实验手册

实验一　异方差的检验与修正

一、实验目的

了解异方差（heteroscedasticity）、Goldfeld - Quandt 检验、Spearman rank correlation 检验、Park 检验、Glejser 检验、Breusch - Pagan 检验、White 检验、加权最小二乘法（weighted least squares，简记 WLS）、模型对数变换法等基本概念及异方差产生的原因和后果。

掌握异方差的检验与修正方法以及如何运用 EViews 软件在实证研究中实现相关检验与修正。

二、基本概念

异方差（heteroscedasticy）就是对同方差假设（assumption of homoscedasticity）的违反。经典回归中同方差是指随着样本观察点 X 的变化，线性模型中随机误差项的方差并不改变，保持为常数。

异方差的检验有图示法及解析法，检验异方差的解析方法的共同思想是，由于不同的观察值随机误差项具有不同的方差，因此检验异方差的主要问题是判断随机误差项的方差与解释变量之间的相关性。

异方差的修正方法有加权最小二乘法和模型对数变换法等，其基本思路是变异方差为同方差，或者尽量缓解方差变异的程度。

三、实验内容及要求

（一）实验内容

根据北京市 1978—1998 年人均储蓄与人均收入的数据资料，若假定 X 为人均收入（元），Y 为人均储蓄（元），通过建立一元线性回归模型分析人均储蓄受人均收入的线性影响，并讨论异方差的检验与修正过程。

（二）实验要求

（1）深刻理解上述基本概念；

（2）思考：异方差的各种检验方法所适用的情况及如何运用加权最小二乘法（WLS）修正异方差？

（3）熟练掌握相关 EViews 操作。

四、实验指导

（一）用 OLS 估计法估计参数

1. 导入数据

打开 EViews 软件，选择“File”菜单中的“New - Workfile”选项，出现“Workfile Range”对话框，在“Workfile frequency”框中选择“Annual”，在“Start date”和“End date”框中分别输入“1978”和“1998”，如图 1 - 1 所示。

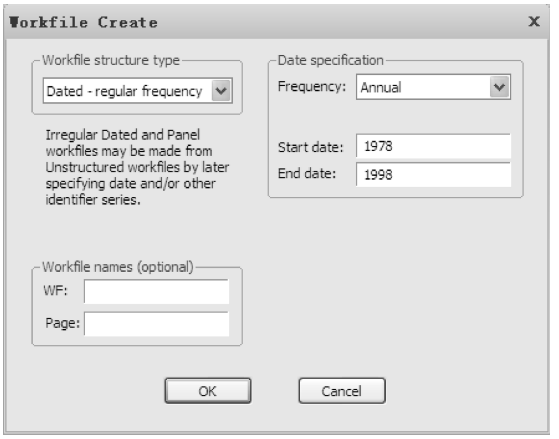


图 1 - 1　建立新文件

然后单击“OK”，弹出如下窗口（见图 1 - 2）。

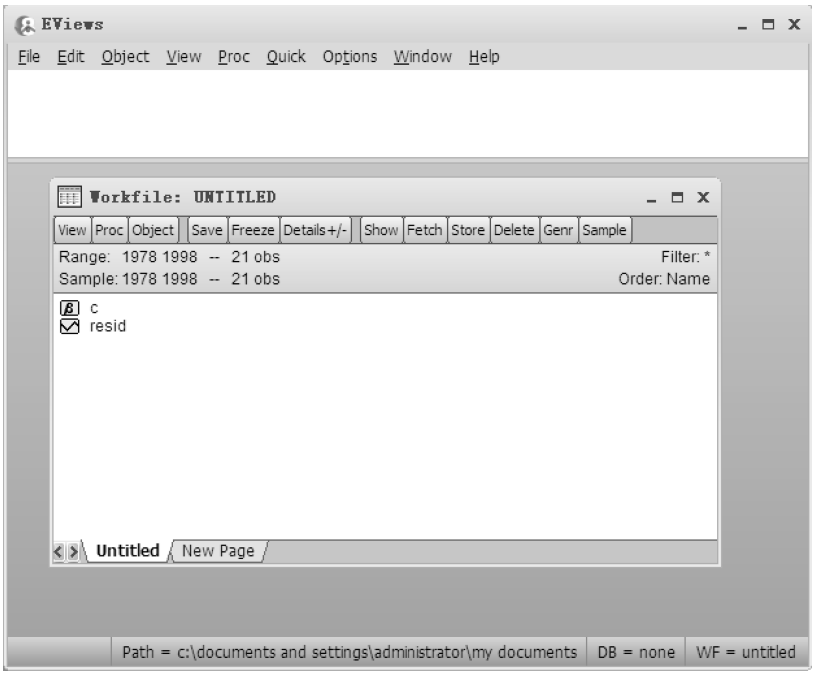


图 1 - 2　建立新文件

选择“File”菜单中的“Import - Read Text - Lotus - Excel”选项，找到要导入的名为 EX3. 2. xls 的 Excel 文档，单击“打开”出现“Excel Spreadsheet Import”对话框并在其中输入“x”和“y”，如图 1 - 3 所示。

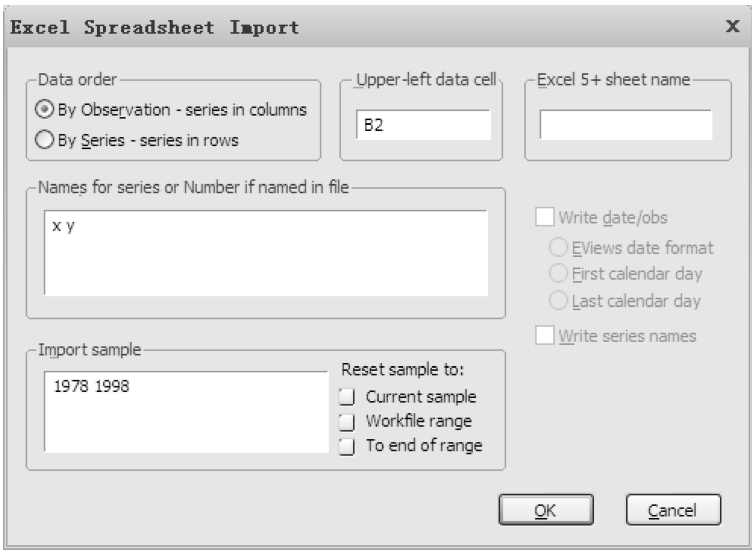


图 1－3　导入数据

再单击“OK”完成数据导入。

2. 回归数据估计方程

设模型为 $Y=\beta_1+\beta_2X+\mu$，在 EViews 命令窗口中输入“LS Y C X”并回车，得到如下结果，见图 1－4。

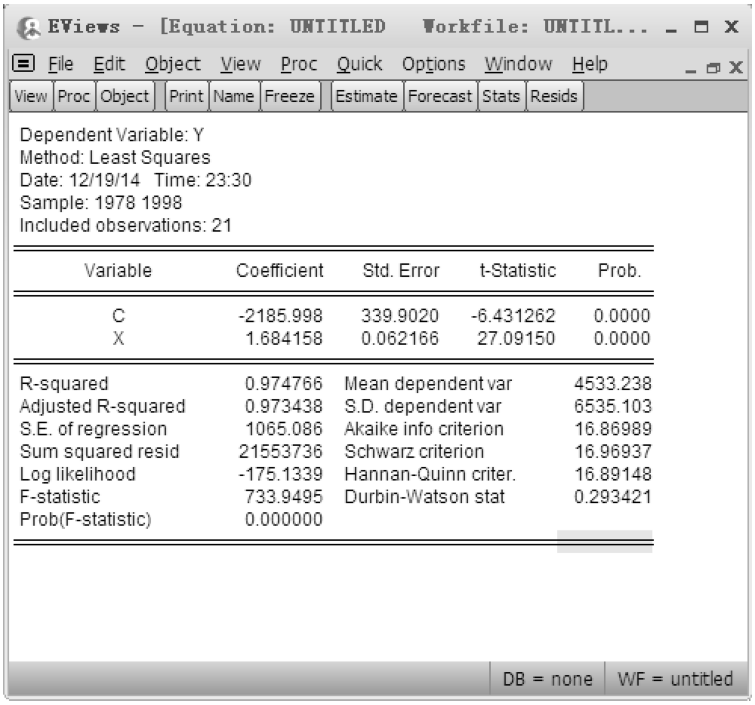


图 1－4　EViews 回归结果

（二）异方差检验

1. 图示法

首先通过“Equation”对话框中“Procs”菜单的“Make Residual Series”命令生成残差序列 E，点击“OK”，如图 1－5 所示。

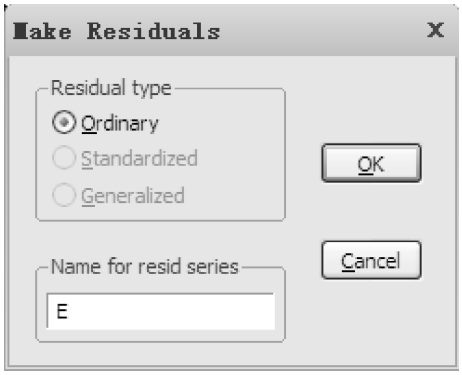


图 1－5　生成残差序列

然后在“Quick”菜单中选“Graph”选项，再在弹出的对话框中输入“X E^2”，并单击“OK”即可得到图 1－6 所示结果。

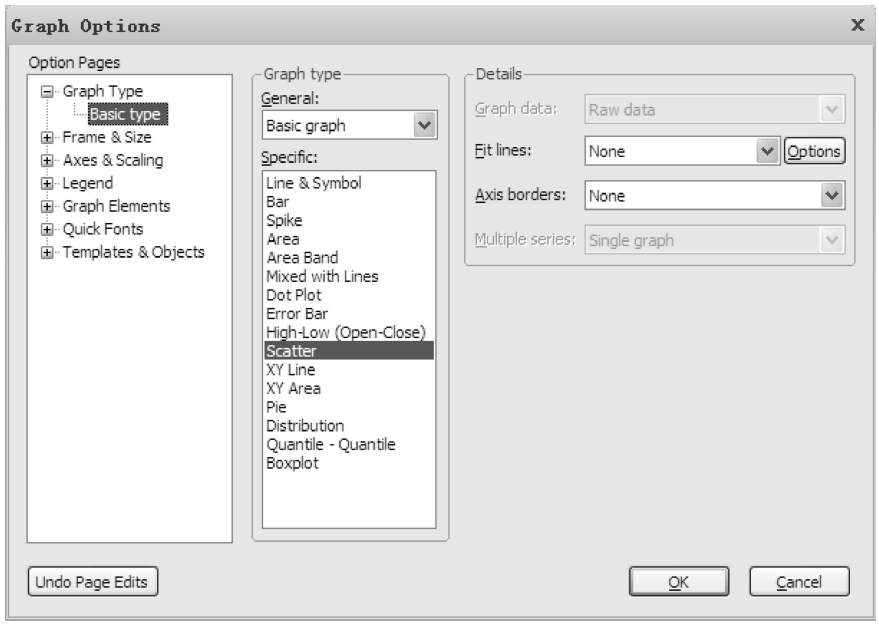


图 1－6　残差序列图示法

再在“Graph Type”框中选择散点图（Scatter Diagram），并单击“OK”即可得到图 1－7 所示的结果。

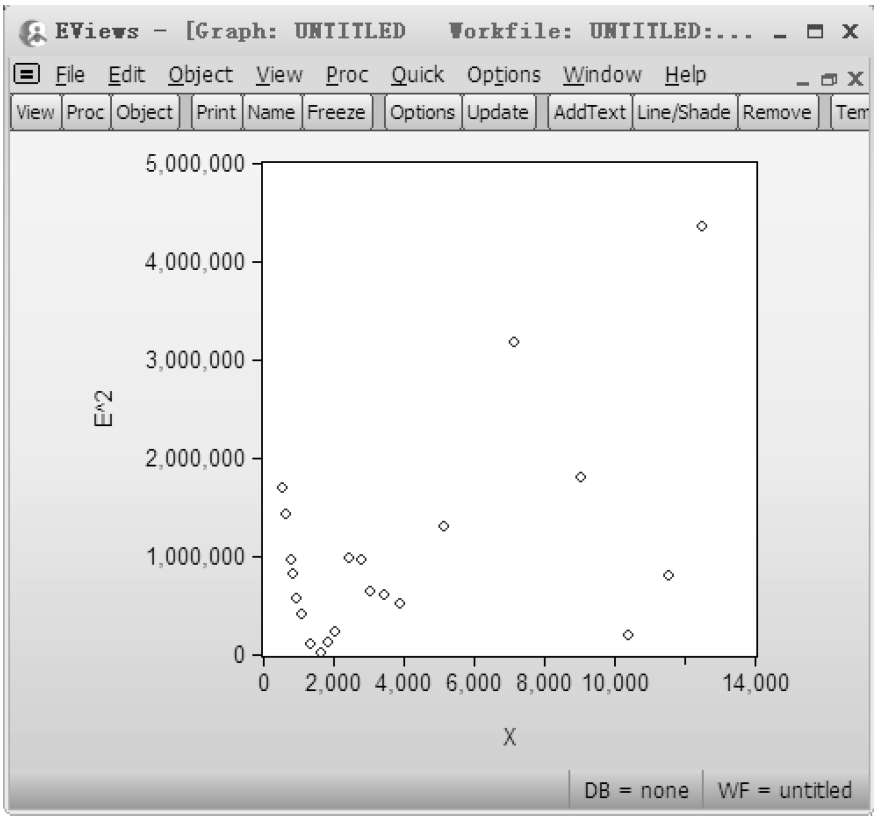


图 1－7　残差序列的散点图

2. Goldfeld－Quandt 检验

首先将时间定义为“1978—1985”，方法如下：在“Workfile”对话框中选择“Procs”菜单的“sample”选项，弹出如下窗口并把期间改为“1978 1985”，如图 1－8 所示。

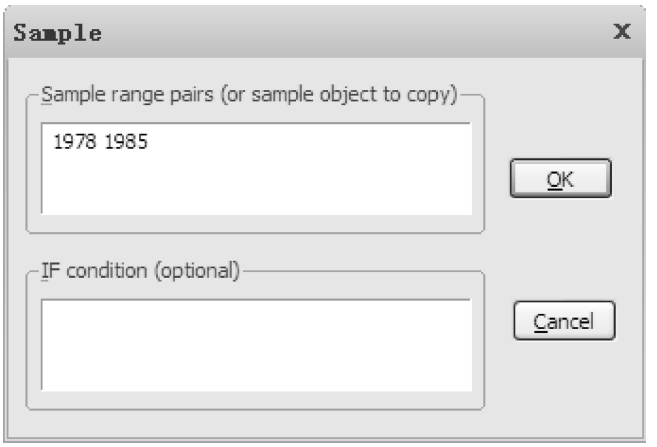


图 1－8　样本范围的设定

再在 EViews 命令区输入命令“LS Y C X”回车得到图 1－9 所示的结果。

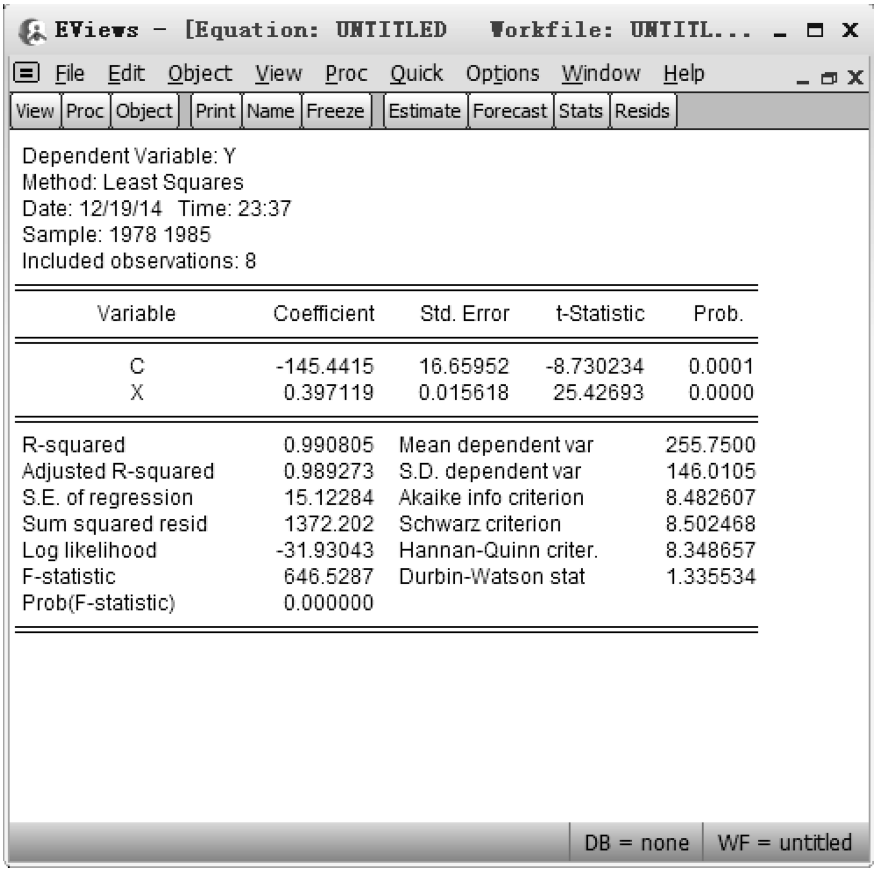


图 1－9　1978—1995 年数据的回归结果

即用 OLS 方法可求得下列结果：

$$Y = -145.4415 + 0.3971X \qquad (1978—1985)$$

$$(-8.7302) \quad (25.4269)$$

$$R^2 = 0.9908 \quad \sum e_1^2 = 1372.202$$

其次用相同的方法将时间定义为“1991—1998”，回归得到如下结果，如图 1－10 所示。

即：

$$Y = -4602.365 + 1.9525X \qquad (1991—1998)$$

$$(-5.0660) \quad (18.4094)$$

$$R^2 = 0.9826 \quad \sum e_2^2 = 5811189$$

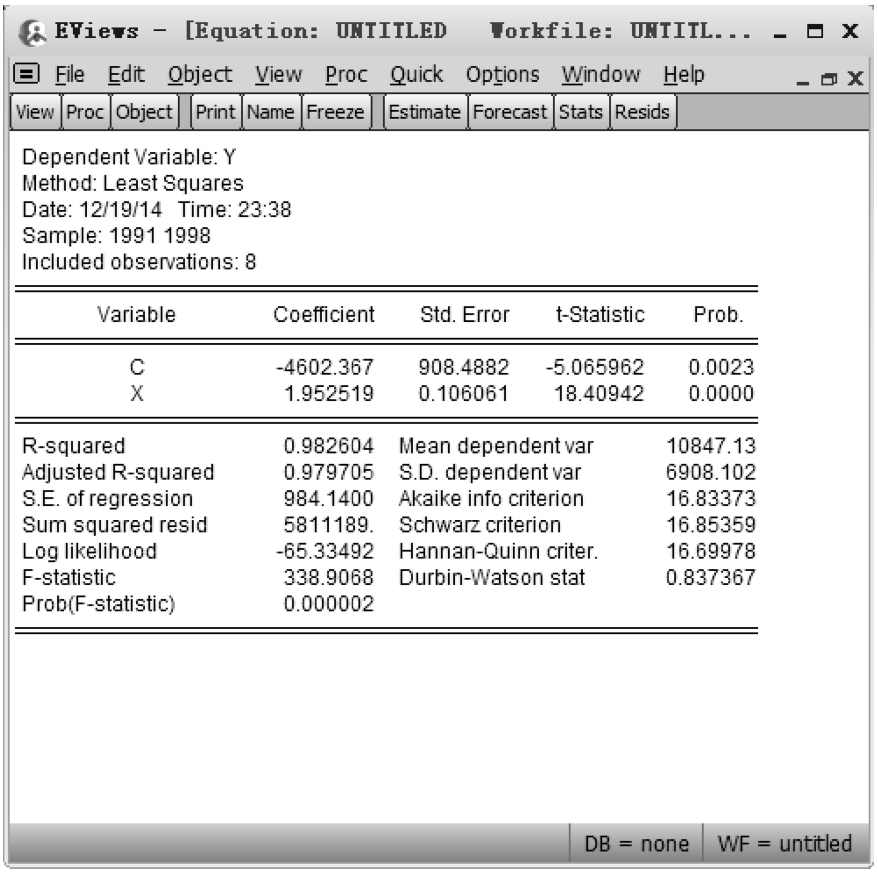


图 1－10　1991—1998 年数据的回归结果

求 F 统计量：$F=\dfrac{\sum e_2^2}{\sum e_1^2}=4334.9370$，查 F 分布表，给定显著性水平 $\alpha=0.05$，得临界值 $F_{0.05}(6,6)=4.28$，比较 $F=4334.9370>F_{0.05}(6,6)=4.28$ 则拒绝 $H_0:\sigma_1^2=\sigma_2^2$，表明随机误差项显著存在异方差。

3. ARCH 检验

在窗口的“View”菜单中选择—“Residual Tests”—“ARCH LM Test”选项，然后在弹出的对话框中选择滞后阶数为 3 阶，即可得到图 1－11。

从中可知 $Obs*R^2=10.186$，P 值为 0.017，表明模型随机误差项存在异方差性。

（三）异方差的修正

1. WLS 估计法

选择“Equation”对话框中“Estimate”菜单的“Option”选项，填入权重 X^(－0.5) 即可得到图 1－12。

EViews - [Equation: UNTITLED Workfile: UNTITLED...

File Edit Object View Proc Quick Options Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Heteroskedasticity Test: ARCH

F-statistic	6.083602	Prob. F(3,14)	0.0072
Obs*R-squared	10.18624	Prob. Chi-Square(3)	0.0170

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/19/14 Time: 23:42
Sample (adjusted): 1981 1998
Included observations: 18 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	242407.2	372654.8	0.650487	0.5259
RESID^2(-1)	1.229916	0.329833	3.728905	0.0022
RESID^2(-2)	-1.408992	0.378753	-3.720081	0.0023
RESID^2(-3)	1.018754	0.327542	3.110302	0.0077

R-squared	0.565902	Mean dependent var	971588.3
Adjusted R-squared	0.472881	S.D. dependent var	1130306.
S.E. of regression	820635.6	Akaike info criterion	30.26668
Sum squared resid	9.43E+12	Schwarz criterion	30.46454
Log likelihood	-268.4001	Hannan-Quinn criter.	30.29396
F-statistic	6.083602	Durbin-Watson stat	2.125335
Prob(F-statistic)	0.007182		

DB = none WF = untitled

图 1－11 ARCH 检验结果

EViews - [Equation: UNTITLED Workfile: UNTITLED::Untit...

File Edit Object View Proc Quick Options Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y
Method: Least Squares
Date: 02/28/15 Time: 00:54
Sample: 1978 1998
Included observations: 21
Weighting series: X^(-0.5)
Weight type: Inverse standard deviation (EViews default scaling)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1378.962	220.3181	-6.258959	0.0000
X	1.481876	0.083896	17.66326	0.0000

Weighted Statistics

R-squared	0.942597	Mean dependent var	2384.938
Adjusted R-squared	0.939575	S.D. dependent var	2695.511
S.E. of regression	832.7163	Akaike info criterion	16.37766
Sum squared resid	13174913	Schwarz criterion	16.47713
Log likelihood	-169.9654	Hannan-Quinn criter.	16.39925
F-statistic	311.9908	Durbin-Watson stat	0.165765
Prob(F-statistic)	0.000000	Weighted mean dep.	1182.530

Unweighted Statistics

R-squared	0.960704	Mean dependent var	4533.238
Adjusted R-squared	0.958636	S.D. dependent var	6535.103
S.E. of regression	1329.123	Sum squared resid	33564815
Durbin-Watson stat	0.237190		

Path = c:\documents and settings\administrator\my documents DB = none WF = untitled

图 1－12 WLS 估计结果

2. 对数变换法

在“Equation”窗口中“Estimate”菜单的对话框直接输入“LOG(Y)C LOG(X)”，再单击“OK”后，就可以得到线性模型对数变换后的结果如图 1－13 所示。

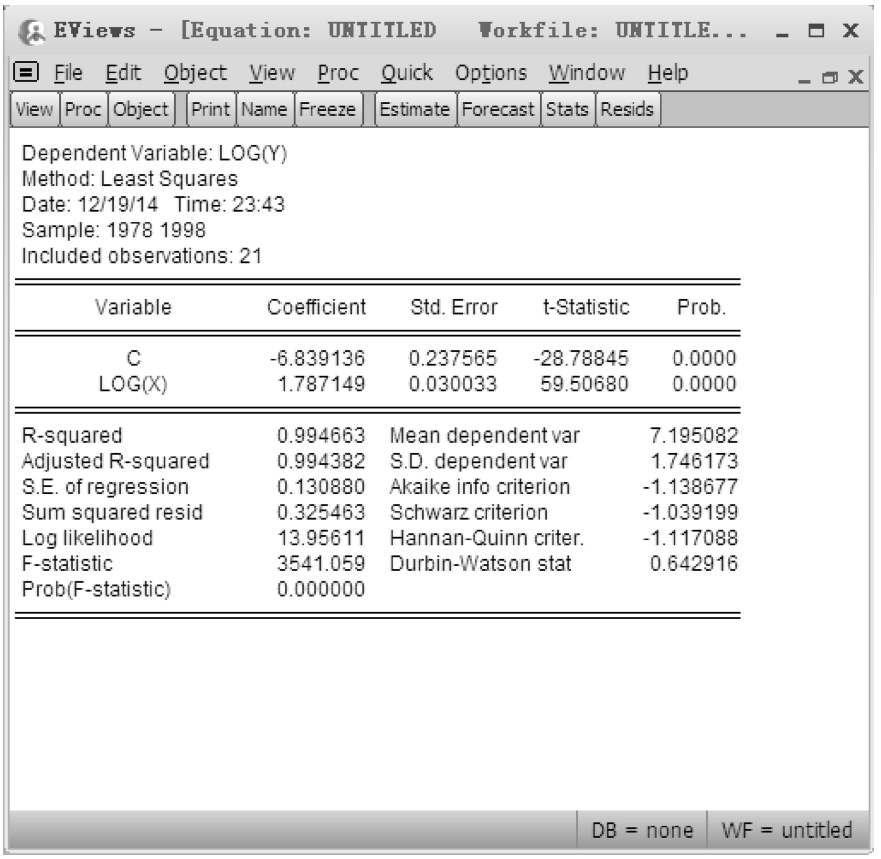


图 1－13　对数变换估计结果

比较上述两种修正方法，对数变换后的结果在拟合优度和系数显著性都要好于加权最小二乘法得到的结果，这说明人均收入与人均储蓄的关系更接近于对数关系。

实验二　虚拟变量在金融数据处理中的作用

一、实验目的

了解虚拟变量、方差分析模型、协方差分析模型、虚拟变量陷阱、季节调整、分段线性回归、级差截距、级差斜率系数、周内效应等基本概念及虚拟变量的引入原则、虚拟变量模型中参数的意思。

掌握虚拟变量模型在回归分析中的应用，及如何在 EViews 中实现相应的操作。

二、基本概念

由于其不能直接度量，为研究方便，可构造一个变量，令其取值为 1 或为 0，取值为 0 时表示某一性质出现（不出现），取值为 1 时表示某性质不出现（出现），该变量即为虚拟变量（dummy variables），也称指标变量（indicator variables）、二值变量（binary variables）、定性变量（qualitative variables）和二分变量（dichotomous variables），通常我们记为 D。一般来说，在虚拟变量的设置中，基础类型、否定类型取值为“0”，称为基底（base）类、基准（benchmark）类或参考（reference）类；而比较类型、肯定类型取值“1”。

