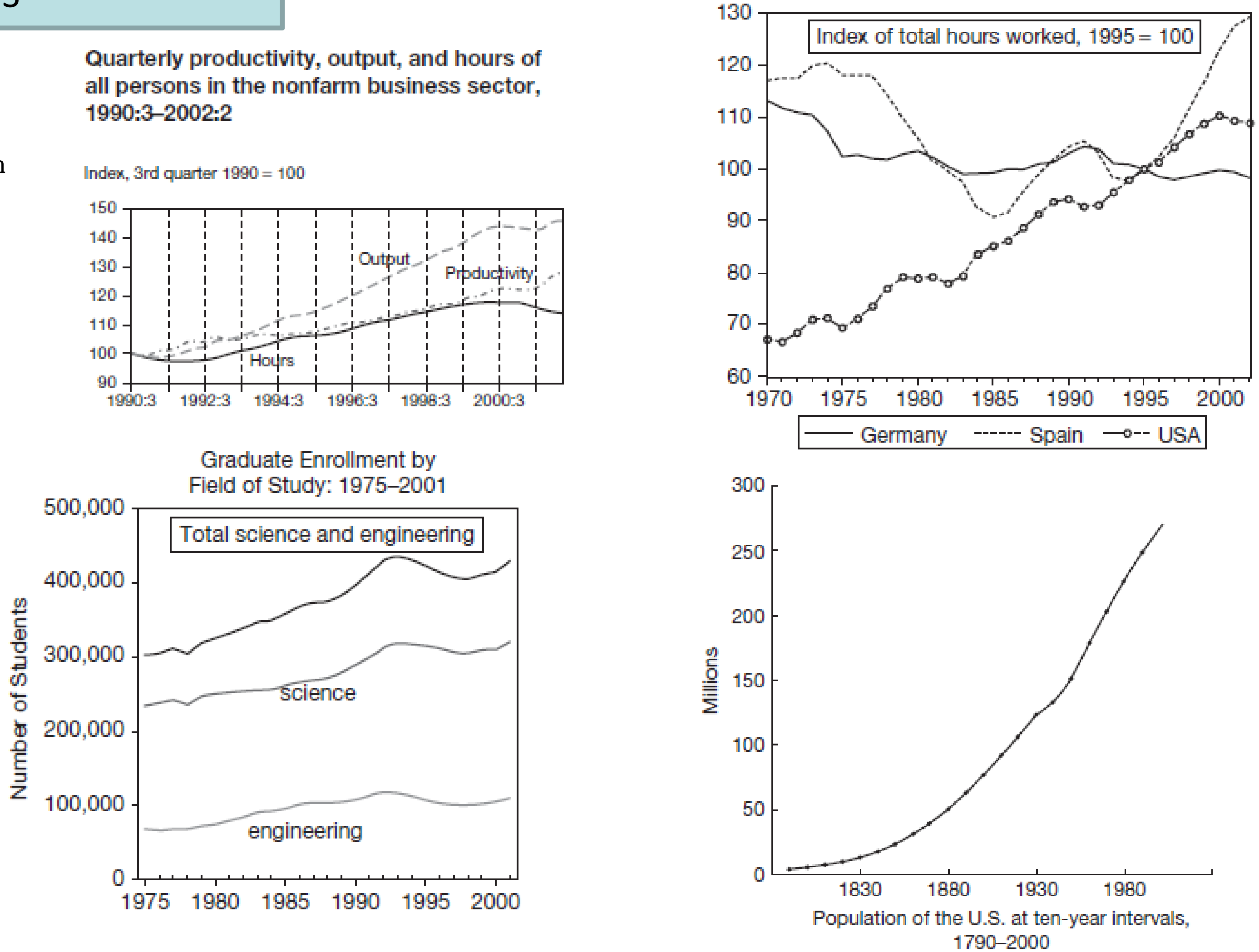
CHAPTER 10
FORECASTING THE LONG TERM: DETERMINISTIC AND STOCHASTIC TRENDS

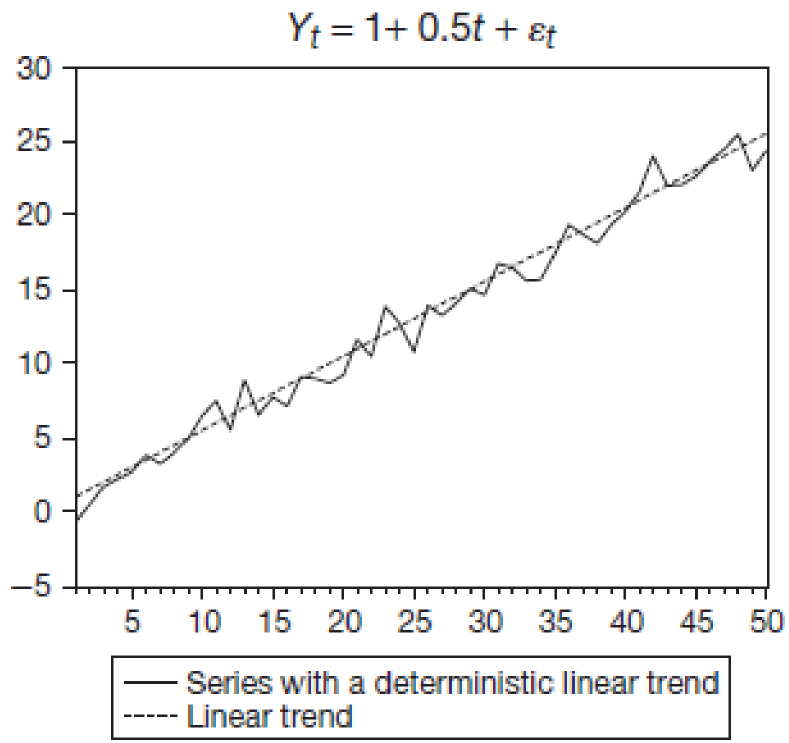
"A trend is a relatively smooth, mostly unidirectional, pattern in the data that arises from the accumulation of information over time."

