1

Data Analysis Based on a Single Time Series by States

1.1 Introduction

Panel data can be viewed as a finite set of time-series data. As an illustration Table 1.1 presents part of the data in POOLG7.wf1, namely *unstacked data*, consisting of a single time series *GDP* from seven countries. Note that this table shows seven time series variables, namely GDP_CAN_t to GDP_US_t.

Based on each time series of GDP by states, various growth models can be considered as presented in Agung (2009a, Chapter 2), starting with classical growth models, namely geometric and exponential growth models, and their extensions. Therefore, based on the seven states, the multivariate growth models should be applied as presented in the following sections.

1.2 Multivariate Growth Models

1.2.1 Continuous Growth Models

In general, let Y_{it} be the observed value of the variable Y for the i-th individual (a country, state, region, agency, community, household or person) at time t, for $i = 1, \ldots, N$, and $t = 1, \ldots, T$. In panel data analysis, the symbol $Y\_i(t)$, Y_i_t, or $Y\_\text{“}Name\text{”}_t$ will be used to indicate the time series variable Y_{it}, such as the variable GDP_Can_t to GDP_US_t in POOL7.wf1. In this chapter, the panel data set will be considered as a finite set of time-series variables. For this reason, the simplest model considered is a multivariate classical growth model with the following general equation.

$$\log(Y\_i_t) = C(i1) + C(i2)^{*}t + \mu_{it} \quad (1.1)$$

where *C(i2)* indicates the exponential growth rate of *Y_i*, that is the growth rate of the variable *Y* for the *i*-th individual (country, state, region, community, household, firm or agency), *C(i1) is* the intercept parameter, and μ_{it} their residuals which, in general, should be autocorrelated (see to Agung 2009a, Chapter 2).

Panel Data Analysis Using EViews, First Edition. I Gusti Ngurah Agung.

Companion website: www.wiley.com/go/panel_data

Table 1.1 *A subset of the unstacked data in POOLG7.wf1*

Year	GDP_CAN	GDP_FRA	GDP_GER	GDP_ITA	GDP_JPN	GDP_UK	GDP_US
1950	6209	4110	3415	2822	1475	5320	8680
1951	6385	4280	3673	3023	1649	5484	9132
1952	6752	4459	4013	3131	1787	5517	9213
1953	6837	4545	4278	3351	1884	5791	9450
1954	6495	4664	4577	3463	1972	5971	9177
1955	6907	4861	5135	3686	2108	6158	9756
1956	7349	5195	5388	3815	2249	6238	9756
1957	7213	5389	5610	3960	2394	6322	9724
1958	7061	5463	5787	4119	2505	6340	9476
1959	7180	5610	6181	4351	2714	6569	9913

Therefore, the basic growth model considered should be a multivariate autoregressive growth model, namely MAR($q_1, \ldots, q_i, \ldots$)_GM = MAR($\boldsymbol{q}$)_GM, with the following general equation, where the error terms ε_{it} would be assumed or accepted to have an $i.i.d.N(0, \sigma_i^2)$, in a theoretical sense. Refer to the special notes presented in Section 2.14.3 (Agung, 2009a).

$$\begin{aligned} &\log(Y\_i_t) = C(i1) + C(i2)^* t + \mu_{it} \\ &\mu_{it} = \rho_{i1}\mu_{i,t-1} + \rho_{i,t-2}\mu_{i,t-2} + \ldots + \rho_{i,t-q_i}\mu_{i,t-q_i} + \varepsilon_{it} \end{aligned} \tag{1.2}$$

However, for a multivariate GLM, the vector of the error terms $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)$, in general, would have a residual correlation matrix, namely CM($\boldsymbol{\varepsilon}$), or a residual covariance matrix, namely $\Sigma(\boldsymbol{\varepsilon})$, which is not a diagonal matrix, and should indicate that the endogenous variables *log(Y_i)* or *Y_i*, for the states $i = 1, 2, \ldots, N$, are correlated in a statistical sense, even though they may not be correlated in a theoretical sense. In other words, the quantitative correlations between all *log(Y_i)* are taken into account in the estimation process.

Example 1.1 Illustrative growth curves

As an illustration, Figure 1.1 presents the growth curves GDP_t of two pairs of neighboring countries, namely (a) *GDP_CAN* and *GDP_US*, and (b) *GDP_FRA* and *GDP_GER*, which clearly show differential characteristics. Corresponding to growth curves, we find that each pair of the five variables *GDP_CAN*,

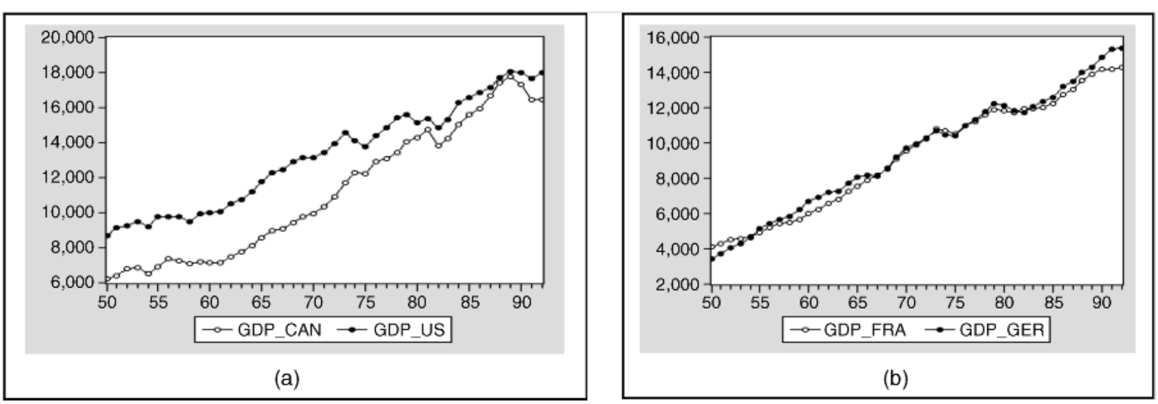


Figure 1.1 *Growth curves of* GDP_CAN, GDP_US, GDP_FRA *and* GDP_GE

GDP_US, *GDP_FRA*, *GDP_GER* and the time *t* variable are significantly positively correlated with a *p*-value = 0.0000. However, unexpected statistical results are obtained based on the model in (1.2), as presented in Example 1.3.

Growth curves are important descriptive statistics in any time series, as well as panel data analyses. Many findings and conclusions can be derived based on descriptive statistical summaries. See various continuous and discontinuous growth curves and time series models presented in Agung (2009a), and the descriptive statistical summaries presented in Agung (2004, 2009b, 2011). For additional illustrations, see the graphical presentations in Leary (2009), and Chambers and Dimson (2009).

Example 1.2 A multivariate classical growth model (MCGM)

Figure 1.2 presents the statistical results based on a MCGM of *GDP_Can*, *GDP_US, GDP_Fra*, and *GDP_Ger*. Its residuals graphs are obtained by selecting View/Residuals/Graphs, as presented in Figure 1.3.

Based on these results, the following notes are presented.

1. Note that the four regressions in the model in fact represent a growth model by states, which has been presented as a multiple regression model or a single time series model using dummy variables of the states in Agung (2009a).
2. Using the standard *t*-test statistic in the output, it can be concluded that *GDP_Can, GDP_US, GDP_Fra* and *GDP_Ger*, have significant positive exponential growth rates of

$$\hat{C}(11) = 0.0273339, \hat{C}(21) = 0.018282, \hat{C}(31) = 0.030681, \text{and } \hat{C}(41) = 0.032058.$$

3. The null hypothesis H_0: $C(11) = C(21) = C(31) = C(41)$ is rejected based on the Chi-square statistic of $\chi_0^2 = 242.8469$ with $df = 3$ and a *p*-value = 0.0000. Therefore, it can be concluded that the growth rates of *GDP* of the four countries have significant differences. The other hypotheses on the growth rates differences can easily be tested using the Wald test.
4. However, note that the MCGM is an inappropriate time series model indicated by the very small Durbin–Watson statistics values of the four regressions, as well as their residuals graphs in Figure 1.3. For this reason, a modified GM will be presented in the following example. Refer also to Chapter 2 in Agung (2009a).

Estimation Method: Least Squares
Date: 08/22/09 Time: 18:49
Sample: 1950 1992
Included observations: 43
Total system (balanced) observations 172

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	8.683015	0.017182	505.3611	0.0000
C(11)	0.027339	0.000704	38.81228	0.0000
C(20)	9.080149	0.011628	780.8541	0.0000
C(21)	0.018282	0.000477	38.34938	0.0000
C(30)	8.414288	0.025252	333.2182	0.0000
C(31)	0.030681	0.001035	29.63753	0.0000
C(40)	8.403292	0.030639	274.2682	0.0000
C(41)	0.032058	0.001256	25.52230	0.0000
Determinant residual covariance		1.82E-11		

Equation: LOG(GDP_CAN)=C(10)+C(11)*T
Observations: 43

R-squared	0.973504	Mean dependent var	9.257129
Adjusted R-squared	0.972858	S.D. dependent var	0.347921
S.E. of regression	0.057320	Sum squared resid	0.134708
Durbin-Watson stat	0.323229		

Equation: LOG(GDP_US)=C(20)+C(21)*T
Observations: 43

R-squared	0.972878	Mean dependent var	9.464070
Adjusted R-squared	0.972216	S.D. dependent var	0.232736
S.E. of regression	0.038794	Sum squared resid	0.061702
Durbin-Watson stat	0.452012		

Equation: LOG(GDP_FRA)=C(30)+C(31)*T
Observations: 43

R-squared	0.955405	Mean dependent var	9.058594
Adjusted R-squared	0.954317	S.D. dependent var	0.394138
S.E. of regression	0.084241	Sum squared resid	0.290960
Durbin-Watson stat	0.060788		

Equation: LOG(GDP_GER)=C(40)+C(41)*T
Observations: 43

R-squared	0.940785	Mean dependent var	9.076509
Adjusted R-squared	0.939340	S.D. dependent var	0.415012
S.E. of regression	0.102214	Sum squared resid	0.428355
Durbin-Watson stat	0.084715		

Figure 1.2 *Statistical results based on a multivariate classical growth model*

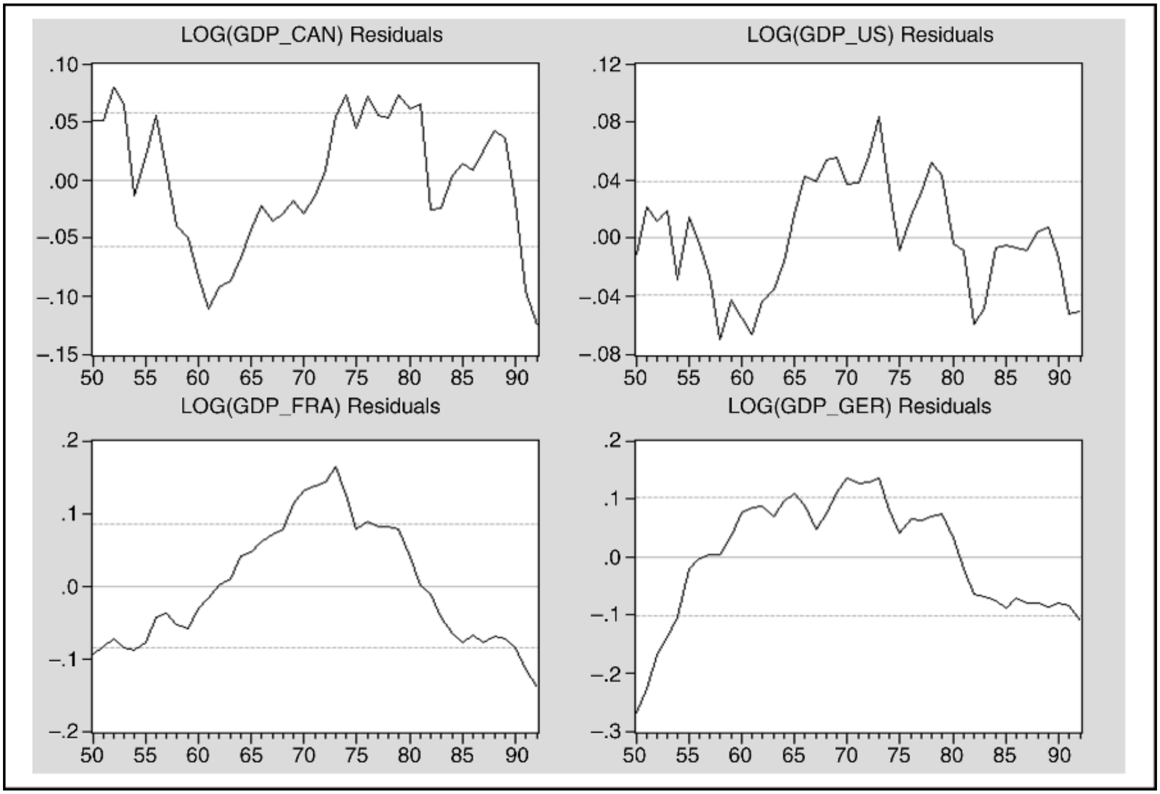


Figure 1.3 *The residuals graphs of the MCGM in Figure 1.2*

5. On the other hand, by observing the residual graphs in Figure 1.3, then it can be said that a polynomial growth model should be explored for each state, such as quadratic regressions of *log(GDP_FRA)* and *log(GDP_GER)* on the time t, and at least third degree polynomials of *GDP_CAN* and *GDP_US* on the time t. Do this as an exercise.
6. As an additional illustration, Table 1.2 presents the correlations between the time t with the dependent variables of each model. Note that each parameter *C(i2)* has exactly the same value of the t-statistic, as well as Prob(t-stat). Compared to the results in Figure 1.2, the following notes and conclusions are made.
   6.1 The testing hypothesis on each *C(i2)*, either a two- or one-sided hypothesis, can be done using the corresponding bivariate correlation. To generalize the results, the set of simple linear regressions can be presented using a correlation matrix of the set of variables considered.

Table 1.2 *Bivariate correlations of time* t *with each of the dependent variables of the multivariate model in Figure 1.2*

	LOG(GDP_CAN)	LOG(GDP_US)	LOG(GDP_FRA)	LOG(GDP_GER)
Time *t*	0.986 663	0.98 635	0.97 745	0.96 994
***t*-stat**	38.81 228	38.34 938	29.63 753	25.52 230
Prob.	0.00 000	0.00 000	0.00 000	0.00 000

6.2 On the other hand, by doing a series of *state-by-state analyses*, we obtain exactly the same set of four regressions as presented in Figure 1.2. For this reason the model presented in Figure 1.2 will be referred to as the *system of independent states*.

Example 1.3 A MAR(1)_GM unexpected result

Figure 1.4 presents the statistical results based on a MAR(**1**)_GM = MAR(1,1,1,1)_GM of the four time series *GDP_CAN, GDP_US*, *GDP_FRA*, and *GDP_GER*. Based on these results, the following findings and notes are presented.

1. The estimate of $C(31) = -0.212\,090$ with a p-value = 0.9358, which should indicate the (adjusted) growth rate of *GDP_FRA*, is an *unexpected result*, since *r(log(GDP_FRA),t)* = 0.97 448 with a p-value = 0.0000 and obtains a simple linear regression function of *LOG(GDP_FRA)* = 8.3841 + $0.0307^{*}t$ as presented in Figure 1.2, with an exponential growth rate of *GDP_FRA* as $r = 0.0307$.
2. This finding indicates the impact of using an AR(1) on the parameter estimates is in fact unpredictable. Nothing is wrong with the model, but the structure of the data set cannot provide acceptable estimates. Compared to the growth curve of *GDP_FRA* in Figure 1.1, the AR(1)_GM of *GDP_FRA* should be considered as an unacceptable or inappropriate time series model for representing the *GDP* of France. The results of the author's experimentation based on the variable *GDP_FRA*, are presented in the following examples.
3. On the other hand, we find the residual matrix correlation, says M(**ε**), is not a diagonal matrix. For comparison, application to the WLS or SUR estimation methods is recommended. Do this as an exercise.
4. For a comparison study, Table 1.3 presents a summary of the statistical results using the series of state-by-state analyses based on the LS AR(1)_GMs. Note that this table shows the coefficients of the time t and the AR(1) terms are exactly the same as those in Figure 1.4, but they have different intercepts. Compare this to the other statistics.

System: SYS01
Estimation Method: Iterative Least Squares
Date: 07/07/09 Time: 15:05
Sample: 1951 1992
Included observations: 43
Total system (balanced) observations 168
Convergence achieved after 2 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	8.675757	0.117154	74.05426	0.0000
C(11)	0.025588	0.003869	6.613016	0.0000
C(12)	0.881066	0.094300	9.343263	0.0000
C(20)	9.080932	0.045963	197.5692	0.0000
C(21)	0.017432	0.001607	10.84940	0.0000
C(22)	0.788836	0.104235	7.567884	0.0000
C(30)	86.20318	1629.510	0.052901	0.9579
C(31)	-0.212090	2.627543	-0.080718	0.9358
C(32)	0.996671	0.035945	27.72798	0.0000
C(40)	8.768484	0.188648	46.48057	0.0000
C(41)	0.020326	0.004748	4.281062	0.0000
C(42)	0.891366	0.035255	25.28374	0.0000
Determinant residual covariance		2.00E-14		

Equation: LOG(GDP_CAN)=C(10)+C(11)*T+[AR(1)=C(12)]
Observations: 42

R-squared	0.991522	Mean dependent var	9.269590
Adjusted R-squared	0.991087	S.D. dependent var	0.342288
S.E. of regression	0.032314	Sum squared resid	0.040724
Durbin-Watson stat	1.510506		

Equation: LOG(GDP_US)=C(20)+C(21)*T+[AR(1)=C(22)]
Observations: 42

R-squared	0.988209	Mean dependent var	9.473482
Adjusted R-squared	0.987604	S.D. dependent var	0.227124
S.E. of regression	0.025287	Sum squared resid	0.024938
Durbin-Watson stat	1.762500		

Equation: LOG(GDP_FRA)=C(30)+C(31)*T+[AR(1)=C(32)]
Observations: 42

R-squared	0.997723	Mean dependent var	9.076151
Adjusted R-squared	0.997607	S.D. dependent var	0.381517
S.E. of regression	0.018665	Sum squared resid	0.013587
Durbin-Watson stat	1.414278		

Equation: LOG(GDP_GER)=C(40)+C(41)*T+[AR(1)=C(42)]
Observations: 42

R-squared	0.996821	Mean dependent var	9.098903
Adjusted R-squared	0.996658	S.D. dependent var	0.392866
S.E. of regression	0.022710	Sum squared resid	0.020114
Durbin-Watson stat	1.596264		

Figure 1.4 *Statistical results based on a MAR(**1**)_GM of the* GDP *of four countries*

Table 1.3 *Summary of the statistical results based on the four LS AR(1)_GMs in Figure 1.4*

Variable	Dependent Variables							
	log(Gdp_Can)		log(Gdp_US)		log(Gdp_Fra)		log(Gdp_Ger)	
	Coef.	*t*-Stat.	Coef.	*t*-Stat.	Coef.	*t*-Stat.	Coef.	*t*-Stat.
C	8.701 345	76.60 812	9.098 363	204.4720	85.99 109	0.052 856	8.788 810	47.76 036
T	0.025 588	6.613 016	0.017 432	10.84 940	−0.212 090	−0.080 718	0.020 326	4.281 062
AR(1)	0.881 066	9.343 263	0.788 836	7.567 884	0.996 671	27.72 798	0.891 366	25.28 374
R-squared	0.991 522		0.988 209		0.997 723		0.996 821	
Adjusted R-squared	0.991 087		0.987 604		0.997 607		0.996 658	
S.E. of regression	0.032 314		0.025 287		0.018 665		0.022 710	
F-statistic	2280.633		1634.300		8545.536		6115.340	
Prob(F-statistic)	0.000 000		0.000 000		0.000 000		0.000 000	
Durbin–Watson stat	1.510 506		1.762 500		1.414 278		1.596 264	

1.2.2 Discontinuous Growth Models

Corresponding to the inappropriate estimate of *C(31)* = − 0.212 090 in Figure 1.4, experimentation should be done based on the data of the *GDP_FRA*. See the following examples.

Example 1.4 An experimentation based on *GDP_FRA*

By using trial-and-error methods, we finally obtain the statistical results in Figure 1.5, based on two sub-samples of sizes 29 and 30, respectively, for T < 31 and T < 32. Based on these results the following findings and notes are presented.

1. Based on the results in Figure 1.5(a), the null hypothesis H_0: $C(2) \leq 0$ is rejected, based on the *t*-statistic of $t_0 = 1.819\,125$ with a *p*-value = 0.0804/2 = 0.0402 < 0.05. Therefore, it can be

Dependent Variable: LOG(GDP_FRA)
Method: Least Squares
Date: 07/08/09 Time: 09:55
Sample: 1950 1992 IF T< 31
Included observations: 29
Convergence achieved after 4 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C	8.494952	0.705860	12.03490	0.0000
T	0.031080	0.017085	1.819125	0.0804
AR(1)	0.941059	0.107208	8.777880	0.0000

R-squared	0.997041	Mean dependent var	8.904769
Adjusted R-squared	0.996814	S.D. dependent var	0.335020
S.E. of regression	0.018911	Akaike info criterion	-5.000398
Sum squared resid	0.009299	Schwarz criterion	-4.858954
Log likelihood	75.50578	Hannan-Quinn criter.	-4.956100
F-statistic	4380.578	Durbin-Watson stat	1.490196
Prob(F-statistic)	0.000000		

Inverted AR Roots .94

(a)

Dependent Variable: LOG(GDP_FRA)
Method: Least Squares
Date: 07/08/09 Time: 09:56
Sample: 1950 1992 IF T< 32
Included observations: 30
Convergence achieved after 4 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C	9.809473	18.01907	0.544394	0.5906
T	0.076572	0.197390	0.387924	0.7011
AR(1)	1.019645	0.102863	9.912682	0.0000

R-squared	0.996862	Mean dependent var	8.920418
Adjusted R-squared	0.996629	S.D. dependent var	0.340169
S.E. of regression	0.019750	Akaike info criterion	-4.916720
Sum squared resid	0.010531	Schwarz criterion	-4.776600
Log likelihood	76.75080	Hannan-Quinn criter.	-4.871895
F-statistic	4288.174	Durbin-Watson stat	1.527986
Prob(F-statistic)	0.000000		

Inverted AR Roots 1.02
Estimated AR process is nonstationary

(b)

Figure 1.5 *Statistical results based on an AR(1)_GM of* GDP_FRA *using two sub-samples* concluded that *GDP_FRA* has a significant positive growth rate of 3.11% within the time period $t=1$ to $t=30$.

2. On the other hand, Figure 1.5(b) shows a note "*Estimated AR process is nonstationary*", which indicates that the AR(1)_GM is an inappropriate time series model within the period $t=1$ to $t=31$. Finally, based on the whole data set, the Inverted AR Root = 1.00 is obtained, however, without the statement "*Estimated AR process is nonstationary*".

Example 1.5 A piece-wise growth model of *GDP_FRA*

As the complement of the AR(1)_GM of *GDP_FRA* for $t<31$, Figure 1.6(a) presents another piece of AR(1)_GM of *GDP_FRA* for $t>=31$, which should be considered an acceptable time series model, in a statistical sense. Note that this model shows that GDP_FRA has a significant positive growth rate of 2.18% based on the *t*-statistic of $t_0=7.425\,573$ with a *p*-value = 0.0000, for $t>=31$, compared to the growth rate of 3.11% for $t<31$. Therefore, based on these findings the growth model of *GDP_FRA* could be presented by a two-piece GM using dummy variables *Dt1* and *Dt2*, which should be generated for the two time periods.

Figure 1.6(b) presents the statistical results based on an acceptable two-piece AR(2)_GM of *log(GDP_FRA)*, in a statistical sense. Based on these results, the following pair of regression functions can be derived.

$$\log(GDP\_FRA) = 8.3274 + 0.0359^{*}t + [AR(1) = 1.2142, AR(2) = -0.3301], \quad \text{for } t < 31$$

$$\log(GDP\_FRA) = 9.0101 + 0.0127^{*}t + [AR(1) = 1.2142, AR(2) = -0.3301], \quad \text{for } t \geq 31$$

Based on these findings, the MAR(**1**)_GM presented in Figure 1.4 should be modified to a MAR (1,1,2,1)_GM, with the statistical results presented in Figure 1.7.

