ECON 515 Time Series Analysis

Stationarity and Unit Roots

Zhan Gao

17 October 2024

Overview

Stationarity

Stationarity

- A time series is essentially an *chronically ordered* collection of random variables.
- $y_1, y_2, \cdots, y_T$ is a random draw from population of an infinite series

$$\cdots, y_{-1}, y_0, y_1, y_2, \cdots$$

- There is *dependence* within the series
- A single realization is observed... but statistical analysis requires *repeated patterns*
  - Recall: analysis of cross-sectional data relies on *independence* across observations

Stationarity

- Roughly speaking, stationarity requires the "*patterns*" to be *invariant* across time.
- *Strict stationarity*
  - $\mathbf{Pr}\left(Y_{t_1}, Y_{t_2}, \cdots, Y_{t_k}\right) = \mathbf{Pr}\left(Y_{t_1+h}, Y_{t_2+h}, \cdots, Y_{t_k+h}\right)$ for any $t_1, \cdots, t_k$, h.
- *Weak stationarity*
  - constant mean $\mu = \mathbb{E}[Y_t]$ and variance $\sigma^2 = \mathbf{Var}(Y_t)$ for all t
  - Autocovariance $\gamma(h) = \mathbf{Cov}(Y_t, Y_{t+h})$ depends only on h.
- Strict stationarity implies weak stationarity if first two moments exist.

Stationarity

- Wold Decomposition (Wold, 1938)
- Any trend-stationary process $\{y_t\}$ can be decomposed into a *deterministic trend* component and a schochastic *stationary process* component:

$$y_t = d_t + \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j},$$

where $a_0 = 1$ and $\sum_{j=0}^{\infty} \alpha_j^2 < K < \infty$, and $\{\epsilon_t\}$ is a serially uncorrelated process defined by inoovations, $\epsilon_t = y_t - \mathbb{E}[y_t|\mathcal{I}_{t-1}]$ where $\mathbb{E}[\epsilon_t^2|\mathcal{I}] = \sigma^2$, $\mathbb{E}[\epsilon_t d_s] = 0$ and $\mathcal{I}_{t-1} = \{y_{t-1}, y_{t-2}, \cdots\}$.

Stationarity

- The key to stationarity: $\{\alpha_j\}$ decays fast enough (*exponentially decay*) so that it is *absolutely summable* (i.e., $\sum_{j=0}^{\infty} |\alpha_j| < \infty$).
- The autocovariance function: $\gamma(h) = \text{Cov}(y_t, y_{t+h}) = \sum_{j=0}^{\infty} \alpha_j \alpha_{j+h} \sigma^2$
- Long run variance: $\text{lrvar}(y_t) = \text{var}\left[\lim_{T\to\infty} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t\right] = \sum_{h=-\infty}^{\infty} |\gamma(h)|$

$$\text{lrvar}(y_t) \le \sigma^2 \sum_{h=0}^{\infty} \sum_{j=0}^{\infty} |a_j a_{j+h}| \le \sigma^2 \left(\sum_{j=0}^{\infty} |a_j| \right)^2 < \infty$$

- The corresponding autocorrelation: $\sum_{j=0}^{\infty} |\rho(h)| \le \sum_{j=0}^{\infty} |\alpha_j| < \infty$

Autocovariance and Autocorrelation

- Useful to check these statistics as first exploratory steps
- Moment estimators:
  - $\hat{\mu}_T = \bar{y}_T = \frac{1}{T}(y_1 + y_2 + \ldots + y_T)$.
  - $\hat{\gamma}_T(h) = \frac{1}{T}\sum_{t=h+1}^{T}(y_t - \bar{y}_T)(y_{t-h} - \bar{y}_T)$
  - $\hat{\rho}_T(h) = \frac{\hat{\gamma}_T(h)}{\hat{\gamma}(0)}$
- Asymptotic properties; Ref. Pesaran (2015) Ch. 14.1 and 14.2.
- Yule-Walker equations; Ref. Pesaran (2015) p.280; AR(2) as a example

$$y_{t-h}y_t = \phi_1 y_{t-h}y_{t-1} + \phi_2 y_{t-h}y_{t-2} + y_{t-h}\varepsilon_t$$

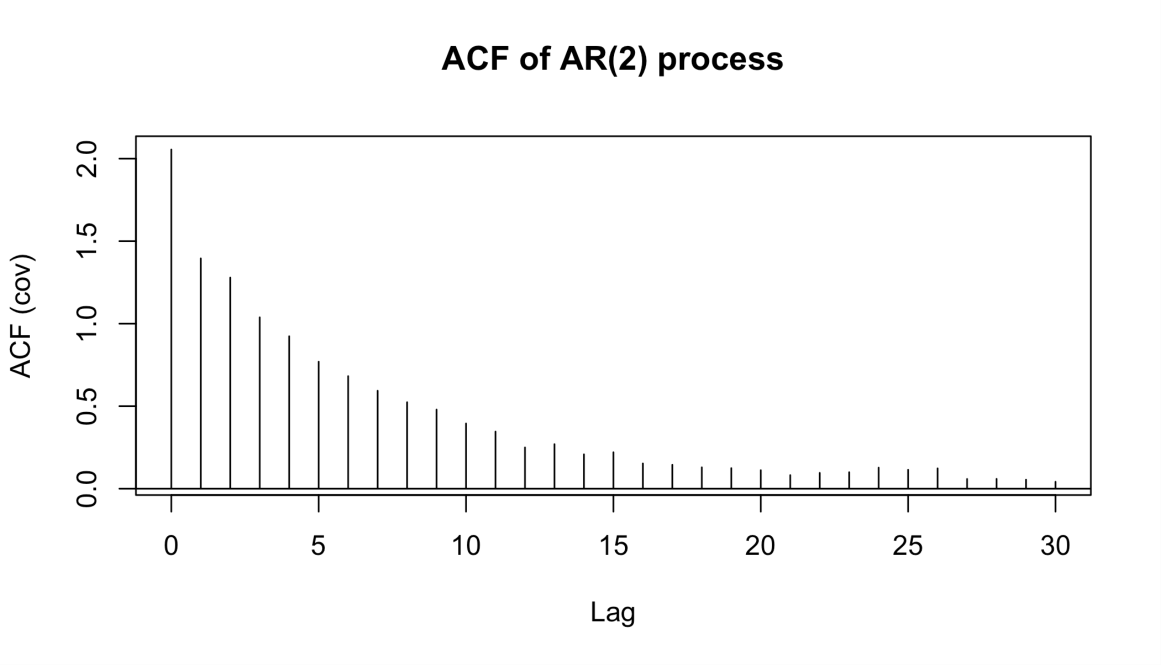
Taking expectation of both sides:

$$\gamma(h) - \phi_1\gamma(h-1) - \phi_2\gamma(h-2) = \begin{cases} 0 & h > 0 \\ \sigma^2 & h = 0 \end{cases}$$

Solve for $\phi_1 = \frac{\gamma(0)\gamma(1) - \gamma(1)\gamma(2)}{\gamma^2(0) - \gamma^2(1)}$ and $\phi_2 = \frac{\gamma(0)\gamma(2) - \gamma^2(1)}{\gamma^2(0) - \gamma^2(1)}$.

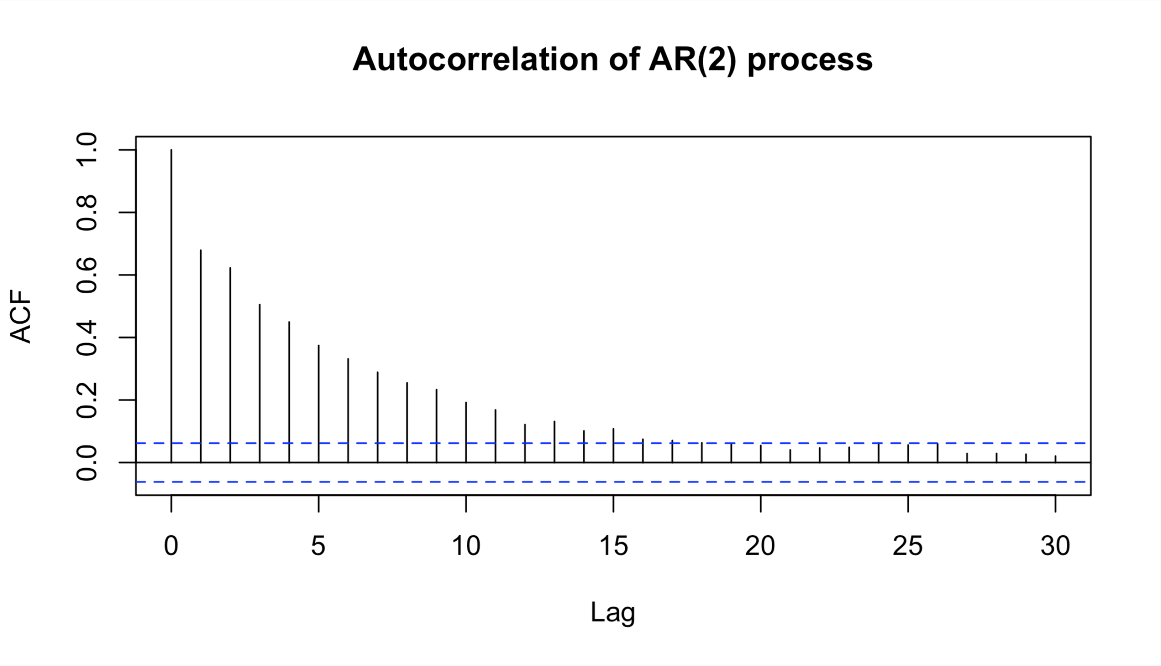
Autocovariance and Autocorrelation

```
set.seed(123)
y <- arima.sim(n = 1000, list(ar = c(0.5, 0.3)))
acf(y, type = "covariance", main = "ACF of AR(2) process")
```



