Unit root problem, solution of difference equations

Simple deterministic model, question of unit root

$$(1-\phi_1 L)X_t = \mu$$

,

$$X_t - \phi_1 X_{t-1} = \mu$$

Solution

$$X_t = A + Bz^t$$

with unknown z and unknown A (clearly $X_{t-1} = A + Bz^{t-1}$)

$$0 = X_t - \phi_1 X_{t-1} - \mu = A + Bz^t - \phi_1(A + Bz^{t-1}) - \mu$$

$$= A + Bz^t - \phi_1\left(A + \frac{B}{z}z^t\right) - \mu$$

$$= \underbrace{-\mu + A(1-\phi_1)}_{=0} + B\underbrace{\left(1 - \frac{\phi_1}{z}\right)}_{=0} z^t$$

$z = \phi_1$, $A = \frac{\mu}{1-\phi_1}$
Elimination B by the initial condition

$$X_t = \frac{\mu}{1-\phi_1} + B\phi_1^t$$

$$X_{t_0} = \frac{\mu}{1-\phi_1} + B\phi_1^{t_0} \longrightarrow B = \left[X_{t_0} - \frac{\mu}{1-\phi_1}\right]\phi_1^{-t_0}$$

$$X_t = \frac{\mu}{1-\phi_1} + \left[X_{t_0} - \frac{\mu}{1-\phi_1}\right]\phi_1^{-t_0}\phi_1^t$$

$$\left(X_t - \frac{\mu}{1-\phi_1}\right) = \left(X_{t_0} - \frac{\mu}{1-\phi_1}\right)\phi_1^{t-t_0}$$

$$\frac{X_t - \frac{\mu}{1-\phi_1}}{X_{t_0} - \frac{\mu}{1-\phi_1}} = \phi_1^{t-t_0}$$

Qualitative difference for $\phi_1 < 1$ and $\phi_1 > 1$ (unit root means $z = \phi_1 = 1$)
$\underline{\phi_1 = 0.8}$: $t_1 - t_0 = 0$, $\phi_1^0 = 1$; $t - t_0 = 1$, $\phi_1^1 = 0.8$; $t - t_0 = 2$, $\phi_1^2 = 0.64$; $t - t_0 = 3$; $\phi_1^3 = 0.52$; $\phi_1^{10} = 0.107$, $\phi_1^{100} = 2.3 \times 10^{-10}$;
$\underline{\phi_1 = 1.2}$: $t - t_0 = 10$, $\phi_1^{10} = 6.19$; $\phi_1^{100} = 82817974.522$;

Unit root signature of nonstationary behaviour $z = \phi_1 = 1$
Take homogeneous equation
$(1 - \phi_1 L)X_t = 0, \longrightarrow (1 - L)X_t = 0$
formally via the "trick" $L = 1/z$ we get characteristic equation
$(1 - \phi_1/z) = 0$ where the unit root is achieved for $z = \phi_1 = 1$

Generalization for AR(p) without stochastic component:

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = 0$$

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \phi_3 X_{t-3} \ldots \phi_p X_{t-p} = 0$$

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 \ldots - \phi_p L^p)X_t = 0$$

expected form of solution

$$X_t = z^t X_0, \qquad t = 0, 1, 2 \ldots$$

$$(1 - \phi_1(1/z) - \phi_2(1/z)^2 - \phi_3(1/z)^3 \ldots - \phi_p(1/z)^p)z^t X_0 = 0$$

$$\left[1 - \phi_1(1/z) - \phi_2(1/z)^2 - \phi_3(1/z)^3 \ldots - \phi_p(1/z)^p\right] z^t X_0 = 0$$

characteristic equation

$$z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \phi_3 z^{p-3} - \ldots - \phi_p = 0$$

generally has exactly p roots

$$(z - \lambda_1)(z - \lambda_2)\ldots(z - \lambda_p) = 0$$

$$\prod_{k=1}^{p} (z - \lambda_k)^k = 0$$

For non-identical (also complex) roots

$$X_t = B_1 \lambda_1^t + B_2 \lambda_2^t + \ldots + B_p \lambda_p^t$$

when roots are identical for e.g. $p = 2$

$$X_t = B_1 \lambda_1^t + B_2 t \lambda_1^t$$

Augmented Dickey-Fuller (ADF) unit root test

(Rozšírený DF test)

The test is widely used by practioners. In statistics, the traditional DickeyFuller test tests (1979) whether a unit root is present in an autoregressive model.