Dependent Variable: LOG(GDP_FRA)
Method: Least Squares
Date: 07/08/09 Time: 16:45
Sample: 1950 1992 IF T> 30
Included observations: 13
Convergence achieved after 4 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C	8.639295	0.118147	73.12313	0.0000
T	0.021810	0.002937	7.426573	0.0000
AR(1)	0.610004	0.158493	3.848772	0.0032

R-squared	0.976634	Mean dependent var	9.458464
Adjusted R-squared	0.971960	S.D. dependent var	0.078091
S.E. of regression	0.013076	Akaike info criterion	-5.636837
Sum squared resid	0.001710	Schwarz criterion	-5.506464
Log likelihood	39.63944	Hannan-Quinn criter.	-5.663634
F-statistic	208.9827	Durbin-Watson stat	1.417305
Prob(F-statistic)	0.000000		

Inverted AR Roots .61

(a)

Dependent Variable: LOG(GDP_FRA)
Method: Least Squares
Date: 07/08/09 Time: 17:56
Sample (adjusted): 1952 1992
Included observations: 41 after adjustments
Convergence achieved after 12 iterations
LOG(GDP_FRA)=(C(11)+C(12)*T)*DT1+(C(21)+C(22)*T)*DT2
+[AR(1)=C(1), AR(2)=C(2)]

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(11)	8.327340	0.098475	84.56308	0.0000
C(12)	0.035929	0.004369	8.223822	0.0000
C(21)	9.010104	0.211207	42.65998	0.0000
C(22)	0.012724	0.005679	2.240634	0.0315
C(1)	1.214156	0.161654	7.510837	0.0000
C(2)	-0.330056	0.173660	-1.900584	0.0656

R-squared	0.998026	Mean dependent var	9.093576
Adjusted R-squared	0.997744	S.D. dependent var	0.368947
S.E. of regression	0.017522	Akaike info criterion	-5.116229
Sum squared resid	0.010746	Schwarz criterion	-4.865462
Log likelihood	110.8827	Hannan-Quinn criter.	-5.024913
Durbin-Watson stat	2.000139		

Inverted AR Roots .80 .41

(b)

Figure 1.6 *Statistical results based on (a) an AR(1)_GM of* GDP_FRA *for t>=31, and (b) a two-piece (discontinuous) AR(1)_GM of* GDP_FRA

System: SYS04
Estimation Method: Iterative Least Squares
Date: 07/08/09 Time: 18:20
Sample: 1951 1992
Included observations: 43
Total system (unbalanced) observations 167
Convergence achieved after 12 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	8.675757	0.117154	74.05426	0.0000
C(11)	0.025588	0.003869	6.613016	0.0000
C(12)	0.881066	0.094300	9.343263	0.0000
C(20)	9.080932	0.045963	197.5692	0.0000
C(21)	0.017432	0.001607	10.84940	0.0000
C(22)	0.788836	0.104235	7.567884	0.0000
C(30)	8.327353	0.098426	84.60564	0.0000
C(31)	0.035928	0.004367	8.227148	0.0000
C(32)	9.010076	0.211183	42.66470	0.0000
C(33)	0.012725	0.005678	2.240981	0.0265
C(34)	1.214152	0.161655	7.510770	0.0000
C(35)	-0.330039	0.173659	-1.900506	0.0593
C(40)	8.768484	0.188648	46.48057	0.0000
C(41)	0.020326	0.004748	4.281062	0.0000
C(42)	0.891366	0.035255	25.28374	0.0000
Determinant residual covariance	1.85E-14			

Equation: LOG(GDP_CAN)=C(10)+C(11)*T+[AR(1)=C(12)]
Observations: 42

R-squared	0.991522	Mean dependent var	9.269590
Adjusted R-squared	0.991087	S.D. dependent var	0.342288
S.E. of regression	0.032314	Sum squared resid	0.040724
Durbin-Watson stat	1.510506		

Equation: LOG(GDP_US)=C(20)+C(21)*T+[AR(1)=C(22)]
Observations: 42

R-squared	0.988209	Mean dependent var	9.473482
Adjusted R-squared	0.987604	S.D. dependent var	0.227124
S.E. of regression	0.025287	Sum squared resid	0.024938
Durbin-Watson stat	1.762500		

Equation: LOG(GDP_FRA)=(C(30)+C(31)*T)*DT1+(C(32)+C(33)*T)*DT2
+[AR(1)=C(34),AR(2)=C(35)]
Observations: 41

R-squared	0.998026	Mean dependent var	9.093576
Adjusted R-squared	0.997744	S.D. dependent var	0.368947
S.E. of regression	0.017522	Sum squared resid	0.010746
Durbin-Watson stat	2.000128		

Equation: LOG(GDP_GER)=C(40)+C(41)*T+[AR(1)=C(42)]
Observations: 42

R-squared	0.996821	Mean dependent var	9.098903
Adjusted R-squared	0.996658	S.D. dependent var	0.392866
S.E. of regression	0.022710	Sum squared resid	0.020114
Durbin-Watson stat	1.596264		

Figure 1.7 *Statistical results based on a MAR(1,1,2,1)_GM of* GDP_Can, GDP_US, GDP_Fra *and* GDP_Ger

1.3 Alternative Multivariate Growth Models

As an extension of all the continuous and discontinuous growth models presented in Agung (Agung, 2009a, Chapters 2 and 3), various multivariate growth models can easily be derived. However, only some selected models will be presented in the following sub-sections.

1.3.1 A Generalization of MAR(p)_GM

As an extension of the MAR(***p***)_GM in (1.2), the following growth model is presented.

$$\begin{aligned} \log(Y\_i) &= F_i(t, C(i^*)) + \mu_{it} \\ \mu_{it} &= \rho_{i1}\mu_{i,t-1} + \rho_{i,t-2}\mu_{i,t-2} + \ldots + \rho_{i,t-p_i}\mu_{i,t-p_i} \end{aligned} \tag{1.3}$$

where $F_i(t,C(i^*))$ can be any functions of t, such as the polynomial and the natural logarithmic of t, either continuous or discontinuous functions, as well as nonlinear with a finite number of parameters, namely $C(i^*)$, for each $i=1,\ldots,N$. Note that any continuous and discontinuous growth models in Agung (2009a) could be inserted for the function $F_i(t,C(i^*))$. For example, as follows:

1.3.1.1 A Polynomial Growth Model

The independent-states system of polynomial growth models has the following equation for $i=1,\ldots,N$.

$$\log(Y\_i_t) = c(i0) + c(i1)^* t + \ldots + c(ik_i)^* t^{k_i} + \mu_{it} \tag{1.4a}$$

1.3.1.2 A Translog Linear Model

The independent-states system of the translog linear growth models has the following equation for $i=1,\ldots,N$.

$$\log(Y\_i_t) = c(i0) + c(i1)^* \log(t) + \mu_{it} \tag{1.4b}$$

1.3.1.3 The Simplest Nonlinear Growth Model

The independent-states system of the simplest nonlinear growth model has the following equation for $i = 1, \ldots, N$.

$$\log(Y\_i_t) = c(i0) + c(i1)^* t^{c(i2)} + \mu_{it} \tag{1.4c}$$

1.3.1.4 A Two-Piece Growth Model

The independent-states system of two-piece growth models has the following equation for $i = 1, \ldots, N$.

$$\log(Y\_i_t) = (c(i10) + c(i11)^* t)^* Dt1 + (c(i20) + c(i21)^* t)^* Dt2 + \mu_{it} \tag{1.4d}$$

where *Dt1* and *Dt2* are dummy variables of two time periods considered, such as $t < t_0$ and $t \geq t_0$, which are defined to be valid for all states or individuals, or all $i = 1, \ldots, N$. To generalize, the dummy variables would be dependent on *i*, namely *Dt(i1)* and *Dt(i2)*, the model can easily be extended to three or more time periods, and the linear function of *t* within each time period could be replaced by other functions of *t*. For further illustration, refer to various discontinuous growth models presented in Chapter 3, (Agung, 2009a), specifically the multivariate models by states and time periods in the general models (3.79) to (3.87).

1.3.2 Multivariate Lagged Variables Growth Models

Corresponding to the MAR(***p***)_GM in (1.2), a multivariate lagged variables growth model, namely MLV(***q***)_GM, may be considered an alternative growth model with the following general equation, where the error terms should also be assumed or accepted in a theoretical sense to have an $i.i.d.N(0, \sigma^2)$.

$$\log(Y\_i_t) = C(i0) + \sum_{j=1}^{q_i} C(ij)^* \log(Y\_i_{t-j}) + C(i, q_i + 1)^* t + \varepsilon_{it} \tag{1.5}$$

Note that the lag variable $log(Y\_i_{t-j})$ is not a cause factor of $log(Y\_i_t)$, but is an up-stream or a predictor variable. Also, the exogenous variables, namely X_i_t and X_i_{t-j}, used in most models are not really the true cause factors of the dependent variable of these models. See the models presented in Section 1.4.

All lagged variables and autoregressive models, in fact, are *dynamic models* (Gujarati, 2003, Gourierroux and Manfort, 1997, Hamilton, 1994, and Kmenta, 1986). Therefore, various models in (1.5) should be considered as *multivariate dynamic growth models* (MDGM), or multivariate *dynamic models with trend*, for $i = 1, \ldots, N$. Wooldridge (2002; 493) presents another type of dynamic model, called *dynamic unobserved effects models*.

Example 1.6 A MLV(1)_GM of GDP? in Figure 1.7

As an alternative multivariate growth model of GDP in Figure 1.7, Figure 1.8 presents the statistical results based on an MLV(1)_GM, where the regression of *GDP_Fra* is a two-piece LV(1)_GM. Based on the results in Figures 1.7 and 1.8, the following findings and notes are presented.

1. The estimates of the parameter C(12) in both models have exactly the same values of 0.881 066, which indicates the first-order autocorrelation of *log(GDP_Can)*. Similarly for the parameters *C*(22) and C(42), respectively, there is first-order autocorrelation of *log(GDP_US)* and *log(GDP_Ger)*.

System: SYS05_LAG
Estimation Method: Least Squares
Date: 08/21/09 Time: 15:10
Sample: 1951 1992
Included observations: 42
Total system (balanced) observations 168

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	1.057433	0.815685	1.296374	0.1968
C(11)	0.003043	0.002651	1.147831	0.2528
C(12)	0.881066	0.094300	9.343263	0.0000
C(20)	1.934999	0.944323	2.049086	0.0422
C(21)	0.003681	0.001951	1.886825	0.0611
C(22)	0.788836	0.104235	7.567884	0.0000
C(310)	-0.116130	0.793076	-0.146430	0.8838
C(311)	-0.001504	0.003765	-0.399557	0.6900
C(312)	1.019645	0.095786	10.64503	0.0000
C(320)	4.084122	2.518689	1.621527	0.1070
C(321)	0.010357	0.005974	1.733656	0.0850
C(322)	0.529470	0.288767	1.833553	0.0687
C(40)	0.972882	0.295011	3.297777	0.0012
C(41)	0.002208	0.001180	1.871744	0.0632
C(42)	0.891366	0.035255	25.28374	0.0000
Determinant residual covariance		2.04E-14		

Equation: LOG(GDP_CAN)=C(10)+C(11)*T+C(12)*LOG(GDP_CAN(-1))
Observations: 42

R-squared	0.991522	Mean dependent var	9.269590
Adjusted R-squared	0.991087	S.D. dependent var	0.342288
S.E. of regression	0.032314	Sum squared resid	0.040724
Durbin-Watson stat	1.510506		

Equation: LOG(GDP_US)=C(20)+C(21)*T+C(22)*LOG(GDP_US(-1))
Observations: 42

R-squared	0.988209	Mean dependent var	9.473482
Adjusted R-squared	0.987604	S.D. dependent var	0.227124
S.E. of regression	0.025287	Sum squared resid	0.024938
Durbin-Watson stat	1.762500		

Equation: LOG(GDP_FRA)=(C(310)+C(311)*T+C(312)*LOG(GDP_FRA(-1)))
*DT1+(C(320)+C(321)*T+C(322)*LOG(GDP_FRA(-1)))*DT2
Observations: 42

R-squared	0.997960	Mean dependent var	9.076151
Adjusted R-squared	0.997676	S.D. dependent var	0.381517
S.E. of regression	0.018391	Sum squared resid	0.012176
Durbin-Watson stat	1.557845		

Equation: LOG(GDP_GER)=C(40)+C(41)*T+C(42)*LOG(GDP_GER(-1))
Observations: 42

R-squared	0.996821	Mean dependent var	9.098903
Adjusted R-squared	0.996658	S.D. dependent var	0.392866
S.E. of regression	0.022710	Sum squared resid	0.020114
Durbin-Watson stat	1.596264		

Figure 1.8 *Statistical results based on a MLV(**1**)_GM of* GDP_Can, GDP_US, GDP_Fra *and* GDP_Ger

2. Corresponding to the regression model of *log(GDP_Fra)*, Figures 1.7 and 1.8 present different types of two-piece growth models. Figure 1.7 presents a two-piece AR(2)_GM, where the autocorrelation of the error terms AR(1) and AR(2) should be valid for the whole time period. On the other hand, Figure 1.8 presents a two-piece LV(1)_GM, where $\hat{C}(312) = 1.019645$ is the AR(1) of *log(GDP_Fra)* for $t < 31$, and $\hat{C}(312) = 0.382358$ is its AR(1) for $t \geq 31$.
3. However, Figure 1.8 presents a negative adjusted growth rate of log(GDP_Fra), for $t < 31$, namely $\hat{C}(311) = -0.001504$ which is an inappropriate estimate. For this reason, the statistical results based on a MLV(1,1,2,1) are presented in Figure 1.9, where the two-piece regressions of *log(GDP_Fra)* is

Estimation Method: Least Squares
Date: 08/21/09 Time: 15:21
Sample: 1951 1992
Included observations: 42
Total system (unbalanced) observations 167

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	1.057433	0.815685	1.296374	0.1968
C(11)	0.003043	0.002651	1.147831	0.2529
C(12)	0.881066	0.094300	9.343263	0.0000
C(20)	1.934999	0.944323	2.049086	0.0422
C(21)	0.003681	0.001951	1.886825	0.0611
C(22)	0.788836	0.104235	7.567884	0.0000
C(310)	0.300243	0.872774	0.344011	0.7313
C(311)	0.000601	0.004205	0.142842	0.8866
C(312)	1.221207	0.192555	6.342122	0.0000
C(313)	-0.253122	0.209560	-1.207873	0.2290
C(320)	3.984467	2.545501	1.565297	0.1196
C(321)	0.009331	0.006145	1.518485	0.1310
C(322)	0.876620	0.492666	1.779340	0.0772
C(323)	-0.333157	0.381130	-0.874129	0.3834
C(40)	0.972882	0.295011	3.297777	0.0012
C(41)	0.002208	0.001180	1.871744	0.0632
C(42)	0.891366	0.035255	25.28374	0.0000
Determinant residual covariance		1.96E-14		

Equation: LOG(GDP_CAN)=C(10)+C(11)*T+C(12)*LOG(GDP_CAN(-1))
Observations: 42

R-squared	0.991522	Mean dependent var	9.269590
Adjusted R-squared	0.991087	S.D. dependent var	0.342288
S.E. of regression	0.032314	Sum squared resid	0.040724
Durbin-Watson stat	1.510506		

Equation: LOG(GDP_US)=C(20)+C(21)*T+C(22)*LOG(GDP_US(-1))
Observations: 42

R-squared	0.988209	Mean dependent var	9.473482
Adjusted R-squared	0.987604	S.D. dependent var	0.227124
S.E. of regression	0.025287	Sum squared resid	0.024938
Durbin-Watson stat	1.762500		

Equation: LOG(GDP_FRA)=(C(310)+C(311)*T+C(312)*LOG(GDP_FRA(-1))
+C(313)*LOG(GDP_FRA(-2)))*DT1+(C(320)+C(321)*T+C(322)
*LOG(GDP_FRA(-1))+C(323)*LOG(GDP_FRA(-2)))*DT2
Observations: 41

R-squared	0.997910	Mean dependent var	9.093576
Adjusted R-squared	0.997467	S.D. dependent var	0.368947
S.E. of regression	0.018568	Sum squared resid	0.011378
Durbin-Watson stat	1.954962		

Equation: LOG(GDP_GER)=C(40)+C(41)*T+C(42)*LOG(GDP_GER(-1))
Observations: 42

R-squared	0.996821	Mean dependent var	9.098903
Adjusted R-squared	0.996658	S.D. dependent var	0.392866
S.E. of regression	0.022710	Sum squared resid	0.020114
Durbin-Watson stat	1.596264		

Figure 1.9 *Statistical results based on a MLV(1,1,2,1)_GM of* GDP_Can, GDP_US, GDP_Fra *and* GDP_Ger

LV(2)_GM, and the two regressions represent positive growth rates of *GDP_Fra*, namely $\hat{C}(311) = 0.000601$ and $\hat{C}(321) = 0.009331$, respectively, for $t < 31$ and $t \geq 31$, adjusted for *log(GDP_Fra(−1))* and *log(GDP_Fra(−2))*.

1.3.3 Multivariate Lagged-Variable Autoregressive Growth Models

As an extension of LVAR(1,1)_GM presented in Agung (2009a), data analysis based on a multivariate lagged variables autoregressive model, MLVAR(***p;q***)_GM, where $\boldsymbol{p} = (p_i)$ and $\boldsymbol{q} = (q_i)$, of the time series Y_i_t, for $i = 1, \ldots, N$, will have the following general equation.

$$\begin{aligned} \log(Y\_i_t) &= C(i0) + \sum_{j=1}^{p_i} C(ij)^* \log(Y\_i_{t-j}) + C(i, p_i + 1)^* t + \mu_{it} \\ \mu_{it} &= \sum_{k=1}^{q_i} \rho_{ik} \mu_{i,t-k} + \varepsilon_{it} \end{aligned} \tag{1.6}$$

Note that for $\boldsymbol{q} = 0$, the MLV(***p***)_GM will be obtained, and the MAR(***q***)_GM obtained for $\boldsymbol{p} = 0$. Various special cases would be obtained, where $p_i = p$ and $q_i = q$ for all $i = 1, \ldots, N$.

1.3.4 Bounded MLVAR(*p;q*)_GM

As an extension of the general MLVAR(***p;q***)_GM in (1.6) as well as the bounded growth model presented in Agung (2009a), the bounded MLVAR(***p;q***)_GM, of the time series Y_i_t, $i = 1, 2, \ldots, N$, has the following general equation.

$$\begin{aligned} \log\left(\frac{Y\_i - Li}{Ui - Y\_i}\right) &= C(i0) + \sum_{j=1}^{p_i} C(ij)^* \log(Y\_i_{t-j}) + C(i, p_i + 1)^* t + \mu_{it} \\ \mu_{it} &= \sum_{k=1}^{q_i} \rho_{ik} \mu_{i,t-k} + \varepsilon_{it} \end{aligned} \tag{1.7}$$

where *Li* and *Ui* are the lower and upper bounds of *Y_i*, which are theoretically selected fixed numbers.

1.3.5 Special Notes

Based on the findings presented previously, the following special notes are presented.

1. Unexpected parameter estimates can be obtained by using autoregressive or lagged variables growth models. In general, by inserting an additional independent variable to a model, we can never predict its impact on the parameter estimates. Refer to the special notes in Agung (2009a, Section 2.14.2). For this reason, one should use the trial-and-error method to develop several acceptable growth models, in both theoretical and statistical senses. Note that this statement also should be applicable for any statistical model.
2. Graphic representations between each of the independent variables and the corresponding dependent variable should be analyzed to evaluate their possible patterns of relationship. Specifically, whether a linear or non-linear model would acceptable. Refer to Chapter 1 in Agung (2009a).

3. Furthermore, residuals analysis should be done to identify the limitation of each model. Refer to Agung (2009a).
4. Corresponding to the relationship between Y_i_t and its lag Y_i_{t-s}, specific for $s=1$, $s=4$, and $s=12$, respectively, if and only if the time series data are annually, quarterly, and monthly data sets, the following notes are presented.
   4.1 The observed values of Y_i_{t-s} and Y_i_t can be considered as the observations before and after a natural-experiment for the *i*-th individual, with a set of environmental variables could be the treatment or experimental factors, namely $\mathbf{Z}_t = (Z1, Z2, \ldots, Zk)_t$. Refer to Section 1.8.
   4.2 The lag variable Y_i_{t-s} should be considered as a *covariate* in any time series models having Y_i_t as the dependent variable. So the "*classical growth model of Y_i_t with a covariate Y_i_{t-s}*" for the *i*-th individual may have the following alternative equations.

$$\log(Y\_i_t) = C(i0) + C(i1)^*t + C(i2)^*Y\_i_{t-s} + \mu_{it} \tag{1.8a}$$

$$\log(Y\_i_t) = C(i0) + C(i1)^*t + C(i2)^*\log(Y\_i_{t-s}) + \mu_{it} \tag{1.8b}$$

1.4 Various Models Based on Correlated States

It is known that stock prices of selected countries have a causal relationship. In this section, as an extension of the previously mentioned models, I consider the models based on correlated states. The definition is that two states are correlated if, and only if, their endogenous variables have a causal relationship. Note that if all variables are assumed or defined to be correlated, then all the time series models presented in Agung (2009a) can easily be applied.

With regards to the time series data by states or unstacked data considered, it is acceptable that growth of a problem indicator or variable of a state (country, region, firm or agency) should be theoretically influenced by the factors of the other state(s). For illustrative examples, at the first stage the *GDP* of two states, namely *GDP_US*, and *GDP_Can* in POOLG7.wf1, are defined to have a causal relationship. Here, two alternative causal relationships are considered, as presented in Figure 1.10, out of a lot of possible models.

Note that Figure 1.10(a) presents the path diagram where *GDP_US* is defined as the cause factor of *GDP_Can*. Based on this path diagram, the *simplest causal model* with trend would have the following system specification.

$$\begin{aligned} GDP\_US &= C(10) + C(11)^*GDP\_US(-1) + C(12)^*t \\ GDP\_Can &= C(20) + C(21)^*GDP\_Can(-1) + C(22)^*GDP\_US \\ &\quad + C(23)^*GDP\_US(-1) + C(24)^*t \end{aligned} \tag{1.9}$$

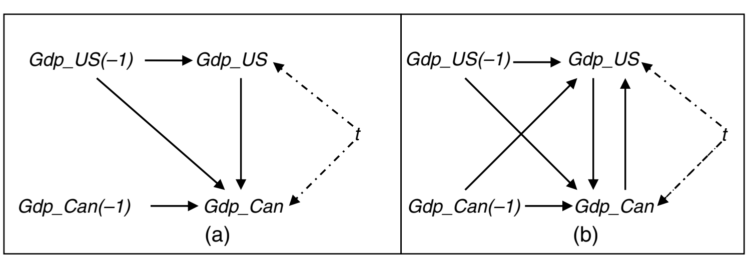


Figure 1.10 *Two alternative causal relationships between* GDP_US *and* GDP_Can

Note that each multiple regression in the model is an additive regression model of its independent variables. For instance, the first regression is an additive model of *GDP_US* on *GDP_US(−1)*, and the time *t*. In other words, this model represents the linear adjusted effects of *GDP_US(−1)*, and the time *t* on *GDP_US*. In fact, there are a lot more models that could be subjectively defined by the researchers. Refer to various time series models presented in Agung (2009a).

On the other hand, Figure 1.10(b) presents the path diagram where *GDP_US* and *GDP_Can* are defined to have a simultaneous causal linear effects or two-way causal effects. Based on this path diagram, the simplest causal model with trend would have the following system specification.

$$\begin{aligned} GDP\_US &= C(10) + C(11)^*GDP\_US(-1) + C(12)^*GDP\_Can \\ &\quad + C(13)^*GDP\_Can(-1) + C(14)^*t \\ GDP\_Can &= C(20) + C(21)^*GDP\_Can(-1) + C(22)^*GDP\_US \\ &\quad + C(23)^*GDP\_US(-1) + C(24)^*t \end{aligned} \tag{1.10}$$

Note that the models (1.9) and (1.10) are not the VAR (Vector Autoregressive) models, since they do not have the same set of independent variables. For this reason, Agung (2009a) has introduced the MAR (Multivariate Autoregressive Model) and the SCM (Seemingly Causal Model) instead of the System Equation Model (SEM), because the term SEM is already used for the structural equation model. The following subsections present empirical examples of SCM and VAR Models.

To generalize, a problem indicator by states may be presented as *Y_s*, for the states $s = 1, \ldots, S$. Then, the relationship between the indicators *Y_s*, $s = 1, \ldots, S$, would be a matter of subjective or expert judgment by the researchers. It could be very difficult to define the path diagram of an SCM based on the *GDP* of the seven states as presented in POOLG7.wf1, even more so for the number of states greater than seven. For this reason, I recommend to all students planning to write theses or dissertations, select only two or three states for the data analysis, since they can apply various MLVAR(*p,q*) models and study the limitations of each model using residual analysis. Note that with a single variable *Y*, one would have to consider the variable *Y*, the time *t*-variable and the categorical state variable, as well as the lagged of *Y*, say *Y(−1), . . . ,Y(−p)* for a selected integer *p*, as well as the indicators *AR(1), . . . ,AR(q)*.

1.4.1 Seemingly Causal Models with Trend

For illustrative purposes, Figure 1.10 presents two alternative theoretically defined SCMs between *GDP_US* and *GDP_Can*. Note that the arrows with dotted lines from the time *t* indicate that this is not a real causal factor. However, the following example presents data analysis based on the model (1.10) only.

Example 1.7 SCMs with trend

Figure 1.11(a) presents statistical results based on a bivariate first-order lagged-variable SCM, namely LV(**1**)_SCM, of *GDP_US* and *GDP_Can*, which show that the error terms of each regression have the first autocorrelation problem, indicated by the small value of its Durbin–Watson statistic. For this reason, Figure 1.11(b) presents statistical results based on its AR(**1**) model, namely LVAR(**1,1**)_SCM, which is acceptable, in both theoretical and statistical senses. Note that these models are not growth models. Based on this output, the following conclusions are derived.

1. The *p*-value = 0.0000 of the parameter *C(12)* in the first regression indicates that *GDP_Can* has a significant positive adjusted linear effect on *GDP_US*, and the *p*-value = 0.0000 of the parameter *C(22)* in the second regression indicates that *GDP_US* also has a significant positive adjusted linear effect on *GDP_Can*.

System: UNTITLED
Estimation Method: Least Squares
Date: 08/13/09 Time: 10:33
Sample: 1951 1992
Included observations: 42
Total system (balanced) observations 84

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	2340.532	678.8094	3.447996	0.0009
C(11)	0.703819	0.097711	7.203047	0.0000
C(12)	0.580407	0.100906	5.751936	0.0000
C(13)	-0.554104	0.099596	-5.563521	0.0000
C(14)	62.09836	20.53535	3.023974	0.0034
C(20)	-1888.823	869.9183	-2.171264	0.0331
C(21)	0.873552	0.070032	12.47358	0.0000
C(22)	0.813339	0.141403	5.751936	0.0000
C(23)	-0.500667	0.159265	-3.143598	0.0024
C(24)	-35.34991	26.51819	-1.333044	0.1866

Determinant residual covariance 1.78E+09

Equation: GDP_US = C(10)+C(11)*GDP_US(-1)+C(12)*GDP_CAN+C(13)
*GDP_CAN(-1)+C(14)*T
Observations: 42

R-squared	0.994199	Mean dependent var	13335.29
Adjusted R-squared	0.993572	S.D. dependent var	2944.502
S.E. of regression	236.0718	Sum squared resid	2062006.
Durbin-Watson stat	1.432991		

Equation: GDP_CAN = C(20)+C(21)*GDP_CAN(-1)+C(22)*GDP_US+C(23)
*GDP_US(-1)+C(24)*T
Observations: 42

R-squared	0.995001	Mean dependent var	11224.90
Adjusted R-squared	0.994460	S.D. dependent var	3754.577
S.E. of regression	279.4560	Sum squared resid	2889540.
Durbin-Watson stat	1.229405		

(a)

System: UNTITLED
Estimation Method: Iterative Least Squares
Date: 08/13/09 Time: 10:36
Sample: 1952 1992
Included observations: 42
Total system (balanced) observations 82
Convergence achieved after 8 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	3705.368	1399.393	2.647840	0.0100
C(11)	0.513887	0.213813	2.403435	0.0189
C(12)	0.598526	0.099965	5.987379	0.0000
C(13)	-0.556744	0.120949	-4.603130	0.0000
C(14)	105.4857	39.25613	2.687115	0.0090
C(15)	0.432322	0.254010	1.701991	0.0932
C(20)	-2473.391	1294.467	-1.910741	0.0601
C(21)	0.769470	0.193263	3.981463	0.0002
C(22)	0.845435	0.141269	5.984574	0.0000
C(23)	-0.395862	0.221602	-1.786366	0.0784
C(24)	-38.84161	43.44929	-0.893953	0.3744
C(25)	0.477109	0.264250	1.805520	0.0753

Determinant residual covariance 1.21E+09

Equation: GDP_US = C(10)+C(11)*GDP_US(-1)+C(12)*GDP_CAN+C(13)
*GDP_CAN(-1)+C(14)*T+[AR(1)=C(15)]
Observations: 41

R-squared	0.994937	Mean dependent var	13437.80
Adjusted R-squared	0.994214	S.D. dependent var	2904.200
S.E. of regression	220.9119	Sum squared resid	1708072.
Durbin-Watson stat	1.921317		

Equation: GDP_CAN = C(20)+C(21)*GDP_CAN(-1)+C(22)*GDP_US+C(23)
*GDP_US(-1)+C(24)*T+[AR(1)=C(25)]
Observations: 41

R-squared	0.995644	Mean dependent var	11342.95
Adjusted R-squared	0.995022	S.D. dependent var	3721.474
S.E. of regression	262.5740	Sum squared resid	2413079.
Durbin-Watson stat	1.876369		

(b)

Figure 1.11 *Statistical results based on the models with trends in (1.10), namely (a) a LV(**1**)_SCM, and (b) a LVAR(**1,1**)_SCM*

2. Therefore, based on the SCM in Figure 1.11(b), it can be concluded that the data supports the hypothesis that *GDP_US* and *GDP_Can* have simultaneous causal linear effects, adjusted for the other independent variables in the model.
3. In order to conduct the unadjusted simultaneous causal effects, the bivariate correlation analysis can easily be applied (Agung, 2006, 2009a, 2011). In this case, $H_0{:}\rho(GDP\_US, GDP\_Can) = 0$ is rejected based on the *t*-statistic of $t_0 = 36.43270$ with a *p*-value = 0.0000.
4. Various univariate and multivariate hypotheses could easily be tested using the Wald test.

