计量经济学实验设计与应用

基于 EViews 软件

胡华平

2023-10-30

仅供学习参考，请勿随意外传
NWSUAF @HHP 2025

目录

前言 **3**

致谢 4

声明 4

1 EViews 入门 **5**

1.1 为何选择 EViews? 6

1.2 怎样理解 Eviews 的运作逻辑? 7

1.3 如何获取 Eviews 官方帮助? 7

1.4 实验内容 8

1.5 实验准备 9

1.5.1 实验软件 9

1.5.2 实验材料 9

1.5.3 实验规则 9

1.6 EViews 快速入门 9

1.7 EViews 菜单操作流程 13

1.7.1 准备数据（data preparing） 14

1.7.2 构建工作文件（workfile creating） 17

1.7.3 构造对象（objects generating） 26

1.7.4 探索性分析（exploratory analyzing） 32

1.7.5 线性回归分析（linear regression analyzing） 46

1.7.6 报告结果输出（output reporting） 48

1.8 Eviews 编程体系及语法 50

1.8.1 命令运行方式 50

1.8.2 命令语法结构 59

1.8.3 常用命令及函数 64

2 一元线性回归 **68**

2.1 实验目的及要求 69

2.2 实验原理 . . . 69
2.2.1 回归的本质是什么？ . . . 69
2.2.2 怎样用样本数据推断总体参数？ . . . 72
2.2.3 最小二乘法有哪些优点？ . . . 73
2.2.4 如何对模型进行统计检验？ . . . 76
2.2.5 如何进行样本外预测？ . . . 83
2.2.6 怎样看懂 EViews 回归分析报告结果？ . . . 86
2.3 实验内容 . . . 89
2.4 实验准备 . . . 90
2.4.1 实验软件 . . . 90
2.4.2 实验材料 . . . 91
2.4.3 实验规则 . . . 92
2.5 主要实验步骤 . . . 94
2.5.1 新建工作文件并导入数据 . . . 94
2.5.2 进行 Eviews 回归分析 . . . 96
2.5.3 构建几个重要变量对象 . . . 99
2.5.4 计算 FF 原序列 (5 个序列 **series**) . . . 103
2.5.5 计算 ff 离差序列 (5 个序列 **series**) . . . 103
2.5.6 计算回归系数的估计值: . . . 105
2.5.7 计算回归拟合值、拟合值的离差、回归残差和残差平方序列 (**series**)： . . . 107
2.5.8 计算回归方程的误差方差及标准差 . . . 108
2.5.9 计算回归系数的样本方差和标准差： . . . 112
2.5.10 进行平方和分解，计算 TSS、ESS 和 RSS，以及各自的自由度（标量） . . . 114
2.5.11 计算相关系数 r 和判定系数 r^2(标量 **scalar**)： . . . 116
2.5.12 回归系数的 t 检验 . . . 116
2.5.13 对回归方程的整体显著性进行 F 假设检验 . . . 118
2.5.14 进行样本外的均值预测、个值预测，并计算置信区间 . . . 120
2.6 附录：prg 源代码 . . . 125
2.7 作业题 . . . 128

3 模型函数形式 **132**
3.1 实验目的及要求 . . . 133
3.2 实验原理 . . . 133
3.2.1 模型设置为有截距还是无截距？ . . . 133
3.2.2 变量单位变换下怎么选择模型？ . . . 134
3.2.3 如何选择模型的函数形式？ . . . 135

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

3.3 实验内容 . . . 137
3.4 实验准备 . . . 137
3.4.1 实验软件 . . . 137
3.4.2 实验材料 . . . 138
3.4.3 实验规则 . . . 138
3.5 主要实验步骤 . . . 141
3.5.1 新建工作文件并导入数据 . . . 141
3.5.2 绘制三个散点图（Y 表示食物支出，X 表示家庭总支出） . . . 141
3.5.3 对 7 类模型进行 Eviews 回归分析 . . . 144
3.5.4 提取各类回归方程的回归系数 . . . 148
3.5.5 给出指定的样本数据点 (X_0,Y_0) . . . 149
3.5.6 得到 7 个线性模型中 Y 对 X 在 (X_0, Y_0) 处的点斜率 . . . 151
3.5.7 得到 7 个线性模型中 Y 对 X 在 (X_0, Y_0) 处的点弹性 . . . 152
3.5.8 得到 7 个线性模型中 Y 对 X 的平均弹性 . . . 154
3.5.9 把所有计算结果制作成表格 . . . 155
3.5.10 对倒数模型 M_6 进行 Eviews 回归分析并提取回归系数 . . . 158
3.5.11 倒数模型 M_6 情形下，计算 Y 相对于 X 的点斜率 ($X_0 = 100, Y_0 = 30$) 点斜率 (slope) . . . 160
3.5.12 倒数模型 M_6 情形下，计算 Y 相对于 X 的平均弹性 (elasticity) . . . 164
3.6 附录：prg 源代码 . . . 164
3.7 作业题 . . . 169

4 多元线性回归 172
4.1 实验目的及要求 . . . 173
4.2 实验原理 . . . 173
4.2.1 基本模型和重要概念的矩阵表达 . . . 173
4.2.2 OLS 估计及 BLUE 性质的矩阵表达 . . . 175
4.2.3 平方和分解与拟合优度的矩阵表达 . . . 179
4.2.4 回归系数显著性检验（t 检验）的矩阵方法实现 . . . 180
4.2.5 样本外预测的矩阵方法实现 . . . 184
4.3 实验内容 . . . 187
4.4 实验准备 . . . 187
4.4.1 实验软件 . . . 187
4.4.2 实验材料 . . . 188
4.4.3 实验规则 . . . 189

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

4.5 主要实验步骤——以对数模型为例式 4.58 190
4.5.1 新建工作文件并导入数据 190
4.5.2 进行对数模型的 Eviews 回归分析 192
4.5.3 构建几个重要变量对象 194
4.5.4 构造 X 矩阵和 Y 矩阵对象 195
4.5.5 计算回归方程的回归系数向量 197
4.5.6 计算回归方程的误差方差及标准差 197
4.5.7 计算回归系数的方差协方差矩阵、系数的样本方差和标准差 (列向量) 200
4.5.8 进行平方和分解，计算 TSS、ESS 和 RSS，以及各自的自由度（标量） 200
4.5.9 计算自变量的相关系数表格、回归方程的判定系数和调整判定系数 204
4.5.10 对回归方程的回归系数进行显著性 t 检验 206
4.5.11 对回归方程的整体显著性进行 F 假设检验 206
4.5.12 进行样本外的均值预测、个值预测，并计算置信区间 208
4.6 附录：prg 源代码 211
4.7 实验作业 213

5 多重共线性问题 217
5.1 实验目的及要求 218
5.2 实验原理 218
5.2.1 什么是模型多重共线性？ 218
5.2.2 如何诊断模型多重共线性问题？ 219
5.2.3 如何修正模型多重共线性问题？ 222
5.3 实验内容 223
5.4 实验准备 224
5.4.1 实验软件 224
5.4.2 实验材料 224
5.4.3 实验规则 225
5.5 主要实验步骤 227
5.5.1 新建工作文件并导入数据 227
5.5.2 采用最小二乘法建立主回归模型 228
5.5.3 侦查多重共线性 231
5.5.4 多重共线性模型矫正 256
5.6 附录：prg 源代码 265
5.7 实验作业 267

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

6 异方差问题 **272**
6.1 实验目的及要求 . . . 273
6.2 实验原理 . . . 273
6.2.1 什么是模型异方差问题？ . . . 273
6.2.2 如何诊断模型异方差问题？ . . . 275
6.2.3 图形识别检验法 . . . 276
6.2.4 正式检验方法：Park 检验法 . . . 277
6.2.5 正式检验方法：Glejser 检验法 . . . 278
6.2.6 正式检验方法：BPG 检验法 . . . 279
6.2.7 正式检验方法：White 检验法 . . . 280
6.2.8 如何矫正模型异方差问题？ . . . 282
6.3 实验内容 . . . 283
6.4 实验准备 . . . 283
6.4.1 实验软件 . . . 283
6.4.2 实验材料 . . . 283
6.4.3 实验规则 . . . 284
6.5 主要实验步骤 . . . 286
6.5.1 导入数据并进行预处理 . . . 286
6.5.2 采用最小二乘法建立主回归模型 . . . 288
6.5.3 侦查模型是否存在异方差 . . . 288
6.5.4 使用加权最小二乘法（WLS）矫正异方差问题 . . . 314
6.5.5 使用 White 校正法矫正异方差问题 . . . 318
6.6 作业题 . . . 321

7 自相关问题 **324**
7.1 实验目的及要求 . . . 325
7.2 实验原理 . . . 325
7.2.1 什么时候模型会出现自相关问题？ . . . 325
7.2.2 随机干扰项自相关有哪些表现形态？ . . . 326
7.2.3 自相关问题模型会有什么后果？ . . . 327
7.2.4 如何诊断模型存在自相关问题？ . . . 329
7.2.5 如何矫正存在自相关问题模型？ . . . 330
7.3 实验内容 . . . 331
7.4 实验准备 . . . 332
7.4.1 实验软件 . . . 332
7.4.2 实验材料 . . . 332

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

7.4.3 实验规则 . . . 333
7.5 主要实验步骤 . . . 335
7.5.1 导入数据并进行预处理 . . . 335
7.5.2 采用最小二乘法建立主回归模型 . . . 337
7.5.3 侦查模型是否存在自相关问题 . . . 337
7.5.4 自相关问题模型的矫正 . . . 351
7.6 作业题：存货案例 . . . 369

8 虚拟变量回归 375
8.1 实验目的及要求 . . . 376
8.2 实验原理 . . . 376
8.2.1 如何把定性变量转换为虚拟变量？ . . . 377
8.2.2 如何理解虚拟变量回归模型？ . . . 379
8.2.3 虚拟变量回归模型有哪些形式？ . . . 380
8.3 实验内容 . . . 382
8.4 实验准备 . . . 382
8.4.1 实验软件 . . . 382
8.4.2 实验材料 . . . 383
8.4.3 实验规则 . . . 384
8.5 主要实验步骤 . . . 385
8.5.1 导入数据并进行预处理 . . . 385
8.5.2 把定性变量设置成虚拟变量体系 . . . 386
8.5.3 操作解读 . . . 389
8.5.4 仅含有虚拟变量 . . . 390
8.5.5 同时含有虚拟变量和定量变量 . . . 411
8.5.6 时间序列季节虚拟变量模型 . . . 427
8.5.7 分段线性回归模型 . . . 430
8.6 作业题 . . . 431

实验课程指导 436

参考文献 437

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

前言

你好，世界 (Hello World！)。

本书的第一次完成稿最早发布于 2019 年，当时完全使用了 `Rmarkdown` 文档编写，并使用 `bookdown` 进行全书渲染。其后我对这本书稿进行了断断续续的修订和维护，并主要以 html 网页版形式公开发布。

当然，html 网页版本的维护和发布是非常方便和有效率的。但是，我一直期望能够再进一步完善书稿后进行正式出版。遗憾的是，word 版本或 LaTex 版本的维护和更新，对我来说太过于耗费时间和精力，也比较麻烦，里面需要学习更多精细的技巧，甚至需要巧妙的（也可能是极其愚蠢的打补丁）。

2022 年底，`R` 世界迎来了 `Quarto` 新生态体系。在观望近一年后，我开始借助 `Quarto` 工具链重新编写了这本书。word 版本或 LaTex 版本的维护和更新变得更加便捷，让我有精力更多关注于书稿内容本身。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

致谢

2016-2019 年期间，R 世界变得更加富有生产力，这几年也是 `Rmarkdown` 文档写作和谢益辉 `down` 系列包大放异彩的时期（包括 `bookdown`、`pagedown`、`blogdown` 等重磅级 R 包）。这里对他给 R 生态带来的巨大贡献表示由衷的致敬和感谢。

2022 年底以后，`Quarto` 新生态开始以更统一化的工作流程重塑写作体系，感谢 `Quarto` 团队给 R 开源世界带来的创新和努力。

本书写作过程中，我在任教的大学里每年度讲授《计量经济学》课程，教学过程中（尤其是实验课环节）很多学生帮助发现了不少笔误、错误、啰嗦之处，或者提出了一些阅读感受和修订建议。他们既是本书稿的第一批读者和用户，也是本书稿不断完善的最大贡献群体之一。这里一并表示感谢。

最后，本书稿也会持续修订，也希望自己能集中精力争取尽快将其正式出版，呈现给更多的读者。

声明

- 本书稿目前仅作为我个人《计量经济学》本科生课程配套指导用书，供大家学习参考和使用。如有引用，请注明本书稿的版权出处：

  胡华平（著）. 计量经济学实验设计与应用——基于 EViews 软件 [Manuscript]. 2023 年 11 月.

- 本人对本书稿具有知识版权，任何他人或机构不得将此书稿用于商业目的。

胡华平（Kevin Hu）

2023 年 11 月

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

1 EViews 入门

仅供学习参考，请勿随意外传
NWSUAF @HHP 2025

1.1 为何选择 EViews?

EViews 全称为 Econometrics Views，是计量经济分析的常用软件之一。1981 年 David Lilien 开发一款名为 MicroTSP（Time Series Processor）分析工具，主要用于大型计算机进行时间序列数据分析。1994 年 QMS（Quantitative Micro Software）公司用 C++ 语言对 MicroTSP 进行重新设计，成功开发出 EViews，使其广泛适用于 Windows 操作系统。2010 年 QMS 公司被 IHS 公司收购[1]。

实际上，计量经济分析软件类型多样，功能特征各有侧重，不同使用群体根据自身情况可以灵活选择。常用的计量经济分析软件就包括 SPSS、EViews、Stata、SAS、Matlab 等，它们也是各有优势和不足。作为参考，不同任务情形和学习阶段应选择合适的软件：

（1）计量经济初级学习者往往以菜单驱动型操作为主，适合的软件包括：

- EViews：属于商业软件，基于 C++ 语言
- SPSS：属于商业软件，基于 Java 语言
- Stata：属于商业软件，基于 C 语言
- …

（2）计量经济中高级学习者则追求更多灵活的命令编程设计和操作，适合的软件包括：

- Matlab：属于商业软件，基于 C++、Java、MATLAB 语言
- SAS：属于商业软件，基于 C 语言
- R：属于开源软件，基于 C、Fortran、R 语言
- Python：属于开源软件
- …

计量经济学课程中使用到的数据类型主要包括截面数据（cross-section data）、时间序列数据（time series data）和面板数据（pannel data）。EViews 软件因为继承了其前身软件 MicroTSP 的主要特征，在计量经济分析中能满足基本功能分析需求，并且拥有十分全面的时间序列分析工具。此外，EViews 软件同时支持菜单型操作和命令代码操作，很适合计量经济学初学者使用，学习门槛相对较低，使用者可以集中于关注计量建模分析的理论应用与验证。而对于中高级使用者，EViews 也可以进行代码编程实现各种定制化的计量分析需求，从而保证软件使用的连贯性和一致性。

[1]参阅 IHS 官方网站：http://www.eviews.com/general/about_us.html

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

1.2 怎样理解 Eviews 的运作逻辑?

（1）使用者通过菜单操作基本就可以实现大部分的 EViews 功能

- 主要的统计和计量分析工具都被菜单化设计到 EViews 软件中。
- 鼠标 + 键盘的操作符合使用者的直观化学习习惯。
- 意味着使用者要熟记各类统计和计量分析操作的准确菜单位置，这一定程度上增加了使用者的记忆负担。

（2）使用者也可以通过运行命令来实现各种 EViews 功能

- 单独编写命令行，独立运行每一行命令。适用于简单分析任务，与菜单操作进行配合使用。
- 批量编写命令代码，整体运行代码文件。适合于复杂的或个性化的分析任务，便于命令代码维护，而且代码可重复使用，可以灵活开发出使用者定制化的分析流程（往往是 EViews 常用分析流程不能满足需求）。
- EViews 具有独树一帜命令代码编写体系，使用者需要理解 EViews 代码规则，避免与其他统计软件代码编写规则的混淆使用。

（3）EViews 是以对象（Object）方式来进行操作结果的管理和维护

- 所有的数据（data）、表格（table）、图片（graph）、方程（equation）等都将保存为对象（Object）。
- 所有的对象（Object）都存放在一个工作文件（workfile）中。
- 所有对象都可以进行复制（copy）、粘贴（paste）、保存（save）、重命名（rename）等操作。

1.3 如何获取 Eviews 官方帮助?

作为一款商业软件，EViews 能得到商业公司持续的软件维护和版本更新，不同 EViews 版本的界面、功能、特性等会有一定的改进和变化。2010 年以后，IHS 公司拥有 EViews 软件的商标和版权，目前发布的最新版本为 EViews 10。本书所用示例均以 EViews 10 版本进行运行和演示，使用者应注意实际使用版本下可能细节上的变化和差异[2]。

[2]需要注意的是，目前 EViews1-6 将不再兼容 Windows 10 操作系统

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

如果正确安装 EViews 10，使用者将得到 IHS 公司官方提供的详尽帮助手册和在线文档。菜单操作的具体获取办法是：运行 EViews 10 ⇒ 点击 Help ⇒ 下拉选择相应帮助手册或文档。

（1）快速帮助手册

- Object Reference
- Basic Command Reference
- Function Reference
- Matrix Reference
- Programming Reference
- Online Tutorials

（2）pdf 帮助文档

- User Guide I.pdf
- User Guide II.pdf
- Command & Programming Ref.pdf
- Object Reference.pdf
- Eviews Illustrated.pdf

1.4 实验内容

利用 EViews 软件进行如下操作：

- EViews 软件的启动；
- 数据的输入、编辑与序列生成；
- 图形分析与描述统计分析；
- 数据文件的存贮、调用。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

1.5 实验准备

1.5.1 实验软件

本次实验需要提前准备好如下软件：

- 统计分析软件 Eviews 9.0 版本及以上
- 公式编辑软件 Mathtype 6.0 版本及以上
- 写作编辑软件 Office Word/Excel 2010 版本及以上
- 浏览器软件 chrome 66.0 版本及以上或 360 极速浏览器 9.5 版本及以上

1.5.2 实验材料

1.5.3 实验规则

1.6 EViews 快速入门

宏观经济数据：表 1.1 给出给出了一国 1980 年第 1 季度到 2012 年第 1 季度间共计 129 个季度，关于 quarter 季度，GDP 国内生产总值（10 亿美元），PCE 人均消费支出（美元），Investment 投资（10 亿美元），Government 政府支出（10 亿美元）等方面的数据。

表 1.1: 宏观经济发展数据

year	quarter	GDP	PCE	Investment	Government
1980	Q1	5903	3797	778	1365
1980	Q2	5782	3710	708	1370
1980	Q3	5772	3750	654	1351
1980	Q4	5878	3800	721	1349
1981	Q1	6001	3821	792	1367
2011	Q1	13228	9377	1751	2514
2011	Q2	13272	9393	1778	2508
2011	Q3	13332	9434	1784	2508
2011	Q4	13429	9482	1876	2481

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 1.1: 宏观经济发展数据

year	quarter	GDP	PCE	Investment	Government
2012	Q1	13502	9550	1903	2462

变量定义和说明见表 1.2 ：

表 1.2: 变量定义及说明

variable	label
year	年
quarter	季度
GDP	国内生产总值（10 亿美元）
PCE	人均消费支出（美元）
Investment	投资（10 亿美元）
Government	政府支出（10 亿美元）

使用 EViews 软件分析该国宏观经济发展情况，一般需要事先将相关经济变量收集和整理成规范化的数据表（见表 1.1 ）。其中，每列代表变量（variables），每行代表样本（observations），数据文件格式为电子表格文件（.xlsx）等。

利用数据导入菜单，可以把预备好的电子表格文件（.xlsx）导入到 EViews 工作文件（workfile）中。导入数据后的 EViews 界面见图 1.1 ：

为了掌握宏观经济的样本数据特征，可以进行 EViews 基本的描述性统计分析。描述性统计分析结果见图 1.2 ：

进一步地，利用绘图操作，可以得到 4 个主要变量的多线图（见图 1.3 ）

同时，还可以绘制对数化的国内生产总值 GDP 与其他变量（PCE,Investment,Government）之间的相关系数表（见图 1.4 ）。

一个重要的工作还包括进行回归建模分析，可以尝试如下的回归模型：

$$GDP = \beta_1 + \beta_2 PCE + \beta_3 Investment + \beta_4 Government + u_i \tag{1.1}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

Group: GROUP_VARS Workfile: MACRO_ECONOMIC::gdp\

View | Proc | Object | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/- | Compare+/- | Transpose+/-

	GDP	PCE	INVESTMENT	GOVERNMENT
1980Q1	5903.4	3796.7	778.3	1365.4
1980Q2	5782.4	3710.5	708.1	1369.7
1980Q3	5771.7	3750.3	654.1	1350.8
1980Q4	5878.4	3800.3	720.6	1349.4
1981Q1	6000.6	3821.1	792.2	1367.3
1981Q2	5952.7	3821.1	754.5	1370.4
1981Q3	6025.0	3836.6	801.3	1367.3
1981Q4	5950.0	3807.6	770.2	1379.9
1982Q1	5852.3	3832.2	690.0	1378.5
1982Q2	5884.0	3845.9	689.4	1386.5
1982Q3	5861.4	3875.4	681.3	1396.0
1982Q4	5866.0	3946.1	620.7	1420.1
1983Q1	5938.9	3984.8	642.8	1430.8
1983Q2	6072.4	4063.9	704.8	1443.0
1983Q3	6192.2	4135.7	752.2	1468.0
1983Q4	6320.2	4201.3	831.4	1443.2
1984Q1				

图 1.1: 导入宏观经济发展数据

Table: TAB_DESCRIPTIVE Workfile: MACRO_ECONOMIC::gdp\

View | Proc | Object | Print | Name | Edit+/- | CellFmt | Grid+/- | Title | Comments+/-