In time series models in econometrics (the application of statistical methods to economics), a **unit root** is a feature of processes that evolve through time that can cause **problems** in statistical inference if it is not adequately dealt with. Such a process is **non-stationary**.

(If the other roots of the characteristic equation lie inside the unit circle that is, have a modulus (absolute value) less than one then the first difference of the process will be stationary).

Other words: series of the first difference (previously considered integrated of order one) turned to be stationary.

R: test available in **adf.test**()
tseries() [package for Time series analysis and computational finance]
The testing procedure is applied to the model

$$\Delta X_t = \underbrace{\alpha}_{\text{drift}} + \underbrace{\delta\, t}_{\text{lin.trend}} + \underbrace{\varrho}_{\text{par. tested}} X_{t-1} + \sum_{j=1}^{p} \psi_j \Delta X_{t-j}$$

$\Delta X_{t-j} = X_{t-j} - X_{t-j-1}$, $\Delta X_t = X_t - X_{t-1}$
number of lags of autoregression; default:
$p = (\text{length of d.s.} - 1)^{1/3}$; (example: length of data series $= 65$, $p = 4$); Swert criterion: $12 \times (length\, of\; d.s./100)^{1/4}$
Alternative form:

$$X_t = \alpha + \delta t + \underbrace{(1+\varrho)}_{\phi_1 \textit{ in former notation}} X_{t-1} + \sum_{j=1}^{p} \psi_j \Delta X_{t-j}$$

$\sum_{j=1}^{p} \psi_j \Delta X_{t-j}$ the autoregression terms absorbing dynamic structure

H$_0$: $\varrho = 0$ for given model **has a unit root**, process is $I(1)$, i.e. the data needs to be differenced to make it stationary;

H$_1$: $\varrho < 1$ i.e. the data is **trend stationary**, process is $I(0)$ and needs to be analyzed by means of using a trend in the regression model instead of differencing the data;

The unit root test is carried out under H$_0$: $\varrho = 0$ against the alternative hypothesis of $\varrho < 0$.

Note 1: *Caution*: Sometimes if data is exponentially trending then you might need to take the log of the data first before differencing it.

Note 2: $\varrho = 0$ means $X_t \sim X_{t-1}$ (i.e. random walk process) test statistic [one-sided left tail test]:

$$ADF_\tau = \frac{\hat{\varrho}}{SE(\hat{\varrho})}$$

(remind that $t_{\text{score}} = \frac{\hat{\varrho}}{SE(\hat{\varrho})}$)
$SE(\hat{\varrho})$ is the usual standard error estimate
critical value: $t^*_{1-\alpha}(n)$

If the test statistic is less than the **critical value** $t^*_{1-\alpha}(n)$, then the H_0 ($\varrho = 0$) is rejected and **no unit root is present**.

Example: A model that includes a constant and a time trend is estimated using sample of $n = 50$ observations and yields the ADF_τ statistic of -4.57. This is more negative than the tabulated $t^* = -3.50$, so at the 95 per cent level the null hypothesis of a unit root will be rejected.