Example 1.8 Translog linear SCMs with trend

The alternative models of those in Figure 1.11, Figure 1.12(a) and (b) present statistical results based on a translog linear LV(**1**)_SCM, and LVAR(**1,1**)_SCM. Based on these results the following notes are presented.

1. The translog linear LVAR(1,1)_SCM is an unacceptable model, in a statistical sense, based on the data set used, since the AR(1) of both regressions are insignificant with such a large *p*-values of 0.82 and 0.42, respectively.
2. In this case, the translog linear LV(**1**)_SCM would be a better model, supported by the fact that each independent variable has a significant adjusted effect on its corresponding dependent variable with sufficiently large DW statistics and their residual graphs, as shown in Figure 1.13. It would not be the best out of all possible models, which have not been explored.
3. Note that this translog-linear LV(**1**)_SCM can be viewed as a bivariate growth model, where *C(14)* indicates the adjusted exponential growth rate of *GDP_US*, and *C(24)* indicates the adjusted exponential growth rate of *GDP_Can*.

System: SYS02
Estimation Method: Least Squares
Date: 08/13/09 Time: 10:53
Sample: 1951 1992
Included observations: 42
Total system (balanced) observations 84

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	2.609122	0.648857	4.021103	0.0001
C(11)	0.735920	0.086255	8.531956	0.0000
C(12)	0.572822	0.088428	6.477804	0.0000
C(13)	-0.597241	0.089866	-6.645922	0.0000
C(14)	0.005485	0.001553	3.532318	0.0007
C(20)	-2.724012	0.882777	-3.085729	0.0029
C(21)	0.916529	0.077391	11.84283	0.0000
C(22)	0.927723	0.143216	6.477805	0.0000
C(23)	-0.546423	0.166390	-3.283987	0.0016
C(24)	-0.004779	0.002146	-2.227028	0.0290

Determinant residual covariance 4.96E-08

Equation: LOG(GDP_US) = C(10)+C(11)*LOG(GDP_US(-1))+C(12)
*LOG(GDP_CAN)+C(13)*LOG(GDP_CAN(-1))+C(14)*T
Observations: 42

R-squared	0.994925	Mean dependent var	9.473482
Adjusted R-squared	0.994376	S.D. dependent var	0.227124
S.E. of regression	0.017033	Sum squared resid	0.010735
Durbin-Watson stat	1.814607		

Equation: LOG(GDP_CAN) = C(20)+C(21)*LOG(GDP_CAN(-1))+C(22)
*LOG(GDP_US)+C(23)*LOG(GDP_US(-1))+C(24)*T
Observations: 42

R-squared	0.996381	Mean dependent var	9.269590
Adjusted R-squared	0.995990	S.D. dependent var	0.342288
S.E. of regression	0.021677	Sum squared resid	0.017385
Durbin-Watson stat	1.688250		

(a)

System: SYS03
Estimation Method: Iterative Least Squares
Date: 08/13/09 Time: 10:57
Sample: 1952 1992
Included observations: 42
Total system (balanced) observations 82
Convergence achieved after 6 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	2.670910	0.721082	3.704031	0.0004
C(11)	0.751711	0.101318	7.419300	0.0000
C(12)	0.561529	0.087604	6.409872	0.0000
C(13)	-0.609972	0.090577	-6.734285	0.0000
C(14)	0.005971	0.001652	3.613867	0.0006
C(15)	0.042296	0.185503	0.228008	0.8203
C(20)	-2.799240	1.021972	-2.739057	0.0078
C(21)	0.903637	0.108494	8.328919	0.0000
C(22)	0.951495	0.149107	6.381282	0.0000
C(23)	-0.549361	0.186524	-2.945253	0.0044
C(24)	-0.004892	0.002609	-1.875427	0.0649
C(25)	0.157224	0.195886	0.802631	0.4249

Determinant residual covariance 4.56E-08

Equation: LOG(GDP_US) = C(10)+C(11)*LOG(GDP_US(-1))+C(12)
*LOG(GDP_CAN)+C(13)*LOG(GDP_CAN(-1))+C(14)*T+[AR(1)=C(15)]
Observations: 41

R-squared	0.995076	Mean dependent var	9.482115
Adjusted R-squared	0.994373	S.D. dependent var	0.222860
S.E. of regression	0.016718	Sum squared resid	0.009782
Durbin-Watson stat	1.844620		

Equation: LOG(GDP_CAN) = C(20)+C(21)*LOG(GDP_CAN(-1))+C(22)
*LOG(GDP_US)+C(23)*LOG(GDP_US(-1))+C(24)*T+[AR(1)=C(25)]
Observations: 41

R-squared	0.996343	Mean dependent var	9.281978
Adjusted R-squared	0.995821	S.D. dependent var	0.336874
S.E. of regression	0.021778	Sum squared resid	0.016599
Durbin-Watson stat	1.773246		

(b)

Figure 1.12 *Statistical results based on (a) translog linear LV(**1**)_SCM, and (b) translog linear LVAR(**1,1**)_SCM; with trend*

4. To generalize, the variable *GDP* could easily be replaced by a variable *Y*. Then, as a modified model, the SCM will be written as pair of nonlinear models as follows:

$$\begin{aligned} Y\_US &= Y\_US(-1)^{C(11)} Y\_Can^{C(12)} Y\_Can(-1)^{C(13)} Exp(C(10) + C(14)^{*} t) \\ Y\_Can &= Y\_Can(-1)^{C(21)} Y\_US^{C(22)} Y\_US(-1)^{C(23)} Exp(C(20) + C(24)^{*} t) \end{aligned} \quad (1.11)$$

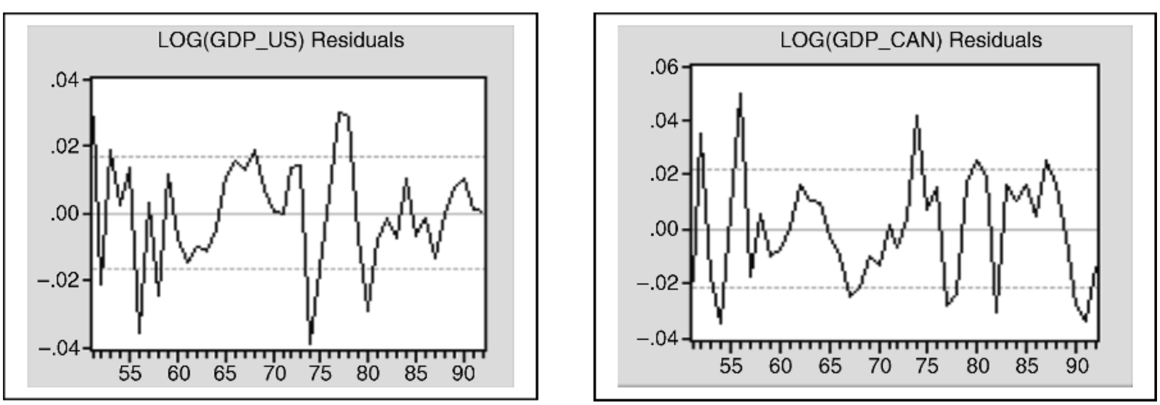


Figure 1.13 *Residual graphs of the LV(**1**)_SCM in Figure 1.11(a)*

5. Finally, for the seven states in POOLG7.wf1, a much more complex path diagram should be developed or defined to represent the theoretical causal model between the seven time series variables. Therefore, based on the path diagram, the system specification of a SCM would easily be written, either as translog linear or nonlinear models. However, by using many independent variables, the error messages of "*near singular matrix*" or "*overflow*", as well as the unexpected estimates of parameters, may be obtained. Refer to special notes in Agung (2009a, Section 2.14).

1.4.2 The Application of the Object "VAR"

EViews provides the object "*VAR*" for conducting the data analysis based on a *vector autoregressive* (VAR) and *vector error correction* (VEC) models, which are special cases of the *multivariate autoregressive* (MAR) models and SCMs, (Agung, 2009a). See the following example.

Example 1.9 A VAR model

Figure 1.14 presents the statistical result based on a VAR model of *log(GDP_US)* and *log(GDP_Can)*, using the lag interval of endogenous "1 1", and exogenous variables "*C T*" with the default options. Based on this result the following notes are presented.

1. Note that this VAR model in fact is a special case of the MAR(1)_GM, where all regressions have exactly the same independent variables. Compared to the path diagram in Figure 1.10, the path diagram of this VAR model presented in Figure 1.15 shows the causal relationship between *log(GDP_US)* and *log(GDP_Can)* is not taken into account.
2. However, the quantitative coefficient of correlations of the independent variables *log(GDP_US(−1))*, *log(GDP_Can)* and the time *t* should be taken into account in the regression analysis, and it is well-known that they have an unpredictable impact on the estimate of the model parameters. Refer to Section 2.14.2 in Agung (2009a).

Vector Autoregression Estimates
Date: 08/13/09 Time: 12:31
Sample (adjusted): 1951 1992
Included observations: 42 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	LOG(GDP_US)	LOG(GDP_...
LOG(GDP_US(-1))	0.902551	0.290894
	(0.11868)	(0.15103)
	[7.60498]	[1.92602]
LOG(GDP_CAN(-1))	-0.154154	0.773517
	(0.08402)	(0.10692)
	[-1.83475]	[7.23425]
C	2.238143	-0.647635
	(0.93168)	(1.18568)
	[2.40226]	[-0.54621]
T	0.005864	0.000661
	(0.00224)	(0.00285)
	[2.62140]	[0.23215]

(a)

R-squared	0.989169	0.992276
Adj. R-squared	0.988313	0.991666
Sum sq. resids	0.022909	0.037102
S.E. equation	0.024553	0.031247
F-statistic	1156.763	1627.292
Log likelihood	98.19675	88.07141
Akaike AIC	-4.485560	-4.003401
Schwarz SC	-4.320067	-3.837908
Mean dependent	9.473482	9.269590
S.D. dependent	0.227124	0.342288

Determinant resid covariance (dof adj.)	2.76E-07
Determinant resid covariance	2.26E-07
Log likelihood	202.1872
Akaike information criterion	-9.247008
Schwarz criterion	-8.916024

```
Estimation Proc:
===============================
LS 1 1 LOG(GDP_US) LOG(GDP_CAN) @ C T

VAR Model:
===============================
LOG(GDP_US) = C(1,1)*LOG(GDP_US(-1)) + C(1,2)*LOG(GDP_CAN(-1)) +
C(1,3) + C(1,4)*T

LOG(GDP_CAN) = C(2,1)*LOG(GDP_US(-1)) + C(2,2)*LOG(GDP_CAN(-1)) +
C(2,3) + C(2,4)*T
```

(b)

Figure 1.14 *Statistical results based on a VAR Model of* log(GDP_US) *and* log(GDP_Can)

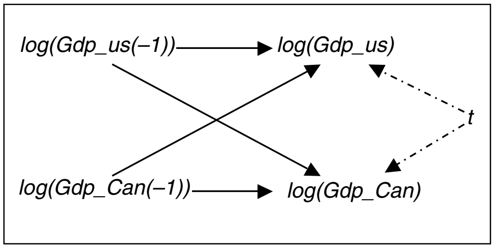


Figure 1.15 *The path diagram of the VAR model in Figure 1.14*

3. Furthermore, note that the symbol *C(i,j)* is used to present the model parameters, and test the hypotheses using the Wald test, that is the *Block Exogeinity Wald Test*.
4. For a model with many endogenous or exogenous variables, applying the object "*System*" is recommended instead of the object "*VAR*", because in general, the good fit multiple regressions in the model would have different sets of independent variables.
5. Refer to Chapter 6 in Agung, (2009a; p. 316), for more detailed notes on various VAR models, as well as their residual analysis, and a special causality test; the VAR Granger Causality/Block Exogeneity Wald Tests.
6. Furthermore, in order to match conditions in previous and recent years, Agung (2009a) proposes special VAR models using the lag interval of endogenous "4 4" for a quarterly data set, and "12 12" for a monthly data set.

Example 1.10 A vector error correction (VEC) model

Figure 1.16 presents the statistical result based on a VEC model by inserting the endogenous variables "*log(GDP_US) log(GDP_Can)*", the lag interval of endogenous "1 1", and exogenous variables "*T*" with the default options. Based on this result the following notes are presented.

1. By inserting the endogenous variables *log(GDP_US)* and *log(GDP_Can)*, the output directly presents two regressions with the first differences *Dlog(GDP_US)* and *Dlog(GDP_Can)* as their dependent variables.
2. Note that, in general, the first difference of *log(Y_s)*, in fact indicates the exponential growth rate of *Y_s*, which can be presented as follows:

$$D\log(Y\_s) = \log(Y\_s_t) - \log(Y\_s_{t-1}) = R_t(Y\_s) \tag{1.12}$$

Then, the two independent variables Dlog(Y_US(−1)) and Dlog(Y_Can(−1)), can be presented as follows:

$$D\log(Y\_s(-1)) = \log(Y\_s_{t-1}) - \log(Y\_s_{t-2}) = R_{t-1}(Y\_s) \tag{1.13}$$

For these reasons, the VEC model in fact presents a bivariate LV(1) model of $R_t(GDP\_US)$ and $R_t(GDP\_Can)$ with exogenous variables.

3. Beside the independent or exogenous variables *C* and *T*, both regressions in the VEC model have a special independent variable, called the *Cointegrating Equation*, namely:

$$CointEg1 = \log(GDP\_US(-1)) + 0.286109 \log(GDP\_Can) - 12.1466$$

4. For more detailed notes on various VEC models, as well as the characteristics of alternative cointegrating equations, refer to Section 6.3 in Agung (2009a).

Vector Error Correction Estimates
Date: 08/13/09 Time: 14:48
Sample (adjusted): 1952 1992
Included observations: 41 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1	
LOG(GDP_US(-1))	1.000000	
LOG(GDP_CAN(-1))	0.286109	
	(0.26786)	
	[1.06813]	
C	-12.11466	
Error Correction:	**D(LOG(GDP...**	**D(LOG(GDP...**
CointEq1	-0.229705	-0.041189
	(0.10087)	(0.13528)
	[-2.27722]	[-0.30447]
D(LOG(GDP_US(-1)))	0.118220	0.203617
	(0.20449)	(0.27426)
	[0.57812]	[0.74243]
D(LOG(GDP_CAN(-1)))	0.013610	0.107740
	(0.19289)	(0.25869)
	[0.07056]	[0.41648]
C	-0.117662	-0.002766
	(0.06156)	(0.08256)
	[-1.91131]	[-0.03350]
T	0.005992	0.000896
	(0.00270)	(0.00362)
	[2.21965]	[0.24751]

R-squared	0.163536	0.060650
Adj. R-squared	0.070596	-0.043723
Sum sq. resids	0.022341	0.040184
S.E. equation	0.024911	0.033410
F-statistic	1.759578	0.581090
Log likelihood	95.87921	83.84437
Akaike AIC	-4.433132	-3.846067
Schwarz SC	-4.224160	-3.637095
Mean dependent	0.016525	0.023027
S.D. dependent	0.025840	0.032703
Determinant resid covariance (dof adj.)		3.02E-07
Determinant resid covariance		2.33E-07
Log likelihood		196.7108
Akaike information criterion		-9.010285
Schwarz criterion		-8.508751

```
Estimation Proc:
===============================
EC(C,1) 1 1 LOG(GDP_US) LOG(GDP_CAN) @T

VAR Model:
===============================
D(LOG(GDP_US)) = A(1,1)*(B(1,1)*LOG(GDP_US(-1)) + B(1,2)*LOG(GDP_CAN(-1)) + B
(1,3)) + C(1,1)*D(LOG(GDP_US(-1))) + C(1,2)*D(LOG(GDP_CAN(-1))) + C(1,3) + C(1,4)*T

D(LOG(GDP_CAN)) = A(2,1)*(B(1,1)*LOG(GDP_US(-1)) + B(1,2)*LOG(GDP_CAN(-1)) + B
(1,3)) + C(2,1)*D(LOG(GDP_US(-1))) + C(2,2)*D(LOG(GDP_CAN(-1))) + C(2,3) + C(2,4)*T

VAR Model - Substituted Coefficients:
===============================
D(LOG(GDP_US)) = - 0.229705011793*( LOG(GDP_US(-1)) + 0.286109023785*LOG
(GDP_CAN(-1)) - 12.1146586662 ) + 0.118220324*D(LOG(GDP_US(-1))) +
0.0136103060211*D(LOG(GDP_CAN(-1))) - 0.117662329679 + 0.00599184978411*T

D(LOG(GDP_CAN)) = - 0.0411894448722*( LOG(GDP_US(-1)) + 0.286109023785*LOG
(GDP_CAN(-1)) - 12.1146586662 ) + 0.203617013369*D(LOG(GDP_US(-1))) +
0.107740204007*D(LOG(GDP_CAN(-1))) - 0.00276581170121 + 0.000896101575108*T
```

Figure 1.16 *Statistical results based on a VEC Model of* log(GDP_US) *and* log(GDP_Can)

1.4.3 The Application of the Instrumental Variables Models

It is not an easy task to define a "good fit" instrumental variables model, since there is no general guide on how to select an acceptable set of instrumental variables corresponding to any defined statistical model. For this reason, Agung (2009a, p. 382) suggests everyone has *two-stages of problems* (TSOP), in demonstrating or developing an instrumental model. First, he/she should develop a model with at least one exogenous variable which is significantly correlated with the residual of the model. Second, he/she has to search to find the best possible set of instrumental variables. For various examples with special notes on instrumental variables models, refer to Chapter 7 in Agung (2009a).

Example 1.11 (A two-stage LSE method)
Figure 1.17 presents the statistical results based on an instrumental variable model with a trend of *log(GDP_US)* and *log(GDP_Can)*. Based on this result the following notes are presented.

1. Figure 1.17(a) presents the statistical results based on a bivariate AR(1,1)_SCM, where both regressions in the model are the simplest AR(1) linear regressions, with the same set of instrumental variables. These

System: SYS05
Estimation Method: Iterative Two-Stage Least Squares
Date: 08/13/09 Time: 15:55
Sample: 1951 1992
Included observations: 43
Total system (balanced) observations 84
Convergence achieved after 12 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	3.627138	0.616561	5.882853	0.0000
C(11)	0.632266	0.065137	9.706704	0.0000
C(12)	0.842680	0.090564	9.304761	0.0000
C(20)	-4.491572	1.349184	-3.329102	0.0013
C(21)	1.451370	0.140951	10.29701	0.0000
C(22)	0.851882	0.082311	10.34959	0.0000
Determinant residual covariance		8.01E-09		

Equation: LOG(GDP_US) = C(10)+C(11)*LOG(GDP_CAN)+[AR(1)=C(12)]
Instruments: C T LOG(GDP_US(-1)) LOG(GDP_CAN(-1))
Observations: 42

R-squared	0.993172	Mean dependent var	9.473482
Adjusted R-squared	0.992822	S.D. dependent var	0.227124
S.E. of regression	0.019243	Sum squared resid	0.014441
Durbin-Watson stat	1.622016		

Equation: LOG(GDP_CAN) = C(20)+C(21)*LOG(GDP_US)+[AR(1)=C(22)]
Instruments: C T LOG(GDP_CAN(-1)) LOG(GDP_US(-1))
Observations: 42

R-squared	0.993355	Mean dependent var	9.269590
Adjusted R-squared	0.993014	S.D. dependent var	0.342288
S.E. of regression	0.028609	Sum squared resid	0.031921
Durbin-Watson stat	1.577089		

(a)

Estimation Method: Iterative Two-Stage Least Squares
Date: 08/13/09 Time: 16:04
Sample: 1952 1992
Included observations: 42
Total system (unbalanced) observations 81
Convergence achieved after 7 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	4.360923	0.645500	6.755880	0.0000
C(11)	-0.254102	0.109895	-2.312221	0.0236
C(12)	0.812034	0.086502	9.387500	0.0000
C(13)	1.245268	0.153931	8.089757	0.0000
C(14)	-0.417304	0.152645	-2.733820	0.0079
C(20)	-1.022586	0.510377	-2.003591	0.0489
C(21)	0.771526	0.093052	8.291338	0.0000
C(22)	0.333348	0.141338	2.358516	0.0211
C(23)	0.199423	0.186186	1.071095	0.2877
Determinant residual covariance		1.39E-07		

Equation: LOG(GDP_US) = C(10)+C(11)*LOG(GDP_US(-1))+C(12)
*LOG(GDP_CAN)+[AR(1)=C(13), AR(2)=C(14)]
Instruments: C T LOG(GDP_US(-1)) LOG(GDP_US(-2)) LOG(GDP_CAN(
-1)) LOG(GDP_US(-3)) LOG(GDP_CAN(-2))
Observations: 40

R-squared	0.993566	Mean dependent var	9.490958
Adjusted R-squared	0.992831	S.D. dependent var	0.218292
S.E. of regression	0.018483	Sum squared resid	0.011957
Durbin-Watson stat	1.781571		

Equation: LOG(GDP_CAN) = C(20)+C(21)*LOG(GDP_CAN(-1))+C(22)
*LOG(GDP_US)+[AR(1)=C(23)]
Instruments: C T LOG(GDP_CAN(-1)) LOG(GDP_CAN(-2)) LOG(GDP_US(
-1))
Observations: 41

R-squared	0.994549	Mean dependent var	9.281978
Adjusted R-squared	0.994107	S.D. dependent var	0.336874
S.E. of regression	0.025861	Sum squared resid	0.024746
Durbin-Watson stat	1.874247		

(b)

Figure 1.17 *Statistical results based on bivariate models (a) AR(1,1)_SCM, and (b) LVAR(1,1;2,1)_SCM, with sets of instrumental variables*

results show that *log(GDP_Can)* and *log(GDP_US)* have significant simultaneous causal effects, since *log(GDP_Can)* has a significant positive effect on *log(GDP_US)* based on the t-statistic of $t_0 = 9.706\,704$ with a p-value $= 0.0000/2 = 0.0000$, and *log(GDP_US)* also has a significant positive effect on *log(GDP_Can)* based on the t-statistic of $t_0 = 10.29\,701$ with a p-value $= 0.0000/2 = 0.0000$. Note that in this case the p-values in the output should be divided by 2 for testing the one-sided hypotheses.

2. Figure 1.17(b) presents the statistical results based on a LVAR(1,1;2,1)_SCM, where the first regression is a LVAR(1,2) model with an exogenous variable *log(GDP_Can)* and the second regression is a LVAR (1,1) model with an exogenous variable *log(GDP_US)*. Compared to the model in Figure 1.14(a), the regressions in this model have different sets of instrumental variables

1.5 Seemingly Causal Models with Time-Related Effects

As the extension of the SCMs with trends in (1.9) and (1.10), the following system equations present SCMs with time-related effects.

1.5.1 SCM Based on the Path Diagram in Figure 1.10(a)

As an extension of the additive model in (1.9), a SCM with time-related effects based on the path diagram in Figure 1.10(a), will have the following system specification. Note that the two-way interaction $t^{*}GDP\_US$

is inserted as an additional independent variable of the second regression to indicate the time-related effect of *DGP_US* on *GDP_Can*. In other words, the effect of *DGP_US* on *GDP_Can* depends on the time *t*.

$$\begin{aligned} GDP\_US &= C(10) + C(11)^{*}GDP\_US(-1) + C(12)^{*}t \\ GDP\_Can &= C(20) + C(21)^{*}GDP\_Can(-1) + C(22)^{*}GDP\_US \\ &\quad + C(23)^{*}GDP\_US(-1) + C(24)^{*}t + C(25)^{*}t^{*}GDP\_US \end{aligned} \tag{1.14}$$

Note that this model is indicating that the effect of *GDP_US* on *GDP_Can* depends on the time *t*, indicated by the following partial derivative:

$$\frac{\partial GDP\_Can}{\partial GDP\_US} = c(24) + c(25)^{*}t$$

1.5.2 SCM Based on the Path Diagram in Figure 1.10(b)

As an extension of the interaction model in (1.14), a SCM with the time-related effects based on the path diagram in Figure 1.10(b), will have the following system specification. Note that the second regressions of the SCMs in (1.14) and (1.15) are identical models.

$$\begin{aligned} GDP\_US &= C(10) + C(11)^{*}GDP\_US(-1) + C(12)^{*}GDP\_Can \\ &\quad + C(13)^{*}GDP\_Can(-1) + C(14)^{*}t + C(15)^{*}t^{*}GDP\_Can \\ GDP\_Can &= C(20) + C(21)^{*}GDP\_Can(-1) + C(22)^{*}GDP\_US \\ &\quad + C(23)^{*}GDPGDP\_US(-1) + C(24)^{*}t + C(25)^{*}t^{*}GDP\_US \end{aligned} \tag{1.15}$$

Example 1.12 A translog linear SCM with time-related effects
We find that the statistical results based on the model in (1.15) present several insignificant independent variables, including the two-way interactions $t^{*}GDP\_Can$ and $t^{*}GDP\_US$. So, by using the trial-and-error method we can finally obtain a good fit translog linear SCM with time-related effects as presented in Figure 1.18,

System: UNTITLED
Estimation Method: Least Squares
Date: 08/13/09 Time: 17:12
Sample: 1951 1992
Included observations: 42
Total system (balanced) observations 84

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	2.520206	0.638031	3.949973	0.0002
C(11)	0.769750	0.083439	9.225269	0.0000
C(12)	0.568780	0.088966	6.393227	0.0000
C(13)	-0.617770	0.092585	-6.672465	0.0000
C(15)	0.000562	0.000163	3.455584	0.0009
C(20)	-2.759283	0.875206	-3.152724	0.0023
C(21)	0.928523	0.079587	11.66681	0.0000
C(22)	0.934923	0.143415	6.519011	0.0000
C(23)	-0.561499	0.166267	-3.377087	0.0012
C(25)	-0.000504	0.000220	-2.291717	0.0248

Determinant residual covariance 4.94E-08

Equation: LOG(GDP_US) = C(10)+C(11)*LOG(GDP_US(-1))+C(12)
*LOG(GDP_CAN)+C(13)*LOG(GDP_CAN(-1)) +C(15)*T
*LOG(GDP_CAN)
Observations: 42

R-squared	0.994869	Mean dependent var	9.473482
Adjusted R-squared	0.994314	S.D. dependent var	0.227124
S.E. of regression	0.017126	Sum squared resid	0.010852
Durbin-Watson stat	1.858029		

Equation: LOG(GDP_CAN) = C(20)+C(21)*LOG(GDP_CAN(-1))+C(22)
*LOG(GDP_US)+C(23)*LOG(GDP_US(-1))+C(25)*T*LOG(GDP_US)
Observations: 42

R-squared	0.996406	Mean dependent var	9.269590
Adjusted R-squared	0.996017	S.D. dependent var	0.342288
S.E. of regression	0.021601	Sum squared resid	0.017265
Durbin-Watson stat	1.723052		

Figure 1.18 *Statistical results based on a reduced model of a modified model in (1.15)*

which is in fact a nonhierarchical reduced model of the modified model in (1.15). Based on these results the following notes are presented.