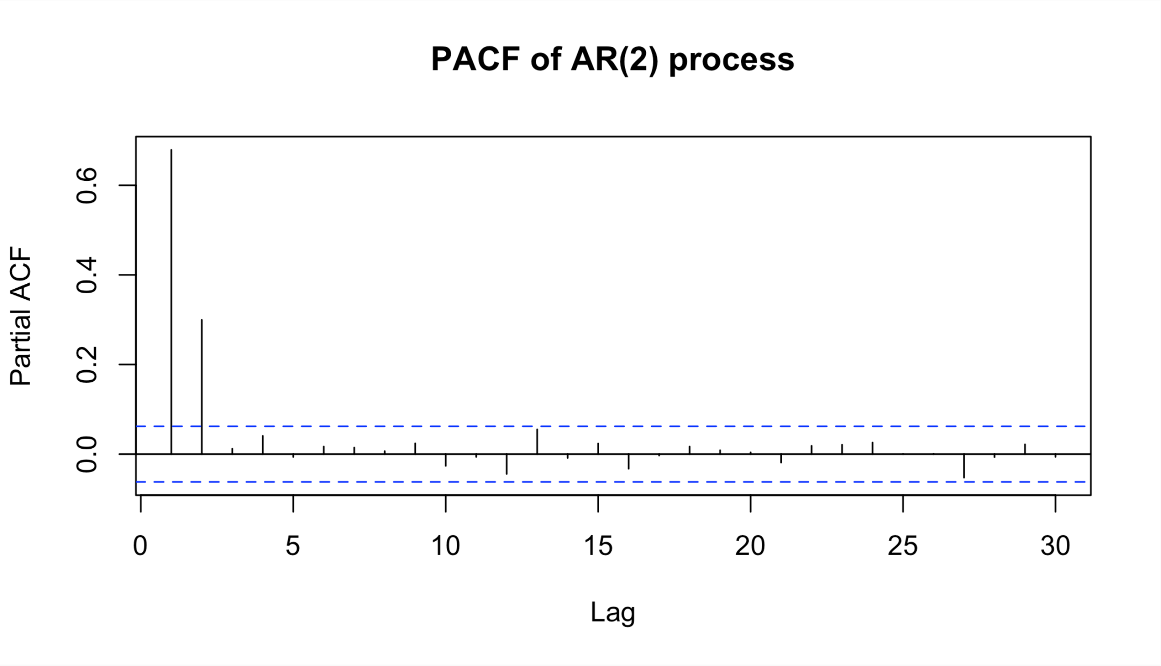
Autocovariance and Autocorrelation

```
acf(y, type = "correlation", main = "Autocorrelation of AR(2) process")
```



Partial Autocorrelation Function (PACF)

```
pacf(y, main = "PACF of AR(2) process")
```



Box-Pierce and Ljung-Box Test

- Test: $H_0 : \rho(1) = \rho(2) = \cdots = \rho(m) = 0$
- Box and Pierce (1970) Q statistic $Q_m = T \sum_{h=1}^{m} \hat{\rho}^2(h) \rightarrow_d \chi^2(m)$
  - Poor finite sample performance
- Ljung and Box (1978) $Q_m^* = T(T+2) \sum_{h=1}^{m} \frac{1}{T-h} \hat{\rho}^2(h) \rightarrow_d \chi^2(m)$

Box-Pierce and Ljung-Box Test

```
Box.test(y, lag = 10, type = "Box-Pierce")
```

```
##
##  Box-Pierce test
##
## data:  y
## X-squared = 1795.5, df = 10, p-value < 2.2e-16
```

```
Box.test(y, lag = 10, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  y
## X-squared = 1805.3, df = 10, p-value < 2.2e-16
```

ARIMA

Autoregressive Moving Average (ARMA) Process

- ARMA(p,q) process: $y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$
- Write the model in lag operator (backshift operator) notation:

$$\phi(L) y_t = \theta(L) \epsilon_t$$

$$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$$

$$\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$$

Autoregressive Moving Average (ARMA) Process

To invert the AR component: the roots of $\phi(z) = 0$ are outside the unit

- $y_t = \alpha(L)\theta(L)\epsilon_t$ where $\alpha(L)\phi(L) = 1$
- The latter is equivalent to $\alpha_i = \phi_1\alpha_{i-1} + \cdots + \phi_p\alpha_{i-p}$, $i = 1, 2, \cdots$ with $\alpha_0 = 1$ and $\alpha_i = 0$ for $i < 0$
- The solution is given in terms of roots of $\phi(z) = \prod_{j=1}^{p}(1 - \lambda_j z) = 0$
- $(1 - \lambda L)^{-1} = 1 + \lambda L + \lambda^2 L^2 + \cdots$ for $|\lambda| < 1$
- ARMA(p,q) $\rightsquigarrow$ MA (∞)

Other the other hand, the invertibility of the MA component $\rightsquigarrow$ a natural approach for esi.

$$y_t \approx \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_S y_{t-S} + \epsilon_t \text{ where } \beta(L)\phi(L) = 1$$

- *Truncation* as order S
- Estimate βs by Ordinary Least Squares (OLS)

Autoregressive Moving Average (ARMA) Process

Estimation of AR process: Yule-Walker / OLS / MLE

- they are asymptotically equivalent

Due to violation of strict exogeneity, OLS has a *small sample bias* (Details in Pesaran (2015) Ch. 14.5 and references therein)

- stationary AR(1) with normally distributed errors (Kendall (1954) and Marriott and Pope (1954))

$$\mathbf{E}\left(\hat{\phi}_{OLS}\right) - \phi = -\frac{1+3\phi}{T} + \mathcal{O}\left(\frac{1}{T^2}\right)$$

- *Bias-correction*
    - Orcutt and Winokur (1969) based on the formula above
    - Higher order autoregressive model - shaman and Stine (1988)
    - (Half) Jackknife bias-correction - Quenouille (1949)

Simulation: Bias of OLS estimator for the AR(1)

```
set.seed(100)
n_vec <- c(10, 20, 30, 50, 100)
num_n <- length(n_vec)
rho <- 0.5
R <- 2000
rho_hat_mat <- matrix(NA, R, num_n) # Initialize result container
for(i in 1:num_n) {
    n <- n_vec[i]
    for(r in 1:R){
        y <- arima.sim(model = list(order = c(1, 0, 0), ar = rho), n = n)
        rho_hat_mat[r, i] <- as.numeric(
            ar(y, order.max = 1, aic = FALSE, method = "ols")$ar
        )
    }
}
```

Simulation: Bias of OLS estimator for the AR(1)

```
data.frame(
    n = n_vec,
    bias = colMeans(rho_hat_mat) - rho,
    formula = -(1 + 3 * rho) / n_vec
)
```

```
##     n        bias     formula
## 1  10 -0.25481602 -0.25000000
## 2  20 -0.12655888 -0.12500000
## 3  30 -0.08788927 -0.08333333
## 4  50 -0.04736005 -0.05000000
## 5 100 -0.02640868 -0.02500000
```

You may try other values of ϕ and T to complete the picture.

Integrated Time Series

- (Loosely speaking) Weakly dependent time series is called *integrated of order 0* (I(0))
- *Integrated of order 1* (I(1)): the first difference $\Delta y_t = y_t - y_{t-1}$ is a I(0) process
- *Integrated of order d* (I(d)): the d-th difference $\Delta^d y_t$ is a I(0) process
- In real financial and economic applications, we rarely witness time series of integration order higher than 2
- `ARIMA(p, d, q)`: After d-th differencing, the time series becomes a stationary `ARMA(p, q)` process

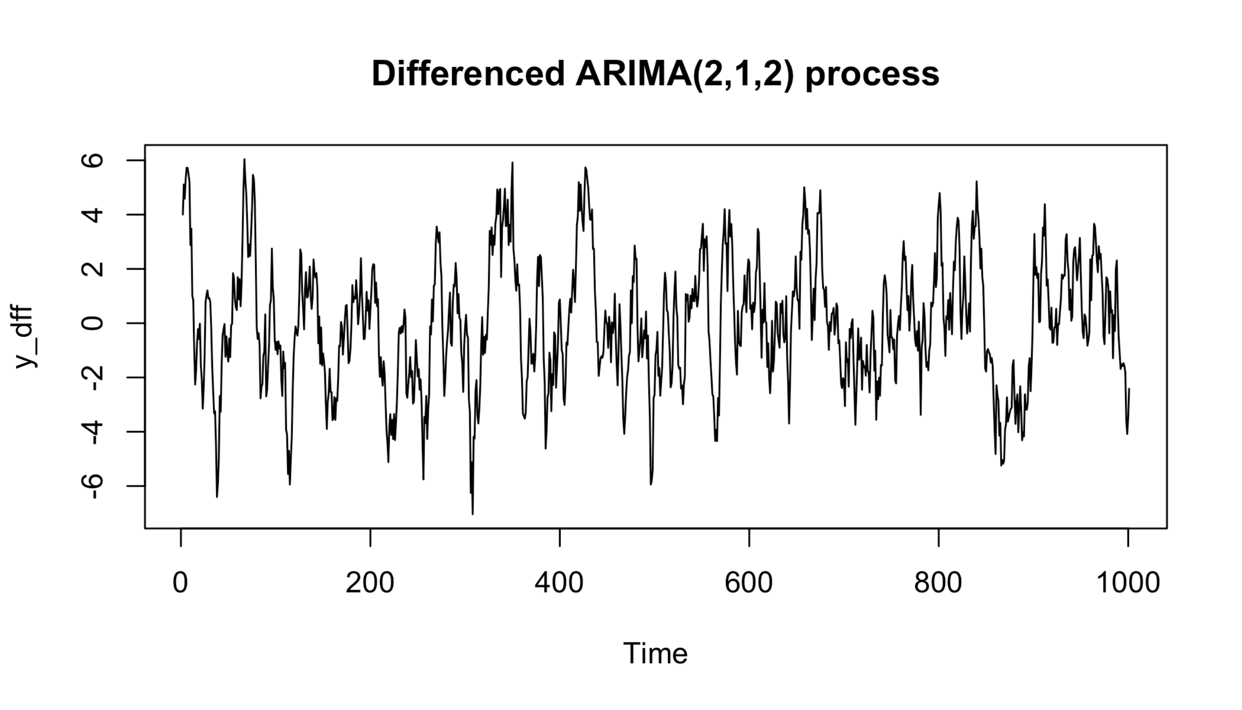
ARIMA(p,d,q)

- Simulate data from ARIMA model: `arima.sim(list(order = c(p,d,q), ar = , ma = )`

```
y ← arima.sim(n = 1000, list(order = c(2, 1, 2), ar = c(0.5, 0.3), ma = c(0.5, 0.3)))
plot(y, type = "l", main = "ARIMA(2,1,2) process")
```

ARIMA(p,d,q)

```r
y_dff <- diff(y)
plot(y_dff, type = "l", main = "Differenced ARIMA(2,1,2) process")
```



ARIMA(p,d,q)

- Estimation: `arima(y, order = c(p,d,q))`

```
arima(y, order = c(2, 1, 2))
```

```
##
## Call:
## arima(x = y, order = c(2, 1, 2))
##
## Coefficients:
##          ar1     ar2     ma1     ma2
##       0.4307  0.3671  0.5369  0.2767
## s.e.  0.0957  0.0894  0.0927  0.0347
##
## sigma^2 estimated as 0.9558:  log likelihood = -1397.29,  aic = 2804.59
```