R implementation

```
> x=arima.sim(n = 1000, list(ma = c(-0.2, 0.2)),sd=1)
> adf.test(x)

Augmented Dickey-Fuller Test

data:  x
Dickey-Fuller = -9.8277, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(x) : p-value smaller than printed p-value

[Note: rejection of null hypothesis that data has
a unit root]
```

```
> x= arima.sim(n = 1000, list(ma = c(-0.2, 0.2)),sd=1)
   + 0.1*(c(1:1000)^3)
> adf.test(x)

        Augmented Dickey-Fuller Test

data:  x
Dickey-Fuller = -0.038, Lag order = 9, p-value = 0.99
alternative hypothesis: stationary

Warning message:
In adf.test(x) : p-value greater than printed p-value
```

KPSS test

Kwiatkowski, Phillips, Schmidt, Shin (1992)

observed time series x_t is expressed as the sum
$x_t = T_t + RW_t + u_t$:

- deterministic trend T_t
- pure random walk RW_t (i.e. $I(1)$), $RW_t = RW_{t-1} + \epsilon_t$
- stationary error term u_t (i.e. $I(0)$)

H$_0$: $\sigma_\epsilon^2 = 0$

observable time series is stationary around a deterministic trend, x_t is $I(0)$

H$_1$: $\sigma_\epsilon^2 > 0$

in R:

```
>   s<-rnorm(1000)
>   kpss.test(s)
>
>   KPSS Test for Level Stationarity
>
>   data:  s
>   KPSS Level = 0.0429, Truncation lag parameter = 7,
>   p-value = 0.1
```

... stationarity cannot be rejected here ...

Second order Stochastic Difference Equation and Business Cycles

Samuelson (1939) applied only second order difference equations to explain the business cycles:

Investment I_t is dependent on the changes in income rather then the level of income

$$I_t = b_1(Y_{t-1} - Y_{t-2})$$

Consumption C_t depends on the past income by the first order nonhomogeneous equation

$$C_t = b_2 Y_{t-1} + b_3$$

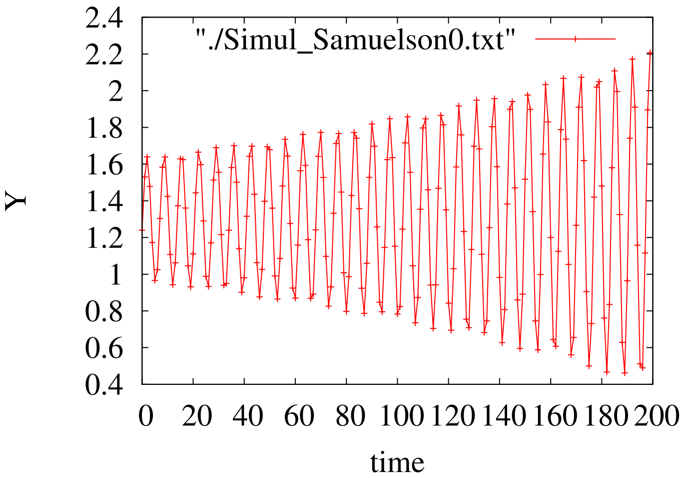
Production/consumption balance with error term