1. The p-value of the two-way interaction $T^*log(GDP\_Can)$ in the first regression indicates that the adjusted effect *log(GDP_Can)* on *log(GDP_US)* is significantly dependent on the time t, and the p-value of the two-way interaction $T^*log(GDP\_US)$ in the second regression also indicates that the adjusted effect *log(GDP_US)* on *log(GDP_Can)* is significantly dependent on the time t.
2. Therefore, based on this SCM, it can be concluded that the data support the hypothesis *log(GDP_US)* and *log(GDP_Can)* have simultaneous causal effects dependent on the time t.

1.6 The Application of the Object POOL

Many students, as well as less experienced analysts, have used the object *POOL* to present statistical results or outputs based on either fixed or random effects models, without considering or discussing the characteristics of the models, not to mention their limitations. For this reason, the following examples present illustrative statistical results with special notes.

The steps of the analysis using the object "*POOL*" are as follows:

1. By selecting *Object/New Objects/Pool ... OK*, the window in Figure 1.19(a) appears.
2. By inserting Cross-Section Identifiers, namely the series *_CAN _US _FRA _GER*, and then clicking "*Estimate*", then options in Figure 1.19(b) appear.

1.6.1 What is a Fixed-Effect Model?

The following example presents the statistical results based on simple multivariate growth models with special notes on the acceptability of the models.

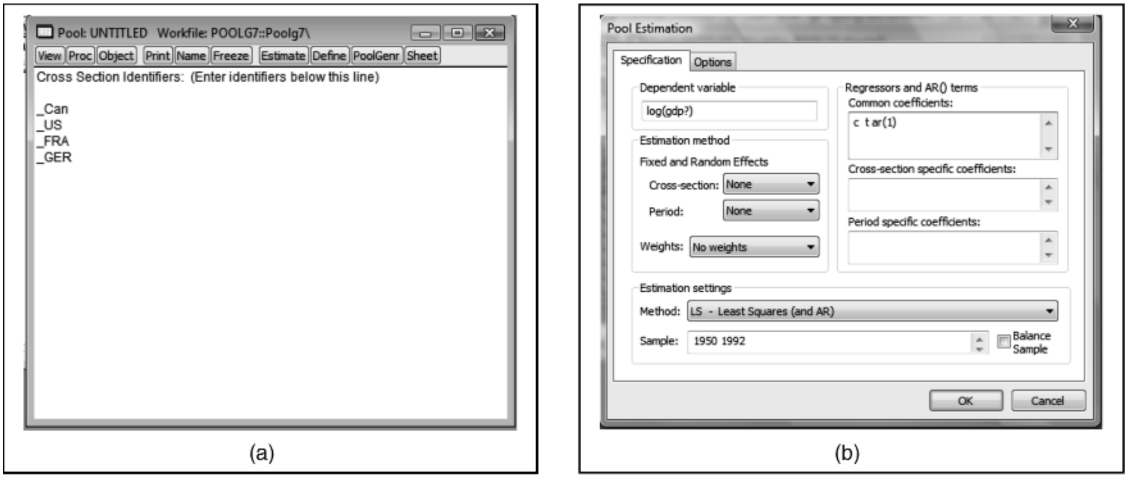


Figure 1.19 The windows and options for using the object "POOL"

Example 1.13 A fixed-effect MAR(1)_GM

For comparison with the statistical results in Figure 1.2, Figure 1.20(a) presents the statistical results by selecting cross-section: *Fixed*, inserting log(*GDP?)* as the dependent variable, and the series "*c t ar(1)*" as the *Regressors and AR() term Common Coefficients*.

Then by selecting *View/Representation*, the estimation equation in Figure 1.20(b) is obtained. Based on these results, the following conclusions and comments are presented.

1. Note that this fixed-effect AR(1) model has special characteristics, where the four regressions in the model have special intercept parameters, namely *C(4) + C(1), C(5) + C(1), C(6) + C(1)*, and *C(7) + C(1)*, respectively, where the parameters *C(4)* to *C(7)* are named as the cross-section fixed-effect parameters.
2. However, we find the equality of the parameters *C(4)*, *C(5)*, *C(6)* and *C(7)* cannot tested using the Wald test.
3. The growth rates of *GDP* are presented by a single parameter of *C*(2) in the four countries, as well as a single autocorrelation of *C*(3) for the four regressions, which should be inappropriate or unrealistic estimates, in a theoretical sense. For this reason, compared to the model in Figure 1.2 along with the model in Figure 1.20, this MAR(1)_GM should be considered unacceptable for representing growth rates of the *GDP* in the four countries.

Example 1.14 A fixed effect MLV(1)_GM

For a comparison with the statistical results in Figure 1.9, Figure 1.21(a) presents the statistical results by selecting Cross-section: *Fixed*, inserting log(*GDP?)* as the dependent variable, and the series "*c t log(GDP?*

Dependent Variable: GDP?
Method: Pooled Least Squares
Date: 07/10/09 Time: 13:04
Sample (adjusted): 1951 1992
Included observations: 42 after adjustments
Cross-sections included: 4
Total pool (balanced) observations: 168
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4683.433	424.2855	11.03840	0.0000
T	268.9511	13.41703	20.04550	0.0000
AR(1)	0.863802	0.038453	22.46383	0.0000
Fixed Effects (Cross)				
_CAN--C	325.1819			
_US--C	2299.204			
_FRA--C	-1570.708			
_GER--C	-1053.678			

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.994053	Mean dependent var	10871.11
Adjusted R-squared	0.993870	S.D. dependent var	3660.970
S.E. of regression	286.6443	Akaike info criterion	14.18942
Sum squared resid	13310721	Schwarz criterion	14.30099
Log likelihood	-1185.911	Hannan-Quinn criter.	14.23470
F-statistic	5415.789	Durbin-Watson stat	1.514880
Prob(F-statistic)	0.000000		

(a)

```
Estimation Command:
=====================
LS(CX=F) GDP? C T AR(1)

Estimation Equations:
=====================
GDP_CAN = C(4) + C(1) + C(2)*T + [AR(1)=C(3)]

GDP_US = C(5) + C(1) + C(2)*T + [AR(1)=C(3)]

GDP_FRA = C(6) + C(1) + C(2)*T + [AR(1)=C(3)]

GDP_GER = C(7) + C(1) + C(2)*T + [AR(1)=C(3)]

Substituted Coefficients:
=====================
GDP_CAN = 325.181869729 + 4683.43277354 + 268.951075506*T + [AR(1)=0.863802466939]

GDP_US = 2299.20386889 + 4683.43277354 + 268.951075506*T + [AR(1)=0.863802466939]

GDP_FRA = -1570.70754267 + 4683.43277354 + 268.951075506*T + [AR(1)=0.863802466939]

GDP_GER = -1053.67819595 + 4683.43277354 + 268.951075506*T + [AR(1)=0.863802466939]
```

(b)

Figure 1.20 *Statistical results based on a fixed effect MAR(**1**)_GM of* GDP_Can, GDP_US, GDP_Fra *and* GDP_Ger

Dependent Variable: LOG(GDP?)
Method: Pooled Least Squares
Date: 07/10/09 Time: 14:11
Sample (adjusted): 1951 1992
Included observations: 42 after adjustments
Cross-sections included: 4
Total pool (balanced) observations: 168

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.684159	0.172521	3.965655	0.0001
T	0.001406	0.000575	2.445828	0.0155
LOG(GDP?(-1))	0.925098	0.020089	46.04897	0.0000
Fixed Effects (Cross)				
_CAN--C	-7.15E-05			
_US--C	0.009831			
_FRA--C	-0.008606			
_GER--C	-0.001153			

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.995548	Mean dependent var	9.229532
Adjusted R-squared	0.995410	S.D. dependent var	0.375014
S.E. of regression	0.025406	Akaike info criterion	-4.472593
Sum squared resid	0.104566	Schwarz criterion	-4.361023
Log likelihood	381.6978	Hannan-Quinn criter.	-4.427312
F-statistic	7244.806	Durbin-Watson stat	1.612986
Prob(F-statistic)	0.000000		

(a)

```
Estimation Command:
=========================
LS(CX=F) LOG(GDP?) C T LOG(GDP?(-1))

Estimation Equations:
=========================
LOG(GDP_CAN) = C(4) + C(1) + C(2)*T + C(3)*LOG(GDP_CAN(-1))

LOG(GDP_US) = C(5) + C(1) + C(2)*T + C(3)*LOG(GDP_US(-1))

LOG(GDP_FRA) = C(6) + C(1) + C(2)*T + C(3)*LOG(GDP_FRA(-1))

LOG(GDP_GER) = C(7) + C(1) + C(2)*T + C(3)*LOG(GDP_GER(-1))

Substituted Coefficients:
=========================
LOG(GDP_CAN) = -7.15294200332e-05 + 0.684159304375 + 0.00140608433156*T + 0.925097700701*LOG(GDP_CAN(-1))

LOG(GDP_US) = 0.0098308713251 + 0.684159304375 + 0.00140608433156*T + 0.925097700701*LOG(GDP_US(-1))

LOG(GDP_FRA) = -0.00860588498651 + 0.684159304375 + 0.00140608433156*T + 0.925097700701*LOG(GDP_FRA(-1))

LOG(GDP_GER) = -0.00115345691855 + 0.684159304375 + 0.00140608433156*T + 0.925097700701*LOG(GDP_GER(-1))
```

(b)

Figure 1.21 *Statistical results based on a fixed effect MLV(**1**)_GM of* GDP_Can, GDP_US, GDP_Fra *and* GDP_Ger

(−1)" as the *Regressors and AR() term Common Coefficients*. Then by selecting *View/Representation*, the estimation equation in Figure 1.21(b) can be obtained. This model also has the same problem as the fixed-effect MAR(**1**)_GM, so it should be considered inappropriate for representing the growth rates of the *GDP* in the four countries.

1.6.2 What is a Random Effect Model?

The AR() terms cannot be used for a random effect model. For this reason, the following examples only present results based on the classical growth model and a random effect MLV(**1**)_GM.

Example 1.15 A random effect multivariate classical growth model: REMCGM
Figure 1.22 presents the statistical results based on a REMCGM of *GDPs* for the four countries, as well as the four regression functions having the same growth rates of *C(2)*. This model is also an inappropriate model in a theoretical sense, aside from the very small value of the weighted DW statistic of 0.078 665. For this reason, a random effect MLV(**1**)_GM is presented in the following example.

Example 1.16 A random effect MLV(1)_GM
For comparison with the statistical results based on the fixed effect MLV(**1**)_GM in Figure 1.21, Figure 1.23 presents the results based on a random effect MLV(**1**)_GM, which shows that the four regression functions have exactly the same coefficients of *C(1), C(2)* and *C(3)*; and *C(4)* = *C(5)* = *C(6)* = *C(7)* = 0. In a theoretical sense, this is the worst model.

Dependent Variable: LOG(GDP?)
Method: Pooled EGLS (Cross-section random effects)
Date: 07/10/09 Time: 15:19
Sample: 1950 1992
Included observations: 43
Cross-sections included: 4
Total pool (balanced) observations: 172
Swamy and Arora estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.618096	0.095592	90.15486	0.0000
T	0.027090	0.000616	43.97470	0.0000
Random Effects (Cross)				
_CAN--C	0.042773			
_US--C	0.248363			
_FRA--C	-0.154467			
_GER--C	-0.136669			

Effects Specification		S.D.	Rho
Cross-section random		0.188634	0.7797
Idiosyncratic random		0.100261	0.2203

Weighted Statistics			
R-squared	0.919193	Mean dependent var	0.744397
Adjusted R-squared	0.918718	S.D. dependent var	0.351667
S.E. of regression	0.100261	Sum squared resid	1.708871
F-statistic	1933.774	Durbin-Watson stat	0.078665
Prob(F-statistic)	0.000000		

Unweighted Statistics			
R-squared	0.755259	Mean dependent var	9.214075
Sum squared resid	6.299071	Durbin-Watson stat	0.021341

```
Estimation Command:
=====================
LS(CX=R) LOG(GDP?) C T

Estimation Equations:
=====================
LOG(GDP_CAN) = C(3) + C(1) + C(2)*T

LOG(GDP_US) = C(4) + C(1) + C(2)*T

LOG(GDP_FRA) = C(5) + C(1) + C(2)*T

LOG(GDP_GER) = C(6) + C(1) + C(2)*T


Substituted Coefficients:
=====================
LOG(GDP_CAN) = 0.0427728268614 + 8.61809606213 + 0.0270899687917*T

LOG(GDP_US) = 0.248362979351 + 8.61809606213 + 0.0270899687917*T

LOG(GDP_FRA) = -0.154467059659 + 8.61809606213 + 0.0270899687917*T

LOG(GDP_GER) = -0.136668746553 + 8.61809606213 + 0.0270899687917*T
```

Figure 1.22 *Statistical results based on a RECGM of* GDP_Can, GDP_US, GDP_Fra *and* GDP_Ger

Dependent Variable: LOG(GDP?)
Method: Pooled EGLS (Cross-section random effects)
Date: 07/10/09 Time: 14:58
Sample (adjusted): 1951 1992
Included observations: 42 after adjustments
Cross-sections included: 4
Total pool (balanced) observations: 168
Swamy and Arora estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.452666	0.087556	5.169991	0.0000
T	0.000666	0.000323	2.059916	0.0410
LOG(GDP?(-1))	0.952062	0.010187	93.45688	0.0000
Random Effects (Cross)				
_CAN--C	0.000000			
_US--C	0.000000			
_FRA--C	0.000000			
_GER--C	0.000000			

Effects Specification		S.D.	Rho
Cross-section random		0.000000	0.0000
Idiosyncratic random		0.025406	1.0000

```
Estimation Command:
=====================
LS(CX=R) LOG(GDP?) C T LOG(GDP?(-1))

Estimation Equations:
=====================
LOG(GDP_CAN) = C(4) + C(1) + C(2)*T + C(3)*LOG(GDP_CAN(-1))

LOG(GDP_US) = C(5) + C(1) + C(2)*T + C(3)*LOG(GDP_US(-1))

LOG(GDP_FRA) = C(6) + C(1) + C(2)*T + C(3)*LOG(GDP_FRA(-1))

LOG(GDP_GER) = C(7) + C(1) + C(2)*T + C(3)*LOG(GDP_GER(-1))


Substituted Coefficients:
=====================
LOG(GDP_CAN) = 0 + 0.452666208449 + 0.000665617512811*T + 0.952061931017*LOG(GDP_CAN(-1))

LOG(GDP_US) = 0 + 0.452666208449 + 0.000665617512811*T + 0.952061931017*LOG(GDP_US(-1))

LOG(GDP_FRA) = 0 + 0.452666208449 + 0.000665617512811*T + 0.952061931017*LOG(GDP_FRA(-1))

LOG(GDP_GER) = 0 + 0.452666208449 + 0.000665617512811*T + 0.952061931017*LOG(GDP_GER(-1))
```

Weighted Statistics			
R-squared	0.995433	Mean dependent var	9.229532
Adjusted R-squared	0.995377	S.D. dependent var	0.375014
S.E. of regression	0.025497	Sum squared resid	0.107270
F-statistic	17980.39	Durbin-Watson stat	1.614198
Prob(F-statistic)	0.000000		

Unweighted Statistics			
R-squared	0.995433	Mean dependent var	9.229532
Sum squared resid	0.107270	Durbin-Watson stat	1.614198

Figure 1.23 *Statistical results based on a random effect MLV(**1**)_GM of* GDP_Can, GDP_US, GDP_Fra *and* GDP_Ger

1.6.3 Special Notes

Based on the statistical results of the multivariate fixed and random effects growth models using the object "POOL" presented previously, please note the following special points.

1. In general, a multivariate fixed and random effects growth model is the worst multivariate growth model by states, in a theoretical sense. This is even more so if it is known that a state should have a discontinuous or piece-wise growth curve, as this may be because of some external factors. Therefore, in general one should use models with heterogeneous slopes or *heterogeneous regressions* (Agung, 2006, 2011). For an additional illustration, Chandrasekaran and Tellis (in Malhotra, 2007, p. 45) present the findings of Golder and Tellis (2004) on piece-wise mean growth rates of a product's life cycle over six time periods; namely during the introduction, takeoff, growth, slowdown, early maturity and late maturity.
2. Referring to the ANCOVA models, in a statistical sense, models that have homogeneous slopes or *homogeneous regressions* (Agung, 2006, 2011) with various intercepts are acceptable. The main objectives of ANCOVA are to test the hypotheses on the *adjusted means differences* of the corresponding dependent variables, which in fact are the hypotheses on *the intercept differences* of the homogeneous regressions considered. However, analysis should be conducted using the object "*System*", instead of the object "*POOL*" – refer to point (2) in Example 1.13. See the following example.

Example 1.17 A MAR(1) ANCOVA growth model

For illustration, Figure 1.24 presents the statistical results based on a MAR(1) ANCOVA growth model, using the object "*System*". Based on these results the following conclusions and notes are presented.

1. Note that the growth rates of *GDP* of the four states are assumed to be equal to $\hat{C}(11) = 0.011266$ which are unacceptable in a theoretical sense. Similarly so for the first autoregressive indicator $\hat{C}(12) = 0.960457$.

System: SYS09
Estimation Method: Iterative Least Squares
Date: 08/19/09 Time: 09:53
Sample: 1951 1992
Included observations: 43
Total system (balanced) observations 168
Convergence achieved after 2 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	9.378796	0.344686	27.20967	0.0000
C(11)	0.011266	0.006555	1.718618	0.0876
C(12)	0.960457	0.009552	100.5540	0.0000
C(20)	9.452469	0.438635	21.54973	0.0000
C(30)	9.057210	0.362113	25.01213	0.0000
C(40)	9.573194	0.549645	17.41705	0.0000

Determinant residual covariance 5.73E-14

Equation: LOG(GDP_US)=C(10)+C(11)*T+[AR(1)=C(12)]
Observations: 42

R-squared	0.987376	Mean dependent var	9.473482
Adjusted R-squared	0.986729	S.D. dependent var	0.227124
S.E. of regression	0.026165	Sum squared resid	0.026699
Durbin-Watson stat	1.957912		

Equation: LOG(GDP_GER)=C(20)+C(11)*T+[AR(1)=C(12)]
Observations: 42

R-squared	0.996285	Mean dependent var	9.098903
Adjusted R-squared	0.996094	S.D. dependent var	0.392866
S.E. of regression	0.024552	Sum squared resid	0.023510
Durbin-Watson stat	1.457365		

Equation: LOG(GDP_UK)=C(30)+C(11)*T+[AR(1)=C(12)]
Observations: 42

R-squared	0.993413	Mean dependent var	9.064748
Adjusted R-squared	0.993075	S.D. dependent var	0.260995
S.E. of regression	0.021719	Sum squared resid	0.018397
Durbin-Watson stat	1.526476		

Equation: LOG(GDP_JPN)=C(40)+C(11)*T+[AR(1)=C(12)]
Observations: 42

R-squared	0.998088	Mean dependent var	8.728910
Adjusted R-squared	0.997990	S.D. dependent var	0.698799
S.E. of regression	0.031331	Sum squared resid	0.038284
Durbin-Watson stat	1.075140		

Wald Test:
System: SYS09

Test Statistic	Value	df	Probability
Chi-square	8.464132	3	0.0373

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(10) - C(40)	-0.194398	0.275429
C(20) - C(40)	-0.120724	0.187951
C(30) - C(40)	-0.515984	0.256885

Restrictions are linear in coefficients.

Figure 1.24 *Statistical results based on an MAR(1) ANCOVA growth model*

System: SYS10
Estimation Method: Iterative Least Squares
Date: 08/19/09 Time: 11:06
Sample: 1951 1992
Included observations: 43
Total system (balanced) observations 168
Convergence achieved after 7 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	9.078944	0.038474	235.9787	0.0000
C(11)	0.018204	0.001356	13.42234	0.0000
C(12)	0.784511	0.103526	7.577923	0.0000
C(20)	8.872253	0.077353	114.6988	0.0000
C(22)	0.903524	0.020657	43.73899	0.0000
C(30)	8.689235	0.051048	170.2162	0.0000
C(32)	0.853583	0.092179	9.260062	0.0000
C(40)	9.303788	0.251365	37.01301	0.0000
C(42)	0.962359	0.007906	121.7262	0.0000

Determinant residual covariance 3.87E-14

Equation: LOG(GDP_US)=C(10)+C(11)*T+[AR(1)=C(12)]
Observations: 42

R-squared	0.988132	Mean dependent var	9.473482
Adjusted R-squared	0.987523	S.D. dependent var	0.227124
S.E. of regression	0.025370	Sum squared resid	0.025101
Durbin-Watson stat	1.743993		

Equation: LOG(GDP_GER)=C(20)+C(11)*T+[AR(1)=C(22)]
Observations: 42

R-squared	0.996809	Mean dependent var	9.098903
Adjusted R-squared	0.996645	S.D. dependent var	0.392866
S.E. of regression	0.022756	Sum squared resid	0.020196
Durbin-Watson stat	1.607262		

Equation: LOG(GDP_UK)=C(30)+C(11)*T+[AR(1)=C(32)]
Observations: 42

R-squared	0.993821	Mean dependent var	9.064748
Adjusted R-squared	0.993504	S.D. dependent var	0.260995
S.E. of regression	0.021035	Sum squared resid	0.017257
Durbin-Watson stat	1.471386		

Equation: LOG(GDP_JPN)=C(40)+C(11)*T+[AR(1)=C(42)]
Observations: 42

R-squared	0.998109	Mean dependent var	8.728910
Adjusted R-squared	0.998012	S.D. dependent var	0.698799
S.E. of regression	0.031155	Sum squared resid	0.037854
Durbin-Watson stat	1.089900		

Wald Test:
System: SYS10

Test Statistic	Value	df	Probability
Chi-square	130.9457	3	0.0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(10) - C(40)	-0.224815	0.241215
C(20) - C(40)	-0.431517	0.242261
C(30) - C(40)	-0.614530	0.240215

Restrictions are linear in coefficients.

Figure 1.25 *Statistical results based on an alternative MAR(**1**) ANCOVA growth model*

2. However, in a statistical sense, this model is an acceptable MANCOVA model of the variables *log(GDP_US), log(GDP_Ger), log(GDP_UK)*, and *log(GDP_JPN)*, where the time t is considered covariate, with the intercept parameters: $C(10)$, $C(20)$, $C(30)$, and $C(40)$.
3. Therefore, various hypotheses on the adjusted means differences of the *log(GDP?)* between any subsets of the four states can easily be tested using the Wald test. For example, H_0: $C(10) = C(20) = C(30) = C(40)$ is rejected based on the Chi-square statistic of $\chi_0^2 = 8.464132$, with $df = 3$ and a p-value $= 0.0373$, which indicates that the four states have significant adjusted means differences of the *log(GDP?)*.
4. As a comparison, Figure 1.25 presents the statistical results based on an alternative MAR(**1**)_GM, under the assumption the time t has the same slopes of $C(11)$, but various intercepts as well as AR(1). For this model, H_0: $C(10) = C(20) = C(30) = C(40)$ is rejected based on the Chi-square statistic of $\chi_0^2 = 130.9457$, with $df = 3$ and a p-value $= 0.0000$.
5. On the other hand, the null hypothesis H_0: $C(12) = C(22) = C(32) = C(42)$ should be considered in comparing this model with the model in Figure 1.24. The null hypothesis is rejected based on the Chi-square statistic of $\chi_0^2 = 10.43420$, with $df = 3$ and a p-value $= 0.0152$. Then, in a statistical sense, this model is a better fit compared to the model in Figure 1.24.
6. Note that the Durbin–Watson statistics of the regressions in Figures 1.21 and 1.24 indicate that other models should be explored, such as the higher order AR models. However, try it as an exercise.

Example 1.18 A MAR(1) heterogeneous growth model

Building on the model in Figure 1.25, as well as for further comparison, Figure 1.26 presents the statistical results based on a MAR(**1**) heterogeneou**s** growth model. Based on these results, note the following:

1. The main objectives of this model are to test the hypotheses of the exponential growth rate differences between the GDPs of the four states, indicated by the parameters $C(11)$, $C(21)$, $C(31)$ and $C(41)$.