	A	B	C	D	E	F
3	描述性统计分析表					
4		GDP	PCE	INVESTMENT	GOVERNMENT	
5	Mean	9637.569	6561.340	1395.501	1984.816	
6	Median	9239.300	6204.000	1282.400	1900.500	
7	Maximum	13502.40	9550.200	2266.300	2570.300	
8	Minimum	5771.700	3710.500	620.7000	1349.400	
9	Std. Dev.	2539.585	1896.230	495.7913	364.6666	
10	Skewness	0.048646	0.131772	0.172480	-0.044492	
11	Kurtosis	1.593279	1.602691	1.588386	1.948627	
12	Jarque-Bera	10.68727	10.86787	11.35013	5.984003	
13	Probability	0.004778	0.004366	0.003430	0.050187	
14	Sum	1243246.	846412.8	180019.6	256041.2	
15	Sum Sq. Dev.	8.26E+08	4.60E+08	31463553	17021666	
16	Observations	129	129	129	129	
17						
18						

图 1.2: 宏观经济发展数据的描述性统计

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

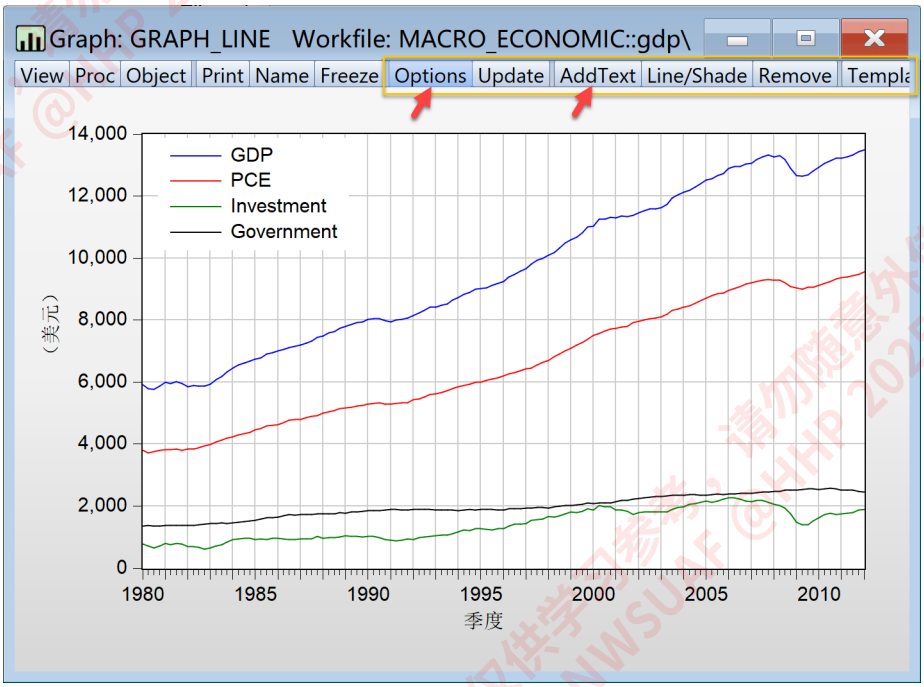


图 1.3: 4 个时间序列变量的多线图

Table: TAB_CORRL Workfile: MACRO_ECONOMIC::gdp\

View Proc Object Print Name Edit+/- CellFmt Grid+/- Title Comments+/-

Correlation

	A	B	C	D	E	F
1		GDP	PCE	INVESTMENT	GOVERNMENT	
2						
3	GDP	1.000000	0.998752	0.937940	0.980474	
4	PCE	0.998752	1.000000	0.927905	0.982601	
5	INVESTMENT	0.937940	0.927905	1.000000	0.872755	
6	GOVERNMENT	0.980474	0.982601	0.872755	1.000000	
7						
8						
9						
10						

图 1.4: 多个变量之间的相关系数表

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$GDP = \hat{\beta}_1 + \hat{\beta}_2 PCE + \hat{\beta}_3 Investment + \hat{\beta}_4 Government + e_i \tag{1.2}$$

利用 EViews 回归分析操作，可以得到如下的回归分析报告结果（见图 1.5 ）

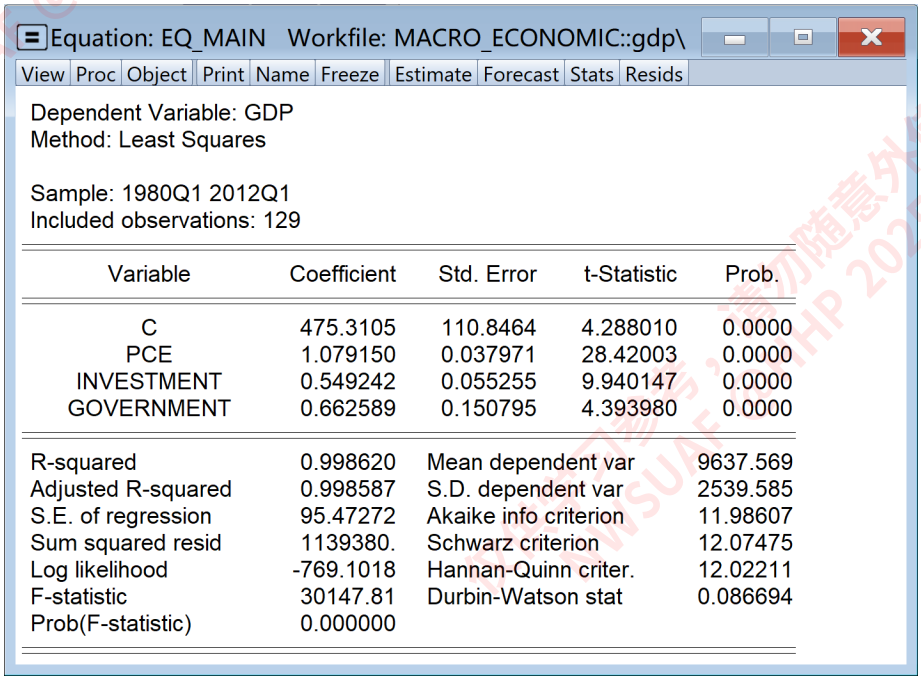


图 1.5: 国内生产总值 ln(GDP) 回归结果

1.7 EViews 菜单操作流程

上述的快速案例速览，简要列出了使用 EViews 进行一次样本数据实证分析的主要步骤过程，是典型的实践操作流程。我们基本可以总结为如下 6 个步骤：

- 准备数据（data preparing）
- 构建工作文件（workfile creating）
- 构造对象（objects generating）
- 探索性分析（exploratory analyzing）
- 线性回归分析（linear regression analyzing）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 报告结果输出（output reporting）

下面将逐条进行简要说明。

1.7.1 准备数据（data preparing）

准备数据（data preparing）属于前期准备工作，包括数据收集、清洗、整理等环节。目标是获得规范的数据表，要求每列都代表变量（variables），每行都代表样本（observations）。在进行正式 EViews 分析之前，提前准备一份规范的数据表是极为重要的。需要做的准备工作包括：

- 数据收集：根据研究目的设计理论模型，模型中的理论变量需要找到对应的现实数据变量。一般情况下，一个理论变量可以对应多个数据变量，但由于实际情况制约（如人力、资金、技术等条件约束），数据变量并不总是可以收集得到。
- 数据清洗：收集得到的数据并不总是完美和理想，比如常常有数据缺失、虚假数据、错误数据等，从而需要进行数据替换、删减、调整等操作。
- 数据整理：清洗后的数据，还需要以某种形式呈现出来。数据整理就是把数据的统计结构（变量和样本）与物理结构（行和列）进行某种匹配整合。

规范的数据表要求具有下面的结构形式 (Hadley 2016)：

- 每列都代表变量（variables）
- 每行都代表样本（observations）

规范的数据表可以称为坐标型数据（Cartesian data），数据的统计结构和物理结构是严格对应的，也即每列代表变量，每行都代表样本，数据点通过行坐标和列坐标一起定位得到。

大部分不规范的数据表则可以称为索引型数据（indexed data），数据的统计结构和物理结构并不是严格对应的，数据点往往是通过查询索引序号而得到（有点像通过偏旁部首查字典的方式）。

一份规范化的数据数据表，具有属于坐标型数据的形式，以表 1.3 为例：

- 每一行都代表一个观测样本（observation），分别从 1980 年 Q1 季度到 2012 年 Q1 季度，共有 129 行（也即 129 个季度）；
- 每一列都代表一个变量（variable），分别是 year 年、quarter 季度、GDP 国内生产总值（10 亿美元）、PCE 人均消费支出（美元）、Investment 投资（10 亿美元）、Government 政府支出（10 亿美元），共有 6 列（也即 6 个变量）。

- 数据表的维度为 (行数 =129; 列数 =6)，统计结构和物理结构严格对应。

表 1.3: 宏观经济发展数据规范数据表

	year	quarter	GDP	PCE	Investment	Government
1	1980	Q1	5903	3797	778	1365
2	1980	Q2	5782	3710	708	1370
3	1980	Q3	5772	3750	654	1351
4	1980	Q4	5878	3800	721	1349
5	1981	Q1	6001	3821	792	1367
125	2011	Q1	13228	9377	1751	2514
126	2011	Q2	13272	9393	1778	2508
127	2011	Q3	13332	9434	1784	2508
128	2011	Q4	13429	9482	1876	2481
129	2012	Q1	13502	9550	1903	2462

而糟糕的数据表往往有很多种形式，下面列出三种形式（见表 1.4 、表 1.5 和表 1.6 ）。它们都属于索引型数据形式，其中：

- 对于表 1.4 ：统计结构中，我们所关心的变量被分散在第 2 列所对应的各行中，而观测样本则被割裂在第一列对应的各行（表示年份）和第 3 列到第 4 列（表示季度）。整个数据表的维度为 (行数 =132; 列数 =6)，统计结构和物理结构不能一一对应，而是通过索引指标才能查到数据点。

- 对于表 1.5 ：统计结构中，我们所关心的变量与观测样本糅合在一起，除第 1 列之外，其他列都是这样的混合体，如 GDP_Q1GDP_Q2GDP_Q3 表示国内生产总值 GDP 分别与 Q1~Q4 季度的混合。同时，观测样本（如 1980 年 Q1 季度）也是被割裂开来了。整个数据表的维度为 (行数 =33; 列数 =17)，统计结构和物理结构不能一一对应，而是通过索引指标才能查到数据点。

- 对于表 1.6 ：该形式的数据表属于典型的索引型数据形式。其中，我们所关系的变量全部作为索引指标之一，出现在第 3 列（variable）；而数据值则全部出现在第 4 列（value）。数据值必须通过观测样本索引指标（第 1 列的 year 和第 2 列的 quarter）与变量索引指标（第 3 列的 variable）才能查询得到。该数据表的维度为 (行数 =516; 列数 =4)，统计结构和物理结构不能一一对应，而是通过索引指标才能查到数据点。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 1.4: 宏观经济发展数据（糟糕数据表 1）

	year	variable	Q1	Q2	Q3	Q4
1	1980	GDP	5903	5782	5772	5878
2	1980	PCE	3797	3710	3750	3800
3	1980	Investment	778	708	654	721
4	1980	Government	1365	1370	1351	1349
5	1981	GDP	6001	5953	6025	5950
125	2011	GDP	13228	13272	13332	13429
126	2011	PCE	9377	9393	9434	9482
127	2011	Investment	1751	1778	1784	1876
128	2011	Government	2514	2508	2508	2481
129	2012	GDP	13502			
130	2012	PCE	9550			
131	2012	Investment	1903			
132	2012	Government	2462			

表 1.5: 宏观经济发展数据（糟糕数据表 2）

	year	GDP_Q1	GDP_Q2	GDP_Q3	GDP_Q4	Government_Q3	Government_Q4
1	1980	5903	5782	5772	5878	1351	1349
2	1981	6001	5953	6025	5950	1367	1380
3	1982	5852	5884	5861	5866	1396	1420
4	1983	5939	6072	6192	6320	1468	1443
5	1984	6443	6554	6618	6672	1500	1532
29	2008	13267	13310	13187	12884	2511	2520
30	2009	12663	12641	12694	12814	2554	2548
31	2010	12938	13058	13140	13216	2570	2552
32	2011	13228	13272	13332	13429	2508	2481
33	2012	13502					

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 1.6: 宏观经济发展数据（糟糕数据表 3）

	year	quarter	variable	value
1	1980	Q1	GDP	5903
2	1980	Q2	GDP	5782
3	1980	Q3	GDP	5772
4	1980	Q4	GDP	5878
5	1981	Q1	GDP	6001
6	1981	Q2	GDP	5953
510	2010	Q3	Government	2570
511	2010	Q4	Government	2552
512	2011	Q1	Government	2514
513	2011	Q2	Government	2508
514	2011	Q3	Government	2508
515	2011	Q4	Government	2481
516	2012	Q1	Government	2462

简单地总结起来就是：

- 一份规范的数据表（见表 1.3 ）可以直接导入到 EViews 的工作文件（workfile）中，并马上可以用于实证建模分析的。
- 一份糟糕的数据表（见表 1.4 、表 1.5 和表 1.6 ）是不能直接导入到 EViews 工作文件（workfile）中（更严格地说，导入进去后是无意义的或无效的）。

1.7.2 构建工作文件 (workfile creating)

工作文件（workfile）是 EViews 所有操作结果的“大容器”，操作结果将“分门别类”地以各种对象（objects）形式留存在这个“大容器”里。工作文件（workfile）的构建跟数据类型（截面数据、时间序列数据、面板数据）很有关系，一般都需要提前设置（当然也可以事后修改）。

1.7.2.1 启动 EViews 软件

在电脑操作系统中，找到 EViews 安装程序，点击运行后将显示 EViews 软件界面（见图 1.6 ）

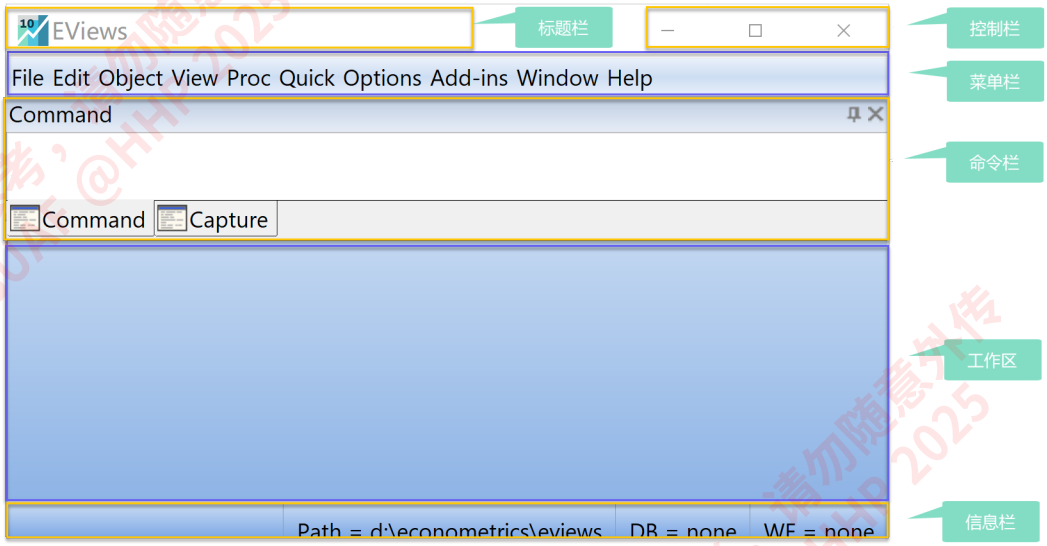


图 1.6: EViews 10 软件界面

其中界面分块情况如下：

- 标题栏：用于显示 EViews 任务的标题信息，并控制窗口状态。
- 菜单栏：菜单栏中排列着按照功能划分的 10 个主菜单选项，用鼠标单击任意选项会出现不同的下拉菜单，显示该部分的具体功能。
- 命令窗口：用户可在光标位置用键盘输入各种 Eviews 命名，并按回车键执行该命令。
- 工作区窗口：操作过程中打开的各子窗口在工作区内显示。
- 状态栏：Eviews 主窗口的底部是状态栏，从左到右分别为：信息框、路径框、当前数据库框和当前工作文件框。

1.7.2.2 创建工作文件（workfile）

进入 EViews 软件后的第一件工作应从创建新的或调入原有的工作文件（workfile）开始。只有新建或调入原有工作文件，EViews 才允许用户输入开始进行数据处理。

EViews 工作文件（workfile）的创建需要提前考虑如下几个方面：

（1）工作文件的结构类型（workfile structure type），一般包括截面数据、时间序列数据、面板数据三种类型

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 对于截面数据，可以选择非结构型/非日期型（Unstructured/Undated）
- 对于时间序列数据，可以选择有规则日期型（Data-regular frequency）
- 对于面板数据，可以选择平衡面板（Balanced Panel）

（2）数据长度或起止时间（data specification），依据数据类型而不同

- 对于截面数据，需要设置观测样本数（observations）
- 对于时间序列数据，需要设置数据频率（frequency）和起止时间（start date 和 end date）
- 对于面板数据，需要设置数据频率（frequency）、起止时间（start date 和 end date）和截面单元数（number of cross sections）

（3）工作文件的命名（workfile names），而且在一个工作文件下可以开展多个分析任务。不同分析任务可以设置不同的子页（page），并可以对子页进行命名以便区分

（4）工作文件的存放地址（path），可以决定存放在电脑本地的具体文件夹位置

此次实验演示的宏观经济发展案例为时间序列数据，我们将构建一个工作文件（workfile）和一个子页（page），具体步骤是（见图 1.7 和图 1.8 ）：

（1）启动 EViews 软件并新建 workfile - 在菜单上依次点击 ⇒ File⇒ New⇒ Workfile - 设置工作文件的结构类型（workfile structure type）- 点击选择有规则日期型（Data-regular frequency）

（2）设置数据长度或起止时间（data specification）- 设置数据频率（frequency）：点击选择季度（Quarterly）- 设置起始时间（start date）：手工输入 `1980/1` - 设置结束时间（end date）：手工输入 `2012/1`

（3）设置工作文件的命名（workfile names）- WF：手工输入 `macro_economic`（建议命名）- page：手工输入 `gdp`（建议命名）

（4）完成设置，查看工作文件，并保存到指定本地地址。

- ⇒ 点击 Ok。
- 此时可以看到已经建好的工作文件视窗。
- 保存工作文件到指定的本地地址（如 `D:\econometrics\eviews\macro_economic.wf1`）。
  - 方法 1（快捷保存）：在工作文件视窗下，同时按键盘 Ctrl+S ⇒ 选择本地路径地址 ⇒ 点击确认

– 方法 2（菜单保存）：在工作文件视窗下，点击菜单栏中的 Save ⇒ 选择本地路径地址 ⇒ 点击确认

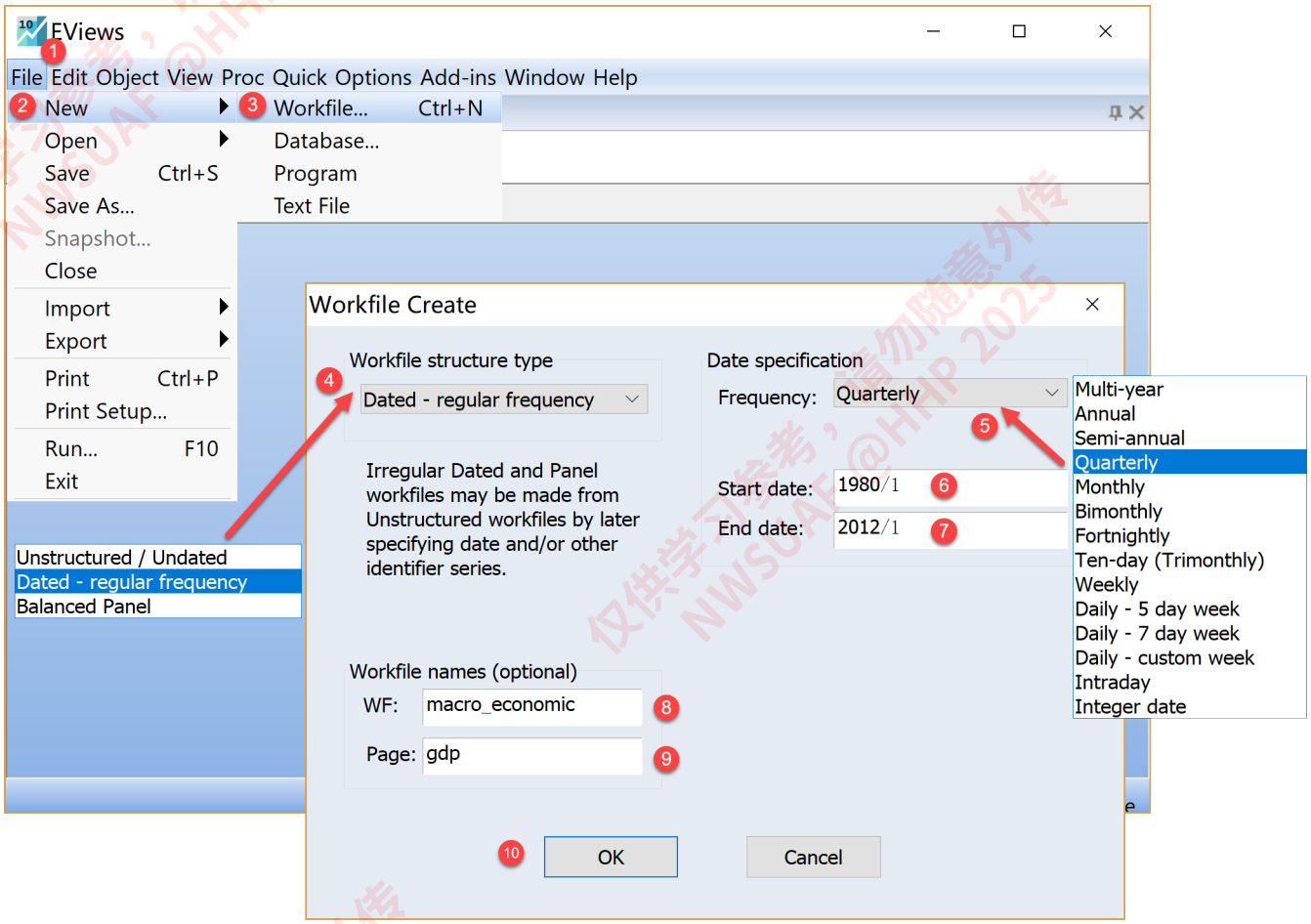


图 1.7: 创建宏观经济案例的工作文件

1.7.2.3 导入数据

建立或调入工作文件以后，可以输入和编辑数据。输入数据的基本方法有三种：

（1）方式 1：通过菜单操作导入（import）数据

- 直接导入（import）外部数据（最为常用），支持的外部数据文件格式包括：表格文件 excel（最为常用的.xls、.xlsx）；文本文件 text（较为常用的.txt、.csv、.dat）；其他文件 …。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