$$Y_t = C_t + I_t + \epsilon_t$$

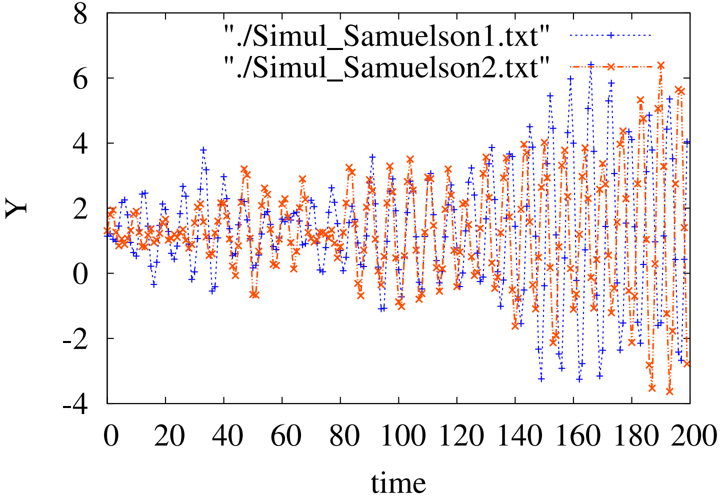
Consequence - accelerator principle

$$Y_t = b_1(Y_{t-1} - Y_{t-2}) + b_2 Y_{t-1} + b_3 + \epsilon_t$$

```
simulation in R

b3<-1.04   b2<-0.2     b1<-1.01
Y1<-1.0    Y2<-1.04    Y0<-1
for(i in 1:200){
      Y2<-Y1
      Y1<-Y0
      Y0<-(b1*(Y1-Y2)+b2*Y1+b3)
      print(paste(Y0))
               }
```




```
b3<-1.04   b2<-0.2   b1<-1.01   Y1<-1.0   Y2<-1.04   Y0<-1
for(i in 1:200){
    Y2<-Y1
    Y1<-Y0
    Y0<-(b1*(Y1-Y2)+b2*Y1+b3) + rnorm(1,mean=0,sd=0.3)
            }
```



"./Simul_Samuelson1.txt"
"./Simul_Samuelson2.txt"
Y
time
8
6
4
2
0
-2
-4
0
20
40
60
80
100
120
140
160
180
200

Solution of the homogeneous (non-stochastic) Samuelson's system

$$Y_t = z^t Y_0$$

$z = ...,\ b_3 = 0$

$$Y_t = b_1(Y_{t-1} - Y_{t-2}) + b_2 Y_{t-1}$$

$$z^t Y_0 = b_1(z^{t-1} Y_0 - z^{t-2} Y_0) + b_2 Y_0 z^{t-1}$$

$$1 = b_1 \left(\frac{1}{z} - \frac{1}{z^2} \right) + b_2 \frac{1}{z}$$

lag operator formally produces $\sim 1/z$

$$b_1 \left(\frac{1}{z} - \frac{1}{z^2} \right) + b_2 \frac{1}{z} - 1 = 0$$

$$\frac{b_1 + b_2}{z} - \frac{b_1}{z^2} - 1 = 0$$

$$z^2 - (b_1 + b_2)z + b_1 = 0$$

$$z_{1,2} = \frac{b_1 + b_2 \pm \sqrt{(b_1 + b_2)^2 - 4b_1}}{2}$$

$b_1 = 1.01,\ b_2 = 0.2,\ b_3 = 1.04$

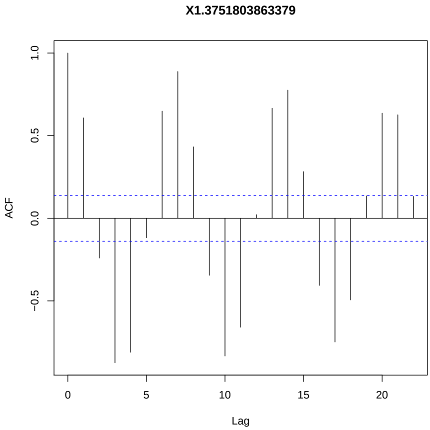
$$z_{1,2} = \frac{1.21 \pm \sqrt{1.21^2 - 4.04}}{2} = 0.605 \pm i0.8024$$

$|z_{1,2}| = 1.00492,\ \arccos(0.605/|z_{1,2}|) = 0.1465(2\pi)$

$$Y_t = Y_0 z^t = (1.00492)^t\ [\cos(\pm(2\pi)0.147t) + i\sin(\pm(2\pi)0.147t)]$$

$$= (1.00492)^t \exp(i2\pi 0.147t) = (1.00492)^t \exp\left(i2\pi\,\frac{t}{6.8}\right)$$

```
Qualitative check of the period
-------------------------------
x 1   0.966046199224
  2   1.02464804231704
  3   1.30411746998738
  4   1.58308761594452
  5   1.63837737060561
  6   1.42351812632883
  7   1.10769578854622
x 8   0.9425585965488
  9   1.06172315539237
 10   1.37270083551048
 11   1.62862762402139
 12   1.62421158120029
 13   1.36038211299075
 14   1.04560865970652
x15   0.931200544124225
```