System: SYS11
Estimation Method: Iterative Least Squares
Date: 08/19/09 Time: 12:12
Sample: 1951 1992
Included observations: 43
Total system (balanced) observations 168
Convergence achieved after 2 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	9.098363	0.044497	204.4720	0.0000
C(11)	0.017432	0.001607	10.84940	0.0000
C(12)	0.788836	0.104235	7.567884	0.0000
C(20)	8.788810	0.184019	47.76036	0.0000
C(21)	0.020326	0.004748	4.281062	0.0000
C(22)	0.891366	0.035255	25.28374	0.0000
C(30)	8.623544	0.025492	338.2892	0.0000
C(31)	0.020563	0.000956	21.51224	0.0000
C(32)	0.709349	0.116626	6.082249	0.0000
C(40)	15.62487	27.31062	0.572117	0.5681
C(41)	-0.041003	0.188326	-0.217721	0.8279
C(42)	0.984128	0.030234	32.55033	0.0000

Determinant residual covariance 3.27E-14

Equation: LOG(GDP_US)=C(10)+C(11)*T+[AR(1)=C(12)]
Observations: 42

R-squared	0.988209	Mean dependent var	9.473482
Adjusted R-squared	0.987604	S.D. dependent var	0.227124
S.E. of regression	0.025287	Sum squared resid	0.024938
Durbin-Watson stat	1.762500		

Equation: LOG(GDP_GER)=C(20)+C(21)*T+[AR(1)=C(22)]
Observations: 42

R-squared	0.996821	Mean dependent var	9.098903
Adjusted R-squared	0.996658	S.D. dependent var	0.392866
S.E. of regression	0.022710	Sum squared resid	0.020114
Durbin-Watson stat	1.596264		

Equation: LOG(GDP_UK)=C(30)+C(31)*T+[AR(1)=C(32)]
Observations: 42

R-squared	0.994162	Mean dependent var	9.064748
Adjusted R-squared	0.993863	S.D. dependent var	0.260995
S.E. of regression	0.020446	Sum squared resid	0.016304
Durbin-Watson stat	1.376546		

Equation: LOG(GDP_JPN)=C(40)+C(41)*T+[AR(1)=C(42)]
Observations: 42

R-squared	0.998137	Mean dependent var	8.728910
Adjusted R-squared	0.998041	S.D. dependent var	0.698799
S.E. of regression	0.030927	Sum squared resid	0.037302
Durbin-Watson stat	1.130107		

Wald Test:
System: SYS11

Test Statistic	Value	df	Probability
Chi-square	2.923708	3	0.4035

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(11) - C(41)	0.058434	0.188333
C(21) - C(41)	0.061329	0.188386
C(31) - C(41)	0.061565	0.188329

Restrictions are linear in coefficients.

Figure 1.26 *Statistical results based on an alternative MAR(**1**) heterogeneous growth model*

2. The null hypothesis H_0: $C(11) = C(21) = C(31) = C(41)$ is accepted based on the Chi-square statistic of $\chi_0^2 = 2.923708$, with $df = 3$ and a p-value $= 0.4035$, which indicates that the four growth rates are insignificantly different in the corresponding populations. Based on this finding then, the model in Figure 1.25 can be considered to be a better fit, in a statistical sense, compared to the model in Figure 1.26. However, in a theoretical sense, would you be very confident in saying that the four growth rates of the GDPs are equal?
3. Compare the growth curve of *log(GDP_JPN)* in Figure 1.27 to the negative estimate of its growth rate, namely $\hat{C}(41) = -0.014003$ in Figure 1.26: this indicates that the model is inappropriate for representing the *GDP_JPN*. So a modified model should be explored. Do this as an exercise: refer to the case of the *GDP_FRA* presented in Examples 1.3 to 1.5.

1.7 Growth Models of Sample Statistics

In many studies, we should consider the time series of sample statistics, such as the mean, median and SD (standard deviation), of groups of individuals based on sample surveys as well as experiments. In general, the symbol *Y_gi(t)* will be used to indicate the time series of a single endogenous variable Y of the i-th individual within the g-th group, for $g = 1, \ldots, G$, and $i = 1, \ldots, I_g$. The panel data file, a set of time series *Y_gi* with the format shown in Table 1.4, where the first group ($g = 1$) contains five individuals and the second group ($g = 2$) contains eight individuals.

Based on this data set the time series of the mean, median and SD of the Y-variable can easily be generated, namely *MY_g, MedY_g*, and *SDY_g*, for $g = 1, \ldots, G$, either using EViews or Excel.

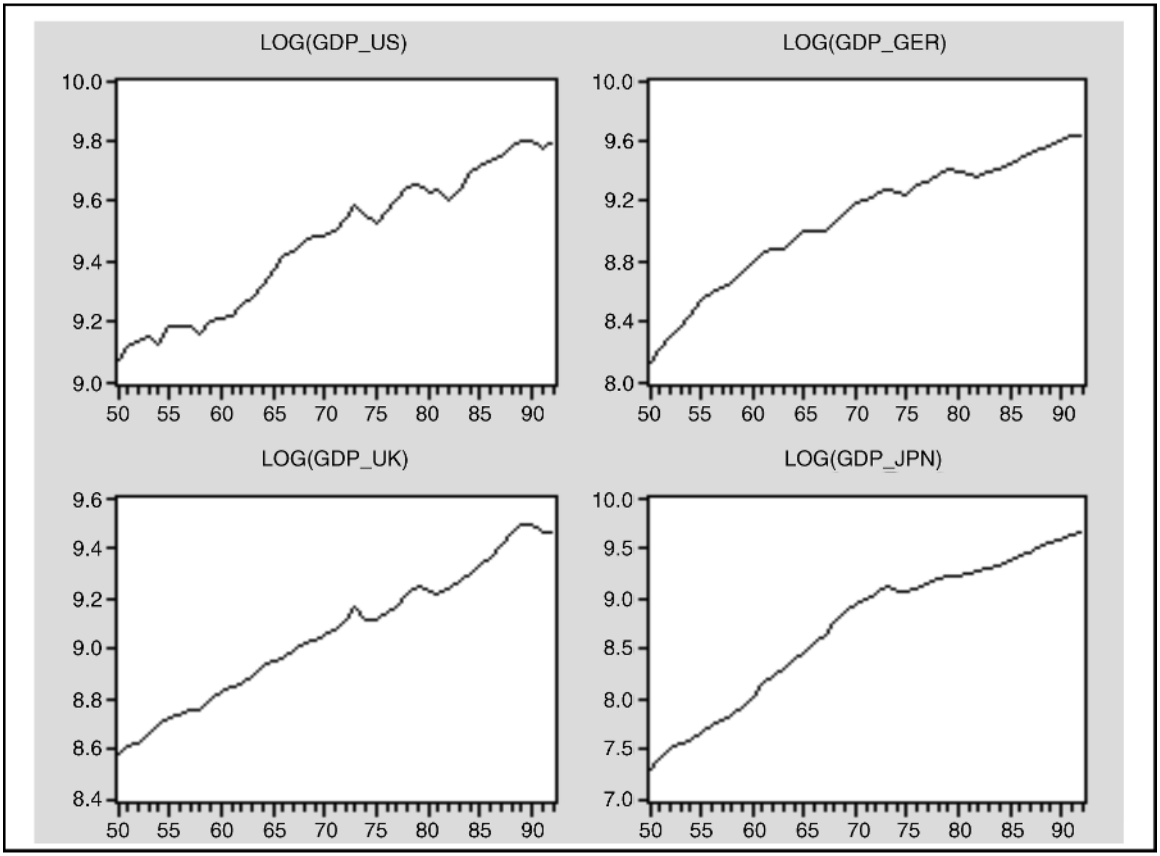


Figure 1.27 *The growth curves of the endogenous variables in Figure 1.26*

A latent variable, or a set of either independent or dependent factors or latent variables, can easily be generated for each group of individuals in order to reduce the dimension of the multivariate considered. For a detailed stepped analysis, refer to Chapter 10 in Agung (2011).

Therefore, all growth models previously presented, as well as their extensions, should be applicable for the time series of the sample statistics *MY_g, MedY_g* and *SDY_g*, as well as latent variables.

Table 1.4 *Illustrated format of a set of time series by two groups of individuals*

Time	g = 1			G = 2		
	Y_11	...	Y_15	Y_21	...	Y_28
1	*Y_11(1)*	...	*Y_15(1)*	*Y_21(1)*	...	*Y_28(1)*
.	...	...	...	...	...	...
t	*Y_11(t)*	...	*Y_51(t)*	*Y_1G(t)*	...	*Y_8G(t)*
.	...	...	...	...	...	...
T	*Y_11(T)*	...	*Y_51(T)*	*Y_1G(T)*		*Y_8G(T)*

Example 1.19 Generating sample statistics using the object "POOL"

For an illustration, the group of *GDP_FRA, GDP_GER* and *GDP_ITA* will be considered for data analysis. The steps of the analysis are as follows:

1. By selecting *Object/New Object/Pool . . . OK*, Figure 1.28(a) appears on the screen.
2. By entering "*_fra _ger _ita*" and selecting *View/Descriptive Statistics . . .*, Figure 1.28(b) is shown on the screen. Then by entering "*gdp?*" and selecting "*Time period specific*", the sample statistics in Figure 1.29 are obtained.
3. Each of the sample statistics can easily be copied to the file. For example, by using the copy-paste method of *Mean GDP?*, the POOLG7.wf1 will have an additional variable *Series01*. Then this variable can be renamed, for example as *M_GDP* or *Mean GDP?* Similarly, this can be done for each of the others.
4. The other copy-paste method can be done using Excel, by copying all sample statistics in Figure 1.29 as an Excel file, then opening the Excel file as an EViews work file.
5. As a result, data analysis based on various models of sample statistics of each group can be easily performed.

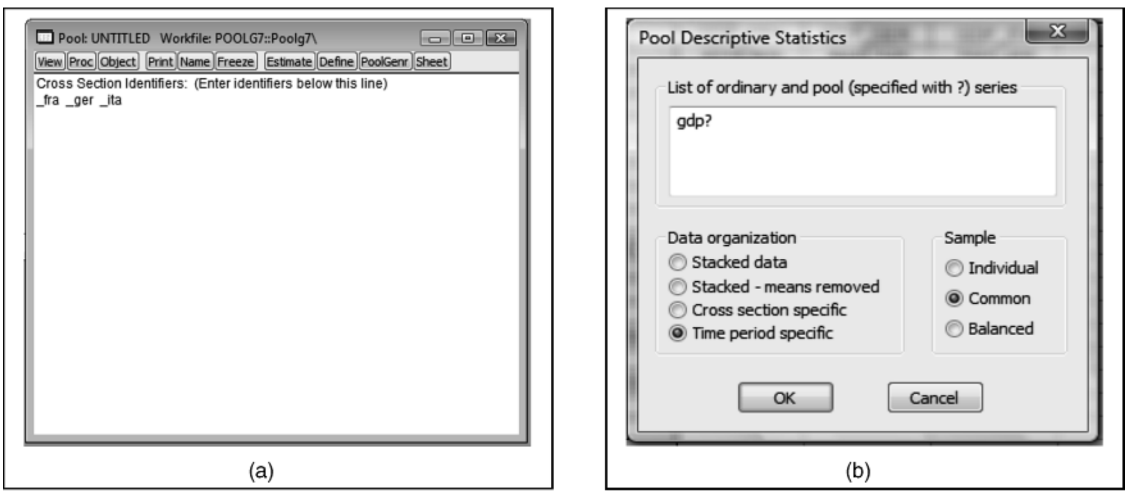
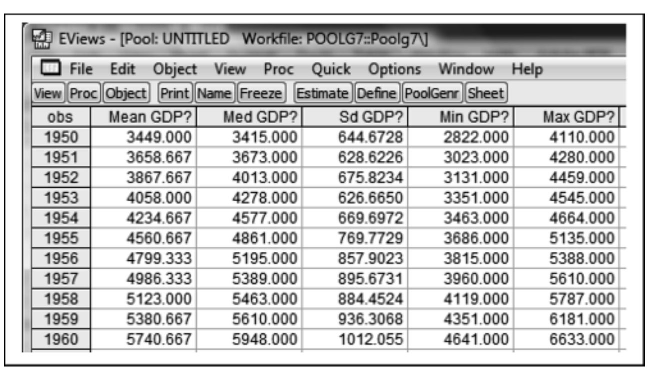


Figure 1.28 *The cross section identifiers and options of the pool descriptive statistics*



obs	Mean GDP?	Med GDP?	Sd GDP?	Min GDP?	Max GDP?
1950	3449.000	3415.000	644.6728	2822.000	4110.000
1951	3658.667	3673.000	628.6226	3023.000	4280.000
1952	3867.667	4013.000	675.8234	3131.000	4459.000
1953	4058.000	4278.000	626.6650	3351.000	4545.000
1954	4234.667	4577.000	669.6972	3463.000	4664.000
1955	4560.667	4861.000	769.7729	3686.000	5135.000
1956	4799.333	5195.000	857.9023	3815.000	5388.000
1957	4986.333	5389.000	895.6731	3960.000	5610.000
1958	5123.000	5463.000	884.4524	4119.000	5787.000
1959	5380.667	5610.000	936.3068	4351.000	6181.000
1960	5740.667	5948.000	1012.055	4641.000	6633.000

Figure 1.29 *Sample statistics of* GDP_FRA, GDP_GER *and* GDP_ITA

1.8 Special Notes on Time-State Observations

Corresponding to time-series models for independent states, as well as the models for dependent or correlated states illustrated previously, the following notes are made.

1. The time-state or time-cross-section observations should have much larger time-point observations compared to the cross-sectional observations.
2. As *a rule of thumb* the time points should be at least five times the total number of the variables in the system specification considered. Most researchers make the number of the units of analysis at least 10 times the number of variables in the model.
3. On the other hand, if panel data has a much larger cross-section observation compared to time-point observations, the following alternative data analysis is suggested.
   3.1 Conduct the analysis based on time series models of sample statistics by groups of states or individuals, such as the mean and standard deviation of the groups, or latent variables as presented in Section 1.7, in addition to the descriptive statistical summaries by groups.
   3.2 If the panel data has only a few time-point observations, then the panel data should be presented or considered as a set of cross-section data by times or a cross-section over times (which will be discussed in Part II). As a special case for a two-year observation, the panel data can be considered a *natural-experimental* data set.

1.9 Growth Models with an Environmental Variable

Suppose Y_i_t is an endogenous time series, say the productivity and return rates of the i-th industry or firm in a state/country, then in general there is an environmental or external time series with the same scores/values for all industries, namely Z_t, such as income per capita, inflation rate, exchange rate of US$, *GDP* and others at the state/country level.

Referring to the *GDPs* of three states in Europe in POOL7 G.wf1, namely $Y\_1_t = GDP\_Fra_t$, $Y\_2_t = DGP\_Ger_t$ and $Y\_3 = GDP\_\_Ita_t$, to generalize, they can be presented as Y_1_t, Y_2_t, and Y_3_t; or Y_a_t, Y_b_t, and Y_c_t; respectively. There should be at least one environmental factor, say Z_t, such as the US$ exchange rate or an external factor out of Europe, which could be defined or judged as a causal factor of the *GDPs* of the three states. Would you consider *GDP_US* or *GDP_Jpn*, or both for use as external factors?

Therefore, corresponding to the MLVAR(***p;q***)_GM in (1.6), the following general MLVAR(***p;q***)_GM with an environmental variable Z_t is made.

$$\begin{aligned}\log(Y\_i_t) &= C(i0) + \sum_{j=1}^{p_i} C(ij)^*\log(Y\_i_{t-j}) + C(i,p_i+1)^*t + C(i,p_i+2)^*Z_t + \mu_{it} \\ \mu_{it} &= \sum_{k=1}^{q_i} \rho_{ik}\mu_{i,t-k} + \varepsilon_{it}\end{aligned} \tag{1.16}$$

However, in general, we know the effect of Z_t on Y_i_t is dependent on t. For this reason, applying the model with a time-related-effect (TRE) is recommended, as follows:

$$\begin{aligned}\log(Y\_i_t) &= C(i0) + \sum_{j=1}^{p_i} C(ij)^*\log(Y\_i_{t-j}) + C(i,p_i+1)^*t \\ &\quad + C(i,p_i+2)^*Z_t + C(i,p_i+3)^*t^*Z_t + \mu_{it} \\ \mu_{it} &= \sum_{k=1}^{q_i} \rho_{ik}\mu_{i,t-k} + \varepsilon_{it}\end{aligned} \tag{1.17}$$

Note that various two- and three-way interaction models have been demonstrated in Agung (2009a) if a vector of environmental variable, namely $\mathbf{Z}_t = (Z1_t, Z2_t, \ldots)$, should be considered. Furthermore, under the assumption that Y_i_t for some $i = 1,2,\ldots,N$, are correlated, then a lot of advanced models, (such as various VAR, VEC and Instrumental Variables Models (IVMs), as well as various TGARCH(*a,b,c*) models), could easily be subjectively defined by a researcher. However, not everyone can always be very sure as to which is the true population model, since unexpected estimates of the model parameters could be obtained as the impact of the multicollinearity of all variables in the model, and these are highly dependent on the data set that happens to be selected or available to the researcher. Refer to special notes presented in Section 2.14 (Agung, 2009a). See the following selected models.

1.9.1 The Simplest Possible Model

The simplest model is a MLVAR(**1,1**) model with an environmental variable and TRE as follows:

$$\begin{aligned} \log(Y\_i_t) &= C(i0) + C(i1)^*\log(Y\_i_{t-1}) + C(i2)^*t + C(i3)^*Z_t + C(i4)^*t^*Z_t + \mu_{it} \\ \mu_{it} &= \rho_{i1}\mu_{i.t-1} + \varepsilon_{it} \end{aligned} \tag{1.18}$$

For an illustration, a hypothetical data set is generated based on the data in POOL7 G.wf1, where *X_1 = GDP_US, Y_1 = GDP_Can*, *X_2 = GDP_Fra, Y_2 = GDP_UK, X_3 = GDP_Ger*, and *Y_3 = GDP_Ita*, and the environmental variable *Z1 = GDP_Jpn* is taken. See the following examples.

For a vector of the environmental variable, namely $\mathbf{Z} = (Z1,\ldots,Zk)$, the model in (1.18) can be extended to a more general M LVAR(**1,1**), as follows:

$$\begin{aligned} \log(Y\_i_t) &= C(i0) + C(i1)^*\log(Y\_i_{t-1}) + F_i(t, Z1, \ldots, Zk) + \mu_{it} \\ \mu_{it} &= \rho_{i1}\mu_{i.t-1} + \varepsilon_{it} \end{aligned} \tag{1.19}$$

where $F_i(t,Z1,\ldots,Zk)$ is a function of the time, *t* and an external or environmental vector $\mathbf{Z} = (Z1,\ldots,Zk)$ with a finite number of parameter for each $i = 1,2,\ldots,N$. Therefore, there would be a lot of possible functions of two-way interaction factors, namely t^*Zk and Zi^*Zj, and a few selected three-way interactions.

On the other hand, specific to the quarterly and monthly data sets, Agung (2009a) proposes two alternative models using the lags Y_i_{t-4} and Y_i_{t-12} respectively, in order to match the conditions in the previous and recent years.

Example 1.20 An application of the system equation

Figure 1.30 presents the statistical results based on a MLVAR(1,1) model in (1.17). The main objective of this model is to test the hypothesis that the effect of the environmental variable *Z1* on the trivariate *(Y_1, Y_2,Y_3)* depends on the time, *t*. Based on these results, see the following notes and comments.

1. Note that the interaction t^*Z1 has a significant effect on each of the variables *Y_1*, *Y_2* and *Y_3*, with a *p*-value of 0.0003, 0.0370 and 0.0102, respectively. It can then be directly concluded that the effect of *Z1* on the trivariate (*Y_1,Y_2,Y_3*) is significantly dependent on the time *t*.
2. On the other hand, if the effects of t^*Z1 on *Y_i* are insignificant for the *i*-th individual, testing the null hypothesis is suggested H_0: $C(13) = C(14) = C(23) = C(24) = C(33) = C(34) = 0$, using the Wald test. Refer to the following example.

System: SYS15
Estimation Method: Seemingly Unrelated Regression
Date: 11/01/09 Time: 17:36
Sample: 1952 1992
Included observations: 42
Total system (balanced) observations 123
Iterate coefficients after one-step weighting matrix
Convergence not achieved after: 1 weight matrix, 584 total coef iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	-2.013163	0.577742	-3.484536	0.0007
C(11)	0.649242	0.079103	8.207511	0.0000
C(12)	0.095458	0.026042	3.665576	0.0004
C(13)	0.544602	0.109193	4.987528	0.0000
C(14)	-0.009671	0.002562	-3.774172	0.0003
C(15)	0.118118	0.146734	0.804976	0.4227
C(20)	0.897583	1.802112	0.498073	0.6195
C(21)	0.123505	0.136029	0.907932	0.3660
C(22)	0.133851	0.057065	2.345572	0.0209
C(23)	0.716732	0.145310	4.932445	0.0000
C(24)	-0.012416	0.005878	-2.112352	0.0370
C(25)	0.734109	0.076490	9.597462	0.0000
C(30)	-0.816335	1.059428	-0.770543	0.4427
C(31)	0.484293	0.176536	2.743304	0.0072
C(32)	0.150537	0.058116	2.590258	0.0110
C(33)	0.544049	0.157995	3.443462	0.0008
C(34)	-0.014455	0.005522	-2.617849	0.0102
C(35)	0.291409	0.176220	1.653669	0.1012

Determinant residual covariance 1.03E-11

Equation: Y_1=C(10)+C(11)*Y_1(-1)+ C(12)*T+C(13)*Z1+C(14)*T*Z1
+[AR(1)=C(15)]
Observations: 41

R-squared	0.998530	Mean dependent var	9.093576
Adjusted R-squared	0.998320	S.D. dependent var	0.368947
S.E. of regression	0.015122	Sum squared resid	0.008004
Durbin-Watson stat	1.712889		

Equation: Y_2=C(20)+C(21)*Y_2(-1)+ C(22)*T+C(23)*Z1+C(24)*T*Z1
+[AR(1)=C(25)]
Observations: 41

R-squared	0.997409	Mean dependent var	9.120614
Adjusted R-squared	0.997039	S.D. dependent var	0.371363
S.E. of regression	0.020209	Sum squared resid	0.014294
Durbin-Watson stat	1.378654		

Equation: Y_3=C(30)+C(31)*Y_3(-1)+ C(32)*T+C(33)*Z1+C(34)*T*Z1
+[AR(1)=C(35)]
Observations: 41

R-squared	0.998144	Mean dependent var	8.889700
Adjusted R-squared	0.997879	S.D. dependent var	0.430718
S.E. of regression	0.019839	Sum squared resid	0.013775
Durbin-Watson stat	1.718009		

Figure 1.30 *Statistical results based on a MLVAR(1,1) Model in (1.18)*

3. On the other hand, a reduced model should be made by deleting either one of *t* and *Z1*, or both, since *t, Z1* and t^*Z1 in many cases are highly correlated, and their impacts on the parameter estimates are unpredictable. Refer to special notes in Section 2.14.3 (Agung, 2009a). In many cases, then it would be found t^*Z1 would have a significant effect in the reduced model.
4. Considering the previous results, note that the AR(1) term is insignificant with a p-value = 0.4227, in the first regression, and $Y\_2(-1)$ is insignificant in the second regression with a p-value = 0.3660. Since their p-values > 0.20; a reduced model should be explored. Do it as an exercise. For the intercept of the third regression, namely *C*(30), it is not a problem.

1.9.2 The Application of VAR and VEC Models

As an extension of the model in (1.17), the application of the VAR and VEC Models are presented in the following examples. Refer to various VAR and VEC models and their limitations presented in Chapter 6 (Agung, 2009a).

Example 1.21 A VAR model using the object "*System*"
Corresponding to the model in Example 1.19, since a single environmental variable *Z1* is defined to be a cause of factors *Y_1, Y_2* and *Y_3*, then *Y_1, Y_2* and *Y_3* should be correlated in a theoretical sense. For this reason a VAR model could be applied. Referring to various VAR models presented in Chapter 6 (Agung, 2009a), then based on the variables *Y_1, Y_2*, *Y_3, Z1* and *t*, a lot of VAR models could easily be derived or defined. However, Agung (2009a; 380) states that the system function (estimation method) is the preferred method used to develop alternative multivariate time series models, since it is more flexible to use for developing a multivariate model where multiple regressions could have different sets of exogenous variables.

System: SYS20
Estimation Method: Seemingly Unrelated Regression
Date: 11/03/09 Time: 13:39
Sample: 1951 1992
Included observations: 42
Total system (balanced) observations 126
Linear estimation after one-step weighting matrix

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	-2.767715	0.510882	-5.417524	0.0000
C(11)	0.509448	0.081353	6.262231	0.0000
C(12)	0.123211	0.048940	2.517595	0.0133
C(13)	0.182611	0.101597	1.797398	0.0751
C(14)	0.023549	0.026241	0.897414	0.3716
C(15)	0.485480	0.088796	5.467380	0.0000
C(16)	-0.002907	0.002519	-1.153915	0.2512
C(20)	-1.364387	0.824546	-1.654713	0.1010
C(21)	-0.309881	0.131300	-2.360093	0.0201
C(22)	0.944625	0.078987	11.95920	0.0000
C(23)	0.086933	0.163974	0.530161	0.5971
C(24)	0.024415	0.042352	0.576476	0.5655
C(25)	0.415637	0.143313	2.900193	0.0045
C(26)	-0.002543	0.004066	-0.625490	0.5330
C(30)	-1.566476	0.726523	-2.156129	0.0334
C(31)	-0.132370	0.115691	-1.144169	0.2552
C(32)	0.189123	0.069597	2.717401	0.0077
C(33)	0.621420	0.144481	4.301053	0.0000
C(34)	0.058630	0.037317	1.571123	0.1192
C(35)	0.458434	0.126276	3.630408	0.0004
C(36)	-0.005686	0.003583	-1.586888	0.1155
Determinant residual covariance		1.04E-11		

Equation: Y_1=C(10)+ C(11)*Y_1(-1)+C(12)*Y_2(-1)+C(13)*Y_3(-1)+C(14)
*T+C(15)*Z1+C(16)*T*Z1
Observations: 42

R-squared	0.998976	Mean dependent var	9.076151
Adjusted R-squared	0.998800	S.D. dependent var	0.381517
S.E. of regression	0.013215	Sum squared resid	0.006112
Durbin-Watson stat	2.045999		

Equation: Y_2=C(20)+C(21)*Y_1(-1)+C(22)*Y_2(-1)+C(23)*Y_3(-1)+C(24)
*T+C(25)*Z1+C(26)*T*Z1
Observations: 42

R-squared	0.997484	Mean dependent var	9.098903
Adjusted R-squared	0.997053	S.D. dependent var	0.392866
S.E. of regression	0.021328	Sum squared resid	0.015921
Durbin-Watson stat	1.612953		

Equation: Y_3=C(30)+C(31)*Y_1(-1)+C(32)*Y_2(-1)+C(33)*Y_3(-1)+C(34)
*T+C(35)*Z1+C(36)*T*Z1
Observations: 42

R-squared	0.998487	Mean dependent var	8.868850
Adjusted R-squared	0.998228	S.D. dependent var	0.446376
S.E. of regression	0.018792	Sum squared resid	0.012360
Durbin-Watson stat	1.697621		

Figure 1.31 *Statistical results based on a VAR model using the object system*

As an illustration and an extension of the model in Figure 1.30, Figure 1.31 presents the results of a VAR model using the system function, or the object "*System*". Based on these results, the following notes and conclusions are made.