Figure 10.1 Economic Time Series with Trends



10.1 Deterministic Trends

Figure 10.2 Time Series with Linear Deterministic Trend



10.1.1 Trend Shapes

Figure 10.3 Common Deterministic Trends

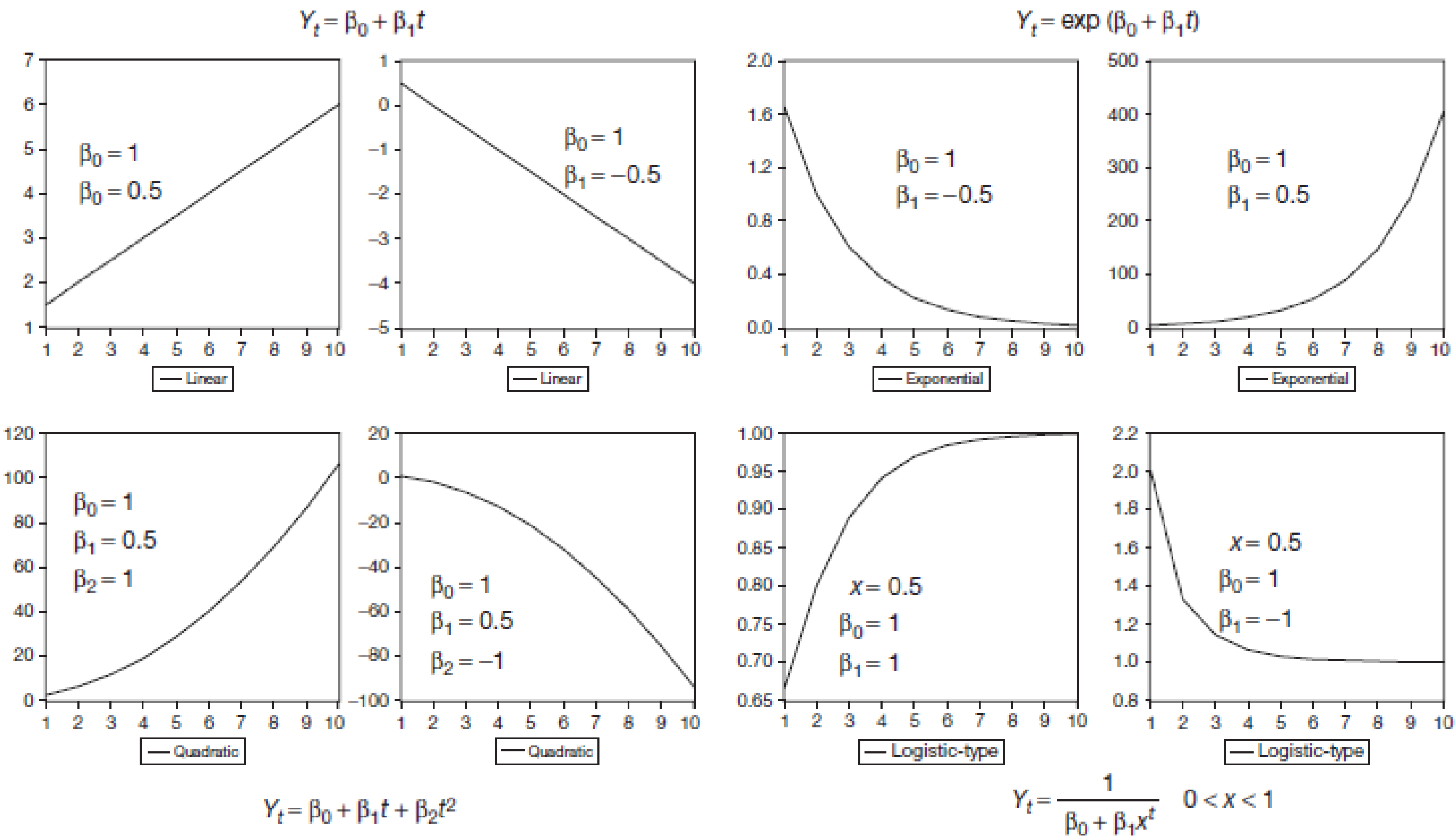
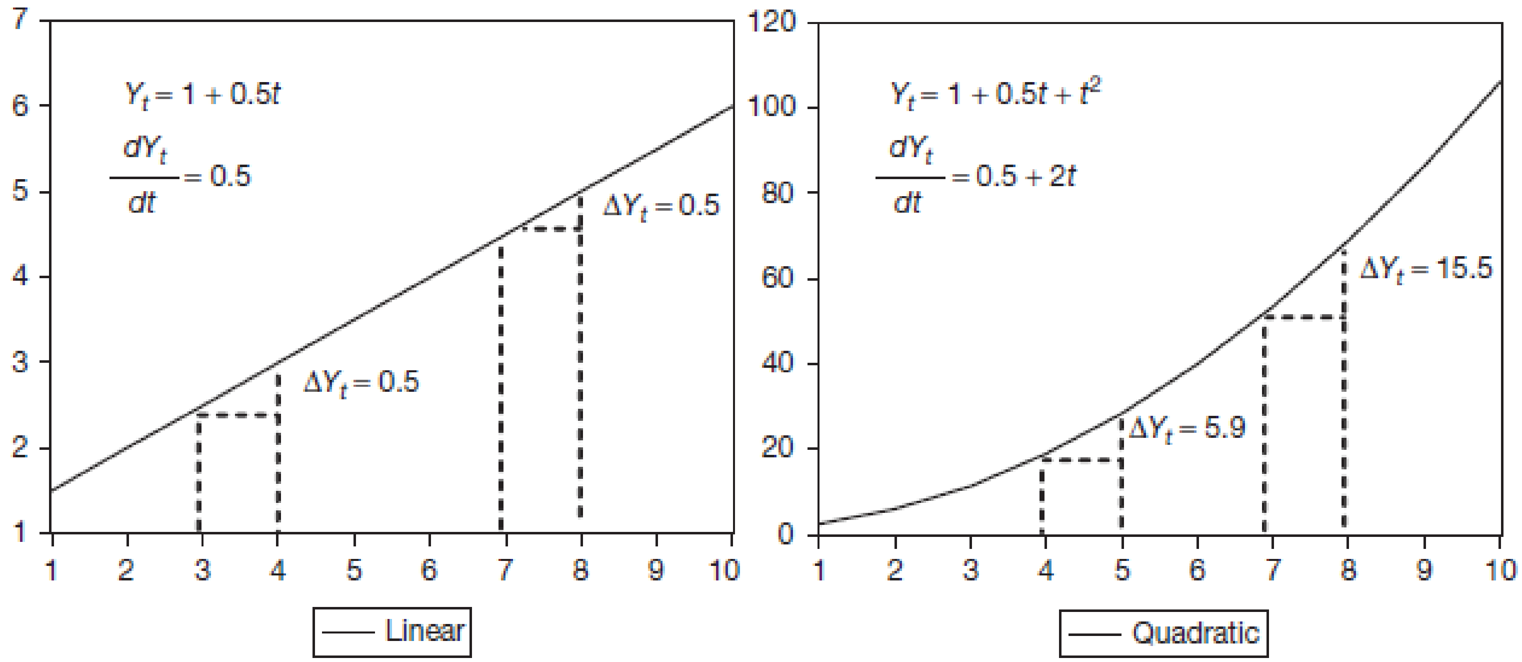


Figure 10.4 Growth Rate with Linear and Quadratic Deterministic Trends



Geometric growth: $P_t = P_0 (1 + r)^t$

Exponential growth: $P_t = P_0 e^{r^*t}$

$(r \approx r^*)$

For a simple linear **trend** series with uncorrelated noise:

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

Unconditional mean

$$E(Y_t) = E(\beta_0 + \beta_1 t + \varepsilon_t) = \beta_0 + \beta_1 t + E(\varepsilon_t) = \beta_0 + \beta_1 t \equiv \mu_t$$

Unconditional variance

$$\sigma_Y^2 = E(Y_t - \mu_t)^2 = E(Y_t - \beta_0 - \beta_1 t)^2 = E(\varepsilon_t)^2 = \sigma_\varepsilon^2 \equiv \gamma_0$$

Autocovariance of order k

$$\gamma_k \equiv E(Y_t - \mu_t)(Y_{t-k} - \mu_{t-k}) = E(Y_t - \beta_0 - \beta_1 t)(Y_{t-k} - \beta_0 - \beta_1(t-k))$$

$$= E(\varepsilon_t \varepsilon_{t-k}) = 0$$

Autocorrelation of order k

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{0}{\sigma_\varepsilon^2} = 0$$

In this example, we see why these models are called **trend stationary**: because they are stationary around a deterministic trend

10.1.3 Optimal Forecasting

Figure 10.5 Home Mortgage Outstanding Debt (Billions of Dollars)

