apply the transformation and get

$$R^{-1}\Gamma = A = \begin{pmatrix} 0.0094 \\ 0.0250 \\ -0.0022 \\ 0.0864 \end{pmatrix}$$

The first entry stands for the 'average mean'. The annual growth rate is around 3.8%.

Stochastic and deterministic seasonality

Idea: test procedures should be developed that have deterministic seasonality as the null hypothesis. Then, the alternative represents cases where purely deterministic seasonal models are mis-specified. This alternative must be understood in more detail.
Warning: *deterministic* seasonality is not non-stochastic. Rather, 'deterministic' is a special case of 'stochastic'. However, current usage identifies 'stochastic seasonality' with stationary and unit-root seasonality and 'deterministic seasonality' with pure dummy patterns.

More on the stationary model IIa

The simple seasonal autoregression ($S = 4$, season s in year τ)

$$z_{s\tau} = \phi_4 z_{s,\tau-1} + \varepsilon_{s\tau}, s = 1, \ldots, 4, \tau = 1, \ldots, T_\tau,$$

describes a stable (asymptotically stationary) model for $|\phi_4| < 1$. Repeated substitution yields:

$$\mathrm{var}\,(z_{s\tau}) = \phi_4^{2\tau}\mathrm{var}\,(z_{s0}) + \sigma_\varepsilon^2 \sum_{j=0}^{\tau-1} \phi_4^{2j}$$

The variance just depends on the history of a specific season. In principle, the variance of starting values may vary across seasons. As τ becomes large, the variance converges to $\sigma^2/(1-\phi_4^2)$.

The model can also be written as

$$\left(1-\phi_4 B^4\right) z_t = \varepsilon_t$$

The operator can be decomposed as (assume a positive ϕ_4)

$$1-\phi_4 B^4 = \left(1-\sqrt[4]{\phi_4}B\right)\left(1+\sqrt[4]{\phi_4}B\right)\left(1+\sqrt[2]{\phi_4}B^2\right)$$

and the inverted roots are clearly $\pm\sqrt[4]{\phi_4}, \pm\sqrt[4]{\phi_4}i$.
The roots are symmetric around 0, on the real and on the imaginary axis. As ϕ_4 approaches one, they approach the unit circle. If $S=12$, the pattern is comparable.

More on the seasonal random walk

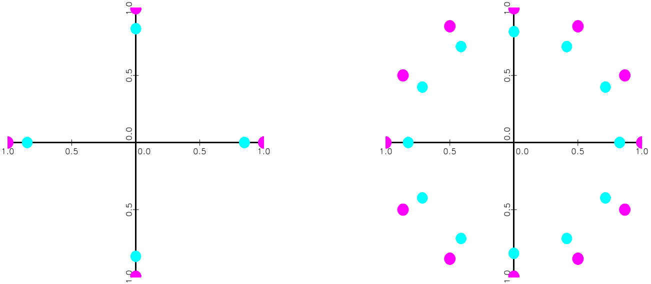
Formally, this is the first-order seasonal autoregression with $\phi_4 = 1$:

$$\begin{aligned} y_{s\tau} &= y_{s,\tau-1} + \varepsilon_{s\tau} \\ &= y_{s0} + \sum_{j=1}^{\tau} \varepsilon_{sj}, \quad s = 1, \ldots, 4, \end{aligned}$$

which yields immediately

$$\mathrm{var}\,(y_{s\tau}) = \mathrm{var}\,(y_{s0}) + \tau\sigma_{\varepsilon}^2.$$

Inverted roots for seasonal autoregressions



$S = 4$ and $S = 12$. Roots for stationary autoregression (cyan) and for seasonal RW (magenta).

A functional limit theorem

For the seasonal random walk (and for any random walk), one may use a special property called the functional central limit theorem (FCLT), here

$$T_\tau^{-1/2} y_{s,r\tau} \Rightarrow \sigma W_s(r), s = 1, \ldots, S, r \in [0,1].$$

The scaled trajectory (in its entirety) converges in distribution to Brownian motion. W_s denotes a standard Brownian motion. (Note that the form in the textbook does not make much sense, as r is missing from the left side)
This holds for any season s, we have S independent Brownian motions as limits (spring motion, summer motion etc.).