ARIMA(p,d,q)

In practice, the order is unknown: Model selection by information criteria

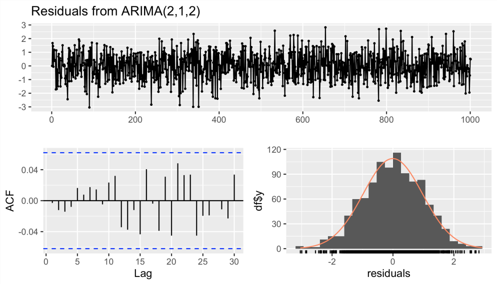
```
forecast::auto.arima(y)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
## Series: y
## ARIMA(2,1,2)
##
## Coefficients:
##          ar1     ar2     ma1     ma2
##       0.4307  0.3671  0.5369  0.2767
## s.e.  0.0957  0.0894  0.0927  0.0347
##
## sigma^2 = 0.9597:  log likelihood = -1397.29
## AIC=2804.59   AICc=2804.65   BIC=2829.13
```

ARIMA(p,d,q)

```
forecast::checkresiduals(forecast::auto.arima(y))
```



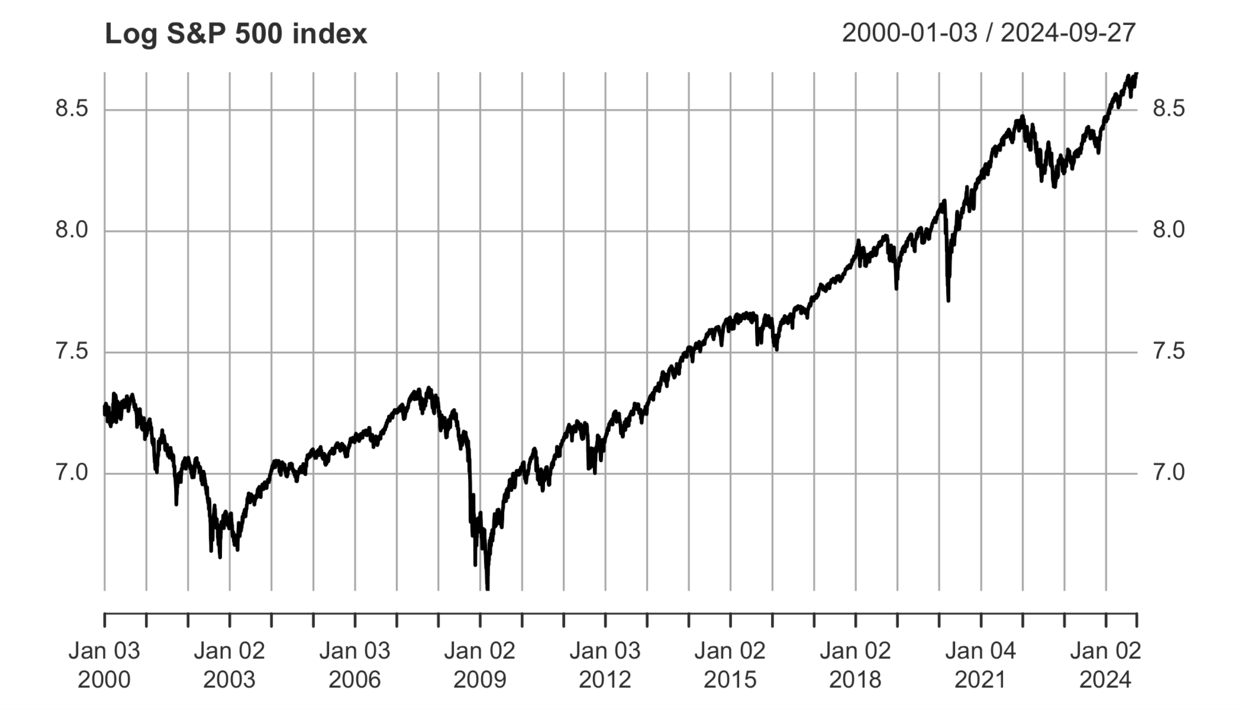

```
##
##      Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,2)
## Q* = 1.8579, df = 6, p-value = 0.9323
##
## Model df: 4.   Total lags used: 10
```

Real data: S&P 500 index

```
SPX <- quantmod::getSymbols("^GSPC",auto.assign = FALSE, from = "2000-01-01")$GSPC.Cl
plot(SPX)
```

Real data: S&P 500 index

```r
lSPX <- log(SPX)
plot(xts::xts(lSPX), main = "Log S&P 500 index")
```



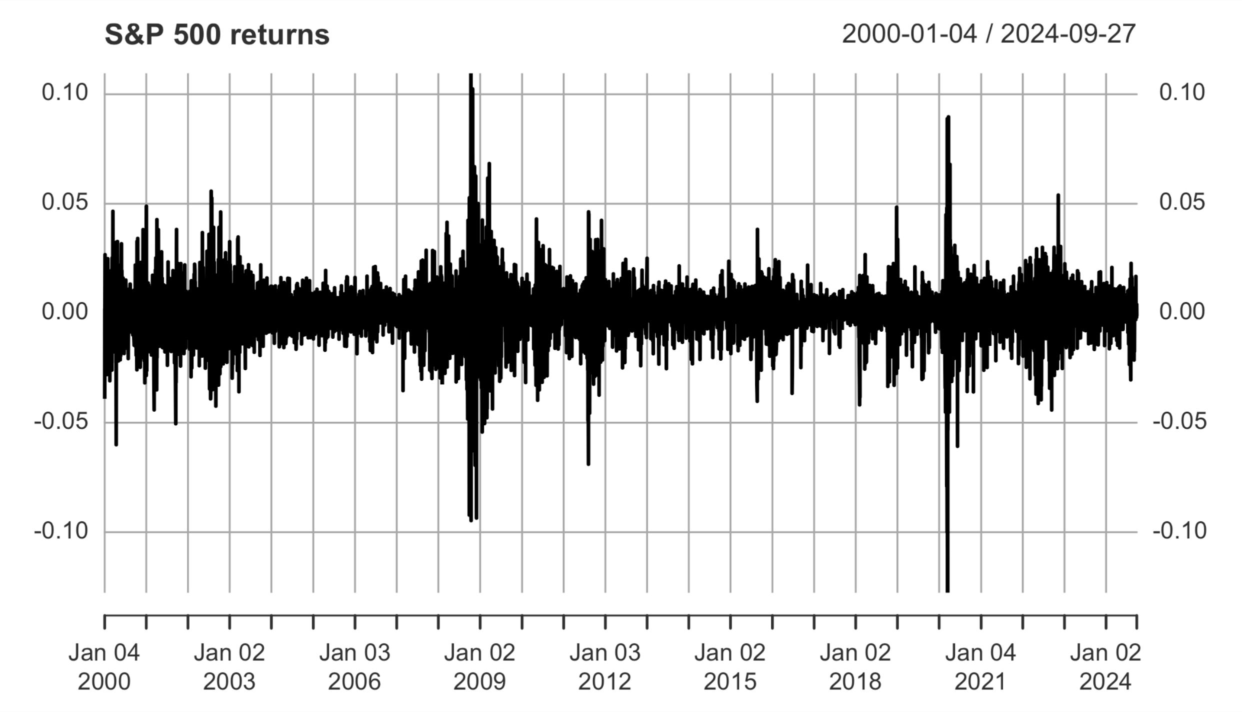
Real data: S&P 500 index

```r
forecast::auto.arima(lSPX)
```

```
## Series: lSPX
## ARIMA(1,1,0) with drift
##
## Coefficients:
##           ar1  drift
##       -0.1008  2e-04
## s.e.   0.0126  1e-04
##
## sigma^2 = 0.0001491:  log likelihood = 18586.31
## AIC=-37166.62   AICc=-37166.62   BIC=-37146.41
```

Real data: S&P 500 index

```
SPX_ret ← diff(lSPX)[-1]
plot(SPX_ret, type = "l", main = "S&P 500 returns")
```



Real data: S&P 500 index

```
forecast::auto.arima(SPX_ret)
```

```
## Series: SPX_ret
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##           ar1   mean
##       -0.1008  2e-04
## s.e.   0.0126  1e-04
##
## sigma^2 = 0.0001491:  log likelihood = 18586.31
## AIC=-37166.62   AICc=-37166.62   BIC=-37146.41
```

Detrending: Hodrick-Prescott filter

Detrending: Hodrick-Prescott filter

- The Hodrick-Prescott (HP) filter is a curve fitting procedure proposed by Hodrick and Prescott (1997) to estimate the trend path of a series.
- y_t is decomposed into a *trend* component and a *cyclical* component

$$y_t = y_t^* + c_t$$

- The HP filter:

$$\min_{y_1^\star, y_2^\star, \ldots y_T^\star} \left[\sum_{t=1}^{T} \left(y_t - y_t^\star\right)^2 + \lambda \sum_{t=2}^{T-1} \left(\Delta^2 y_{t+1}^\star\right)^2 \right]$$

  where λ is a tuning parameter.
- Conventional choice: $\lambda = 1600$ for quarterly data, $\lambda = 100$ for annual data.