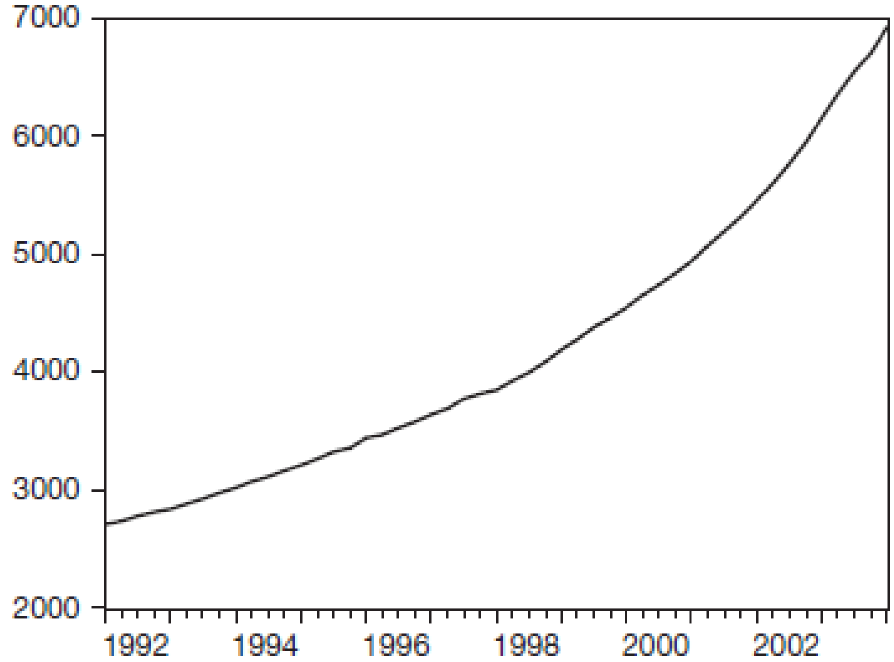


Table 10.1 Deterministic Trend Specifications

Trend model	Adjusted R-squared	AIC	SIC
$Y_t = \beta_0 + \beta_1 t$	0.9358	14.292	14.369
$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2$	0.9945	11.847	11.963
$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$	0.9994	9.520	9.674
$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$	0.9997	**8.827**	**9.020**
$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \beta_5 t^5$	0.9997	8.861	9.093
$Y_t = \beta_0 \exp(\beta_1 t + \beta_2 t^2)$		10.382	10.498
$Y_t = \beta_0 \exp(\beta_1 t + \beta_2 t^2 + \beta_3 t^3)$		9.933	10.087
$Y_t = \beta_0 \exp(\beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4)$		9.114	9.307
$Y_t = \beta_0 \exp(\beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \beta_5 t^5)$		8.855	9.087
$\log Y_t = \beta_0 + \beta_1 t + \beta_2 t^2$	0.9986	-6.343	-6.227
$\log Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$	0.9997	-7.874	-7.719
$\log Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$	0.9997	-7.906	-7.713
$\log Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \beta_5 t^5$	0.9997	-7.974	**-7.743**
$\log Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \beta_5 t^5 + \beta_6 t^6$	0.9997	**-7.977**	-7.707

The best fit may not translate in best forecasts

We should apply the *parsimony principle* – *Occam's* (1287–1347) *razor*

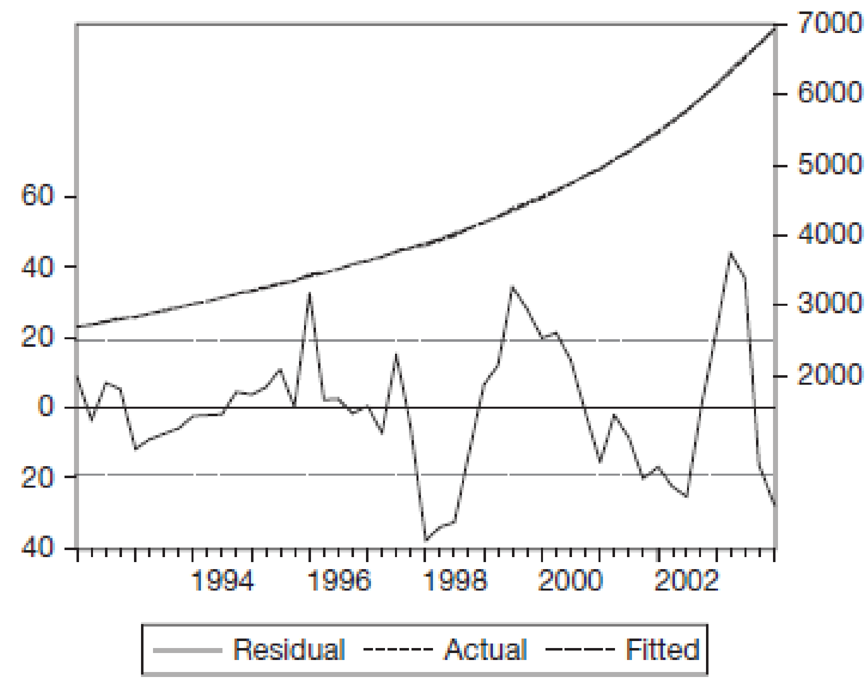
Ptolemy (90–168) stated, "We consider it a good principle to explain the phenomena by the simplest hypothesis possible."

Table 10.2 Least Squares Estimation of a Polynomial Trend

Dependent Variable: DEBT
Method: Least Squares
Sample: 1992:1 2004:1
Included observations: 49

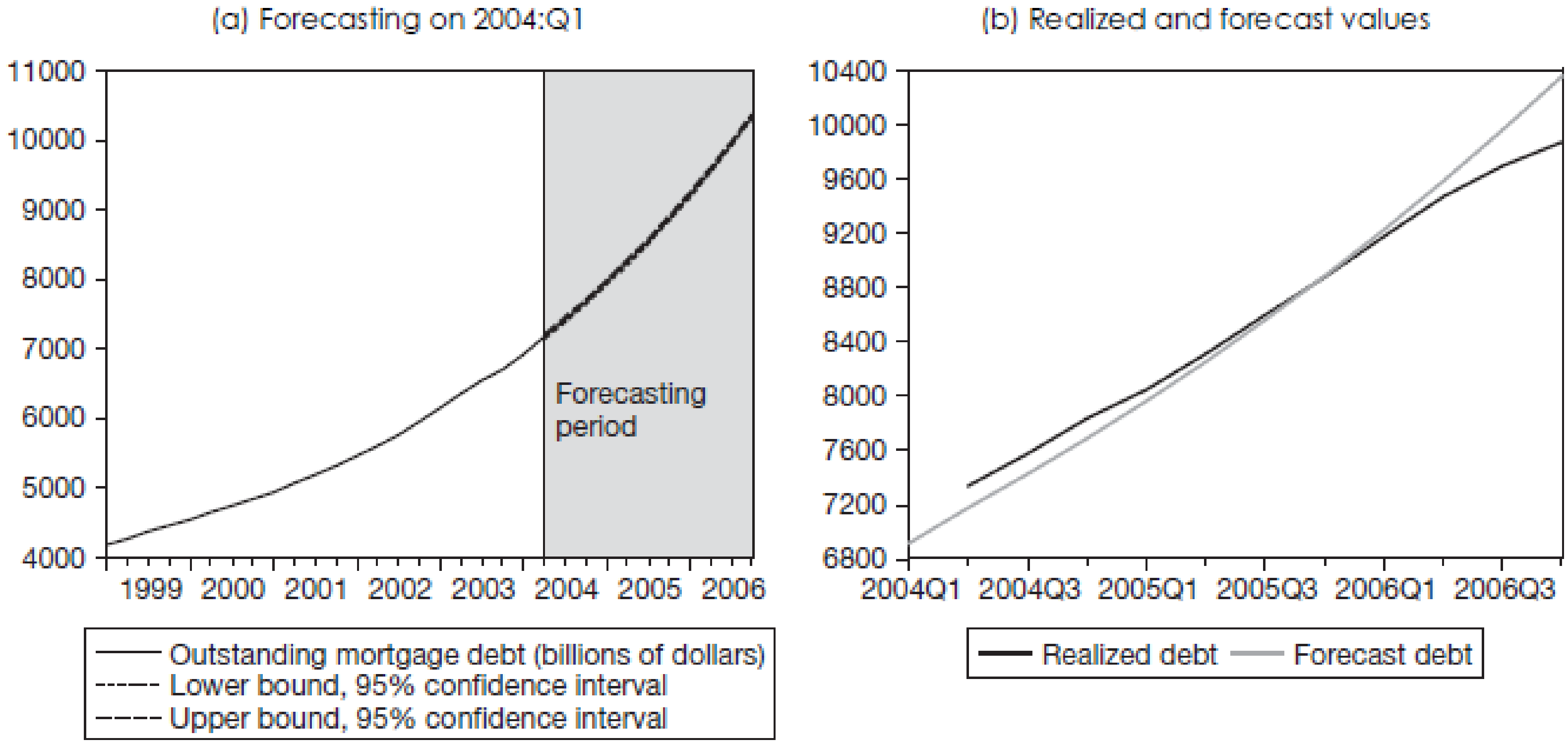
Variable	Coefficient	Std. Error	t–Statistic	Prob.
C	2668.571	15.45319	172.6874	0.0000
TRND	31.63280	4.190147	7.549329	0.0000
TRND^2	1.065196	0.335967	3.170541	0.0028
TRND^3	-0.032256	0.010048	-3.210235	0.0025
TRND^4	0.000688	9.97E-05	6.897960	0.0000