The genesis of limit theorems

Laws of large numbers (LLN) are used to establish that, under certain conditions

$$\bar{x}_T = \frac{1}{T} \sum_{t=1}^{T} x_t \to \mathrm{E}x.$$

Central limit theorems (CLT) are used to establish that, under certain conditions (note that LLN is used)

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \bar{x}_T) \Rightarrow N(0, \sigma^2).$$

Functional central limit theorems establish joint convergence for the function $b(r)$ on $[0,1]$ that is defined by T–vectors of all partial sums from a CLT

$$b(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{rT} (x_t - \bar{x}), r = \frac{1}{T}, \frac{2}{T}, \ldots$$

(and some interpolation for other r) to a stochastic process on $[0,1]$, usually a form of Brownian motion. Because of ongoing refinement, convergence works for all r in $[0,1]$. Because Brownian motion is normally distributed for any fixed r, traditional CLTs will then hold for any fixed r and particularly for $r = 1$.

Some more limits

$$T_\tau^{-3/2} \sum_{\tau=1}^{T_\tau} y_{s\tau} \Rightarrow \sigma \int_0^1 W_s(r)dr$$

$$T_\tau^{-1} \sum_{\tau=1}^{T_\tau} \varepsilon_{s\tau} y_{s,\tau-1} \Rightarrow \sigma^2 \int_0^1 W_s(r)dW_s(r)$$

$$T_\tau^{-2} \sum_{\tau=1}^{T_\tau} y_{s\tau}^2 \Rightarrow \sigma^2 \int_0^1 W_s^2(r)dr$$

These are limit properties but not *functional*. The expressions on the right are real random variables. The first one is Gaussian, the second one is a transformed chi-square, the last one is non-standard. These are derived using FCLT and the property that weighted sums converge to integrals.

What are the limit properties used for?

These limit properties are used to derive asymptotic properties of *estimators* for parameters of non-stationary seasonal processes of type IIb (unit roots).

They are also useful for deriving asymptotic properties of *test statistics*.

Here, test statistics are presented with the null of a purely deterministic seasonal model and a *unit-root alternative*. In practice, the reverse idea is more common (next section).

Ideas for seasonal stationarity tests

1. One may use an unobserved-components model that assumes unit-root components. If the variance of the seasonal component is zero, this is evidence for the deterministic null hypothesis.
2. One may take seasonal differences of the data. If the deterministic model is correct, seasonal differences have no seasonality and they have unit roots in their moving-average representation.

Both ideas have been taken up in the literature. They lead to comparable test statistics that also have comparable distributions under the deterministic-stationary null.

The Canova-Hansen test

Formally, the trigonometric representation of the deterministic model can be written as

$$y_t = \sum_{s=1}^{S} F_s' A \delta_{st} + z_t,$$

where F_s' is a row from the transformation matrix R. If $A = A_t$ is allowed to vary over time like a random walk, the variable y_t will belong to the alternative and have seasonal unit roots (but will not be a simple SRW). Canova and Hansen motivated this idea by a UC model.

Some properties of the Canova-Hansen test

The test statistic is constructed on the *Lagrange-multiplier* (LM) principle. The model is estimated under the null (deterministic), and 'improvements of fit' in the direction of seasonal unit roots are explored.

There are versions that test for specific unit roots $(1, -1, i)$ only.

The limit distribution is defined by an integral over a Brownian bridge (starts at 0 for $r = 0$ and comes down to 0 again for $r = 1$).

This law is sometimes called the *van Mises distribution*, for the standard version a $VM(S)$ law with S degrees of freedom. One has to consult special tables for the VM significance points.

The Caner test

This test is very similar to the Canova-Hansen test. The main difference is that CANER bases his derivation on a parametric model. He considers the autoregressive representation

$$\phi(B)y_t = \sum_{s=1}^{S} F_s' A \delta_{st} + \varepsilon_t,$$

where again under the alternative A_t is time-dependent and contains unit roots. CANER uses the observation that, under the null, seasonal differences $\Delta_S y_t$ have a unit root in the moving-average representation.