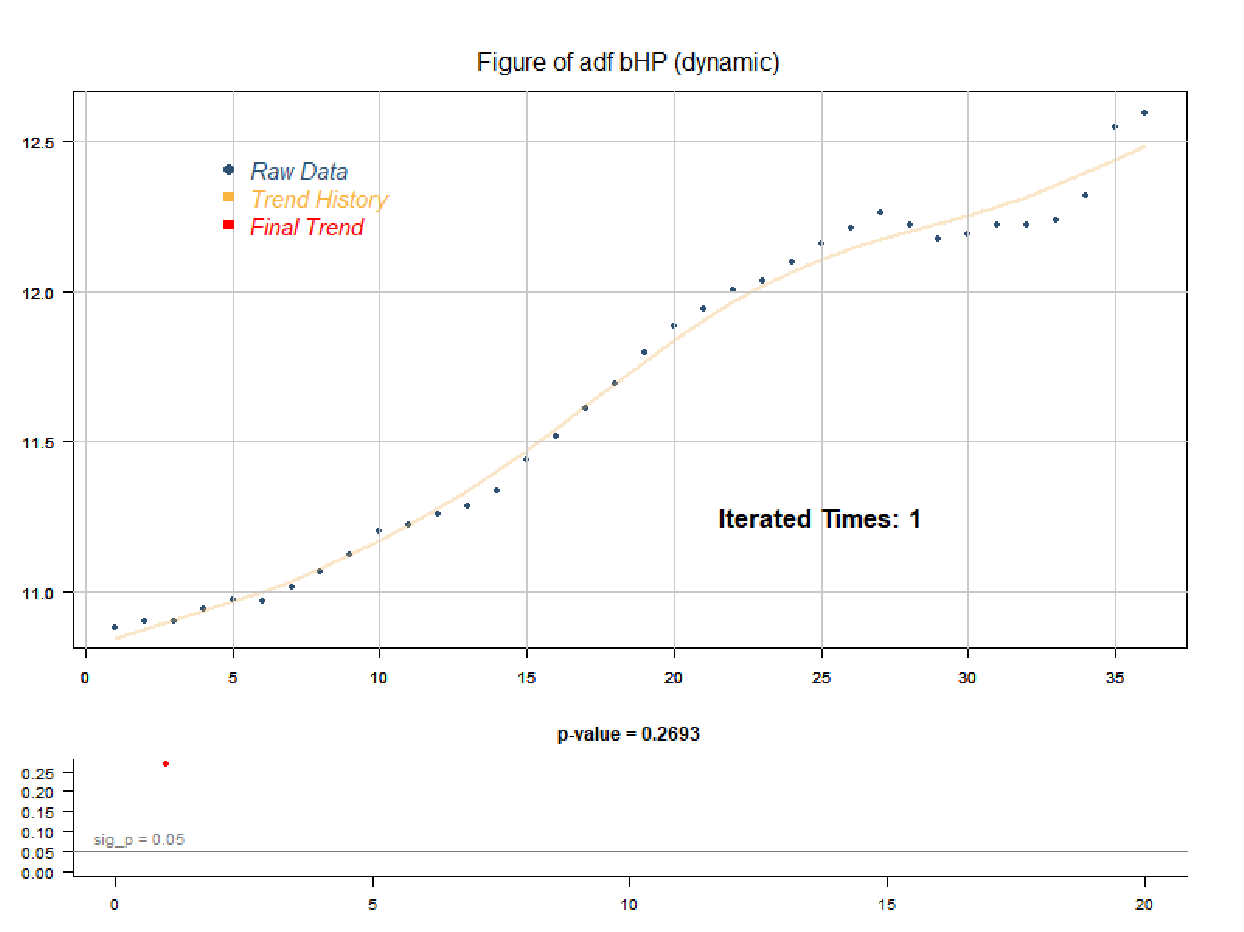
Boosted HP filter

- Phillips, P. C., & Shi, Z. (2021). Boosting: Why you can use the HP filter. *International Economic Review*, 62(2), 521-570.
- *Iterate* the HP filter to fully remove the trend

```
data(IRE) # load the data 'IRE'
lam <- 100 # tuning parameter for the annual data
# raw HP filter
bx_HP <- bHP::BoostedHP(IRE, lambda = lam, iter= FALSE)
# stopping stands for the condition of the terminal of iteration
# by BIC
bx_BIC <- bHP::BoostedHP(IRE, lambda = lam, iter= TRUE, stopping = "BIC")
# by ADF
bx_ADF <- bHP::BoostedHP(IRE, lambda = lam, iter= TRUE, stopping = "adf")
```

Boosted HP filter

```
#Dynamic Demonstration
plot(bx_ADF, iteration_location = "upright", interval_t = 0.8)
```

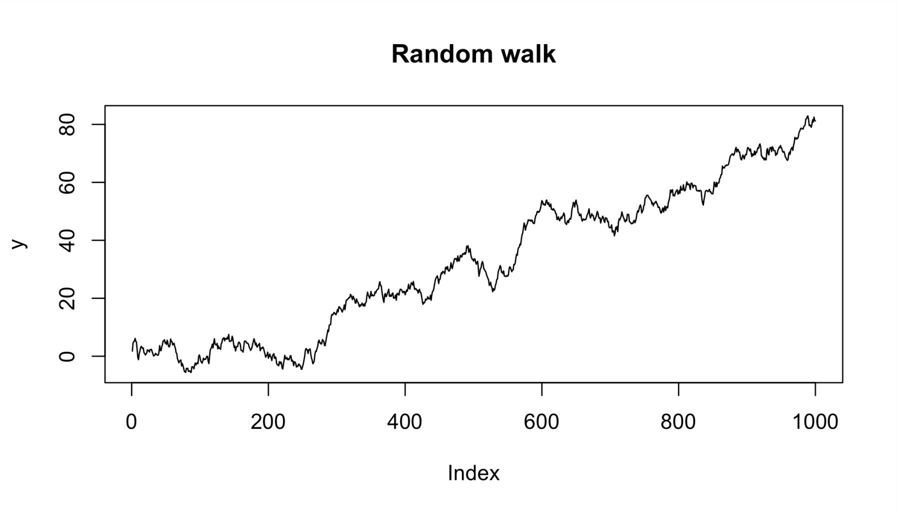


Unit Roots

Random walks

$$E_t[y_{t+h}] = E[y_{t+h}|y_t, y_{t-1}, \ldots] = y_t$$

- Related to efficient market hypothesis (EMH)
  - Samuelson (1965, Nobel 1970), Fama (1970, Nobel 2013)
  - Fama v.s. Thaler
- A simple model of random walk (with no drift) is the AR(1): $y_t = y_{t-1} + \epsilon_t$ with $\epsilon_t \sim iid(0, \sigma^2)$
  - AR(1) coefficient $\rho = 1$ (*unit root*) and no drift $\mu = 0$
  - `eps <- rnorm(n); y <- cumsum(eps)`



Random walks: Properties

- For simplicity, initial value $y_0 = 0$
- Shocks are *permenant*: $y_t = \epsilon_1 + \epsilon_2 + \cdots + \epsilon_t$
- *Nonstationarity*: $\mathrm{E}[y_t] = 0$; $\mathrm{var}[y_t] = t\sigma^2$; $\mathrm{cov}[y_t, y_s] = \min(t, s)\sigma^2$
- $\mathrm{E}_t[y_{t+h}] = \mathrm{E}[y_{t+h} | y_t, y_{t-1}, \ldots] = y_t$
- Compared to stationary AR(1)
  - The best mean prediction $E_t[y_{t+h}] = \beta^h y_t$ for $h > 0$
  - *Mean reversion* $E_t[y_{t+h}] \to 0$ as $h \to \infty$
  - *Diminishing shocks* $y_t = \sum_{q=0}^{t-1} \rho^q \varepsilon_{t-q}$
- Usual LLN and CLT don't apply in the presence of unit roots

Unit root Test

- Consider $y_t = \rho y_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.(0, \sigma^2)$
- Want to test the null hypothesis $H_0 : \rho = 1$ against the alternative $H_1 : |\rho| < 1$
- Let $\hat{\rho} = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}y_t}{T^{-1}\sum_{t=1}^{T} y_{t-1}^2}$ be the OLS estimator of ρ
- The t-statistic: $t_\rho = \frac{\hat{\rho}-1}{\hat{\sigma}_{\hat{\rho}}}$
  - $\hat{\sigma}_{\hat{\rho}} = s_T^2(\sum_{t=1}^{T} y_{t-1}^2)^{-1}$ where $s_T^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \hat{\rho}y_{t-1})^2$.
  - In the cross-sectional setup / stationary time series: t-statistic $\rightarrow N(0,1)$
  - In the presence of unit root: the limiting distribution of t_ρ is *non-standard*

Alternative representation

- Subtract y_{t-1} on both sides,

$$\Delta y_t = \tau y_{t-1} + \epsilon_t$$

  where $\tau = \rho - 1$.
- The hypotheses become $H_0 : \tau = 0$ against $H_1 : \tau < 0$.
- The t-statistic from OLS estimation of τ, t_τ, is exactly the same as t_ρ
- As a historical convention, most statistical software, such as the `urca` package in R, adopt the τ representation

Dickey-Fuller Test

- Dicky and Fuller (1979, 1981) study the asymptotic distribution of the t-statistic.
- The limiting distribution of t_ρ is a *stable* distribution
  - *non-standard*: Unlike Bernoulli/uniform/normal/etc distribution, we don't have an analytical form of the density of DF distrbution
  - It can be easily approximated by *Monte Carlo simulation*.
- Steps:
  1. Generate data under the null hypothesis $y_t = y_{t-1} + u_t$
  2. Compute the t-statistic t_ρ
  3. Repeat 1. and 2. for R times, with R large enough (say 10,000 or 20,000)
  4. Calculate the critical values (for T) by finding the empirical quantile of the R t-statistics

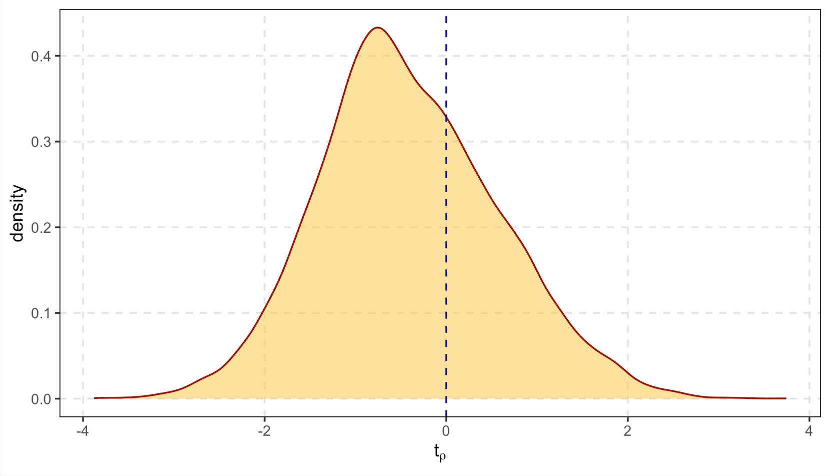
Dickey-Fuller Test

- `DF_sim` generates the t-statistics under H_0 / H_1 depending on ρ_0

```
DF_sim <- function(rho_0, n = 1000, num_rep = 20000) {
  test_stat <- rep(NA, num_rep)
  t0 <- Sys.time()
  for(r in 1:num_rep) {
      if (rho_0 == 1) {
          x <- cumsum(rnorm(n))
      } else {
          x <- arima.sim(model = list(order = c(1, 0, 0), ar = rho_0), n = n)
      }
      rho_hat <- lsfit(x[-n], x[-1], intercept = FALSE)$coefficients
      sigma2 <- mean((x[-1] - rho_hat * x[-n])^2) / (sum(x[-n]^2))
      test_stat[r] <- (rho_hat - rho_0) / sqrt(sigma2)
  }
  return(test_stat)
}
```