许多按月度或季度数据表示的金融时间序列，常呈现出季节变化的规律性，如公司销售额、通货膨胀率、节假日储蓄额等。在研究中，有时需要消除季节性因素的影响，即需要进行季节调整（seasonal adjustment），利用虚拟变量进行季节调整是较为简单的一种。另外，在金融理论中，常常会出现一种情况：当某影响因素越过某一临界值，或时间过了某一临界点之后，因变量对影响因素的变化率将发生变化，在图形中就表现为斜率不同的两段连续折线，利用虚拟变量模型进行分段线性回归可有效地解决在分界点处两边因变量取值不相等，与理论图形不一致这个问题。

三、实验内容及要求

（一）实验内容

利用上海股票市场上证指数 1997 年 1 月 1 日—2004 年 12 月 31 日的日收盘价数据，共 1 926 个观测值，通过建立虚拟变量模型，来检验实行涨跌停板制度后的上海股票市场是否存在周内效应。

（二）实验要求

（1）深刻理解上述基本概念；

（2）思考：虚拟变量模型的各种不同运用情形，及虚拟变量法与邹式检验的异同；

（3）熟练掌握相关 EViews 操作；

（4）根据表 4－10 我国 1994 年第一季度至 2004 年第四季度的零售物价指数（以上年同期为 100）数据，请利用虚拟变量模型进行季节调整（如果需要）。

四、实验指导

（一）简单理论回顾

根据 Fama 的有效市场理论，在有效市场中，由于股票价格能够及时地反映所有的信息，因此股价将会呈现出随机波动的特征。并且在有效市场中，由于投资者能够随时获取所需要的信息，因此将不存在套利的机会，股票的价格将反映价值。按照有效市场理论，一周内每天的收益率将是随机波动、没有规律的。

自从 Fama1965 年提出有效市场概念以来，各国学者分别利用各国的证券市场数据对其进行了实证检验，结果发现许多与有效市场理论相背离的现象，周内效应就是其中之一。

在我国，许多学者也利用上海股票市场、深圳股票市场的数据对周内效应进行了检验，检验结果大多表明存在周内效应。下面我们将利用虚拟变量模型对这一现象进行实证检验。

（二）实证检验

1. 数据说明

我们利用的是上海股票市场上证指数 1997 年 1 月 1 日—2004 年 12 月 31 日的日收盘价数据，共 1 926 个观测值。之所以采用 1997 年来的日数据，是因为 1996 年 12 月 16 日股票市场实施了涨跌停板制度，而上证指数也具有广泛的代表性。关于收益率的计算，我们采用的是连续收益率法，计算公式如下：

$$R_t = \ln P_t - \ln P_{t-1}$$

其中 P_t 为 t 时期的收盘价，P_{t-1} 为 $t-1$ 时期的收盘价。

2. 数据导入

打开 EViews 软件，选择“File”菜单中的“New－Workfile”选项，出现“Workfile Range”对话框，在“Workfile frequency”框中选择“Undated or irregular”，在“Start

observation”和“End observation”框中分别输入“1”和“1926”，单击“OK”，出现一个新的工作文件。

然后选择“File”菜单中的“Import－Read Text－Lotus－Excel”选项，找到要导入的名为 EX4. 3. xls 的 Excel 文档，单击“打开”出现“Excel Spreadsheet Import”对话框并在其中输入“D1”“D2”“D4”“D5”和“R”，如图 2－1 所示。

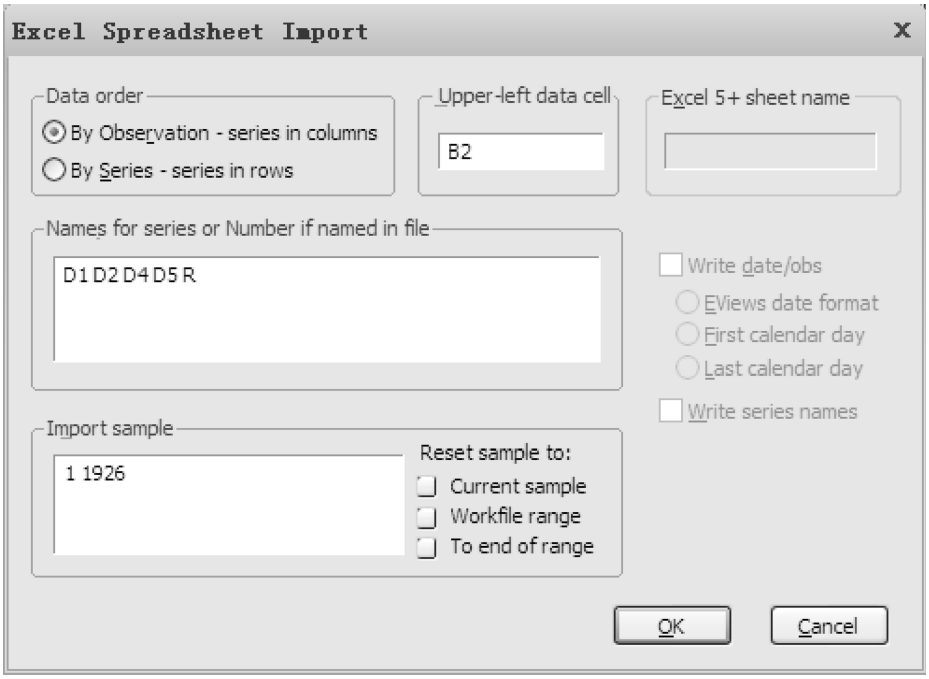


图 2－1　数据导入

再单击“OK”完成数据导入。

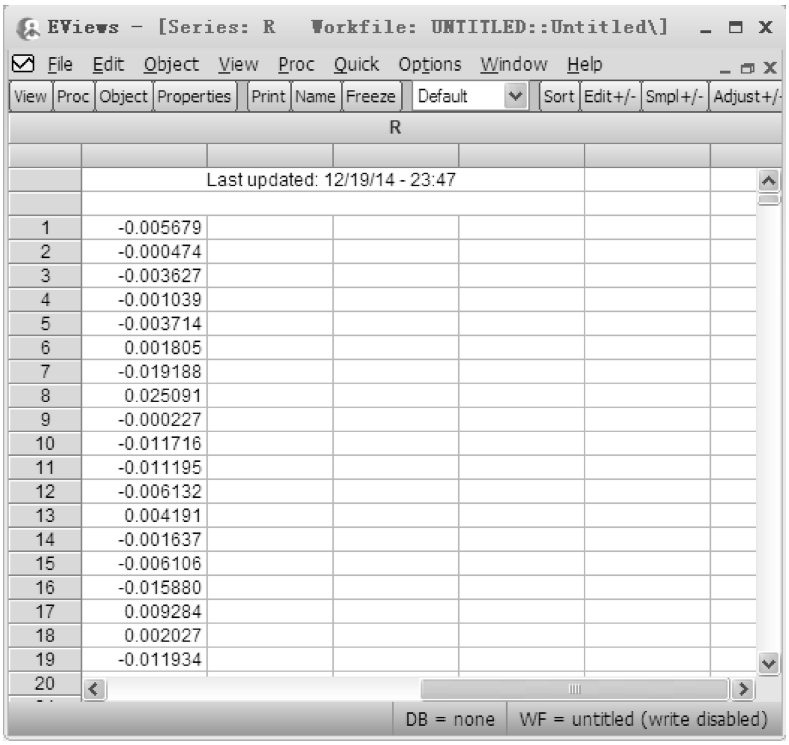
3. 数据统计特征描述

下面分别描述上证指数 1997 年 1 月 1 日—2004 年 12 月 31 日收益率(R)的统计特征以及日收益率按时间排列的序列图，具体操作如下：

首先双击“R”序列，出现如图 2－2 所示的窗口。

其次选择“Series：R”窗口的“View”菜单中的“Descriptive－Statistic－histogram and stats”选项，可以得到日收益率(R)的统计特征图，如图 2－3 所示。

图 2－3 中数据可列表如表 2－1 所示。



1	-0.005679
2	-0.000474
3	-0.003627
4	-0.001039
5	-0.003714
6	0.001805
7	-0.019188
8	0.025091
9	-0.000227
10	-0.011716
11	-0.011195
12	-0.006132
13	0.004191
14	-0.001637
15	-0.006106
16	-0.015880
17	0.009284
18	0.002027
19	-0.011934
20	

图 2－2　序列 *R* 的统计数据

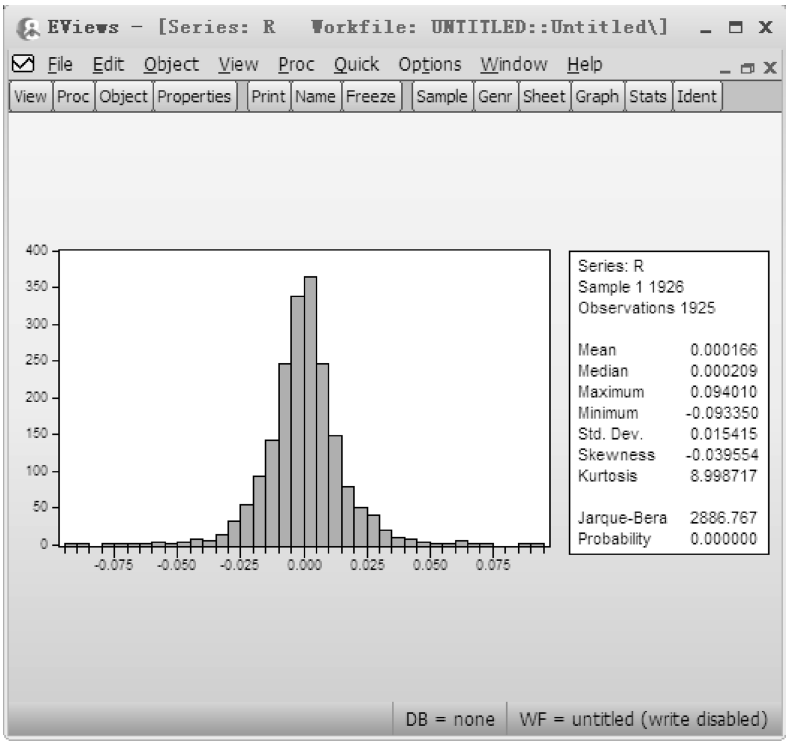


图 2－3　日收益率(*R*)的统计特征图

表 2－1　上证指数日收益率统计特征描述

均　值	0.000166	中位数	0.000209
最大值	0.094010	最小值	－0.093350
标准差	0.015415	偏　度	－0.039554
峰　度	8.998717	Jarque－Bera 值	2886.767

再选择“Series：R”窗口的“View”菜单中的“Line Graph”选项，可得到日收益率线性图，如图 2－4 所示。

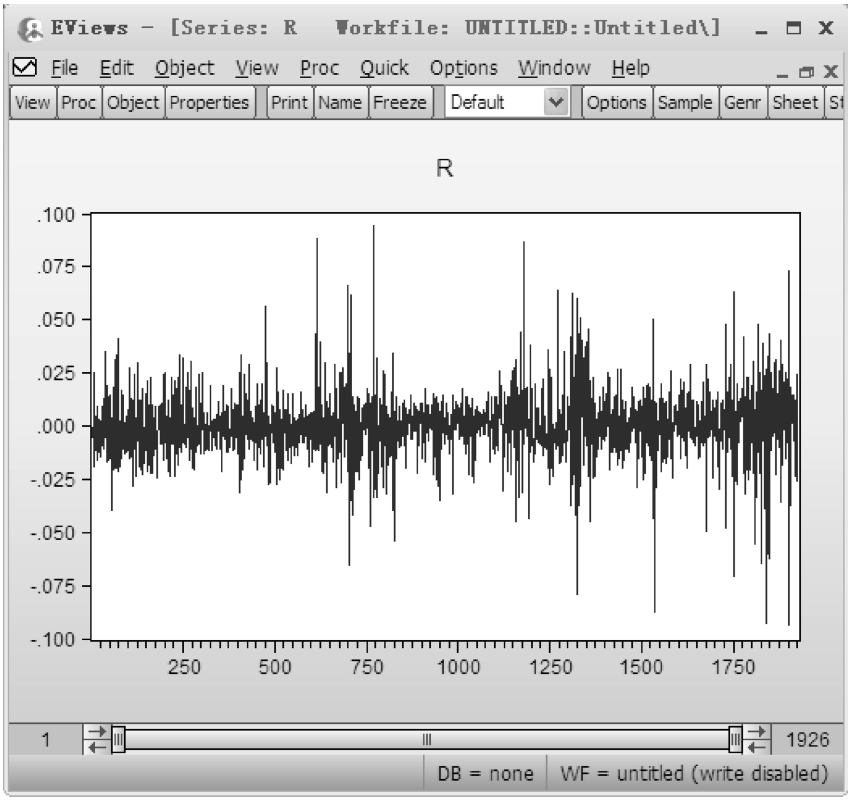


图 2－4　日收益率(R)的线性图

4. 回归数据估计方程

我们建立如下的虚拟变量模型：

$$R_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{4t} + \alpha_4 D_{5t} + \varepsilon_t$$

其中，R_t 表示 t 时刻的收益率，虚拟变量 $D_{it}(i=1, 2, 4, 5)$ 的取值在每周的第 i 天（一周五天）取值为 1，其余时刻取值为 0。从模型可以看到，α_0 表示的是周三的平均收益，而 α_1、α_2、α_3、α_4 分别表示的是星期一、星期二、星期四、星期五与星期三平均

收益的差异。若差异是显著的，则可以表明上海股票市场存在周内效应。

使用 EViews 软件对上述模型进行 OLS 回归，具体操作如下：

在 EViews 命令窗口中输入“LS R C D1 D2 D4 D5”并回车，得到如图 2－5 所示结果。即：（括号内为相应的 t 值）

$$R_t = 0.001107 - 0.001495D_{1t} - 0.000793D_{2t} - 0.001982D_{4t} - 0.000445D_{5t}$$
$$(1.41) \qquad (-1.35) \qquad (-0.72) \qquad (-1.78) \qquad (-0.40)$$

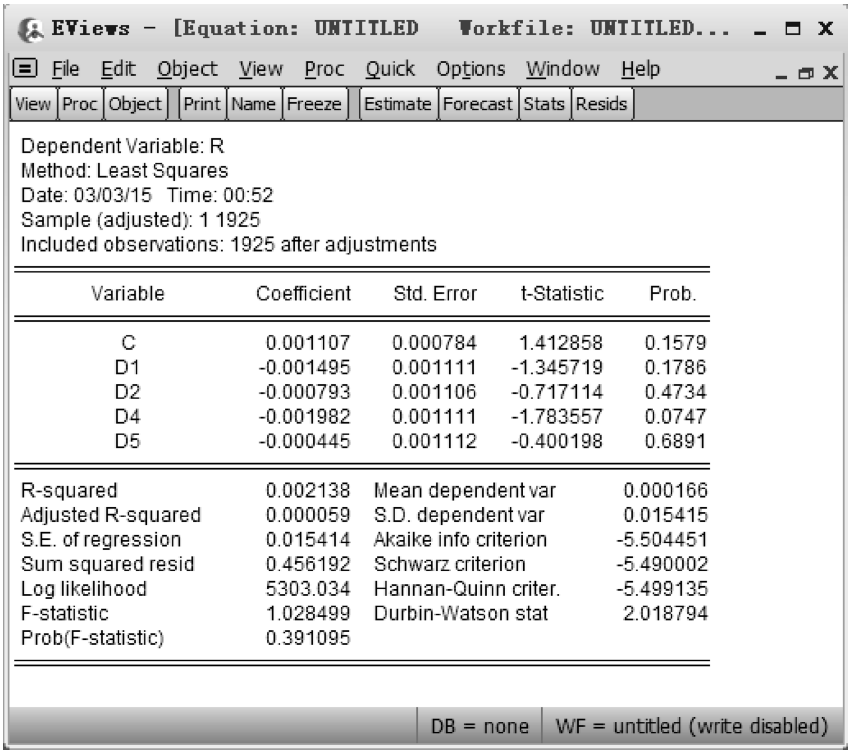


图 2－5　虚拟变量模型回归结果

对模型各系数估计值进行联合 F 检验，看各系数值是否同时为零，结果的到 F 值为 1.03，其概率值为 0.39，因此不能拒绝各系数值同时为零的假设，则可以得出结论，上海股票市场不存在周内效应。

实际上，这个模型是相当粗糙的，因此结论也可能不具有代表性。因为在使用最小二乘法进行回归时，我们假定随机误差项满足同方差的假定，但大量的研究表明，金融数据中这一假定往往不能得到满足。更深入的研究涉及自回归条件异方差模型（ARCH）以及扩展的自回归条件异方差模型（GARCH），我们将在后面的章节中予以介绍，有兴趣的同学那时也可以重新对上海股票市场的周内效应作出检验。

实验三　金融数据的平稳性检验

一、实验目的

理解经济时间序列存在的不平稳性，掌握 ADF 检验平稳性的方法。认识不平稳的序列容易导致伪回归问题，掌握为解决伪回归问题引出的协整检验，协整的概念和具体的协整检验过程。协整描述了变量之间的长期关系，为了进一步研究变量之间的短期均衡是否存在，掌握误差纠正模型方法。理解变量之间的因果关系的计量意义，掌握格兰杰因果检验方法。

二、基本概念

如果一个随机过程的均值和方差在时间过程上都是常数，并且在任何两时期的协方差值仅依赖于该两时期间的距离或滞后，而不依赖于计算这个协方差的实际时间，就称它为平稳的。强调平稳性是因为将一个随机游走变量（即非平稳数据）对另一个随机游走变量进行回归可能导致荒谬的结果，传统的显著性检验将告知我们变量之间的关系是不存在的。这种情况就称为“伪回归”。

有时虽然两个变量都是随机游走的，但它们的某个线形组合却可能是平稳的，在这种情况下，我们称这两个变量是协整的。

因果检验用于确定一个变量的变化是否为另一个变量变化的原因。

三、实验内容及要求

（一）实验内容

用 EViews 来分析上海证券市场 A 股成分指数（简记 SHA）和深圳证券市场 A 股成分指数（简记 SZA）之间的关系，包括：

（1）对数据进行平稳性检验；

（2）协整检验；

（3）因果检验；

（4）误差纠正机制 ECM。

（二）实验要求

在认真理解本章内容的基础上，通过实验掌握 ADF 检验平稳性的方法，具体的协整检验过程，掌握格兰杰因果检验方法，以及误差纠正模型方法。

四、实验指导

（一）对数据进行平稳性检验

首先导入数据，将上海证券市场 A 股成分指数记为 SHA，深圳证券市场 A 股成分指数记为 SZA（若已有 wf1 文件则直接打开该文件）。

在 workfile 中按住 ctrl 选择要检验的二变量，右击，选择“open-as group”。此时可在弹出的窗口中对选中的变量进行检验。检验方法有：

（1）画折线图。“View”—“graph”—“line”，如图 3－1 所示。

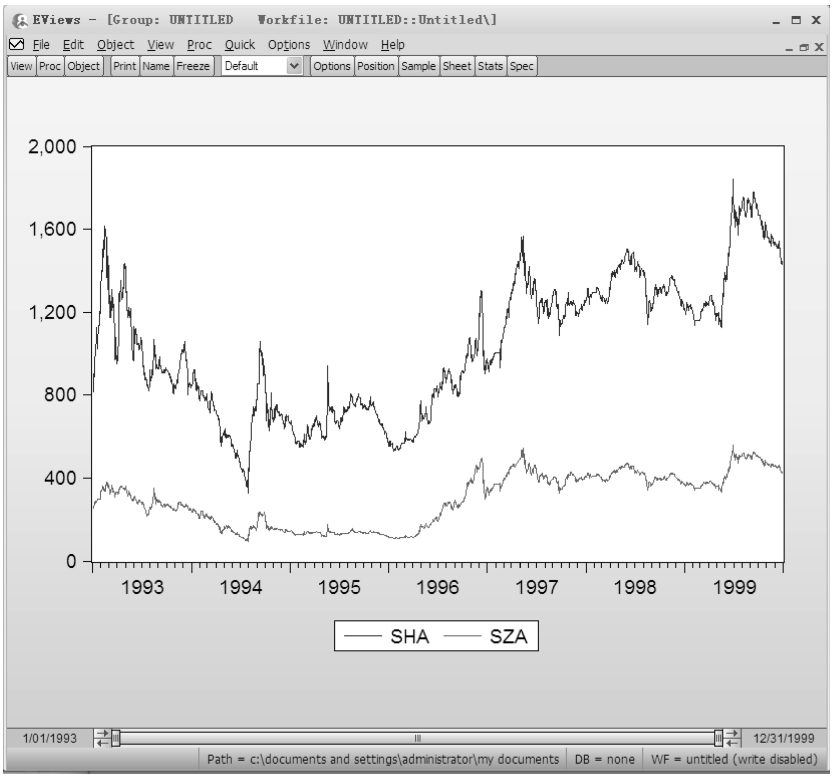


图 3－1　SHA 和 SZA 原始数值线性图

（2）画直方图。在 workfile 中按住选择要检验的变量，右击，选择“open”，或双击选中的变量，“view”-“descriptive statistic”-“histogram and stats”；注意到图中的 JB 统计量，其越趋向于 0，则图越符合正态分布，也就说明数据越平稳，如图 3－2 和图 3－3 所示。

（3）用 ADF 检验。方法一：“view”—“unit root test”；方法二：点击菜单中的“quick”—“series statistic”—“unit root test”。分析原则即比较值的大小以及经验法则。点击“OK”，如图 3－4、表 3－1、表 3－2 所示。

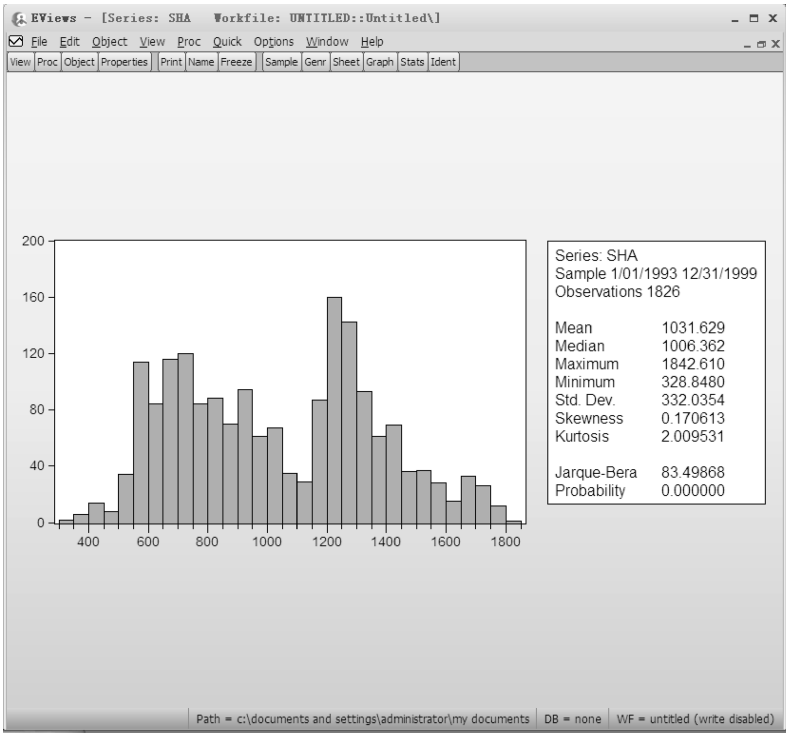


图 3－2　SHA 原始数值直方图

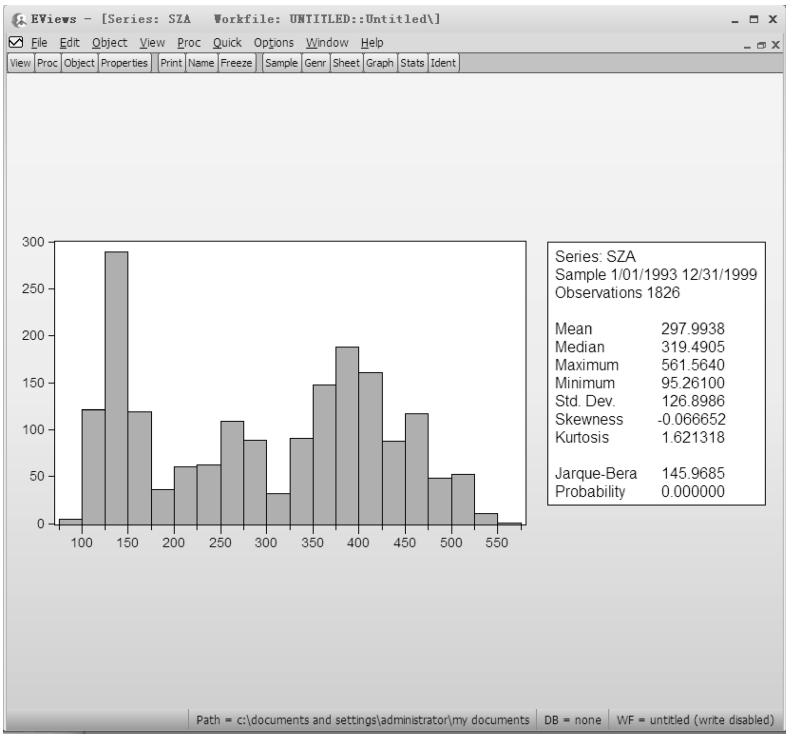


图 3－3　SZA 原始数值直方图

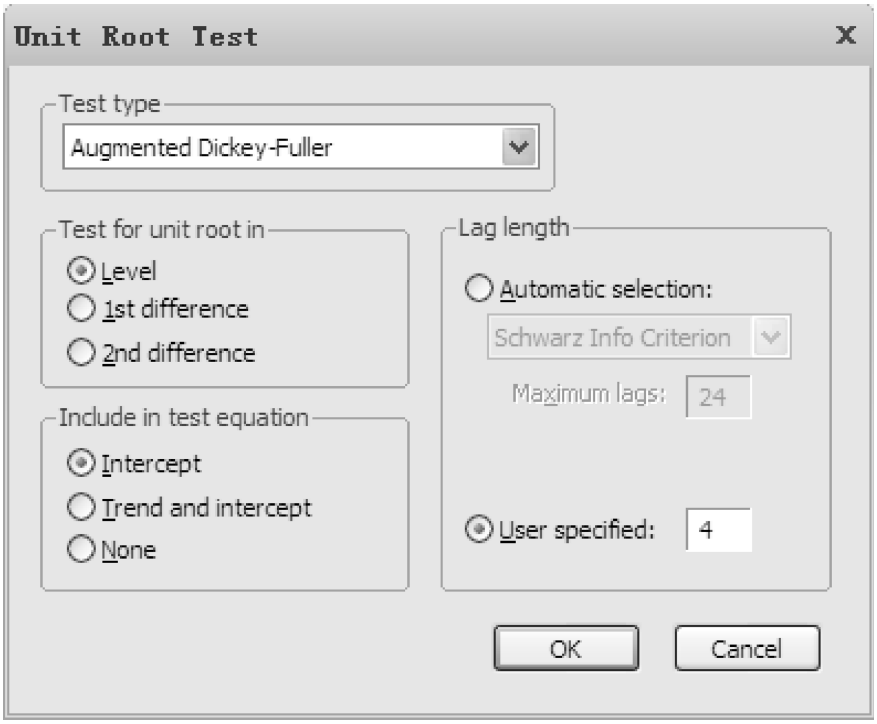


图 3－4　单位根检验对话框

表 3－1　SHA 数值的 ADF 检验结果

Null Hypothesis: SHA has a unit root

Exogenous: Constant

Lag Length: 4 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.824806	0.3687
Test critical values:	1% level	-3.433739	
	5% level	-2.862924	
	10% level	-2.567554	

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

（续表）

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SHA)
Method: Least Squares
Date: 12/19/14 Time: 23:02
Sample (adjusted): 1/08/1993 12/31/1999
Included observations: 1821 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SHA(-1)	-0.003575	0.001959	-1.824806	0.0682
D(SHA(-1))	-0.038736	0.023427	-1.653464	0.0984
D(SHA(-2))	-0.010797	0.023308	-0.463217	0.6433
D(SHA(-3))	0.111127	0.023287	4.772149	0.0000
D(SHA(-4))	0.062380	0.023399	2.665901	0.0077
C	3.943077	2.121673	1.858476	0.0633