Some properties of the Caner test

Like the Canova-Hansen test, the Caner test is derived from the LM principle. The limiting distribution of the test statistic under the null is again a $VM(3)$ distribution, if $S = 4$ and both the root at -1 and the root pair at $\pm i$ are tested for. If also the unit root at $+1$ is under scrutiny, the limiting law will be $VM(4)$. Generally, the experience of many researchers is that tests based on parametric models are slightly more reliable than those based on non-parametric concepts. Thus, the Caner test may be better than the Canova-Hansen test.

The Tam-Reinsel test

Tam and Reinsel preferred to derive a test solely on the idea of moving-average unit roots. If, in the special ARMA model

$$\Phi(B)\Delta_S y_t = \theta^*(B)(1-\theta_S B^S)\varepsilon_t,$$

(assuming well-behaved Φ and θ^*) we find $\theta_S = 1$, then the factor Δ_S cancels from both sides and y_t has a representation without seasonal or other unit roots. Under the alternative $\theta_S \neq 1$, y_t has all S seasonal unit roots.

Some properties of the Tam-Reinsel test

Like the Canova-Hansen and Caner tests, the construction uses the LM principle. The limit distribution is also a scaled version of $VM(S)$.

The Tam-Reinsel test is less popular than the other tests, as it does not allow to investigate the occurrence of *some* of the unit roots. In many applications, the researcher does *not* want to consider the unit root at 1 together with the seasonal roots.

Seasonal integration

In modern language, one does not use words like 'seasonal ARIMA' or 'seasonal non-stationarity' but rather seasonal integration. A precise definition is:

Definition

The process y_t, observed at S equally spaced time intervals per year, is seasonally integrated of order d, in symbols $y_t \sim SI(d)$, if $\Delta_S^d y_t$ is a stationary, invertible ARMA process.

Usually, $d > 1$ is never used in applications. An SI(1) process has S unit roots, among them the root at $+1$.

Seasonal random walk with drift

The process

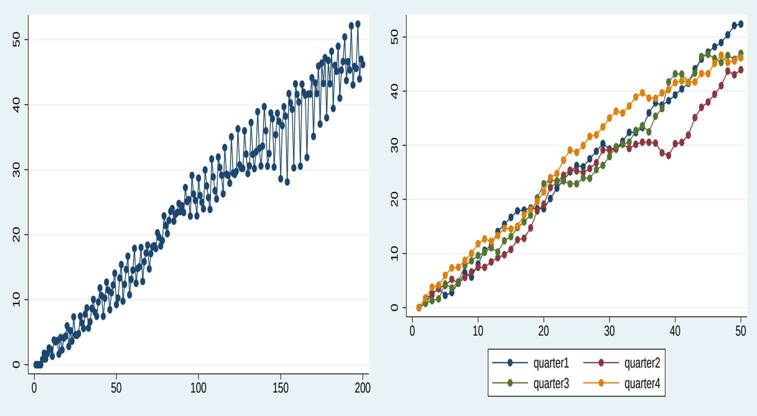
$$\Delta_S y_t = \gamma + \varepsilon_t, \quad t = 1, \ldots, T,$$

with ε_t IID$(0, \sigma^2)$, is called the *SRW with drift* γ. It is clearly first-order seasonally integrated SI(1). By continuous substitution, one gets

$$y_{s\tau} = \gamma\tau + y_{s0} + \sum_{j=1}^{\tau} \varepsilon_{sj}, s = 1, \ldots, S, \tau = 1, \ldots, T_\tau,$$

with possibly different seasonal starting values y_{s0}.

SRW with drift: simulated trajectory

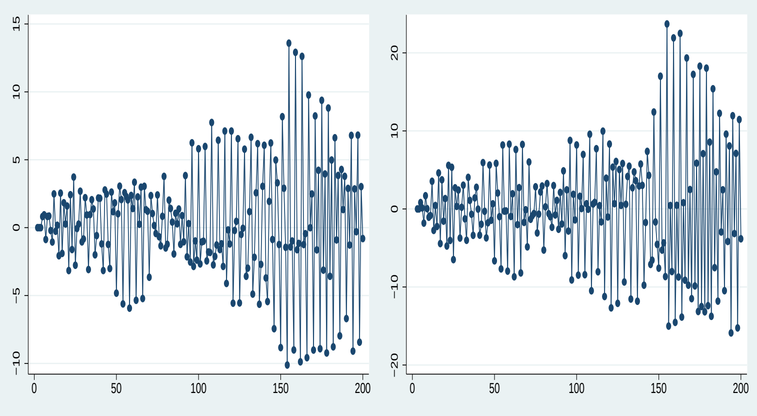


Starting patterns survive for a while and change permanently. The trend dominates for larger γ, here $\gamma = \sigma^2 = 1$.