Comparison Samuelson's accelerator ($b_3 = 0$) with ARIMA(p,d,.)

$$Y_t = b_1(Y_{t-1} - Y_{t-2}) + b_2 Y_{t-1}$$

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1-L)^d Y_t = 0$$

<u>Try $p = 1, d = 1$</u>

$$(1-\phi_1 L)(1-L)Y_t = (1-\phi_1 L)(Y_t - Y_{t-1})$$

$$= Y_t - Y_{t-1} - \phi_1(LY_t - LY_{t-1})$$

$$= Y_t - Y_{t-1} - \phi_1(Y_{t-1} - Y_{t-2})$$

<u>Try $p = 2, d = 1$</u>

$$(1-\phi_1-\phi_2 L)(1-L)Y_t = (1 - L - L\phi_1 + L^2\phi_1 - L^2\phi_2 + L^3\phi_2)$$

$$= Y_t - Y_{t-1} - \phi_1 Y_{t-1} + \phi_1 Y_{t-2} - \phi_2 Y_{t-2} + \phi_2 Y_{t-3})$$

Try $p = 2, d = 0$

$$(1 - \phi_1 L - \phi_2 L^2) Y_t = Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2}$$

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = 0$$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2}$$

$$Y_t = b_1(Y_{t-1} - Y_{t-2}) + b_2 Y_{t-1} = \underbrace{(b_1 + b_2)}_{\phi_1} Y_{t-1} + \underbrace{(-b_1)}_{\phi_2} Y_{t-2}$$

Conclusion: Model is transformable to ARIMA(2,0,q)
i.e. $d = 0$ model

Box Jenkins metodology

1. identification of model, mostly AR(1)
   - making sure that model is stationary
   - using plots of ACF, and PACF
2. estimation of the model parameters: say $x_t = 0.68x_{t-1} + \epsilon_t$, $\hat{\sigma} = 11.24$
   - using the maximum likelihood
   - non-linear least-squares
3. diagnostics of the model: prediction abilities, Ljung-Box test

Model identification by means of autocorrelation function:

	AR(p)	MA(q)	ARMA(p, q)
ρ_k ACF	k_0 does not exist ρ_k is U shaped	$k_0 = q$	k_0 does not exist ρ_k becomes U-shaped after first q -
ρ_{kk} PACF	$k_0 = p$	k_0 does not exist ρ_{kk} is limited by U-shaped curve	k_0 does not exist ρ_{kk} becomes U-shaped after the firs

Table: U denotes the curve of geometrically or sin-like decaying amplitudes.

Rule 1 If the series has positive autocorrelations out to a high number of lags, then it probably needs a higher order of differencing.

Rule 2 If the lag-1 autocorrelation is zero or negative, or the autocorrelations are all small a patternless, then the series does not need a higher order of differencing. If the lag-1 autocorrelation is -0.5 or more negative, the series may be overdifferenced. Beware of overdifferencing.

Rule 3 The optimal order of differencing is then often the order at which the standard deviation is lowest.

Rule 4 A model with no orders of differencing assumes that the original series is stationary.

Rule 5 A model with no orders of differencing normally includes a constant term

Rule 6 If the part of PACF of the differenciated series displays a sharp cutoff and/or the lag 1 autocorrelation is positive, i.e. the series appears slightly "underdifferenced" then consider adding one or more AR terms to the model

Rule 7 The autocorrelation function ACF of the differenced series displays a sharp cutoff and or the lag-1 autocorrelation function is negative .

Rule 8 It is possible for an AR term and MA term to cancel each other's effects, so if a mixed AR-MA model seems to fit data also try a model with one fewer AR term.