R-squared	0.018447	Mean dependent var	0.295316
Adjusted R-squared	0.015743	S.D. dependent var	27.87568
S.E. of regression	27.65538	Akaike info criterion	9.480807
Sum squared resid	1388148.	Schwarz criterion	9.498952
Log likelihood	-8626.275	Hannan-Quinn criter.	9.487501
F-statistic	6.822257	Durbin-Watson stat	2.001095
Prob(F-statistic)	0.000003		

表 3－2 SZA 数值的 ADF 检验结果

Null Hypothesis: SZA has a unit root
Exogenous: Constant
Lag Length: 4 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.386897	0.5902
Test critical values:	1% level	-3.433739	
	5% level	-2.862924	
	10% level	-2.567554	

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

（续表）

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SZA)
Method: Least Squares
Date: 12/19/14 Time: 23:03
Sample (adjusted): 1/08/1993 12/31/1999
Included observations: 1821 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SZA(-1)	-0.001999	0.001441	-1.386897	0.1656
D(SZA(-1))	-0.028638	0.023396	-1.224056	0.2211
D(SZA(-2))	0.029664	0.023325	1.271755	0.2036
D(SZA(-3))	0.084650	0.023327	3.628817	0.0003
D(SZA(-4))	0.081428	0.023390	3.481380	0.0005
C	0.667786	0.466362	1.431905	0.1523

R-squared	0.015405	Mean dependent var	0.087348
Adjusted R-squared	0.012693	S.D. dependent var	7.839108
S.E. of regression	7.789199	Akaike info criterion	6.946643
Sum squared resid	110119.0	Schwarz criterion	6.964788
Log likelihood	-6318.918	Hannan-Quinn criter.	6.953337
F-statistic	5.679524	Durbin-Watson stat	1.998663
Prob(F-statistic)	0.000033		

粗略观察数据并不平稳。此时应对数据取对数（取对数的好处在于：即可以将间距很大的数据转换为间距较小的数据，也便于后面的取差分），再对新变量进行平稳性检验。点击 EViews 中的“quick”—“generate series”，键入 logsha＝log（sha），同样的方法得到 logsza。此时，logsha 和 logsza 为新变量，对其进行平稳性检验方法如上，发现也是不平稳的，如图 3－5 所示。

用 ADF 方法检验 logsha 和 logsza 的平稳性。通过比较检验值和不同显著性下的关键值来得出结论。如表 3－3 和表 3－4 所示，检验值大于关键值，则得出数据不平稳，反之平稳。

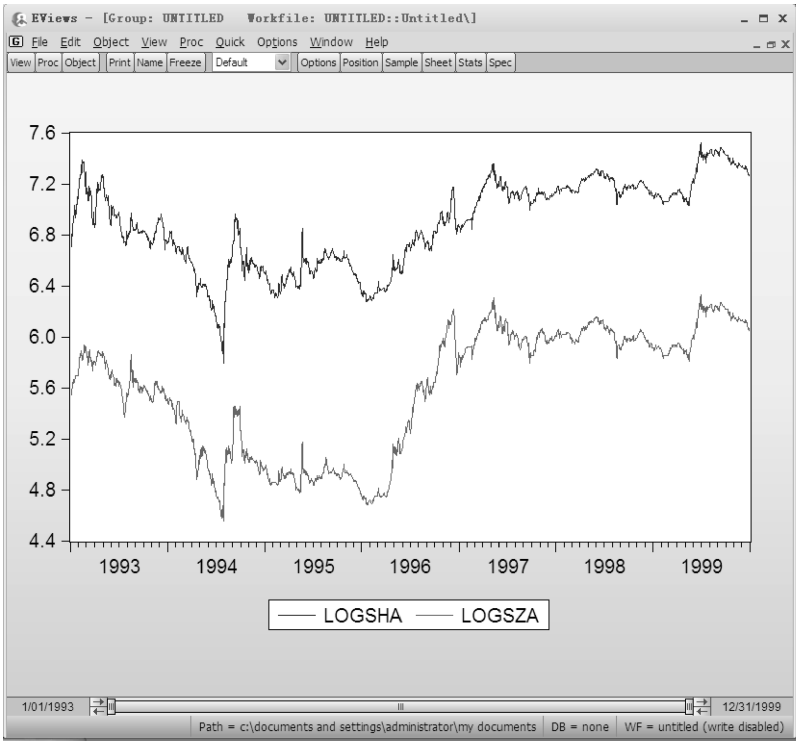


图 3－5 SHA 和 SZA 对数值线性图

表 3－3 SHA 对数值的 ADF 检验结果

Null Hypothesis: LOGSHA has a unit root

Exogenous: Constant

Lag Length: 4 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.795526	0.3831
Test critical values:	1% level	-3.433739	
	5% level	-2.862924	
	10% level	-2.567554	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(LOGSHA)

Method: Least Squares

Date: 12/19/14 Time: 23:04

Sample (adjusted): 1/08/1993 12/31/1999

（续表）

Included observations: 1821 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGSHA(-1)	-0.003583	0.001995	-1.795526	0.0727
D(LOGSHA(-1))	-0.034725	0.023459	-1.480261	0.1390
D(LOGSHA(-2))	0.020525	0.023427	0.876128	0.3811
D(LOGSHA(-3))	0.065236	0.023404	2.787354	0.0054
D(LOGSHA(-4))	0.034323	0.023421	1.465476	0.1430
C	0.024892	0.013751	1.810156	0.0704

R-squared	0.008123	Mean dependent var	0.000254
Adjusted R-squared	0.005391	S.D. dependent var	0.029001
S.E. of regression	0.028923	Akaike info criterion	-4.245075
Sum squared resid	1.518313	Schwarz criterion	-4.226929
Log likelihood	3871.140	Hannan-Quinn criter.	-4.238380
F-statistic	2.972845	Durbin-Watson stat	2.001003
Prob(F-statistic)	0.011179		

表 3－4　SZA 对数值的 ADF 检验结果

Null Hypothesis: LOGSZA has a unit root

Exogenous: Constant

Lag Length: 4 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.236119	0.6608
Test critical values:	1% level	-3.433739	
	5% level	-2.862924	
	10% level	-2.567554	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(LOGSZA)

Method: Least Squares

Date: 12/19/14　Time: 23:04

（续表）

Sample (adjusted): 1/08/1993 12/31/1999

Included observations: 1821 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGSZA(-1)	-0.001645	0.001331	-1.236119	0.2166
D(LOGSZA(-1))	-0.010639	0.023402	-0.454600	0.6495
D(LOGSZA(-2))	0.043671	0.023391	1.866982	0.0621
D(LOGSZA(-3))	0.033284	0.023393	1.422825	0.1550
D(LOGSZA(-4))	0.078284	0.023392	3.346659	0.0008
C	0.009404	0.007463	1.260037	0.2078

R-squared	0.009984	Mean dependent var	0.000252
Adjusted R-squared	0.007257	S.D. dependent var	0.027998
S.E. of regression	0.027897	Akaike info criterion	-4.317335
Sum squared resid	1.412468	Schwarz criterion	-4.299190
Log likelihood	3936.934	Hannan-Quinn criter.	-4.310641
F-statistic	3.660782	Durbin-Watson stat	2.001713
Prob(F-statistic)	0.002675		

（二）协整检验

首先要提取残差：点击菜单中的“quick”—“estimate equation”键入“logsha c logsza”，得到结果如表 3－5 所示。

表 3－5　logsza 对 logsha 的最小二乘回归

Dependent Variable: LOGSHA

Method: Least Squares

Date: 12/19/14　Time: 23:05

Sample: 1/01/1993 12/31/1999

Included observations: 1826

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.185265	0.026985	118.0392	0.0000
LOGSZA	0.661851	0.004811	137.5733	0.0000

（续表）

R-squared	0.912098	Mean dependent var	6.883358
Adjusted R-squared	0.912050	S.D. dependent var	0.340928
S.E. of regression	0.101107	Akaike info criterion	-1.744184
Sum squared resid	18.64600	Schwarz criterion	-1.738149
Log likelihood	1594.440	Hannan-Quinn criter.	-1.741958
F-statistic	18926.43	Durbin-Watson stat	0.041307
Prob(F-statistic)	0.000000		

接着在窗口中点击“procs”—“make residual series”来对残差 resid01 进行提取和保存；然后对残差进行 ADF 检验（方法同上），得到结果如表 3－6 所示。你会发现数据通过了检验，残差 resid01 是平稳的，所以 logsha 同 logsza 有协整关系。

表 3－6　残差 resid01 的 ADF 检验结果

Null Hypothesis: RESID01 has a unit root

Exogenous: Constant

Lag Length: 4 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.132316	0.0009
Test critical values:	1% level	-3.433739	
	5% level	-2.862924	
	10% level	-2.567554	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(RESID01)

Method: Least Squares

Date: 12/19/14　Time: 23:06

Sample (adjusted): 1/08/1993 12/31/1999

Included observations: 1821 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01(-1)	-0.019808	0.004793	-4.132316	0.0000
D(RESID01(-1))	-0.089306	0.023497	-3.800810	0.0001

（续表）

D(RESID01(-2))	-0.020115	0.023563	-0.853691	0.3934
D(RESID01(-3))	0.064304	0.023497	2.736735	0.0063
D(RESID01(-4))	0.022089	0.023396	0.944140	0.3452
C	9.14E-05	0.000476	0.192199	0.8476

R-squared	0.023020	Mean dependent var	8.71E-05
Adjusted R-squared	0.020329	S.D. dependent var	0.020512
S.E. of regression	0.020303	Akaike info criterion	-4.952841
Sum squared resid	0.748139	Schwarz criterion	-4.934695
Log likelihood	4515.561	Hannan-Quinn criter.	-4.946146
F-statistic	8.553192	Durbin-Watson stat	1.996742
Prob(F-statistic)	0.000000		

接下来以同样的方法协整 logsza c logsha，得到残差 resid02，经过检验也是平稳的，如表 3－7 所示。

表 3－7　残差 resid02 的 ADF 检验结果

Null Hypothesis: RESID02 has a unit root

Exogenous: Constant

Lag Length: 4 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.900100	0.0021
Test critical values:	1% level	-3.433739	
	5% level	-2.862924	
	10% level	-2.567554	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(RESID02)

Method: Least Squares

Date: 12/19/14　Time: 23:09

（续表）

Sample (adjusted): 1/08/1993 12/31/1999

Included observations: 1821 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID02(-1)	-0.017724	0.004544	-3.900100	0.0001
D(RESID02(-1))	-0.095416	0.023495	-4.061081	0.0001
D(RESID02(-2))	-0.024582	0.023577	-1.042621	0.2973
D(RESID02(-3))	0.059774	0.023511	2.542356	0.0111
D(RESID02(-4))	0.022353	0.023395	0.955429	0.3395
C	-0.000105	0.000652	-0.160597	0.8724

R-squared	0.022832	Mean dependent var	-9.79E-05
Adjusted R-squared	0.020140	S.D. dependent var	0.028126
S.E. of regression	0.027841	Akaike info criterion	-4.321324
Sum squared resid	1.406845	Schwarz criterion	-4.303179
Log likelihood	3940.566	Hannan-Quinn criter.	-4.314630
F-statistic	8.481765	Durbin-Watson stat	1.996185
Prob(F-statistic)	0.000000		

（三）因果检验

在 workfile 中同时选中“logsha”和“logsza”，右击，选择“open”—“as group”，在弹出的窗口中点击“view”—“granger causality”并选择滞后阶数（此处我们根据以往的实证检验结果选择滞后值为 5），点击 OK，结果如表 3－8 所示。

表 3－8　格兰杰因果检验结果

Pairwise Granger Causality Tests

Date: 12/19/14　　Time: 23:11

Sample: 1/01/1993 12/31/1999

Lags: 5

Null Hypothesis:	Obs	F-Statistic	Prob.
LOGSZA does not Granger Cause LOGSHA	1821	3.60184	0.0030
LOGSHA does not Granger Cause LOGSZA		0.70399	0.6205

根据 F 检验值和相应的 P 值，在本例中即 logsza 是 logsha 变化的原因；而 logsha 不影响 logsza。

（四）误差纠正机制 ECM(error correction mechanism)

即使两个变量之间有长期均衡关系，但在短期内也会出现失衡（例如受突发事件的影响）。此时，我们可以用 ECM 来对这种短期失衡加以纠正。

具体做法是：首先要提取残差，从“sha c sza”中提取残差“resid03”，接着点击“quick”—“estimate equation”，在弹出得窗口中输入：“d(sha)c d(sza)resid03(－1)”。Resid03(－1)中的(－1)指的是滞后一阶，结果如表 3－9 所示。

表 3－9　误差修正模型结果

Dependent Variable: D(SHA)
Method: Least Squares
Date: 12/19/14　Time: 23:18
Sample (adjusted): 1/04/1993 12/31/1999
Included observations: 1825 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.109030	0.468941	0.232503	0.8162
D(SZA)	2.462137	0.059863	41.12931	0.0000
RESID03(-1)	-0.021581	0.004824	-4.473995	0.0000

R-squared	0.484705	Mean dependent var	0.348548
Adjusted R-squared	0.484139	S.D. dependent var	27.89010
S.E. of regression	20.03164	Akaike info criterion	8.834145
Sum squared resid	731107.4	Schwarz criterion	8.843202
Log likelihood	-8058.157	Hannan-Quinn criter.	8.837486
F-statistic	856.9180	Durbin-Watson stat	2.172798
Prob(F-statistic)	0.000000		

resid03(－1)的系数为－0.021 581，且通过了 t 检验(4.648231＞2)，其表明 sha 的实际值与长期或均衡值之间的差异约有 2.158 1%得以纠正。从这也可以看出 resid03(－1)的系数必须为负值。

从表面上看，SZA 对 SHA 的影响要更强一点，SHA 对 SZA 的依赖也更多一点，但总体看来两个市场的联系还是很紧密的。SZA 走在前面的原因可能是因为深圳的地理位置，与海外市场联系更密切一些。所以海外市场大市变化的信息最先传递和影

响深圳市场，经过一段时间，蔓延到内陆地区。从整体上看，就形成 SHA 跟在 SZA 后面变动的局面。而两个市场的投资者包括投资理念等各方面都是类似的，总体对价格信息的表现也大同小异，两个市场相关度很高可以理解。

值得指出的是，目前一般认为，深市股指是随沪市股指而动，与我们上面的检验结论相反。但应该注意到的是，我们上边研究中的样本范围为 1993—1999 年，而现在的情况已经发生了很大变化。所以，若要研究当前股指的联动效应，需选择最新的样本范围。有兴趣的同学不妨一试，看是否会得出新的结论。

实例四　基于 Microfit 的 ARDL 模型应用

一、实验目的

理解 ARDL 模型的原理与应用条件，运用 ARDL 模型，估计变量之间长期关系的系数。注意，只有当能够确定第一步所估计的变量间的长期关系是肯定存在的，而不是伪回归，那么才能应用该模型进行估计。

二、基本概念

ARDL(autoregressive distributed lag)称为自回归分布滞后模型。ARDL 模型的一大优点，就是我们不用管变量是否同阶单整，都可以用 ARDL 模型来检验变量之间的长期关系，而这是标准的协整检验所做不到的。

三、实验内容及要求

（一）实验内容

运用 ARDL 模型研究美国非耐用消费品支出 LC(取对数形式)与真实可支配收入 LY(取对数形式)，通胀率 DP 之间的关系，数据为 1960 年第一季度到 1994 年第一季度的季度数据。

（二）实验要求

在认真理解本章内容的基础上，通过实验掌握 ARDL 模型的实际应用方法，并熟悉 Microfit 软件的基本使用方法。

四、实验指导

ARDL 模型的主要优点在于不管回归项是 $I(0)$ 还是 $I(1)$，都可以进行检验和估计。

首先，我们调用 Microfit 软件读入 EX6.1 的数据文件。对原始数据进行取对数作差分的处理。

由于观测值是季度数据，ARDL 模型中最大滞后阶数取 4 阶，利用 1960 年第一季度到 1992 年第四季度的样本区间进行估计，1993 年第一季度到 1994 年第一季度的数据进行预测。

对应于 ARDL(4,4,4)中变量 LC，LY 和 DP 的误差修正模型(ECM)如下：

$$DLC_t = \alpha_0 + \sum_{i=1}^{4} b_i DLC_{t-i} + \sum_{i=1}^{4} d_i DLY_{t-i} + \sum_{i=1}^{4} e_i DPI_{t-i} + \delta_1 LC_{t-1} + \delta_2 LY_{t-1} + \delta_3 PI_{t-1} + u_t \tag{6.4}$$

检验的原假设是：变量间不存在稳定的长期关系，即：$H_0: \delta_1 = \delta_2 = \delta_3 = 0$。

备择假设是：$H_1: \delta_1 \neq 0$ 或 $\delta_2 \neq 0$ 或 $\delta_3 \neq 0$。

检验 δ_1、δ_2、δ_3 联合显著的统计量就是我们熟悉的 F 统计量。为了计算 F 统计量，在 Microfit 中选择 Single，在编辑窗中输入：

DLC　INPT　DLC{1－4}　DLY{1－4}　DPI{1－4}

选择估计样本期 1960 年第一季度到 1992 年第四季度，按 START，然后点击 OK，就得到了用 OLS 估计的一阶差分的回归结果，这个结果对我们没有直接的用途。按 CLOSE 回到选择菜单，选“2. Move to hypothesis testing menu”，如图 4－1 所示。

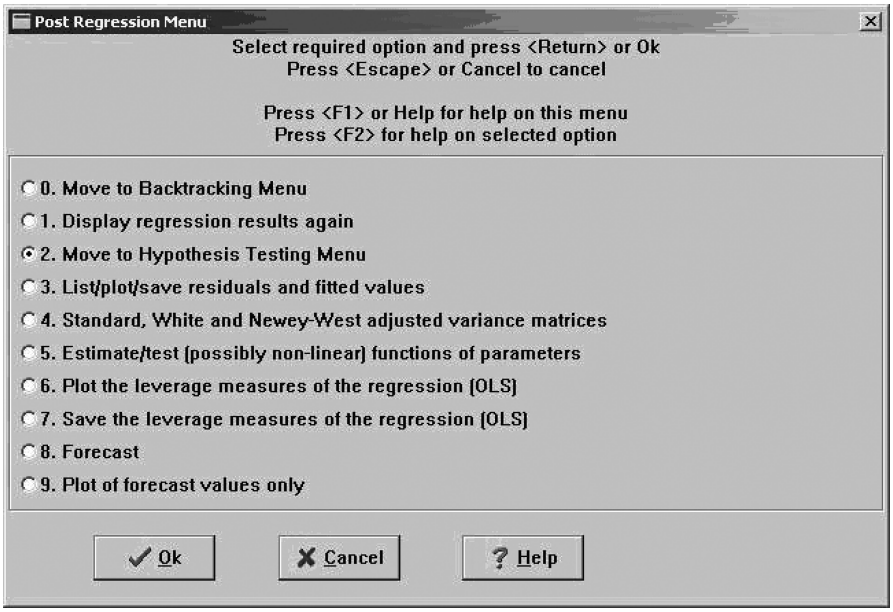


图 4－1　选择假设检验

点击“OK”出现以下窗口，如图 4－2 所示。

选“6. Variable addition test”，点击“OK”后在“Input text”窗口中输入长期变量的滞后值：

LC(－1)　LY(－1)　PI(－1)

点击“OK”后出现以下的估计结果，如图 4－3 所示。

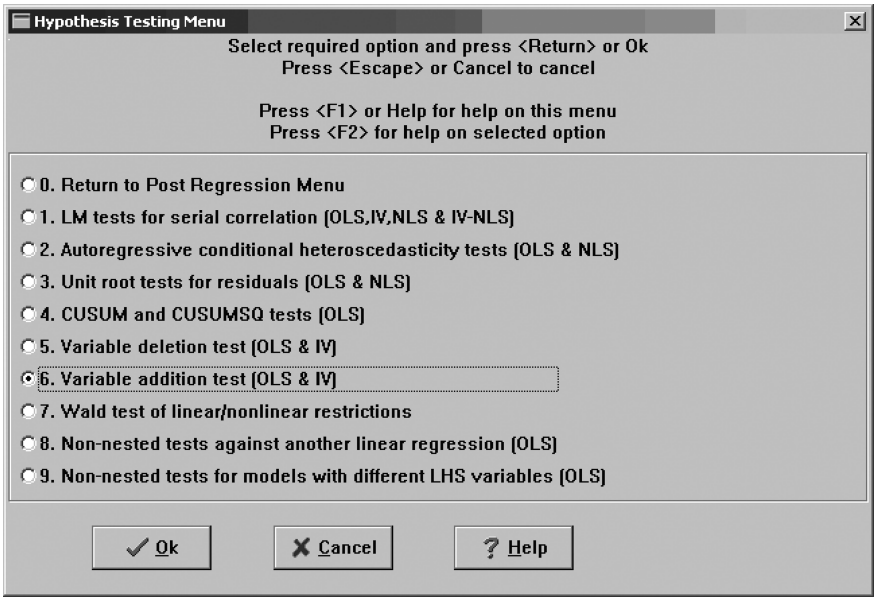


图 4－2　假设检验窗口

```
                    Variable Addition Test (OLS case)
*******************************************************************************
 Dependent variable is DLC
 List of the variables added to the regression:
 LC(-1)          LY(-1)          PI(-1)
 126 observations used for estimation from 1961Q3 to 1992Q4
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 INPT                       .18833            .061053              3.0847[.003]
 DLC(-1)                    .22537             .10246              2.1996[.030]
 DLC(-2)                  -.045573             .10512             -.43353[.665]
 DLC(-3)                    .20189             .10364              1.9480[.054]
 DLC(-4)                  -.067611            .098364             -.68736[.493]
 DLY(-1)                    .10464            .072392              1.4455[.151]
 DLY(-2)                  -.087969            .074826             -1.1757[.242]
 DLY(-3)                  -.012725            .075130             -.16937[.866]
 DLY(-4)                  -.082102            .071981             -1.1406[.257]
 DPI(-1)                    .24041             .11833              2.0316[.045]
 DPI(-2)                    .29154             .12176              2.3943[.018]
 DPI(-3)                   .025833             .11587              .22296[.824]
 DPI(-4)                   .034089            .095690              .35624[.722]
 LC(-1)                   -.12997            .046251             -2.8101[.006]
 LY(-1)                    .088399            .032299              2.7369[.007]
 PI(-1)                   -.31301            .091620             -3.4164[.001]
*******************************************************************************
 Joint test of zero restrictions on the coefficients of additional variables:
 Lagrange Multiplier Statistic     CHSQ( 3)=  16.2601[.001]
 Likelihood Ratio Statistic        CHSQ( 3)=  17.4093[.001]
 F Statistic                    F(  3, 110)=   5.4329[.002]
*******************************************************************************
```

图 4－3　假设检验结果

F 统计量出现在该结果的最后一行。F 统计量用于检验原假设：所有水平变量的系数为零(即水平变量之间不存在长期关系)，我们记作 $F(LC \mid LY, PI)=5.43$。在原假设 $H_0: \delta_1=\delta_2=\delta_3=0$ 成立时，F 统计量服从一个非标准的分布，而不管 LC、

LY 和 PI 是 $I(0)$ 还是 $I(1)$ 过程。Pesaran 已经计算出了该检验的临界范围表，参见附表 5。查表可知，在 95% 的置信水平该 F 统计量的范围在 3.793～4.855。因为 $F(LC \mid LY, PI)=5.43$ 超出了临界值上界，我们就能拒绝 LC，LY 和 PI 之间没有长期关系的原假设，而不管它们是 $I(0)$ 还是 $I(1)$ 过程。

以上我们已经证明 LY 和 PI 对 LC 有长期的影响。同时，我们还要考虑：LC 和 PI 对 LY 是否有长期的影响？LC 和 LY 对 PI 是否有长期的影响？重复以上的过程，只是把因变量 DLC 分别替换成 DLY 和 DPI，我们可以得到以下结果：

$$F(LY \mid LC, PI)=2.631 \quad F(PI \mid LY, LC)=1.359$$

以上两个统计量都低于临界范围的下界 3.793，从而我们无法拒绝原假设：在 DLY 和 DPI 作为因变量的方程中，加入长期变量是不显著的。同样，该结论的成立不依赖它们是 $I(0)$ 还是 $I(1)$ 过程。

以上的检验结果显示：LC 和 LY，PI 之间存在长期关系，LY 和 PI 对 LC 有长期的影响。

现在，我们用 Microfit 软件中的 ARDL 选项来估计变量间的长期系数以及相应的误差修正模型 ECM。在 Univariate 菜单中选择 6. ARDL approach to cointegration，清空编辑窗口，然后键入：

LC　LY　PI　&　INPT

选择样本期 1960 年第一季度到 1992 年第四季度进行估计，最大滞后值取 4，点击 START。Microfit 软件估计了 125 个回归方程，即共 $(4+1)^{2+1}$ 个回归方程，并提供了以下选择菜单，如图 4－4 所示。

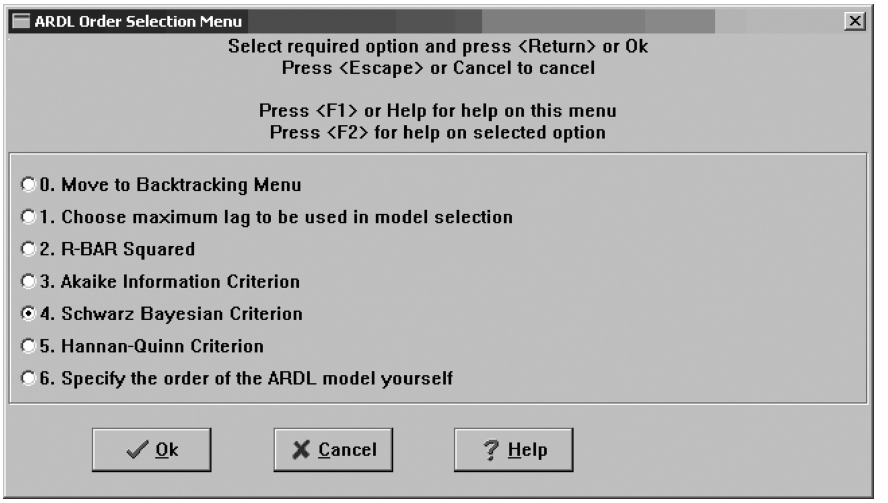


图 4－4　ARDL 选择菜单

用 SBC 准则选择的模型为 ARDL(1，2，0)，用 AIC 准则选择的模型为 ARDL(2,2,3)，估计的变量之间长期关系的系数分别如图 4－5 和图 4－6 所示。

```
                    Autoregressive Distributed Lag Estimates
            ARDL(1,2,0) selected based on Schwarz Bayesian Criterion
*******************************************************************************
 Dependent variable is LC
 127 observations used for estimation from 1961Q2 to 1992Q4
*******************************************************************************
 Regressor               Coefficient       Standard Error         T-Ratio[Prob]
 LC(-1)                      .91210            .035045            26.0265[.000]
 LY                          .28429            .056735             5.0108[.000]
 LY(-1)                    -.031577            .077772            -.40603[.685]
 LY(-2)                     -.19179            .054818            -3.4986[.001]
 PI                         -.22807            .057159            -3.9901[.000]
 INPT                        .11742            .046931             2.5019[.014]
*******************************************************************************
 R-Squared                      .99925   R-Bar-Squared                   .99922
 S.E. of Regression          .0056198   F-stat.    F(  5, 121)   32083.0[.000]
 Mean of Dependent Variable    6.6849   S.D. of Dependent Variable      .20060
 Residual Sum of Squares     .0038214   Equation Log-likelihood       480.9132
 Akaike Info. Criterion      474.9132   Schwarz Bayesian Criterion    466.3807
 DW-statistic                  1.7919   Durbin's h-statistic      1.2766[.202]
*******************************************************************************
```