Some properties of the SRW with drift

The SRW with drift really consists of S different and independent drifting random walks that alternate with the seasons. All S random walks have identical drift.
The deterministic component of a SRW with drift consists of the linear trend and of the seasonal pattern in the starting values.
Usual *first* differences of a seasonal random walk eliminate the trend and yield a series with strong seasonal pattern. The series still has $S-1$ seasonal unit roots.
Second-order seasonal differences or differences Δ_k with $k < S$ eliminate the trend and some though not all seasonal unit roots.

Differences of seasonal random walks



First-order differences of the SRW (left) and second-order seasonal differences (right) applied to drifting seasonal random walks.

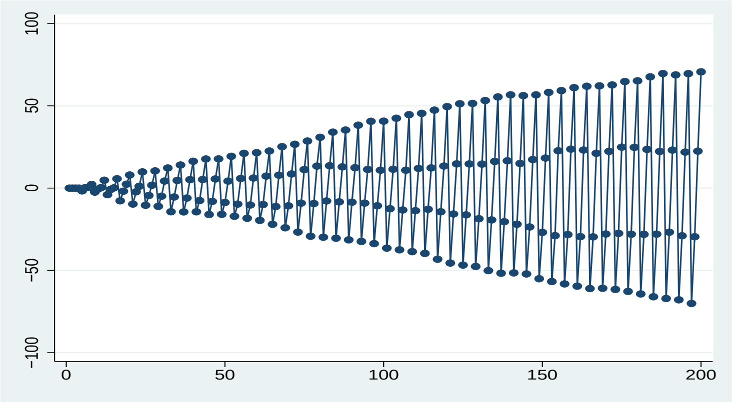
Seasonal dummies plus SRW

The model

$$\Delta_S y_{s\tau} = \gamma_s + \varepsilon_{s\tau}, \quad s = 1, \ldots, S, \quad \tau = 1, \ldots, T_\tau,$$

plays a role as an auxiliary model in tests. It is not a useful model *per se*, as it implies divergent seasonal trends. This is similar to the trend plus random-walk model in the Dickey-Fuller test: an auxiliary construction that does not describe the data in a plausible way, as it implies a quadratic trend.

SRW plus dummies: simulated trajectory



This simulation uses the textbook values $\Gamma = (-1.5, -0.5, 0.5, 1.5)'$. The variable is not $SI(1)$, as $\Delta_4 y$ is not stationary!

The general SI(1) process

According to definition, all SI(1) processes are described by the model

$$\phi(B)\Delta_S y_t = \gamma + \theta(B)\varepsilon_t,$$

with lag orders p for the AR and q for the MA polynomial. Then, y can be expressed as

$$y_{s\tau} = \gamma\tau + y_{s0} + \sum_{j=1}^{\tau} z_{sj}, s = 1, \ldots, S, \tau = 1, \ldots, T_\tau,$$

with z a stationary ARMA process. The S alternating seasons are no more independent.

Transforming seasonal variables

For many methods and derivations, it is convenient to apply transformations to the observed seasonal variables. Particularly, one may apply the transformations that were used for the seasonal dummies (assume $S = 4$):

$$\begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \\ y_t^{(4)} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-3} \\ y_{t-2} \\ y_{t-1} \\ y_t \end{pmatrix}$$

The transformation matrix is the transpose of the R that was introduced before.