Dickey-Fuller Test

```
t_stat_ur ← DF_sim(1)
p ← ggplot(data.frame(tstat = t_stat_ur), aes(x = tstat)) +
    geom_density(color = "#990000", fill = "#FFC72C", alpha = 0.5) +
    geom_vline(aes(xintercept = 0), color = "navyblue", linetype ="dashed")
```




```
quantile(t_stat_ur, c(0.025, 0.05, 0.95, 0.975)) # Critical values
```

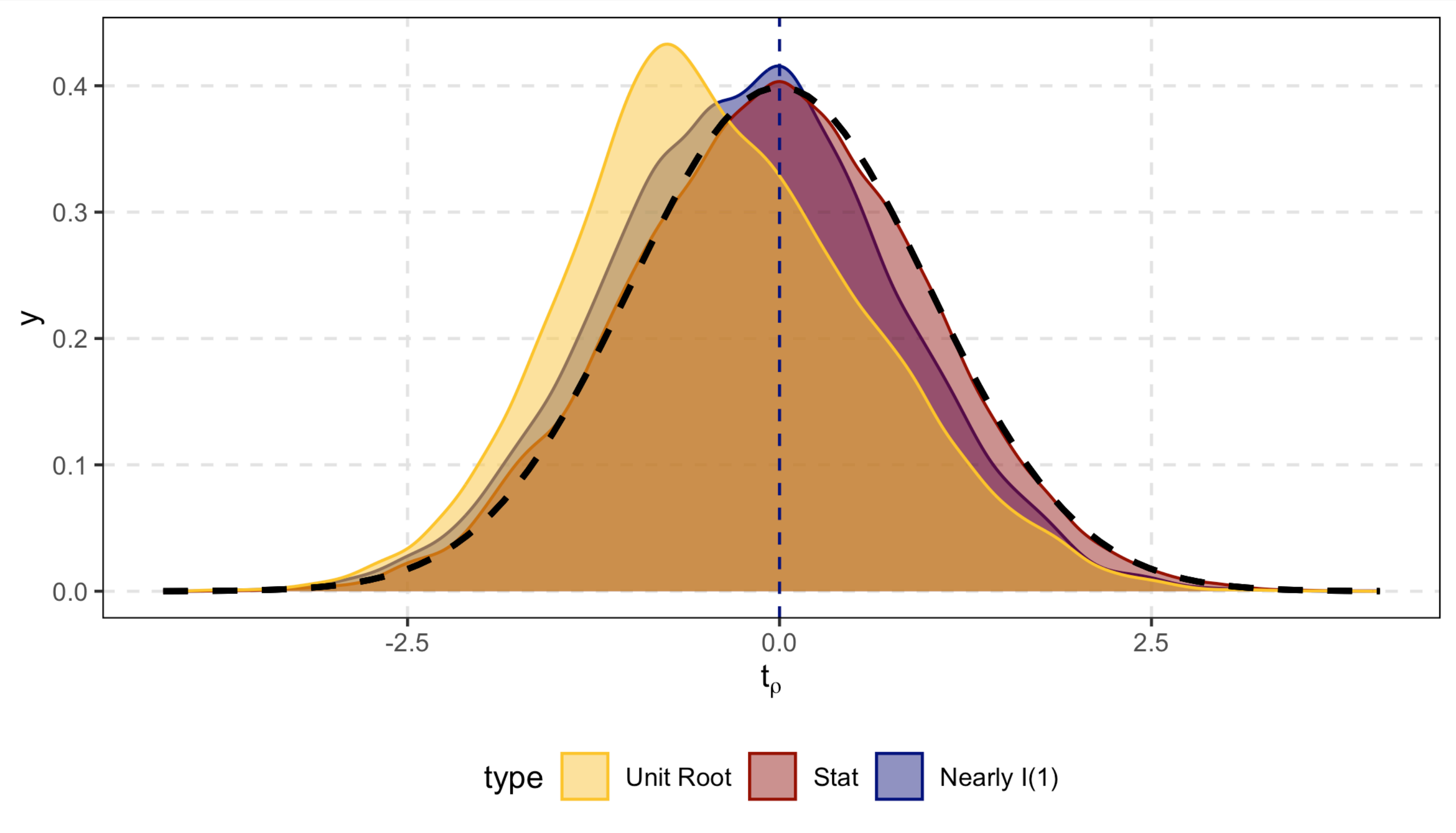
```
##      2.5%        5%       95%     97.5%
## -2.232765 -1.941602  1.312006  1.659953
```

Dickey-Fuller Test

- Compare with empirical distribution of t-statistics with different ρ_0

```
t_stat_stat ← DF_sim(0.5)
t_stat_near_ur ← DF_sim(0.99)
df ← reshape2::melt(list("Unit Root" = t_stat_ur, "Stat" = t_stat_stat, "Nearly I(1)'
names(df) ← c("tstat", "type")
p ← ggplot(df, aes(x = tstat, fill = type, color = type)) +
    geom_density(alpha = 0.5) +
    stat_function(fun = dnorm, color = "black", linetype = "dashed", linewidth = 1 ) -
    geom_vline(aes(xintercept = 0), color = "navyblue", linetype ="dashed") +
    scale_fill_manual(breaks = c("Unit Root", "Stat", "Nearly I(1)"), values = c("#FF(
    scale_color_manual(breaks = c("Unit Root", "Stat", "Nearly I(1)"), values = c("#FI
```

Dickey-Fuller Test



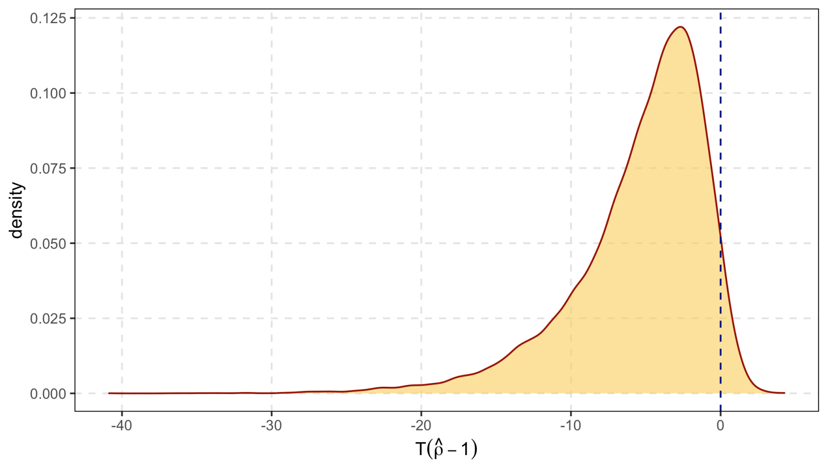
Dicky-Fuller Test

- *Dickey-Fuller (DF) distribution*

$$T(\hat{\rho}-1) = \frac{T^{-1}\sum_{t=1}^{T} X_{t-1}u_t}{T^{-2}\sum_{t=1}^{T} X_{t-1}^2} \Rightarrow \frac{\int_0^1 W(r)dW(r)}{\int_0^1 [W(r)]^2 dr}$$

as $T \to \infty$ by Functional CLT (*Ref. White (2001) Asymp. Theory for Econometricians*), where $W(\cdot)$ is a standard Brownian motion.

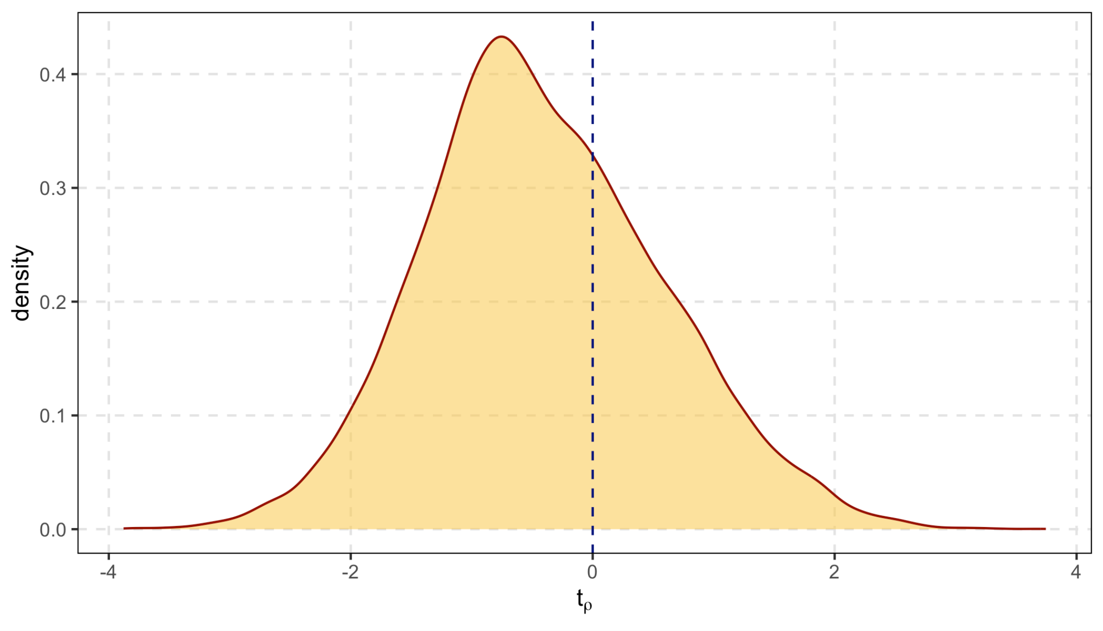
  - Super consistency with rate of convergence T instead of the regular $\sqrt{T}$ rate



Dickey-Fuller Test

- The limit distribution of t_ρ:

$$t_\rho \Rightarrow \frac{\int_0^1 W(r)\, dW(r)}{\left\{\int_0^1 [W(r)]^2 dr\right\}^{1/2}}$$



Dickey-Fuller Test

- Implement DF test with mature packages

```r
library(urca, quietly = TRUE)
n <- 100
y <- arima.sim( n = n, list(order = c(0,1,0) ) )
DFtest <- ur.df( y, type = "none", lags = 0 )
summary(DFtest)
```

- One-sided test
- The t-statistic is usually negative
- Pay attention to the critical values
- The more negative is the t-statistic, the stronger is the evidence of rejection

Dickey-Fuller Test

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1)

Residuals:
```

Dickey-Fuller Test

An example when the null is false

```
n <- 100
y <- arima.sim( n = n, list(ar = 0.5 ) )
DFtest <- ur.df( y, type = "none", lags = 0 )
```

Dickey-Fuller Test

```
summary(DFtest)
```

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1)

Residuals:
```

Dickey-Fuller Test

```
SPX ← quantmod::getSymbols("^GSPC",auto.assign = FALSE, from = "2000-01-01")$GSPC.Cl(
lSPX ← log(SPX)
DFtest ← ur.df( lSPX, type = "none", lags = 0 )
```

