Econometrics II – Chapter 7.3, Heij et al.

Linear Time Series Models for NonStationary data

Marius Ooms

Tinbergen Institute Amsterdam

Contents

- Modelling nonstationarity, decomposition
- Deterministic trend and Stochastic trend
- ARIMA model, I(1) processes
- Unit root tests
- Prediction with ARIMA model

Transforming to Stationarity, Trends and Seasons

Transforming non-stationary series to stationarity: Modelling Trends and Seasonality.
The concept of **stationarity is crucial** because when a series is nonstationary,

1. Mean, variance, covariance, correlation and partial correlation lose their meaning,
2. Important identification and estimation methods do not work
3. Standard asymptotic results for statistical inference do not apply to $\widehat{\phi}$. Standars CLT is not applicable: No asymptotic normality, no $\sqrt{n}$ convergence.

Nonstationarity and Decomposition of time series

We distinguish two types of non-stationarity: **deterministic nonstationarity**, and **stochastic**, random-walk-type **nonstationarity**. Both deterministic and stochastic nonstationarity can apply to the **trends** and/or **seasonal components**.

Additive **decomposition of a time series** into stationary and nonstationary components:

$$y_t = T_t + S_t + R_t$$

T_t: trend(-cycle), S_t: seasonal component, R_t: stationary component. Mean (and variance) of T_t and S_t evolve over time. Convention: $E(S_t + S_{t+1} + S_{t+2} + S_{t+3}) = 0$ for quarterly data.

Use log transformation to get multiplicative decomposition.

Deterministic Trend (DT)

The simplest deterministic trend model is the **linear time trend**:

$$y_t = \alpha + \beta t + \varepsilon_t,$$

so that mean growth β is derived from

$$\Delta y_t = \beta + \varepsilon_t - \varepsilon_{t-1}$$

and h-step prediction $y_{t+h} : \hat{y}_{n+h} = a + b(n+h)$ and

$$E[(y_{n+h} - \hat{y}_{n+h})^2] \sim \sigma_\varepsilon^2, h \to \infty, h/n \to 0$$

Prediction interval stays finite for $h \to \infty$

Example: DT dLog Dow Jones

Log Dow Jones Index Industrials Weekly 1896-2004

Dependent Variable: LOGDJIND
Method: Least Squares
Date: 02/25/04 Time: 17:40
Sample: 11/04/1896 1/27/2004
Included observations: 5596

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.435230	0.012212	281.3038	0.0000
@TREND(11/04/1896)	0.000885	3.78E-06	234.1009	0.0000

R-squared	0.907380	Mean dependent var	5.910911
Adjusted R-squared	0.907363	S.D. dependent var	1.500914
S.E. of regression	0.456822	Akaike info criterion	1.271313
Sum squared resid	1167.393	Schwarz criterion	1.273683
Log likelihood	-3555.134	F-statistic	54803.25
Durbin-Watson stat	0.003266	Prob(F-statistic)	0.000000

Stochastic Trend

The simplest stochastic trend model is the **random walk model (with drift)**:

$$\Delta y_t = \alpha + \varepsilon_t$$

so that

$$y_t = y_1 + \alpha(t-1) + \sum_{s=2}^{t} \varepsilon_s,$$

but

$$E[(y_{n+h} - \hat{y}_{n+h})^2] \sim h\sigma_\varepsilon^2, h \to \infty, h/n \to 0$$

Prediction interval becomes infinitely large at rate $h^{1/2}$. (Compare the root h-law often applied for standard deviations of predictions of logs of stock prices).

Example: RW + drift Log Dow Jones

Returns (dLog) Dow Jones Index Industrials Weekly 1896-2004

Dependent Variable: D(LOGDJIND)
Method: Least Squares
Date: 02/25/04 Time: 18:07
Sample(adjusted): 11/11/1896 1/27/2004
Included observations: 5595 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000993	0.000349	2.845951	0.0044

R-squared	0.000000	Mean dependent var	0.000993
Adjusted R-squared	0.000000	S.D. dependent var	0.026106
S.E. of regression	0.026106	Akaike info criterion	-4.453093
Sum squared resid	3.812560	Schwarz criterion	-4.451908
Log likelihood	12458.53	Durbin-Watson stat	1.957510

Unifying Deterministic and Stochastic Trends

Consider the deterministic trend plus AR(1):

DT:	$y_t = \mu_1 + \mu_2 t + u_t$	$u_t \sim$ stationary AR, reverting to trend
ST:	$y_t = \mu_1 + \mu_2 t + u_t$	$u_t \sim$ RW, AR process with a 'unit root'

In both cases Δy_t is stationary whereas y_t is trending.
The crucial difference is the role of ε_t in AR ($u_t = \phi u_{t-1} + \varepsilon_t$) or RW ($\Delta u_t = \varepsilon_t$). The models coincide when $\phi = 1$.