The seasonal moving average $y^{(1)}$

The first transformed variable is defined as

$$y_t^{(1)} = y_t + y_{t-1} + y_{t-2} + y_{t-3},$$

4 times a seasonal average. If y is SI(1), then $y^{(1)}$ has no seasonal unit roots but it has the unit root at $+1$. It is $I(1)$. If y is a SRW, then $y^{(1)}$ is a random walk. $y^{(1)}$ can also be written as

$$y_t^{(1)} = \left(1 + B + B^2 + B^3\right) y_t = (1 + B)(1 + B^2)y_t.$$

The alternating average $y^{(2)}$

The second transformed variable is defined as

$$y_t^{(2)} = y_t - y_{t-1} + y_{t-2} - y_{t-3},$$

an alternating seasonal average. If y is $SI(1)$, then $y^{(2)}$ does not have unit roots at $+1$ nor at $\pm i$ but a root at -1. If y is a SRW, then $y^{(2)}$ is a 'random jump'. $y^{(2)}$ can also be written as

$$y_t^{(2)} = \left(1 - B + B^2 - B^3\right) y_t = (1 - B)(1 + B^2) y_t.$$

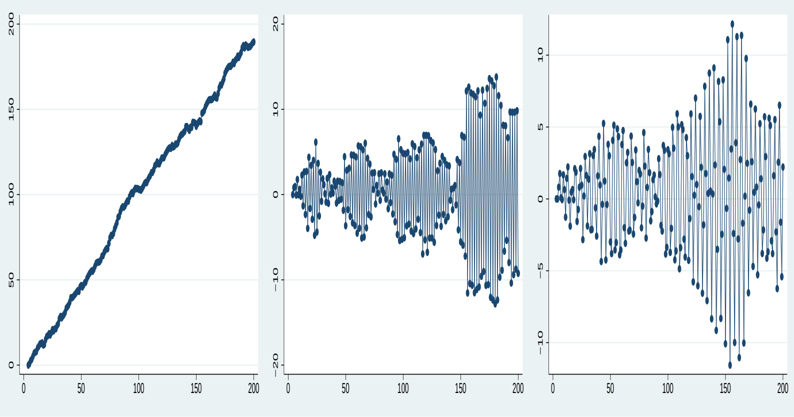
The half-year differences $y^{(3)}$ and $y^{(4)}$

The third and fourth transformed variables are defined as

$$y_t^{(3)} = y_t - y_{t-2},$$
$$y_t^{(4)} = -y_{t-1} + y_{t-3},$$

half-year differences $\Delta_2 y_t$ and $\Delta_2 y_{t-1}$. If y is $SI(1)$, these do not have unit roots at ± 1 but they have roots at $\pm i$. If y is a SRW, these are two independent, alternating random-jump processes.

Transformed variables: simulated trajectories



From left to right: $y^{(1)}$, $y^{(2)}$, and $y^{(3)}$, if original y is a SRW with drift.

Testing seasonal integration

Idea: in all *seasonal unit root tests*, the null hypothesis is the $SI(1)$ model (or some generalization), while the alternative is a model without seasonal unit roots (or without some of them), although possibly with seasonal deterministic patterns (or also seasonal stationary patterns).
The limit theorems from the *seasonal stationarity tests* are again the basis for the limit distributions of test statistics. Rather than on Brownian bridge integrals (VM distributions), they now depend on integrals over (transforms of) Brownian motion proper. These are generalizations of the Dickey-Fuller distributions.

Dickey-Hasza-Fuller test

The oldest and simplest test is based on the model

$$y_t = \phi_S y_{t-S} + \varepsilon_t,$$

where $\phi_S = 1$ under the $SI(1)$ null hypothesis. In practice, the equation will be 'augmented' with deterministic regressors (dummies or a constant only) and with lagged stationary seasonal differences $\Delta_S y$. Usually, the model is re-written as

$$\Delta_S y_t = \alpha_S y_{t-S} + \varepsilon_t,$$

and $\alpha_S = 0$ under the SRW or $SI(1)$ null.

The DHF test statistic

In order to test the null $\alpha_S = 0$ against the alternative $\alpha_S < 0$ in the model

$$\Delta_S y_t = \alpha_S y_{t-S} + \varepsilon_t,$$

one may use the t–statistic for the coefficient α_S. As in the Dickey-Fuller test, this t–statistic is not t–distributed under the null.