图 4－5　ARDL(1,2,0)估计结果

```
                    Autoregressive Distributed Lag Estimates
            ARDL(2,2,3) selected based on Akaike Information Criterion
*******************************************************************************
 Dependent variable is LC
 127 observations used for estimation from 1961Q2 to 1992Q4
*******************************************************************************
 Regressor               Coefficient       Standard Error         T-Ratio[Prob]
 LC(-1)                      .99052            .088563            11.1843[.000]
 LC(-2)                     -.11651            .085463            -1.3633[.175]
 LY                          .26694            .056676             4.7098[.000]
 LY(-1)                    -.012513            .077669            -.16110[.872]
 LY(-2)                     -.16621            .060825            -2.7326[.007]
 PI                         -.18461            .080063            -2.3058[.023]
 PI(-1)                     .085733            .089530             .95759[.340]
 PI(-2)                     .063324            .086584             .73136[.466]
 PI(-3)                     -.25269            .078663            -3.2123[.002]
 INPT                        .15989            .047712             3.3511[.001]
*******************************************************************************
 R-Squared                      .99932   R-Bar-Squared                   .99927
 S.E. of Regression          .0054148   F-stat.    F(  9, 117)   19200.3[.000]
 Mean of Dependent Variable    6.6849   S.D. of Dependent Variable      .20060
 Residual Sum of Squares     .0034305   Equation Log-likelihood       487.7664
 Akaike Info. Criterion      477.7664   Schwarz Bayesian Criterion    463.5455
 DW-statistic                  1.9835
*******************************************************************************
```

图 4－6　ARDL(2,2,3)估计结果

两个模型估计的结果很近似。但是可以看出，用 AIC 准则选择的模型 ARDL(2，2,3)估计的标准误差(Standard Error)要比用 SBC 准则选择的模型 ARDL(1,2,0)估计的标准误差小。要得到长期估计的误差修正模型 ECM，在“Post ARDL Model Selection Menu”中选 3，如图 4－7 所示。

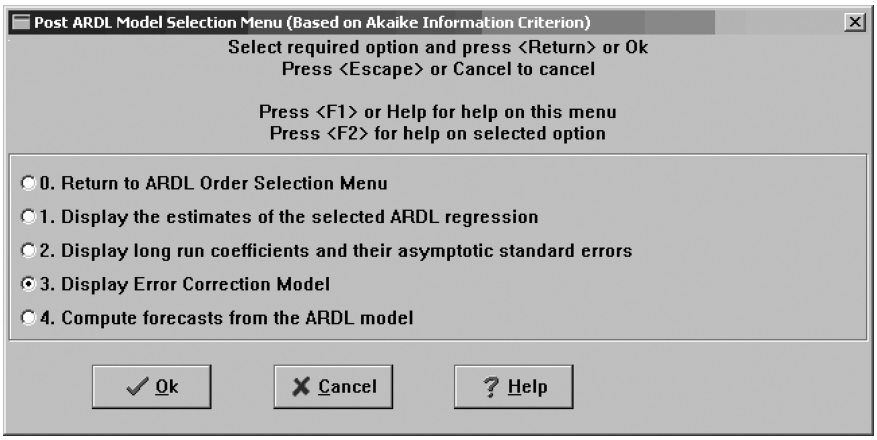


图 4－7　选择建立误差修正模型

用 AIC 准则选择的误差修正模型的结果如图 4－8 所示。

```
          Error Correction Representation for the Selected ARDL Model
          ARDL(2,2,3) selected based on Akaike Information Criterion
*******************************************************************************
 Dependent variable is dLC
 127 observations used for estimation from 1961Q2 to 1992Q4
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 dLC1                       .11651            .085463             1.3633[.175]
 dLY                        .26694            .056676             4.7098[.000]
 dLY1                       .16621            .060825             2.7326[.007]
 dPI                       -.18461            .080063            -2.3058[.023]
 dPI1                       .18936            .087764             2.1576[.033]
 dPI2                       .25269            .078663             3.2123[.002]
 dINPT                      .15989            .047712             3.3511[.001]
 ecm(-1)                   -.12599            .036172            -3.4832[.001]
*******************************************************************************
 List of additional temporary variables created:
 dLC = LC-LC(-1)
 dLC1 = LC(-1)-LC(-2)
 dLY = LY-LY(-1)
 dLY1 = LY(-1)-LY(-2)
 dPI = PI-PI(-1)
 dPI1 = PI(-1)-PI(-2)
 dPI2 = PI(-2)-PI(-3)
 dINPT = INPT-INPT(-1)
 ecm = LC   -.70016*LY +   2.2877*PI   -1.2690*INPT
*******************************************************************************
 R-Squared                      .46234   R-Bar-Squared                   .42098
 S.E. of Regression          .0054148   F-stat.     F(  7, 119)   14.3727[.000]
 Mean of Dependent Variable  .0055870   S.D. of Dependent Variable    .0071160
 Residual Sum of Squares     .0034305   Equation Log-likelihood       487.7664
 Akaike Info. Criterion      477.7664   Schwarz Bayesian Criterion    463.5455
 DW-statistic                  1.9835
*******************************************************************************
 R-Squared and R-Bar-Squared measures refer to the dependent variable
 dLC and in cases where the error correction model is highly
 restricted, these measures could become negative.
```

图 4－8　AIC 准则选定的误差修正模型结果

除了 DLC1 的系数外，其他系数都是统计显著的。ARDL 模型也通过了所有的诊断检验。误差修正项 ECM 的系数，估计为－0.12599(0.036172)，是统计上高度显著的，并且有正确的符号(负号)。这表明了一个向均衡收敛的合适的速度。误差修正项系数越大(绝对值)，表明经济受到冲击以后(once shocked)，向均衡回复的速度越快。

在给定过去和当前真实可支配收入和通胀变化的条件下，以上的误差修正模型也能用来预测消费的变化，在“Post ARDL Model Selection Menu”中选 4，如图 4－9 所示。

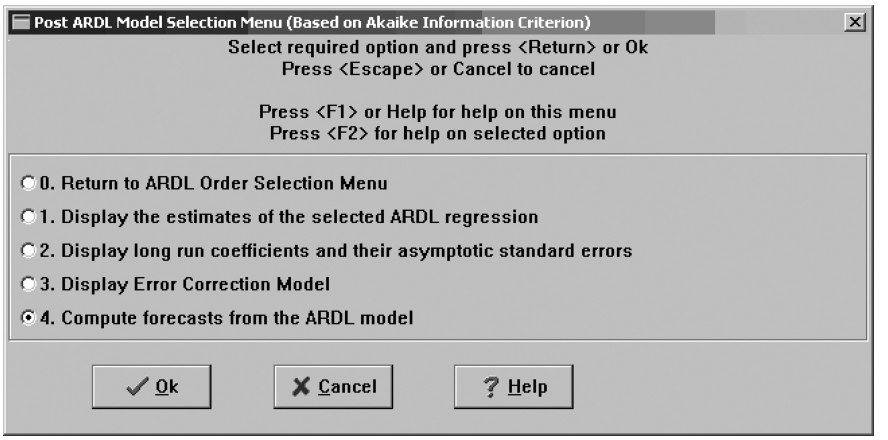


图 4－9　选择利用 ARDL 模型预测

然后点击“OK”，得到结果如图 4－10 所示。

```
                    Dynamic forecasts for the change in LC
*******************************************************************************
 Based on 127 observations from 1961Q2 to 1992Q4.
 ARDL(2,2,3) selected using Akaike Information Criterion.
 Dependent variable in the ARDL model is LC included with a lag of 2.
 List of other regressors in the ARDL model:
 LY              LY(-1)          LY(-2)          PI              PI(-1)
 PI(-2)          PI(-3)          INPT
*******************************************************************************
 Observation             Actual              Prediction              Error
  1993Q1               -.0053759              .0034962             -.0088721
  1993Q2                .0065768              .0026170              .0039598
  1993Q3                .0090587              .0078788              .0011798
  1993Q4                .0066042              .0048395              .0017647
  1994Q1                 .010188              .0078575              .0023304
*******************************************************************************

             Summary Statistics for Residuals and Forecast Errors
*******************************************************************************
                      Estimation Period            Forecast Period
                       1961Q2 to 1992Q4            1993Q1 to 1994Q1
*******************************************************************************
 Mean                      -.1712E-8                    .7253E-4
 Mean Absolute              .0040512                    .0036214
 Mean Sum Squares           .2701E-4                    .2087E-4
 Root Mean Sum Squares      .0051973                    .0045679
*******************************************************************************
```

图 4－10　预测结果

预测期间(1993Q1 to 1994Q1)误差平方和均值的平方根(root mean sum squares)大约为每季度0.45%,优于估计期间(1961Q2 to 1992Q4)0.519%的水平。然而,模型没有预测到1993年第一季度非耐用品消费支出的下降。

实验五　ARIMA 模型的概念和构造

一、实验目的

了解 AR,MA 以及 ARIMA 模型的特点,了解三者之间的区别联系,以及 AR 与 MA 的转换,掌握如何利用自相关系数和偏自相关系数对 ARIMA 模型进行识别,利用最小二乘法等方法对 ARIMA 模型进行估计,利用信息准则对估计的 ARIMA 模型进行诊断,以及如何利用 ARIMA 模型进行预测。掌握在实证研究如何运用 EViews 软件进行 ARIMA 模型的识别、诊断、估计和预测。

二、基本概念

所谓 ARIMA 模型,是指将非平稳时间序列转化为平稳时间序列,然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。ARIMA 模型根据原序列是否平稳以及回归中所含部分的不同,包括移动平均过程(MA)、自回归过程(AR)、自回归移动平均过程(ARMA)以及 ARIMA 过程。

在 ARIMA 模型的识别过程中,我们主要用到两个工具:自相关函数(简称 ACF),偏自相关函数(简称 PACF)以及它们各自的相关图(即 ACF、PACF 相对于滞后长度描图)。对于一个序列$\{Y_t\}$来说,它的第 j 阶自相关系数(记作 ρj)定义为它的 j 阶自协方差除以它的方差,即 $\rho_j=\gamma_j/\gamma_0$,它是关于 j 的函数,因此我们也称之为自相关函数,通常记 ACF(j)。偏自相关函数 PACF(j)度量了消除中间滞后项影响后两滞后变量之间的相关关系。

三、实验内容及要求

(一) 实验内容

根据 1991 年 1 月—2005 年 1 月我国货币供应量(广义货币 M2)的月度时间数据来说明在 EViews 软件中如何利用 BJ 方法论建立合适的 ARIMA(p, d, q)模型,并利用此模型进行数据的预测。

(二) 实验要求

(1) 深刻理解上述基本概念;

(2) 思考:如何通过观察自相关,偏自相关系数及其图形,利用最小二乘法,以及信息准则建立合适的 ARIMA 模型;如何利用 ARIMA 模型进行预测;

(3) 熟练掌握相关 EViews 操作。

四、实验指导

(一) ARIMA 模型的识别

1. 导入数据

打开 EViews 软件，选择“File”菜单中的“New - Workfile”选项，出现“Workfile Range”对话框，在“Workfile frequency”框中选择“Monthly”，在“Start date”和“End date”框中分别输入“1991：01”和“2005：01”，然后单击“OK”，选择“File”菜单中的“Import - Read Text - Lotus - Excel”选项，找到要导入的名为 EX6. 2. xls 的 Excel 文档，单击“打开”出现“Excel Spreadsheet Import”对话框并在其中输入相关数据名称(M2)，再单击“OK”，完成数据导入。

2. 模型的识别

首先利用 ADF 检验，确定 d 值，判断 M2 序列为 2 阶非平稳过程(由于具体操作方法我们在第五章中予以说明，此处略)，即 d 的值为 2，将两次差分后得到的平稳序列命名为 $W2$；下面我们来看 $W2$ 的自相关、偏自相关函数图。打开 $W2$ 序列，点击“View”—“Correlogram”菜单，会弹出如图 5 - 1 所示的窗口，我们选择滞后项数为 36，然后点击“OK”，就得到了 $W2$ 的自相关函数图和偏自相关函数图，如图 5 - 2 所示。

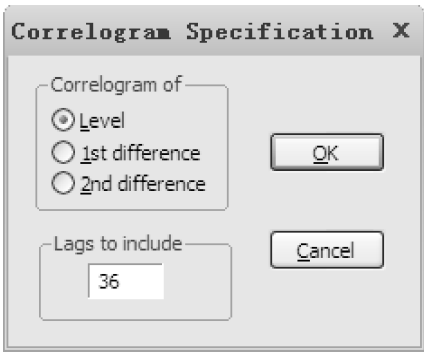
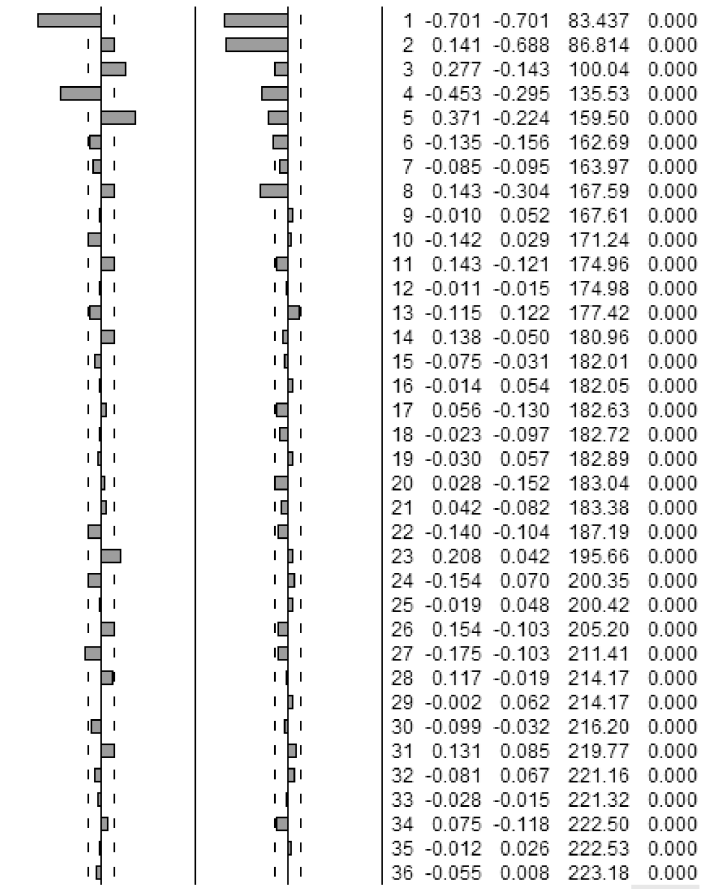


图 5 - 1　自相关形式设定

从 $W2$ 的自相关函数图和偏自相关函数图中我们可以看到，他们都是拖尾的，因此可设定为 ARMA 过程。$W2$ 的自相关函数 1—5 阶都是显著的，并且从第 6 阶开始下降很大，数值也不太显著，因此我们先设定 q 值为 5。$W2$ 的偏自相关函数 1—2 阶都很显著，并且从第 3 阶开始下降很大，因此我们先设定 p 的值为 2，于是对于序列 $W2$，我们初步建立了 ARMA(2，5)模型。



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.701	-0.701	83.437	0.000
		2	0.141	-0.688	86.814	0.000
		3	0.277	-0.143	100.04	0.000
		4	-0.453	-0.295	135.53	0.000
		5	0.371	-0.224	159.50	0.000
		6	-0.135	-0.156	162.69	0.000
		7	-0.085	-0.095	163.97	0.000
		8	0.143	-0.304	167.59	0.000
		9	-0.010	0.052	167.61	0.000
		10	-0.142	0.029	171.24	0.000
		11	0.143	-0.121	174.96	0.000
		12	-0.011	-0.015	174.98	0.000
		13	-0.115	0.122	177.42	0.000
		14	0.138	-0.050	180.96	0.000
		15	-0.075	-0.031	182.01	0.000
		16	-0.014	0.054	182.05	0.000
		17	0.056	-0.130	182.63	0.000
		18	-0.023	-0.097	182.72	0.000
		19	-0.030	0.057	182.89	0.000
		20	0.028	-0.152	183.04	0.000
		21	0.042	-0.082	183.38	0.000
		22	-0.140	-0.104	187.19	0.000
		23	0.208	0.042	195.66	0.000
		24	-0.154	0.070	200.35	0.000
		25	-0.019	0.048	200.42	0.000
		26	0.154	-0.103	205.20	0.000
		27	-0.175	-0.103	211.41	0.000
		28	0.117	-0.019	214.17	0.000
		29	-0.002	0.062	214.17	0.000
		30	-0.099	-0.032	216.20	0.000
		31	0.131	0.085	219.77	0.000
		32	-0.081	0.067	221.16	0.000
		33	-0.028	-0.015	221.32	0.000
		34	0.075	-0.118	222.50	0.000
		35	-0.012	0.026	222.53	0.000
		36	-0.055	0.008	223.18	0.000

图 5－2　W2 自相关函数图和偏自相关函数图

（二）模型的估计

点击“Quick”—“Estimate Equation”，会弹出如图 5－3 所示的窗口，在“Equation Specification”空白栏中键入“ W2 C MA(1) MA(2) MA(3) MA(4) MA(5) AR(1) AR(2)”，在“Estimation Settings”中选择“LS－Least Squares(NLS and ARMA)”，然后点击“OK”，得到如表 5－1 所示的估计结果。

可以看到，除常数项外，其他解释变量的系数估计值在 15%的显著性水平下都是显著的。

（三）模型的诊断

点击“View”-“Residual test”-“Correlogram－Q－statistics”，在弹出的窗口中选择滞后阶数为 36，点击“OK”，就可以得到 Q 统计量，此时为 30.96，p 值为 0.367，因此不能拒绝原假设，可以认为模型较好地拟合了数据。

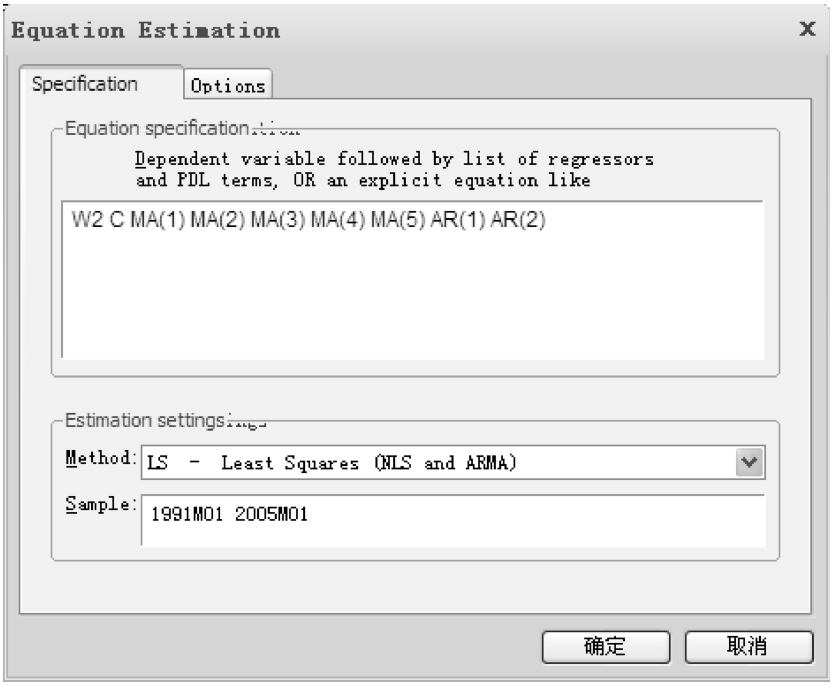


图 5－3　回归方程设定

表 5－1　ARMA(2, 5)回归结果

Dependent Variable: W2

Method: Least Squares

Date: 12/21/14　　Time: 02:05

Sample (adjusted): 1991M05 2005M01

Included observations: 165 after adjustments

Convergence achieved after 50 iterations

MA Backcast: 1990M12 1991M04

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.44319	10.74065	1.344723	0.1807
AR(1)	-0.995579	0.055305	-18.00153	0.0000
AR(2)	-0.837713	0.047357	-17.68914	0.0000
MA(1)	-0.436708	0.096208	-4.539223	0.0000
MA(2)	0.175063	0.104359	1.677513	0.0954
MA(3)	-0.880075	0.052403	-16.79446	0.0000
MA(4)	0.322618	0.100005	3.226012	0.0015
MA(5)	0.190454	0.096508	1.973453	0.0502

（续表）

R-squared	0.805361	Mean dependent var	16.34363
Adjusted R-squared	0.796682	S.D. dependent var	2309.544
S.E. of regression	1041.391	Akaike info criterion	16.78177
Sum squared resid	1.70E+08	Schwarz criterion	16.93236
Log likelihood	-1376.496	Hannan-Quinn criter.	16.84290
F-statistic	92.80278	Durbin-Watson stat	2.041303
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.50-.77i	-.50+.77i		
Inverted MA Roots	.77+.22i	.77-.22i	-.30	-.40-.91i
	-.40+.91i			

我们再来看是否存在一个更好的模型。我们的做法是增加模型的滞后长度，然后根据信息值来判断。表 5－2 是我们试验的几个 p、q 值的 AIC 信息值。

表 5－2　不同 p、q 值的 AIC 信息值

p	2	3	4	2	2	2	3	3	3	4	4	4
q	5	5	5	6	7	8	6	7	8	6	7	8
AIC	16.78	16.75	16.77	16.76	16.76	16.77	16.77	16.78	16.79	16.75	16.79	16.78

可以看到，根据 AIC 信息值，我们应选择 $p=3$，$q=5$，或 $p=4$，$q=6$，但是按照后者建立的模型中有的解释变量的系数估计值是不显著的，而按照前者建立的模型其解释变量的系数值都是显著的（如表 5－3 所示），因此我们最终建立的模型是 ARMA(3,5)。

（四）模型的预测

点击“Forecast”，会弹出如图 5－4 所示的窗口。在 EViews 中有两种预测方式：“Dynamic”和“Static”，前者是根据所选择的一定的估计区间，进行多步向前预测；后者是只滚动的进行向前一步预测，即每预测一次，用真实值代替预测值，加入估计区间，再进行向前一步预测。我们首先用前者来估计 2003 年 1 月—2005 年 1 月的 $W2$，在“Sample range for forecast”空白栏中键入“2003：01 2005：01”（如图 5－4 所示），选择“Dynamic”，其他的一些选项诸如预测序列的名称以及输出结果的形式等，我们可以根据目的自行选择，不再介绍，点击“OK”，得到如图 5－5 所示的预测结果。

表 5－3 ARMA(3,5)回归结果

Dependent Variable: W2

Method: Least Squares

Date: 12/21/14 Time: 02:28

Sample (adjusted): 1991M06 2005M01

Included observations: 164 after adjustments

Convergence achieved after 38 iterations

MA Backcast: 1991M01 1991M05

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.49885	9.044380	1.603079	0.1110
AR(1)	-1.695549	0.115673	-14.65814	0.0000
AR(2)	-1.509088	0.142387	-10.59848	0.0000
AR(3)	-0.595019	0.110749	-5.372669	0.0000
MA(1)	0.291050	0.115673	2.516134	0.0129
MA(2)	-0.213353	0.073117	-2.917948	0.0040
MA(3)	-0.634653	0.050426	-12.58573	0.0000
MA(4)	-0.392415	0.101135	-3.880121	0.0002
MA(5)	0.484989	0.079475	6.102419	0.0000

R-squared	0.815177	Mean dependent var	16.18107
Adjusted R-squared	0.805638	S.D. dependent var	2316.617
S.E. of regression	1021.315	Akaike info criterion	16.74888
Sum squared resid	1.62E+08	Schwarz criterion	16.91900
Log likelihood	-1364.409	Hannan-Quinn criter.	16.81794
F-statistic	85.45525	Durbin-Watson stat	2.024308
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.48+.76i	-.48-.76i	-.74	
Inverted MA Roots	.72-.14i	.72+.14i	-.40-.91i	-.40+.91i
	-.93			

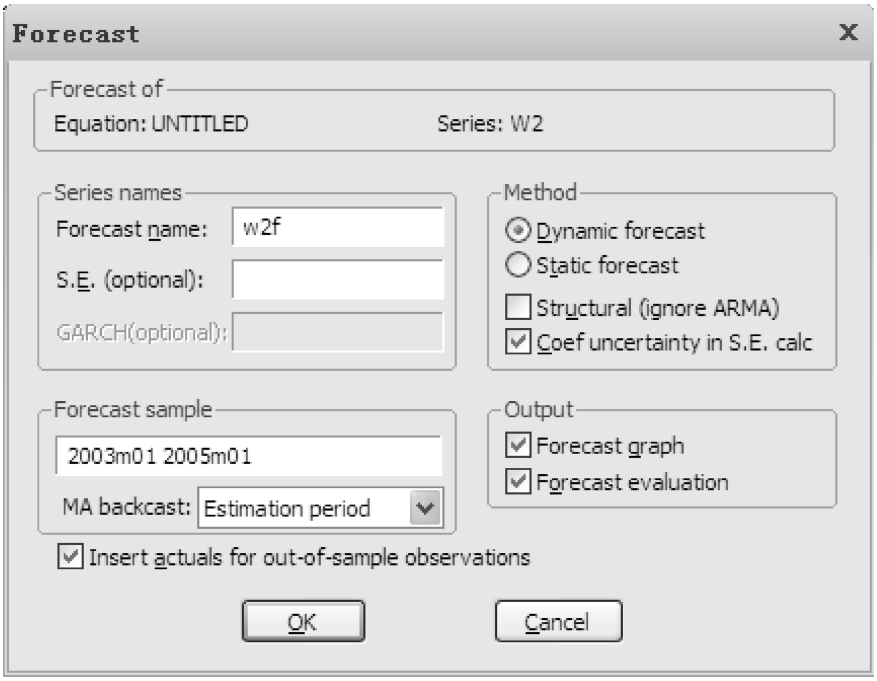


图 5－4　ARMA(3,5)模型预测设定

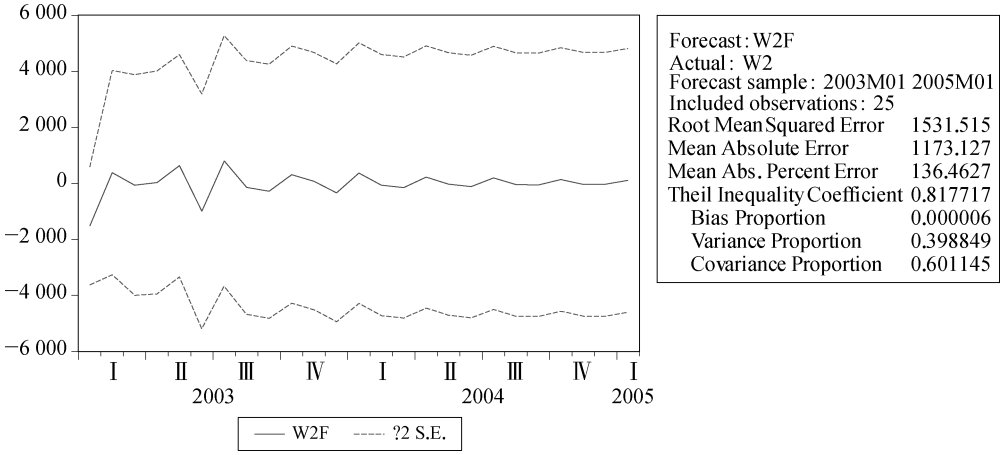
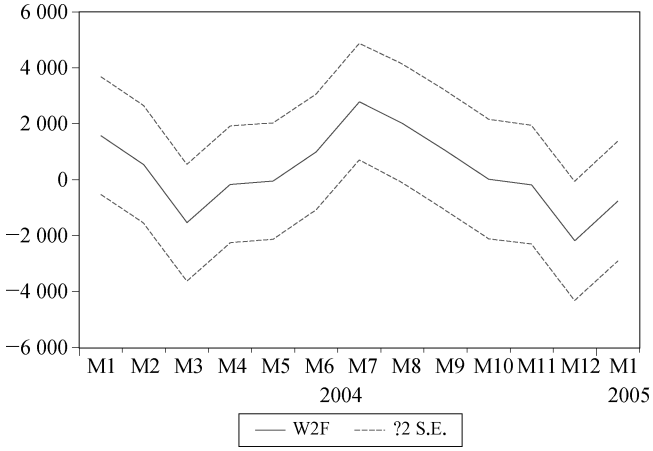