1. The model represents a VAR model by entering "1 1" as the lag interval of endogenous. Refer to alternative lag intervals of alternative VAR models presented in Chapter 6 (Agung, 2009a), as well as the limitation of a VAR model compared to the system equation.
2. Since it is defined that the effect of *Z1* on *(Y_1,Y_2,Y_3)* depends on time, *t*, then the null hypothesis H_0: $C(15) = C(16) = C(25) = C(26) = C(35) = C(36) = 0$ should be tested at the first stage of testing the hypothesis. The null hypothesis can then be rejected based on the Chi-square test of $\chi_0^2 = 40.50483$ with $df = 6$ and a p-value $= 0.000$. Then we can conclude that the effect of *Z1* on *(Y_1,Y_2,Y_3)* is significantly dependent on the time *t*, adjusted or conditional for all other variables in the model.
3. Since some of the independent variables have large *p*-values, a reduced model should be explored. So, in general, three multiple regressions having different sets of independent variables are obtained. Therefore, the reduced model is not a VAR model anymore. Try this as an exercise.
4. In order to keep having a reduced VAR model, then one or two of the variables *t, Z1* or t^*Z1 should be deleted from the three regressions. However, note that each of the variables has significant positive or negative adjusted effects on *Y_3*, at a significance level of $\alpha = 0.10$. So, in a statistical sense, it is not wise to delete one of the variables from the third regression.
5. Based on each of the regressions, the following findings are derived.
   5.1 Based on the first regression, at a significance level of $\alpha = 0.10$, t^*Z1 has insignificant effect on *Y_1*, however, the null hypothesis H_0: $C(15) = C(16) = 0$ is rejected based on the Chi-square test

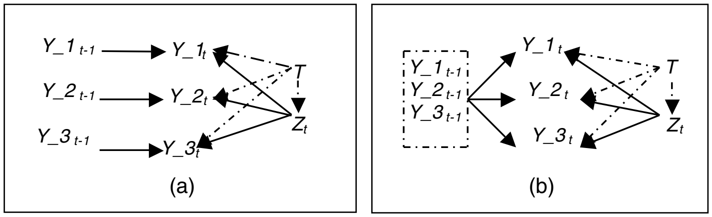


Figure 1.32 *Path diagrams of the models in (a) Figure 1.30, and (b) Figure 1.31*

of $\chi_0^2 = 537.53410$ with $df = 2$ and a p-value = 0.000. It can then be concluded that the effect of *Z1* on *Y_1* is significantly dependent on the time *t*, specifically the effect depends on a linear function of *t*, namely [*C(15)* + *C(16)***t*] adjusted for all variables in the model, not just the independent variable of the first regression.

5.2 On the other hand, even though each of *t* and *t***Z1* has an insignificant adjusted effect, it is found that the variables *t, Z1* and *t***Z1* have significant joint adjusted effects on *Y_1*, since the null hypothesis H_0: $C(14) = C(15) = C(16) = 0$ is rejected based on the Chi-square test of $\chi_0^2 = 42.67729 = 42.67\,729$ $df = 3$ and a p-value = 0.000. Based on this conclusion, if a reduced model should be obtained then, at most, two of the variables *t, Z1* and *t***Z1* can be deleted.

5.3 Similar analysis can easily be done based on the other two regressions. Do it as an exercise.

6. For a graphical illustration, Figure 1.32(a) and (b), respectively, presents the path diagrams of the models in Figures 1.30 and 1.31. Based on these diagrams, the following notes are made.

6.1 Note that Figure 1.32(a) shows that Y_i_{t-1} has a direct effect on each Y_i_t only, but Figure 1.32(b) shows that the trivariate $(Y\_1, Y\_2, Y\_3)_{t-1}$ has direct effect on Y_i_t.

6.2 The effect of Z_t on each endogenous variable Y_i_t, which is defined to be dependent on the time *t*, is represented as an arrow from *t* to Z_t, and then from Z_t to Y_i_t, and in the regression indicated by the term *(i3)* + *c(i4)***t)***Z* in Figure 1.30, and in Figure 1.31 by *(C(i5)* + *c(i6)***t)***Z*.

6.3 The possible causal effects between *Y_1,Y_2* and *Y_3* are not identified, however, their quantitative correlations are taken into account in the estimation process. If they should have a type of causal effect, then a new model should be defined; either additive, two- or three-way interaction models. Refer to the models demonstrated in Agung (2009a), as well as the following chapter.

7. As an extension of the model in Figure 1.31, we might consider Z_t as a function of *t*, then the following general model would also need to be considered.

$$
\begin{aligned}
y\_1 &= c(10) + c(11)^* y\_1(-1) + c(12)^* Y\_2(-1) + c(13)^* Y\_3(-1) + c(14)^* t + c(15)^* z1 + c(16)^* t^* z1 \\
y\_2 &= c(20) + c(21)^* Y\_1(-1) + c(22)^* Y\_2(-1) + c(23)^* Y\_3(-1) + c(24)^* t + c(25)^* z1 + c(26)^* t^* z1 \\
y\_3 &= c(30) + c(31)^* y\_1(-1) + c(32)^* Y\_2(-1) + c(33)^* Y\_3(-1) + c(34)^* t + c(35)^* z1 + c(36)^* t^* z1 \\
Z &= c(40) + F(t)
\end{aligned}
\tag{1.20}
$$

where *F(t)* is a function of the time, *t* with a finite number of parameters, without a constant parameter, such as $F(t) = C(41)^* log(t)$, and $F(t) = C(41)^* t + C(42)^* t^2 + \ldots + C(4k)^* t^k$.

Vector Error Correction Estimates
Date: 11/04/09 Time: 11:08
Sample (adjusted): 1952 1992
Included observations: 41 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1
Y_1(-1)	1.000000
Y_2(-1)	-0.188327
	(0.10802)
	[-1.74350]
Y_3(-1)	-0.697651
	(0.17410)
	[-4.00729]
C	-1.176237

Error Correction:	D(Y_1)	D(Y_2)	D(Y_3)
CointEq1	-0.559820	-0.313246	-0.154371
	(0.09342)	(0.16058)	(0.15937)
	[-5.99236]	[-1.95076]	[-0.96864]
D(Y_1(-1))	0.081726	-0.135375	0.207249
	(0.14843)	(0.25512)	(0.25320)
	[0.55061]	[-0.53063]	[0.81852]
D(Y_2(-1))	0.121231	0.116966	0.163731
	(0.11383)	(0.19566)	(0.19419)
	[1.06499]	[0.59780]	[0.84316]
D(Y_3(-1))	-0.276045	-0.307600	-0.164907
	(0.11378)	(0.19557)	(0.19410)
	[-2.42613]	[-1.57286]	[-0.84962]
C	-3.352072	-2.258550	-1.128369
	(0.64536)	(1.10927)	(1.10092)
	[-5.19409]	[-2.03607]	[-1.02493]
T	-0.004842	-0.063879	-0.006958
	(0.01300)	(0.02235)	(0.02218)
	[-0.37242]	[-2.85840]	[-0.31372]
Z1	0.375213	0.261613	0.129550
	(0.07155)	(0.12298)	(0.12206)
	[5.24404]	[2.12724]	[1.06139]
T*Z1	-0.000315	0.005822	0.000392
	(0.00129)	(0.00221)	(0.00219)
	[-0.24487]	[2.63512]	[0.17875]

R-squared	0.691509	0.534601	0.315831
Adj. R-squared	0.626071	0.435880	0.170704
Sum sq. resids	0.005403	0.015962	0.015722
S.E. equation	0.012795	0.021993	0.021827
F-statistic	10.56746	5.415263	2.176242
Log likelihood	124.9793	102.7720	103.0817
Akaike AIC	-5.706306	-4.623024	-4.638131
Schwarz SC	-5.371951	-4.288668	-4.303776
Mean dependent	0.029314	0.034892	0.035789
S.D. dependent	0.020924	0.029282	0.023969

Determinant resid covariance (dof adj.)	2.36E-11
Determinant resid covariance	1.23E-11
Log likelihood	340.4857
Akaike information criterion	-15.29199
Schwarz criterion	-14.16354

Figure 1.33 *Statistical results based on a VEC model*

Example 1.22 A VEC model

Figure 1.33 presents the statistical results based on a VEC model of the first differences between endogenous variables *DY_1, DY_2* and *DY_3*, exogenous variables *t, Z1* and t^*Z1, and "1 1" as the lag interval of endogenous variables. Refer to the characteristics of various VEC models and their limitations presented in Chapter 6 (Agung, 2009a).

1.9.3 Application of ARCH Model

Various TGARCH(a,b,c) time series models along with their limitations have been presented in Agung (2009a). For this reason, this section only presents the example of an ARCH(1) = TGARCH(1,0,0) model.

Example 1.23 A reduced ARCH(1) model

Figure 1.34 presents the statistical results based on a reduced ARCH(1) model, where its full mean model is presented in Figure 1.30. Based on these results, note the following:

1. Note that the regression of *Y_2* has only two independent variables, namely *Y_2(−1)* and t^*Z1, compared to the other two hierarchical regression models.
2. Based on the output, it can easily be derived that the effect of *Z1* on *(Y_1,Y_2,Y_3)* is significantly dependent on the time *t*. Otherwise, it can be tested using the Wald test.
3. The data supports that error terms have a multivariate Student's *t*-distribution based on z-Statistic of $Z_0 = 0.108\,608$ with a *p*-value = 0.9135.

System: SYS27
Estimation Method: ARCH Maximum Likelihood (Marquardt)
Covariance specification: Constant Conditional Correlation
Date: 11/04/09 Time: 13:19
Sample: 1951 1992
Included observations: 42
Total system (balanced) observations 126
Disturbance assumption: Student's t distribution
Presample covariance: backcast (parameter =0.7)
Convergence achieved after 162 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(10)	-1.858822	0.775986	-2.395431	0.0166
C(11)	0.690433	0.115008	6.003364	0.0000
C(12)	0.087609	0.040721	2.151449	0.0314
C(13)	0.490258	0.161171	3.041859	0.0024
C(14)	-0.008910	0.004033	-2.209131	0.0272
C(20)	1.103364	0.387836	2.844924	0.0044
C(21)	0.876077	0.046077	19.01343	0.0000
C(24)	0.000274	0.000157	1.746065	0.0808
C(30)	-0.605071	1.049288	-0.576649	0.5642
C(31)	0.667993	0.165157	4.044583	0.0001
C(32)	0.091821	0.063656	1.442456	0.1492
C(33)	0.361825	0.196797	1.838565	0.0660
C(34)	-0.008852	0.006192	-1.429544	0.1528
Variance Equation Coefficients				
C(35)	0.000268	0.000199	1.345405	0.1785
C(36)	-0.251133	0.520293	-0.482676	0.6293
C(37)	0.000375	0.000279	1.341809	0.1797
C(38)	0.231851	0.652811	0.355158	0.7225
C(39)	0.000392	0.000148	2.646134	0.0081
C(40)	-0.077905	0.445042	-0.175050	0.8610
C(41)	0.501524	0.259212	1.934804	0.0530
C(42)	0.511081	0.224333	2.278228	0.0227
C(43)	0.272723	0.253580	1.075494	0.2822
t-Distribution (Degree of Freedom)				
C(44)	52.61603	484.4564	0.108608	0.9135

Log likelihood	339.2191	Schwarz criterion	-14.10647
Avg. log likelihood	2.692215	Hannan-Quinn criter.	-14.70926
Akaike info criterion	-15.05805		

Equation: Y_1=C(10)+ C(11)*Y_1(-1)+C(12)*T+C(13)*Z1+C(14)*T*Z1

R-squared	0.998580	Mean dependent var	9.076151
Adjusted R-squared	0.998427	S.D. dependent var	0.381517
S.E. of regression	0.015133	Sum squared resid	0.008473
Durbin-Watson stat	1.495363		

Equation: Y_2=C(20)+C(21)*Y_2(-1)+C(24)*T*Z1

R-squared	0.996801	Mean dependent var	9.098903
Adjusted R-squared	0.996637	S.D. dependent var	0.392866
S.E. of regression	0.022782	Sum squared resid	0.020242
Durbin-Watson stat	1.560383		

Equation: Y_3=C(30)+C(31)*Y_3(-1)+C(32)*T+C(33)*Z1+C(34)*T*Z1

R-squared	0.998161	Mean dependent var	8.868850
Adjusted R-squared	0.997962	S.D. dependent var	0.446376
S.E. of regression	0.020149	Sum squared resid	0.015022
Durbin-Watson stat	1.516532		

Covariance specification: Constant Conditional Correlation
GARCH(i) = M(i) + A1(i)*RESID(i)(-1)^2
COV(i,j) = R(i,j)*@SQRT(GARCH(i)*GARCH(j))

Tranformed Variance Coefficients

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	0.000268	0.000199	1.345405	0.1785
A1(1)	-0.251133	0.520293	-0.482676	0.6293
M(2)	0.000375	0.000279	1.341809	0.1797
A1(2)	0.231851	0.652811	0.355158	0.7225
M(3)	0.000392	0.000148	2.646134	0.0081
A1(3)	-0.077905	0.445042	-0.175050	0.8610
R(1,2)	0.501524	0.259212	1.934804	0.0530
R(1,3)	0.511081	0.224333	2.278228	0.0227
R(2,3)	0.272723	0.253580	1.075494	0.2822

```
Estimation Command:
=====================
ARCH(DERIV=AA, TDIST) @CCC C ARCH(1)

Estimated Equations:
=====================
Y_1=C(10)+ C(11)*Y_1(-1)+C(12)*T+C(13)*Z1+C(14)*T*Z1
Y_2=C(20)+C(21)*Y_2(-1)+C(24)*T*Z1
Y_3=C(30)+C(31)*Y_3(-1)+C(32)*T+C(33)*Z1+C(34)*T*Z1

Substituted Coefficients:
=====================
Y_1=-1.859+ 0.690*Y_1(-1)+0.088*T+0.490*Z1-0.009*T*Z1
Y_2=1.103+0.876*Y_2(-1)+0.0003*T*Z1
Y_3=-0.605+0.668*Y_3(-1)+0.092*T+0.362*Z1-0.009*T*Z1

Variance and Covariance Representations:
=====================
GARCH(i) = M(i) + A1(i)*RESID(i)(-1)^2
COV(i,j) = R(i,j)*@SQRT(GARCH(i)*GARCH(j))
```

```
Variance and Covariance Equations:
=====================
GARCH1 = C(35) + C(36)*RESID1(-1)^2
GARCH2 = C(37) + C(38)*RESID2(-1)^2
GARCH3 = C(39) + C(40)*RESID3(-1)^2

COV1_2 = C(41)*@SQRT(GARCH1*GARCH2)
COV1_3 = C(42)*@SQRT(GARCH1*GARCH3)
COV2_3 = C(43)*@SQRT(GARCH2*GARCH3)

Substituted Coefficients:
=====================
GARCH1 = 0.000267768110001 - 0.251132976834*RESID1(-1)^2
GARCH2 = 0.000374611760699 + 0.231851204332*RESID2(-1)^2
GARCH3 = 0.000392378253889 - 0.0779045947747*RESID3(-1)^2

COV1_2 = 0.501523819388*@SQRT(GARCH1*GARCH2)
COV1_3 = 0.511080942222*@SQRT(GARCH1*GARCH3)
COV2_3 = 0.272723253637*@SQRT(GARCH2*GARCH3)
```

Figure 1.34 *Statistical results based on an ARCH(1) model*

1.9.4 The Application of Instrumental Variables Models

Based on the variables *Y_1, Y_2, Y_3,Y_1(−1), Y_2(−1), Y_3(−1), Z1* and the time *t* used in previous models, a lot of instrumental variables models can easily be subjectively defined. Corresponding to an instrumental variables model (IVM), Agung (2009a) states that there would be two stages of problems

(TSOP) in defining instrumental variables models, since the true population model can never be known and there is no general rule as to how to select the best possible set of instrumental variables.

Example 1.24 An application of the 2SLS estimation method
As an extension of the model in Figure 1.32, under the assumption that *Y_1, Y_2* and *Y_3* are correlated, and we define that the effect of *Z1* on the trivariate *(Y_1,Y_2,Y_3)* depends on the time *t*, then by using *trial-and-error methods*, the statistical results presented in Figure 1.34 are obtained using the 2SLS. Based on these results, the following notes and conclusions are made.

1. Even though each of *Y_2* and Y_3 has an insignificant adjusted effect on *Y_1*, the joint effects of *Y_2* and *Y_3* on *Y_1* are significant, since the null hypothesis H_0: $C(11) = C(12) = 0$ is rejected based on a Chi-square test of $\chi_0^2 = 324.1802$ with $df = 2$ and a p-value = 0.000. The same conclusions are obtained based on the other two regressions. Therefore, we can conclude that the data supports the assumption that variables *Y_1, Y_2* and *Y_3* are correlated.
2. In a statistical sense, a reduced model should be explored, since one of the independent variables in each regression has a p-value > 0.20 (or 0.25). Do it as an exercise.

Example 1.25 An application of the 3SLS estimation method
As a modification of the MAR(1)_IVM in Figure 1.35, the following system specification is considered.

$$
\begin{aligned}
y\_1 &= c(10) + c(11)^* y\_2 + c(12)^* Y\_3 + [ar(1) = c(13)]\,@\,c\,z1\,@\,t\,t^* z1 \\
y\_2 &= c(20) + c(21)^* Y\_1 + c(22)^* Y\_3 + [ar(1) = c(23)]\,@\,c\,z1\,@\,t\,t^* z1 \\
y\_3 &= c(30) + c(31)^* Y\_1 + c(32)^* Y\_2 + [ar(1) = c(33)]\,@\,c\,z1\,@\,t\,t^* z1
\end{aligned}
\tag{1.21}
$$

However, an error message of "*Near Singular Matrix*" is obtained so trial-and-error methods should be applied to delete one or two of the variables from the model in (1.21). Finally, an unexpected good fit model is obtained, in a statistical sense, since each of the independent variables has significant adjusted effect with a p-value = 0.000, as presented in Figure 1.36.

System: SYS28
Estimation Method: Iterative Two-Stage Least Squares
Date: 11/04/09 Time: 15:51
Sample: 1951 1992
Included observations: 43
Total system (balanced) observations 126
Convergence achieved after 22 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	0.107289	1.086964	0.098706	0.9215
C(11)	0.626658	0.435041	1.440455	0.1525
C(12)	0.365074	0.339414	1.075602	0.2844
C(13)	0.824692	0.062984	13.09361	0.0000
C(20)	1.609625	0.672274	2.394299	0.0183
C(21)	0.201697	0.351989	0.573021	0.5678
C(22)	0.640546	0.303956	2.107366	0.0373
C(23)	0.776692	0.068356	11.36242	0.0000
C(30)	-2.091905	0.534275	-3.915405	0.0002
C(31)	0.205612	0.417156	0.492891	0.6230
C(32)	0.997769	0.435167	2.292841	0.0237
C(33)	0.753684	0.080540	9.357851	0.0000

Determinant residual covariance 7.79E-14

Equation: Y_1=C(10)+ C(11)*Y_2+C(12)*Y_3+[AR(1)=C(13)]
Instruments: C T Z1 T*Z1 Y_1(-1) Y_2(-1) Y_3(-1)
Observations: 42

R-squared	0.998316	Mean dependent var	9.076151
Adjusted R-squared	0.998183	S.D. dependent var	0.381517
S.E. of regression	0.016262	Sum squared resid	0.010049
Durbin-Watson stat	1.542501		

Equation: Y_2=C(20)+C(21)*Y_1 +C(22)*Y_3+[AR(1)=C(23)]
Instruments: C T Z1 T*Z1 Y_2(-1) Y_1(-1) Y_3(-1)
Observations: 42

R-squared	0.997385	Mean dependent var	9.098903
Adjusted R-squared	0.997179	S.D. dependent var	0.392866
S.E. of regression	0.020868	Sum squared resid	0.016548
Durbin-Watson stat	1.662941		

Equation: Y_3=C(30)+C(31)*Y_1+C(32)*Y_2+[AR(1)=C(33)]
Instruments: C T Z1 T*Z1 Y_3(-1) Y_1(-1) Y_2(-1)
Observations: 42

R-squared	0.997115	Mean dependent var	8.868850
Adjusted R-squared	0.996887	S.D. dependent var	0.446376
S.E. of regression	0.024905	Sum squared resid	0.023569
Durbin-Watson stat	1.713668		

Figure 1.35 *Statistical results based on a MAR(**1**) instrumental variables model, using the 2SLS estimation method*

System: SYS29
Estimation Method: Three-Stage Least Squares
Date: 11/04/09 Time: 16:35
Sample: 1951 1992
Included observations: 43
Total system (balanced) observations 126
Iterate coefficients after one-step weighting matrix
Convergence achieved after: 1 weight matrix, 467 total coef iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	2.370430	0.412806	5.742231	0.0000
C(11)	-0.494364	0.142092	-3.479179	0.0007
C(12)	1.263649	0.123065	10.26818	0.0000
C(13)	0.752916	0.086334	8.721014	0.0000
C(20)	4.679244	0.765912	6.109373	0.0000
C(21)	-1.947988	0.347784	-5.601151	0.0000
C(22)	2.492620	0.320799	7.770038	0.0000
C(23)	0.750676	0.086492	8.679110	0.0000
C(30)	-2.230141	0.518800	-4.298655	0.0000
C(32)	1.217461	0.055994	21.74282	0.0000
C(33)	0.753724	0.053324	14.13487	0.0000
Determinant residual covariance		4.36E-15		

Equation: Y_1=C(10)+ C(11)*Y_2+C(12)*Y_3+[AR(1)=C(13)]
Instruments: C Z1 @T T*Z1 Y_1(-1) Y_2(-1) Y_3(-1)
Observations: 42

R-squared	0.995587	Mean dependent var	9.076151
Adjusted R-squared	0.995239	S.D. dependent var	0.381517
S.E. of regression	0.026326	Sum squared resid	0.026336
Durbin-Watson stat	1.735812		

Equation: Y_2=C(20)+C(21)*Y_1 +C(22)*Y_3+[AR(1)=C(23)]
Instruments: C Z1 @T T*Z1 Y_2(-1) Y_1(-1) Y_3(-1)
Observations: 42

R-squared	0.983786	Mean dependent var	9.098903
Adjusted R-squared	0.982506	S.D. dependent var	0.392866
S.E. of regression	0.051963	Sum squared resid	0.102606
Durbin-Watson stat	1.733474		

Equation: Y_3=C(30)+C(32)*Y_2+[AR(1)=C(33)]
Instruments: C Z1 @T T*Z1 Y_3(-1) Y_2(-1)
Observations: 42

R-squared	0.996452	Mean dependent var	8.868850
Adjusted R-squared	0.996270	S.D. dependent var	0.446376
S.E. of regression	0.027261	Sum squared resid	0.028983
Durbin-Watson stat	1.678885		

Figure 1.36 *Statistical results based on a MAR(**1**) instrumental variables model, using the 3SLS estimation method*

Note that the regression of *Y_3* only has a single independent variable *Y_2*. The reader could try deleting other variable(s) from the model in (1.20), including modifying the instrumental variables, but leaving the AR(1) terms as they are.

1.10 Models with an Environmental Multivariate

If an endogenous variable by states, namely Y_i_t, $i = 1, \ldots, N$, is known or defined to be effected by the same environmental multivariate, say $\mathbf{Z}_t = (Z1_t, \ldots, Zk_t, \ldots)$, then the set of Y_i_t, $i = 1, \ldots, N$, should be correlated, including the possibility of having causal relationships for some states. As an illustration, the following section presents selected models using two endogenous variables *Y_1* and *Y_2*, which could easily be extended to three or more states.

1.10.1 Bivariate Correlation and Simple Linear Regressions

Data analysis based on the bivariate correlation of Y_i_t and Y_j_t, the simple linear regression of Y_i_t on Y_j_t, and the simple linear regression of Y_j_t on Y_i_t, would give exactly the same values of the t-statistic for testing the hypothesis that Y_j_t is a causal factor of Y_i_t, as well as Y_i_t and Y_j_t having simultaneous causal effects.

Based on these findings, it can be concluded that correlation analysis can be used to test the hypothesis stated earlier. On the other hand, it could be said that independent of a model, the independent variable may not be a causal factor of the corresponding dependent variable. Note that the causal relationship between any pair of variables should be derived based on a strong theoretical foundation: it is not based on the conclusion of testing a hypothesis. See the following example.

Example 1.26 Special findings

Figure 1.37(a), (b) and (c), respectively, present the statistical results based on the bivariate correlation of Y_1_t and Y_2_t, the simple linear regression (SLR) of Y_1_t on Y_2_t, and the SLR of Y_2_t on Y_1_t, which show exactly the same values of the t-statistic of $t_0 = 46.39045$.

Covariance Analysis: Ordinary
Date: 11/18/09 Time: 08:21
Sample: 1950 1992
Included observations: 43

Correlation / t-Statistic / Probability	Y_1	Y_2
Y_1	1.000000	

Y_2	0.990608	1.000000
	46.39045	-----
	0.0000	-----

(a)

Dependent Variable: Y_1
Method: Least Squares
Date: 11/18/09 Time: 08:22
Sample: 1950 1992
Included observations: 43

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.519552	0.184257	2.819716	0.0074
Y_2	0.940785	0.020280	46.39045	0.0000

R-squared	0.981305	Mean dependent var	9.058594
Adjusted R-squared	0.980849	S.D. dependent var	0.394138
S.E. of regression	0.054544	Akaike info criterion	-2.934225
Sum squared resid	0.121977	Schwarz criterion	-2.852309
Log likelihood	65.08584	Hannan-Quinn criter.	-2.904017
F-statistic	2152.074	Durbin-Watson stat	0.173408
Prob(F-statistic)	0.000000		

(b)

Dependent Variable: Y_2
Method: Least Squares
Date: 11/18/09 Time: 08:24
Sample: 1950 1992
Included observations: 43

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.372242	0.203867	-1.825905	0.0752
Y_1	1.043070	0.022485	46.39045	0.0000

R-squared	0.981305	Mean dependent var	9.076509
Adjusted R-squared	0.980849	S.D. dependent var	0.415012
S.E. of regression	0.057433	Akaike info criterion	-2.831016
Sum squared resid	0.135238	Schwarz criterion	-2.749099
Log likelihood	62.86684	Hannan-Quinn criter.	-2.800807
F-statistic	2152.074	Durbin-Watson stat	0.177446
Prob(F-statistic)	0.000000		

(c)

Figure 1.37 *Statistical results based on (a) covariance analysis of* Y_1_t *and* Y_2_t*, (b) the SLR of* Y_1_t *on* Y_2_t*, and (c) the SLR of* Y_1_t *on* Y_2_t

Based on these results, the following notes and comments are presented.