The distribution of the DHF statistic

Dickey-Fuller's test statistic has an asymptotic law that is characterized as

$$\int_0^1 W(r)dW(r) / \left(\int_0^1 W^2(r)dr \right)^{1/2},$$

where W is standard Brownian motion. Similarly, the DHF statistic has the asymptotic null distribution

$$\sum_{s=1}^{S} \int_0^1 W_s(r)dW_s(r) / \left(\sum_{s=1}^{S} \int_0^1 W_s^2(r)dr \right)^{1/2},$$

with S different Brownian motions.

DHF is rarely used nowadays

The problem of the DHF test is that, under the null hypothesis, one has exactly S unit roots. Under the alternative, one has no unit root. This is very restrictive, as some people may wish to test for specific seasonal or non-seasonal unit roots. The HEGY test by Hylleberg, Granger, Engle, Yoo can do this. Therefore, it is the most customary test.

The HEGY test

Hylleberg, Engle, Granger, Yoo (HEGY) suggest to build on the regression ($S = 4$)

$$\Delta_4 y_t = \pi_1 y_{t-1}^{(1)} - \pi_2 y_{t-1}^{(2)} - \pi_3 y_{t-2}^{(3)} - \pi_4 y_{t-1}^{(3)} + \varepsilon_t,$$

where $y^{(j)}$ are the transformed variables introduced above. It is important that the four regressors together form a one-one transform of $(y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4})'$, and that every AR(4) model can be re-written in this form, whether it has unit roots or not. [Other authors use other sign conventions for the coefficients]

All HEGY coefficients are zero

We see immediately that, if all p coefficients in

$$\Delta_4 y_t = \pi_1 y_{t-1}^{(1)} - \pi_2 y_{t-1}^{(2)} - \pi_3 y_{t-2}^{(3)} - \pi_4 y_{t-1}^{(3)} + \varepsilon_t,$$

are zero, then y is a seasonal random walk and is therefore SI(1). One can use the F–statistic for such a joint test. As expected, the null distribution of this F–statistic is not an F distribution but it depends on integrals over four Brownian motion terms. Such an F test is at least as informative as the DHF test.

The first coefficient π_1 is 0

If $\pi_1 = 0$, one has

$$\Delta_4 y_t = -\pi_2 y_{t-1}^{(2)} - \pi_3 y_{t-2}^{(3)} - \pi_4 y_{t-1}^{(3)} + \varepsilon_t.$$

The factor $1 - B$ is contained in Δ_4 but also in the construction of the variables $y^{(2)}$ and $y^{(3)}$. Therefore, the AR operator $\phi(z)$ in

$$\phi(B) y_t = \Delta_4 y_t + \pi_2 \Delta(1+B^2) y_{t-1} + \pi_3 \Delta(1+B) y_{t-2} + \pi_4 \Delta(1+B) y_{t-1}$$

has a unit root at $+1$. The t–test on $\pi_1 = 0$ is just a variant of the usual Dickey-Fuller test. Same tables are to be consulted.

The second coefficient π_2 is 0

If $\pi_2 = 0$, one has

$$\Delta_4 y_t = \pi_1 y^{(1)}_{t-1} - \pi_3 y^{(3)}_{t-2} - \pi_4 y^{(3)}_{t-1} + \varepsilon_t.$$

The factor $1 + B$ is contained in Δ_4 but also in the construction of the variables $y^{(1)}$ and $y^{(3)}$. Therefore, the AR operator $\phi(z)$ in

$$\phi(B) y_t = \Delta_4 y_t - \pi_1 (1+B)(1+B^2) y_{t-1} + \pi_3 \Delta (1+B) y_{t-2} + \pi_4 \Delta (1+B) y_{t-1}$$

contains a unit root at -1. The t–test for $\pi_2 = 0$ tests for the fast semi-annual cycle root. Critical values are the same as for the Dickey-Fuller test.

The coefficients π_3 and π_4 are 0

If $\pi_3 = \pi_4 = 0$, one has

$$\Delta_4 y_t = \pi_1 y_{t-1}^{(1)} - \pi_2 y_{t-1}^{(2)} + \varepsilon_t.$$

The factor $1 + B^2$ is contained in Δ_4 but also in the construction of the variables $y^{(1)}$ and $y^{(2)}$. Therefore, the AR operator $\phi(z)$ in

$$\phi(B) y_t = \Delta_4 y_t - \pi_1 (1 + B)(1 + B^2) y_{t-1} + \pi_2 \Delta (1 + B^2) y_{t-1}$$

contains the complex unit roots $\pm i$. The F–statistic for $\pi_3 = \pi_4 = 0$ tests for the annual seasonal frequency and needs special tables for significance points. Separate t–tests for $\pi_3 = 0$ and $\pi_4 = 0$ should not be used.