Dickey-Fuller Test

```
summary(DFtest)
```

```


###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none



Call:
lm(formula = z.diff ~ z.lag.1 - 1)

Residuals:
```

Estimation with Transformed Data

- Judging stationary based on tests is subject to testing errors - *pretesting issue*
- Many applied economists are inclined to transform a potentially nonstationary time series into a stationary time series, in order to circumvent the inconvenience brought by nonstationarity (*stationarization*)
  - To facilitate the stationarization, **FRED** assigns a transformation code (TCODE), as recommended transformation of potentially nonstationary time series into stationary ones.

```r
head(as.tibble(read.csv("./data/fred_md_data.csv"))[, 1:10])
```

```
## # A tibble: 6 × 10
##   sasdate   RPI W875RX1 DPCERA3M086SBEA CMRMTSPLx RETAILx INDPRO IPFPNSS IPFINAL
##   <chr>   <dbl>   <dbl>           <dbl>     <dbl>   <dbl>  <dbl>   <dbl>   <dbl>
## 1 Transf…    5       5               5          5       5      5       5       5
## 2 7/1/19… 2711.   2538.             21.5   289368.  21423.   27.7    28.7    27.4
## 3 8/1/19… 2722.   2548.             21.6   287421.  21396.   27.8    29.0    27.7
## 4 9/1/19… 2739.   2565.             21.6   284734.  21343.   28.1    29.1    27.7
## 5 10/1/1… 2755.   2580.             21.6   292581.  21714.   28.2    29.3    27.9
## 6 11/1/1… 2760.   2585.             21.7   286944.  21470.   28.4    29.4    28.0
## # ℹ 1 more variable: IPCONGD <dbl>
```

```r
head(as.tibble(fbi::fredmd_description)[, 2:5])
```

```
## # A tibble: 6 × 4
##   tcode ttype                                        fred           description
##   <fct> <fct>                                        <chr>          <chr>
## 1 5     First difference of natural log: ln(x)-ln(x-1) RPI            Real Perso…
## 2 5     First difference of natural log: ln(x)-ln(x-1) W875RX1        Real perso…
## 3 5     First difference of natural log: ln(x)-ln(x-1) DPCERA3M086S… Real perso…
## 4 5     First difference of natural log: ln(x)-ln(x-1) CMRMTSPLx      Real Manu.…
## 5 5     First difference of natural log: ln(x)-ln(x-1) RETAILx        Retail and…
## 6 5     First difference of natural log: ln(x)-ln(x-1) INDPRO         IP Index
```

Estimation with Transformed Data

- Is *stationarization* a sound practice?
- Suppose y_t follows AR(1) process $y_t = \rho y_{t-1} + \epsilon_t$ where $0 < \rho \leq 1$
- Take the first difference: $\Delta y_t = y_t - y_{t-1}$. What happens if we regress Δy_t on Δy_{t-1}?

$$y_t - y_{t-1} = \rho(y_{t-1} - y_{t-2}) + \epsilon_t - \epsilon_{t-1}$$

- The OLS estimator of ρ is not only biased but also inconsistent to ρ
  - Note that $\mathbf{cov}(y_{t-1} - y_{t-2}, \epsilon_t - \epsilon_{t-1}) = -\sigma^2 \neq 0$.
- When $\rho = 1$:

$$\hat{\rho} = \frac{T^{-1} \sum \epsilon_t \epsilon_{t-1}}{T^{-1} \sum \epsilon_{t-1}^2} \to_p 0$$

  instead of 1.
- The level data regression and the differenced data regression are about two different relationships. One does not imply the other.

Extensions of Pure Random Walk

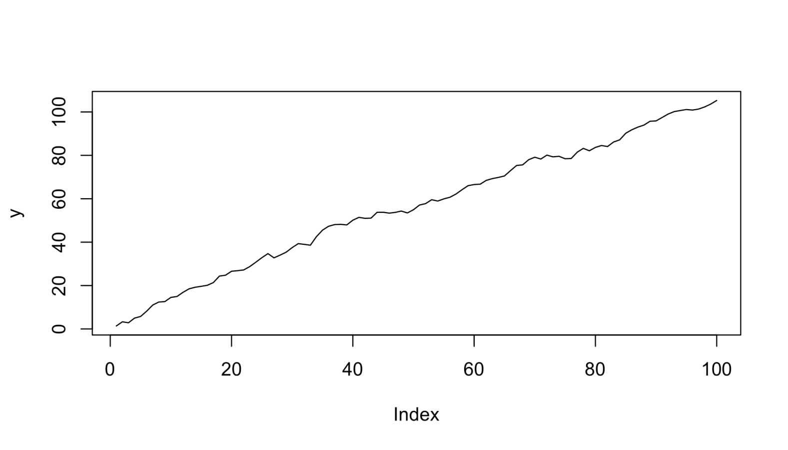
Random walk with *drift*

- AR(1) with AR coefficient $\rho = 1$ and a non-zero drift $\mu \neq 0$:

$$y_t = \mu + y_{t-1} + \epsilon_t$$

  where $\epsilon_t \sim iid(0, \sigma^2)$.

- Assume initial value $y_0 = 0$: $y_t = t\mu + \sum_{s=1}^{t} \epsilon_s$
  - *linear deterministic trend* + stochastic trend component (pure random wallk)
- $E[y_t] = t\mu$, $\text{var}[y_t] = t\sigma^2$, $E_t[y_{t+h}] = h\mu + y_t$ for $h > 0$



Random Walk with Drift

Under the null hypothesis $\mu = 0$ and $\rho = 1$,

$$T(\hat{\rho} - 1) \Rightarrow \frac{\int_0^1 W(r)\, dW(r) - W(1) \int_0^1 W(r)\, dr}{\int_0^1 [W(r)]^2 dr - \left[\int_0^1 W(r)\, dr\right]^2} = \frac{\int_0^1 \bar{W}(r)\, d\tilde{W}(r)}{\int_0^1 \left[\tilde{W}(r)\right]^2 dr}$$

- this distribution is even more strongly skewed than that for the case without drift
- for $T > 25$, 95% of the time the estimated $\hat{\rho}$ will be less than unity
- the limit distribution of t-statistic changes accordingly

$$t_\rho = \frac{\hat{\rho} - 1}{\hat{\sigma}_{\hat{\rho}}} \Rightarrow \frac{\int_0^1 \tilde{W}(r)\, d\tilde{W}(r)}{\left\{\int_0^1 \left[\tilde{W}(r)\right]^2 dr\right\}^{1/2}}.$$

- joint test of $\mu = 0$ and $\rho = 1$ with Wald form of F test

Random Walk with Drift

Under the null hypothesis $\mu \neq 0$ and $\rho = 1$, the limit distribution radically changes:

$$\begin{bmatrix} \sqrt{T}\left(\hat{\mu} - \mu\right) \\ T^{3/2}\left(\hat{\rho} - 1\right) \end{bmatrix} \Rightarrow N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \mu/2 \\ \mu/2 & \mu^2/3 \end{bmatrix}^{-1}\right)$$

- The limit distribution exactly the same as the limit distribution of the OLS estimator of the model

$$y_t = \mu + \delta t + \epsilon_t$$

- Recall that $y_t = \mu + \underbrace{(t-1)\mu + \sum_{s=1}^{t-1} \epsilon_s}_{y_{t-1}} + \epsilon_t$
  - the (nontrivial) time trend asymptotically dominates the stochastic trend
  - in large samples, it behaves as if the regressor y_{t-1} is replaced by the deterministic trend
- This specification cannot differentiate between the deterministic trend and the random walk with drift

Random Walk with Drift and Trend

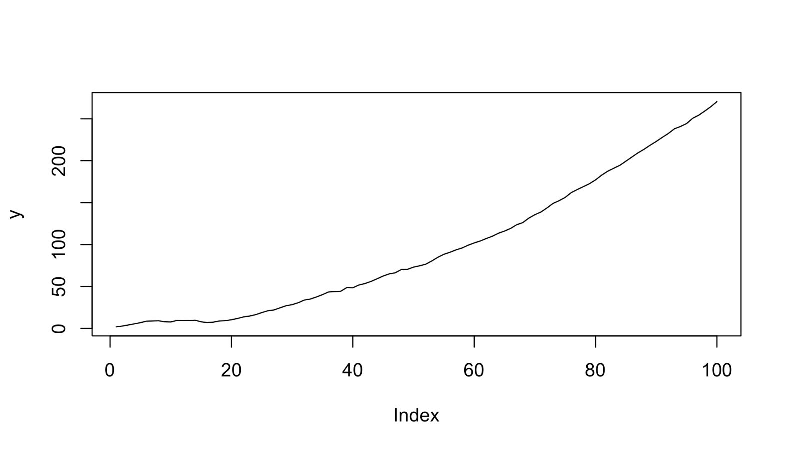
Now we add a linear deterministic trend term in the specification:

$$y = \mu + \delta t + \rho y_{t-1} + \epsilon_t$$

- Assume initial value $y_0 = 0$ and $\rho = 1$:

$$\begin{aligned} y_t &= \mu t + \delta(1 + 2 + \cdots + t) + \varepsilon_1 + \cdots + \varepsilon_t \\ &= \underbrace{\mu t + \frac{\delta}{2} t(t+1)}_{\text{deterministic trend}} + \underbrace{\varepsilon_1 + \cdots + \varepsilon_t}_{\text{stochastic component}} \end{aligned}$$

- Quadratic trend: $E[y_t] = \mu t + \frac{\delta}{2} t(t+1)$.



Random Walk with Drift and Trend

$$y = \mu + \delta t + \rho y_{t-1} + \epsilon_t$$

- As in previous case, under the null hypothesis $\rho = 1$ and $\delta = 0$
- The regressor y_{t-1} asymp. equivalent to a time trend $\rightsquigarrow$ multicollinearity
- The idea: subtract $\mu(t-1)$ from y_{t-1}

$$\begin{aligned} y_t = & \quad (1-\rho)\,\mu + \rho\left(y_{t-1} - \mu\,(t-1)\right) + (\delta + \rho\mu)\,t + \epsilon_t \\ = & \quad \mu^* + \rho^* \xi_{t-1} + \delta^* t + \epsilon_t, \end{aligned}$$

where $\rho^* = \rho$, ξ_t is a random walk without drift.