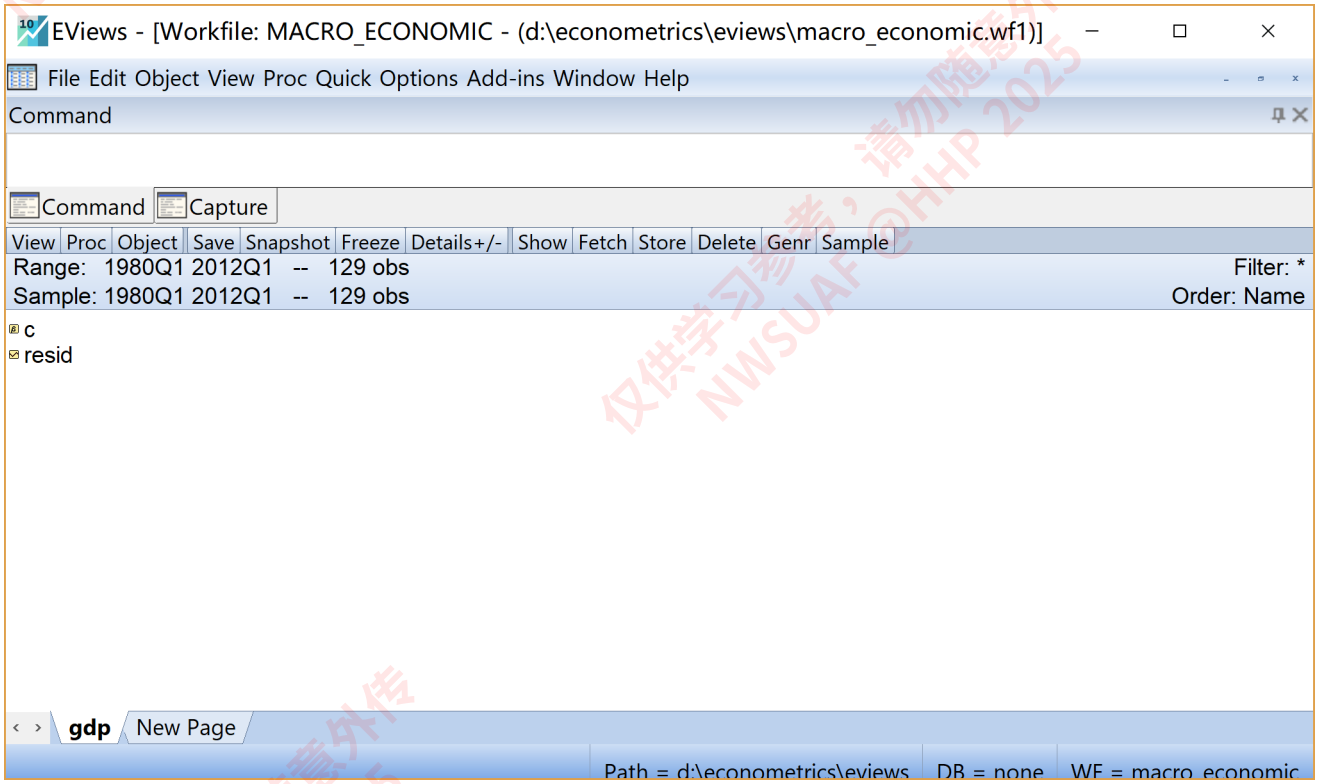


图 1.8: 宏观经济案例的工作文件视窗

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 从 EViews 数据库中抓取（fetch）数据（较为常用）
- 其他方式的导入 …

（2）方式 2：在工作文件下手工输入数据（较少使用）

- 手工一个一个数据输入（重复且单调）
- 复制粘贴数据，利用操作系统的剪贴板功能把外部数据复制并粘贴过来（对外部数据格式有一定要求，出错率较高）

（3）方式 3：在命令窗口下，使用 EViews 代码输入数据（较少使用，此处细节略过）

此次实验演示的宏观经济发展案例，我们采取方式 1，直接导入已经规范化的外部表格数据（.xlsx）文件，具体操作过程为：

（1）通过菜单操作读取外部文件（见图 1.9 ）：

- 在菜单上依次点击：⇒ File ⇒ Import ⇒ Import from file …
- 选择本地数据文件：选择文件路径 `d:/econometrics/eviews/macro_economic.xlsx`（根据文件位置而定）

（2）设置数据导入的引导界面：（如果数据足够规范，基本无需调整各项默认设置，见图 1.10 ）。

- step 1 of 4：设置并确认数据读取范围
- step 2 of 4：设置并确认变量名和行数据说明
- step 3 of 4：设置并确认缺失值和列数据说明
- step 4 of 4：设置并确认导入方法、选项、数据结构等
- 完成导入：点击 finish

（3）检查数据是否正确导入，以序列对象形式查看个别数据。在工作文件视窗下：

- 双击打开并查看序列对象gdp

（4）检查数据是否正确导入，以组对象形式查看全部数据（见图 1.11 ）。

- 选择 4 个主要变量。工作文件视窗中：按住键盘 CTRL 键，鼠标左键依次点击选择对象gdp、pce、investment、government

- 构造组对象。在工作文件的选定区（有底域阴影）上：
  - 点击右键 ⇒ 选择 `Open` ⇒ 选择 `as Group`
  - 得到未命名组对象UNTITLED
- 命名组对象并保存。在未命名组对象UNTITLED 视窗下
  - 点击菜单栏中的 `Name`
  - 输入命名 `group_vars` （建议命名）
  - 点击 `Ok`
- 查看组对象。在工作文件视窗中，双击查看已经命名的组对象group_vars

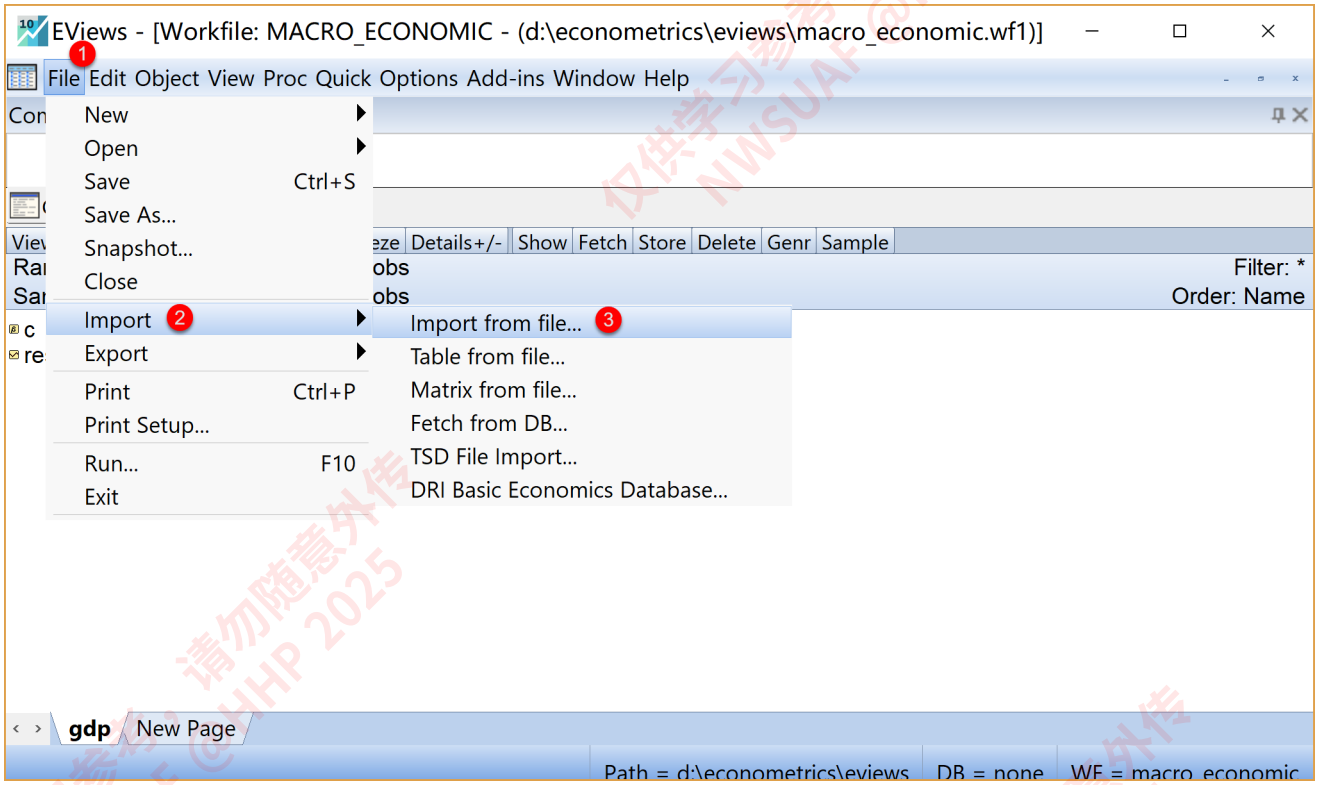


图 1.9: 导入外部 xlsx 数据文件 1：菜单

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

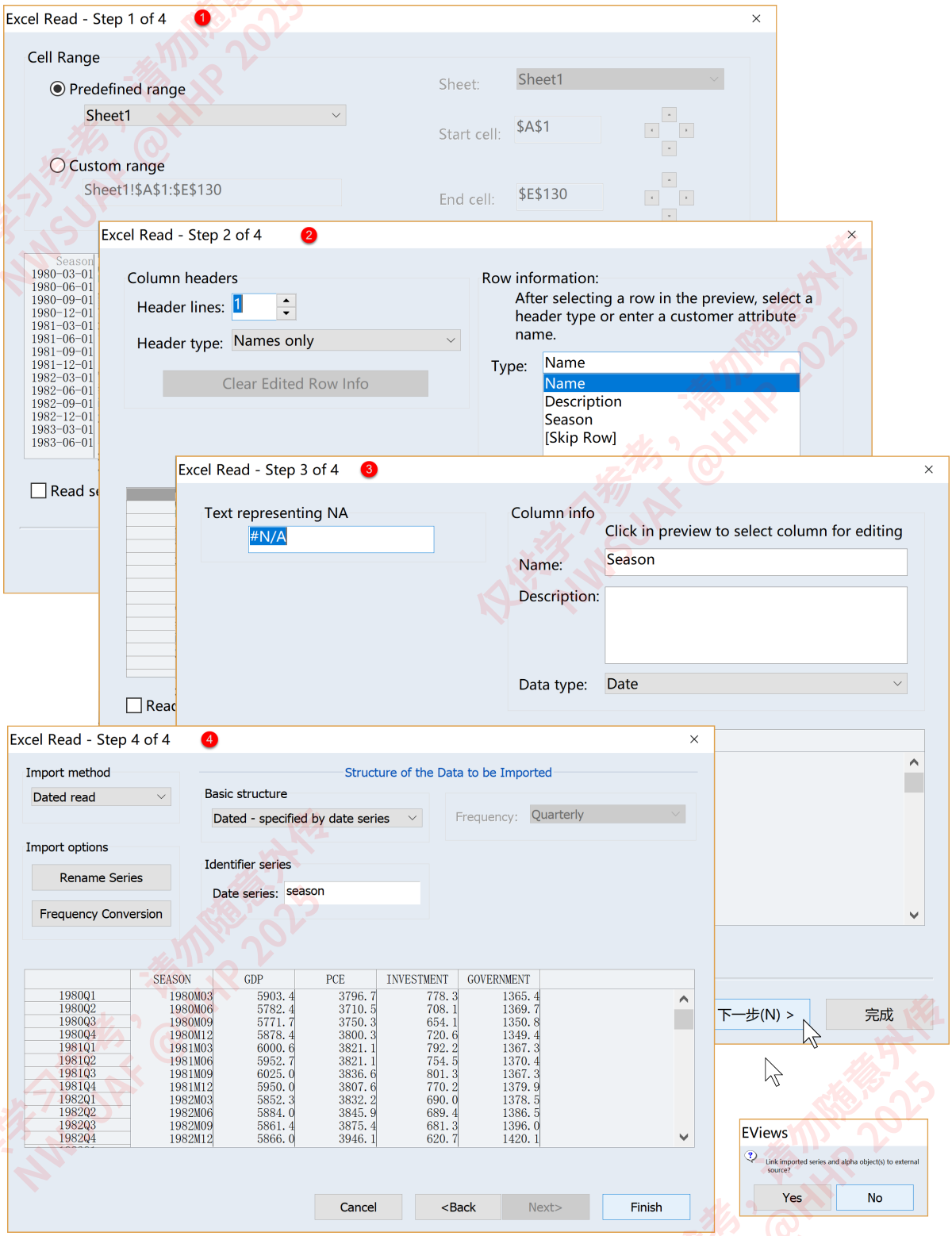


图 1.10: 导入外部 xlsx 数据文 2：引导

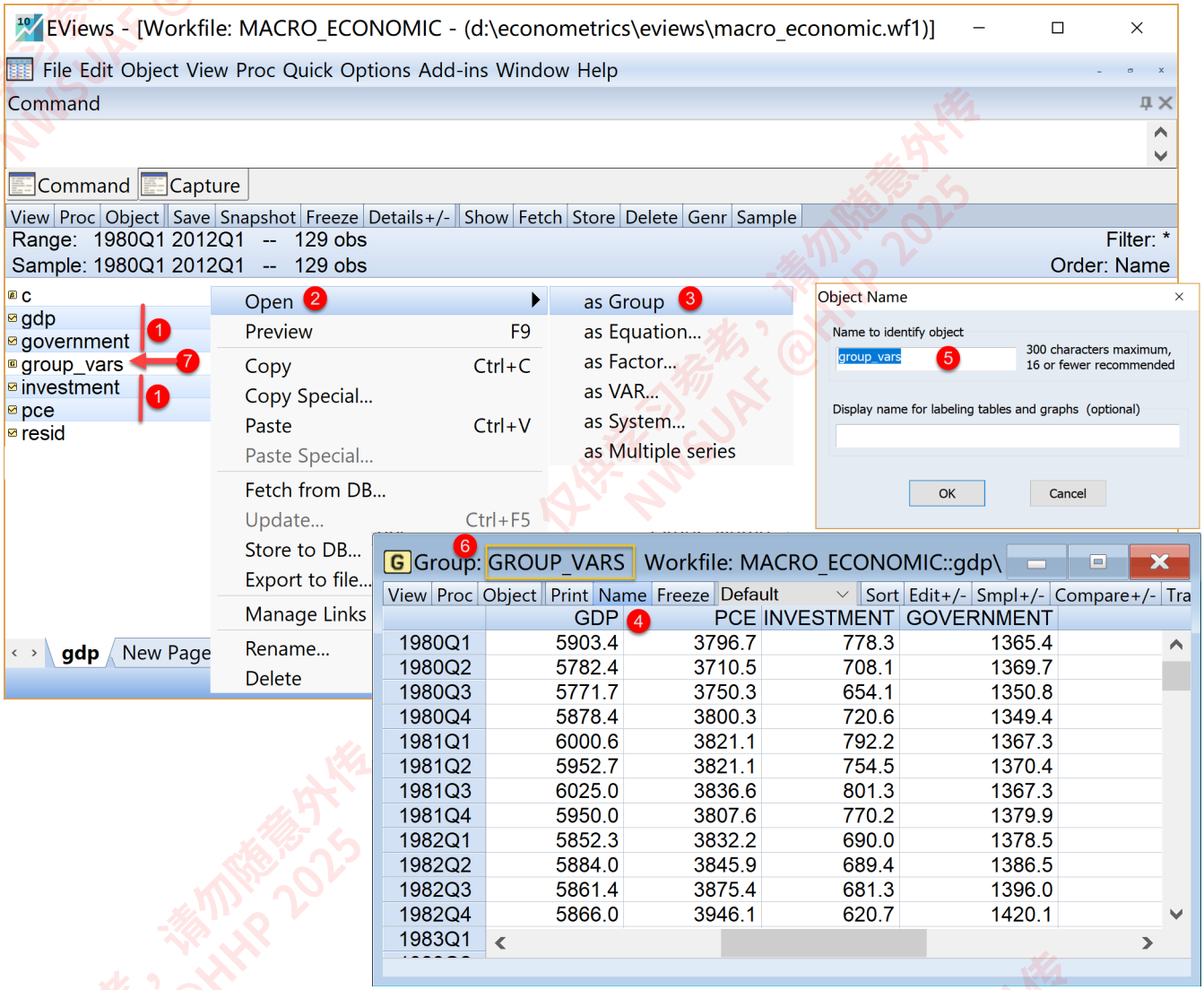


图 1.11: 导入外部 xlsx 数据文件 3：检查

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

1.7.3 构造对象（objects generating）

既然 EViews 是以对象（Object）方式来进行操作结果的管理和维护，那么所有的 EViews 操作过程几乎都会“凡走过必留下痕迹”，这些可见的“痕迹”就是 EViews 对象（object）。简言之，EViews 操作中的所有信息都存放在对象（object）里。同时，构造对象（objects generating）是一系列 EViews 操作过程的副产品：

- 随着分析操作环节的增加，各种对象都会相应产生，如表格（table）、图片（graph）、方程（equation）等。
- 如有必要，可能还会基于某个对象产生另一个新对象，如从方程（equation）对象中提取得到系数（beta）对象；或基于 X 变量的序列对象（series）生成新的 ln(X) 序列（series）对象。

1.7.3.1 EViews 对象（object）的“全家福”

对于 EViews 初学者，很有必要系统地了解 EViews 对象（object）都有哪些[3]。其中，尤其需要关注那些最为常用的（对于初级或中级使用者而言）EViews 对象：

表 1.7: 常用 EViews 对象列表及说明

star	icon	name_eng	name_chn	role
4*		Alpha	字母数字（序列型）	表达包含字母数字的变量
4*		Coef	系数（向量型）	表示方程 (Equation) 或系统 (System) 的参数
5*		Equation	方程	表表示方程的估计、检验或预测
5*		Graph	图形	表示图形输出结果
5*		Group	组	表示序列型对象 (包括 Series/ Alpha 等) 的一个集合体
4*		Matrix	矩阵	表示矩阵 (二维数组)
4*		Scalar	标量	表示标量 (也即一个数值)
5*		Series	序列（数值型）	表示时间序列变量，其元素需为数值
5*		Table	表格	表示表格输出结果

[3]可以参考在线帮助文档Object View And Procedure Reference

star	icon	name_eng	name_chn	role
4*		Vector	列向量（数值）	表示数值类型的列向量

其他一些不太常用（对于初级或中级使用者而言）EViews 对象包括：

表 1.8: 其他不太常用 EViews 对象列表及说明

star	icon	name_eng	name_chn	role
		Factor	因子	表示因子分析结果
		Logl	对数似然	表示对数似然分析结果
		Model	模型	表示联立方程的预测、模拟等
3*		Pool	数据池	表示面板数据 (同时含有时间和截面单元)
		Rowvector	行向量	表示行向量（一维数组）
3*		Sample	样本	表示观测样本的集合体
		Spool	对象池	表示包含各种输出对象的容器池
		Sspace	设置状态空间模型	表示状态空间模型分析结果
		String	字符串	表示标准字符串表达式 (不同于字符串值)
		Svector	列向量（字符串）	表示字符串类型的向量
2*		Sym	对称矩阵	表示对称矩阵
		System	系统	表示估计方程的系统 (与 Model 类似，但不能用于模拟仿真分析)
		Text	文本	表示任意文本信息
		Userobj	自定义对象	用于用户自主设置的对象
2*		Valmap	值映射	表示映射关系，给序列型对象 (Series/Alpha) 的值赋予标签
		Var	向量自回归	表示向量自回归和误差矫正模型分析结果