Rule 9 If there is a unit root in the AR part of the model the sum of AR coefficients is almost exactly 1

Rule 10 If the long-term forecasts appear erratic or unstable, there may be a unit root.

Rule 11 If the series has a strong and consistent seasonal pattern, then you should use an order of seasonal differencing.

Autoregression of simulated time series of the price formation process

Model of random process: $p_0,\ p_1,\ p_2,\ p_3,\ \ldots,\ p_t,\ \ldots$

$$\text{if } p_t > p_{\mathrm{fun}}\ \{\tilde{p}_{t+1} = p_t \times 0.99\}$$
$$\text{else} \quad \tilde{p}_{t+1} = p_t$$

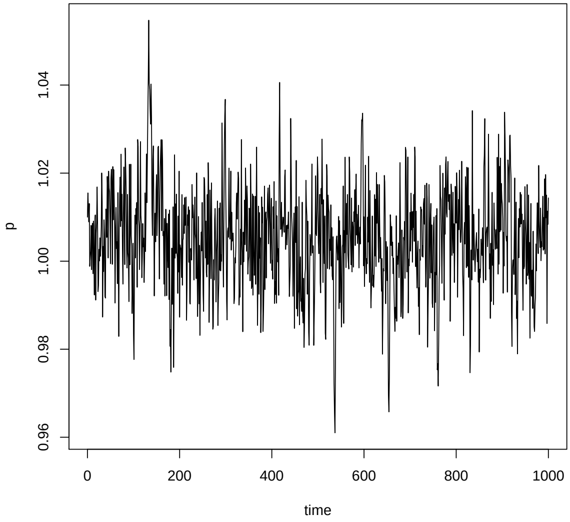
$$\text{if } \tilde{p}_{t+1} < p_{\mathrm{fun}}\ \{\underline{p}_{t+1} = \tilde{p}_{t+1} \times 1.01\}$$
$$\text{else} \quad \underline{p}_{t+1} = \tilde{p}_{t+1}$$

$$p_{t+1} = \underline{p}_{t+1} \exp[\ N(0, 0.01)\ ]$$

R project: simulation of price dynamics

```
pfun<-1             #  "fundamental" price

p<-(pfun+0.01)      #  initial price setting

vp<-p               #  auxilliary initial vector setting for pl

Ns<-1000            #  number of steps

time=c(0:Ns)        #  vector of times

for(i in 1:Ns){

if(p>pfun) p<-(p * 0.99)   # deterministic reduction

if(p<pfun) p<-(p * 1.01)   # deterministic increase

p<-(p*exp(rnorm(1,mean=0,sd=0.01))) #randomization
vp<-cbind(vp,p)
              }

vp<-as.vector(vp)
```

```
pdf("tp_simul.pdf")
plot(time,vp, type="l", pch=21, xlab="time", ylab="p")
dev.off()
```



Testing of simulated price series for stationarity

```
adf.test(vp)

Augmented Dickey-Fuller Test

data:  vp
Dickey-Fuller = -9.0344, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(vp) : p-value smaller than printed p-value

conclusion: reject unit root
```