图 5－5　Dynamic 预测方式结果

图 5－5 中，实线代表的是 W2 的预测值，两条虚线则提供了 2 倍标准差的置信区间。可以看到，正如我们在前面所讲的，随着预测时间的增长，预测值很快趋向于序列的均值(接近 0)。图的右边列出的是评价预测的一些标准，如平均预测误差平方和的平方根(RMSE)，Theil 不相等系数及其分解。可以看到，Theil 不相等系数为 0.82，表明模型的预测能力不太好，而对它的分解表明偏误比例很小，方差比例较大，说明实际序列的波动较大，而模拟序列的波动较小，这可能是由于预测时间过长。

下面我们再利用“Static”方法来估计 2004 年 1 月—2005 年 1 月的 W2(操作过程略),我们可以得到如图 5－6 所示的结果。从图中可以看到,“Static”方法得到的预测值波动性要大;同时,方差比例的下降也表明较好地模拟了实际序列的波动,Theil 不相等系数为 0.62,其中协方差比例为 0.70,表明模型的预测结果较理想。



Forecast: W2F	
Actual: W2	
Forecast sample: 2004M01 2005M01	
Included observations: 13	
Root Mean Squared Error	1280.967
Mean Absolute Error	1061.574
Mean Abs. Percent Error	872.9085
Theil Inequality Coefficient	0.618320
Bias Proportion	0.045897
Variance Proportion	0.246135
Covariance Proportion	0.707968

图 5－6　Static 预测方式结果

实验六　VAR 模型的概念和构造

一、实验目的

理解 VAR 模型的概念，掌握 VAR 模型的形式和特点，掌握 VAR 模型的识别、估计、检验和预测，了解似然比检验法，掌握脉冲响应的作用和应用，掌握使用 EViews 软件进行相关的检验。

二、基本概念

VAR 模型即向量自回归模型由希姆斯(C. A. Sims)提出，在一个含有 n 个方程(被解释变量)的 VAR 模型中，每个被解释变量都对自身以及其他被解释变量的若干期滞后值回归，若令滞后阶数为 k，则 VAR 模型的一般形式可用下式表示：

$$Z_t = \sum_{i=1}^{k} A_i Z_{t-i} + V_t$$

其中，Z_t 表示由第 t 期观测值构成的 n 维列向量，A_i 为 $n * n$ 系数矩阵，V_t 是由随机误差项构成的 n 维列向量，其中随机误差项 $V_i(i=1, 2, \cdots, n)$ 为白噪音过程，且满足 $E(v_{it}v_{jt})(i, j=1, 2, \cdots, n$，且 $i \neq j)$。

对某变量全部滞后项系数的联合检验能够告诉我们该变量是否对被解释变量有显著的影响，但是不能告诉我们这种影响是正还是负，也不能告诉我们这种影响发生作用所需要的时间。为解决这一问题，经常应用的方法是测量脉冲响应。脉冲响应度量的是被解释变量对单位冲击的响应。

三、实验内容及要求

（一）实验内容

在 EViews 软件中利用 VAR 模型对我国货币政策的有效性进行检验。取我国狭义货币供应量 M1，商品零售物价指数 P，以及代表产出水平的国内生产总值 GDP 的季度数据，时间为 1994 年第一季度到 2004 年第二季度。所有的数据我们都取它们的增长率，以保证序列的平稳性。

（二）实验要求

(1) 深刻理解 VAR 模型的基本概念，以及脉冲响应的基本概念；

(2) 思考如何建立适当的 VAR 模型，如何利用 VAR 模型进行预测；

(3) 熟练掌握相关 EViews 操作。

四、实验指导

（一）导入数据

打开 EViews 软件，点击“File”—“New - Workfile”选项，出现“Workfile Range”对话框，在“Workfile frequency”框中选择“Quarterly”，在“Start date”和“End date”框中分别输入“1994：1”和“2004：2”，然后单击“OK”。点击“File”—“Import - Read Text - Lotus - Excel”，找到要导入的名为 EX6. 3. xls 的 Excel 文档，单击“打开”出现“Excel Spreadsheet Import”对话框并在其中输入“CPI”“GDP”和“M1”，单击“OK”，完成数据导入。

为保证序列的平稳性，所有的数据我们都取它们的增长率。在命令框输入命令 genr CPIDL＝DLOG(CPI)，生成 CPI 增长率序列 CPIDL。同样的方法生成 GDPDL、M1DL 序列。

（二）建立模型

点击“Quick”-“Estimate VAR”，弹出如图 6 - 1 所示的窗口。在左边“VAR Specification”中我们选择“Unrestricted VAR”，滞后长度我们从一阶试起，在右边“Endogenous”空白栏中我们键入变量名称“cpidl m1dl gdpdl”，在“Exdogenous”空白栏中保留常数项“C”，点击“OK”，即可以得到估计结果。

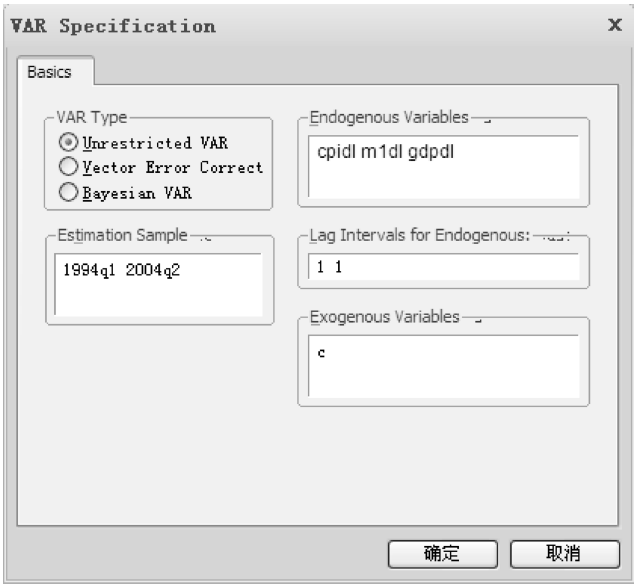


图 6 - 1　VAR 模型设定

在选择滞后项时，我们应用信息准则，表 6 - 1 中是我们试验的几个滞后项（根据金融理论，货币效应时滞在一年左右，所以我们选择最大 5 阶）及相应的信息值。

表 6－1　不同滞后值下的 AIC 值和 SC 值

滞后值	1	2	3	4	5
AIC 值	－10.32	－11.53	－12.45	－12.73	－13.20
SC 值	－9.81	－10.64	－11.16	－11.03	－11.09

由表 6－1，根据 AIC 信息准则，我们应选择滞后项为 5，根据 SC 信息准则，我们应选择滞后项为 3，考虑到 3 阶后 AIC 值下降较缓，因此我们根据 SC 值选择滞后项为 3，然后进行估计，得到如表 6－2 所示的结果。

表 6－2　VAR 模型估计结果

Vector Autoregression Estimates
Date: 02/24/15　Time: 21:34
Sample (adjusted): 1995Q1 2004Q2
Included observations: 38 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	CPIDL	M1DL	GDPDL
CPIDL(-1)	0.231545	-0.454025	0.331327
	(0.19939)	(0.36350)	(0.54581)
	[1.16128]	[-1.24902]	[0.60704]
CPIDL(-2)	0.076421	-0.969225	0.476782
	(0.16326)	(0.29764)	(0.44692)
	[0.46809]	[-3.25635]	[1.06682]
CPIDL(-3)	0.079005	0.154853	0.697243
	(0.18940)	(0.34529)	(0.51846)
	[0.41714]	[0.44847]	[1.34484]
M1DL(-1)	0.203608	-0.460036	0.171646
	(0.11549)	(0.21056)	(0.31616)
	[1.76295]	[-2.18485]	[0.54291]
M1DL(-2)	-0.289403	-0.125967	-0.049720
	(0.10444)	(0.19041)	(0.28590)
	[-2.77099]	[-0.66157]	[-0.17391]

（续表）

M1DL(-3)	-0.003083	-0.173310	-0.023864
	(0.12024)	(0.21921)	(0.32914)
	[-0.02564]	[-0.79063]	[-0.07250]
GDPDL(-1)	0.092482	-0.002143	-1.110726
	(0.04169)	(0.07600)	(0.11412)
	[2.21838]	[-0.02820]	[-9.73287]
GDPDL(-2)	0.114096	-0.142067	-0.792235
	(0.04670)	(0.08514)	(0.12784)
	[2.44308]	[-1.66858]	[-6.19690]
GDPDL(-3)	0.066997	-0.053859	-0.439512
	(0.03172)	(0.05782)	(0.08682)
	[2.11235]	[-0.93144]	[-5.06215]
C	-0.002466	0.071970	0.071507
	(0.00991)	(0.01807)	(0.02713)
	[-0.24885]	[3.98290]	[2.63549]
R-squared	0.638976	0.607271	0.848922
Adj. R-squared	0.522933	0.481036	0.800361
Sum sq. resids	0.007643	0.025402	0.057272
S.E. equation	0.016521	0.030120	0.045227
F-statistic	5.506358	4.810661	17.48161
Log likelihood	107.8004	84.97973	69.53331
Akaike AIC	-5.147388	-3.946302	-3.133332
Schwarz SC	-4.716444	-3.515358	-2.702388
Mean dependent	0.003255	0.033820	0.014668
S.D. dependent	0.023920	0.041811	0.101221
Determinant resid covariance (dof adj.)		4.06E-10	
Determinant resid covariance		1.62E-10	
Log likelihood		266.5314	
Akaike information criterion		-12.44902	
Schwarz criterion		-11.15619	

在表 6－2 中我们也可以看到，在同一变量不同的滞后项，有的是显著的，有的是不显著的，有的符号是相反的，验证了我们所说的 VAR 模型是缺乏理论依据的，我们无法直接得出某种结论。

首先，对于物价 CPI，上期的货币供应量对其的影响是显著的，并且系数为正，与理论相符，说明货币供应量的增加将使物价水平上升。

其次，对于货币供应量来说，GDP 和物价水平对其影响不显著，说明货币供应量不受上期的产出和物价水平的影响，是一个独立的外生变量，由央行控制，不受实体经济要素的影响。

再次，对于 GDP，上期的货币供应量对其没有影响，这也施的稳健的货币政策效果是不大的，而上期物价水平则对产出有显著的正的影响。

（三）检验脉冲响应

在 EViews 软件点击“Impulse”菜单，就会弹出如图 6－2 所示的窗口。

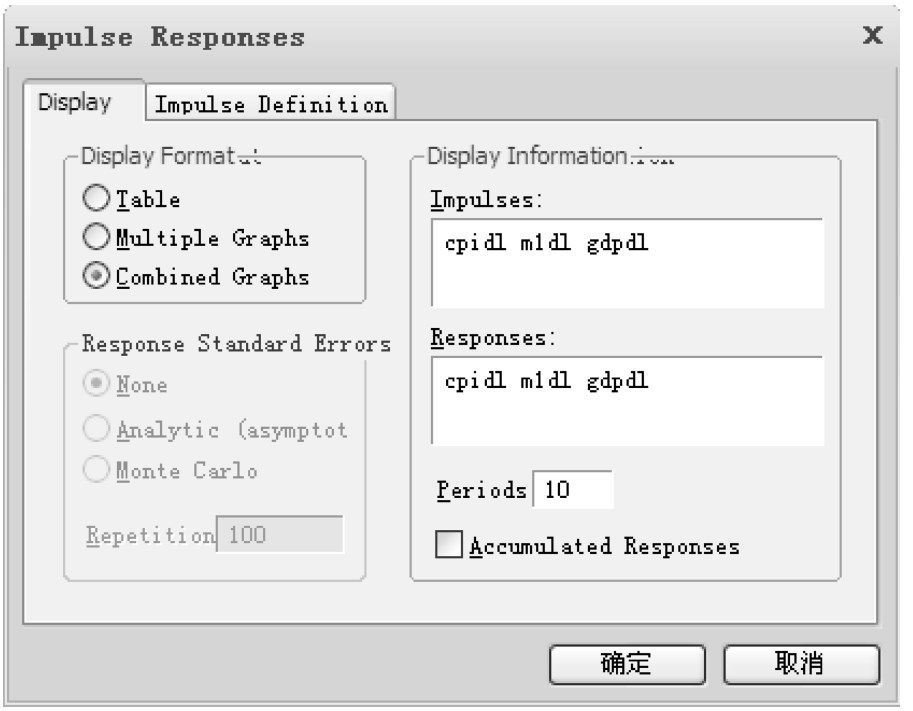


图 6－2　VAR 脉冲响应设定

我们选择对三个变量都进行脉冲响应测试，冲击也分别来自三个变量，然后选择时期为 10(其他各项可根据需要选择)，点击“OK”，得到如图 6－3 所示的脉冲测试结果。

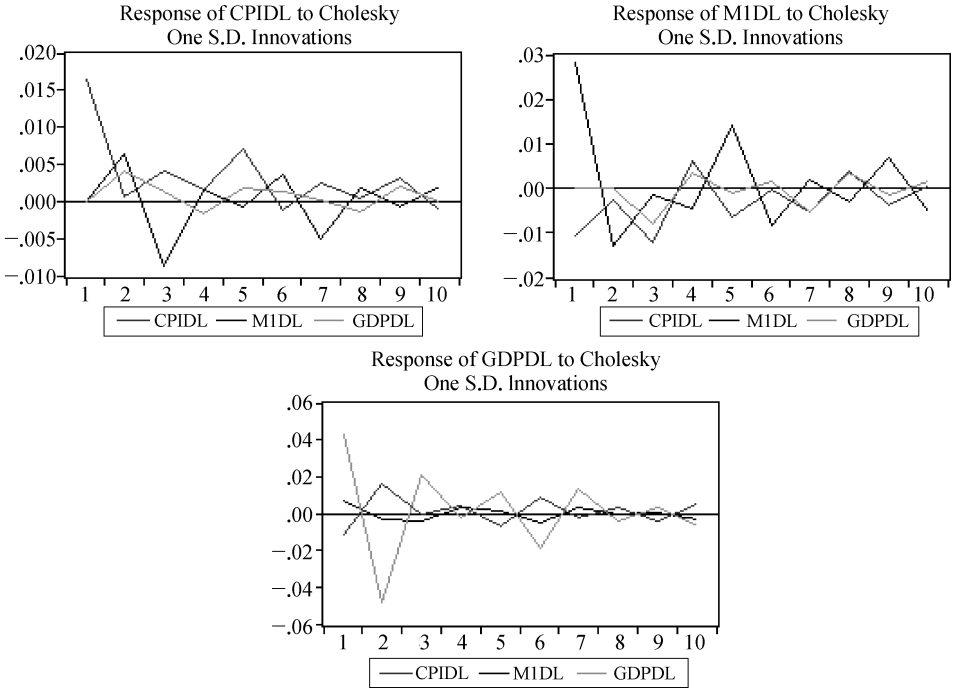


图 6－3　VAR 脉冲响应结果

实验七　(G)ARCH 模型在金融数据中的应用

一、实验目的

理解自回归异方差(ARCH)模型的概念及建立的必要性和适用的场合。

了解(G)ARCH 模型的各种不同类型,如 GARCH－M 模型(GARCH in Mean),EGARCH 模型(Exponential GARCH)和 TARCH 模型(Threshold ARCH)。掌握对(G)ARCH 模型的识别、估计及如何运用 EViews 软件在实证研究中实现。

二、基本概念

p 阶自回归条件异方程 ARCH(p)模型,其定义由均值方程(7.1)和条件方差方程(7.2)给出:

$$y_t = \beta x_t + \varepsilon_t \tag{7.1}$$

$$h_t = \mathrm{var}(\varepsilon_t \mid \Omega_{t-1}) = a_0 + a_1\varepsilon_{t-1}^2 + a_2\varepsilon_{t-2}^2 + \cdots + a_p\varepsilon_{t-p}^2 \tag{7.2}$$

其中,Ω_{t-1} 表示 $t-1$ 时刻所有可得信息的集合,h_t 为条件方差。方程(7.2)表示误差项 ε_t 的方差 h_t 由两部分组成:一个常数项和前 p 个时刻关于变化量的信息,用前 p 个时刻的残差平方表示(ARCH 项)。

广义自回归条件异方差 GARCH(p, q)模型可表示为:

$$y_t = \beta x_t + \varepsilon_t \tag{7.3}$$

$$h_t = \mathrm{var}(\varepsilon_t \mid \Omega_{t-1}) = a_0 + a_1\varepsilon_{t-1}^2 + \cdots + a_p\varepsilon_{t-p}^2 + \lambda_1 h_{t-1} + \cdots + \lambda_q h_{t-q} \tag{7.4}$$

三、实验内容及要求

(一) 实验内容

以上证指数和深证成分指数为研究对象,选取 1997 年 1 月 2 日—2002 年 12 月 31 日共 6 年每个交易日上证指数和深证成份指数的收盘价为样本,完成以下实验步骤:

(1) 沪深股市收益率的波动性研究;

(2) 股市收益波动非对称性的研究;

(3) 沪深股市波动溢出效应的研究。

(二) 实验要求

(1) 深刻理解本章的概念;

(2) 对实验步骤中提出的问题进行思考;

(3) 熟练掌握实验的操作步骤,并得到有关结果。

四、实验指导

(一) 沪深股市收益率的波动性研究

1. 描述性统计

(1) 导入数据,建立工作组。打开 EViews 软件,选择"File"菜单中的"New Workfile"选项,在"Workfile frequency"框中选择"undated or irregular",在"Start observation"和"End observation"框中分别输入"1"和"1444",单击"OK"。选择"File"菜单中的"Import - Read Text - Lotus - Excel"选项,找到要导入的名为 EX6. 4. xls 的 Excel 文档完成数据导入。

(2) 生成收益率的数据列。在 EViews 窗口主菜单栏下的命令窗口中键入如下命令:genr rh=log(sh/sh(−1)),回车后即形成沪市收益率的数据序列 *rh*,同样的方法可得深市收益数剧序列 *rz*。

(3) 观察收益率的描述性统计量。双击选取"rh"数据序列,在新出现的窗口中点击"View"-"Descriptive Statistics"-"Histogram and Stats",则可得沪市收益率 *rh* 的描述性统计量,如图 7 - 1 所示。

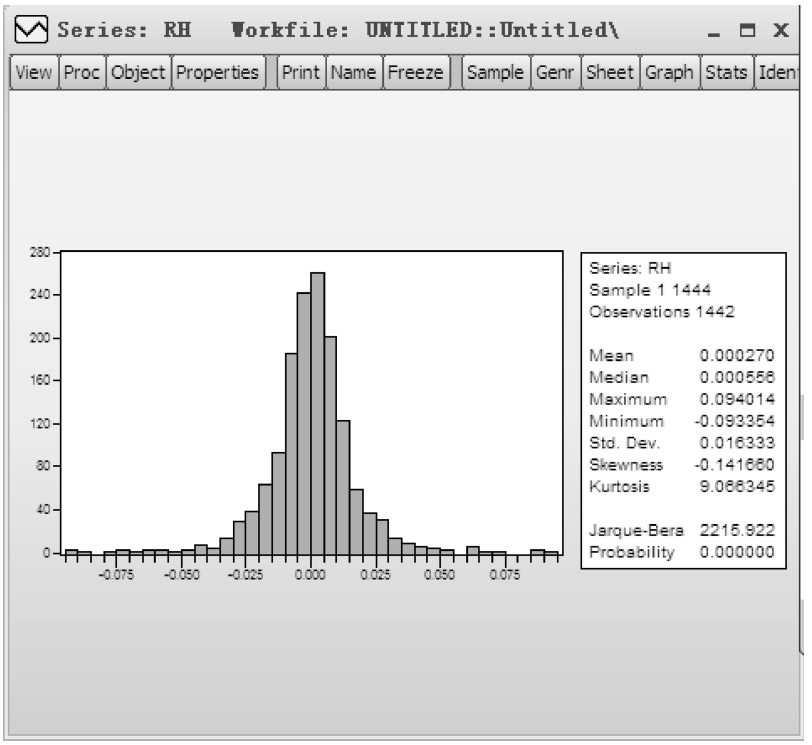


图 7 - 1　沪市收益率 *rh* 的描述性统计量

同样的步骤可得深市收益率 *rz* 的描述性统计量。观察这些数据，我们可以发现：样本期内沪市收益率均值为 0.027%，标准差为 1.63%，偏度为－0.146，左偏峰度为 9.07，远高于正态分布的峰度值 3，说明收益率 *rh* 具有尖峰和厚尾特征。JB 正态性检验也证实了这点，统计量为 2 216，说明在极小水平下，收益率 *rh* 显著异于正态分布；深市收益率均值为－0.012%，标准差为 1.80%，偏度为－0.027，左偏峰度为 8.172，收益率 *rz* 同样具有尖峰厚尾特征。深市收益率的标准差大于沪市，说明深圳股市的波动更大。

2. 平稳性检验

再次双击选取 *rh* 序列，点击“View”—“Unit Root Test”，出现如图 7－2 所示的窗口。

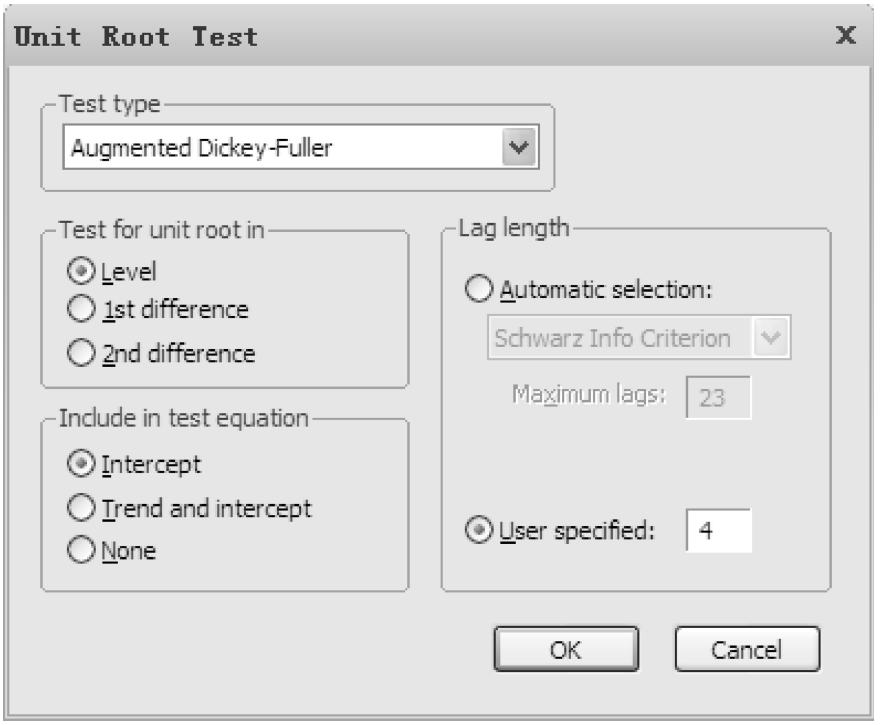


图 7－2　单位根检验

对该序列进行 ADF 单位根检验，选择滞后 4 阶，带截距项而无趋势项，所以采用窗口的默认选项，得到如图 7－3 所示的结果。

同样对 *rz* 做单位根检验后，得到如图 7－4 所示的结果。

在 1%的显著水平下，两市的收益率都拒绝随机游走的假设，说明是平稳的时间序列数据。这个结果与国外学者对发达成熟市场波动性的研究一致。Pagan(1996)和 Bollerslev(1994)指出：金融资产的价格一般是非平稳的，经常有一个单位根(随机游走)，而收益率序列通常是平稳的。

Series: RH Workfile: UNTITLED::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ident

Augmented Dickey-Fuller Unit Root Test on RH

Null Hypothesis: RH has a unit root
Exogenous: Constant
Lag Length: 4 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-16.38279	0.0000
Test critical values:	1% level	-3.434692	
	5% level	-2.863345	
	10% level	-2.567780	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RH)
Method: Least Squares
Date: 02/24/15 Time: 22:25
Sample (adjusted): 7 1443
Included observations: 1437 after adjustments

图 7－3 *rh* ADF 检验结果

Series: RZ Workfile: UNTITLED::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ident

Augmented Dickey-Fuller Unit Root Test on RZ

Null Hypothesis: RZ has a unit root
Exogenous: Constant
Lag Length: 4 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-15.95064	0.0000
Test critical values:	1% level	-3.434692	
	5% level	-2.863345	
	10% level	-2.567780	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RZ)
Method: Least Squares
Date: 02/24/15 Time: 22:26
Sample (adjusted): 7 1443
Included observations: 1437 after adjustments

图 7－4 *rz* ADF 检验结果

3. 均值方程的确定及残差序列自相关检验

通过对收益率的自相关检验，我们发现两市的收益率都与其滞后 15 阶存在显著的自相关，因此对两市收益率 r_t 的均值方程都采用如下形式：

$$r_t = c + ar_{t-15} + \varepsilon_t \tag{7.5}$$

（1）对收益率做自回归。在 EViews 主菜单中选择“ Quick”—“ Estimation Equation”，出现如图 7－5 所示的窗口。

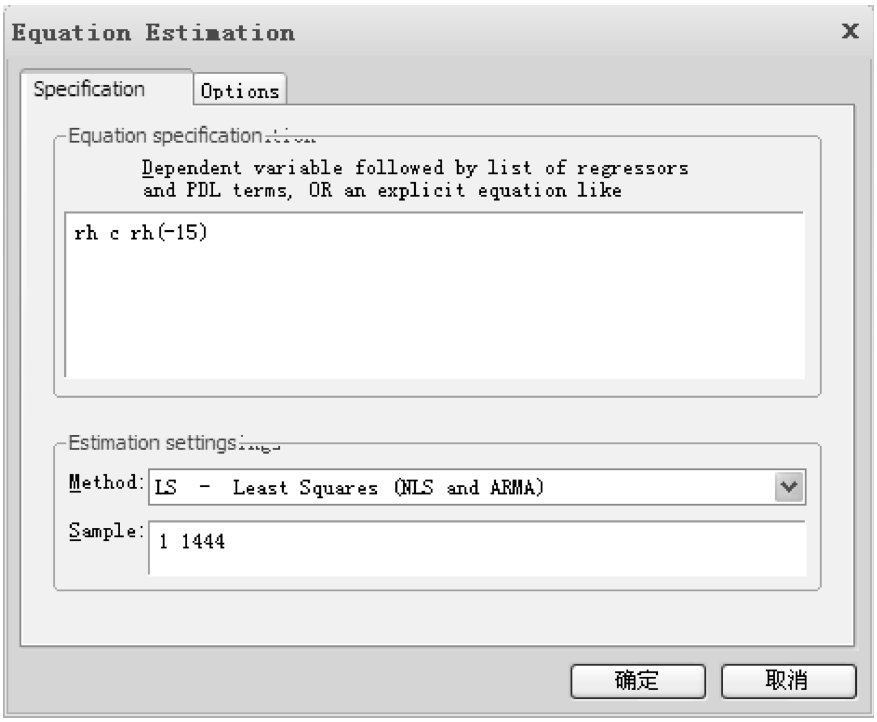


图 7－5　对收益率 *rh* 做自回归

在“Method”中选择 LS(即普通最小二乘法)，然后在“Estimation settings”上方空白处输入图 7－5 所示变量，单击“OK”，则出现图 7－6 所示的结果。

（2）用 Ljung－Box Q 统计量对均值方程拟和后的残差及残差平方做自相关检验。点击“View”—“Residual Test”—“Correlogram－Q－statistics”，选择 10 阶滞后，则可得沪市收益率 *rh* 残差项的自相关系数 *acf* 值和 *pacf* 值，如图 7－7 所示。

点击“View”—“Residual Test”—“Correlogram Squared Residuals”，选择 10 阶滞后，则可得沪市收益率 *rh* 残差平方的自相关系数 *acf* 值和 *pacf* 值，如图 7－8 所示。

采用同样的方法，可得深市收益率 *rz* 的回归方程及残差、残差平方的 *acf* 值和 *pacf* 值。结果表明两市的残差不存在显著的自相关，而残差平方有显著的自相关。

Equation: UNTITLED　Workfile: UNTITLED::U...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RH
Method: Least Squares
Date: 02/24/15 Time: 22:30
Sample (adjusted): 17 1443
Included observations: 1427 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000226	0.000431	0.524763	0.5998
RH(-15)	0.094697	0.026301	3.600507	0.0003

R-squared	0.009015	Mean dependent var	0.000253
Adjusted R-squared	0.008320	S.D. dependent var	0.016348
S.E. of regression	0.016279	Akaike info criterion	-5.396425
Sum squared resid	0.377654	Schwarz criterion	-5.389048
Log likelihood	3852.349	Hannan-Quinn criter.	-5.393670
F-statistic	12.96365	Durbin-Watson stat	2.024254
Prob(F-statistic)	0.000328		

图 7－6　收益率 *rh* 回归结果

Equation: UNTITLED　Workfile: UNTITLED::U...