Differences between ST and DT

* In the DT case, the influence of ε_t on y_{t+k} dies away as k increases and the series **reverts to the trend line** after a shock. In the ST case, the influence of ε_t persists for any $k > 0$: No 'reversion to trend'.
* In practice the difference is subtle, diagnostics based on time series plots, correlograms of y_t and Δy_t and (partial) scatterplots of $y_t, \Delta y_t$, $y_{t-1}, \Delta y_{t-1}$ are **not** sufficient.
* Information from S(P)ACF: r_k for y_t does not decrease exponentially for DT or ST, only for Δy_t. Moreover for DT only: $\sum_1^\infty r_k \approx -0.5$ for Δy_t: induced MA "unit root", cf. ACF MA(1) and *Exercise (1)*: Exercise 7.13.
* Nelson and Plosser (1982, JME, 139-162) suggest many US macro time series to be of the ST type using statistical **unit root tests** developed by Dickey and Fuller.

Det. Trend + AR(1) Log Dow Jones

Returns (dLog) Dow Jones Index Industrials Weekly 1896-2004

Dependent Variable: LOGDJIND
Method: Least Squares
Date: 02/25/04 Time: 17:41
Sample(adjusted): 11/11/1896 1/27/2004
Included observations: 5595 after adjusting endpoints
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.924902	0.729835	4.007623	0.0001
@TREND(11/04/1896)	0.001052	0.000188	5.584583	0.0000
AR(1)	0.998668	0.000764	1306.829	0.0000

R-squared	0.999698	Mean dependent var	5.911304
Adjusted R-squared	0.999698	S.D. dependent var	1.500760
S.E. of regression	0.026102	Akaike info criterion	-4.453111
Sum squared resid	3.809768	Schwarz criterion	-4.449556
Log likelihood	12460.58	F-statistic	9243842.
Durbin-Watson stat	1.956337	Prob(F-statistic)	0.000000

Inverted AR Roots	1.00

Extensions of Stochastic Trend, I(1) processes

ARIMA(p, d, q) model:

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

The “purely stochastic” process y_t is **Integrated of order d**, $I(d)$ if it requires differencing d times to become a stationary and invertible ARMA process. An I(1) process requires differencing once and is called **difference stationary**. A process that is differenced too many times is called **overdifferenced**: I(-1).

Unit root Tests, algebra

"Tests for an AR unit root ($\phi = 1$)" provide a "formal criterion" to choose between stochastic trends and deterministic trends in the context of AR models. Consider the general "first order" trend model

$$y_t = \mu_1 + \mu_2 t + u_t, \quad u_t = \phi u_{t-1} + \varepsilon_t.$$

Exercise (2): Show that we can rewrite this general trend model as

$$y_t = [\mu_1(1-\phi) + \phi\mu_2] + \mu_2(1-\phi)t + \phi y_{t-1} + \varepsilon_t,$$

Dickey-Fuller test regression

By introducing $\rho = \phi - 1$ and subtracting y_{t-1} :
Dickey-Fuller test regression form:

$$\Delta y_t = [-\mu_1 \rho + (\rho + 1)\mu_2] - \mu_2 \rho t + \rho y_{t-1} + \varepsilon_t.$$

Note $\phi(1) = \rho = 0$ for ST. So, to test for ST, we look at the coefficient of y_{t-1}.

Interpretation Dickey-Fuller test regression:

Main Idea: Test for (negative) partial correlation between growth rates and lagged levels. This idea can be extended to AR(p) models with $p > 1$.

Exercise (3): derive the "Augmented Dickey-Fuller" test regression when u_t follows an AR(2), using the "unit root" factorisation of $\phi(z)$: $\phi(z) = \phi(1)z + \rho(z)(1 - z)$, cf. Heij et al. p. 598.