The three coefficients π_2, π_3, π_4 are 0

If $\pi_2 = \pi_3 = \pi_4 = 0$, one has

$$\Delta_4 y_t = \pi_1 y_{t-1}^{(1)} + \varepsilon_t.$$

The autoregressive operator $\phi(B)$ has the unit root at -1 and the complex roots at $\pm i$. There may be a unit root at $+1$ or not. The F–test for $\pi_2 = \pi_3 = \pi_4 = 0$ tests for the presence of *all* seasonal unit roots. When it rejects, this is evidence on either no unit root at -1 or no unit root at $\pm i$. The distribution of the F–statistic under the null has been tabulated by HEGY.

Deterministic terms in the HEGY test

Usually, either a constant or seasonal dummies are added to the HEGY regression, for example

$$\Delta_4 y_t = \sum_{s=1}^{4} \gamma_s \delta_{st} + \pi_1 y_{t-1}^{(1)} - \pi_2 y_{t-1}^{(2)} - \pi_3 y_{t-2}^{(3)} - \pi_4 y_{t-1}^{(3)} + \varepsilon_t,$$

This is done, as in the Dickey-Fuller test, to take care of the purely deterministic seasonality without (some) unit roots under the alternative. The dummies improve test properties. They also affect significance points for the HEGY F and t statistics.

Note that the model with $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0$ but $\gamma_s \neq 0$ is highly implausible.

Lag augmentation in the HEGY test

If errors are not white noise, distributions are invalid. To get white-noise errors, one can augment the equation by lags of $\Delta_4 y_t$. These lags are stationary under null and alternative.

$$\Delta_4 y_t = \sum_{s=1}^{4} \gamma_s \delta_{st} + \pi_1 y_{t-1}^{(1)} - \pi_2 y_{t-1}^{(2)} - \pi_3 y_{t-2}^{(3)} - \pi_4 y_{t-1}^{(3)} + \sum_{j=1}^{p} \psi_j \Delta_4 y_{t-j} + \varepsilon_t$$

Some algebra shows that every autoregression of order $p+4$ can be transformed into this form (ignoring the dummies).

Example: Austrian industrial production

Data for log production (without taking first differences) is for 1957-2009. AIC and also BIC recommend three additional augmenting lags, and we estimate the regression

$$\begin{aligned}\Delta_4 y_t &= \sum_{s=1}^{4} \gamma_s \delta_{st} + \pi_1 y_{t-1}^{(1)} - \pi_2 y_{t-1}^{(2)} - \pi_3 y_{t-2}^{(3)} - \pi_4 y_{t-1}^{(3)} \\ &\quad + \sum_{j=1}^{3} \psi_j \Delta_4 y_{t-j} + \varepsilon_t\end{aligned}$$

by OLS. First, we analyze the t–statistics for π_1 and π_2, and then the F–statistic for $\pi_3 = \pi_4 = 0$.

The statistic $t(\pi_1)$ is 2.10. Using the usual Dickey-Fuller μ, we see that this is insignificant. There is evidence on a unit root at $+1$, as expected.

The statistic $t(\pi_2)$ is 2.74. According to HEGY, we revert its sign. The literature gives a critical 5% value at -3.11 and a critical 10% value at -2.54 . Because $-2.54 > -2.74 > -3.11$, the unit root at -1 is rejected at 10% but not at 5%.

The statistic $F(\pi_3, \pi_4)$ is 8.08. This is larger than the 5% significance point by HEGY of 6.57, though smaller than the 1% point of 8.79 . The unit root pair at $\pm i$ is rejected at the usual 5% level.

No seasonal unit root at $\pm i$ but some evidence on a unit root at -1 and convincing evidence on a unit root at $+1$. The joint F–test $F(\pi_2, \pi_3, \pi_4)$ has a 1% point of 7.63, which is surpassed by the observed value of 8.48. Thus, the joint test would tend to reject all seasonal unit roots.