- under H_0, the limit distribution of the OLS estimator of the hypothetical regression:

$$\begin{bmatrix} T^{1/2}\left(\hat{\mu}^* - 0\right)/\sigma \\ T\left(\hat{\rho}^* - 1\right) \\ T^{3/2}\left(\hat{\delta}^* - \mu\right)/\sigma \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \int_0^1 W(r)\,dr & 1/2 \\ \int_0^1 W(r)\,dr & \int_0^1 [W(r)]^2 dr & \int_0^1 rW(r)\,dr \\ 1/2 & \int_0^1 rW(r)\,dr & 1/3 \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int_0^1 W(r)\,dW(r) \\ \int_0^1 r dW(r) \end{bmatrix}$$

- The limit distribution of $\hat{\rho}$ can be obstained by $T(\hat{\rho}^* - 1) = T(\hat{\rho} - 1)$
- The distribution does not depend on μ and σ; whether $\mu = 0$ or not does not matter.

Discussion of Specifications of DF tests

- Three specifications of DF tests:
  - Random walk: $y_t = y_{t-1} + \epsilon_t$
  - Random walk with drift: $y_t = \mu + y_{t-1} + \epsilon_t$
  - Random walk with drift and trend: $y_t = \mu + \delta t + y_{t-1} + \epsilon_t$
- Each specification leads to a different asymptotic distribution, and thus provides different critical values.
- Which is the "right" one to use?
  - If you have a specific null hypothesis about the process - go for it
  - If not, fit a specification, that is a plausible description of the data under both the null hypothesis and the alternative.
  - For example, if you observe a obvious trend - use the random walk with drift and trend

Implementation

enough math... let's implement the DF test with different specifications

```
n ← 100
x ← 1 + rnorm(n) # mu = 1, sigma = 1
y_drift ← cumsum(x)
DFtest_drift ← ur.df( y_drift, type = "drift", lags = 0 )
```

```
n ← 100
x ← 0.2 + 0.05*(1:n) + rnorm(n) # mu = 1, sigma = 1
y_trend ← cumsum(x)
DFtest_trend ← ur.df( y_trend, type = "trend", lags = 0 )
```

Implementation

- The packages uses the representation

$$\Delta y_t = \mu + \tau y_{t-1} + \epsilon_t$$

- The null hypothesis is $\tau = 0$ and the alternative is $\tau < 0$

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################
```

- `tau2` refers to τ
- `phi1` refers to joint null hypothesis of $\mu = 0$ and $\tau = 0$; This statistic is non-negative. The bigger is the value, the stronger is the evidence of rejection.
- These joint tests are two-sided

Implementation

- For the trend specification, the packages uses the representation

$$\Delta y_t = \mu + \delta t + \tau y_{t-1} + \epsilon_t$$

- The null hypothesis is $\tau = 0$ and the alternative is $\tau < 0$

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################
```

- `tau3` refers to τ
- `phi2` refers to joint null hypothesis of $\mu = \tau = 0$
- `phi3` refers to joint null hypothesis of $\mu = \tau = \delta = 0$
- Different specifications have different critical values

Augmented Dickey-Fuller (ADF) Test

- The asymptotic distribution of the DF test is based on the assumption that the error term has *homoskedasticity* and *no serial correlation*
  - The assumption $\epsilon_t \sim iid(0, \sigma^2)$ was maintained
- To cope with the violation of the assumption of no serial correlation, the augmented Dicky-Fuller (ADF) test adds more differenced lag terms Δy_{t-j}, $j = 1, 2, \cdots, p$.
- Three specifications:
  - None: $\Delta y_t = \tau y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{t-j} + \epsilon_t$
  - With drift: $\Delta y_t = \mu + \tau y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{t-j} + \epsilon_t$
  - With drift and trend: $\Delta y_t = \tau + \delta t + \rho y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{t-j} + \epsilon_t$
- The lag terms are supposed to absorb serial correlation in the error term
- The number of lags can be decided by AIC or BIC.
- Under the null $\tau = 0$ - these are AR(p) for Δy_t; If y_t is I(1), then the AR(p) for Δy_t is stationary

Augmented Dickey-Fuller (ADF) Test

Let's walk through the AR(2) process to see how ADF works.

$$\begin{aligned} y_t = & \quad \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \epsilon_t \\ = & \quad (\varphi_1 + \varphi_2)\, y_{t-1} + \varphi_2\, (y_{t-2} - y_{t-1}) + \epsilon_t \\ = & \quad \rho y_{t-1} + \phi_1 \Delta y_{t-1} + \epsilon_t \end{aligned}$$

- Under the null hypothesis that there exists one unit root - $\rho = 1$ and $|\phi_1| < 1$
  - then we have $\Delta y_t = \phi_1 \Delta y_{t-1} + \epsilon_t := u_t$ where u_t admits a MA (∞) representation with LRV $\lambda^2 = \sigma^2 \left(\frac{1}{1-\phi_1}\right)^2$
  - including the lagged term Δy_{t-1} aborbs the serial correlation and separates the clean error ϵ_t out
- Limit distribution of the OLS estimator of $y_t = \rho y_{t-1} + \phi_1 \Delta y_{t-1} + \epsilon_t$:

$$T(\hat{\rho} - 1) \Rightarrow \frac{\lambda\sigma \int_0^1 W(r)\, dW(r)}{\lambda^2 \int_0^1 [W(r)]^2 dr} = \frac{\sigma \int_0^1 W(r)\, dW(r)}{\lambda \int_0^1 [W(r)]^2 dr}$$

- $\frac{1}{1-\hat{\phi}_1} \rightarrow_p \frac{1}{1-\phi_1} = \frac{\lambda}{\sigma}$; then $\frac{T(\hat{\rho}-1)}{1-\hat{\phi}_1} \Rightarrow \textbf{DF dist.}$ Similar results for the t-statistic.

Augmented Dickey-Fuller (ADF) Test

- The insight we learned from AR(2) extends to higher order
- Implementation is straightforward with the `urca` package

```
y ← arima.sim( model = list(order = c(3,1,1), ar = c(0.4, 0.2, 0.2), ma = 0.5), n = 1
df ← ur.df(y, type = "trend", lags = 10, selectlags="AIC" )
```

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend
```

Phillips-Perron (PP) Test

- Phillips and Perron (1988) handle *heteroskedasticity* and *no serial correlation*
  - stick to the simple AR(1) setup; no lags included

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

- serially correlated error term: $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$ where $\epsilon_t \sim iid(0, \sigma^2)$.
- The PP test statistic involves the *long-run variance*
  - naturally arise in the presence of serial correlation
  - semiparametric framework
  - nonparametric estimation of the long-run variance

Long-run Variance

- Recall: for generic time series X_t,

$$\begin{aligned}\text{var}[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t] &= \frac{1}{T}E[(\sum_{t=1}^{T}X_t)^2] \\ &= \frac{1}{T}E[\sum_{t=1}^{T}X_t^2 + 2\sum_{t=1}^{T}\sum_{j>1}^{T-j}X_tX_{t+j}] \\ &= \gamma_0 + 2\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\gamma_j\end{aligned}$$

- Long run variance: $\textbf{lrvar}(y_t) = \textbf{var}\left[\lim_{T\to\infty}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\right]$
  - can be defined for any time series, not necessarily stationary ones - as long as the variance above exists
  - compared to the plain variance γ_0, it takes the serial correlation into consideration
- For stationary process $X_t = C(L)\epsilon_t$,

$$\lambda^2 := \text{lrvar}(X_t) = \sum_{h=-\infty}^{\infty}|\gamma(h)| = \sigma^2\sum_{h=0}^{\infty}\sum_{j=0}^{\infty}|c_jc_{j+h}| = \sigma^2\left(\sum_{j=0}^{\infty}|c_j|\right)^2 = [\sigma^2C(1)]^2$$

Long-run Variance

OLS with Serially Correlated Error

- Gauss-Markov theorem is gone
- Simple regression: $\sqrt{T}(\hat{\beta} - \beta_0) = \sqrt{T} \times \frac{\Sigma(x_t - \bar{x})\varepsilon_t}{\Sigma(x_t - \bar{x})^2} = \frac{T^{-1/2}\sum(x_t - \bar{x})\varepsilon_t}{T^{-1}\sum(x_t - \bar{x})^2}$
- The numerator is $x_t \epsilon_t$ is serially correlated in general if ϵ is serially correlated
- The asymp. variance must be adjusted: $\frac{\mathbf{lrvar}[x_t \epsilon_t]}{(\mathbf{var}[x_t])^2}$ instead of simple $\mathbf{var}(\epsilon_t)/\mathbf{var}(x_t)$

Long-run Variance

Estimation of long-run variance

- Recall that $\lambda^2 = \rho_0 + 2\sum_{h=1}^{\infty}\gamma(h)$
- Impossible to estimate $\gamma(h)$ accurately for large h given the sample size is T
- Rely on the convergence property $\sum_{h=1}^{\infty}|\gamma(h)| < \infty$, the lrvar is approximated by truncation at some p (with $p/T \to 0$ for consistency): $\hat{\lambda}^2 = \hat{\gamma}_0 + 2\sum_{h=1}^{p} w_p(h)\hat{\gamma}(h)$
- $w_p(h)$ is the *kernel weight*
  - Nadaraya–Watson kernel: $w_p(h) = 1$
  - Bartlett kernel: $w_p(h) = 1 - h/(p+1)$ (Newey and West, 1987)
- The estimator for lrvar based on Bartlett kernel is also called Newey-West (NW) estimator
  - Phillips (1987): NW est. is consistent if $p \to \infty$ and $p/T^{1/4} \to 0$.
  - in practice, choosing p can be tricky

Phillips-Perron (PP) Test

Let's get back to unit root tests...

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

- serially correlated error term: $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$ where $\epsilon_t \sim iid(0, \sigma^2)$.
- Under the null $\rho = 1$, Phillips and Perron (1988) should that

$$T(\hat{\rho} - 1) \Rightarrow \frac{\int_0^1 W(r)dW(r)}{\int_0^1 [W(r)]^2 dr} + \frac{(1/2)\left\{\lambda^2 - \gamma_0\right\}}{\lambda^2 \int_0^1 \left[W\left(r\right)\right]^2 dr}$$

  where the additional term arises from the serial correlation.
- The idea is to find the finite sample counter part of it and move it to the left-hand side
- Note $T^2 \frac{\hat{\sigma}_{\hat{\rho}}}{s_T^2} = \left(T^{-2} \sum X_{t-1}^2\right)^{-1} \Rightarrow \left(\lambda^2 \int_0^1 \left[W\left(r\right)\right]^2 dr\right)^{-1}$; λ and γ_0 can consistently estimated
- Then appending $T^2 \frac{\hat{\sigma}_{\hat{\rho}}}{s_T^2} \times \frac{\hat{\lambda} - \hat{\gamma}_0}{2}$ to $T(\hat{\rho} - 1)$ gives us the DF distribution in the limit

Phillips-Perron (PP) Test

- *PP ρ test* (in `urca` this corresp. to `type = "Z-alpha"`)

$$T(\hat{\rho}-1) - T^2 \frac{\hat{\sigma}^2_{\hat{\rho}}}{s^2_T} \frac{\hat{\lambda}^2 - \hat{\gamma}_0}{2} \Rightarrow \frac{\int_0^1 W(r)\, dW(r)}{\int_0^1 [W(r)]^2 dr}$$

- *PP t test* (in `urca` this corresp. to `type = "Z-tau"`):

$$\sqrt{\frac{\hat{\gamma}_0}{\hat{\lambda}^2}} t_T - \left\{ T^2 \frac{\hat{\sigma}^2_{\hat{\rho}}}{s^2_T} \right\}^{1/2} \frac{\hat{\lambda}^2 - \hat{\gamma}_0}{2\sqrt{\hat{\lambda}^2}} \Rightarrow \frac{\int_0^1 W(r)\, dW(r)}{\left\{ \int_0^1 [W(r)]^2 dr \right\}^{1/2}}$$

- For the specifications with drift and trend, the PP test statistics has the same limiting distribution as those derived for DF tests.

Phillips-Perron (PP) Test

Implementation of PP test is similar as before using `urca` package