Unit Root test for trending data, practice

Unit root testing using Dickey-Fuller regression for trending data:

H_0	:	$\phi = 1 \Leftrightarrow \rho = 0$ ST
H_1	:	$-1 < \phi < 1 \Leftrightarrow -2 < \rho < 0$ DT
test-statistic	:	$t : \rho = 0$ or $F : \mu_2\rho = 0 \wedge \rho = 0$
critical value t	:	$\tau_{ct}(\alpha, n), \tau_{ct}(.05, \infty) = -3.4,$ Exhibit 7.16
critical value F	:	$F_{ct}(0.05, 100) = 6.5,$ Exibit 7.16
conclusion	:	reject ST against DT when $t < \tau_{ct}(\alpha, n)$
Power	:	t asymptotically normal under $H_1,$ and $P(t < \tau(\alpha)\|H_1) \to 1, n \to \infty$: consistency Inevitable: power low if ϕ close to 1 and n small

Example: DF test trending Log Dow Jones

Augmented Dickey-Fuller Unit Root Test on LOGDJIND

Null Hypothesis: LOGDJIND has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.899565	0.6546
Test critical values:	1% level	-3.959678	
	5% level	-3.410608	
	10% level	-3.127081	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGDJIND)
Method: Least Squares
Date: 02/25/04 Time: 17:45
Sample(adjusted): 11/25/1896 1/27/2004
Included observations: 5593 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGDJIND(-1)	-0.001452	0.000764	-1.899565	0.0575
D(LOGDJIND(-1))	0.020812	0.013367	1.556905	0.1195
D(LOGDJIND(-2))	0.044448	0.013367	3.325219	0.0009
C	0.005362	0.002716	1.974263	0.0484
@TREND(11/04/1896)	1.49E-06	7.10E-07	2.092154	0.0365

R-squared	0.003154	Mean dependent var	0.001001
Adjusted R-squared	0.002441	S.D. dependent var	0.026108
S.E. of regression	0.026076	Akaike info criterion	-4.454697
Sum squared resid	3.799650	Schwarz criterion	-4.448771
Log likelihood	12462.56	F-statistic	4.420447
Durbin-Watson stat	2.001370	Prob(F-statistic)	0.001440

Dickey-Fuller test for data without trend

Dickey-Fuller test for data without trend

In practice there might be theoretical reasons to exclude the possibility of a drift in y_t.

If $\mu_2 = 0$ both under H_0 (no drift) and H_1 (mean reversion), one should omit the trend in D-F test regression and apply τ_c instead of τ_{ct} for critical values, to increase power of test. See also table 7.6.

Another application of a D-F test regression without trend in the test of I(2) vs. I(1), where $\Delta\Delta y_t$ is the dependent variable and we do not expect a trend in Δy_t.

(A)DF test nontrending Returns Dow Jones

Augmented Dickey-Fuller Unit Root Test on D(LOGDJIND)

Null Hypothesis: D(LOGDJIND) has a unit root
Exogenous: Constant
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-50.07399	0.0001
Test critical values:	1% level	-3.431339	
	5% level	-2.861862	
	10% level	-2.566984	

*MacKinnon (1996) one-sided p-values.