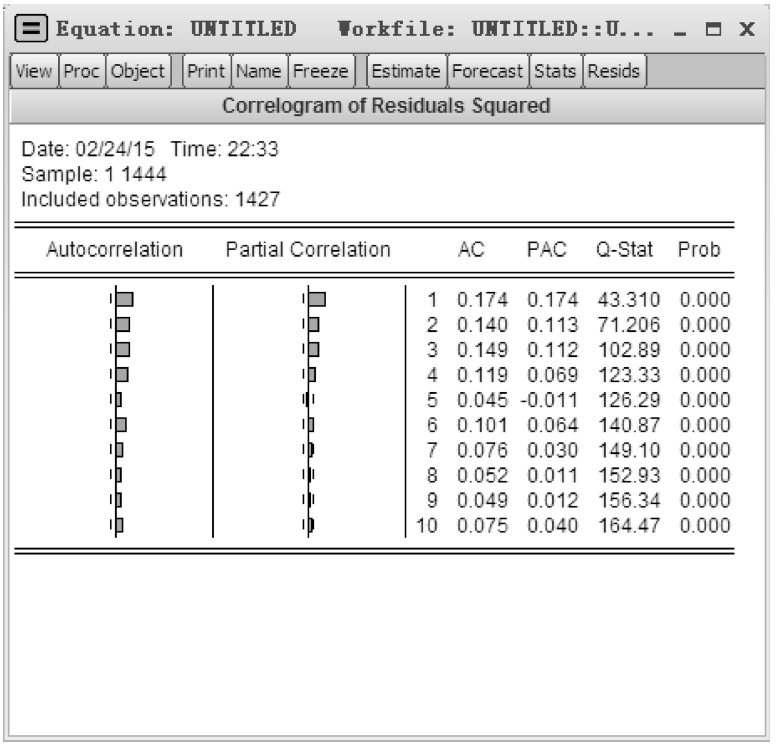
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Correlogram of Residuals

Date: 02/24/15 Time: 22:32
Sample: 1 1444
Included observations: 1427
Q-statistic probabilities adjusted for 1 dynamic regressor

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.012	-0.012	0.2167	0.642
		2	-0.034	-0.034	1.8888	0.389
		3	0.011	0.011	2.0766	0.557
		4	0.058	0.057	6.8801	0.142
		5	-0.019	-0.017	7.3833	0.194
		6	0.006	0.009	7.4284	0.283
		7	0.025	0.023	8.3585	0.302
		8	-0.038	-0.040	10.461	0.234
		9	-0.047	-0.045	13.664	0.135
		10	0.012	0.007	13.880	0.179

图 7－7　沪市收益率 *rh* 残差项的自相关系数 *acf* 值和 *pacf* 值



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.174	0.174	43.310	0.000
		2	0.140	0.113	71.206	0.000
		3	0.149	0.112	102.89	0.000
		4	0.119	0.069	123.33	0.000
		5	0.045	-0.011	126.29	0.000
		6	0.101	0.064	140.87	0.000
		7	0.076	0.030	149.10	0.000
		8	0.052	0.011	152.93	0.000
		9	0.049	0.012	156.34	0.000
		10	0.075	0.040	164.47	0.000

图 7－8　沪市收益率 *rh* 残差平方的自相关系数 *acf* 值和 *pacf* 值

(3) 对残差平方做线性图。对 *rh* 进行回归后在命令栏输入命令：genr res1＝resid^2，得到 *rh* 残差平方序列 *res*1，用同样的方法得到 *rz* 残差平方序列 *res*2。双击选取序列 *res*1，在新出现的窗口中选择“View”—“Line Graph”，得到 *res*1 的线性图如图 7－9 所示。

同理得到 *rz* 残差平方线状图，如图 7－10 所示。

可见 ε_t^2 的波动具有明显的时间可变性和波动丛簇性，适合用 GARCH 类模型来建模。

(4) 对残差进行 ARCH－LM Test。依照步骤(1)，再对 *rh* 做一次滞后 15 阶的回归，在出现的“Equation”窗口中点击“View”—“Residual Test”—“ARCH LM Test”，选择一阶滞后，得到如图 7－11 所示的结果。

对 *rz* 方程回归后的残差项同样可做 ARCH－LM Test，结果表明残差中 ARCH 效应是很显著的。

4. GARCH 类模型建模

(1) GARCH(1,1)模型估计结果。点击“Quick”—“Estimate Equation”，在出现的窗口中“Method”选项选择“ARCH”，可以得到如图 7－12 所示的对话框。

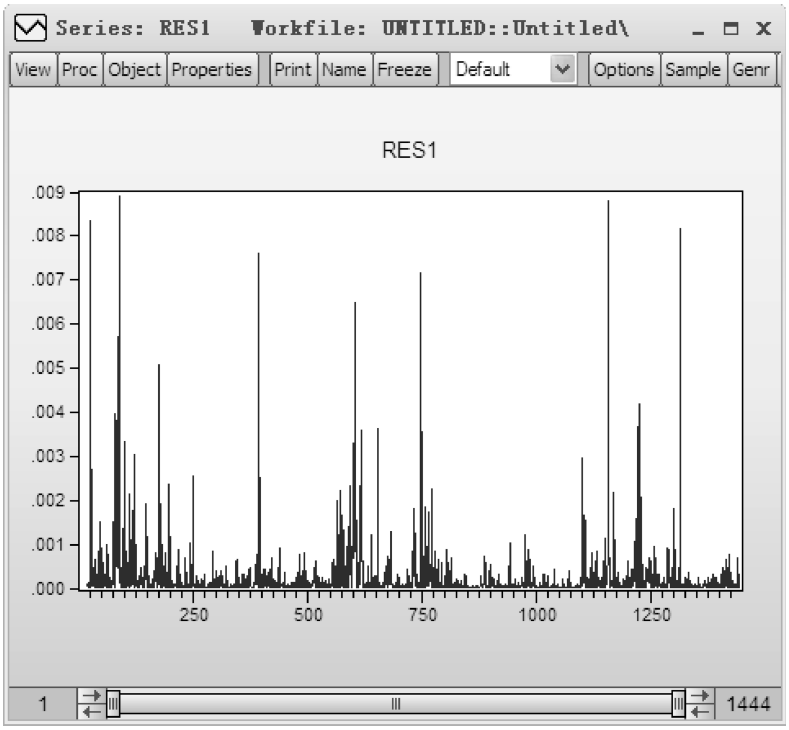


图 7－9　*rh* 残差平方线状图

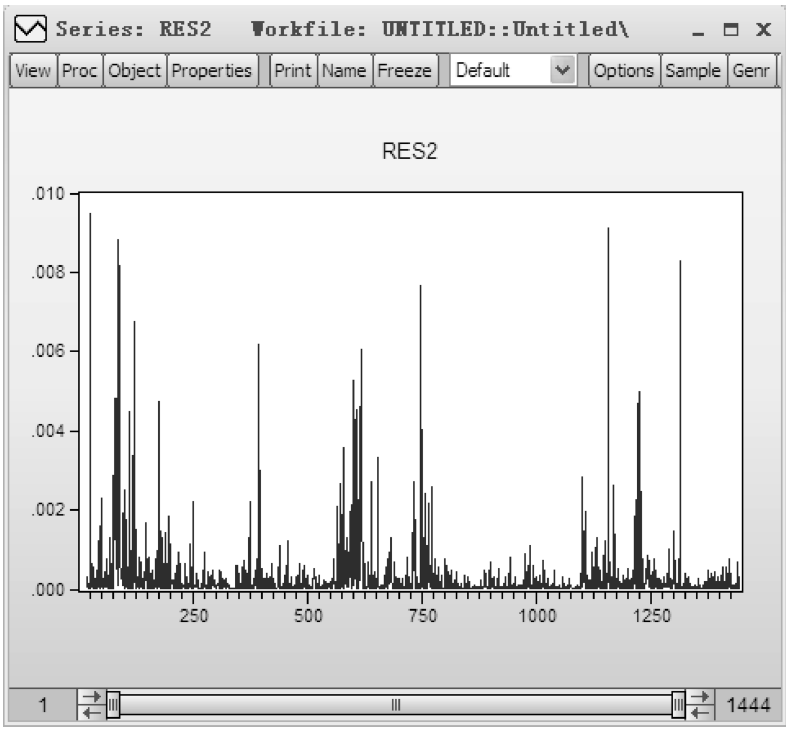


图 7－10　*rz* 残差平方线状图

Equation: UNTITLED　Workfile: UNTITLED::U...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: ARCH

F-statistic	44.48012	Prob. F(1,1424)	0.0000
Obs*R-squared	43.19340	Prob. Chi-Square(1)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 02/24/15　Time: 22:47
Sample (adjusted): 18 1443
Included observations: 1426 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000219	2.09E-05	10.48264	0.0000
RESID^2(-1)	0.174043	0.026096	6.669342	0.0000

R-squared	0.030290	Mean dependent var	0.000265
Adjusted R-squared	0.029609	S.D. dependent var	0.000755
S.E. of regression	0.000743	Akaike info criterion	-11.56967
Sum squared resid	0.000787	Schwarz criterion	-11.56229
Log likelihood	8251.178	Hannan-Quinn criter.	-11.56692
F-statistic	44.48012	Durbin-Watson stat	2.039221
Prob(F-statistic)	0.000000		

图 7－11　*rh* ARCH－LM Test

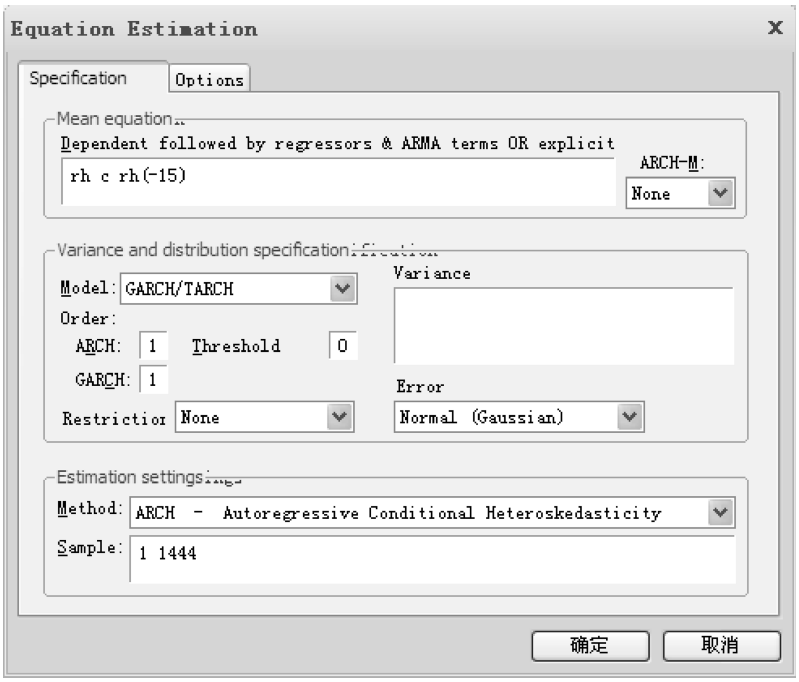


图 7－12　Equation Specification 窗口

在这个对话框中要求用户输入建立 GARCH 类模型相关的参数："Mean Equation Specification"栏需要填入均值方差的形式；"ARCH－M term"栏需要选择 ARCH－M 项的形式，包括方差、标准差和不采用三种；"ARCH Specification"栏需要选择 ARCH 和 GARCH 项的阶数，以及估计方法包括 GARCH、TARCH 和 EGARCH 等等；"Variance Regressors"栏需要填如结构方差的形式，由于 EViews 默认条件方差方程中包含常数项，因此在此栏中不必要填入"C"。我们现在要用 GARCH(1,1)模型建模，以沪市为例，只需要在"Mean Equation Specification"栏输入均值方差"RH C RH(－15)"，其他选择默认即可，得到如图 7－13 和图 7－14 所示的结果。

可见，沪深股市收益率条件方差方程中 ARCH 项和 GARCH 项都是高度显著的，表明收益率序列具有显著的波动丛簇性。沪市中 ARCH 项和 GARCH 项系数之和为 0.98，深市也为 0.98，均小于 1。因此，GARCH(1,1)过程是平稳的，其条件方差表现出均值回复(mean-reversion)，即过去的波动对未来的影响是逐渐衰减。

Equation: UNTITLED　Workfile: UNTITLED::Unt...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RH
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 02/24/15　Time: 22:50
Sample (adjusted): 17 1443
Included observations: 1427 after adjustments
Convergence achieved after 18 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.23E-06	0.000283	-0.007891	0.9937
RH(-15)	0.059355	0.020212	2.936626	0.0033
Variance Equation				
C	8.72E-06	1.45E-06	6.009939	0.0000
RESID(-1)^2	0.176309	0.015220	11.58366	0.0000
GARCH(-1)	0.807551	0.013689	58.99450	0.0000

R-squared	0.007547	Mean dependent var	0.000253
Adjusted R-squared	0.006850	S.D. dependent var	0.016348
S.E. of regression	0.016292	Akaike info criterion	-5.645039
Sum squared resid	0.378214	Schwarz criterion	-5.626597
Log likelihood	4032.735	Hannan-Quinn criter.	-5.638152
Durbin-Watson stat	2.023588		

图 7－13　沪市收益率 GARCH(1,1)模型估计结果

Equation: UNTITLED　Workfile: UNTITLED::U...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RZ
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 02/24/15　Time: 22:51
Sample (adjusted): 17 1443
Included observations: 1427 after adjustments
Convergence achieved after 17 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.000578	0.000327	-1.767382	0.0772
RZ(-15)	0.057618	0.021915	2.629194	0.0086
Variance Equation				
C	6.34E-06	1.29E-06	4.914888	0.0000
RESID(-1)^2	0.122121	0.009786	12.47964	0.0000
GARCH(-1)	0.863589	0.009235	93.50794	0.0000

R-squared	0.005937	Mean dependent var	-0.000173
Adjusted R-squared	0.005240	S.D. dependent var	0.017933
S.E. of regression	0.017886	Akaike info criterion	-5.476727
Sum squared resid	0.455883	Schwarz criterion	-5.458285
Log likelihood	3912.645	Hannan-Quinn criter.	-5.469839
Durbin-Watson stat	1.914371		

图 7－14　深市收益率 GARCH(1,1)模型估计结果

(2) GARCH－M(1,1)估计结果。依照前面的步骤只要在“ARCH－M term”栏选择方程作为 ARCH－M 项的形式，即可得到 GARCH－M(1,1)模型的估计结果，如图 7－15 和图 7－16 所示。

可见，沪深两市均值方程中条件方差项 GARCH 的系数估计分别为 5.937 671 和 5.162 608，而且都是显著的。这反映了收益与风险的正相关关系，说明收益有正的风险溢价。而且上海股市的风险溢价要高于深圳，这说明上海股市的投资者更加厌恶风险，要求更高的风险补偿。

(二) 股市收益波动非对称性的研究

1. TARCH 模型估计结果

在图 7－12 的“ARCH Specification”下拉列表中选择“TARCH”，即可得到 *rh*、*rz* 的 TARCH 模型估计结果，如图 7－17 和图 7－18 所示。

Equation: UNTITLED　Workfile: UNTITLED::Untit...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RH
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 02/24/15 Time: 23:17
Sample (adjusted): 17 1443
Included observations: 1427 after adjustments
Convergence achieved after 35 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	5.393201	2.229364	2.419166	0.0156
C	-0.000901	0.000470	-1.919788	0.0549
RH(-15)	0.058052	0.019591	2.963217	0.0030
Variance Equation				
C	9.53E-06	1.64E-06	5.816806	0.0000
RESID(-1)^2	0.186266	0.016265	11.45214	0.0000
GARCH(-1)	0.796661	0.014985	53.16511	0.0000

R-squared	0.015188	Mean dependent var	0.000253
Adjusted R-squared	0.013805	S.D. dependent var	0.016348
S.E. of regression	0.016234	Akaike info criterion	-5.648340
Sum squared resid	0.375302	Schwarz criterion	-5.626210
Log likelihood	4036.091	Hannan-Quinn criter.	-5.640075
Durbin-Watson stat	2.016243		

图 7－15　沪市收益率 GARCH－M(1,1)模型估计结果

Equation: UNTITLED　Workfile: UNTITLED::Untit...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RZ
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 02/24/15 Time: 23:16
Sample (adjusted): 17 1443
Included observations: 1427 after adjustments
Convergence achieved after 30 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	4.594723	2.053713	2.237276	0.0253
C	-0.001463	0.000545	-2.685754	0.0072
RZ(-15)	0.057913	0.021850	2.650517	0.0080
Variance Equation				
C	6.49E-06	1.32E-06	4.935549	0.0000
RESID(-1)^2	0.123269	0.009872	12.48647	0.0000
GARCH(-1)	0.862239	0.009246	93.25964	0.0000

R-squared	0.009552	Mean dependent var	-0.000173
Adjusted R-squared	0.008161	S.D. dependent var	0.017933
S.E. of regression	0.017860	Akaike info criterion	-5.479087
Sum squared resid	0.454225	Schwarz criterion	-5.456956
Log likelihood	3915.328	Hannan-Quinn criter.	-5.470822
Durbin-Watson stat	1.908661		

图 7－16　深市收益率 GARCH－M(1,1)模型估计结果

Equation: UNTITLED Workfile: UNTITLED::Untit...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RH
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 02/24/15 Time: 23:11
Sample (adjusted): 17 1443
Included observations: 1427 after adjustments
Convergence achieved after 17 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*RESID(-1)^2*(RESID(-1)<0) + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.000255	0.000319	-0.798658	0.4245
RH(-15)	0.065431	0.021635	3.024390	0.0025
Variance Equation				
C	7.24E-06	1.17E-06	6.214853	0.0000
RESID(-1)^2	0.110386	0.013645	8.090044	0.0000
RESID(-1)^2*(RESID(-1)<0)	0.085070	0.018400	4.623231	0.0000
GARCH(-1)	0.832203	0.011047	75.33515	0.0000

R-squared	0.007258	Mean dependent var	0.000253
Adjusted R-squared	0.006561	S.D. dependent var	0.016348
S.E. of regression	0.016294	Akaike info criterion	-5.649883
Sum squared resid	0.378324	Schwarz criterion	-5.627753
Log likelihood	4037.191	Hannan-Quinn criter.	-5.641618
Durbin-Watson stat	2.022234		

图 7－17 沪市收益率 TARCHT(1,1)模型估计结果

Equation: UNTITLED Workfile: UNTITLED::Untit...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RZ
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 02/24/15 Time: 23:13
Sample (adjusted): 17 1443
Included observations: 1427 after adjustments
Convergence achieved after 25 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*RESID(-1)^2*(RESID(-1)<0) + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.000785	0.000355	-2.209213	0.0272
RZ(-15)	0.056475	0.022264	2.536640	0.0112
Variance Equation				
C	5.73E-06	1.23E-06	4.678372	0.0000
RESID(-1)^2	0.096538	0.009723	9.929007	0.0000
RESID(-1)^2*(RESID(-1)<0)	0.047731	0.014347	3.327039	0.0009
GARCH(-1)	0.868181	0.008742	99.31512	0.0000

R-squared	0.005212	Mean dependent var	-0.000173
Adjusted R-squared	0.004514	S.D. dependent var	0.017933
S.E. of regression	0.017893	Akaike info criterion	-5.478775
Sum squared resid	0.456216	Schwarz criterion	-5.456645
Log likelihood	3915.106	Hannan-Quinn criter.	-5.470510
Durbin-Watson stat	1.913130		

图 7－18 深市收益率 TARCH(1,1)模型估计结果

在 TARCH 中，$\varepsilon_{t-1}^2 d_{t-1}$ 项的系数估计值都大于 0，而且都是显著的。这说明沪深股市中坏消息引起的波动比同等大小的好消息引起的波动要大，沪深股市都存在杠杆效应。

2. EGARCH 模型估计结果

在图 7－12 的“ARCH Specification”下拉列表中选择“EGARCH”，则可得到 rh、rz 的 EGARCH 模型估计结果，分别如图 7－19 和图 7－20 所示。

Equation: UNTITLED　Workfile: UNTITLED::Untit...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RH
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 02/24/15　Time: 23:06
Sample (adjusted): 17 1443
Included observations: 1427 after adjustments
Convergence achieved after 18 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.000277	0.000304	-0.911354	0.3621
RH(-15)	0.054489	0.020987	2.596320	0.0094
Variance Equation				
C(3)	-0.510279	0.057507	-8.873385	0.0000
C(4)	0.279250	0.022156	12.60396	0.0000
C(5)	-0.051769	0.011142	-4.646175	0.0000
C(6)	0.964145	0.005937	162.3956	0.0000

R-squared	0.006399	Mean dependent var	0.000253
Adjusted R-squared	0.005702	S.D. dependent var	0.016348
S.E. of regression	0.016301	Akaike info criterion	-5.667942
Sum squared resid	0.378651	Schwarz criterion	-5.645811
Log likelihood	4050.076	Hannan-Quinn criter.	-5.659677
Durbin-Watson stat	2.021974		

图 7－19　沪市收益率 EGARCH(1,1)模型估计结果

在 EGARCH 中，$\dfrac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}$ 项的系数估计值都小于零。在估计结果中，沪市为－0.051846，深市为－0.032059，而且都是显著的，这也说明了沪深股市中都存在杠杆效应。

Equation: UNTITLED Workfile: UNTITLED::Untit...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RZ
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 02/24/15 Time: 23:04
Sample (adjusted): 17 1443
Included observations: 1427 after adjustments
Convergence achieved after 25 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.000775	0.000340	-2.277718	0.0227
RZ(-15)	0.049784	0.022380	2.224550	0.0261
Variance Equation				
C(3)	-0.370392	0.045651	-8.113551	0.0000
C(4)	0.236776	0.015847	14.94152	0.0000
C(5)	-0.032125	0.008887	-3.614971	0.0003
C(6)	0.976575	0.005094	191.7096	0.0000

R-squared	0.004831	Mean dependent var	-0.000173
Adjusted R-squared	0.004132	S.D. dependent var	0.017933
S.E. of regression	0.017896	Akaike info criterion	-5.487200
Sum squared resid	0.456391	Schwarz criterion	-5.465069
Log likelihood	3921.117	Hannan-Quinn criter.	-5.478935
Durbin-Watson stat	1.913409		

图 7－20 深市收益率 EGARCH(1,1)模型估计结果

（三）沪深股市波动溢出效应的研究

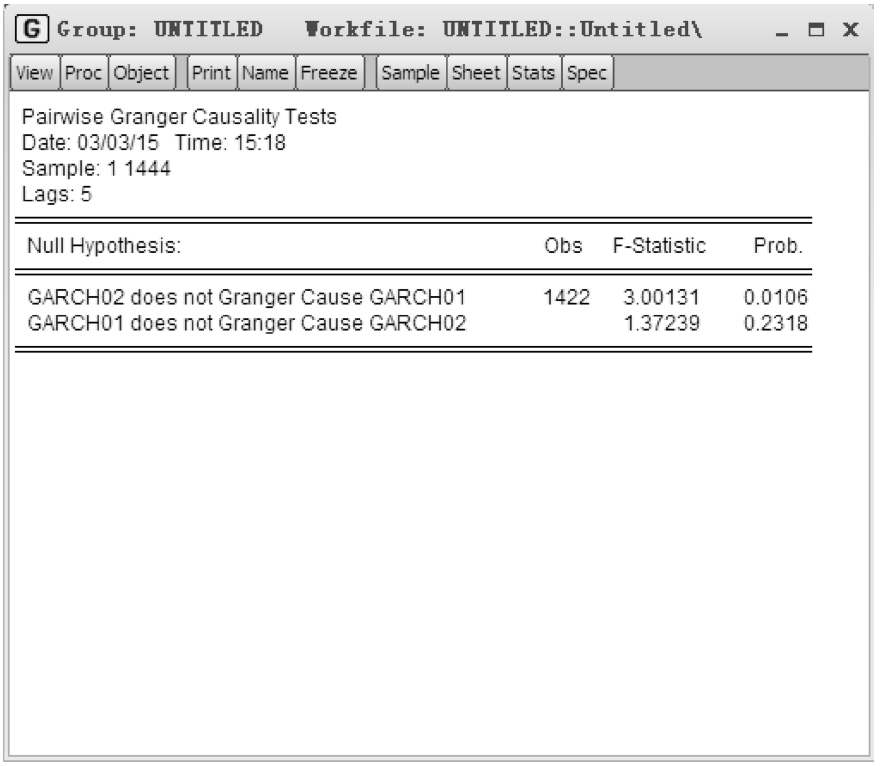
当某个资本市场出现大幅波动的时候，就会引起投资者在另外的资本市场的投资行为的改变，将这种波动传递到其他的资本市场。这就是所谓的“溢出效应”。例如“9·11”恐怖袭击后，美国股市的大震荡引起欧洲及亚洲股市中投资者的恐慌，从而引发了当地资本市场的大动荡。接下来我们将检验深沪两市之间的波动是否存在“溢出效应”。

1. 检验两市波动的因果性

（1）提取条件方差。重复前面 GARCH－M 模型建模的步骤，选择主菜单栏“Procs”下的“Make GARCH Variance Series”，得到 *rh* 回归方程残差项的条件方差数据序列 GARCH01，同样的步骤 *rz* 得到回归方程残差项的条件方差数据序列 GARCH02。

（2）检验两市波动的因果性。在“Workfile”中同时选中“GARCH01”和

“GARCH02”，点击鼠标右键，选择“Open”—“As Group”，在弹出的窗口中点击“View”—“Granger Causality”，并选择滞后阶数 5，得到如图 7－21 所示的结果。



Null Hypothesis:	Obs	F-Statistic	Prob.
GARCH02 does not Granger Cause GARCH01	1422	3.00131	0.0106
GARCH01 does not Granger Cause GARCH02		1.37239	0.2318

图 7－21　Granger 因果检验

可见，我们不能拒绝原假设：上海的波动不能因果深圳的波动。但是可以拒绝原假设：深圳的波动不能因果上海的波动。这初步证明沪深股市的波动之间存在溢出效应，且是不对称，单向的，表明是由于深圳市场的波动导致了上海市场的波动，而不是相反。

2. 修正 GARCH－M 模型

在沪市 GARCH－M 模型的条件方差方程中加入深市波动的滞后项，应该会改善估计结果。在“Equation Specification”窗口中，按图 7－22 所示输入如下变量，即在模型的条件方差方程中加入了深市波动的滞后项。

点击“OK”，则得到加入滞后项 GARCH02 后沪市 GARCH－M 模型重新估计的结果，如图 7－23 所示。

与前面图 7－15 结果比较可见，加入滞后项后，沪市 GARCH－M 模型中均值方程的 GARCH 项估计值变大，而且更加显著，并且估计的标准误差缩小了。这说明在