在宏观经济发展模型案例中，我们就可以看到这些常见 EViews 对象的容貌（参看节 1.6 中的快速案例，以及图 1.12 ）：

（1）序列（Series）对象：GDP、PCE、Government、Investment、…

（2）组（Group）对象：group_vars

（3）表格（Table）对象：tab_descriptive、tab_corrl

（4）图形（Graph）对象：graph_line

（5）方程（Equation）对象：eq_gdp

（6）系统保留对象（一旦建立工作文件，就会自动产生的对象）：

- 用来装载回归方程系数的系数（Coef）对象：C
- 用来装载回归方程残差的序列（Series）对象：resid

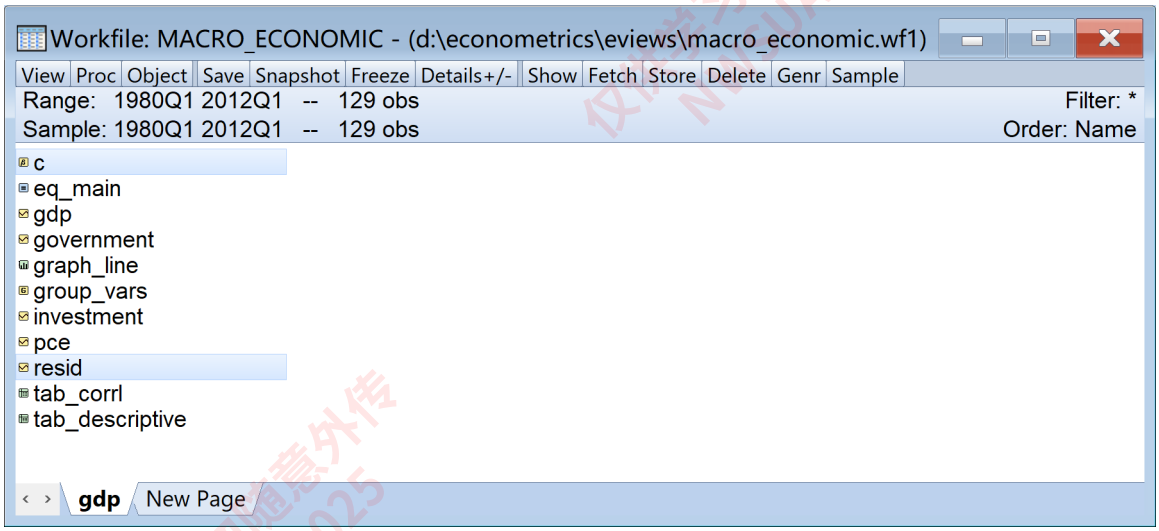


图 1.12: 快速案例中的 EViews 对象一览

1.7.3.2 EViews 分析中常用的对象构造情形

给定特定的数据分析案例，所有的 EViews 操作过程要么会直接产生特定对象，要么迫于分析工作要求而需要手动生成特定对象。这些常见的对象构造情形可以初步概括为：

第一，自动生成新对象。每一个具体的 EViews 分析操作，都会自动生成新对象。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 回归方程的分析操作，会直接生成一个方程对象。
- 探索性分析的绘图操作，会直接生成一个图形对象。
- 探索性分析的相关系数表计算，会生成得到一个表格对象。
- 数据的矩阵运算，会直接生成一个矩阵对象。
- …

第二，手动生成新对象。根据分析任务需要，生成某些有助于分析工作进行的新对象。

（1）通过功能菜单操作，直接生成新对象。

- 直接生成新对象是较为常用的，使用情形包括：直接手动构造一个新序列用于更灵活的建模分析；手动构造一个空白向量，用于后续提取并存放重要数值（如若干个对比模型的弹性系数）；…
- 例如，宏观经济发展案例中，在导入 Excel 数据后，我们可能还需要额外增加一个序列对象（series）时间趋势 `trend` 来调整样本回归模型式 1.2 ，其中变量 `trend` 的取值为 $trend \in \{0, 1, 2, \cdots, 128\}$。因为原数据表.xlsx 中并没有该数据，需要手动产生一个新的序列对象trend（变量名可以自己设置）

（2）通过某些形式的公式运算，生成新对象。

- 常用的公式运算情形包括:对数变换 `log()`;倒数变换 `1/X`;差分变换 `d()`;滞后变量运算 `x(-1)`;…
- 例如，宏观经济发展案例中，在导入 Excel 数据后，可能还需要使用对数化国内生产总值 `log(GDP)` 变量来调整样本回归模型式 1.2 或用于其他目的的分析。因此需要通过公式变换 $ln\_gdp = log(gdp)$，从而产生一个新的序列对象log_gdp（变量名可以自己设置）

（3）通过提取已存在对象的部分或全部信息，生成新对象。

- 常用的信息提取情形包括：提取回归报告中的某些系数; 提取方差协方差矩阵；…
- 例如，宏观经济发展案例中，为了在撰写报告时使用样本回归模型式 1.2 中的 * 政府支出的产出弹性，常常需要提取回归方程eq_gdp 中的系数值

对于大部分常用 Eviews 对象（如图 1.13 ）的创建工作，可以通过菜单操作来实现：

（1）打开对象创建引导界面：

- 依次点击 ⇒ Object ⇒ New object…
- 选择并设置对象：
- 手动选择对象类型：复选框（Type of object）选择
- 命名对象：文本框（Name for object）输入
- 完成操作：点击 OK

（2）编辑对象数值或属性

（3）工作文件窗口中查看结果

下面将展示宏观经济发展案例中，新增创建一个序列对象（series）时间趋势 `trend`（取值为 $trend \in \{0, 1, 2, \cdots, 128\}$）的基本步骤（见图 1.13 ）。

（1）打开对象创建引导界面。在工作视窗下：

- 依次点击菜单栏，⇒ Object ⇒ New object…

（2）选择并设置对象。在对象创建引导视图下：

- 手动选择对象类型：复选框（Type of object）选择 `Series`
- 命名对象：文本框（Name for object）输入 `trend`
- 完成操作：点击 OK

（3）查看并编辑对象数值。在序列对象视窗下：

- 点击菜单栏中的 Edit+/-
- 双击单元格
- 手动填写或编辑数值
- 保存编辑操作，按键盘快捷组合键 Ctrl+S

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

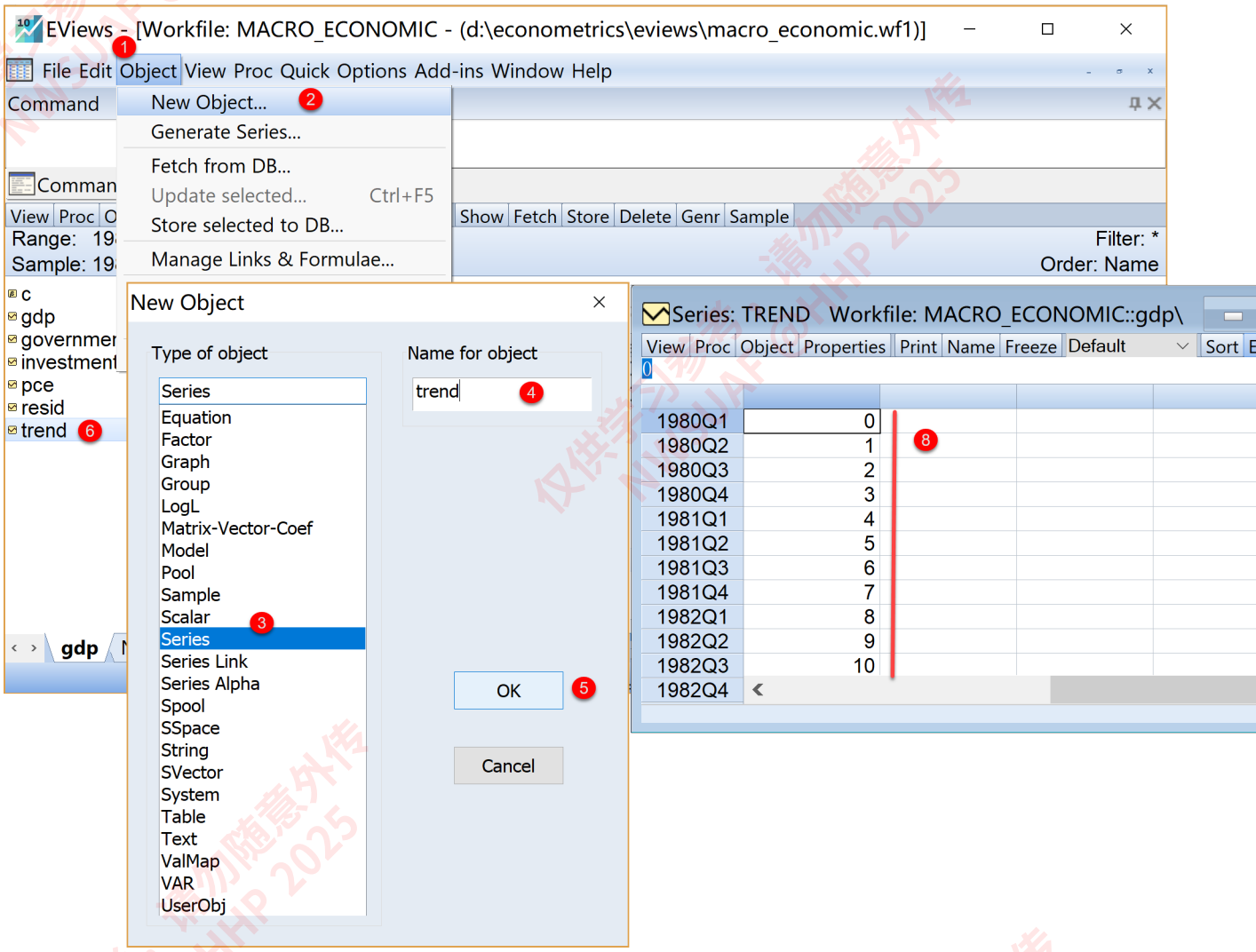


图 1.13: 常用 Eviews 对象的菜单式创建操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

1.7.4 探索性分析 (exploratory analyzing)

探索性分析（exploratory analyzing）是借助统计学基本原理对研究问题进行探索性分析。分析用到的一些统计学基本概念包括均值、中位数、众数、极大值、极小值、标准差、正态性、偏度、峰度、显著度等。探索性分析一般以形象化的数据可视化（data visualization）方式呈现数据特征，主要是通过表格（table）或图形（graph）的形式展现变量之间的特征、趋势、和关系。

1.7.4.1 图表可视化的一般性原则

常用的表格形式包括：描述性统计表、相关系数表等。

- 绘制表格主要的目的是展示确定的数值大小、汇总统计关系等
- 原始数据可以构造形成一张表格
- 计算得到的统计量也可以构造形成一张表格
- 特定的分析操作，往往意味着可以得到特定的表格。例如，回归分析能得到回归分析报告表；自相关分析就能得到自相关表；…

常用的图形形式包括：线形图（line）、描点图（dot）、柱状图（bar）、散点图（scatter）、箱线图（boxplot）等。图形可视化的关键在于——以合适的形式，合理地呈现必要的信息!。因此，选择合适的图形形式至为重要，而数据的类型、变量的多少等都会影响图形类型的选择。

图形设计需要考虑的一些基本原则有：图中要展示多少个变量？每个变量要展示多少数据点？数据点是否以时间、项目或分组的方式呈现？

根据不同的图形绘制目标，具体图形选择应视需要而定（见表 1.9 ）：

- 分析数据比较（Data Comparison）时，可选择循环面积图 (Circular Area Chart)、折线图 (Line Chart)、柱状图 (Column Chart)、簇状-折线图 (Line Chart)、不等宽柱状图 (Variable Width Column Chart)、表格型嵌图 (Table with Embedded Charts)、簇状-条形图 (Bar Chart)、簇状-柱状图 (Column Chart) 等图形
- 分析数据构成（Data Composition）时，可选择饼形图 (Pie Chart)、瀑布图 (Waterfall Chart)、百分比多层构成图 (Components of Components)、堆栈百分比柱状图 (Stacked 100% column chart)、堆栈柱状图 (Stacked Column Chart)、堆栈百分比面积图 (Stacked 100% Area Chart)、堆栈面积图 (Stacked Area Chart) 等图形

- 分析数据分布（Data Distribution）时，可选择柱状直方图 (Column Histogram)、线状直方图 (Line Histogram)、散点图 (Scatter Chart) 等图形
- 分析数据关系（Data Relationship）时，可选择 3D 面积图 (3D Area Chart)、散点图 (Scatter Chart)、泡泡图 (Bubble chart) 等图形

表 1.9: 探索数据比较的图形选择

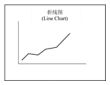
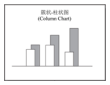
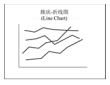
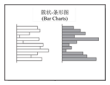
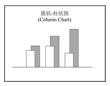
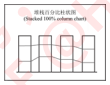
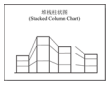
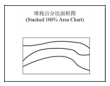
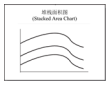
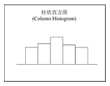
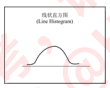
id	purpose	icon	name	characters
1	数据比较		循环面积图 (Circular Area Chart)	时期间比较-> 时期数较多-> 周期性时期
2	数据比较		折线图 (Line Chart)	时期间比较-> 时期数较多-> 非周期性时期
3	数据比较		柱状图 (Column Chart)	时期间比较-> 时期数较少-> 分组较少
4	数据比较		簇状-折线图 (Line Chart)	时期间比较-> 时期数较少-> 分组较多
5	数据比较		不等宽柱状图 (Variable Width Column Chart)	条目间比较-> 每个条目有两个变量->

表 1.9: 探索数据比较的图形选择

id	purpose	icon	name	characters
6	数据比较		表格型嵌图 (Table with Embedded Charts)	条目间比较-> 每个条目有一个变量-> 分组较多
7	数据比较		簇状-条形图 (Bar Chart)	条目间比较-> 每个条目有一个变量-> 分组较少-> 条目较多
8	数据比较		簇状-柱状图 (Column Chart)	条目间比较-> 每个条目有一个变量-> 分组较少-> 条目较少
1	数据构成		饼形图 (Pie Chart)	构成无时期变化-> 简单占比构成
2	数据构成		瀑布图 (Waterfall Chart)	构成无时期变化-> 累积构成或落差构成
3	数据构成		百分比多层构成图 (Components of Components)	构成无时期变化-> 多层次子类的百分比构成

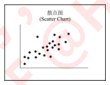
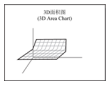
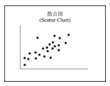
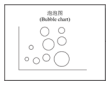
仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 1.9: 探索数据比较的图形选择

id	purpose	icon	name	characters
4	数据构成		堆栈百分比柱状图 (Stacked 100% column chart)	构成有时期变化-> 时期较少-> 仅关心成分之间的相对差异
5	数据构成		堆栈柱状图 (Stacked Column Chart)	构成有时期变化-> 时期较少-> 仅关心成分之间的绝对差异
6	数据构成		堆栈百分比面积图 (Stacked 100% Area Chart)	构成有时期变化-> 时期较多-> 仅关心成分之间的相对差异
7	数据构成		堆栈面积图 (Stacked Area Chart)	构成有时期变化-> 时期较多-> 仅关心成分之间的绝对差异
1	数据分布		柱状直方图 (Column Histogram)	一个变量的分布-> 数据点较少
2	数据分布		线状直方图 (Line Histogram)	一个变量的分布-> 数据点较多

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 1.9: 探索数据比较的图形选择

id	purpose	icon	name	characters
3	数据分布		散点图 (Scatter Chart)	两个变量的分布
1	数据关系		3D 面积图 (3D Area Chart)	三个变量间的分布
2	数据关系		散点图 (Scatter Chart)	两个变量间的关系
3	数据关系		泡泡图 (Bubble chart)	三个变量间的关系

1.7.4.2 EViews 图表可视化工具集

EViews 软件中的表格可视化，大多数“隐藏”在每一项分析操作的背后。这就意味着：表格可视化应该更加重视特定的数据分析操作（例如序列自相关分析、或者异方差分析等）；然后，根据分析结果，获得一份粗糙的原始表格；最后，根据需要进行表格的细节修改和完善。

与表格可视化不同，EViews 软件提供了丰富的图形可视化工具箱（见表 1.10 ），与上述图形选择的一般性原则一致，需要综合考虑图形绘制目的、变量数据特征（变量个数、变量类型、数据分组、数据趋势、时期关系 …）等。

EViews 软件中图形可视化的基本操作过程包括：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

（1）选择变量数据（包括选择变量、数据分组、样本区间等）

- 一般是选择工作文件中的组（Group）对象或序列（Series）对象

（2）选择图形类型（选定什么样的数据，就会给出怎样的工具箱）

- 根据选择变量数据的不同，图形工具箱的界面会有差异
- EViews 给出可选的图形包括：饼图 (Pie)、不等宽柱状图 (XY Bar)、带形图 (Area Band)、点图 (Dot Plot)、钉形图 (Spike)、二维面积图 (XY Area)、高低图 (High-Low)、混合图 (Mixed)、季节图 (Seasonal Graph)、路径图 (XY Line)、密度图 (Distribution)、面积图 (Area)、泡泡图 (Bubble Plot)、散点图 (Scatter)、条形图 (Bar)、误差图 (Error Plot)、线图 (line & Symbol)、箱线图 (Boxplot)、正态图 (Quantile-Quantile) 等

（3）设置图形细节

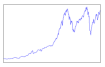
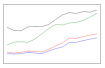
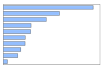
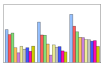
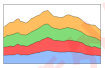
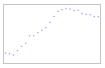
- 设置图形的颜色、尺寸、边框等
- 设置图形的坐标轴、网格线等
- 设置图例
- 设置图形要素
- 设置字体
- ……

（4）保存或输出图形对象

- 命名图形对象
- 保存到本地文件或数据库
- 导出为其他格式（.png/.jpg/.pdf/…）

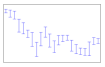
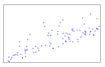
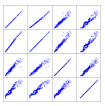
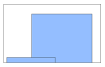
仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 1.10: EViews 图形工具箱及示例

id	name_eng	name_chn	icon	mark
1	line & Symbol	线图		表现变化趋势，包含 1 个序列对象，数据点较多
1	line & Symbol	线图		表现变化趋势，包含 4 个序列对象，数据点较多
2	Bar	条形图		表现分组差异，水平条形，包含 1 个序列对象，数据分组较少
2	Bar	条形图		表现分组差异，垂直条形，包含 1 个序列对象，数据分组较少
2	Bar	条形图		表现分组差异，垂直条形，包含若干序列对象，数据分组较少
3	Spike	钉形图		表现数值大小，包含 1 个序列对象，数据分组较少
4	Area	面积图		表现覆盖面积，包含 1 个序列对象，数据点较多
4	Area	面积图		表现覆盖面积，包含 4 个序列对象，数据点较多
5	Area Band	带形图		表现面积差异，包含 2 个序列对象，数据点较多
6	Mixed	混合图		表现多样信息，包含多个序列对象，数据点较多
7	Dot Plot	点图		表现数值大小，包含 1 个序列对象，数据点较多

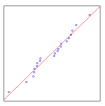
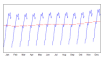
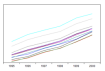
仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 1.10: EViews 图形工具箱及示例

id	name_eng	name_chn	icon	mark
8	Error Plot	误差图		表现数值差异，包含 2 个序列对象，数据点较少
8	Error Plot	误差图		表现数值差异，包含 3 个序列对象，数据点较少
9	High-Low	高低图		表现数值位置关系，包含 2 个序列对象（最高价、最低价），数据点较少
9	High-Low	高低图		表现数值位置关系，包含 3 个序列对象（开盘价、最高价、最低价），数据点较少
10	Scatter	散点图		表现变量关系，包含 2 个序列对象，数据点较多
10	Scatter	散点图		表现变量关系，包含多个序列对象，数据点较多
11	Bubble Plot	泡泡图		表现变量关系，包含 3 个序列对象，数据点较多
12	XY Line	路径图		表现数据变化路径，包含 2 个序列对象，数据点较多
13	XY Area	二维面积图		表现 2D 覆盖面积，包含个序列对象，数据点较多
14	XY Bar	不等宽柱状图		表现分组差异，垂直条形，包含若干个序列对象，数据分组较少

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 1.10: EViews 图形工具箱及示例

id	name_eng	name_cn	icon	mark
15	Pie	饼图		表现数据构成关系，包含 1 个序列对象，数据分组较少
16	Distribution	密度图		表现数据分布密度，直方图，包含 1 个序列对象，数据点较多（自动分组）
17	Quantile-Quantile	正态图		表现数据分布密度，正态 QQ 图，包含 1 个序列对象，数据点较多
18	Boxplot	箱线图		表现数据分布形态，包含若干序列对象，数据点较多
19	Seasonal Graph	季节图		表现数据循环关系，包含 1 个序列对象，数据点较多，分组较少（在同一月份不同年份上的变化）
19	Seasonal Graph	季节图		表现数据循环关系，包含 1 个序列对象，数据点较多，分组较少（在同一年份不同月份上的变化）

下面将利用 EViews 菜单操作，对宏观经济发展案例进行简要的探索性分析。该部分内容将有助于快速了解 EViews 图表可视化的基本操作思路和流程。无论是表格可视化，还是图形可视化，如下的简要案例操作及过程都具有普遍的代表性，读者可以照此举一反三，而要达到熟练地推而广之则必须经历更多的实战摸索和训练。

1.7.4.3 EViews 表格可视化：描述性统计分析表的制作

首先，对宏观经济发展案例中的主要变量（GDP 国内生产总值（10 亿美元）、PCE 人均消费支出（美元）、Investment 投资（10 亿美元）、Government 政府支出（10 亿美元））进行基本描述性统计分析，并得到表格（table）对象形式的描述性统计分析表。然后，对表格（table）对象形式的描述性统计分析表进行美学设计，修改表格细节，实现指定要求的三线表形式。

进行基本描述性统计分析（EViews 操步作骤见图 1.14 ）：

（1）在工作文件视窗中，单击选择已经命名的组对象group_vars

（2）在工作文件视窗中，点击菜单栏中的 Quick ⇒ 选择 Group Statistics ⇒ 选择 Descriptive Statistics ⇒ 选择 Common sample