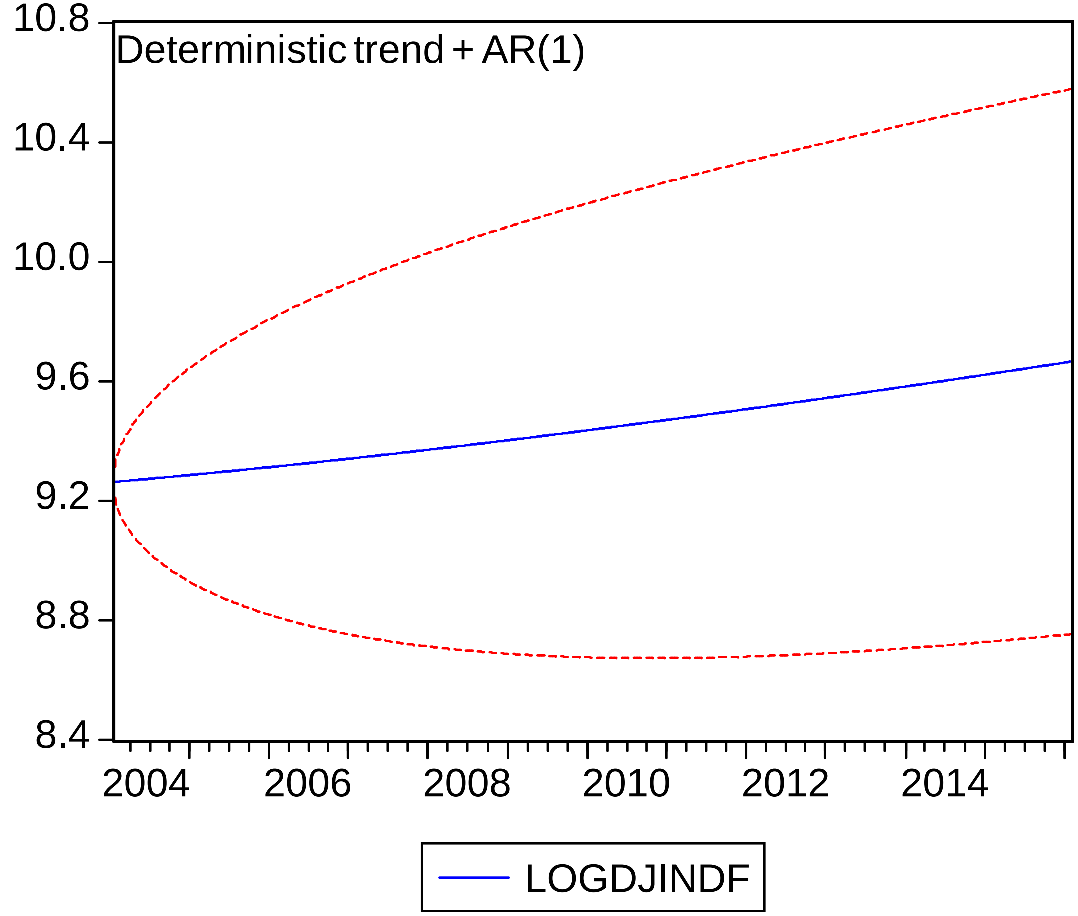
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGDJIND,2)
Method: Least Squares
Date: 02/25/04 Time: 17:44
Sample(adjusted): 11/25/1896 1/27/2004
Included observations: 5593 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LOGDJIND(-1))	-0.936123	0.018695	-50.07399	0.0000
D(LOGDJIND(-1),2)	-0.043745	0.013361	-3.274109	0.0011
C	0.000937	0.000349	2.684011	0.0073

R-squared	0.490468	Mean dependent var	4.53E-06
Adjusted R-squared	0.490286	S.D. dependent var	0.036532
S.E. of regression	0.026082	Akaike info criterion	-4.454612
Sum squared resid	3.802691	Schwarz criterion	-4.451057
Log likelihood	12460.32	F-statistic	2690.426
Durbin-Watson stat	2.001286	Prob(F-statistic)	0.000000