R-squared	0.999765	Mean dependent var	4184.486
Adjusted R-squared	0.999743	S.D. dependent var	1188.904
S.E. of regression	19.04364	Akaike info criterion	8.827795
Sum squared resid	15957.05	Schwarz criterion	9.020838
Log likelihood	-211.2810	F-statistic	46759.77
Durbin-Watson stat	0.731080	Prob(F-statistic)	0.000000



Durbin-Watson: $d = 0.73 < 1 \approx d_{L,\alpha}$, we suspect residual autocorrelation

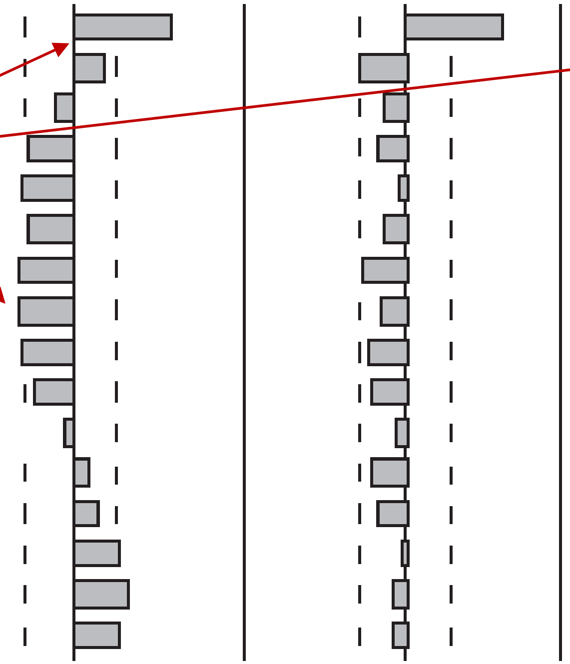
Figure 10.6 Forecast of Outstanding Mortgage Debt Based on Table 10.2



Recall: Durbin-Watson $d = 0.73 < 1 \approx d_{L,\alpha}$, we suspect residual autocorrelation

Figure 10.7 Correlograms of the Residuals of the Fourth Polynomial Trend Model

Sample: 1992:1 2004:1
Included observations: 49



Lag	AC	PAC	Q-Stat	Prob
1	0.608	0.608	19.270	0.000
2	0.197	-0.276	21.323	0.000
3	-0.084	-0.118	21.706	0.000
4	-0.268	-0.169	25.680	0.000
5	-0.300	-0.042	30.796	0.000
6	-0.274	-0.116	35.174	0.000
7	-0.323	-0.270	41.366	0.000
8	-0.326	-0.152	47.862	0.000
9	-0.304	-0.223	53.620	0.000
10	-0.244	-0.210	57.425	0.000
11	-0.049	-0.045	57.583	0.000
12	0.096	-0.204	58.210	0.000
13	0.177	-0.161	60.393	0.000
14	0.296	-0.004	66.666	0.000
15	0.338	-0.063	75.081	0.000
16	0.298	-0.064	81.792	0.000

Suspicion confirmed

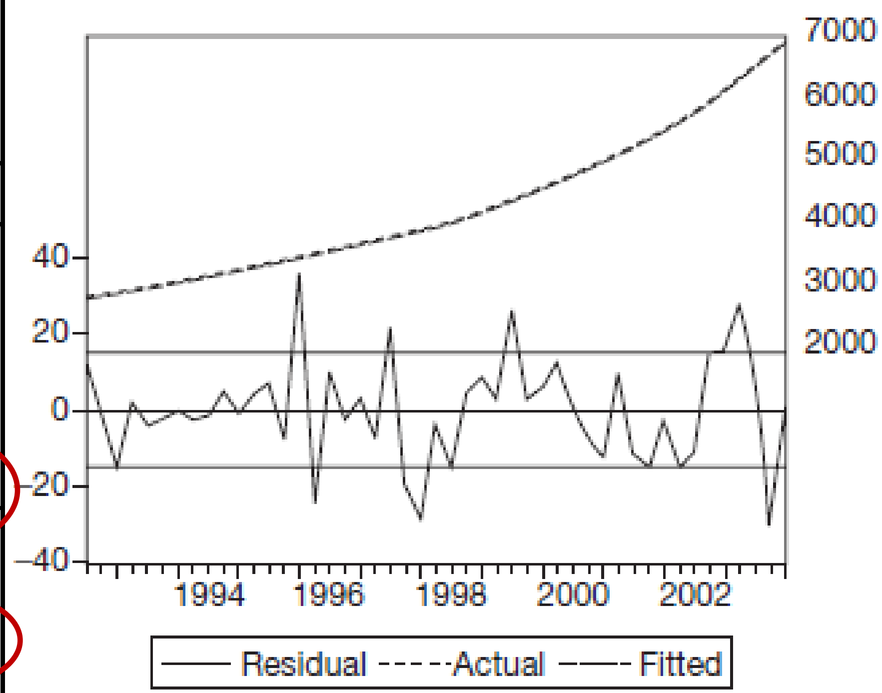
Table 10.3 Least Squares Estimation of Trend and AR Model

Modelo misto é melhor

Dependent Variable: DEBT
Method: Least Squares
Sample(adjusted): 1992:3 2004:1
Included observations: 47 after adjusting endpoints
Estimation settings: tol = 0.00010, derivs = accurate mixed (linear)
Convergence achieved after 8 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2661.391	50.60255	52.59402	0.0000
TRND	33.90561	11.62180	2.917415	0.0058
TRND^2	0.870626	0.831794	1.046684	0.3015
TRND^3	-0.026140	0.023088	-1.132204	0.2643
TRND^4	0.000625	0.000217	2.874052	0.0065
AR(1)	0.826760	0.152925	5.406303	0.0000
AR(2)	-0.281622	0.161740	-1.741197	0.0893