（3）确认变量对象，⇒ 点击 Ok

（4）得到 Group 描述性统计分析视窗group_vars

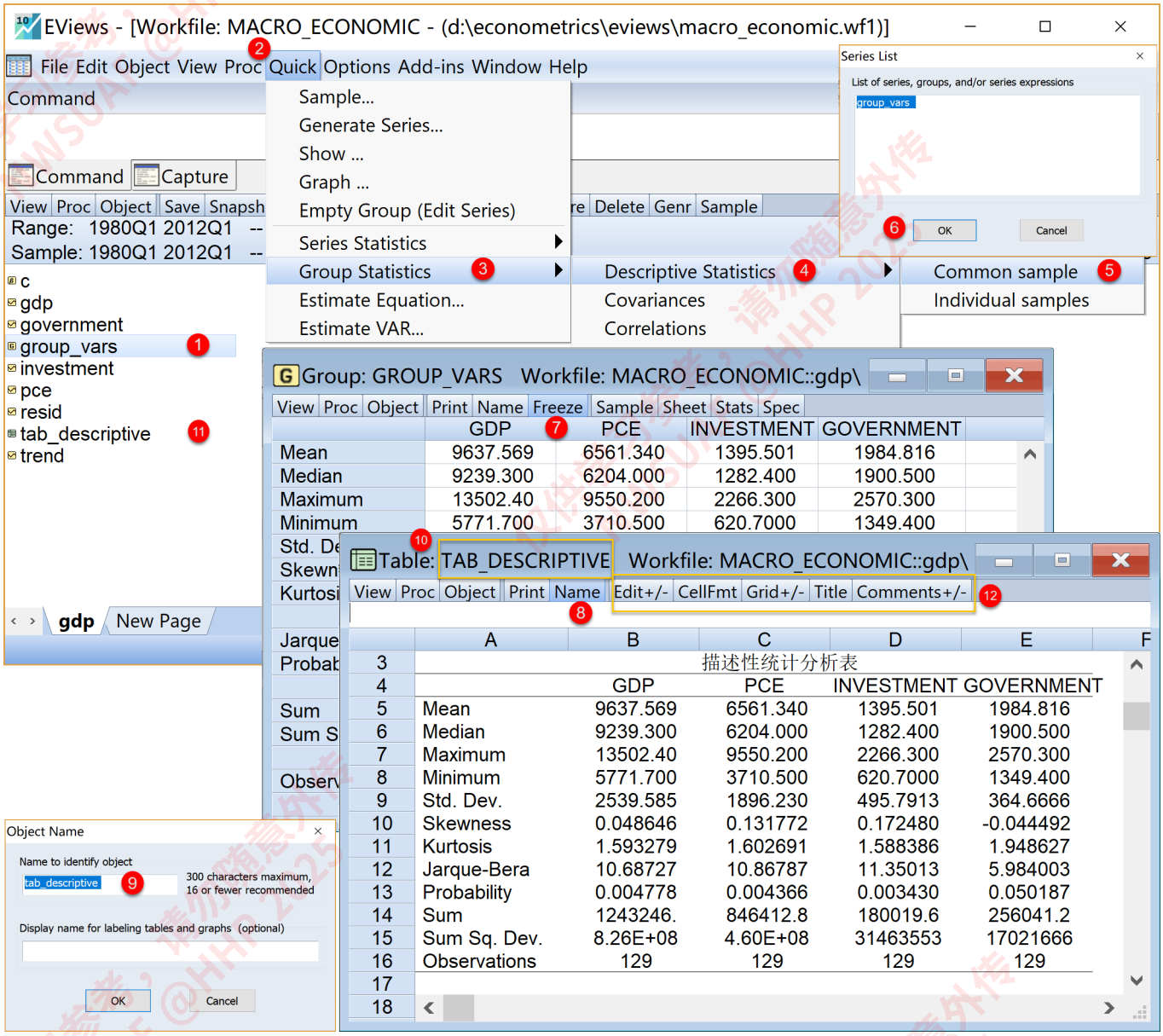
（5）将描述性分析结果转换成表格。在 Group 描述性统计分析视窗group_vars 下，⇒ 点击 Freeze⇒ 得到表格对象视窗UNTITLED

- 给表格对象命名。在表格对象视窗UNTITLED 视窗下，⇒ 点击 Name
- 填写表格对象名称。在 Object Name 窗口中，⇒ 输入 tab_descriptive（建议命名）⇒ 点击 Ok
- 查看保存的表格对象。在工作文件视窗中，确认可以看到已经保存的表格对象tab_descriptive

进行美学设计，修改表格细节，实现三线表形式的表格要求。（此处要求把初始表格调整为三线表，具体细节修改过程略。）。

- 工作文件视窗中，双击打开表格对象tab_descriptive
- 点击菜单 Edit+/-，可以编辑修改单元格的内容
- 点击菜单 CellFmt，可以调整单元格的大小、颜色、边框等
- 点击菜单 Title，可以填写表格标题
- 点击菜单 Grid+/-，可以显示/隐藏网格线
- ...

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025



	GDP	PCE	INVESTMENT	GOVERNMENT
Mean	9637.569	6561.340	1395.501	1984.816
Median	9239.300	6204.000	1282.400	1900.500
Maximum	13502.40	9550.200	2266.300	2570.300
Minimum	5771.700	3710.500	620.7000	1349.400
Std. Dev.	2539.585	1896.230	495.7913	364.6666
Skewness	0.048646	0.131772	0.172480	-0.044492
Kurtosis	1.593279	1.602691	1.588386	1.948627
Jarque-Bera	10.68727	10.86787	11.35013	5.984003
Probability	0.004778	0.004366	0.003430	0.050187
Sum	1243246.	846412.8	180019.6	256041.2
Sum Sq. Dev.	8.26E+08	4.60E+08	31463553	17021666
Observations	129	129	129	129

图 1.14: 构造主要变量的描述性统计分析表（三线表）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

1.7.4.4 EViews 图形可视化：多线图（Line-Symbol）的绘制

下面将展示 EViews 软件在多变量下的线图（Line-Symbol）的绘制过程。对宏观经济发展案例中的主要变量（GDP 国内生产总值（10 亿美元）、PCE 人均消费支出（美元）、Investment 投资（10 亿美元）、Government 政府支出（10 亿美元））进行简单的图形可视化，绘制简单的（标准化的）多变量线图（Line-Symbol）的主要过程如下：（见图 1.15 ）

选择组（Group）对象。在工作文件视窗下，

- 鼠标单击选择已经命名的组对象group_vars。
- 点击菜单栏中的 `Quick` ⇒ 选择 `Graph`
- 确认变量对象，点击 `Ok`

设置图形引导界面。在引导窗口的目录 `Graph Options` 中，分别点击设置如下子选项页：

（1）Graph type（图类型）子选项页，主要用于修改图形类型

- General（通用）下拉框：下拉选择 Basic graph
- Specific（设置）复选框（比较常用的选项设置）：点击选择 Line-Symbol
- Detail（细节）选项区，有多个设置下拉框选择

> 提示
>
> - Graph data（图形数据）下拉框：默认设置
> - Orientation（方向）下拉框：默认设置
> - Axis Borders（坐标轴边框）下拉框：默认设置
> - Multiple series（多序列）下拉框：下拉选择 Single graph

（2）Frame & Size（图框和大小）子选项页，主要用于设置图形的颜色/边框、大小/缩进等：默认设置

（3）Axes & Scaling（图轴和度量）子选项页，主要用于设置数据尺度、数轴标签、网格线等：默认设置

（4）Lgend（图例）子选项，主要用于设置图例相关选项：默认设置

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

（5）Graph Element（图形元素）子选项，主要用于设置线型/符号、填充块等：默认设置

（6）Quick Fonts（快速字体）子选项页，用于调整字体：默认设置

（7）Templates & Objects（样式和对象），用于套用样式等：默认设置

（8）Graph Updating（图形更新）子选项页，用于更新修改：默认设置

（9）完成设置。点击 `Ok`。将得到未命名图形对象UNTITLED

命名并保存图形对象。在未命名图形对象UNTITLED 窗口中

- 点击菜单栏 `Name` ⇒ 输入命名 `graph_line` ⇒ 点击 `Ok`
- 可以看到工作文件中，将出现已命名并保存的图形对象graph_line

此外，根据研究报告撰写的特定要求，可以对图形进行个性化美学设计。美化图形，修改细节，使其满足个性化美学要求（见图 1.15 ，具体步骤此处略）。

（1）工作文件中，双击打开已保存的图形对象graph_line

（2）操作并编辑各种图表细节，在图形对象graph_line 视窗中：

- 点击菜单 `name`，可以修改图形对象命名
- 点击菜单 `Freeze`，可以设置图形是否冻结或可更新（如手工调整样本区间等）
- 点击菜单 `Options`，可以修改各种图形元素（这是最常用的细节修改内容）

提示

- `Graph Type`，用于修改图形类型
- `Frame & Size`，用于修改图形大小、边框等
- `Axes & Scaling`，用于修改图形坐标轴、网格线、标尺等
- `Legend`，用于修改图形的图例
- `Graph Elements`，用于修改图形的数值标签、线型、符号等
- `Quick Font`，用于修改图形的字体
- `Templates & Objects`，用于快速套用图形样式
- `Graph Updating`，用于图形修改更新

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

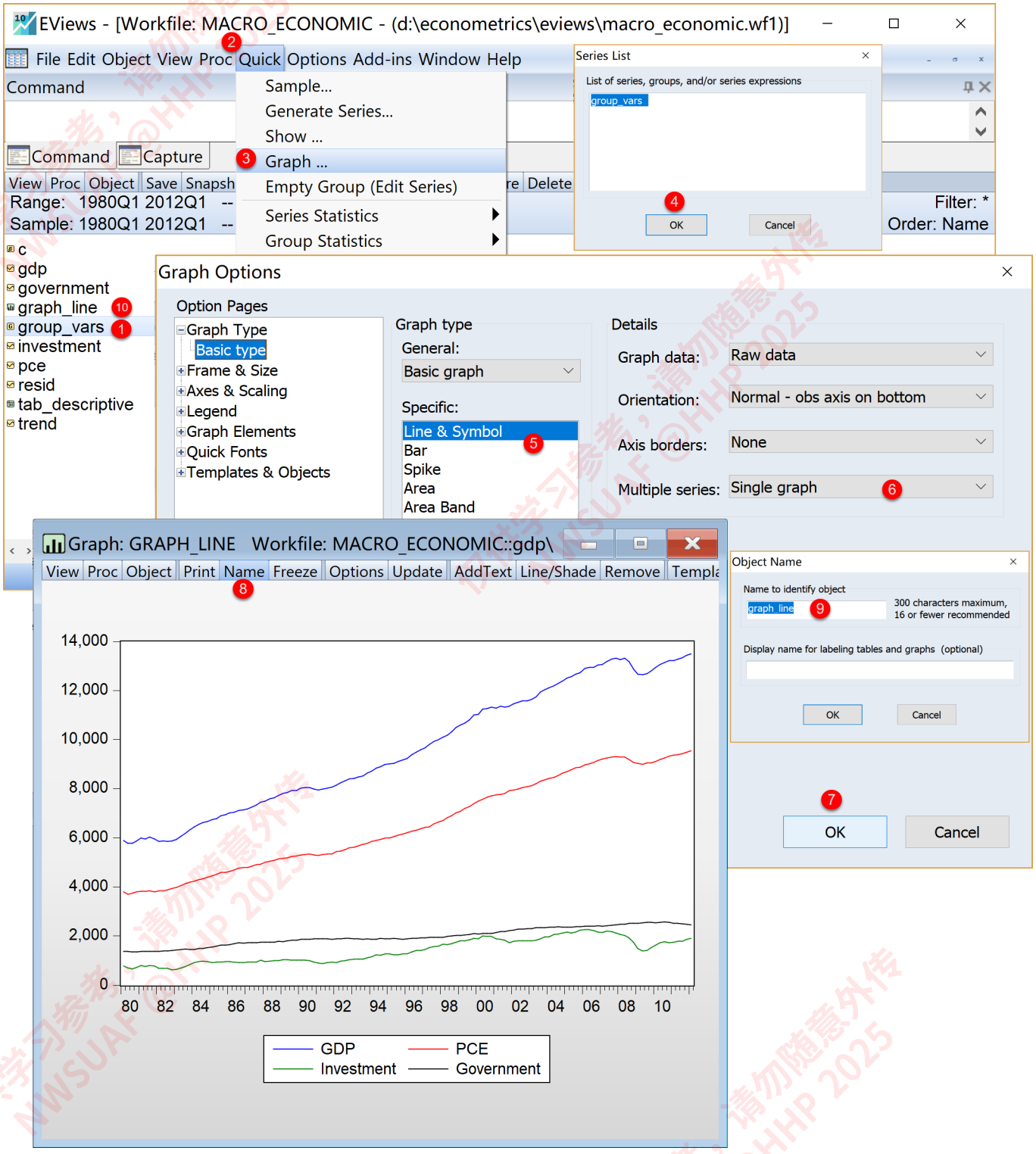


图 1.15: 主要变量的简单多线图（Line-Symbol）绘制

- 点击菜单 `Update`，可以更新图形修改结果
- 点击菜单 `AddTex`，可以给图形添加文本
- …

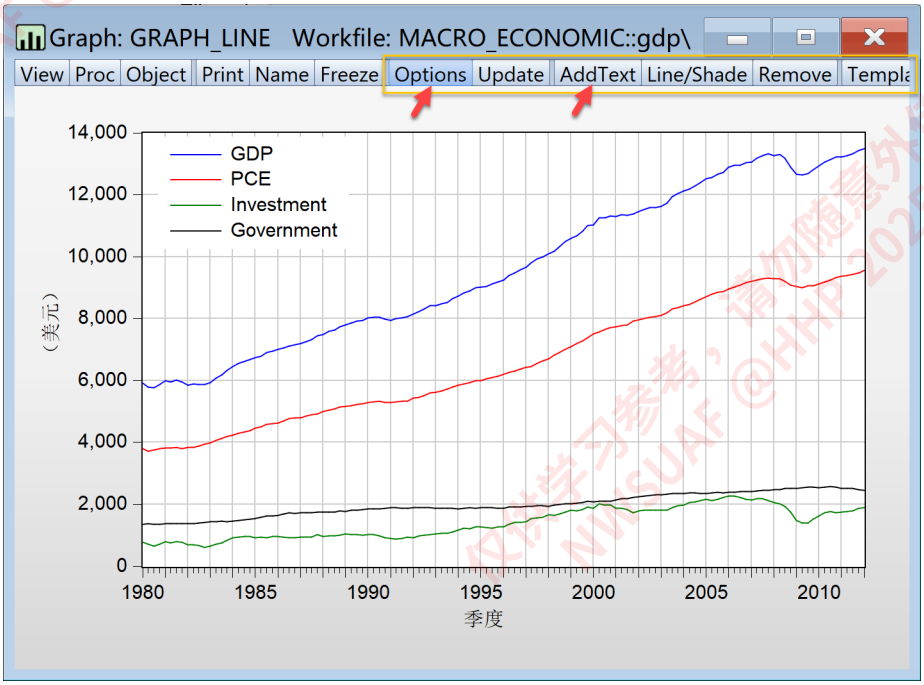


图 1.16: 主要变量的个性化多线图（Line Symbol）绘制

1.7.5 线性回归分析（linear regression analyzing）

线性回归分析（linear regression analyzing）是用来探究现象之间因果作用关系的最重要工具之一。线性回归分析属于一类有助于直观地理解复杂问题的理想化分析工具，一般需要较严苛的假设条件。

EViews 进行线性回归分析中需要考虑的问题可繁可简，取决于问题的复杂程度、研究的细致程度、条件的满足程度等等。一般而言，线性回归分析都需要进行：模型选择和调试、统计学检验（如 t 检验、F 检验等）、计量经济学检验（如多重共线性、异方差、自相关等检验）、预测性检验、稳健性检验等。

为了简便起见，下面将对宏观经济发展案例中的对数模型（见节节 1.6 中的样本回归模型式 1.2 ）进行 EViews 线性回归分析。Eviews 菜单操作的具体步骤为（见图 1.17 ）：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

（1）进入方程分析的设置菜单。

- 在工作文件视窗下，点击菜单栏中的 Quick ⇒ Estimation Equation ⇒ 将看到方程估计的引导界面

（2）完成引导设置。方程估计的引导界面 Equation Estimation 下，进行如下设置：

提示

（a）设置方程。Equation Estimation ⇒ Specification

- 方程设置 Equation specification：依次输入变量 gdp c pce investment government（注意变量之间的空格，以及截距 c）
- 估计方法 Estimation settings：
  - 方法 Method：下拉选择 LS - Least Squares (NLS and ARMA)
  - 样本 Sample：默认设置

（b）设置细节选项。Equation Estimation ⇒ Options

- 协方差设置 Coefficient Covariance ⇒ （细节略）
- 权重设置 Weights ⇒ （细节略）
- 优化算子设置 Optimization ⇒ （细节略）

（c）完成设置，点击 OK ⇒ 将得到未命名的方程对象UNTITLED

（3）命名并保存回归方程。

- 在未命名的方程对象UNTITLED 视窗下，点击菜单栏 Name ⇒ 输入命名 eq_main（建议命名）⇒ 完成命名，点击 Ok
- 此时，在工作文件视窗下可以看到已经保存命名的方程对象eq_main

提示

- 构建并比较多个类似方程时，可以复制、粘贴、重命名多个方程对象
- 可以随时修改方程的属性，在方程对象视窗下，点击菜单栏中的 Estimate ⇒ 重新设

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

置 Equation Estimation

1.7.6 报告结果输出（output reporting）

Eviews 操作结果是以对象（object）形式存储在工作文件（workfile）中，在撰写研究报告时，往往需要呈现这些分析结果和主要结论。报告结果输出（output reporting）是指把 EViews 分析的结果导出或呈现在研究报告中的过程或操作。根据研究报告内容的实际需要，EViews 的对象可能需要完整输出，也可能只是提取对象中的部分元素。但无论如何，进行合理的选取、整合、呈现、展示是极为必要的。例如：

- 调整图形（graph）和表格（table）的各种美学细节以符合论文或报告的格式要求，如字体、颜色、线条、形状、图例、坐标等
- 选取并整合成新的表格，如三线表、多模型对比表
- 把一个 EViews 原始的回归分析报告（raw report），处理成简要报告（simple report），一般表达为三行或四行，包括样本回归方程（第一行）、回归系数的标准误（第二行）、t 值（第三行）、拟合情况（第四行，包括拟合优度 R^2，F 检验值、DW 检验值等）
- …

对于如下的回归方程：

$$GDP_i = + \beta_1 + \beta_2 PCE_i + \beta_3 Government_i + \beta_4 Investment_i + u_i \tag{1.3}$$

其回归报告，可以简要写成：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

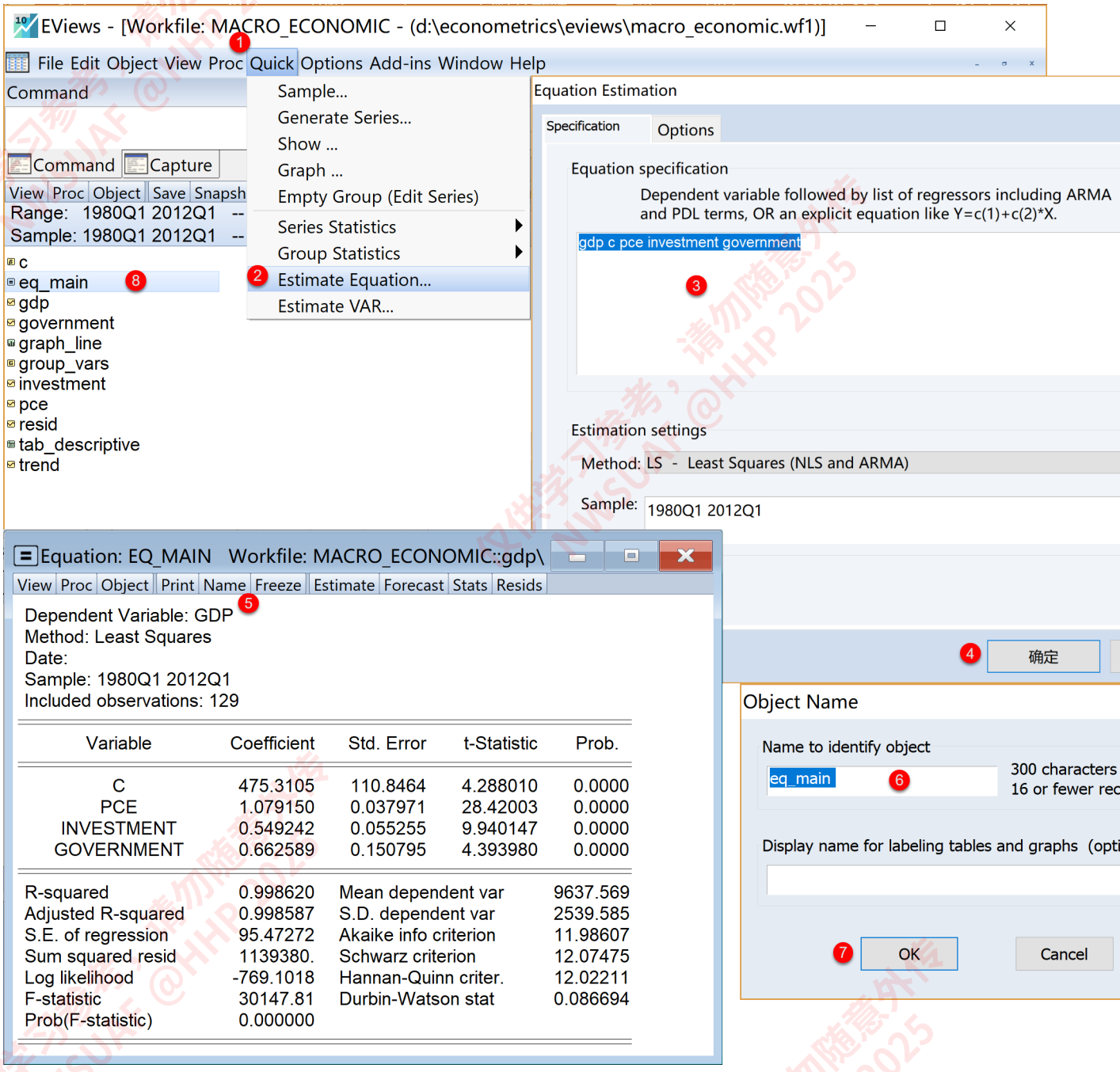


图 1.17: 主回归模型 Eviews 操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{array}{lll}
\widehat{GDP} = & +475.31 & +1.08PCE_i \\
(s) & (110.8464) & (0.0380) \\
(t) & (+4.29) & (+28.42) \\
(cont.) & +0.66Government_i & +0.55Investment_i \\
(s) & (0.1508) & (0.0553) \\
(t) & (+4.39) & (+9.94) \\
(over) & n = 129 & \hat{\sigma} = 95.4727 \\
(fit) & R^2 = 0.9986 & \bar{R}^2 = 0.9986 \\
(Ftest) & F^* = 30147.81 & p = 0.0000
\end{array} \tag{1.4}
$$