条件方差方程中加入深市波动的滞后项是恰当的。此时沪市收益率的 GARCH－M 效应更加明显了，风险（波动性）与收益之间的正相关关系更加显著。

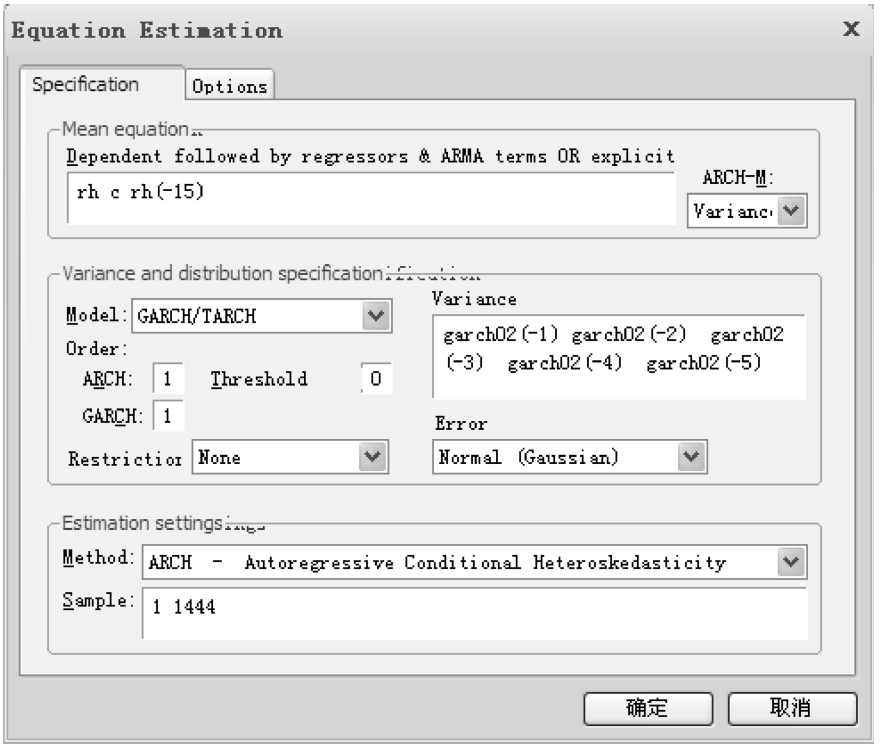


图 7－22　修正 GARCH－M 模型

我们运用 GARCH 类模型，对沪深股市收益率的波动性、波动的非对称性，以及波动之间的溢出效应做了全面的分析。通过分析，基本可以得出以下结论：

第一，沪深股市收益率都存在明显的 GARCH 效应。

第二，沪深股市都存在明显的 GARCH－M 效应，而且沪市的正向风险溢价要高于深市，反映了上海股市的投资者比深圳的投资者更加厌恶风险。

第三，沪深股市都存在明显的杠杆效应，反映了在我国股票市场上坏消息引起的波动要大于好消息引起的波动。

第四，沪深股市之间波动存在溢出效应，而且是单向的，深市的波动将引起沪市的波动，加入深市波动的模型将有助于提高沪市风险溢价的水平。

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RH
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 03/03/15 Time: 15:23
Sample (adjusted): 22 1443
Included observations: 1422 after adjustments
Convergence achieved after 13 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1) + C(7)*GARCH02(-1)
+ C(8)*GARCH02(-2) + C(9)*GARCH02(-3) + C(10)*GARCH02(-4) +
C(11)*GARCH02(-5)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	6.545762	1.890266	3.462879	0.0005
C	-0.000836	0.000374	-2.234814	0.0254
RH(-15)	0.051521	0.020765	2.481217	0.0131
Variance Equation				
C	1.12E-05	2.01E-06	5.590507	0.0000
RESID(-1)^2	0.260882	0.025649	10.17125	0.0000
GARCH(-1)	0.717634	0.032231	22.26506	0.0000
GARCH02(-1)	-0.445244	0.140664	-3.165312	0.0015
GARCH02(-2)	0.809545	0.304196	2.661263	0.0078
GARCH02(-3)	-0.568817	0.302198	-1.882264	0.0598
GARCH02(-4)	-0.234161	0.257296	-0.910082	0.3628
GARCH02(-5)	0.435975	0.112437	3.877497	0.0001

R-squared	0.017513	Mean dependent var	0.000244
Adjusted R-squared	0.016128	S.D. dependent var	0.016373
S.E. of regression	0.016241	Akaike info criterion	-5.680029
Sum squared resid	0.374275	Schwarz criterion	-5.639342
Log likelihood	4049.501	Hannan-Quinn criter.	-5.664831
Durbin-Watson stat	2.000181		

图 7－23 沪市 GARCH－M(加入滞后项 GARCH02)的估计结果

实验八　联立方程模型在金融数据中的应用

一、实验目的

了解内生变量、外生变量的定义及区别，了解联立性偏误的定义，从而理解普通最小二乘法不能用于估计联立方程模型的原因。

掌握联立方程模型的常用估计方法，尤其是两阶段最小二乘法（“TSLS”）的估计方法，以及如何运用 EViews 软件在实证研究中实现。

二、基本概念

由模型系统决定其取值的变量称为内生变量。内生变量受模型中其他变量的影响，也可能影响其他内生变量，即内生变量既可以是被解释变量，也可以是解释变量。由模型系统以外的因素决定其取值的变量称为外生变量。外生变量只影响系统内的其他变量，而不受其他变量的影响，因此在方程中只能做解释变量，不能做被解释变量。

用普通最小二乘法（OLS）对经典线形回归模型进行回归将得到最优线性无偏估计量。但在结构式模型中，由于内生变量既可作为解释变量又可作为被解释变量，经典线性回归模型的一个基本假设——解释变量与随机误差项不相关——将得不到满足，因此若仍对结构式模型中的每个结构方程分别运用 OLS 进行估计，所得到的参数估计值将是有偏和不一致的，即存在联立性偏误或联立方程偏误。

三、实验内容及要求

（一）实验内容

根据 1997 年 1 月—2004 年 3 月的货币供应量（M0、M1、M2）与股票价格的有关数据，利用两阶段最小二乘法估计由股票价格与货币供应量形成的联立方程模型（这里以上证综合指数代表股票价格），从而检验流通中现金 M0、狭义货币 M1、广义货币 M2 作为货币供应量与上证指数的关系。

（二）实验要求

（1）理解本章有关概念；

（2）思考：在何时应建立联立方程模型，并运用有关的估计方法；若此时运用了普通最小二乘法，结果如何；

（3）熟练掌握两阶段最小二乘法在 EViews 中的操作。

四、实验指导

（一）根据有关定义及经济原理建立联立方程模型

$$SCI_t = \alpha 0 + \alpha_1 M_t + \alpha_2 M_{t-6} + u_{1t} \tag{8.1}$$

$$M_t = \beta_0 + \beta_1 SCI_t + \beta_2 \text{II} V_t + \beta_3 IR_t + \beta_4 R_t + u_{2t} \tag{8.2}$$

其中，M_t 代表第 t 月的货币需求量，$\text{II} V_t$ 代表第 t 月的工业增加值，IR_t 代表第 t 月的通货膨胀率，R_t 代表第 t 月的一年期存款利率（模型具体构建过程见教材）。

我们将在 EViews 中利用两阶段最小二乘法估计上述联立方程模型，这个过程主要分两个步骤：首先利用普通最小二乘法求得内生变量的拟合值，然后用拟合值代替内生变量再利用两阶段最小二乘法求得结构参数估计值。我们将以 $M0$ 代表货币量说明模型在 EViews 中的估计过程，然后对于 $M1$、$M2$ 仅列出结果。

（二）导入数据

打开 EViews 软件，选择“File”菜单中的“New Workfile”选项，在“Workfile frequency”框中选择“Monthly”，在“Start date”和“End date”框中分别输入“1997：01”和“2004：03”，单击“OK”。选择“File”菜单中的“Import - Read Text - Lotus - Excel”选项，找到要导入的名为 EX7.1.xls 的 Excel 文档完成数据导入，建立相应的工作组，如图 8 - 1 所示。

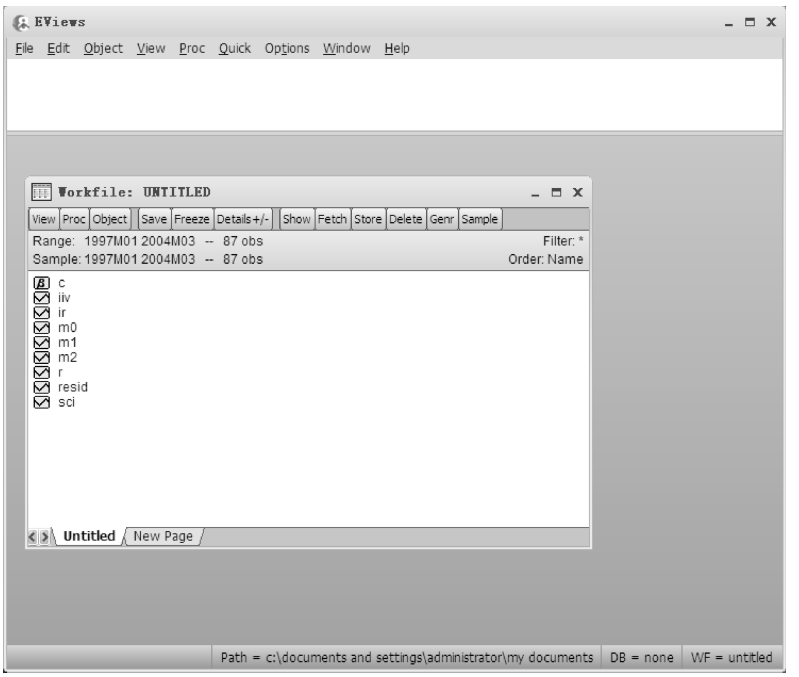


图 8 - 1　数据导入

（三）估计结构式方程(8.1)参数

在菜单中选择“Quick”—“Estimation equation”，出现如图 8－2 所示的窗口。

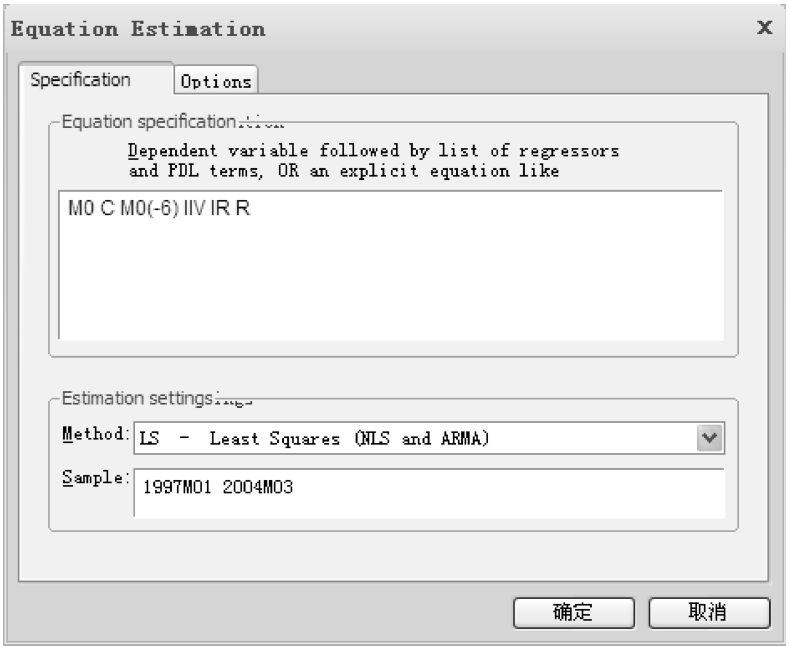


图 8－2　回归方程设定

在“Method”中选择 LS(即普通最小二乘法)，然后在“Estimation Settings”上方空白处首先输入被解释变量 M0，接着输入作为解释变量的外生变量(各变量的下标已除去)，注意不要忘记常数项。单击“OK”，则出现如图 8－3 所示的结果。

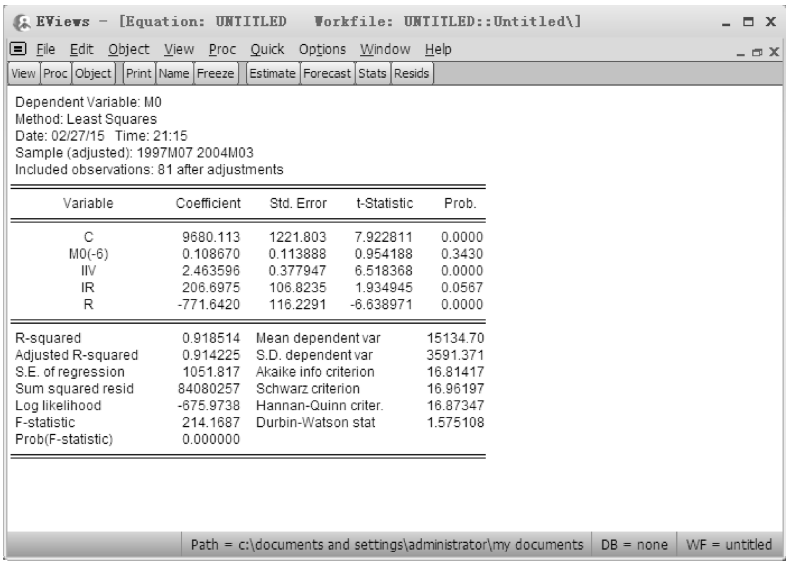


图 8－3　回归方程估计结果

我们得到如下的估计结果(括号内为 t 统计量,下同):

$$\hat{M}_{0t}=9\ 680.11+2.46\,\mathrm{II}V_t+206.70IR_t-771.64R_t+0.11M_{0t-6}$$
$$(7.92)\qquad(6.52)\qquad(1.93)\qquad(-6.54)\qquad(0.95)$$

点击“Quick”菜单下的“Generate Series”,得到如图 8-4 所示的窗口。

在“Enter Equation”下面的空白栏中键入如图 8-4 中的方程,就可以得到 $M0$ 的拟合值“m0fitted”。

点击“Quick”—“Estimate Equation”,在“Method”中选择“TSLS”(两阶段最小二乘法),将出现如图 8-5 所示的窗口。

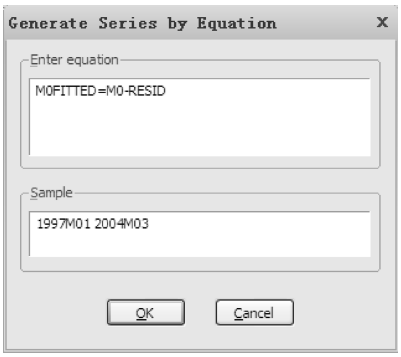


图 8-4　快速生成序列“ mofitted”

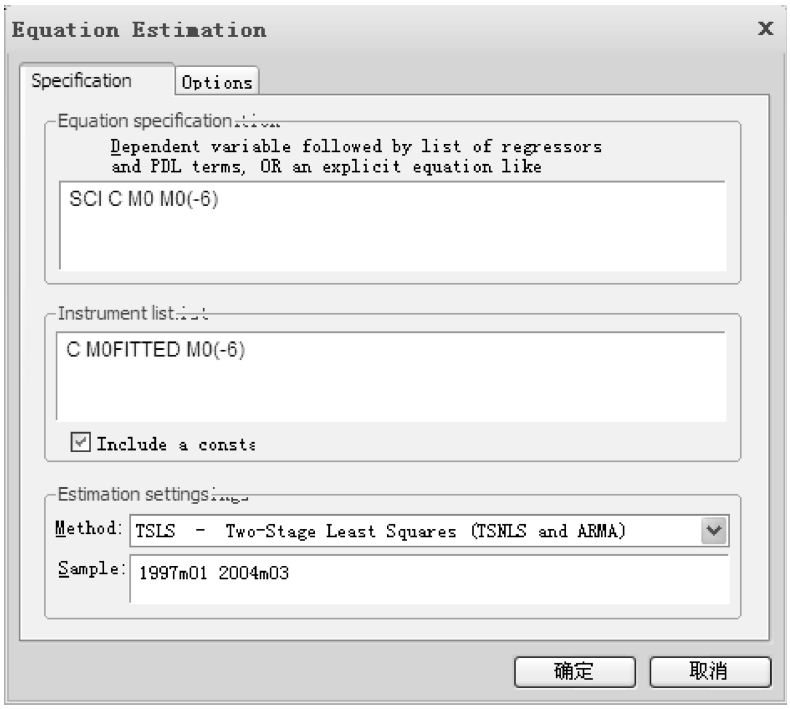


图 8-5　选择两阶段最小二乘法估计方程

在“Instrument List”上方的空白栏中按结构式方程(8.1)输入相应的变量,在其下方的空白栏中输入图示的工具变量,然后点击“OK”,就可以得到结构式方程(8.1)参数的两阶段最小二乘估计值:

$$\hat{SCI}_t=1022.49-0.001M_{0t}+0.04M_{0t-6}$$
$$(7.39)\qquad(-0.04)\qquad(1.32)$$

（四）估计结构式方程(8.2)参数

在菜单中选择“Quick”—“Estimate Equation”，出现如图 8－6 所示窗口。

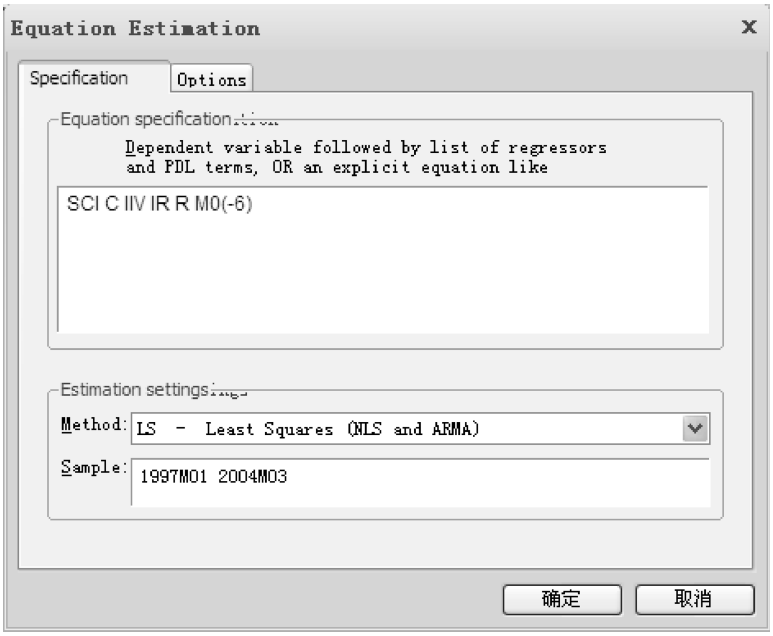


图 8－6　回归方程设定

在“Method”中选择“LS”(即普通最小二乘法)，然后在“Estimation Settings”上方空白处首先输入被解释变量 SCI，接着输入作为解释变量的图示外生变量，单击“OK”，得到如图 8－7 所示的结果。

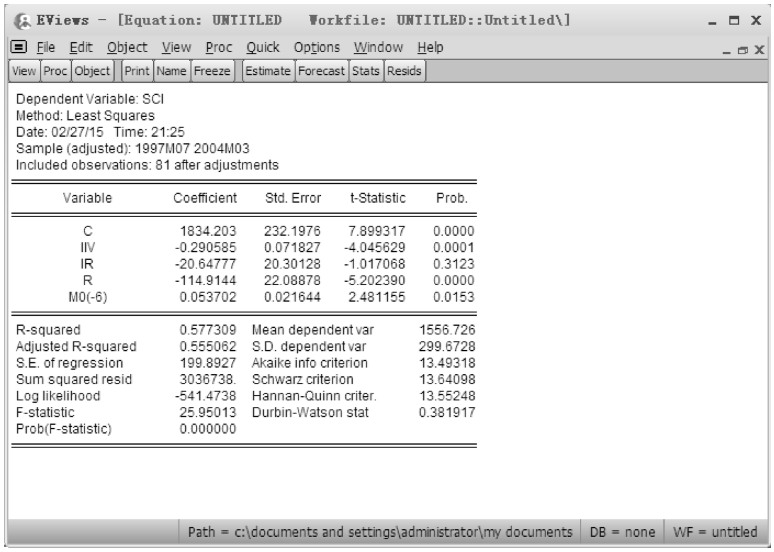


图 8－7　回归方程结果

选择“Quick”菜单下的“Generate series”菜单，将出现如图 8－8 所示的窗口。

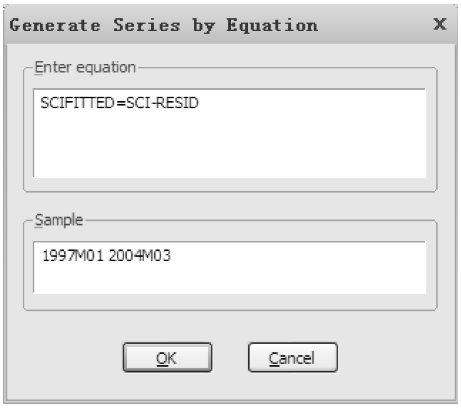


图 8－8　快速生成序列“ scifitted”

在“Enter Equation”下面的空白栏中键入如图 8－8 的方程，就可以得到 *sci* 的拟合值“scifitted”。

点击“Quick”—“Estimate equation”，在“Method”中选择“TSLS”（两阶段最小二乘法），将出现如图 8－9 所示的窗口。

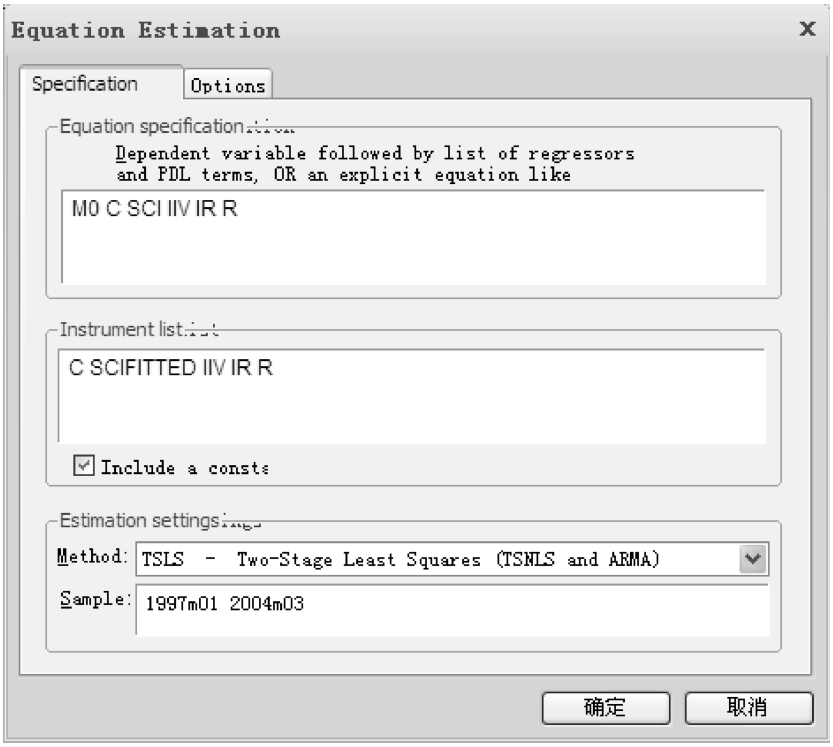


图 8－9　选择两阶段最小二乘法估计方程

在“Instrument list”上方的空白栏中按结构式方程(8.2)输入相应的变量,在其下方的空白栏中输入图示的工具变量。点击“OK”,就可以得到结构式方程(8.2)参数两阶段最小二乘估计值:

$$\hat{M}_{0t}=5\ 968.43+2.02SCI_t+3.05\text{Ⅱ}V_t+248.48IR_t-539.10R_t$$
$$(1.07)\quad(0.85)\quad(7.98)\quad(1.77)\quad(-1.44)$$

同样,对于狭义货币 M1 作为货币量代表,我们可以估计模型得到:

$$\hat{SCI}_t=1\ 239.33+0.01M_{1t}-0.01M_{1t-6}$$
$$(11.14)\quad(0.38)\quad(-0.22)$$

$$\hat{M}_{1t}=24\ 823.27-1.68SCI_t+17.42\text{Ⅱ}V_t-3\ 009.34R_t-163.79IR_t$$
$$(5.44)\quad(-0.93)\quad(0.43)\quad(-8.30)\quad(-0.56)$$

对于广义货币 M2 作为货币量代表,同样可以得到估计模型:

$$\hat{SCI}_t=1\ 210.96+0.04M_{2t}-0.03M_{2t-6}$$
$$(8.49)\quad(1.24)\quad(-1.17)$$

$$\hat{M}_{2t}=320\ 877.8-115.12SCI_t+32.31\text{Ⅱ}V_t-24\ 362.84R_t-4\ 935.05IR_t$$
$$(2.70)\quad(-2.26)\quad(4.06)\quad(-3.06)\quad(-1.76)$$

(五) 分析

可以看出,无论是流通中现金 M0、狭义货币 M1,还是广义货币 M2,无论是当月值还是过去第 6 个月的值,在对股票价格的解释中,他们的系数都是不显著的。因此,可以认为货币供应量对股票指数影响微乎其微。另一方面,股票指数在对流通中现金 M0、狭义货币 M1 的解释中,其系数也是不显著的,但在对广义货币的解释中,股票指数的系数则是显著的,因此,可以认为,股票指数对流通中现金 M0、狭义货币 M1 是没有影响的,而对广义货币量则是有影响的。

实验九　基于 EViews 的 ARDL 模型和 ECM 模型应用

一、实验目的

理解 ARDL 模型与 ECM 模型的原理与应用条件，掌握如何在 EViews 软件中实现。运用 ARDL 模型估计变量之间长期关系的系数，运用 ECM 模型估计变量之间短期误差修正关系的系数。需要注意的是，只有当变量之间确定存在长期协整关系时，才能使用 ARDL 模型与 ECM 模型进行后续估计。

二、基本概念

ARDL(Autoregressive Distributed Lag)模型称为自回归分布滞后模型，其优点在于：一方面，对数据的样本容量要求较低；另一方面，无论各研究变量同为 I(0)、同为 I(1)，或是 I(0)与 I(1)过程的混合，都可以使用 ARDL 模型来测算变量之间的长期关系，得出一致而有效的回归结果。对 ARDL 模型进行简单的线性变换，可以得到 ECM 模型。ECM(Error Correction Model)全称为误差修正模型，主要描述变量由短期波动向长期均衡调整的过程，能够有效解释短期波动行为。

三、实验内容及要求

运用 ARDL 模型与 ECM 模型研究天然气期货价格 NG、美元兑人民币汇率 UC、上证综合指数 SH 与上海空气质量指数 AQI 对上海碳配额现货价格 SHEA 的具体影响过程。数据为 2013 年 11 月 26 日至 2018 年 1 月 31 日之间的有效日数据。

要求在认真理解书中相关章节内容的基础上，通过实验掌握 ARDL 模型与 ECM 模型的实际应用方法，并熟悉 EViews 的基本使用方法。

四、实验指导

(一)准备步骤

(1)工作表建立，数据导入。

打开 EViews 并选择建立新的 Workfile，由于样本数据并非完全连续，因而我们在“Structure type”一项下选择“Unstructured/Undated”，输入观测值个数 590。完成文件建立后，可以使用 Proc 功能下的“Import”导入初始数据文件 EX9. xlsx 中的五个样本序列：SHEA、NG、UC、SH 和 AQI。

(2)数据对数处理。

Genr 功能可以对序列作不同处理。我们对序列作对数处理，在编辑框中输入："LNSHEA=log(SHEA)"。其中，LNSHEA 为对数化后序列名称，log()为过程函数，SHEA 为原始序列。对其他变量作相同处理后，得到序列：LNSHEA、LNNG、LNUC、LNSH 和 LNAQI。

（二）描述性统计

选中所有对数化后序列，右键"Open-as group"，在 View 功能下选择"Descriptive Stats"，可以得到各序列的描述性统计结果。描述性统计的结果使我们对数据质量、序列整体趋势有大致的了解。