R-squared	0.999863	Mean dependent var	4246.766
Adjusted R-squared	0.999842	S.D. dependent var	1173.817
S.E. of regression	14.73993	Akaike info criterion	8.355601
Sum squared resid	8690.616	Schwarz criterion	8.631155
Log likelihood	-189.3566	F-statistic	48613.50
Durbin-Watson stat	2.070406	Prob(F-statistic)	0.000000
Inverted AR Roots	.41-.33i	.41+.33i	

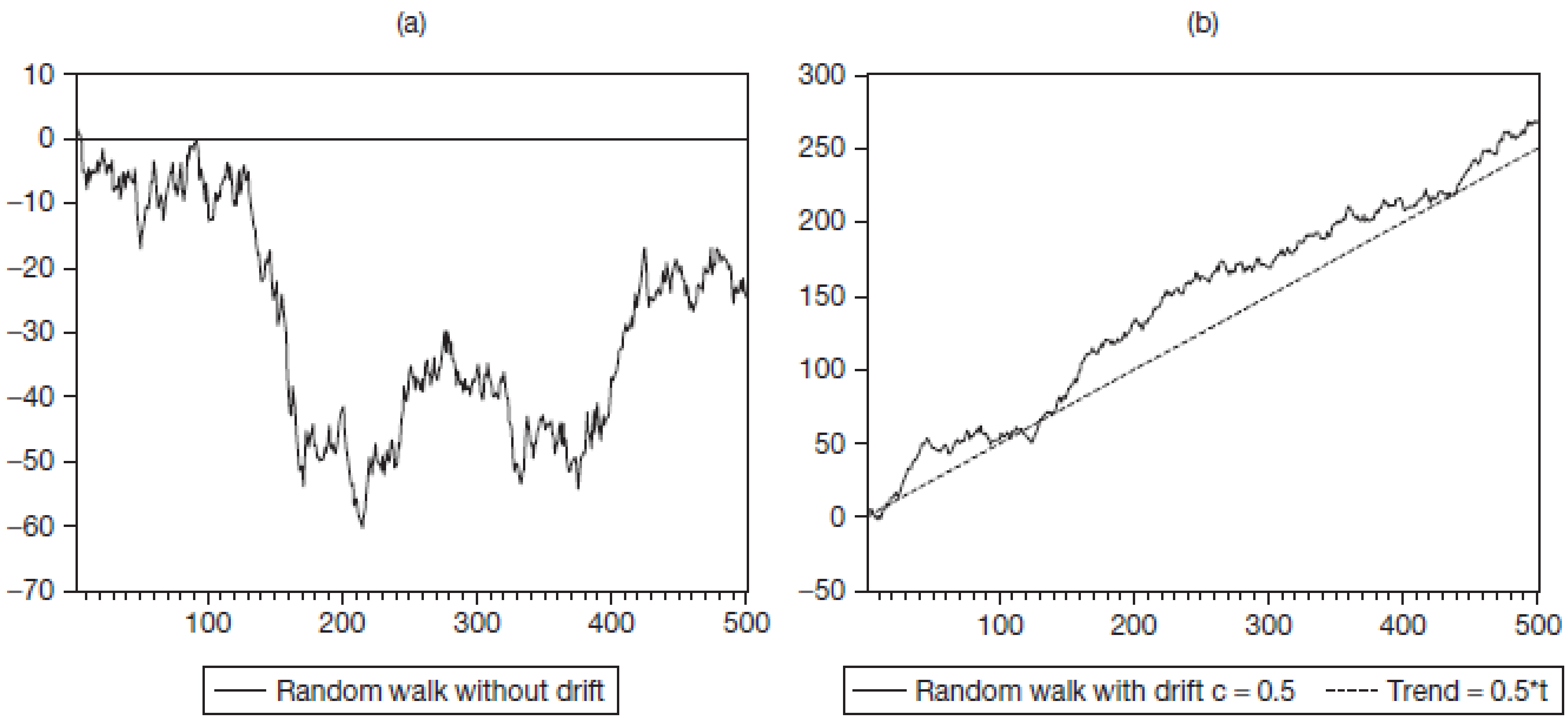


10.2 Stochastic Trends

Random walk without drift such as $Y_t = Y_{t-1} + \varepsilon_t$.

Random walk with drift such as $Y_t = c + Y_{t-1} + \varepsilon_t$, where c is a drift.

Figure 10.8 Random Walk without Drift and with Drift



$$E(\varepsilon_t) = 0,\ E(\varepsilon_t)^2 = \sigma_\varepsilon^2, \text{ and } E(\varepsilon_{t-i}\varepsilon_{t-j}) = 0 \quad i \neq j$$

Random walk without drift ($Y_0 = 0$):

$$Y_t = Y_{t-1} + \varepsilon_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots + \varepsilon_1$$

Unconditional mean

$$\mu \equiv E(Y_t) = E(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1) = 0$$

Unconditional variance

$$\sigma_Y^2 = E(Y_t - \mu)^2 = E(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1)^2 = t\sigma_\varepsilon^2$$

Autocovariance of order k

$$\begin{aligned} \gamma_{t,t-k} &\equiv E(Y_t - \mu)(Y_{t-k} - \mu) \\ &= E(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_{t-k} + \cdots + \varepsilon_1)(\varepsilon_{t-k} + \varepsilon_{t-k-1} + \cdots + \varepsilon_1) \\ &= (t-k)\sigma_\varepsilon^2 \end{aligned}$$

Autocorrelation of order k

$$\rho_{t,t-k} = \frac{\gamma_{t,t-k}}{\sigma_{Y_t}\sigma_{Y_{t-k}}} = \frac{(t-k)\sigma_\varepsilon^2}{\sqrt{t\sigma_\varepsilon^2}\sqrt{(t-k)\sigma_\varepsilon^2}} = \sqrt{\frac{t-k}{t}} \longrightarrow 1$$

Random walk with drift ($Y_0 = 0$):

$$Y_t = c + Y_{t-1} + \varepsilon_t = ct + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots + \varepsilon_1$$

Unconditional mean

$$\mu_t \equiv E(Y_t) = E(ct + \varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1) = ct$$

Unconditional variance

$$\sigma_Y^2 = E(Y_t - \mu_t)^2 = E(Y_t - ct)^2 = E(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1)^2 = t\sigma_\varepsilon^2$$

Autocovariance of order k

$$\gamma_{t,t-k} \equiv E(Y_t - \mu_t)(Y_{t-k} - \mu_{t-k}) = E(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_{t-k} + \cdots + \varepsilon_1)(\varepsilon_{t-k} + \varepsilon_{t-k-1} + \cdots + \varepsilon_1) = (t-k)\sigma_\varepsilon^2$$

Autocorrelation of order k

$$\rho_{t,t-k} = \frac{\gamma_{t,t-k}}{\sigma_{Y_t}\sigma_{Y_{t-k}}} = \frac{(t-k)\sigma_\varepsilon^2}{\sqrt{t\sigma_\varepsilon^2}\sqrt{(t-k)\sigma_\varepsilon^2}} = \sqrt{\frac{t-k}{t}} \longrightarrow 1 \quad \text{for } t \text{ large}$$