1.8 Eviews 编程体系及语法

1.8.1 命令运行方式

EViews 提供了菜单和命令两种方式进行数据分析和处理。事实上，只要能够通过窗口菜单完成的功能，就都可以通过命令窗口中输入相应编程命令来执行实现。根据任务的复杂性、重复性、繁重性等的不同，EViews 可以通过如下方式完成分析任务：

（1）方式 1：完全窗口菜单驱动，也即通过选择特定菜单栏功能进行分析操作

- 适合于简单的、一次性的、工作量小的分析任务
- 适合于 EViews 软件初学者，可以更加集中于统计分析和计量分析本身

> 提示
>
> 方式 1 的优缺点：
>
> - 优点：图形用户界面使得操作十分直观，常用对象窗口和菜单相对固定，简单的几次反复练习后便会形成稳定的“记忆路径”
> - 不足：操作效率较低，难以在底层上实现“掌控一切”的状态

（2）方式 2：完全程序命令驱动，也即通过执行特定编程命令来进行分析操作，具体又可分为两种情形：

方式 2.1：交互式命令驱动，也即在命令视窗中逐条地编写程序代码，逐条运行各个编程命令

- 类似于“看一步走一步”的模式，命令编程不需要十分严格的逻辑安排
- 只要有特定分析需要，就随时编写命令，随时执行编程命令，得到想要的分析结果
- 适合于相对复杂、有一定重复性，但又不太繁重的分析任务

> 提示
>
> 方式 2.1 的优缺点：
>
> - 优点：灵活、自由，可以与菜单操作（方式 1）互相结合——只要需要就可随时登台“串场”发挥作用！
> - 不足：代码管理困难，条理性较差，需要繁琐的手工编写和操作

方式 2.2：批量式命令驱动，也即在程序文件中系统地编写程序代码，自动运行全部编程命令

- EViews 程序文件为.prg 文件后缀，是一类通用文本文件，所有代码都以文本形式进行编写、组织和管理。
- 类似于“照着剧本演戏”的模式，所有“剧本”内容（程序代码）都是经过精心设计和安排的
- 一开始分析任务就极为明确，程序命令有整体“剧情”，各命令间的联系、顺序等逻辑性较强
- 适合于较为复杂、重复性较高、极为繁重的分析任务

> 提示
>
> 方式 2.2 的优缺点：
>
> - 优点：代码管理方便，代码可重复使用，代码逻辑性强，分析操作过程的条理性很清晰
> - 不足：需要系统掌握 EViews 编程语言，学习使用成本较大；需要较强的代码管理和统计分析逻辑，容易出现各类运行报错。

（3）方式 3：菜单窗口和程序命令并行驱动，也即同时结合了功能菜单和编程命令来开展相关分析工作，往往都是菜单驱动加上视窗命令

提示

方式 3 的优缺点：

- 菜单驱动的优点是简单直观（类似于“所见即所得”），不足是功能菜单的操作效率较为低下
- 命令视窗的优点是底层操作（类似于“我思故我在”），不足是命令语法的学习曲线较为陡峭，记忆负担大，代码写作要求较高

1.8.1.1 菜单窗口和程序命令并行驱动

如前所述，菜单窗口和程序命令并行驱动，同时结合了功能菜单和编程命令来开展相关分析工作，是数据分析实践中最常用到的、最受初学者欢迎的操作方式之一。这种操作方式，首先需要掌握功能菜单驱动的数据分析操作，同时还要求了解并使用最基本的 EViews 编程语言进行简单代码编写和运行分析。总体看来，这种操作方式下，以功能菜单操作为主，以视窗命令操作为辅——数据分析者将主要利用菜单操作完成主要的常规性分析任务，而一些非常规的、需要定制化的分析任务则交给视窗命令去执行。

下面通过宏观经济发展案例演示基本操作步骤，将展示“功能菜单操作-视窗命令操作”的交互执行。其中数据分析的目标任务有两个：一是得到一个向量（Vector）对象，用于存放判定系数 R^2 和调整判定系数；二是得到一个标量（Scalar）对象，用于存放给定显著性水平 $\alpha = 0.05$，且自由度为 $f = 128$ 时，理论查表 t 值 $t_{1-\alpha/2}(f) = t_{0.975}(128)$。

假设前期已完成如下菜单操作工作：

- （见节 1.7.2 ）已经成功创建了工作文件macro_economic
- （见节 1.7.2 ）已经成功导入数据
- （见节 1.7.5 ）已经通过菜单操作进行双对数回归分析，并得到方程对象eq_main
- 双击方程对象eq_main 可以查看 EViews 给出的回归分析报告。可以看到，回归方程拟合优度中的判定系数 $R^2 = 0.998620$，调整判定系数判定系数 $\bar{R}^2 =$

目标任务 1（见图 1.18 ）：通过视窗命令操作，提取出双对数回归方程中的判定系数 R^2 和调整判定系数 $\bar{R}^2$，并将这两个数值分别存放在一个向量（Vector）对象vec_fitness 中。

（1）在命令视窗中：

- 输入代码行 `vector(2) vec_fitness(1)=eq_main.@r2`
- 鼠标光标停留在该代码行上 ⇒ 然后按 Enter 键，便可以运行该代码行!

（2）在工作文件视窗下：

- 将看到生成了一个向量对象fitness
- 双击查看fitness，发现该向量含有两个元素，其中第一个元素为 0.998620（也即回归方程的判定系数 R^2 值），第二个元素为 0.000000

（3）在命令视窗中：

- 输入代码行 `vec_fitness(2)=eq_main.@rbar2`
- 鼠标光标停留在该代码行上 ⇒ 然后按 Enter 键，便可以运行该代码行!

（4）在工作文件视窗下：

- 双击查看vector_fitness，发现该向量含有两个元素，其中第一个元素为 0.998620（也即回归方程的判定系数 R^2 值），第二个元素为 0.998587（也即回归方程的调整判定系数 $\bar{R}^2$ 值）

目标任务 2（见图 1.19 ）：通过视窗命令操作，得到一个标量（Scalar）对象t_value，用于存放给定显著性水平 $\alpha = 0.05$，且自由度为 $f = 125$ 时，理论查表 t 值 $t_{1-\alpha/2}(f) = t_{0.975}(125)$。

（1）在命令视窗中：

- 输入代码行 `scalar t_value=@qtdist(0.975,125)`
- 在该代码行上按 Enter 键

（2）在工作文件视窗下中：

- 将看到生成了一个标量对象t_value
- 双击查看t_value，发现该标量含有一个元素，其数值为 2（也即计算表明，在给定显著性水平 $\alpha = 0.05$，且自由度为 $f = 125$ 时，理论查表 t 值 $t_{1-\alpha/2}(f) = t_{0.975}(125) = 2$）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

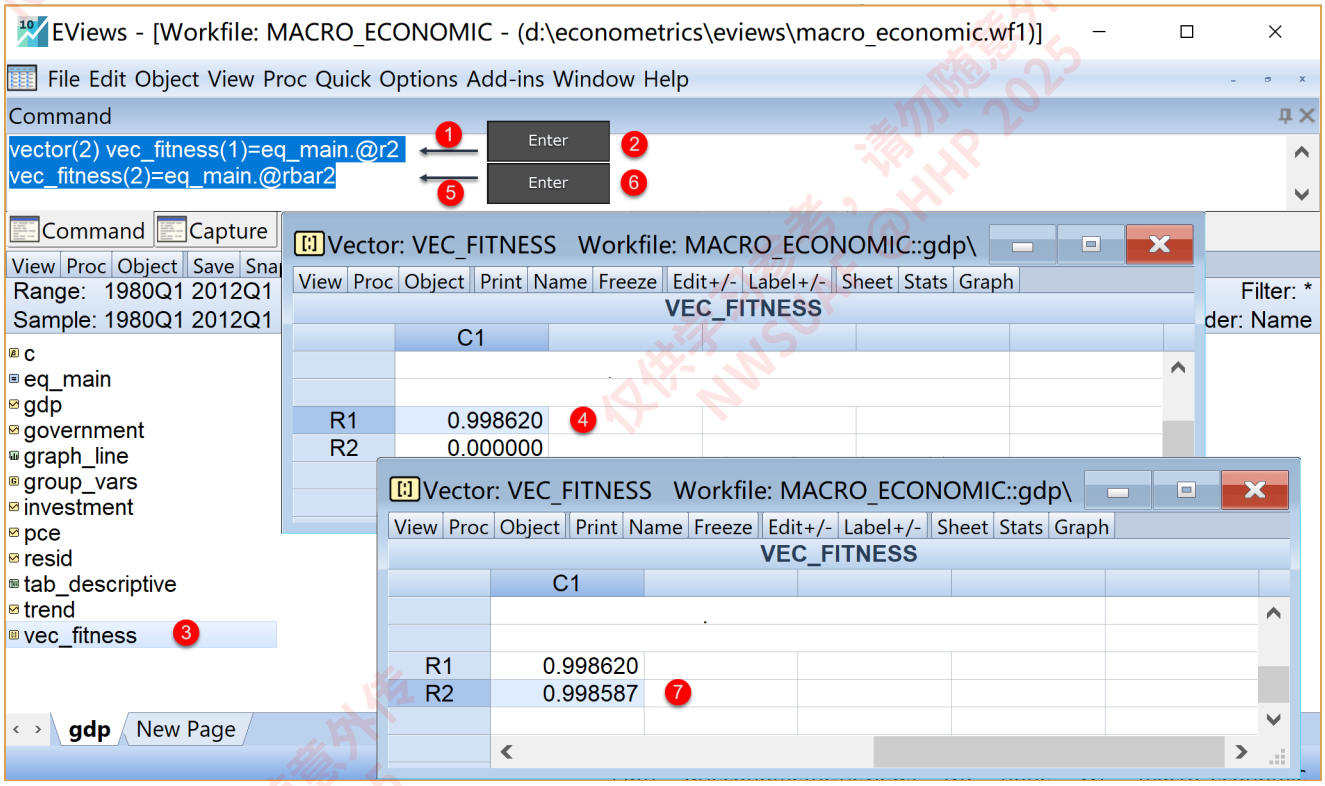


图 1.18: 通过命令视窗提取回归方程的判定系数 R^2 和调整判定系数 $\bar{R}^2$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

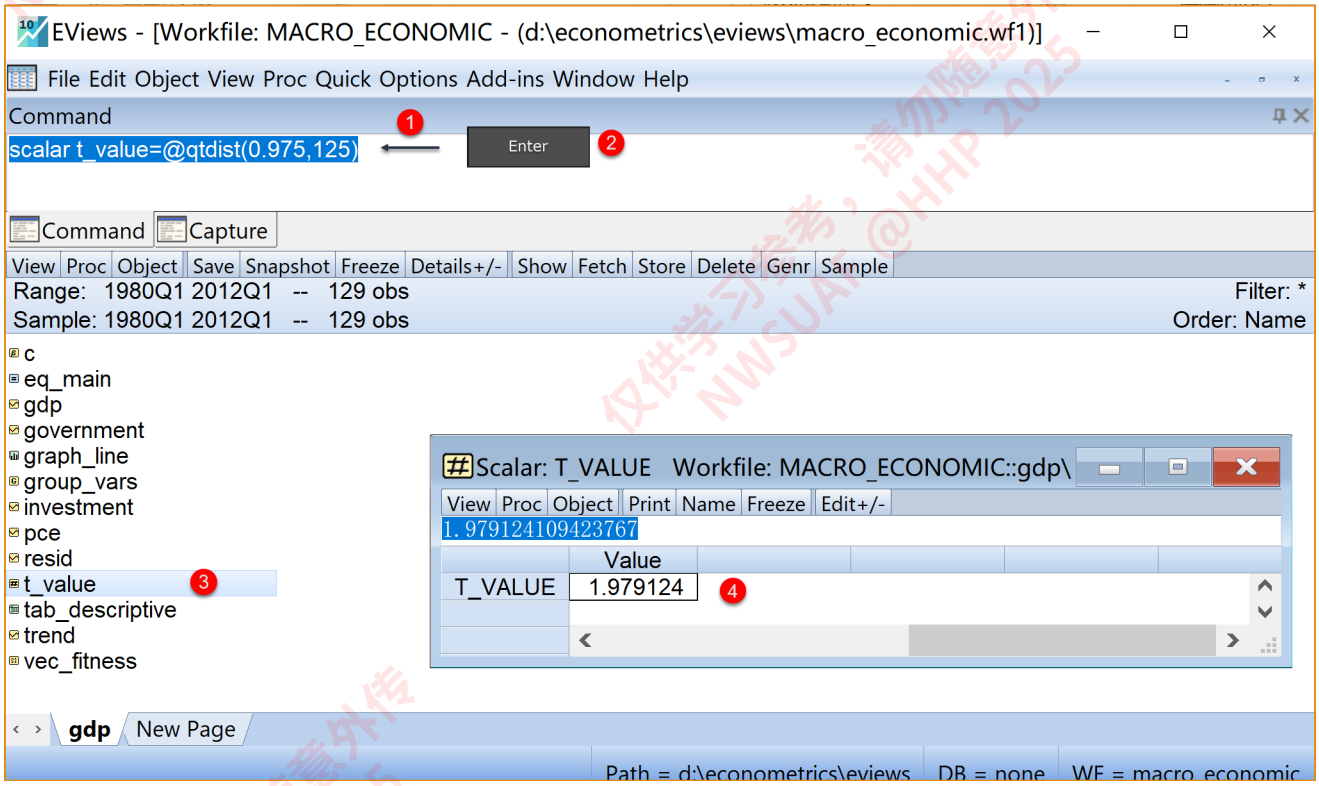


图 1.19: 通过命令视窗计算得到理论 t 值 $t_{1-\alpha/2}(f) = t_{0.975}(125)$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

1.8.1.2 批量式命令驱动

下面将简要介绍（方式 2.2）批量式命令驱动运行模式的基本操作。作为 EViews 高级学习技巧，很有必要理解 EViews 程序文件（文件后缀名为.prg）的创建和运行操作：

（1）创建程序文件（见图 1.20 ）。

- 在工作文件视窗下，依次点击选择菜单栏 File ⇒ New ⇒ Program。
- 命名并保存程序文件。在程序文件视窗下，点击菜单栏 Save ⇒ 确定文件存放地址（如本地文件夹”d:\econometrics\eviews” 下），输入文件名 `first_program`（建议命名，可修改）⇒ 完成，点击 OK
- 查看已保存的程序文件并双击打开。在本地文件夹下（如”d:\econometrics\eviews”）可以看到程序文件first_program.prg，可双击打开

（2）使用程序文件视窗的功能菜单。

> 提示
>
> 在程序文件视窗（Program:first_grogram）下，可以看到如下功能菜单
>
> - 运行菜单（Run）：用于运行全部代码
> - 打印菜单（Pringt）
> - 保存菜单（Save）和另存菜单（SaveAs）
> - 剪切菜单（Cut）、拷贝菜单（Copy）和粘贴菜单（Paste）
> - 插入文本菜单（InsertTxt）
> - 查找菜单（Find）和替换菜单（Replace）
> - 打包命令菜单（Wrap+/-）：用于封装命令
> - 代码行号菜单（LineNum+/-）：用于显示或隐藏代码行号
> - 加密菜单（Encrypt）：用于给程序文件加密

（3）在程序文件中编写、管理和运行 EViews 代码。

- 根据 EViews 编程语言规则（见节 1.8.2 ），直接以文本形式编写程序代码。

- 执行整个程序文件。在检查确认整个程序文件的全部代码后，可以直接运行整个程序文件。在程序文件视窗（Program:first_grogram）下，点击菜单栏中的 Run
- 执行部分代码行。选中已经编写的代码命令行，点击鼠标右键，选择 Run Selected。选中已经编写的代码命令行，点击鼠标右键，还可以看到其他一些常用的快捷菜单。

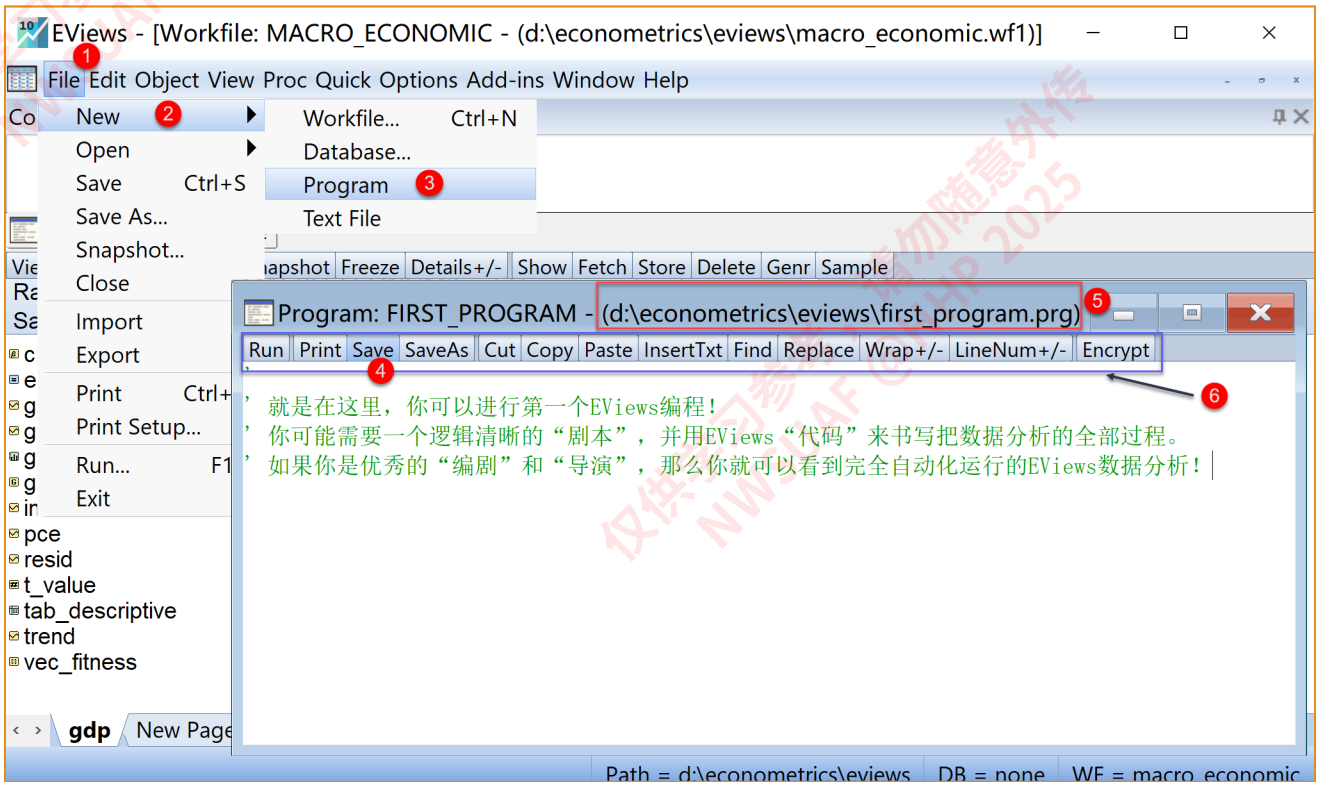


图 1.20: EViews 程序文件（.prg）的创建操作

下面所示为 EViews 程序文件my_first_program.prg 的代码文本，运行该 EViews 程序文件，将可以完全地、自动地再现第一章中“宏观经济发展案例”主要分析步骤，包括：

（1）创建工作文件。（工作文件名 =development，子页命名 =gdp）（效果与节 1.7.2.2 相同）

（2）导入外部数据。（效果与节 1.7.2.3 相同）

（3）构造组（Group）对象group_vars。（效果与节 1.7.4.3 相同）

（4）得到描述性分析的表格（Table）对象tab_descriptive。（效果与节 1.7.4.3 相同）

（5）得到堆栈多线图（Line-Symbol）的图形对象graph_line。（效果与节 1.7.4.3 相同）

（6）进行双对数线性回归模型分析，得到方程（Equation）对象eq_main。（效果与节 1.7.5 相同）

（7）提取出双对数回归方程中的判定系数 R^2 和调整判定系数 $\bar{R}^2$，并将这两个数值分别存放在一个向量（Vector）对象vec_fitness 中。（效果与节 1.8.1.1 相同）

（8）计算得到给定显著性水平 $\alpha = 0.05$，且自由度为 $f = 125$ 时，理论查表 t 值 $t_{1-\alpha/2}(f) = t_{0.975}(125)$，并将这个计算值存放到一个标量（Scalar）对象t_value 中。（效果与节 1.8.1.1 相同）

```
'================================================================================
'说明：以下为EViews编程文件my_first_program.prg的代码
'  将展示第一章中“宏观经济发展案例”主要分析步骤的“批量式命令驱动”实现方法：
'  其中符号'起头的行为注释行，其他为EViews命令行。
'================================================================================

'[命令行1]：创建工作文件（工作文件名=development，子页命名=gdp）。
'  其中数据为季度数据，起始于1980年1季度，结束于2012年1季度。

wfcreate(wf=development,page=gdp)  q  1980/1 2012/1

'[命令行2]：导入外部数据，路径为d:\economitrics\macro-economic.xlsx。

import d:\economitrics\macro-economic.xlsx

'[命令行3]：构造组对象，名为group_vars，由序列gdp; pce; investment; government构成。

group group_vars gdp pce investment government

'[命令行4]：得到表格（表格命名为tab_descriptive），
'  <--组对象group_vars进行描述性统计分析（Descriptive statistics）。

freeze(tab_descriptive) group_vars.stats

'[命令行5]：进行绘图，图形名为graph_line线图（堆栈图）<--对组对象group_vars

graph graph_line.line(s)  group_vars
```

```
'[命令行6]：构建方程，给方程命名为eq_main执行线性回归
'  其中，线性方程为 gdp ~ c+ pce+ investment+ government

equation eq_main.ls gdp c pce investment government

'构造向量（含有2个元素），并将向量命名为vec_fitness，其中：
' [命令行7]：向量的第1个元素将用于提取并存放方程eq_main中的判定系数$R^2$
' [命令行8]：向量的第1个元素将用于提取并存放方程eq_main中的调整判定系数$ \bar{R}
^2$

vector(2) vec_fitness(1)=eq_main.@r2
vec_fitness(2)=eq_main.@rbar2

'[命令行9]：构造标量，标量命名为t_value，
'  利用函数@qtdist(p,f)计算理论t值（a=0.05，f=125）

scalar t_value=@qtdist(0.975,125)

'================================================================================
```