（三）平稳性检验

在将数据导入模型进行实证分析之前，首先要判断样本是否符合模型使用要求，即对各个序列的平稳性进行检验。我们点开 LNSHEA 序列，在 View 功能中选择"Unit root test"，本实验选用 ADF 检验法及 SIC 信息准则，首先考察 LNSHEA 在"Level"，即 I(0)过程下的平稳性情况，观察数据趋势后选择包含截距项而不包含时间趋势项，设定见图 9－1，结果见图 9－2。

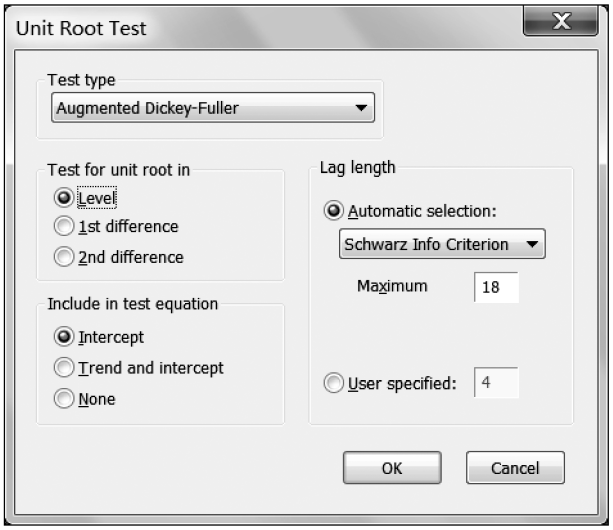


图 9－1　LNSHEA 平稳性检验设定

对于 LNSHEA 序列，ADF 方法给出统计量值－1.33，高于各显著性水平下的临界值，说明 LNSHEA 在 I(0)下并不平稳。我们可以按如上步骤重新对 LNSHEA 进行单位根检验，但选择"1st difference"而不是"Level"，考察 LNSHEA 的一阶差分项，即 LNSHEA 在 I(1)过程下是否平稳。完成所有平稳性检验工作后，我们发现，LNSHEA、LNNG、LNUC 和 LNSH 在 I(1)过程下达到平稳，LNAQI 在 I(0)过程下

Null Hypothesis: LNSHEA has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.332223	0.6160
Test critical values:	1% level	-3.441223	
	5% level	-2.866228	
	10% level	-2.569326	

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

图 9－2　LNSHEA 平稳性检验结果

平稳，符合 ARDL 模型的要求。

（四）ARDL 边限协整检验

在这一步骤中，我们通过构建完整的 ARDL 模型来考察各变量间是否存在长期协整关系。打开 EViews 上方的 Quick 功能，点击“Estimate Equation”，在 Method 中选择 ARDL 方法。

首先，我们在 Specification 下的主编辑框内依次写出因变量 4 四个回归的自变量：“LNSHEA LNNG LNUC LNSH LNAQI”。勾选 Automatic Select，利用软件自动确定各变量的滞后期，由于本实验所用样本为日数据，故为因变量与自变量设置最大滞后项均为 5。关于是否包含截距项及时间趋势项，则需要经过试验，本实验在比较后选择仅包含截距项，设定如图 9－3 所示。上文中用于选择最优滞后期的信息准则可在 Options 下进行选择，首先选择 AIC，见图 9－4。

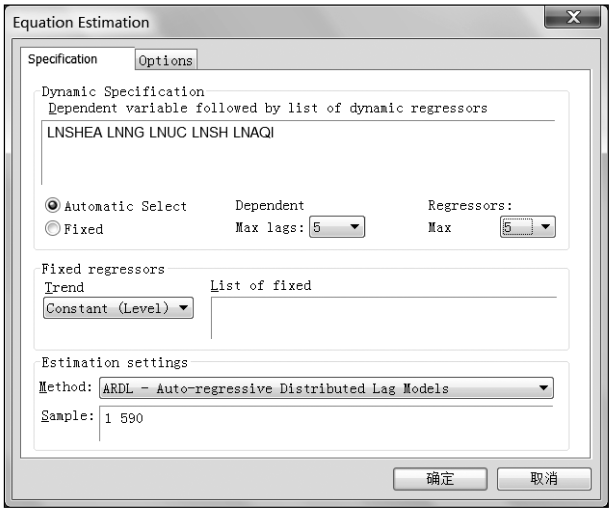


图 9－3　ARDL 模型设定

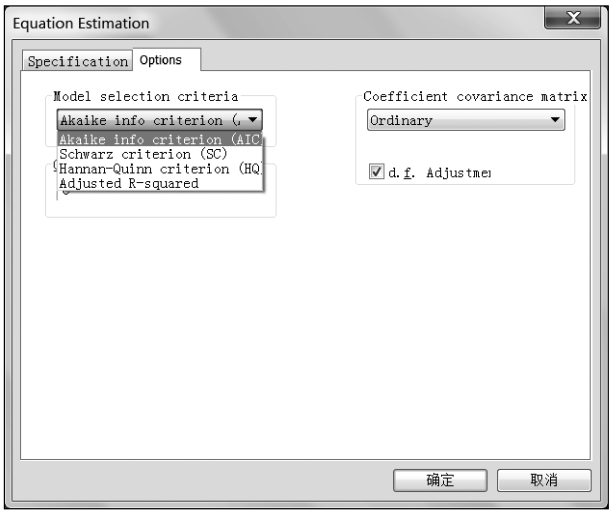


图 9－4　信息准则设定

根据 AIC 准则构建的模型为 ARDL(5,1,3,5,2),回归的标准误差“S. E. of regression”为 0.061062。界面上方的 Estimate 功能可以使我们在保留模型基本设定的情况下对个别选项进行调整。更换为 SIC 准则后,模型变为 ARDL(1,0,0,0,0),回归的标准误差为 0.062518,高于 AIC 准则下自动选取的模型,因此,本实验将根据 ARDL(5,1,3,5,2)进行协整检验与后续的系数估计。

完成了 ARDL 模型的选取与构建后,我们在模型界面打开 View 功能,在“Coefficient diagnostics”下选择“Bounds test”,对模型 F(LNSHEA|LNNG, LNUC, LNSH, LNAQI)进行边限协整检验,得到结果如图 9－5 所示。

ARDL Bounds Test
Sample: 6 590
Included observations: 585
Null Hypothesis: No long-run relationships exist

Test Statistic	Value	k
F-statistic	5.639288	4

Critical Value Bounds

Significance	I0 Bound	I1 Bound
10%	2.45	3.52
5%	2.86	4.01
2.5%	3.25	4.49
1%	3.74	5.06

图 9－5　边限协整检验结果

结果得到的 F 统计量值大于各显著水平下的上界值，可以认为本实验的因变量与自变量之间存在长期协整关系。

（五）长期与短期系数估计

为考察变量之间具体的系数关系并进行分析，使用 View 功能下的“Coefficient diagnostics”，选择“Cointegration and long run form”。结果如图 9－6、图 9－7 所示。

Long Run Coefficients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNNG	3.322083	0.578025	5.747299	0.0000
LNUC	8.487621	2.789462	3.042745	0.0025
LNSH	0.913023	0.541679	1.685543	0.0924
LNAQI	-0.288065	0.328773	-0.876180	0.3813
C	-22.441794	7.149424	-3.138965	0.0018

图 9－6　长期系数估计

Cointegrating Form

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LNSHEA(-1))	0.005125	0.040880	0.125363	0.9003
D(LNSHEA(-2))	-0.084624	0.041052	-2.061366	0.0397
D(LNSHEA(-3))	-0.069162	0.040791	-1.695514	0.0905
D(LNSHEA(-4))	-0.069454	0.040527	-1.713744	0.0871
D(LNNG)	0.284762	0.071713	3.970858	0.0001
D(LNUC)	2.758916	0.953897	2.892258	0.0040
D(LNUC(-1))	-2.625756	1.424160	-1.843722	0.0657
D(LNUC(-2))	3.012435	0.959887	3.138322	0.0018
D(LNSH)	0.027579	0.137896	0.199996	0.8416
D(LNSH(-1))	0.128293	0.197587	0.649296	0.5164
D(LNSH(-2))	-0.291042	0.196722	-1.479458	0.1396
D(LNSH(-3))	0.004430	0.195106	0.022707	0.9819
D(LNSH(-4))	0.321628	0.137614	2.337173	0.0198
D(LNAQI)	-0.007834	0.006809	-1.150510	0.2504
D(LNAQI(-1))	0.015034	0.006738	2.231343	0.0261
CointEq(-1)	-0.028125	0.007312	-3.846564	0.0001

Cointeq = LNSHEA - (3.3221*LNNG + 8.4876*LNUC + 0.9130*LNSH -0.2881*LNAQI -22.4418)

图 9－7　短期系数估计

（六）模型稳定性检验

在 EViews 的 9.0 及以上版本中，仅能检验 OLS 模型的稳定性，因此，我们需要进行一定的转换。

回到模型 ARDL（5，1，3，5，2）的操作界面，在 View 功能下选择“Representations”，选择“Estimation Equation”下方的公式内容，复制到剪切板备用。重新打开 Quick 下的“Estimate Equation”，选用 LS 方法，将上述复制内容粘贴到编辑框中，如图 9－8 所示。

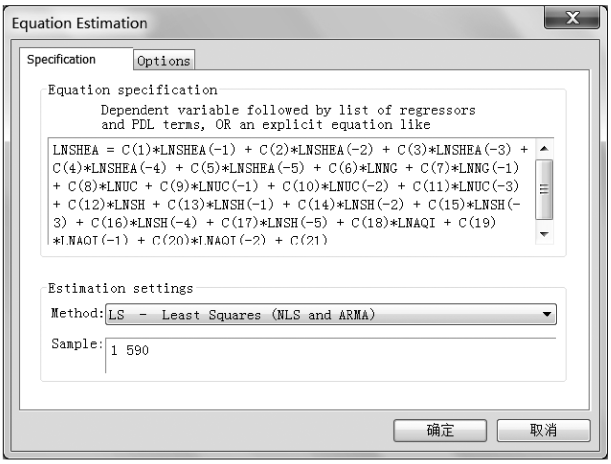


图 9－8　OLS 设定

根据此方式得到的模型，在本质上与 ARDL 完全一致。使用 View 功能下的“Stability diagnostics”，选择“Recursive estimates”，即可使用 CUSUM 或 CUSUMSQ 等方法对模型稳定性进行检验，如图 9－9 所示。

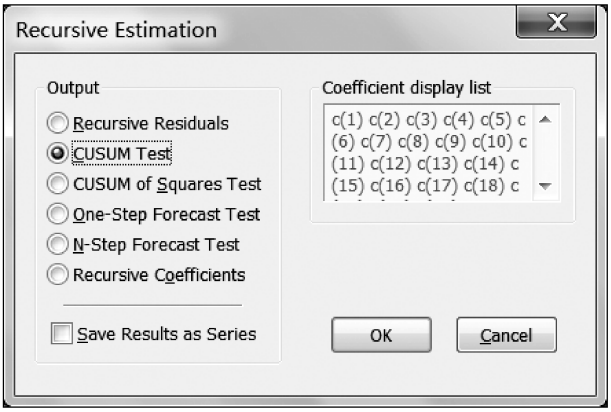


图 9－9　稳定性检验设定

（七）自相关检验

同样在 View 功能下，利用“Residual diagnostics”中的“Serial correlation LM test”，根据标准误差选择出最优滞后期，得到结果见图 9－10。

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.232864	Prob. F(2,562)	0.7923
Obs*R-squared	0.484387	Prob. Chi-Square(2)	0.7849

图 9－10　自相关检验结果

F 统计量的 p 值为 0.7923，接受原假设，说明本实验的模型不存在自相关问题。

实验十　面板数据分析及回归模型应用

一、实验目的

理解面板数据的特点及运用，通过具体实验案例掌握在 EViews10 软件中面板数据的处理方式、模型的回归操作及结果解读等。

二、基本概念

在经济研究中，除了单独的横截面和时间序列分析之外，我们也常常需要将两者结合同时分析，在时间序列上选取多个截面进行观测，这时候需要用到面板数据。面板数据是一种二维数据，即 n 个个体在两个或多个观测时期 T 上观测到的数据，如果 n 远大于 T，即观测个体大于时间，我们称之为短面板，反之为长面板。在本实验当中，我们选用的是短面板数据。

三、实验内容及要求

运用面板数据模型及多元回归模型，在现金-现金流敏感性模型下判断中小板企业是否存在融资约束。现金-现金流敏感性模型主要研究企业持有现金变动与现金流之间的关系，该模型认为，企业存在外部融资约束时会从现金流中提取现金，因而两者正向变动。同时，模型中还存在未来投资机会（Growth）、公司规模（Size）、资本支出（Expen）、非现金营运资本变动（ΔNWC）、短期借款变动（ΔSD）等其他影响现金变动的控制变量。本实验中样本数据为 2008—2015 年中小板上市企业的微观非平衡面板数据，数据来源于 CSMAR 数据库。

在充分理解面板数据的含义和特点的基础上，通过本实验操作掌握面板数据模型的应用与结果解读，并了解 EViews 软件的使用方法。

四、实验指导

（一）准备步骤

（1）数据导入。

打开 EViews10 软件，在“File”栏目下选择“Open”的“Foreign Data as Workfile”，如图 10－1 所示。

注意：在其他面板数据模型实验中，选择本方式导入需要保证已经将原始样本数据整理为标准的面板数据格式。

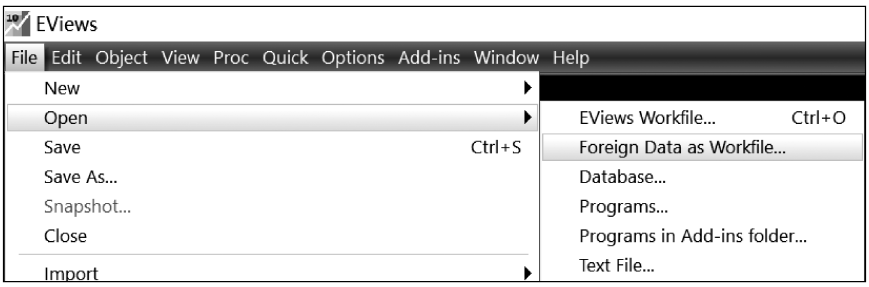


图 10－1　外部数据导入

选择本实验数据文件 EX10. xlsx，在第一、二、三步保持软件原设定不做修改，点击“OK”进入导入第四步(Step 4 of 5)。注意在 Column info 按以下要求修改相应列的数据类型。鼠标点击到窗口下方表格的第一列，修改“stkcd”(股票代码)为 character(字符型)(如图 10－2 所示)；点击第二列，修改“accper”(会计期间)为 Date(日期型)；本实验中其他变量设置为 Number(数字型)。

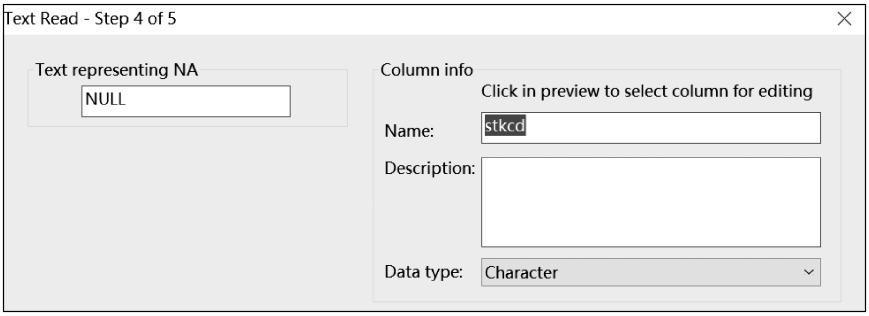


图 10－2　数据类型选择步骤

在数据导入第五步，按照图 10－2 进行设置。注意设置为 Dated Panel，并且横截面序列的 ID 为 stkcd(股票代码)，时间序列的 ID 为 accper(会计期间)，如图 10－3 所示。点击 Finish 完成面板数据的导入工作。弹出的窗口选择“NO”。

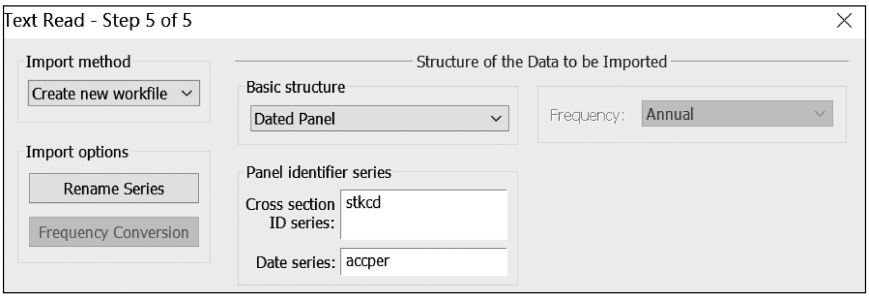


图 10－3　横截面与时间序列选择

(2)数据预处理——缩尾(可选)。

在实证部分下载的数据样本往往比较大,难免会出现极端值的情况。为避免异常值对实证结果的影响,在研究当中通常对连续变量进行缩尾处理,将变量超出特定百分位范围的那部分数值替换为其特定百分位的数值,可以进行双侧或者单侧缩尾处理。

本实验采用双侧缩尾,如果软件没有 winsorising 功能,在缩尾之前需要在 EViews 官网或者软件内下载插件“Trim”。(Add-ins)

插件下载:菜单栏“Add-ins”“Download Add-ins”(见图 10－4)。在插件选项中找到“Trim”插件下载安装(install)(见图 10－5)。

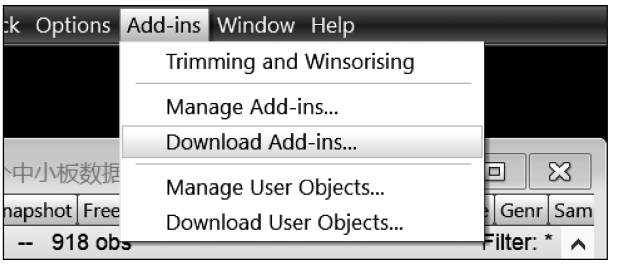


图 10－4　插件下载

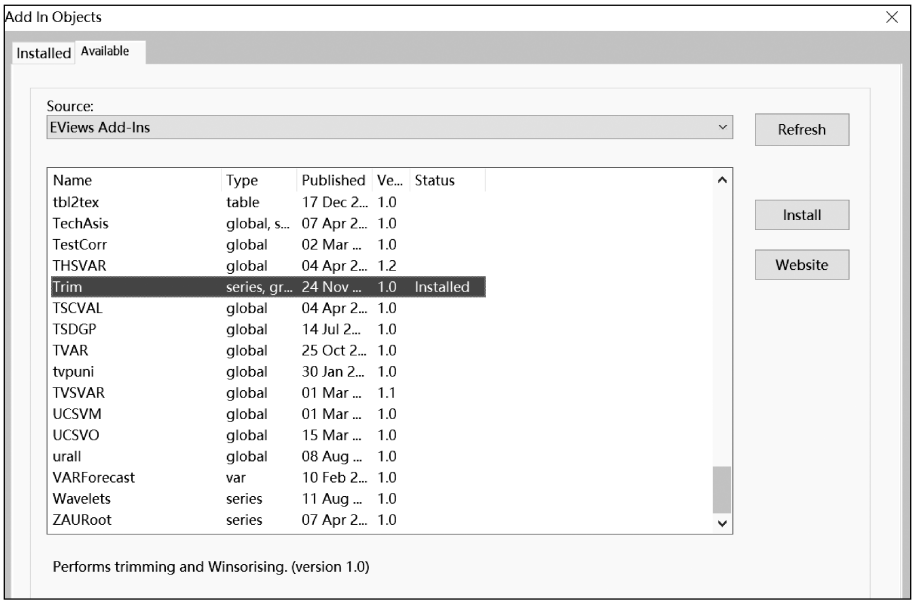


图 10－5　选择缩尾和截尾插件

软件介绍："Trim"插件主要用来进行 Trimming(截尾)和 Winsorising(缩尾)操作。

对变量进行缩尾处理：

双击变量名称，打开需要缩尾的变量序列(本图打开的是 CASH)，然后点击"Add-ins"下"Triming and Winsorising"栏目，出现如图 10－6 所示的窗口。更改"Techniques"处为"Winsorising"，选择合适的百分比进行缩尾，并自行修改定义新序列名称，随即得到缩尾后的 CASH 序列。按照上述步骤，依次对实验中其他变量进行缩尾处理，生成缩尾后的新序列，用于后续回归。

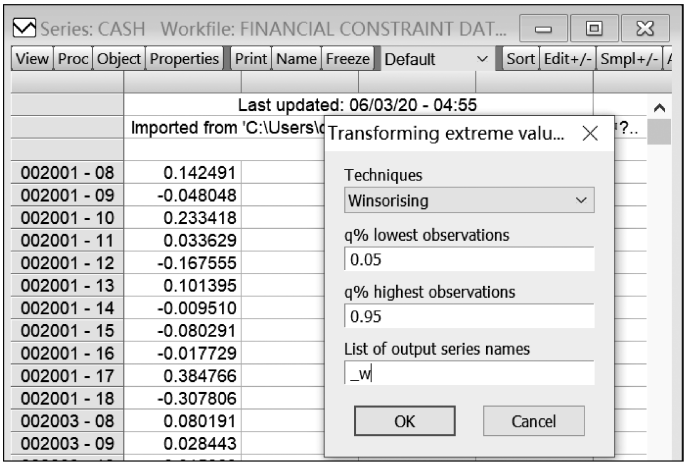


图 10－6　对序列缩尾处理

(3)描述性统计分析。

"Quick"—"Group Statistics"— "Descriptive Statistics"—"Common sample"，在窗口中键入本实验的主要变量名称，以空格隔开，可以得到变量的描述性统计分析(见图 10－7)。通过描述性统计分析，我们可以对样本数据的特征和数据的波动有大致的了解。

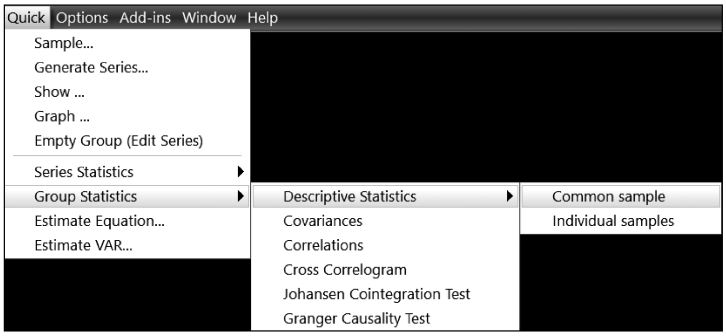


图 10－7　数据描述性统计分析

(4)主要变量相关系数。

在 EViews10 上方菜单栏中选择“Quick”菜单，找到 Group Statistics 栏目下 Correlations 选项，来查看多变量间的相关系数，见图 10－8。

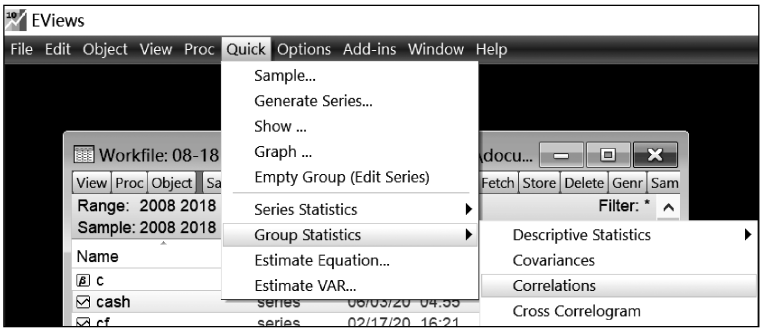


图 10－8　相关系数分析

在对话框中键入主要变量的名称，以空格隔开，点击“OK”，即得到主要变量之间的相关系数结果(见图 10－9)。相关系数结果显示，本实验中 7 个变量之间相关系数都不高，均在 0.8 以下，因而可以排除多重共线性的情况。

View Proc Object Print Name Freeze Sample Sheet Stats Spec

Correlation

	CASH	CF	GROWTH	NWC	SD	EXPEN	SIZE
CASH	1.000000	0.150587	0.078414	-0.063972	-0.045910	0.148213	0.078877
CF	0.150587	1.000000	0.285228	-0.267495	0.226557	-0.207098	-0.002142
GROWTH	0.078414	0.285228	1.000000	-0.200155	0.481383	-0.418031	0.033667
NWC	-0.063972	-0.267495	-0.200155	1.000000	-0.689581	0.229124	-0.046626
SD	-0.045910	0.226557	0.481383	-0.689581	1.000000	-0.303081	0.114008
EXPEN	0.148213	-0.207098	-0.418031	0.229124	-0.303081	1.000000	0.133742
SIZE	0.078877	-0.002142	0.033667	-0.046626	0.114008	0.133742	1.000000

图 10－9　相关系数结果

(5)面板数据回归。

在菜单栏中选择“Quick”—“Estimate Equation”，在窗口中键入基准模型的变量(见图 10－10)，其他界面设定保持不变，点击“OK”得到回归结果。

实证结果如图 10－11 所示，现金流的系数在 1%显著性水平上为正(0.17)，表明样本企业存在明显的融资约束。

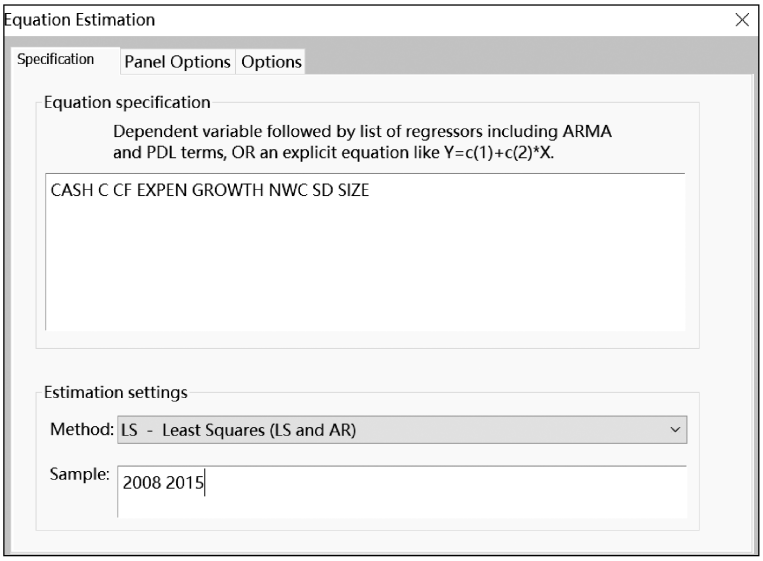


图 10－10　回归模型设定

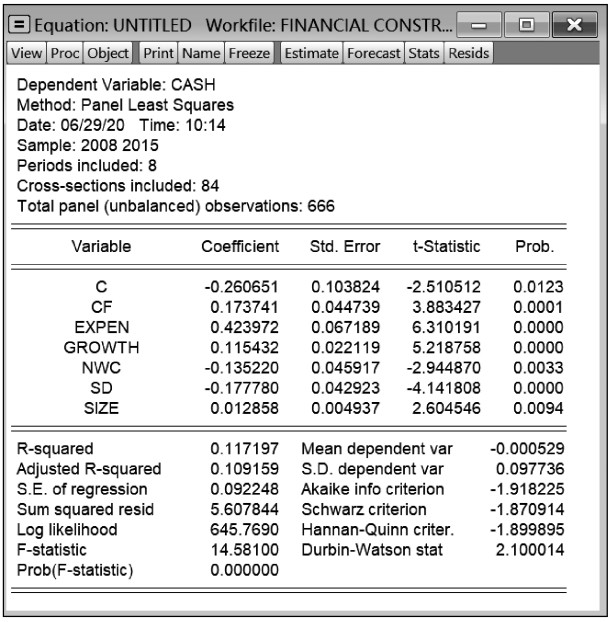


图 10－11　面板数据回归结果