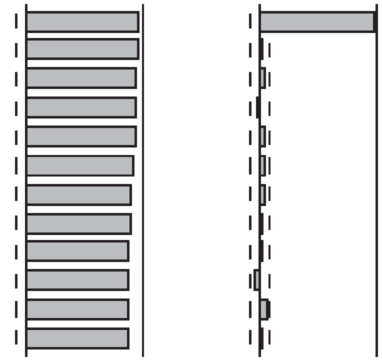
RWs are not covariance stationary

10.2.2 Stationarity Properties

Figure 10.9 Autocorrelograms of Random Walks

(a) Random Walk Without Drift (Figure 10.8a)

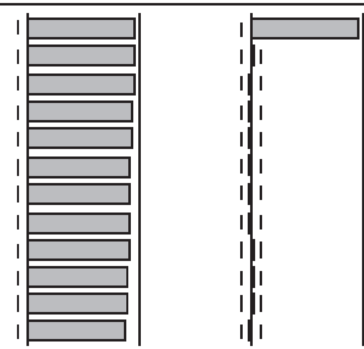
Sample: 2 500
Included observations: 499



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.990	0.990	492.08	0.000
		2	0.980	-0.014	975.14	0.000
		3	0.970	0.020	1449.7	0.000
		4	0.960	-0.029	1915.4	0.000
		5	0.951	0.024	2372.9	0.000
		6	0.942	0.027	2822.8	0.000
		7	0.934	0.023	3265.7	0.000
		8	0.925	0.004	3701.7	0.000
		9	0.917	0.001	4131.1	0.000
		10	0.909	-0.036	4553.2	0.000
		11	0.901	0.042	4968.9	0.000
		12	0.893	0.009	5378.5	0.000

(b) Random Walk with Drift (Figure 10.8b)

Sample: 2 500
Included observations: 499



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.993	0.993	495.20	0.000
		2	0.986	0.000	984.67	0.000
		3	0.980	-0.006	1468.4	0.000
		4	0.973	-0.004	1946.4	0.000
		5	0.966	-0.014	2418.6	0.000
		6	0.959	-0.014	2884.9	0.000
		7	0.952	-0.011	3345.1	0.000
		8	0.944	-0.019	3799.2	0.000
		9	0.937	0.015	4247.4	0.000
		10	0.930	0.000	4689.7	0.000
		11	0.923	0.006	5126.4	0.000
		12	0.916	-0.004	5557.3	0.000

10.2.2.1 Testing for Unit Root

Recall: RW $Y_t = Y_t + \varepsilon_t$ is an "AR(1)" with $\phi = 1$

As this "AR(1)" is not stationary, we need a different test for

$$\begin{cases} H_0: \phi = 1 \\ H_1: \phi < 1 \end{cases}$$

The statistic is still $\dfrac{\hat{\phi}-1}{\hat{\sigma}_{\hat{\phi}}}$

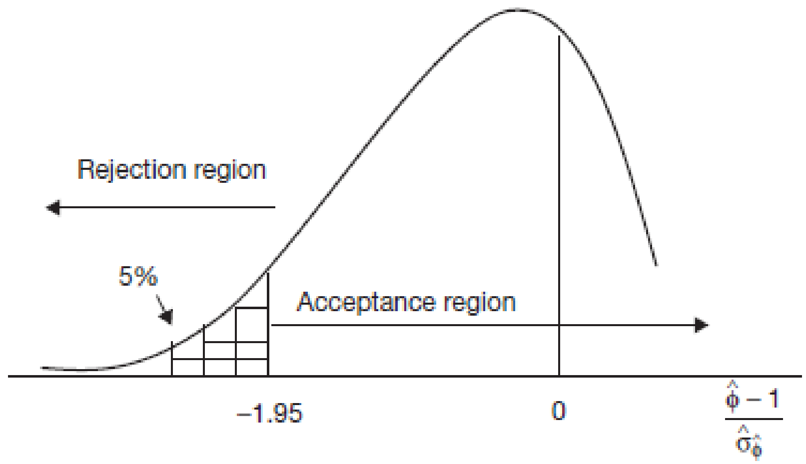
but it's distribution is non-standard.

We have to use the tabulated **Dickey-Fuller test**.

We have three cases:
I – mean zero no trend
II – nonzero mean no trend
III – trend

Table 10.4 Dickey-Fuller Critical Values for a 5% Critical Region

Example: Case I, $T = 25$, $\Pr\left(\frac{\hat{\phi}-1}{\hat{\sigma}_{\hat{\phi}}} \le -1.95\right) = 5\%$



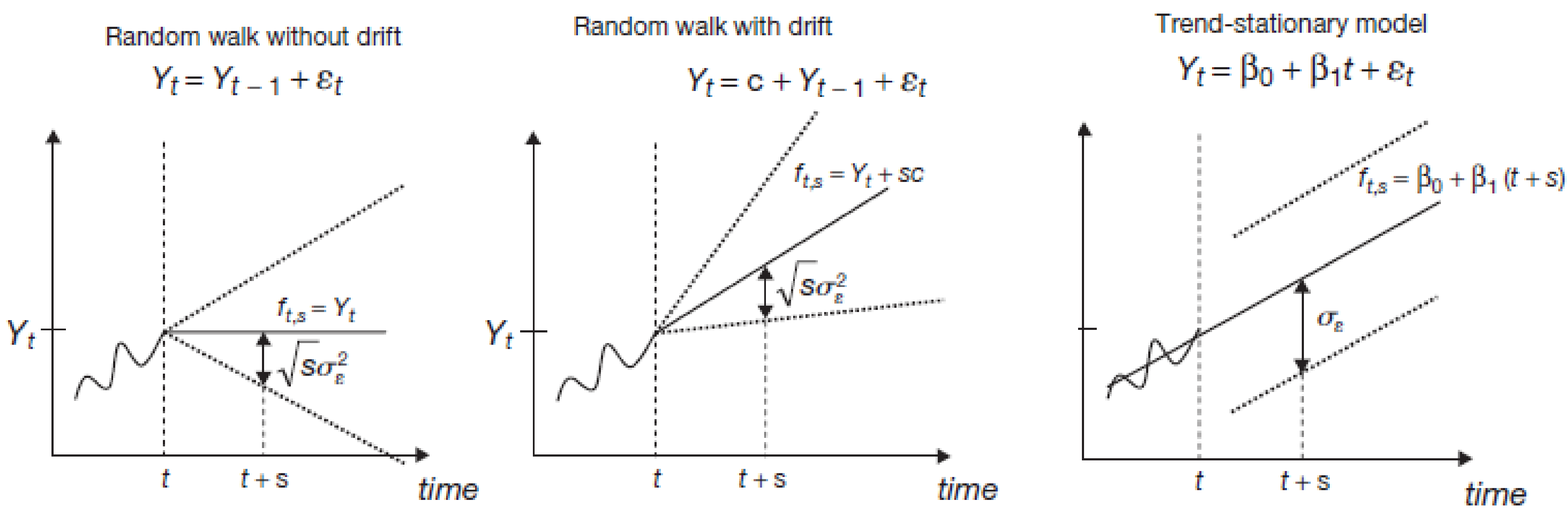
	Case I	Case II	Case III
Sample size T	$H_0: Y_t = Y_{t-1} + \varepsilon_t$ $H_1: Y_t = \phi Y_{t-1} + \varepsilon_t$	$H_0: Y_t = Y_{t-1} + \varepsilon_t$ $H_1: Y_t = c + \phi Y_{t-1} + \varepsilon_t$	$H_0: Y_t = c + Y_{t-1} + \varepsilon_t$ $H_1: Y_t = c + \alpha t + \phi Y_{t-1} + \varepsilon_t$
25	–1.95	–3.00	–3.60
50	–1.95	–2.93	–3.50
100	–1.95	–2.89	–3.45
250	–1.95	–2.88	–3.43
500	–1.95	–2.87	–3.42
∞	–1.95	–2.86	–3.41