```
y ← arima.sim( model = list(order = c(3,1,1), ar = c(0.4, 0.2, 0.2), ma = 0.5), n = 
df ← ur.pp(y, type = "Z-tau", model = "trend")
```

```
#####################################
# Phillips–Perron Unit Root Test #
#####################################

Test regression with intercept and trend
```

Discussion

- Both ADF and PP test the null of a unit root against the alternative of stationarity
- Unit root tests may have low power against relevant alternatives
- Useful to perform tests of the null of stationarity as well as tests of the null of a unit root.

Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test

- Kwiatkowski, Phillips, Schmidt and Shin (1992) device a test *under the null of stationarity*
- Suppose the series can be decomposed into a *random walk* and a *staionary error*

$$y_t = w_t + u_t,$$

where $\Delta w_t = v_t$ for some $v_t \sim iid(0, \sigma_v^2)$ and $u_t = C(L)\epsilon_t$ for $\epsilon_t \sim iid(0, \sigma^2)$.

- Null hypothesis $H_0 : \sigma_v^2 = 0$
  - under which $w_t = w_0$ become a constant determined by the initial value
  - the regression is thus $y_t = w_0 + u_t$ (intercept only)
- Alternative $H_1 : \sigma_v^2 > 0$: the regression is a linear deterministic trend plus a random walk.
- The KPSS test statistic

$$KPSS = \frac{1}{T^2 \hat{\lambda}^2} \sum_{t=1}^{T} (\sum_{j=1}^{t} \hat{u}_j)^2 \begin{cases} \Rightarrow \int_0^1 [V(r)]^2 dr & \text{under } H_0 \\ = O_p(T/p) \rightarrow_p \infty & \text{under } H_1 \end{cases}$$

where $V(r) = W(r) - rW(1)$ is a standard Brownian bridge, $W(r)$ is a standard

Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test

- The baseline setup can be extended to include drift and trend

$$y_t = \mu + \delta t + w_t + u_t,$$

- Under the null hypothesis $H_0 : \sigma_v^2 = 0$, the regression is trend stationary
  $y_t = \mu_w + \delta t + u_t$ where w_0 is absorbed in the drift
- Regress y_t on a constant and a time trend to obtained residuals $\hat{u}_t$

$$KPSS \begin{cases} \Rightarrow \int_0^1 \left[V_2\left(r\right)\right]^2 dr & \text{under } H_0 \\ = O_p(T/p) \rightarrow_p \infty & \text{under } H_1 \end{cases}$$

where $V_2(r)$ is a second-level Brownian bridge defined by

$$V_2\left(r\right) = W\left(r\right) + \left(2r - 3r^2\right) W\left(1\right) + \left(-6r + 6r^2\right) \int_0^1 W\left(s\right) ds$$

Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test

- The KPSS test can also be implemented in the `urca` package

```
y ← arima.sim( model = list(order = c(3,1,1), ar = c(0.4, 0.2, 0.2), ma = 0.5), n = 1
ur.kpss(y, type = "tau",  lags = "short") ▷ summary()
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 7 lags.
##
## Value of test-statistic is: 1.7178
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
```

Alternative package: `tseries`

- Phillips-Perron (PP) Test
  - `tseries::pp.test(x)`
- Augmented Dickey-Fuller (ADF) Test
  - `tseries::adf.test(x)`
- Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test
  - `tseries::kpss.test(x)`

Bubble Testing

Bubble Testing

- The alternative hypothesis of conventional unit root tests is the stationary regime
- Financial bubbles and crisese have been witnessed in history: Financial crises often preceded by asset market bubbles
- How to detect bubbles?
- Essentially, want to test the null hypothesis: $\rho = 1$ (unit root) versus alternative hypothesis: $\rho > 1$ (explosive)

Past practice

- Diba and Grossman (1989):
  - unit-root test on first differenced price level (Δp_t)
  - cointegration test on stock price (p_t) and dividend series (d_t)
  - found no evidence of bubbles in historical data
- Evans (1991)
  - showed standard tests fail to detect explosive bubbles due to periodic collapse

Bubble Testing

Phillips, Wu and Yu (2011) Approach

- Applied right-tailed augmented Dickey-Fuller (ADF) test
- Used forward recursive rolling windows to improve power
  - Bubble is a transient phenomenon.
- Found strong evidence of explosive characteristics in p_t for 1990s data
- Cannot deal with multiple bubbles

Phillips, Shi and Yu (2015, PSY) Approach

- Proposed generalized sup ADF test (GSADF)
- Allows flexible starting and ending points for rolling windows
- Uses recursive backward regression technique for date stamping
- Using long historical monthly data, identify three big historical bubbles: 1890's, 1929, and 2001

PSY Test

Test the existence of exuberance behavior

- Rolling window:starting from r_1^{th} fraction of the sample and ending at the r_2^{th} fraction

$$\Delta y_t = \alpha_{r_1,r_2} + \beta_{r_1,r_2} y_{t-1} + \sum_{i=1}^{k} \psi_{r_1,r_2}^{i} \Delta y_{t-i} + \varepsilon_t$$

- ADF statistic based on the regression is denoted by $ADF_{r_1}^{r_2}$,
- Generalized sup ADF test statistic (varying both the starting and the ending point)

$$GSADF\left(r_0\right) = \sup_{\substack{r_2 \in [r_0,1] \\ r_1 \in [0, r_2 - r_0]}} \left\{ ADF_{r_1}^{r_2} \right\}$$

- Limiting distribution under the null

$$\sup_{\substack{r_2 \in [r_0,1] \\ r_1 \in [0, r_2 - r_0]}} \left\{ \frac{\frac{1}{2} r_w \left[W(r_2)^2 - W(r_1)^2 - r_w \right] - \int_{r_1}^{r_2} W(r) dr [W(r_2) - W(r_1)]}{r_w^{1/2} \left\{ r_w \int_{r_1}^{r_2} W(r)^2 dr - \left[\int_{r_1}^{r_2} W(r) dr \right]^2 \right\}^{1/2}} \right\}$$

PSY Test

Date-stamping Strategies

- Backward SADF test: performs a sup ADF test on a backward expanding sample sequence where the endpoint of each sample is fixed at r_2,

$$BSADF_{r_2}(r_0) = \sup_{r_1 \in [0, r_2 - r_0]} ADF_{r_1}^{r_2}$$

- compare $BSADF_{r_2}(r_0)$ to the critical value of the sup ADF statistic based on $\lfloor Tr_2 \rfloor$ observations for each $r_2 \in [r_0, 1]$

$$\hat{r}_e = \inf_{r_2 \in [r_0, 1]} \left\{ r_2 : BSADF_{r_2}(r_0) > scv_{r_2}^{\beta_T} \right\}$$

$$\hat{r}_f = \inf_{r_2 \in [\hat{r}_e + \delta \log(T)/T, 1]} \left\{ r_2 : BSADF_{r_2}(r_0) < scv_{r_2}^{\beta_T} \right\}$$

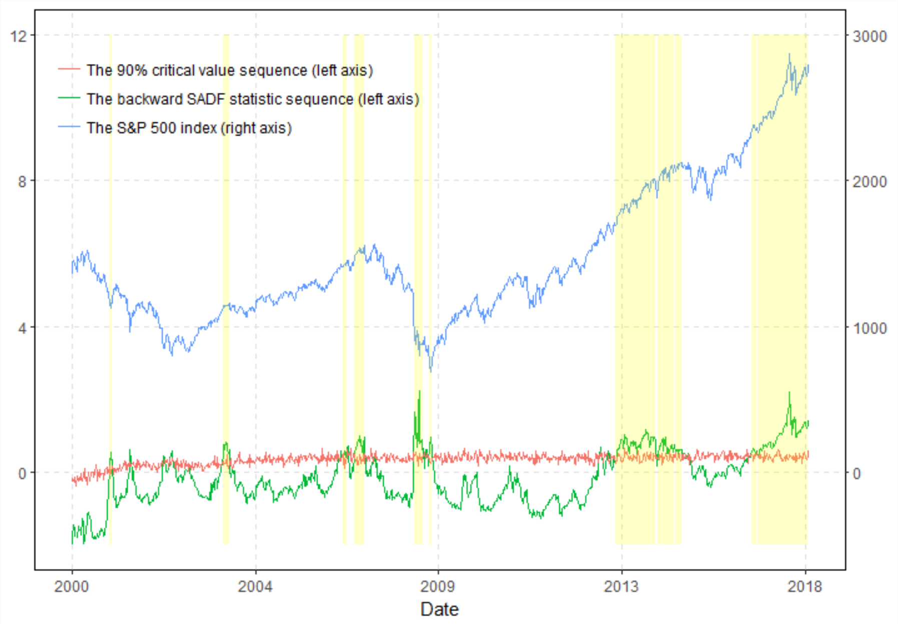
where $scv_{r_2}^{\beta_T}$ is the $100(1 - \beta_T)\%$ critical value of the sup ADF statistic based on $\lfloor Tr_2 \rfloor$ observations.

Discussion

- The test statistic is based on ADF test. Take into consideration of the multiple testing issue
- Reduced form by nature
- Work as a real time monitoring system
- In use in central banks

Implementation PSY Test

- `BubbleTest` based on the `MultipleBubbles` package
    - A wrapper with date-stamping function and visualization



- Recent R package `psymonitor`
    - Install from Github source
- Computationally intensive due to Monte Carlo simulation

Acknowledgements: Some parts of this set of slides are dependent on Professor Zhentao Shi's lecture notes at Georgia Tech (ECON 4160).