Selection of the "best" arima model

based on the Aikake information criterion

The Akaike information criterion is a measure of the **relative goodness of fit** of a statistical model. It is grounded in the concept of information entropy, in effect offering a relative measure of the **information lost** when a given model is used to describe reality. In the general case, the AIC is $AIC = 2k - 2\ln(L)$, where k is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model. To apply AIC in practice, we start with a set of candidate models, and then find the models corresponding AIC values. There will almost always be information lost due to using one of the **candidate models** to represent the "true" model. We wish to select, from among n candidate models, the model that minimizes the estimated **information loss**. Denote the AIC values of the candidate models by AIC_1, AIC_2, AIC_3, $\ldots$, AIC_n. Let $AIC_{\min}$ be the minimum of those values. Then $e^{(AIC_{\min} - AIC_i)/2}$ can be interpreted as the relative probability that the i-th model minimizes the (estimated) information loss.

```
##########################################################################
Fit an ARIMA model to a univariate time series.
Usage
arima(x, order = c(0, 0, 0),
      seasonal = list(order = c(0, 0, 0), period = NA),
      ....
      optim.control = list(), kappa = 1e6)
```

```
# Example: R code for selection of the "best" model
# Aikake's information criterion, select among 10 candidate models

a1=arima(vp,order=c(0,0,0))
a2=arima(vp,order=c(0,0,1))
a3=arima(vp,order=c(0,1,0))
a4=arima(vp,order=c(1,0,0))
a5=arima(vp,order=c(1,1,0))
a6=arima(vp,order=c(0,1,1))
a7=arima(vp,order=c(1,0,1))
a8=arima(vp,order=c(1,1,1))
a8=arima(vp,order=c(0,0,2))
a9=arima(vp,order=c(0,2,0))
a10=arima(vp,order=c(2,0,0))

veca=as.vector(cbind(a1$aic,a2$aic,a3$aic,a4$aic,
               a5$aic,a6$aic,a7$aic,a8$aic,a9$aic,a10$aic))

which.min(veca)
```

Model Outputs:

```
> which.min(veca)
[1] 4
```

Check the best:

```
>  a4
Call:
arima(x = vp, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.3941     1.0042
s.e.  0.0290     0.0005

sigma^2 estimated as 0.0001101:  log likelihood = 3140.99,
aic = -6275.98
```

!!! intercept indicates fundamental price p_fun=1

```
### Check another one = not the best
> a3
Call:
arima(x = vp, order = c(0, 1, 0))

sigma^2 estimated as 0.0001581:
log likelihood = 2957.18,
aic = -5912.36
```

Forecast by Arima

```
predict.Arima {stats} R Documentation

Forecast from ARIMA fits
Description: Forecast from models fitted by arima.

predict(object, n.ahead = 1, newxreg = NULL,
        se.fit = TRUE, ...)
Arguments:
object    The result of an arima fit.

n.ahead   The number of steps ahead for
          which prediction is required.
```

Make the prediction:

```
> a4=arima(x=vp, order=c(1,0,0))
> predict(a4,n.ahead=6)
\$pred
Time Series:
Start = 1002
End = 1007
Frequency = 1
[1] 1.005989 1.004901 1.004472 1.004303 1.004236
1.004210

\$se
Time Series:
Start = 1002
End = 1007
Frequency = 1
[1] 0.01049478 0.01128027 0.01139739 0.01141547
0.01141828 0.01141871
```

Testing of forecast

```
  vp1=vp[1:990]
> a4=arima(vp1,order=c(1,0,0))
> vp2=vp[991:1000]
> pre=predict(a4,n.ahead=10)

Time Series:
Start = 991
End = 1000
Frequency = 1
 [1] 1.002855 1.003630 1.003936 1.004057 1.004104 1.004123 1.004131 1.004133
 [9] 1.004135 1.004135

\$se
Time Series:
Start = 991
End = 1000
Frequency = 1

 [1] 0.01048887 0.01127568 0.01139327 0.01141146 0.01141429 0.01141473
 [7] 0.01141480 0.01141481 0.01141481 0.01141481

> vp2
 [1] 1.0091107 0.9884819 0.9970753 1.0119451 0.9983571 1.0193618 1.0258751
 [8] 1.0078596 1.0182782 1.0172228
```

```
vpred=pre\$pred[1:10]

> vp2-vpred
 [1]  0.006256168 -0.015148120 -0.006860733  0.007888411 -0.005747281
 [6]  0.015238707  0.021744591  0.003726142  0.014143558  0.013087774

compare errors with a mean standard deviation

(vp2-vpred)/mean(se1)
 [1]  0.5533614 -1.3398593 -0.6068355  0.6977342 -0.5083501  1.3478718
 [7]  1.9233207  0.3295793  1.2510053  1.1576206
```