Table 10.5 Unit Root Testing Procedure

	OLS-Estimated model (standard error of $\hat{\phi}$)	Ratio: $\frac{\hat{\phi}-1}{\hat{\sigma}_{\hat{\phi}}}$	Dickey-Fuller critical value at 5% level (Table 10.4)	Decision
Fig 10.8a				
Case I $H_0: Y_t = Y_{t-1} + \varepsilon_t$ $H_1: Y_t = \phi Y_{t-1} + \varepsilon_t$	$Y_t = 0.9988 Y_{t-1} + \hat{\varepsilon}_t$ (0.0025)	$\frac{0.9988-1}{0.0025} = -0.4341$	−1.95	$-1.95 < -0.43$ fail to reject $H_0 \Rightarrow$ unit root
Case II $H_0: Y_t = Y_{t-1} + \varepsilon_t$ $H_1: Y_t = c + \phi Y_{t-1} + \varepsilon_t$	$Y_t = -0.325 +$ $+ 0.9905 Y_{t-1} + \hat{\varepsilon}_t$ (0.0051)	$\frac{0.9905-1}{0.0051} = -1.8530$	−2.87	$-2.87 < -1.85$ fail to reject $H_0 \Rightarrow$ unit root
Fig. 10.8b				
Case III $H_0: Y_t = c + Y_{t-1} + \varepsilon_t$ $H_1: Y_t = c + \alpha t + \phi Y_{t-1} + \varepsilon_t$	$Y_t = 0.9167 + 0.0082 t +$ $+ 0.9830 Y_{t-1} + \hat{\varepsilon}_t$ (0.0077)	$\frac{0.9830-1}{0.0077} = -2.1853$	−3.42	$-3.42 < -2.18$ fail to reject $H_0 \Rightarrow$ unit root

10.2.3 Optimal Forecast

Figure 10.10 Differences in the Forecasts of Random Walk and Trend-Stationary Processes



1. Unit roots imply the need for differencing
2. We may need to difference *d* times *d* = 1, 2, …
3. Instead of a WN we could have an ARMA noise

This means that we can generalize the model to an **ARIMA(*p,d,q*)**:

$$\Phi_p(L)\Delta^d Y_t = \theta_q(L)\varepsilon_t$$

In order to account for such a structure we use the so-called

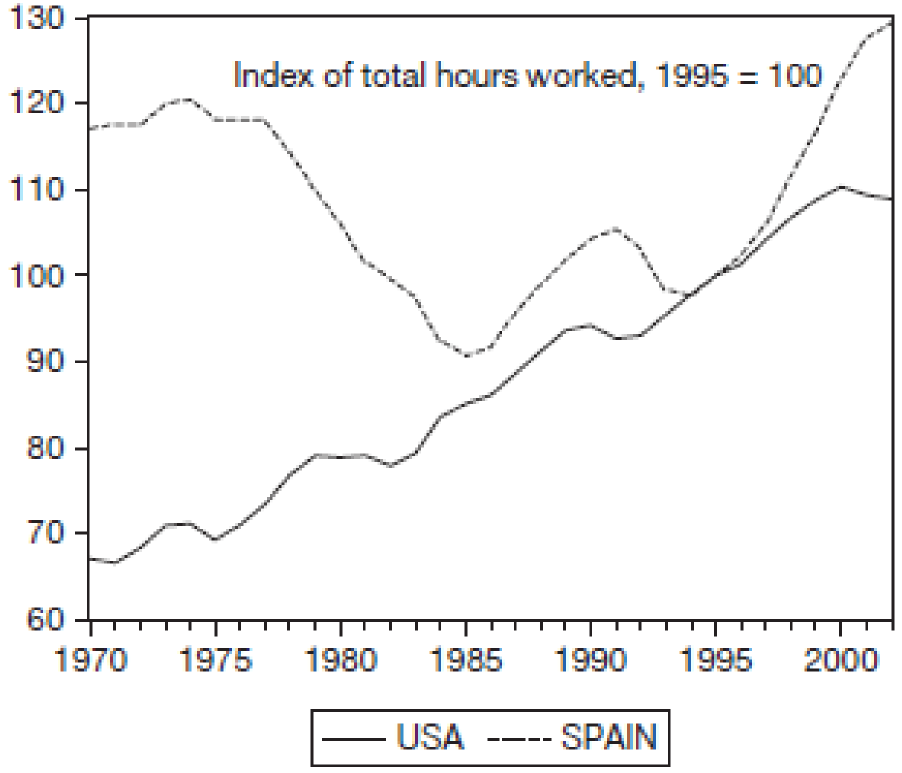
Augmented Dickey-Fuller (ADF) Test

We simply need to augment the regression with the lagged differences necessary to destroy autocorrelation:

$$\Delta Y_t = c + \beta Y_{t-1} + \alpha_1 \Delta Y_{t-1} + \alpha_2 \Delta Y_{t-2} + \alpha_3 \Delta Y_{t-3} + \cdots + \varepsilon_t$$

An example:

Figure 10.11 Index of Total Hours Worked in United States and Spain



It seems wise to test for a UR with intercept (Type II)

Table 10.6 Augmented Dickey-Fuller Unit Root Test on Spain

ADF Test Statistic	0.140798	1% Critical Value*	-3.6496
		5% Critical Value	-2.9558
		10% Critical Value	-2.6164

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SPAIN)
Method: Least Squares
Sample(adjusted): 1971 2002
Included observations: 32 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SPAIN(-1)	0.008178	0.058084	0.140798	0.8890
C	-0.494473	6.272498	-0.078832	0.9377

R-squared	0.000660	Mean dependent var	0.384840
Adjusted R-squared	-0.032651	S.D. dependant var	3.254600
S.E. of regression	3.307307	Akaike info criterion	5.290607
Sum squared resid	328.1483	Schwarz criterion	5.382215
Log likelihood	-82.64971	F-statistic	0.019824
Durbin-Waston stat	0.437263	Prob(F-statistic)	0.888970

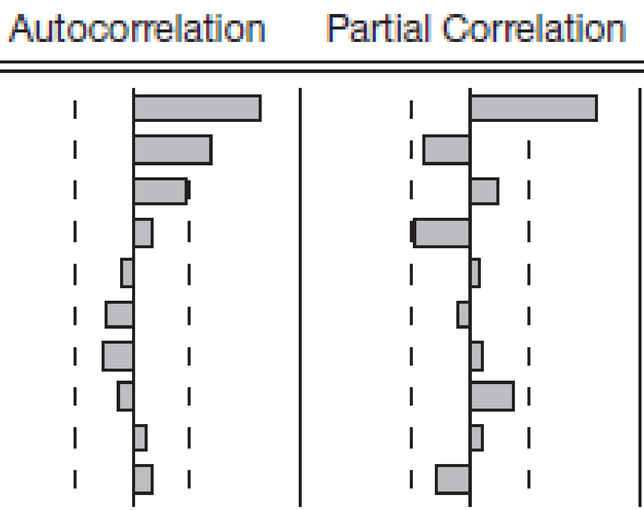
This is the AR(1) with $\beta = (\phi - 1)$

We fail to reject the Unit Root

But the residuals are not yet white

Figure 10.12 Correlogram of Residuals

Sample: 1971 2002
Included observations: 32



	AC	PAC	Q-Stat	Prob
1	0.779	0.779	21.279	0.000
2	0.487	-0.303	29.881	0.000
3	0.315	0.160	33.613	0.000
4	0.115	-0.340	34.125	0.000
5	-0.067	0.041	34.303	0.000
6	-0.156	-0.076	35.324	0.000
7	-0.182	0.058	36.761	0.000
8	-0.086	0.265	37.097	0.000
9	0.073	0.065	37.350	0.000
10	0.107	-0.202	37.921	0.000

How many differencing should be needed?