1. Even though the regressions have very small DW-statistics, because their R-squares are very large, namely $R^2 = 0.981\,217$, then the SLR should be considered to be a good fit.
2. As an alternative analysis, Figure 1.38(a) presents the statistical results based on system equations of two SLRs using the LS estimation method, where both SLRs also show the same values of the *t*-statistic of $t_0 = 46.39\,045$. Thus the results of these system equations can be represented by the result of the covariance analysis in Figure 1.37(a). In other words, the simultaneous causal effects of *Y_1* and *Y_2* tested using the system equations in Figure 1.38(a) can be substituted by covariance analysis.

Estimation Method: Least Squares
Date: 11/18/09 Time: 09:02
Sample: 1950 1992
Included observations: 43
Total system (balanced) observations 86

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	0.519552	0.184257	2.819716	0.0060
C(11)	0.940785	0.020280	46.39045	0.0000
C(20)	-0.372242	0.203867	-1.825905	0.0715
C(21)	1.043070	0.022485	46.39045	0.0000

Determinant residual covariance 1.67E-07

Equation: Y_1=C(10)+C(11)*Y_2
Observations: 43

R-squared	0.981305	Mean dependent var	9.058594
Adjusted R-squared	0.980849	S.D. dependent var	0.394138
S.E. of regression	0.054544	Sum squared resid	0.121977
Durbin-Watson stat	0.173408		

Equation: Y_2=C(20)+C(21)*Y_1
Observations: 43

R-squared	0.981305	Mean dependent var	9.076509
Adjusted R-squared	0.980849	S.D. dependent var	0.415012
S.E. of regression	0.057433	Sum squared resid	0.135238
Durbin-Watson stat	0.177446		

(a)

Estimation Method: Seemingly Unrelated Regression
Date: 11/18/09 Time: 09:34
Sample: 1950 1992
Included observations: 43
Total system (balanced) observations 86
Linear estimation after one-step weighting matrix

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	0.438979	0.127950	3.430871	0.0009
C(11)	0.949662	0.014068	67.50334	0.0000
C(20)	-0.461398	0.141554	-3.259524	0.0016
C(21)	1.052913	0.015598	67.50334	0.0000

Determinant residual covariance 3.78E-12

Equation: Y_1=C(10)+C(11)*Y_2
Observations: 43

R-squared	0.981217	Mean dependent var	9.058594
Adjusted R-squared	0.980759	S.D. dependent var	0.394138
S.E. of regression	0.054671	Sum squared resid	0.122547
Durbin-Watson stat	0.176200		

Equation: Y_2=C(20)+C(21)*Y_1
Observations: 43

R-squared	0.981217	Mean dependent var	9.076509
Adjusted R-squared	0.980759	S.D. dependent var	0.415012
S.E. of regression	0.057567	Sum squared resid	0.135870
Durbin-Watson stat	0.176219		

(b)

Figure 1.38 *Statistical results based on (a) a system equation of SLRs of* Y_1_t *and* Y_2_t*, and (b) a MAR(**1**) of* Y_1_t *on* Y_2

3. For a comparison, Figure 1.38(b) presents the statistical results using the SUR estimation method, which show different values of the *t*-statistic of $t_0 = 67.50\,334$ for both SLRs.
4. Compared to the regressions in Figure 1.37(b), the four regressions in Figure 1.38 also have very small DW-statistics, but very large R-squares.
5. Further analysis can easily be done based on MLV(***p***), MAR(***q***) or MLVAR(***p,q***) models using the variables *Y_1* and *Y_2*, which will have much larger DW-statistics. Do it as an exercise.

1.10.2 Simple Models with an Environmental Multivariate

Since Y_1_t and Y_2_t are correlated, specifically linearly correlated, then simple models of Y_1_t and Y_2_t with an environmental multivariate $\boldsymbol{Z}_t$ and the time t as independent variables will have the following general equation.

$$\begin{aligned} Y\_1_t &= C(10) + C(11)^* Y\_2_{t-j} + F(t, Z_t, C(1^*)) + \mu_{1t} \\ Y\_2_t &= C(20) + C(21)^* Y\_1_{t-j} + F(t, Z_t, C(2^*)) + \mu_{1t} \end{aligned} \tag{1.22}$$

for a subscript $j \geq 0$, where $F(t, \boldsymbol{Z}_t, ^*)$ can be any functions $Z1_t, \ldots, Zk_t, \ldots$, and the time t, including some selected two- and three-ways of their interactions, with a finite number of parameters but no constant parameter. For instance, if the effect of *Z1* on *Y_i* depends on *Z2*, then $Z1^*Z2$ should be used as an independent variable or a term of the function $F(t, \boldsymbol{Z}_t, ^*)$. Note that there would be a lot of possible time-series models.

However, the following four groups of models will be considered, corresponding to selected forms of the function $F(t, \boldsymbol{Z}_t, ^*)$, such as follows:

1. Additive models or two functions, namely $F(t, \boldsymbol{Z}_t, ^*) = F_1(t, ^*) + F_2(\boldsymbol{Z}_t, ^*)$.
   Refer to the models (1.4a) to (1.4d) for the alternative functions of $F_1(t, ^*)$, and $F_2(\boldsymbol{Z}_t, ^*)$ can be additive or interaction functions of the components of $\boldsymbol{Z}_t$.
2. Models with trend: $F(t, \boldsymbol{Z}_t, C(i^*) = C(i2)^* t + F_2(\boldsymbol{Z}_t, C(ik)),\ i = 1,2$, and $k > 2$.
3. Models with Trend-Related Effects (TRE):

$$F(t, \boldsymbol{Z}_t, C(i^*) = C(i2)^* t + F_2(\boldsymbol{Z}_t, C(ik)) + t^* F_2(\boldsymbol{Z}_t, C(i^*)), \quad \text{for } i = 1, 2, \text{and } k > 2.$$

4. Models without the time t: $F(t, \boldsymbol{Z}_t, ^*) = F_2(\boldsymbol{Z}_t, ^*)$. Refer to all seemingly causal models (SCMs) presented in Chapter 4 (Agung, 2009a).

Comparing these to the models in Figure 1.38, the models in (1.22), for $j = 0$ in fact show that Y_1_t and Y_2_t have simultaneous causal relationships.

To generalize, the following general model can be applied

$$\begin{aligned} G_1(Y\_1_t) &= C(10) + C(11)^* Y\_2_{t-j} + F(t, Z_t, C(1^*)) + \mu_{1t} \\ G_2(Y\_2_t) &= C(20) + C(21)^* Y\_1_{t-j} + F(t, Z_t, C(2^*)) + \mu_{1t} \end{aligned} \tag{1.23}$$

where $G_i(Y\_i_t)$ is a function of Y_i_t having no parameter, such as $G_i(Y\_i_t) = Y\_i_t$, $log(Y\_i_t)$ or $log[(Y\_i_t\text{-}Li)/(Ui\text{-}Y\_i_t)]$, where *Li* and *Ui* are the lower and upper bounds of Y_i_t, which should be subjectively selected by the researchers.

If there are correlated endogenous variables for three states, namely *Y_1, Y_2* and *Y_3*, then the simple models of $G_i(Y\_i_t)$, $i = 1,2$ and 3 will have the general equation as follows:

$$\begin{aligned} G_1(Y\_1_t) &= C(10) + C(11)^*Y\_2_{t-j} + C(12)^*Y\_3_{t-j} + F(t, Z_t, C(1^*)) + \mu_{1t} \\ G_2(Y\_2_t) &= C(20) + C(21)^*Y\_1_{t-j} + C(22)^*Y\_3_{t-j} + F(t, Z_t, C(2^*)) + \mu_{1t} \\ G_3(Y\_3_t) &= C(30) + C(31)^*Y\_1_{t-j} + C(32)^*Y\_2_{t-j} + F(t, Z_t, C(3^*)) + \mu_{1t} \end{aligned} \tag{1.24}$$

1.10.3 The VAR Models

1.10.3.1 Basic General VAR Models

For illustration, a VAR model of *Y_1* and *Y_2* with "1 *p*" as the lag intervals for the endogenous variables, and the time *t* and $\mathbf{Z}_t$ as exogenous variables will be considered. The model considered has the following general equation.

$$\begin{aligned} Y\_1_t &= C(110) + \sum_{j=1}^{p} C(11j)^*Y\_1_{t-j} + \sum_{j=1}^{p} C(12j)^*Y\_2_{t-j} + F(t, Z_t, C(13^*)) + \varepsilon_{1t} \\ Y\_2_t &= C(210) + \sum_{j=1}^{p} C(21j)^*Y\_1_{t-j} + \sum_{j=1}^{p} C(22j)^*Y\_2_{t-j} + F(t, Z_t, C(23^*)) + \varepsilon_{1t} \end{aligned} \tag{1.25}$$

1.10.3.2 Special VAR Interaction Models

With multivariate environmental variables, it is generally known that an effect of at least one of its components on the endogenous variables depends on the other component(s). Under these criteria, this section presents three alternative VAR interaction models of *Y_1*, *Y_2* and *Y_3* with the lag intervals for the endogenous: "1 1", and the environmental variables *Z1* and *Z2*, such as follows:

1. A VAR interaction model with trend:

$$\begin{aligned} Y\_i = {} & C(i0) + C(i1)^*Y\_1(-1) + C(i2)^*Y\_2(-1) + C(i3)^*Y\_3(-1) \\ & + C(i4)^*t + C(i5)^*Z1 + C(i6)^*Z2 + C(i7)^*Z1^*Z2 \\ for\ i = {} & 1,2,3 \end{aligned} \tag{1.26}$$

2. A VAR interaction model with time-related effects:

$$\begin{aligned} Y\_i = {} & C(i0) + C(i1)^*Y\_1(-1) + C(i2)^*Y\_2(-1) + C(i3)^*Y\_3(-1) \\ & + C(i4)^*t + C(i5)^*Z1 + C(i6)^*Z2 + C(i7)^*Z1^*Z2 \\ & + C(i8)^*t^*Z1 + C(i9)^*t^*Z2 + C(i10)^*t^*Z1^*Z2 \\ for\ i = {} & 1,2,3 \end{aligned} \tag{1.27}$$

3. A VAR interaction model without the time *t*:

$$\begin{aligned} Y\_i = {} & C(i0) + C(i1)^*Y\_1(-1) + C(i2)^*Y\_2(-1) + C(i3)^*Y\_3(-1) \\ & + C(i4)^*Z1 + C(i5)^*Z2 + C(i6)^*Z1^*Z2 \\ for\ i = {} & 1,2,3 \end{aligned} \tag{1.28}$$

Corresponding to these VAR interaction models, note the following:

1. In practice, a reduced model obtained would be a good fit in a statistical sense, because the three variables *Z1, Z2* and *Z1*Z2* are highly or significantly correlated in general. More so for the independent variables of the model in (2.27).
2. Since it is defined that the effects of *Z1 (Z2)* on *Y_i, i = 1,2* and *3* depend on *Z2 (Z1)* in a theoretical sense, then the interaction *Z1*Z2* should be used in the model, as well as in the reduced model(s). So a reduced model should be obtained by deleting either *Z1* or *Z2*, or both *Z1* and *Z2*. Note that a model can be considered an acceptable or good fit, even though some of its independent variables have insignificant adjusted effects.
3. Note that three models here are hierarchical two- and three-way interaction models. However, corresponding to the earlier notes, an empirical acceptable model obtained would be non-hierarchical in general. See the following example.

Example 1.27 A reduced VAR interaction model

Figure 1.39 presents the statistical results based on two reduced models of the VAR interaction model in (1.26). Based on these results, the following conclusions and notes are made.

1. By using the full model in (1.26), each of the independent variables *Z1, Z2* and *Z1∗Z2* has insignificant adjusted effects. By deleting either *Z1* or *Z2*, the results in Figure 1.39(a) and (b) are obtained.

Vector Autoregression Estimates
Date: 11/17/09 Time: 13:53
Sample (adjusted): 1951 1992
Included observations: 42 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	Y_1	Y_2	Y_3
Y_1(-1)	0.491574	-0.332933	-0.140681
	(0.07191)	(0.13066)	(0.11241)
	[6.83628]	[-2.54814]	[-1.25148]
Y_2(-1)	0.138289	0.957298	0.220475
	(0.04156)	(0.07551)	(0.06497)
	[3.32759]	[12.6773]	[3.39357]
Y_3(-1)	-0.085190	-0.214036	0.337287
	(0.10994)	(0.19976)	(0.17187)
	[-0.77488]	[-1.07144]	[1.96245]
C	2.285288	4.147653	4.394102
	(1.24927)	(2.26998)	(1.95300)
	[1.82929]	[1.82717]	[2.24992]
T	-0.002228	0.002815	0.004795
	(0.00155)	(0.00282)	(0.00242)
	[-1.43840]	[1.00008]	[1.97991]
Z1	-0.005993	-0.105799	-0.174082
	(0.11600)	(0.21078)	(0.18135)
	[-0.05166]	[-0.50194]	[-0.95994]
Z1*Z2	0.023509	0.025581	0.027959
	(0.00535)	(0.00972)	(0.00836)
	[4.39597]	[2.63254]	[3.34432]

(a)

Vector Autoregression Estimates
Date: 11/17/09 Time: 14:31
Sample (adjusted): 1951 1992
Included observations: 42 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	Y_1	Y_2	Y_3
Y_1(-1)	0.493659	-0.332149	-0.128280
	(0.07233)	(0.13155)	(0.11174)
	[6.82556]	[-2.52493]	[-1.14802]
Y_2(-1)	0.136375	0.953908	0.205518
	(0.04279)	(0.07783)	(0.06611)
	[3.18684]	[12.2557]	[3.10854]
Y_3(-1)	-0.094441	-0.216255	0.283936
	(0.11487)	(0.20893)	(0.17747)
	[-0.82216]	[-1.03507]	[1.59991]
C	2.264351	3.244793	3.073037
	(0.40767)	(0.74149)	(0.62984)
	[5.55438]	[4.37606]	[4.87909]
T	-0.001988	0.003159	0.006566
	(0.00198)	(0.00360)	(0.00306)
	[-1.00489]	[0.87810]	[2.14866]
Z2	0.020480	0.096814	0.240974
	(0.11596)	(0.21091)	(0.17915)
	[0.17662]	[0.45904]	[1.34510]
Z1*Z2	0.021824	0.014612	0.004119
	(0.00872)	(0.01586)	(0.01347)
	[2.50330]	[0.92153]	[0.30583]

(b)

Figure 1.39 *Statistical results based on two reduced models of the model in (1.26)*

2. The model in Figure 1.39(a) is a better model, in a statistical sense, since the effect of the interaction $Z1^{*}Z2$ on each *Y_i*, $i = 1,2$ and *3* is significant, based on *t*-statistics greater than 2.6. Based on this model, the following conclusions are derived.
   2.1 The data supports the hypothesis stated that the effect of *Z1 (Z2)* on each *Y_i* depends on *Z2 (Z1)*.
   2.2 An disadvantage of this model is *Z1* has a negative adjusted effect on each *Y_i*, in fact *Z1* and *Y_i* are significantly positive correlated, which shows the unexpected impact of the multicollinearity between the independent variables, specifically between *Z1* and $Z1^{*}Z2$.
   2.3 Furthermore, since *Z1* has insignificant adjusted effect on each *Y_i*, based on such a small *t*-statistics, then *Z1* could be deleted. Try it as an exercise.
3. On the other hand, based on the results in Figure 1.39(b) we draw the following conclusions.
   3.1 Since the interaction $Z1^{*}Z2$ has a significant adjusted effect on *Y_1*, it cannot be deleted from the VAR model.
   3.2 Since *Z2* has an insignificant adjusted effect on each *Y_i*, based on such a small *t*-statistics, then *Z2* could be deleted. The reduced model obtained would be the same as the reduced model by deleting *Z1* from the model in Figure 1.39(a). We find the final reduced model can be considered the best fit, conditional for the data used.

Example 1.28 Additional analyses for a VAR model

As an illustration, the VAR model in Figure 1.39(a) will be referred to. EViews provides so many alternative options for doing additional analyses for a VAR model. By selecting *View/Residuals Tests*, the options in Figure 1.40(a) shown on the screen, and Figure 1.40(b) obtained by selecting *View/Lag Structure*. However, only several analyses will be demonstrated, such as follows:

1. *Residual Analysis*
   1.1 *Residual Autocorrelation Tests*

Figure 1.41 presents the two statistics for testing the residual autocorrelation, which shows the null hypothesis, no residual autocorrelation up to lag 4, is accepted. As a result, the VAR model does not have the autocorrelation problem.

   1.2 *Basic Assumptions of Residuals*

Figure 1.42 shows that the null hypothesis, residuals are multivariate normal, is accepted. So it can be concluded that the data supports a basic assumption of the residuals. The other assumption is the

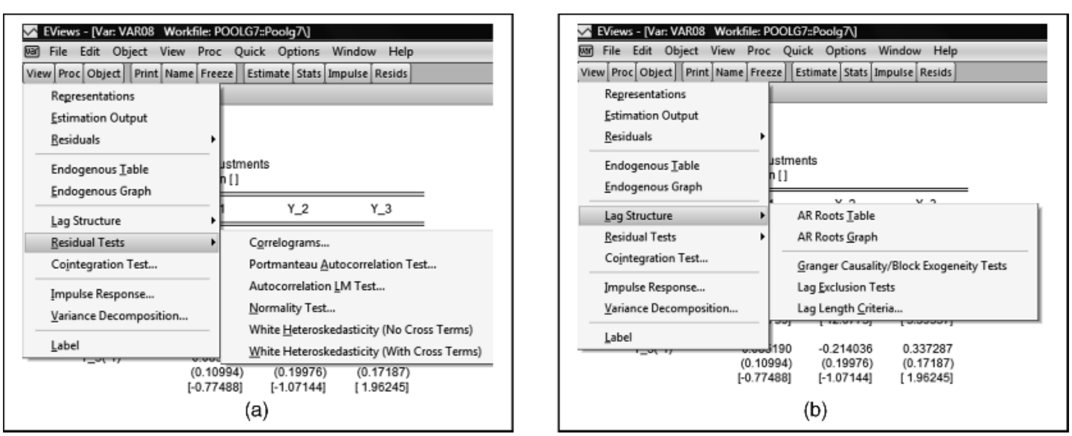


Figure 1.40 *Options for residual and lag structure, using EViews 6 or 7 Beta*

VAR Residual Portmanteau Tests for Autocorrelations
Null Hypothesis: no residual autocorrelations up to lag h
Date: 11/19/09 Time: 09:42
Sample: 1950 1992
Included observations: 42

Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	3.150951	NA*	3.227804	NA*	NA*
2	14.16143	0.5867	14.78880	0.5402	16
3	23.67465	0.5382	25.03382	0.4605	25
4	34.52254	0.4428	37.02358	0.3312	34

*The test is valid only for lags larger than the VAR lag order.
df is degrees of freedom for (approximate) chi-square distribution
*df and Prob. may not be valid for models with exogenous variables

VAR Residual Serial Correlation LM T...
Null Hypothesis: no serial correlation ...
Date: 11/19/09 Time: 09:37
Sample: 1950 1992
Included observations: 42

Lags	LM-Stat	Prob
1	5.867162	0.7531
2	12.27446	0.1983
3	8.502711	0.4844
4	10.25424	0.3303

Probs from chi-square with 9 df.

Figure 1.41 *The residual autocorrelation tests for the VAR model in Figure 1.39(a)*

VAR Residual Normality Tests
Orthogonalization: Cholesky (Lutkepohl)
Null Hypothesis: residuals are multivariate normal
Date: 11/19/09 Time: 09:52
Sample: 1950 1992
Included observations: 42

Component	Skewness	Chi-sq	df	Prob.
1	0.274744	0.528389	1	0.4673
2	-0.235376	0.387813	1	0.5335
3	-0.156196	0.170780	1	0.6794
Joint		1.086983	3	0.7802

Component	Kurtosis	Chi-sq	df	Prob.
1	2.234073	1.026626	1	0.3110
2	4.403388	3.446619	1	0.0634
3	2.912283	0.013465	1	0.9076
Joint		4.486710	3	0.2135

Component	Jarque-Bera	df	Prob.
1	1.555016	2	0.4595
2	3.834432	2	0.1470
3	0.184245	2	0.9120
Joint	5.573693	6	0.4726

Figure 1.42 *The residual normality tests for the VAR model in Figure 1.39(a)*

heterokedasticity of the residuals, which can easily be done by selecting the *Residual Tests/White heterokedasticity(*)*. Corresponding to the testing of the basic assumptions of the residuals, refer to the special notes and comments presented in Section 2.14.3 (Agung, 2009a).

2. *The Lag Structure*

2.1 *The AR Roots*

By selecting *Lag Structure/AR Roots*, it is found that the three AR Roots are strictly less than one. Then we can conclude that the VAR satisfies the stability condition.

2.2 *Granger Causality Tests*

By selecting *Lag Structure/Granger Causality/Block Exogeneity Wald Tests*, the results in Figure 1.43 are obtained for making conclusions of the corresponding tests. For example, *Y_1* and *Y_2* have significant Granger causalities with the *p*-values of 0.0009 and 0.0108, respectively, but *Y_1* and *Y_3* have insignificant Granger causalities with *p*-values of 0.4384 and 0.2108, respectively.

However, the three variables *Y_1, Y_2* and *Y_3* have significant Granger causalities with *p*-values of 0.0029, 0.0015 and 0.0019, respectively.

2.3 *The VAR Lag Exclusion Wald Tests*

Based on the results in Figure 1.44 we can conclude that the first lags *Y_1(−1)*, *Y_2(−1)* and *Y_3(−1)* have significant joint effects on each of *Y_1*,*Y_2* and *Y_3*, as well as on the trivariate *(Y_1,Y_2,Y_3)*.

VAR Granger Causality/Block Exogeneity Wald Tests
Date: 11/18/09 Time: 15:28
Sample: 1950 1992
Included observations: 42

Dependent variable: Y_1

Excluded	Chi-sq	df	Prob.
Y_2	11.07285	1	0.0009
Y_3	0.600436	1	0.4384
All	11.66778	2	0.0029

Dependent variable: Y_2

Excluded	Chi-sq	df	Prob.
Y_1	6.493032	1	0.0108
Y_3	1.147988	1	0.2840
All	12.99838	2	0.0015

Dependent variable: Y_3

Excluded	Chi-sq	df	Prob.
Y_1	1.566190	1	0.2108
Y_2	11.51633	1	0.0007
All	12.56475	2	0.0019

Figure 1.43 *Statistical results for the VAR Granger causality tests*

VAR Lag Exclusion Wald Tests
Date: 11/19/09 Time: 15:54
Sample: 1950 1992
Included observations: 42

Chi-squared test statistics for lag exclusion:
Numbers in [ ] are p-values

	Y_1	Y_2	Y_3	Joint
Lag 1	92.81195	187.7680	28.13310	293.2393
	[0.000000]	[0.000000]	[3.41e-06]	[0.000000]
df	3	3	3	9

Figure 1.44 *The VAR lag exclusion Wald tests*

However, in general, the joint effects of the exogenous variables of a VAR cannot be tested using the VAR model. For this reason, Agung (2009a) recommends applying the object *System*, instead of the VAR model, since by using the object *System*, each regression in the model can have a different set of independent variables, and various hypotheses can easily be tested using Wald tests.

2.4 *The Lag Order Selection Criteria*

By selecting *Lag Structure/Lag Length Criteria* . . . , and then insert the lags to include = 2 . . . *OK*, the results in Figure 1.45 are obtained. These results show that 1 (one) is the lag order selected by the five criteria. Therefore, we can conclude that the VAR model is best based on these five criteria.

VAR Lag Order Selection Criteria
Endogenous variables: Y_1 Y_2 Y_3
Exogenous variables: C T Z1 Z1*Z2
Date: 11/19/09 Time: 11:01
Sample: 1950 1992
Included observations: 41

Lag	LogL	LR	FPE	AIC	SC	HQ
0	291.6335	NA	2.40e-10	-13.64066	-13.13912	-13.45803
1	355.7553	106.3485*	1.64e-11*	-16.32953*	-15.45185*	-16.00992*
2	359.8311	6.163353	2.13e-11	-16.08932	-14.83549	-15.63275

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Figure 1.45 *The VAR lag order selection criteria using two lags*

VAR Lag Order Selection Criteria
Endogenous variables: Y_1 Y_2 Y_3
Exogenous variables: C T Z1 Z1*Z2
Date: 11/19/09 Time: 11:14
Sample: 1950 1992
Included observations: 38

Lag	LogL	LR	FPE	AIC	SC	HQ
0	288.6945	NA	9.53e-11	-14.56287	-14.04574	-14.37888
1	331.8760	70.45396	1.59e-11	-16.36189	-15.45691*	-16.03991
2	338.3978	9.611067	1.86e-11	-16.23146	-14.93863	-15.77148
3	351.5898	17.35788	1.57e-11	-16.45209	-14.77141	-15.85412
4	366.8876	17.71331*	1.22e-11*	-16.78356*	-14.71503	-16.04759*
5	375.2866	8.398951	1.43e-11	-16.75192	-14.29555	-15.87796

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Figure 1.46 *The VAR lag order selection criteria using five lags*

However, we find that by inserting lags to include greater than 2, contradictory conclusions could be obtained. As an illustration, Figure 1.46 presents another result using five lags, which shows that 1 (one) is the lags' order selected by the SC criterion only. On the other hand, an error message is obtained, "*Near singular matrix*", using 10 lags. These contradictory findings lead to a great problem, since there are 9 (nine) alternatives results using 1–9 lags to consider. Based on the author's point of view, the simplest possible model should be the best selection. As an exercise, do the analysis based on a VAR model using the lags interval for the endogenous "4 4" and "1 4".

1.10.3.3 Special Notes and Comments

Corresponding to the environmental multivariate $\mathbf{Z}_t = (Z1_t, \ldots, ZK_t)$, which has been defined or known to be the causal factor of the set of the endogenous variables $Y\_i_t,\ i = 1, \ldots, N$ of the N-states (individuals), the following special notes and comments are made.