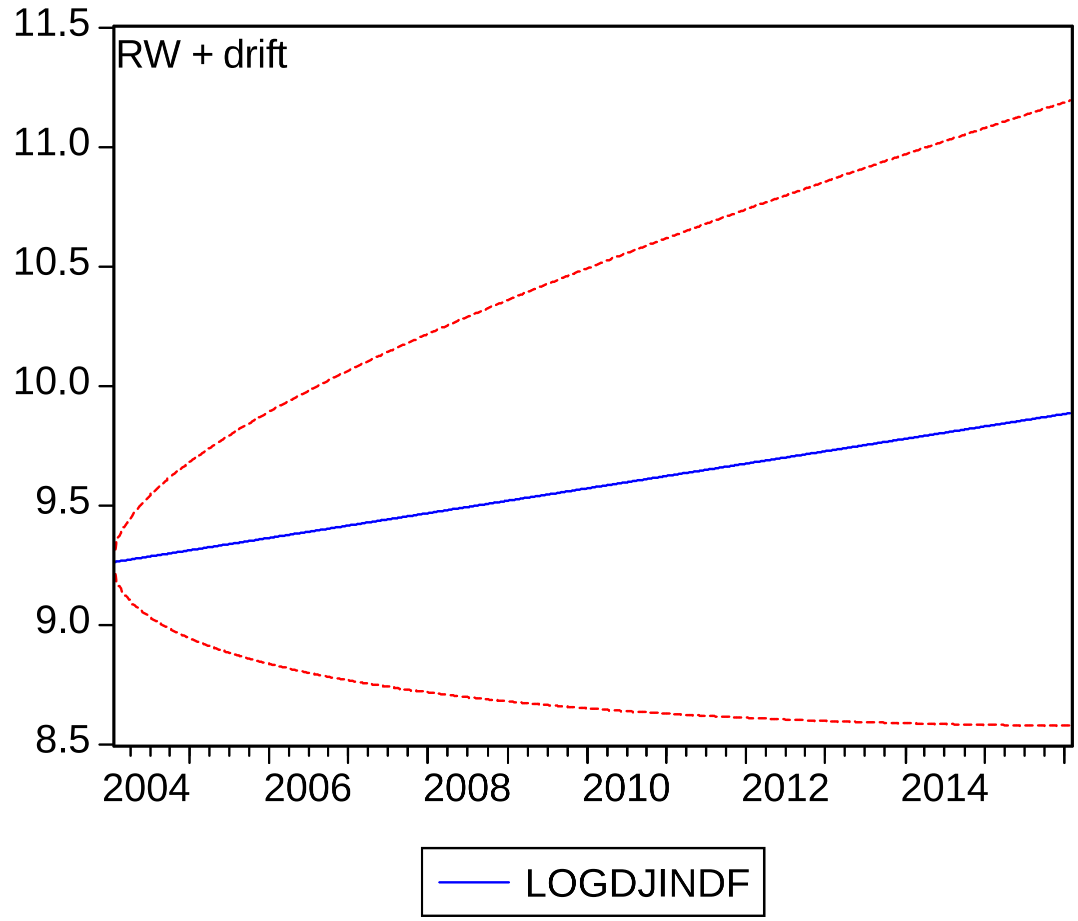
Prediction DT+AR(1) log Dow Jones

Log Dow Jones Index Weekly 1896-2004, forecast 2004-2016



Prediction ST log Dow Jones

Log Dow Jones Index Weekly 1896-2004, forecast 2004-2016



Prediction with an ARIMA model

In case y_t is a nonstationary ARIMA($p, 1, q$) model, implying $z_t = \Delta y_t$, is a stationary ARMA(p,q), we predict y_t as an (integrated) cumulated sum of forecasts for z_t, i.e partial sums of z_t. If y_t is ARIMA(1,1,0), then $z_t = \phi z_{t-1} + \varepsilon_t$ with MA(∞) form: $z_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots$, and $y_{n+1} = y_n + z_{n+1}$, $y_{n+2} = y_{n+1} + z_{n+2}$ etc.

Exercise (4): show optimal forecasts of y_t are given by:

$$\widehat{y}_{n+h} = y_n + (a_h - 1)\Delta y_n, \text{ with forecast error variance:}$$

$$\text{Var}(e_{n+h}) = \sigma_\varepsilon^2 \left[1 + \sum_{j=1}^{h-1} a_{h-j}^2 \right].$$

where $a_h = 1 + \sum_{i=1}^{h} \psi_i = \frac{(1-\phi^{h+1})}{(1-\phi)}$.

Prediction error variance ARIMA(1,1,0)

In this case, as h increases, the variance increases monotonically with h. The forecasts keep becoming more imprecise as h increases. $\mathbb{V}ar(e_{n+h})$ eventually grows proportionally to h and s.e.(e_{n+h}) grows proportionally to $h^{1/2}$, a **square-root law** often applied for standard deviations of predictions of logs of stock prices.
This is another **key difference** between **I(0)** and **I(1)** processes.

Summary interval prediction ARIMA($p, 1, q$)

Multistep interval prediction:

1. derive first $h-1$ coeffients ψ_j in MA(∞) representation of $(1-L)y_t$, $\psi(L) = \phi(L)^{-1}\theta(L)$
2. construct partial sum series of ψ_j: $a_j = 1 + \sum_{i=1}^{j} \psi_i$, ($a(L) = (1-L)^{-1}\psi(L)$)
3. Var($e_{n+h}) = \sigma_\varepsilon^2(1 + \sum_{j=1}^{h-1} a_j^2)$

Variance of long run prediction increases linearly in h, for large h.

ARIMA estimation in Eviews

NB: Eviews automatically deletes first p observations at the beginning of the sample. Exact ML is not possible.
There are two ways to estimate ARMA(p,q) model. Beware of the interpretation of the constant term.
Example:

```
ls y c y(-1) ’ c constant in regression (7.17)
ls y c AR(1) ’ c mean of y
```

To estimate ARIMA model use “Auto-Series”: an Eviews-expression in place of a series. This allows forecasting for “differenced” and “level” series using one menu.

ARIMA forecasting in Eviews

E.g. for $(1-L)^n(1-L^s)y_t$ use:

Example:

```
ls d(y,n,s) c MA(1) SMA(s) 'c mean of d(y,n,s)
forecast(d)  'multistep forecast of d(y,n,s)
forecast(u)  'multistep forecast of y
forecast(s)  'forecast regression part only
```

Note: Eviews results are unreliable if $\theta(z)=0$ contains roots near the unit circle.