通过以上 EViews 程序文件my_first_program.prg，可以初步对 EViews 编程语言及命令体系有一个直观的认知，并可以利用数据进行实证练习。当然，里面的很多 EViews 编程命令与代码，会令初学者不知所以然（尽管只需要修改外部数据文件路径，就能完全再现实验分析结果）。因此，本章后续内容将会帮助梳理和总结 EViews 编程的命令语法结构，以及常用命令及函数。

1.8.2 命令语法结构

实际上 EViews 编程语言源自于一种比较古老的编程语言，最早于 1965 年由 Robert Hall 开发而成。EViews 编程语言有其自身的体系性和逻辑性。如果具备一定的编程基础，可以很快掌握基本的 EViews 编程技巧。总体来看，学习 EViews 编程语言需要把握如下几个基本原则：

（1）Eviews 编程的“前台”是各种 EViews 工作文件对象（workfile object，见节 1.7.3.1 ），这些都是在“前台”——工作文件视窗下——“可见的”。

- 这些 EViews 工作文件对象（workfile object）包括：序列（Series）对象、方程（Equation）对象、表格（Table）对象、图形（Graph）对象、标量（Scalar）对象等
- 大部分 Eviews 编程代码都会直接作用于或指向这些 EViews 工作文件对象（workfile object）
- Eviews 编程代码直接产生的各种 EViews 工作文件对象，也往往就是数据分析的直接成果，将用于报告撰写，也是数据分析的最重要目标之一。

（2）Eviews 编程的“后台”是命令视窗或者编程文件.prg，所有 EViews 命令代码在这里被编写、管理和运行。

- 其中的一些 EViews 命令代码，可能只是辅助性地帮助实现前台的结果——这些命令代码本身并不会直接“可见的”，也即不会直接作用或直接指向“EViews 工作文件对象”。
- EViews 开发有一系列编程变量（programming variables），它们大多只是在“后台”发挥桥梁或表征作用，从不直接在“前台”现身。

> 提示
>
> 编程变量（programming variables）示例：
>
> - 控制变量（control variables），用以代表数值型的变量，用感叹号! 做前缀表达。例如，编写并运行命令`!x=7`，表示生成一个数值型变量 `x`，其值为 7。然而即便运行此命令行，在工作文件视窗下，也不会产生任何 EViews 工作文件对象——但控制变量 `x` 确确实实是隐身在后台，并可以发挥作用的！
> - 字符变量（string variables），用以代表字符型的变量，用百分号% 做前缀表达。例如，编写并运行命令`%model="gdp~ c + investment + pce + government"`，表示生成一个字符型变量 model，其值为`"gdp~ c + investment + pce + government"`。同理，即便运行此命令行，在工作文件视窗下，也不会产生任何 EViews 工作文件对象——也即字符变量 `model` 是隐身在后台发挥作用的！

（3）作为初级或中级 Eviews 编程学习者，应该重点关注数据分析目标的实现而不是程序运行效率的提高。也就是说：

- 紧密围绕“可见的”工作文件对象（object）进行 EViews 代码编程，直接得到可见的分析结果
- 为了提高编程代码的运行效率，或者为了改善编程代码的简洁性（很多编程者都会被认为有“代码洁癖”倾向），或者为了数据分析的代码可重复性使用，诸如此类目标的实现往往都需要对代

码进行“费尽心机”的后台安排和管理。如果这些目标并不那么重要，这些“令人尊敬”的编程努力也就完全有理由可以暂时放弃。

因此，假定 Eviews 编程学习者关注的重点是数据分析目标的实现，下面将展示 EViews 编程语言中基于对象（object）的一种标准语法结构（注意空格的使用）：

```
action(action_opt)  object_name.view_or_proc(options_list)
```

该语法结构的大意是：将会执行一项动作（action），这项动作的执行必须按照若干标准（action_opt),这项动作将会施加在一个对象（名为 object_name）的某种可执行状态上（view_or_proc），对象上的这种可执行状态可以是对象的视图（view）或者程序（proc），并且可能会要求提供这种状态下的具体要求（options_list），而这种状态（view_or_proc）的出现或执行也需要具体的参数条件。其中：

（1）action 表示动作（action Commands）命令，包括：

- show 表示显示结果
- do 表示执行指定程序
- freeze 表示从对象视窗中生成表格或图形
- print 表示打印对象视窗

（2）action_opt 表示动作（action Commands）命令的各种选项设置（options）

（3）object_name 表示 EViews 的对象名（object name），例如：- 名为 eq_main 的方程对象eq_main - 名为 gdp 的序列对象gdp - 名为 graph_line 的图形对象graph_line

（4）view_or_proc 表示对象可用于执行的视窗或程序命令（View and Procedure Commands）。对于不同的对象（object），其可用于执行的视窗或程序命令，都不尽相同。例如，序列（Series）对象ser01 可用于执行的视窗（View）命令包括：

- 显示为表单和描述性统计，分别是：ser01.sheet；ser01.stats
- 显示为图形，包括：ser01.line；ser01.bar；ser01.hist
- 显示为表单，包括：ser01.freq；ser01.correl；ser01.lrvar；ser01.uroot；ser01.vratio 2 4 8；ser01.bdstest；…
- …

（5）(options_list) 表示对象可用于执行视窗或程序命令（View and Procedure Commands）的各种选项设置（options）

（6）arg_list 表示视窗或程序命令下的语法参数（arguments）

- 例如，方程对象eq_main 的线性回归程序表达为 `eq_main.ls`，就需要提供线性回归方程参数：`gdp c investment pce government`，也即方程 $gdp = \hat{\beta}_1 + \hat{\beta}_2 investment + \hat{\beta}_3 pce + \hat{\beta}_4 government + e_i$

下面将展示 3 条典型的 EViews 编程命令行，它们的语法结构类似，但运行结果则迥然不同：

```
'说明：以下三个EViews命令行，将简要展示EViews编程的语法结构及其特征。
'      其中，符号'起头的行，为注释行，其他为EViews命令行。
' ============================================================================
' 命令行1：生成线性回归方程对象eq_main01
equation eq_main01.ls gdp c investment pce government

' 命令行2：生成并打开线性回归方程对象eq_main02
show eq_main02.ls gdp c investment pce government

' 命令行3：生成线性回归方程对象eq_main03，并将回归报告保存为表格对象tab_output
freeze(tab_output) eq_main03.ls gdp c investment pce government
' ============================================================================
```

以上 3 条 EViews 编程命令行的具体解读如下：

命令行 1 解读：将生成线性回归方程对象eq_main01

- `equation` 表示声明（declare）一个方程对象
- `eq_main01` 表示对象名（object_name）
- `.ls` 表示对方程对象执行一个叫 `ls`（也即线性回归分析）的程序（proc）
- `gdp c investment pce government` 表示执行 `ls`（也即线性回归分析）程序（proc）需要调用的参数（此处为设定回归方程）

命令行 2 解读：将生成并打开线性回归方程对象eq_main02

- `show` 表示执行一个名为 `show`（也即展示结果）的动作（action）
- `eq_main02` 表示对象名（object_name）
- `.ls` 表示对方程对象执行一个叫 `ls`（也即线性回归分析）的程序（proc）
- `gdp c investment pce government` 表示执行 `ls`（也即线性回归分析）程序（proc）需要调用的参数（此处为设定回归方程）

命令行 3 解读：将生成线性回归方程对象eq_main03，并将回归报告保存为表格对象tab_output

- `freeze(tab_output)` 表示执行一个名为 `freeze`（此处为保存表格）的动作（action）；括号内的 `tab_output` 为执行该动作的选项设置（action_opt），此处表示保存的表格命名为 `tab_output`
- `eq_main03` 表示对象名（object_name）
- `.ls` 表示对方程对象执行一个叫 `ls`（也即线性回归分析）的程序（proc）
- `gdp c investment pce government` 表示执行 `ls`（也即线性回归分析）程序（proc）需要调用的参数（此处为设定回归方程）

总体而言，以上演示只是一种常用的典型语法结构，是基于对象（object）进行操作。实际上，EViews 编程语法结构有多种多样，有的是标准化语法结构，有的是简化语法结构，还有更多的约定式语法结构，深入掌握的话需要进行大量记忆，或者查看 EViews 官方的命令参考书 (IHS 2017)。无论如何，围绕对象（object）进行 EViews 编程是一条切实有效的学习途径，而 EViews 官方在线帮助文档[4]提供了全面的对象参考（object reference），并分类列出了全部的相关命令，对 EViews 编程很有帮助：

- 对象的声明（Declaration）
- 对象的方法（Methods）
- 对象的视窗（Views）
- 对象的可执行程序（Procs）
- 对象的可返回数据（Data Members）

[4] http://www.eviews.com/help/helpintro.html#page/content%2Fcobjpreface.html%23

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

1.8.3 常用命令及函数

本书后续章节相关部分还将涉及 EViews 编程的初级或中级技能。为了便于后续内容的学习，本书总结和梳理了常用的 EViews 函数命令（见表 1.11 ），可以更好地配合 EViews 代码编写和运行，帮助完成计量经济学实证分析任务。

表 1.11: 常用 EViews 函数命令

序号	分类	函数命令	作用	说明
1	运算符	+	加	
2	运算符	-	减	
3	运算符	*	乘	
4	运算符	/	除	
5	运算符	^	幂次	
6	运算符	=	等于	
7	常数	@pi	圆周率 π	$3.14159\cdots$
8	基础数学函数	@abs(x)	绝对值	$\mid X \mid$
9	基础数学函数	@exp(x)	自然数 e 的幂次	e^X
10	基础数学函数	@inv(x)	倒数	$1/X$
11	基础数学函数	@log(x)	自然对数	$log_e(X)$
12	基础数学函数	@log10(x)	10 为底的对数	$log_{10}(X)$
13	基础数学函数	@recode()	重新编码	
14	基础数学函数	@round(x)	取整	
15	基础数学函数	@sqrt(x)	平方根	$\sqrt{X}$
16	时间序列函数	@d(x)	一阶差分	Δ

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 1.11: 常用 EViews 函数命令

序号	分类	函数命令	作用	说明
17	时间序列函数	`@lag(x,n)`	n 阶滞后	X_{t-n}
18	时间序列函数	`@trend()`	构造趋势序列	$0, 1, 2, \cdots$
19	描述性统计函数	`@cor(x,y)`	相关系数	r
20	描述性统计函数	`@cov(x,y)`	协方差	$Cov(X, Y)$
21	描述性统计函数	`@gmean(x)`	几何平均数	$\bar{X}_G$
22	描述性统计函数	`@hmean(x)`	调和平均数	$\bar{X}_H$
23	描述性统计函数	`@max(x)`	极大值	Max
24	描述性统计函数	`@mean(x)`	简单平均数	$\bar{X}$
25	描述性统计函数	`@median(x)`	中位数	M_e
26	描述性统计函数	`@min(x)`	极小值	Min
27	描述性统计函数	`@obs(x)`	样本数	n
28	描述性统计函数	`@quantile(x,q)`	分位数	$Q_1; Q_3$
29	描述性统计函数	`@skew(x)`	偏度	
30	描述性统计函数	`@stdev(x)`	标准差	S
31	描述性统计函数	`@sum(x)`	求和	$\sum X_i$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 1.11: 常用 EViews 函数命令

序号	分类	函数命令	作用	说明
32	描述性统计函数	`@sumsq(x)`	平方和	$\sum X_i^2$
33	描述性统计函数	`@var(x)`	方差	S^2
34	密度; 概率函数	`@cchisq(x,f)`	累积卡方分布函数	$G_{\chi^2}(X,f)$
35	密度; 概率函数	`@dchisq(x,f)`	卡方分布的密度函数	$g_{\chi^2}(X,f)$
36	密度; 概率函数	`@qchisq(p,f)`	卡方值	右侧大值 $\chi^2_{(1-\alpha)}(p,f)$
37	密度; 概率函数	`@cfdist(x,f1,f2)`	累积 F 分布函数	$G_F(X,f1,f2)$
38	密度; 概率函数	`@dfdist(x,f1,f2)`	F 分布的密度函数	$g_F(X,f1,f2)$
39	密度; 概率函数	`@qfdist(p,f1,f2)`	F 值	右侧大值 $F_{(1-\alpha)}(p,f1,f2)$
40	密度; 概率函数	`@ctdist(x,f)`	累积 t 分布函数	$G_t(X,f)$
41	密度; 概率函数	`@dtdist(x,f)`	t 分布的密度函数	$g_t(X,f)$
42	密度; 概率函数	`@qtdist(p,f)`	t 值	右侧正值 $t_{(1-\alpha/2)}(p,f)$
43	密度; 概率函数	`@cnorm(x)`	累积 t 分布函数	$G_N(X)$
44	密度; 概率函数	`@dnorm(x)`	t 分布的密度函数	$g_N(X)$
45	密度; 概率函数	`@qnorm(p)`	t 值	右侧正值 $N_{(1-\alpha/2)}(p)$
46	特定表达式	`ar`	自回归误差设置项	p 阶自回归 $AR(p)$
47	特定表达式	`ma`	移动平均误差设置项	q 阶移动平均 $MA(q)$
48	特定表达式	`@expand`	自动产生的虚拟变量	$Edu_{D1}; Edu_{D2}; \cdots$

表 1.11: 常用 EViews 函数命令

序号	分类	函数命令	作用	说明
49	矩阵函数	`@ones(n,k)`	生成元素全为 1 的 $n*k$ 维矩阵	$\mathbf{1}_{n,k}$
50	矩阵函数	`@identity(n,n)`	生成 $n*n$ 维单位矩阵	$\mathbf{I}_{n,n}$
51	矩阵函数	`@transpose(X)`	对矩阵 $\mathbf{X}$ 进行转置	$\mathbf{X}'$
52	矩阵函数	`@det(X)`	得到方阵 $\mathbf{X}$ 的行列式	$\mid \mathbf{X} \mid$
53	矩阵函数	`@eigenvalues(X)`	得到对称矩阵的特征值向量	
54	矩阵函数	`@inverse(X)`	对方阵 $\mathbf{X}$ 求逆	$\mathbf{X}^{-1}$
55	矩阵函数	`@ediv(X,Y)`	两个矩阵对应元素点除	X_{ij}/Y_{ij}
56	矩阵函数	`@getmaindiagonal(X)`	得到矩阵 $\mathbf{X}$ 的主对角线元素(向量)	

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

2 一元线性回归

仅供学习参考，请勿随意外传
NWSUAF @HHP 2025

2.1 实验目的及要求

目的：熟悉一元回归建模和分析的基本原理。

要求：熟练利用 EViews 软件进行一元回归分析，并进行计量经济检验；能快速看懂 EViews 分析报告，并理解报告内容的内部关系。

2.2 实验原理

- 经济建模要点
- 参数估计
- 参数推断
- 拟合优度（判定系数）
- 模型检验（经济学检验、统计学检验 _t 检验和 F 检验）

2.2.1 回归的本质是什么?

回归是对个体值 Y_i 有“回归”到的条件期望值 $(Y|X_i)$ 这一现象的形象表达，这也是回归的本质。

一般所说的线性回归，是指条件期望值 $(Y|X_i)$ 的轨迹表现为直线的情形，也就意味着在给定不同的 X_i，个体值 Y_i 有“回归”到一条直线的趋势。

此外，回归分析在根本上是对宏大的、规则化的、超现实的**总体**真实规律的探索和认知，但实际上我们只能通过局部的、破碎的、现实的**样本**来对其加以猜测和推断。这一“故事主线”是始终贯穿于回归分析全部过程中。

定义 2.1 (回归)**.** 通过解释变量 X_i（在重复抽样中）的已知或设定值，去估计和（或）预测应变量 Y_i 的总体均值。

定义 2.2 (概率)**.** 注意区分两种概率定义：

- **无条件概率**：不受 X_i 变量取值影响下，Y_i 出现的可能性，记为：$p(Y_i)$
- **条件概率**：给定变量 X_i 的取值条件下，Y_i 出现的可能性，记为：$p(Y_i|X_i)$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

定义 2.3 (期望)**.** 注意区分两种期望定义：

无条件期望：不受 X_i 变量取值影响下，变量 Y_i 的期望值。

- 离散变量情形下记为: $E(Y) = \sum_{i=1}^{N} Y_i \cdot p(Y_i)$
- 连续变量情形下记为：$E(Y) = \int Y \cdot g(Y)dY$

条件期望：在给定变量 X_i 的取值条件下，变量 Y_i 的期望值。

- 离散变量情形下记为: $E(Y|X_i) = \sum_{i=1}^{N} (Y_i|X_i) \cdot p(Y_i|X_i)$
- 连续变量情形下记为：$E(Y|X) = \int (Y|X) \cdot g(Y|X)dY$

定义 2.4 (总体回归线)**. 总体回归线（PRC）**：是给定变量 X_i 值时 Y_i 的条件期望值的轨迹。

定义 2.5 (总体回归函数)**. 总体回归函数（PRF）**：是用来刻画总体回归线的数学函数。

- 一般地可以表达为：$E(Y|X_i) = f(X_i)$。
- 如果总体回归线表现为直线，则可以用如下函数表达：$E(Y|X_i) = f(X_i) = \beta_1 + \beta_2 X_i$

定义 2.6 (样本回归线)**. 样本回归线（SRC）**：是通过对样本数据拟合后得到的一条抽象化的曲线。

定义 2.7 (样本回归函数)**. 样本回归函数（SRF）**是样本回归曲线的数学函数形式。

- 一般地可以表达为：$\hat{Y}_i = f(X_i)$。
- 如果样本回归曲线表现为直线，则可以用如下函数表达：$\hat{Y}_i = f(X_i) = \hat{\beta}_1 + \hat{\beta}_2 X_i$

定义 2.8 (随机干扰项和总体回归模型)**. 随机干扰项**：是总体回归函数 SRF 中忽略掉的，但又影响着 Y_i 的其他全部变量的**替代物**。它是 Y_i 与其条件期望的离差，也即 $u_i = Y_i - E(Y|X_i)$。

总体回归模型（PRM）：用确定性的总体回归函数和非系统性的随机干扰项共同来刻画变量 Y_i 的一种假象模型。

- 如果没有明确的总体回归函数，一般表达为：$Y_i = E(Y|X_i) + u_i$
- 如果总体回归函数为直线形式，则表达为：$Y_i = \beta_1 + \beta_2 X_i + u_i$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

定义 2.9 (残差和样本回归模型)**. 残差**：是样本回归函数忽略掉的，但又影响着 Y_i 的全部其他变量的**替代物**。它是样本观测值 Y_i，与样本回归线上的点的取值 $\widehat{Y}_i$ 之间的离差，也即：$e_i = Y_i - \widehat{Y}_i$

样本回归模型（SRM）：用拟合的样本回归函数和未能拟合的残差共同来刻画样本观测值 Y_i 的一种假象模型。

- 如果无明确拟合的样本回归函数，一般表达为：$Y_i = f(X_i) + e_i$
- 如果拟合样本回归函数为直线形式，则表达为：$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$

如果说回归的本质是对总体规律的认知，那么人们脑海中马上浮出的问题往往还有：

- 什么情况下我们认为总体是能被认知的？
- 以及怎样才能说我们完全知悉了总体规律？

看来用样本数据推断总体参数，必须首先交代“故事背景”：

对于线性总体回归模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{2.1}$$

显然，模型式 2.1 等式左边的因变量 Y_i 背后的决定规律是一个谜团，研究者希望破解和揭晓它。然而，计量分析活动的前提是 Y_i 要“能被认知”。我们一般认为 Y_i“能被认知”，至少需要进行下面几个步骤的理解：

- Y_i 是一个客观存在物，不以研究者主观意志而改变
- Y_i 是一个随机变量，它背后有值得探索的规律和秘密
- 一个随机变量只有通过概率分布，才能够被人们所理解和认知
- 一个随机变量的概率分布，可以用参数进行充分的、抽象的表达
- 如果概率分布及其参数都知道了，这个随机变量的规律也被揭晓了。

那么，模型式 2.1 等式右边就是研究者的主观化的构建，期望通过它实现对等式左边的客观存在物 Y_i 的规律破解。它是计量经济学“建模”的核心，需要注意的是：

- 研究者**直接**主观构建的确定性部分为 $\beta_1 + \beta_2 X_i$，也就是大家普遍所知的直线系统[1]

[1] 实际上，设定可以十分的多样化

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 研究者**间接**主观构建的随机性部分为 u_i。它更像是上一个行为 “衍生物” 或 “副产品”
- 确定性部分 $\beta_1 + \beta_2 X_i$ 和随机性部分为 u_i，存在某种 “此消彼长” 的关系
- 一个 “理想化的好模型”，应该是研究者不能够再从 u_i 中提取出任何系统性信息用以解释 Y_i。也即，u_i 只有随机变量的数字化 “信号”，但没有更多的认知化 “信息”[2]

2.2.2 怎样用样本数据推断总体参数?