Table 10.7 Augmented Dickey-Fuller Unit Root Test on Spain

ADF Test Statistic	-2.153659	1% Critical Value*	-3.6576
		5% Critical Value	-2.9591
		10% Critical Value	-2.6181

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SPAIN)
Method: Least Squares
Sample(adjusted): 1972 2002
Included observations: 31 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SPAIN(-1)	-0.079919	0.037109	-2.153659	0.0400
D(SPAIN(-1))	0.864257	0.115315	7.494743	0.0000
C	8.661998	3.982150	2.175206	0.0382

R-squared	0.667569	Mean dependent var	0.384577
Adjusted R-squared	0.643824	S.D. dependent var	3.308399
S.E. of regression	1.974467	Akaike info criterion	4.290240
Sum squared resid	109.1586	Schwarz criterion	4.429013
Log likelihood	-63.49872	F-statistic	28.11405
Durbin-Watson stat	1.737471	Prob(F-statistic)	0.000000

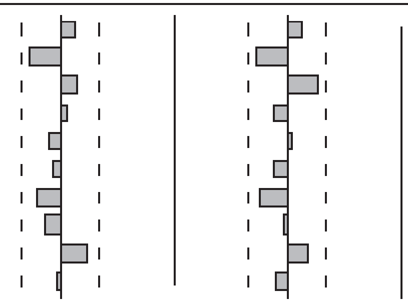
We fail to reject the Unit Root

We can assume the residuals are white and proceed to the estimation

Correlogram of Residuals

Sample: 1972 2002
Included observations: 31

Autocorrelation | Partial Correlation



	AC	PAC	Q-Stat	Prob
1	0.129	0.129	0.5661	0.452
2	-0.272	-0.293	3.1666	0.205
3	0.148	0.257	3.9699	0.265
4	0.060	-0.119	4.1085	0.392
5	-0.114	0.022	4.6232	0.464
6	-0.070	-0.115	4.8221	0.567
7	-0.221	-0.260	6.9045	0.439
8	-0.128	-0.044	7.6381	0.470
9	0.239	0.188	10.284	0.328
10	-0.022	-0.112	10.308	0.414

Table 10.8 Least Squares Estimation of $\Delta Y_t = \alpha_1 \Delta Y_{t-1} + \varepsilon_t$

This is the ARIMA(1,1,0)

Dependent Variable: D(SPAIN)
Method: Least Squares
Sample(adjusted): 1972 2002
Included observations: 31 after adjusting endpoints
Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.789202	0.113642	6.944641	0.0000

R-squared	0.611151	Mean dependent var	0 384577
Adjusted R-squared	0.611151	S.D. dependent var	3.308399
S.E. of regression	2.063043	Akaike info criterion	4.317967
Sum squared resid	127.6844	Schwarz criterion	4.364225
Log likelihood	-65.92849	Durbin-Watson stat	1.532094
Inverted AR Roots	.79		

Table 10.9 Four-Step-Ahead Forecast of Index of Total Hours Worked in Spain

$$\sigma^2_{t+1|t} = \sigma^2_\varepsilon$$

$$e^*_{t,2} = e_{t,2} + e^*_{t,1} = \varepsilon_{t+2} + \phi\,\varepsilon_{t+1} + \varepsilon_{t+1} = \varepsilon_{t+2} + (\phi + 1)\varepsilon_{t+1}$$

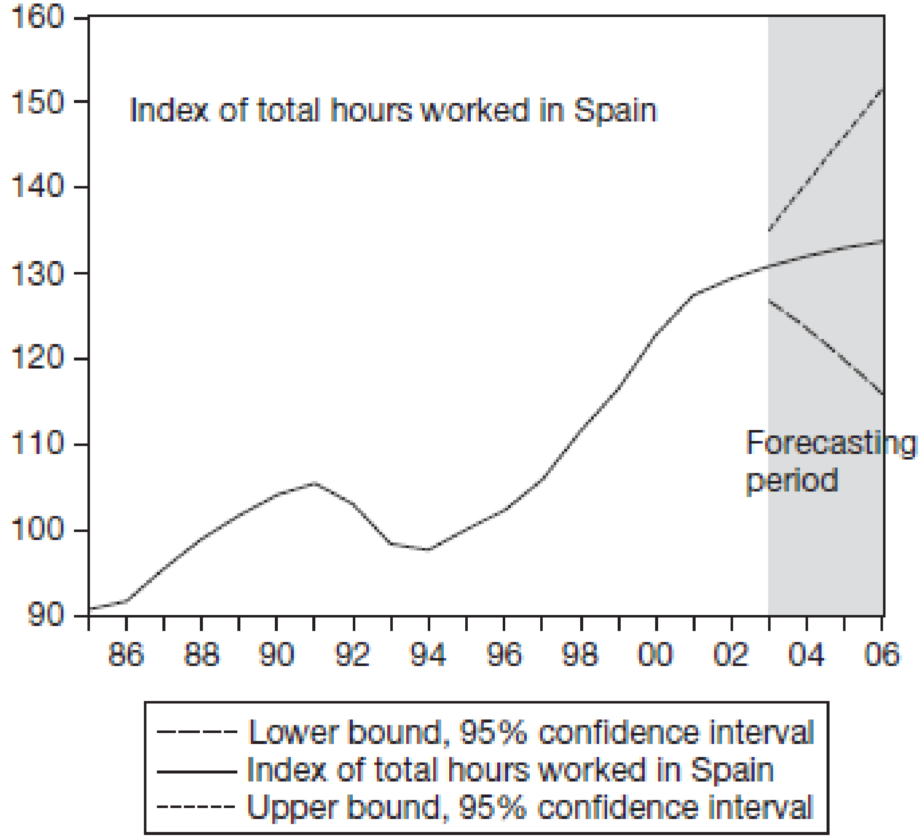
$$\sigma^2_{t+2|t} = \sigma^2_\varepsilon(1 + (\phi + 1)^2)$$

Forecasts for differences

Forecasts for levels

$t = 2002$	$\Delta Y_{2002} = 1.88$		$Y_{2002} = 129.35$	
Forecasting horizon	$f_{2002,s} = \phi^s \Delta Y_{2002}$	$\sigma_{t+s\|2002}$	$f^*_{2002,s} = f_{2002,s} + f^*_{2002,s-1}$	$\sigma^*_{t+s\|2002}$
$s = 1$ 2003	1.48	2.07	130.83	2.07
$s = 2$ 2004	1.17	2.64	132.01	4.25
$s = 3$ 2005	0.92	2.93	132.93	6.56
$s = 4$ 2006	0.73	3.10	133.66	8.90

Figure 10.13 Forecast of Index Total Hours Worked in Spain with 95% Confidence Bands



Forecasting error always diverges

Slope converges to differenced series estimated mean