1. In theoretical sense, the variables $Y\_i_t,\ i = 1, \ldots, N$ should be correlated variables. Therefore the whole set of $(N + K)$ variables, namely $Y\_i_t,\ i = 1, \ldots, N$, and $Zk,\ k = 1, \ldots, K$, can be viewed as single time-series data containing $(N + K)$ variables.
2. For a small number of $(N + K)$, say 2–5, then all models presented in Agung (2009a) and Section 1.8, should be applicable. Following the step-by-step methods presented in Agung (2009a), everyone should have no difficulty in doing the data analysis.
3. On the other hand for a large N, reducing the dimension is recommended using the following alternative methods.
   3.1 To defined groups of states (individuals), using either the judgmental method or cluster analysis, then the groups' statistics, such as the means and SDs, can be considered as the derived time series for further time-series data analysis.
   3.2 To reduce the dimension using factor analysis. Then the time series latent variables models would be applied. Refer to Chapter 10 in Agung (2011).
4. Similarly, for a large K of the environmental multivariate. However, note that some of its components might not be correlated, in a theoretical sense.
5. Furthermore, the environmental variables can be dummy variables of the time periods, thereby piecewise time-series models should be applied.

1.11 Special Piece-Wise Models

As an illustration, the panel data used are the data of daily stock prices of 15 individuals (agencies or industries) consisting of eight banks and seven mining companies, used by one of the author's advisories, namely Mulia (2010), for her thesis. The symbols B_ and M_ respectively, are used to identify the stock prices (*SP*) of the banks and mining companies. Furthermore, we see there are two break time points, which represent the time points of the *auto-rejection regulations* or *price limitations* of the stock prices. The objectives of the analysis are to study the differences of the statistics, such as growth rates, variances (volatilities) and means of *SP*, between 15 days before and after each break point so that four time periods to be considered in the analyses. For a better graphical presentation of the statistical results, the break points are set at $Day = 0$ and $Day = 40$, so that the growth curves of each individual stock price (*SP*) are not very far apart. See the following examples.

1.11.1 The Application of Growth Models

For a preliminary information of the data set, and further data analysis, Figure 1.47(a) and (b) present the scatter graphs of *(Mean_Bank, Day)* and *(Mean_Mining,Day)* with their *Nearest Neighbor Fit Curves*. The individual time series, namely B_ and M_, can easily be presented. Try it as an exercise.

Example 1.29 A four-piece classical growth model
Figure 1.48 presents the statistical results based on a four-piece classical growth model of the mean stock prices of eight banks, namely *Mean_Bank*, using four dummy variables, namely *D1, D2, D3*, and *D4*. Based on these statistical results, the following notes and conclusions are made.

1. The regression in Figure 1.48 represents four classical growth functions, as follows:

$$\log(Mean\_Bank) = 7.089238 - 0.030904^{*}Day,\ \ for\ Period = 1$$
$$\log(Mean\_Bank) = 7.463123 - 0.020809^{*}Day,\ \ for\ Period = 2$$
$$\log(Mean\_Bank) = 7.401909 + 0.002723^{*}Day,\ \ for\ Period = 3$$
$$\log(Mean\_Bank) = 7.724826 - 0.006886^{*}Day,\ \ for\ Period = 4$$

2. By using the Wald test, we discover that the growth rate 15 days before the $Day = 0$ is smaller than after $Day = 0$, based on the t-statistic of $t_0 = -2.355\,063$ with $df = 54$ and a p-value $= 0.0222/2 = 0.0111$,

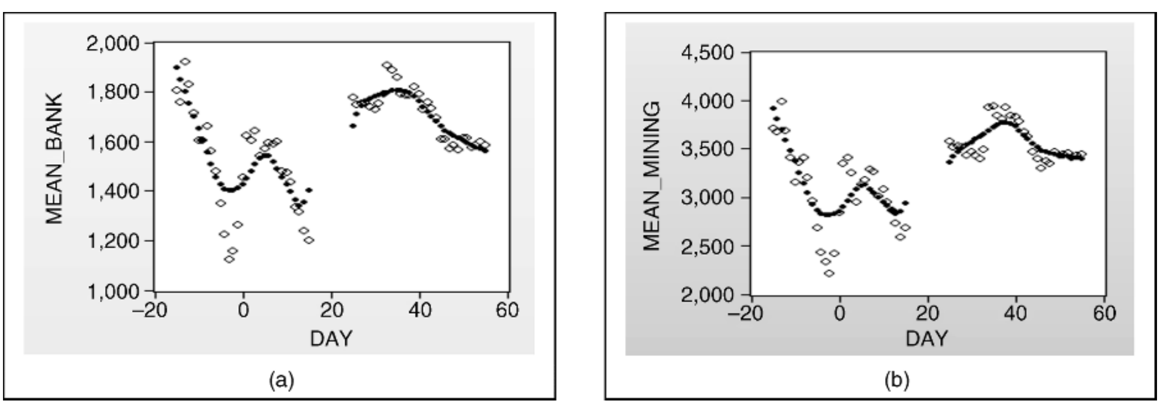


Figure 1.47 *Scatter graphs of* (Mean_Bank, Day) *and* (Mean_Mining, Day) *with their nearest neighbor fit curves*

Dependent Variable: LOG(MEAN_BANK)
Method: Least Squares
Date: 12/06/09 Time: 15:31
Sample: 10/08/2008 2/04/2009
Included observations: 62
LOG(MEAN_BANK)=(C(10)+C(11)*DAY)*D1+(C(20)+C(21)*DAY)*D2
+(C(30)+C(31)*DAY)*D3+(C(40)+C(41)*DAY)*D4

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	7.089238	0.025358	279.5657	0.0000
C(11)	-0.030904	0.002880	-10.72862	0.0000
C(20)	7.463123	0.028860	258.6009	0.0000
C(21)	-0.020809	0.003174	-6.555834	0.0000
C(30)	7.401909	0.094553	78.28350	0.0000
C(31)	0.002723	0.002880	0.945356	0.3487
C(40)	7.724826	0.152974	50.49751	0.0000
C(41)	-0.006886	0.003174	-2.169335	0.0345

R-squared	0.843170	Mean dependent var	7.376567
Adjusted R-squared	0.822840	S.D. dependent var	0.126189
S.E. of regression	0.053113	Akaike info criterion	-2.912860
Sum squared resid	0.152336	Schwarz criterion	-2.638391
Log likelihood	98.29865	Hannan-Quinn criter.	-2.805096
Durbin-Watson stat	1.161126		

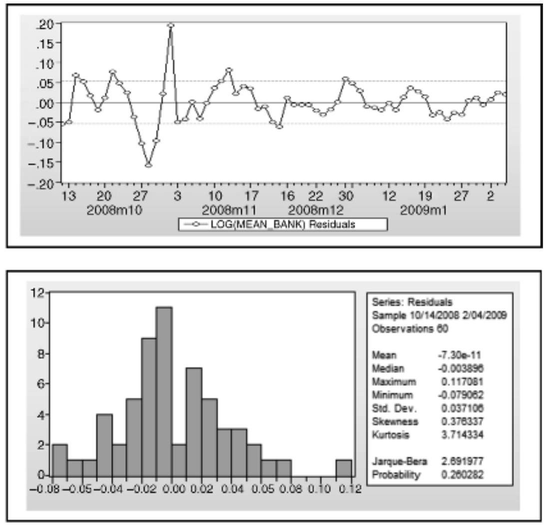


Figure 1.48 *Statistical results based on a four-piece growth model of* Mean_Bank

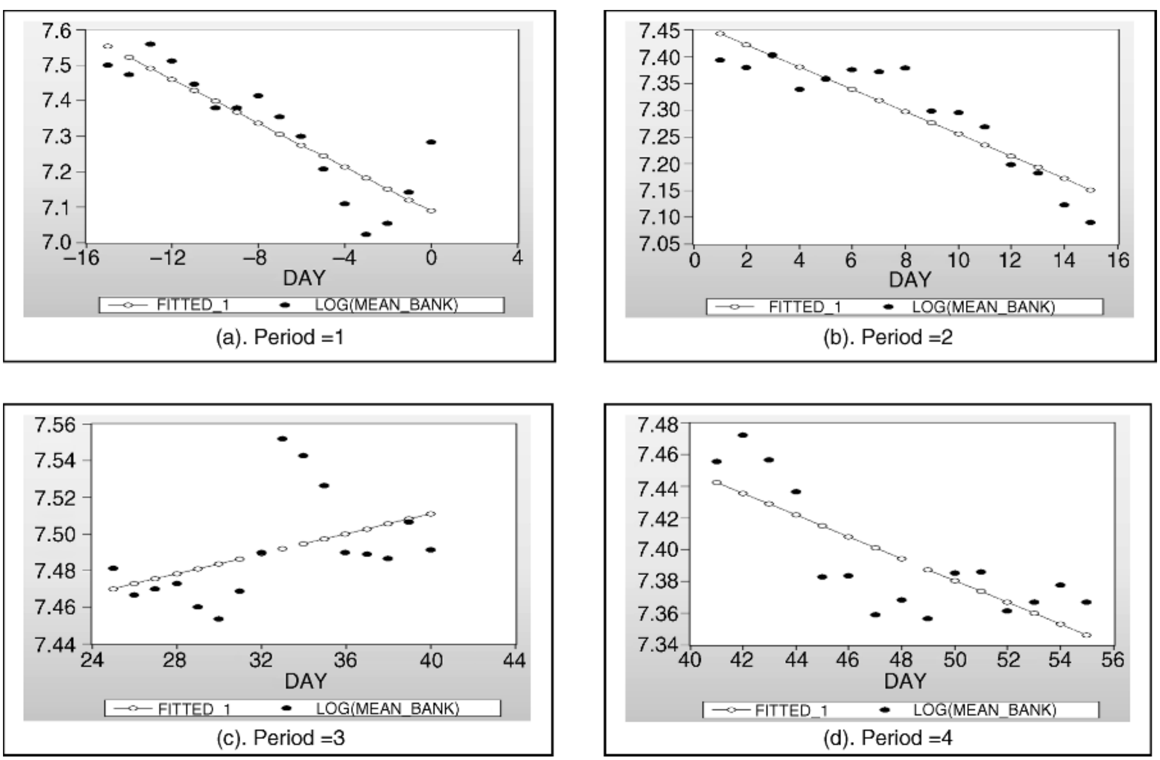


Figure 1.49 *Scatter graphs of the four-piece regression in Figure 1.48 and its fitted values, by time period*

and the growth rate 15 days before the $Day = 40$ is smaller than after $Day = 40$, based on the t-statistic of $t_0 = 2.241\,761$ with $df = 54$ and a p-value $= 0.0291/2 = 0.01\,455$. For a comparison, see the following example.

3. The residuals graph indicates that a nonlinear regression of *log(Mean_Bank)* on *Day* should be explored. On the other hand, Figure 1.49 clearly shows that a polynomial regression may be applied within each of the four time periods considered. See Example 1.30.
4. However, based on the $R^2 = 0.843\,170 > 80\%$, it could be concluded that the independent variables are good predictors for *log(Mean_Bank)*.
5. Note that exactly the same analysis can easily be conducted based on the *SP* of each individual, as well as the *Mean_Mining*.

Example 1.30 AR(2) four-piece growth model

By taking into account the autocorrelation of the classical growth model in Figure 1.49, Figure 1.50 presents the statistical results based on an AR(2) four-piece growth model, as a comparison.

Example 1.31 The nearest neighbor fit of *log*(*Mean_Bank*)

Figure 1.51 presents the scatter graph of *(log(Mean_Bank),Day)* with its *Nearest Neighbor Fit* by the time periods. The four graphs clearly show that nonlinear models should be applied within each time period. See the following example.

Example 1.32 A four-piece polynomial growth model

By using trial-and-error methods, statistical results are obtained based on a four-piece polynomial growth model presented in Figure 1.52. Based on these results, note the following:

1. Compared to the classical growth model in Figure 1.48 and the AR(2) growth model in Figure 1.49, this polynomial growth model has the largest value of $R^2 = 0.957\,224$. So, in a statistical sense, this model

Dependent Variable: LOG(MEAN_BANK)
Method: Least Squares
Date: 12/06/09 Time: 13:58
Sample (adjusted): 10/14/2008 2/04/2009
Included observations: 60 after adjustments
Convergence achieved after 12 iterations
LOG(MEAN_BANK)=(C(10)+C(11)*DAY)*D1+(C(20)+C(21)*DAY)*D2
+(C(30)+C(31)*DAY)*D3+(C(40)+C(41)*DAY)*D4+[AR(1)=C(1),AR(2)=C(2)]

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	7.272001	0.079335	91.66159	0.0000
C(11)	-0.010064	0.014238	-0.706835	0.4830
C(20)	7.350336	0.086670	84.80814	0.0000
C(21)	-0.014022	0.008803	-1.592873	0.1175
C(30)	7.626986	0.233346	32.68531	0.0000
C(31)	-0.003712	0.007029	-0.528084	0.5998
C(40)	7.682881	0.374227	20.53002	0.0000
C(41)	-0.005994	0.007862	-0.762305	0.4495
C(1)	1.241128	0.138661	8.950818	0.0000
C(2)	-0.448722	0.143253	-3.132367	0.0029

R-squared	0.914152	Mean dependent var	7.372923
Adjusted R-squared	0.898700	S.D. dependent var	0.126643
S.E. of regression	0.040308	Akaike info criterion	-3.433543
Sum squared resid	0.081235	Schwarz criterion	-3.084486
Log likelihood	113.0063	Hannan-Quinn criter.	-3.297008
Durbin-Watson stat	1.971303		

Inverted AR Roots	.62-.25i	.62+.25i

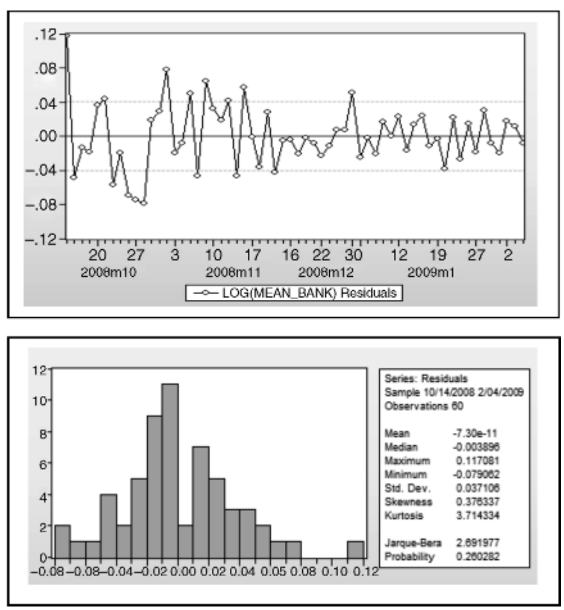


***Figure 1.50** Statistical results based on an AR(2) four-piece growth model of* Mean_Bank

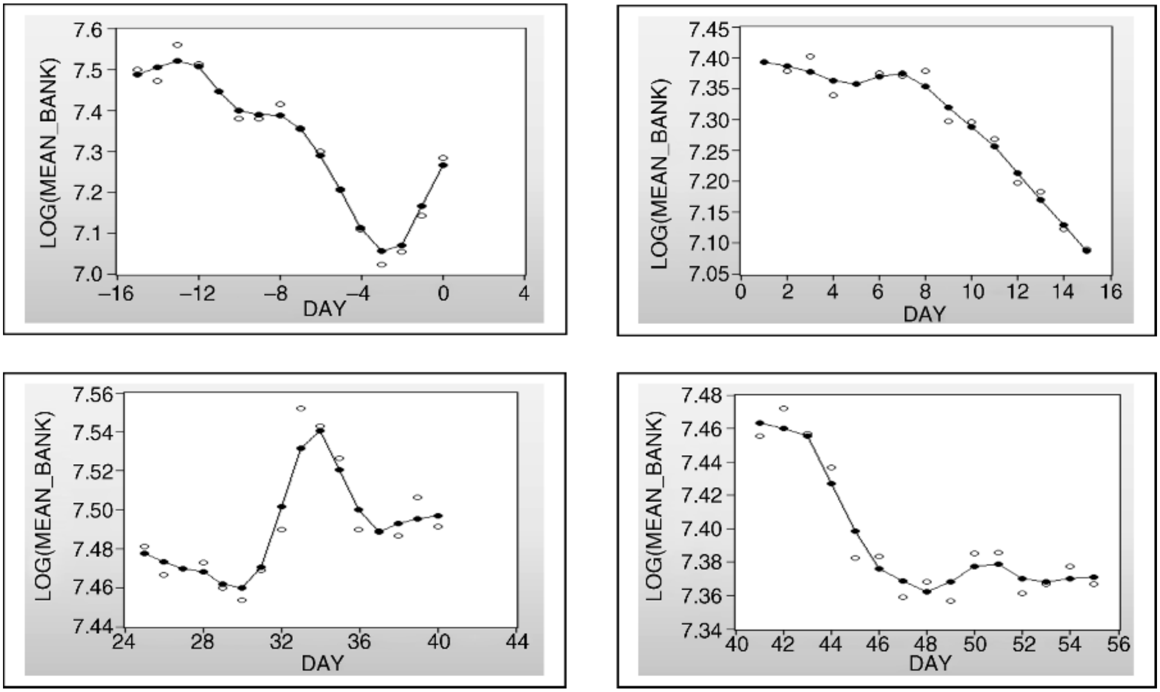


Figure 1.51 *Scatter graphs of (*log(Mean_Bank),Day*) with its nearest neighbor fit by time period*

Dependent Variable: LOG(MEAN_BANK)
Method: Least Squares
Date: 12/06/09 Time: 21:11
Sample: 10/08/2008 2/04/2009
Included observations: 62
LOG(MEAN_BANK)=(C(10)+C(11)*DAY+C(12)*DAY^2+C(13)*DAY^3+C(14)*DAY^4)*D1+(C(20)+C(21)*DAY+C(22)*DAY^2)*D2+(C(30)+C(31)*DAY+C(32)*DAY^2+C(33)*DAY^3+C(34)*DAY^4)*D3+(C(40)+C(42)*DAY^2+C(43)*DAY^3)*D4

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	7.273184	0.026987	269.5056	0.0000
C(11)	0.193225	0.026699	7.237188	0.0000
C(12)	0.055096	0.007600	7.249808	0.0000
C(13)	0.004689	0.000773	6.063391	0.0000
C(14)	0.000130	2.56E-05	5.081061	0.0000
C(20)	7.362563	0.026770	275.0267	0.0000
C(21)	0.014683	0.007699	1.907029	0.0628
C(22)	-0.002218	0.000468	-4.740573	0.0000
C(30)	46.35971	27.27787	1.699535	0.0960
C(31)	-4.792701	3.435079	-1.395223	0.1696
C(32)	0.218807	0.160933	1.359620	0.1806
C(33)	-0.004387	0.003325	-1.319543	0.1935
C(34)	3.26E-05	2.56E-05	1.276490	0.2082
C(40)	8.448153	0.361846	23.34736	0.0000
C(42)	-0.001244	0.000473	-2.628171	0.0116
C(43)	1.62E-05	6.53E-06	2.482523	0.0168

R-squared	0.957224	Mean dependent var	7.376567
Adjusted R-squared	0.943276	S.D. dependent var	0.126189
S.E. of regression	0.030054	Akaike info criterion	-3.953987
Sum squared resid	0.041550	Schwarz criterion	-3.405049
Log likelihood	138.5736	Hannan-Quinn criter.	-3.738460
Durbin-Watson stat	1.375620		

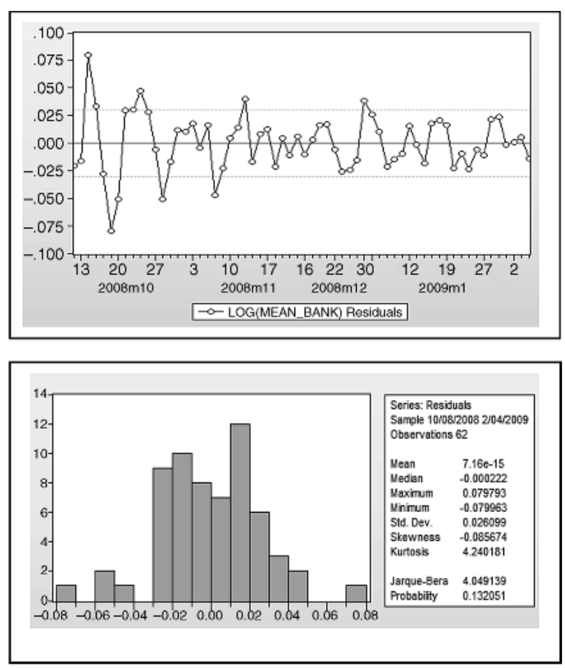


Figure 1.52 *Statistical results based on a four-piece polynomial growth model of* Mean_Bank

should be considered the best of the three growth models, even though it is a standard multiple regression. In other words, independent variables are the best predictors for *log(Mean_Bank)*.

2. The multiple regression in Figure 1.52 in fact represents the following four polynomial regressions within the four time periods, namely *Period* = 1, 2, 3 and 4, respectively.

$$
\begin{aligned}
\log(Mean\_Bank) &= 7.273 + 0.193^{*}t + 0.055^{*}t^{2} + 0.004689^{*}t^{3} + 0.000130^{*}t^{4} \\
\log(Mean\_Bank) &= 7.363 + 0.015^{*}t - 0.002218^{*}t^{2} \\
\log(Mean\_Bank) &= 46.360 - 4.793^{*}t + 0.219^{*}t^{2} - 0.004387^{*}t^{3} + 3.3e - 05^{*}t^{4} \\
\log(Mean\_Bank) &= 8.448 - 0.001^{*}t^{2} + 1.6e - 05^{*}t^{3}
\end{aligned}
$$

1.11.2 Equality Tests by Classifications

The option "*Equality Tests by Classifications*" provides the statistics for testing a hypothesis on the difference of the mean, median or variance of single variables between groups of individuals/objects generated by one or more classification or treatment factors.

Example 1.33 Test for equality of variances

As an illustration, Figure 1.53(a) and (b) presents the statistical results for testing the equality of variances of the variable *B_1* (*SP* for Bank-1), 15 days before and after the first and second break point, respectively, indicated by the *Period* = 1, 2, and *Period* = 3, 4. Based on these results, the following notes and conclusions are presented.

1. Based on the *F*-test, it can be concluded that the variances of *B_1* have significant differences between 15 days before and after each break point, namely at *Day* = 0 and *Day* = 40. Therefore, the *volatilities* of the *B_1s* stock prices before and after each break point have significant differences.
2. However, the Siegel–Tukey test should be questionable, since it has such a very small value compared to the others, specifically in Figure 1.53(b).

Test for Equality of Variances of B_1
Categorized by values of PERIOD
Date: 12/07/09 Time: 08:32
Sample: 10/08/2008 2/04/2009 IF PERIOD<3
Included observations: 31

Method	df	Value	Probability
F-test	(14, 15)	9.379213	0.0001
Siegel-Tukey		2.040246	0.0413
Bartlett	1	14.47209	0.0001
Levene	(1, 29)	9.325066	0.0048
Brown-Forsythe	(1, 29)	7.331016	0.0112

Category Statistics

PERIOD	Count	Std. Dev.	Mean Abs. Mean Diff.	Mean Abs. Median Diff.	Mean Tukey-Siegel Rank
1	16	315.2380	242.1875	234.3750	12.75000
2	15	102.9332	85.55556	85.00000	19.46667
All	31	246.1723	166.3978	162.0968	16.00000

Bartlett weighted standard deviation: 237.7305

(a)

Test for Equality of Variances of B_1
Categorized by values of PERIOD
Date: 12/07/09 Time: 08:36
Sample: 10/08/2008 2/04/2009 IF PERIOD>2
Included observations: 31

Method	df	Value	Probability
F-test	(14, 15)	5.045106	0.0036
Siegel-Tukey		0.186106	0.8524
Bartlett	1	8.179716	0.0042
Levene	(1, 29)	3.843049	0.0596
Brown-Forsythe	(1, 29)	3.699104	0.0643

Category Statistics

PERIOD	Count	Std. Dev.	Mean Abs. Mean Diff.	Mean Abs. Median Diff.	Mean Tukey-Siegel Rank
3	16	176.8945	126.5625	121.8750	16.32292
4	15	78.75520	62.22222	51.66667	15.65556
All	31	233.9872	95.43011	87.90323	16.00000

Bartlett weighted standard deviation: 138.4904

(b)

Figure 1.53 *Test for equality of the variances of* B_1*, 15 days before and after two break points*

Test for Equality of Means of B_1
Categorized by values of PERIOD
Date: 12/07/09 Time: 08:39
Sample: 10/08/2008 2/04/2009 IF PERIOD<3
Included observations: 31

Method	df	Value	Probability
t-test	29	-1.780006	0.0856
Satterthwaite-Welch t-test*	18.35150	-1.828579	0.0838
Anova F-test	(1, 29)	3.168421	0.0856
Welch F-test*	(1, 18.3515)	3.343701	0.0838

*Test allows for unequal cell variances

Analysis of Variance

Source of Variation	df	Sum of Sq.	Mean Sq.
Between	1	179065.9	179065.9
Within	29	1638958.	56515.80
Total	30	1818024.	60600.81

Category Statistics

PERIOD	Count	Mean	Std. Dev.	Std. Err. of Mean
1	16	2631.250	315.2380	78.80950
2	15	2783.333	102.9332	26.57723
All	31	2704.839	246.1723	44.21385

(a)

Test for Equality of Means of B_1
Categorized by values of PERIOD
Date: 12/07/09 Time: 08:38
Sample: 10/08/2008 2/04/2009 IF PERIOD>2
Included observations: 31

Method	df	Value	Probability
t-test	29	7.525814	0.0000
Satterthwaite-Welch t-test*	21.00718	7.695656	0.0000
Anova F-test	(1, 29)	56.63788	0.0000
Welch F-test*	(1, 21.0072)	59.22312	0.0000

*Test allows for unequal cell variances

Analysis of Variance

Source of Variation	df	Sum of Sq.	Mean Sq.
Between	1	1086292.	1086292.
Within	29	556208.3	19179.60
Total	30	1642500.	54750.00

Category Statistics

PERIOD	Count	Mean	Std. Dev.	Std. Err. of Mean
3	16	3181.250	176.8945	44.22363
4	15	2806.667	78.75520	20.33450
All	31	3000.000	233.9872	42.02534

(b)

Figure 1.54 *Test for equality of the means of* B_1*, 15 days before and after two break points*

Example 1.34 Test for equality of means

In addition to the testing of variances presented in Figure 1.53(a) and (b), Figure 1.54(a) and (b) presents the statistical results for testing the equality of means of *B_1*, 15 days before and after the two break points, respectively. Corresponding to heterogeneity of the variances in a statistical sense, then the Welch *F*-test should be used to making the conclusion of the testing hypothesis on the means differences. In this case, however, the other tests also give exactly the same conclusion, at the significance level of either 5 or 10%.

On the other hand, the cell-mean model is not an appropriate time-series model generally – refer to Section 4.3.1 in Agung (2009a). So I cannot recommend conducting a test on the mean differences of a time series between long time periods: this is similarly so for testing equality of medians. We recommend the reader study and test their growth differences.