用样本数据估计总体参数，有三种指导理念完全不同方法：

- 最小二乘法（Least Square，LS）：随机抽样情形下，使得拟合的样本回归线与样本观测点之间的欧几里得空间距离最小化，那么就有理由认为基于这次样本拟合得到的参数估计值是对真实参数的一个**可信推断**。
- 极大似然法（Maximum Likelihood，ML）：随机抽样情形下，如果一次实验样本在下次抽样的时候（假想的）能最大可能再次出现（似然函数值最大），那么就有理由认为基于这次实验样本得到的参数估计值是对真实参数的一个**可信推断**。
- 矩估计法（Method of Moment，MM）：随机抽样情形下，如果样本能很好的继承总体特征，那么就有理由认为通过样本矩等于总体矩[3]，求解得到的参数估计值是对真实参数的一个**可信推断**。

下面重点介绍最小二乘法（LS）在线性回归模型中的使用。

对于如下线性模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad \text{(PRM)} \tag{2.2}$$

其中：

$$u_i \sim N(0, \sigma^2)$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \quad \text{(SRM)} \tag{2.3}$$

[2]可以想象一下传统的收音机（模拟数字信号），在每周三下午全部频道调频静默时，收音机的喇叭只有无线波的 “滋滋滋滋”（信号），但却没有我们喜爱主播磁性的声音或甜美音乐（信息）

[3]矩（moment）的概念最早来自于物理学，简单可理解为 “特征”。一个随机变量可以有多个矩（多个维度的特征），例如样本均值 $\bar{Y}$ 为样本一阶矩，总体期望 E(Y_i) 就是总体一阶矩

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

对模型式 2.3 求解如下最小化过程：

$$\begin{aligned} Min(Q) &= \sum (Y_i - \hat{Y}_i)^2 \\ &= \sum \left(Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_i)\right)^2 \\ &= f(\hat{\beta}_1, \hat{\beta}_2) \end{aligned}$$

可以得到模型式 2.2 的回归参数估计值：

$$\begin{aligned} \hat{\beta}_2 &= \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \\ \hat{\beta}_1 &= \frac{n\sum X_i^2 Y_i - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \end{aligned} \quad \text{(FF solution)} \tag{2.4}$$

以上求解公式也称为公式（Favorite Five，FF）。此外我们也可以得到如下的离差公式 (favorite five，ff)：

$$\begin{aligned} \hat{\beta}_2 &= \frac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\beta}_1 &= \bar{Y}_i - \hat{\beta}_2 \bar{X}_i \end{aligned} \quad \text{(ff solution)} \tag{2.5}$$

进一步地，可以计算得到随机干扰项 u_i 的真是方差 σ^2 的 LS 估计量 $\hat{\sigma}^2$：

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \tag{2.6}$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-1} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-1} \tag{2.7}$$

2.2.3 最小二乘法有哪些优点?

在正态经典线性回归模型假设（N-CLRM）下，回归斜率系数 $\hat{\beta}_2$ 最小二乘估计量的总体方差/标准差，及其样本方差/标准差公式如下：

$$var(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \frac{1}{\sum x_i^2} \cdot \sigma^2 \tag{2.8}$$

$$s.e.(\hat{\beta}_2) = \sigma_{\hat{\beta}_2} = \sqrt{\frac{1}{\sum x_i^2}} \cdot \sigma \tag{2.9}$$

$$S^2_{\hat{\beta}_2} = \frac{1}{\sum x_i^2} \cdot \hat{\sigma}^2 \tag{2.10}$$

$$S_{\hat{\beta}_2} = \sqrt{\frac{1}{\sum x_i^2}} \cdot \hat{\sigma} \tag{2.11}$$

在正态经典线性回归模型假设（N-CLRM）下，回归截距系数 $\hat{\beta}_1$ 最小二乘估计量的总体方差/标准差，及其样本方差/标准差公式如下：

$$var(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n\sum x_i^2} \cdot \sigma^2 \tag{2.12}$$

$$s.e.(\hat{\beta}_1) = \sigma_{\hat{\beta}_1} = \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}} \cdot \sigma \tag{2.13}$$

$$S^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n\sum x_i^2} \cdot \hat{\sigma}^2 \tag{2.14}$$

$$S_{\hat{\beta}_1} = \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}} \cdot \hat{\sigma} \tag{2.15}$$

定理 2.1 (高斯-马尔可夫定理 (Gauss-Markov Theorem))**.** *在正态经典线性回归模型假设（N-CLRM）下，采用普通最小二乘法（OLS），得到的估计量 $\hat{\beta}$，是真实参数 β 最优的、线性的、无偏估计量（BLUE）。记为：*

$$\xrightarrow{N\text{-}CLRM} OLS(\hat{\beta}) \xrightarrow{BLUE} \beta$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

线性性. 根据离差公式，可以得到：

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum k_i Y_i$$

其中，$k_i = \frac{x_i}{\sum x_i^2}$。

容易证明，$\frac{x_i}{\sum x_i^2}$ 不全为 0，所以 $\hat{\beta}_2$ 是线性的。

同理可以证明：

$$\hat{\beta}_1 = \bar{Y}_i - \hat{\beta}_2 \bar{X}_i = \sum w_i Y_i$$

其中，$w_i = \frac{1}{n} - \bar{X}_i k_i$。

容易证明，$\frac{1}{n} - \bar{X}_i k_i$ 不全为 0。

所以 $\hat{\beta}_1$ 是线性的。 □

无偏性. 根据离差公式，可以得到：

$$\begin{aligned}\hat{\beta}_2 &= \sum k_i Y_i \\ &= \sum k_i (\beta_1 + \beta_2 X_i + u_i) \\ &= \beta_2 + \sum k_i u_i\end{aligned}$$

其中，$k_i = \frac{x_i}{\sum x_i^2}$，且可以证明，$\sum k_i = 0$。

因此有：

$$\begin{aligned}E(\hat{\beta}_2) &= \beta_2 + E(\sum k_i u_i) \\ &= \beta_2 + \sum k_i E(u_i) \\ &= \beta_2\end{aligned}$$

同理可以证明：

$$\begin{aligned}E(\hat{\beta}_1) &= \beta_1 + E(\sum w_i u_i) \\ &= \beta_1 + \sum w_i E(u_i) \\ &= \beta_1\end{aligned}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

其中，$w_i = \frac{1}{n} - \bar{X}_i k_i$。 □

方差最小/最优性. 已知：

$$var(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \frac{1}{\sum x_i^2} \cdot \sigma^2 var(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n \sum x_i^2} \cdot \sigma^2$$

假设存在用其他方法估计的线性无偏估计量：

$$\hat{\beta}_2^* = \sum (k_i + d_i) Y_i = \sum c_i Y_i$$
$$\hat{\beta}_1^* = \sum (w_i + g_i) Y_i = \sum h_i Y_i$$

其中，d_i 和 g_i 是不全为 0 的常数，则可以证明：

$$var(\hat{\beta}_2^*) \geq var(\hat{\beta}_2)$$
$$var(\hat{\beta}_1^*) \geq var(\hat{\beta}_1)$$

□

2.2.4 如何对模型进行统计检验?

在经典回归模型假设（CLRM）下，采用最小二乘法（LS），对样本数据进行拟合，得到的样本回归线，已经是最优的了（残差平方和最小），并且样本回归系数理论上也是总体参数的最优线性无偏估计量（BLUE）。为什么我们还需要对回归模型进行统计检验？

一个直接的回答是：样本数据来自于总体，样本是随机抽样得到的。对于一份特定的抽样数据，最小二乘法（LS）拟合的样本回归线此时确实是最优的，其回归系数也确实是总体参数的最优线性无偏估计量（BLUE）。但是，这一份随机样本数据得到的全部结论，究竟能多大程度地“推断”或“证实”模型的真实参数？这一问题仍旧悬而未决。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

2.2.4.1 平方和分解及拟合优度

首先需要确定的工作是，采用最小二乘法（LS），样本回归线对这一组数据拟合得有多好？

通过对 Y_i 的变异及其来源的分解，我们可以得到有关拟合优度评价的启发。其中 TSS 表示总离差平方和，ESS 表示回归平法和，RSS 表示残差平方和。

$$(Y_i - \bar{Y}_i) = (\hat{Y}_i - \bar{Y}_i) + (Y_i - \hat{Y}_i) \tag{2.16}$$

$$y_i = \bar{y}_i + e_i \tag{2.17}$$

$$\sum y_i^2 = \sum \hat{y}_i^{\,2} + \sum e_i^2 \tag{2.18}$$

$$TSS = ESS + RSS \tag{2.19}$$

进一步地，可以得到方差分析表（ANOVA）：

表 2.1: 一元线性回归模型的方差分析表（ANOVA）

变异来源	平方和符号 SS	平方和计算公式	自由度 df	均方和符号 MSS	均方和计算公式
回归平方和	ESS	$\sum(\hat{Y}_i - \bar{Y}_i)^2 = \sum \hat{y}_i^2$	1	MSS_{ESS}	$ESS/df_{ESS} = \hat{\beta}_2^2 \sum x_i^2$
残差平方和	RSS	$\sum(Y_i - \hat{Y}_i)^2 = \sum e_i^2$	$n-2$	MSS_{RSS}	$RSS/df_{RSS} = \frac{\sum e_i^2}{n-2}$
总平方和	TSS	$\sum(Y_i - \bar{Y}_i)^2 = \sum y_i^2$	$n-1$	MSS_{TSS}	$TSS/df_{TSS} = \frac{\sum y_i^2}{n-1}$

定义 2.10 (拟合优度)**.** 评价样本回归线对一组数据拟合程度的指标，称为**拟合优度**（goodness of fit）。其中最常用的指标是**判定系数**（coefficient of determination），记为 r^2（一元回归中）或 R^2（多元回归中）。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

根据拟合优度的定义定义 2.10 ，判定系数 r^2 可以通过如下几个公式计算得到：

$$\begin{aligned} r^2 &= \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \\ &= 1 - \frac{RSS}{TSS} = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2} \end{aligned}$$

$$r^2 = \frac{ESS}{TSS} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{(\hat{\beta}_2 x_i)^2}{\sum y_i^2} = \hat{\beta}_2^2 \left(\frac{\sum x_i^2}{\sum y_i^2}\right) = \hat{\beta}_2^2 \frac{S_X^2}{S_Y^2}$$

$$r^2 = \hat{\beta}_2^2 \left(\frac{\sum x_i^2}{\sum y_i^2}\right) = \left(\frac{\sum x_i y_i}{\sum y_i^2}\right)^2 \left(\frac{\sum x_i^2}{\sum y_i^2}\right) = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2}$$

需要注意的是，一元回归分析中判定系数 r^2 与简单相关系数 r_{X_i,Y_i} 存在某种联系（ρ_{X_i,Y_i} 为总体相关系数）：

$$\rho_{X_i,Y_i} = \frac{cov(X_i, Y_i)}{\sqrt{var(X_i)var(Y_i)}} = \frac{E\left((X_i - EX_i)(Y_i - EY_i)\right)}{\sqrt{E(X_i - EX_i)^2 E(Y_i - EY_i)^2}} \tag{2.20}$$

$$r_{X_i,Y_i} = \frac{Cov(X_i, Y_i)}{S_{X_i} \cdot S_{Y_i}} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} \tag{2.21}$$

$$\left[r_{(X_i,Y_i)}\right]^2 \simeq r^2 \tag{2.22}$$

2.2.4.2 回归系数的显著性检验——t 检验

定义 2.11 (回归系数的显著性检验)**.** 对于总体线性回归模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

利用 OLS 方法，通过样本数据回归，推断回归参数是否是否等于 0 的一种统计学检验。

在正态经典线性回归模型假设（N-CLRM）下，由于随机干扰项服从 $u_i \sim i.i.d\ N(0, \sigma^2)$，使得 $\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})$ 和 $\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$。因此，我们可以借助这一概率分布特征，对最小二乘法参数估计量 $(\hat{\beta}_2, \hat{\beta}_1)$ 进行统计学检验（显著性检验）。具体的做法是：

- 列明模型环境
- 提出理论假设
- 构造出一个可以计算得到的样本统计量（t 统计量）
- 借助假设检验流程进行分析、计算、比较，并得出显著性检验结论

因为：

$$u_i \sim i.i.d\ N(0, \sigma^2) \tag{2.23}$$

可以证明：

$$\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2}) \tag{2.24}$$

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1}) \tag{2.25}$$

又因为：

$$\sigma_{\hat{\beta}_2} = \sqrt{\frac{1}{\sum x_i^2}} \cdot \sigma \tag{2.26}$$

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{\sum X_i^2}{n \sum x_i^2}} \cdot \sigma \tag{2.27}$$

容易证明：

$$Z_{\hat{\beta}_2} = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \sim Z(0, 1) \tag{2.28}$$

$$Z_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim Z(0, 1) \tag{2.29}$$

同时，又因为：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$S_{\hat{\beta}_2} = \sqrt{\frac{1}{\sum x_i^2}} \cdot \hat{\sigma} \tag{2.30}$$

$$S_{\hat{\beta}_1} = \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}} \cdot \hat{\sigma} \tag{2.31}$$

所以进一步可以证明：

$$t^*_{\hat{\beta}_2} = \frac{\hat{\beta}_2 - \beta_2}{S_{\hat{\beta}_2}} \sim t(n-2) \tag{2.32}$$

$$t^*_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t(n-2) \tag{2.33}$$

对于模型：

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i \\ u_i &\sim i.i.d\ N(0, \sigma^2) \\ Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \end{aligned}$$

对总体参数 (β_2, β_1) 分别提出如下的假设：

$$\beta_2 : \begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{cases}$$

$$\beta_1 : \begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

根据原假设 H_0，可以得到：

$$t^*_{\hat{\beta}_2} = \frac{\hat{\beta}_2 - \beta_2}{S_{\hat{\beta}_2}} = \frac{\hat{\beta}_2}{S_{\hat{\beta}_2}} = \frac{\hat{\beta}_2}{\sqrt{\frac{1}{\sum x_i^2}} \cdot \hat{\sigma}} \sim t(n-2) \tag{2.34}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$t^*_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{\sqrt{\frac{\sum X_i^2}{n\sum x_i^2}} \cdot \hat{\sigma}} \sim t(n-2) \tag{2.35}$$

若给定显著性水平 $\alpha = 0.01$ 和自由度 $(n-2)$，很快可以得到 t 分布的查表 t 值，也即 $t_{(1-\alpha/2)}(n-2)$。然后比较样本 t 统计量 $(t^*_{\hat{\beta}_2}, t^*_{\hat{\beta}_1})$ 与理论 t 分布查的表 t 值 $(t_{(1-\alpha/2)}(n-2))$ 的关系。

根据如下法则做出参数 β_2 的显著性检验结论：

- 如果 $t^*_{\hat{\beta}_2} > t_{(1-\alpha/2)}(n-2)$，则表明参数 β_2 的 t 检验在 $\alpha = 0.01$ 水平下是**显著**的，也即显著地拒绝 $H_0 : \beta_2 = 0$，从而接受 $H_1 : \beta_2 \neq 0$。
- 如果 $t^*_{\hat{\beta}_2} \leq t_{(1-\alpha/2)}(n-2)$，则表明参数 β_2 的 t 检验在 $\alpha = 0.01$ 水平下是**不显著**的，也即不能显著地拒绝 $H_0 : \beta_2 = 0$，从而只能暂时接受 $H_0 : \beta_2 = 0$。

同样地，根据如下法则做出参数 β_1 的显著性检验结论：

- 如果 $t^*_{\hat{\beta}_1} > t_{(1-\alpha/2)}(n-2)$，则表明参数 β_1 的 t 检验在 $\alpha = 0.01$ 水平下是**显著**的，也即显著地拒绝 $H_0 : \beta_1 = 0$，从而接受 $H_1 : \beta_1 \neq 0$。
- 如果 $t^*_{\hat{\beta}_1} \leq t_{(1-\alpha/2)}(n-2)$，则表明参数 β_2 的 t 检验在 $\alpha = 0.01$ 水平下是**不显著**的，也即不能显著地拒绝 $H_0 : \beta_1 = 0$，从而只能暂时接受 $H_0 : \beta_1 = 0$。

2.2.4.3 模型整体显著性检验——F 检验

2.2.4.4 模型整体显著性检验

对于总体线性回归模型，

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

利用 LS 方法，通过样本数据回归，推断模型的线性关系是否成立的一种统计学检验。

在正态经典线性回归模型假设（N-CLRM）下，随机干扰项服从 $u_i \sim i.i.d\ N(0, \sigma^2)$。因此，我们可以借助这一概率分布特征，利用最小二乘法参数估计量 $(\hat{\beta}_2, \hat{\beta}_1)$ 结果，对总体回归模型进行整体显著性检验——F 检验。具体的做法是：

- 列明模型环境
- 提出理论假设

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 构造出一个可以计算得到的样本统计量（F 统计量）
- 借助假设检验流程进行分析、计算、比较，并得出显著性检验结论

因为:

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_i + u_i \\
u_i &\sim i.i.d\ N(0, \sigma^2)
\end{aligned}
$$

在 N-CLRM 下容易证明:

$$
Y_i \sim i.i.d\ N(\beta_1 + \beta_2 X_i, \sigma^2)
$$

进一步可以证明：

$$
ESS = \sum (\hat{Y}_i - \bar{Y})^2 \sim \chi^2(df_{ESS}) \tag{2.36}
$$

$$
RSS = \sum (Y_i - \hat{Y}_i)^2 \sim \chi^2(df_{RSS}) \tag{2.37}
$$

从而得到:

$$
\begin{aligned}
F^* &= \frac{ESS/df_{ESS}}{RSS/df_{RSS}} = \frac{MSS_{ESS}}{MSS_{RSS}} \\
&= \frac{\sum (\hat{Y}_i - \bar{Y})^2/df_{ESS}}{\sum (Y_i - \hat{Y}_i)^2/df_{RSS}} \\
&\sim F(df_{ESS}, df_{RSS})
\end{aligned} \tag{2.38}
$$

对于模型：

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_i + u_i \\
u_i &\sim i.i.d\ N(0, \sigma^2) \\
Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i
\end{aligned}
$$

对线性模型的斜率参数 β_2 提出如下的假设：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$\beta_2 : \begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{cases}$$

根据原假设 H_0，可以得到：

$$\begin{aligned} F^* &= \frac{ESS/df_{ESS}}{RSS/df_{RSS}} = \frac{MSS_{ESS}}{MSS_{RSS}} \\ &= \frac{\sum(\hat{Y}_i - \bar{Y})^2/df_{ESS}}{\sum(Y_i - \hat{Y}_i)^2/df_{RSS}} \\ &= \frac{\hat{\beta}_2^{\,2}\sum x_i^2}{\sum e_i^2/(n-2)} \\ &= \frac{\hat{\beta}_2^{\,2}\sum x_i^2}{\hat{\sigma}^2} \end{aligned}$$

若给定显著性水平 $\alpha = 0.01$ 和样本数 (n)，很快可以得到 F 分布的查表 F 值，也即 $F_{(1-\alpha)}(1, n-2)$，然后比较其与样本 F 统计量 F^* 的关系。

根据如下法则做出总体回归模型整体显著性检验结论：

- 如果 $F^* > F_{(1-\alpha)}(1, n-2)$，则表明总体回归模型的 F 检验在 $\alpha = 0.01$ 水平下是**显著**的，也即显著地拒绝 $H_0 : \beta_2 = 0$，从而接受 $H_1 : \beta_2 \neq 0$。
- $F^* \leq F_{(1-\alpha)}(1, n-2)$，则表明总体回归模型的 F 检验在 $\alpha = 0.01$ 水平下是**不显著**的，也即不能显著地拒绝 $H_0 : \beta_2 = 0$，从而只能暂时接受 $H_0 : \beta_2 = 0$。

2.2.5 如何进行样本外预测?

样本外预测包括均值预测（mean prediction）和个值预测（individual prediction）两个内容。

定义 2.12 (均值预测)**.** 均值预测（mean prediction），是一种对样本外条件均值的预测。正式地，给定 X_0，预测 Y_i 的条件均值，也即 $E(Y|X = X_0)$

定义 2.13 (个值预测)**.** 个值预测（individual prediction），是一种对样本外个体值的预测。正式地，给定 X_0，预测 Y_i 的个别值，也即 $(Y_0|X = X_0)$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

定义 2.14 (点预测)**.** 是通过一次性的样本数据拟合结果，得到的关于样本外因变量的、可信的单一个值预测（point forecast）。

定义 2.15 (区间预测)**.** 是利用随机变量的分布特性，借助于一次性的样本数据拟合结果，得到的关于样本外因变量的、可信的取值范围预测（interval forecast）。

事实上，无论是进行均值预测还是个值预测，关键思路是我们将能够证明：

- 样本回归函数的拟合值 $\hat{Y}_0$ 很独特，它既是条件均值 $E(Y|X = X_0)$ 的无偏估计，也是个值 $(Y_0|X = X_0)$ 的无偏估计。也就意味着 $\hat{Y}_0$ 是它们二者的无偏的**点预测值**。
- 随机变量 $\hat{Y}_0$ 和 Y_0 都服从正态分布，并且基于此我们可以构造并计算出均值预测的置信区间和个值预测的置信区间

2.2.5.1 样本外均值预测

对于模型：

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i \\ u_i &\sim i.i.d\ N(0, \sigma^2) \\ Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \end{aligned}$$

容易证明随机变量 $\hat{Y}_0$ 的期望和方差分别为：

$$\begin{aligned} \mu_{\hat{Y}_0} &= E(\hat{Y}_0) \\ &= E(\hat{\beta}_1 + \hat{\beta}_2 X_0) \\ &= \beta_1 + \beta_2 X_0 \\ &= E(Y|X_0) \end{aligned} \tag{2.39}$$

$$\begin{aligned} var(\hat{Y}_0) &= \sigma^2_{\hat{Y}_0} \\ &= E(\hat{\beta}_1 + \hat{\beta}_2 X_0) \\ &= \sigma^2 \left(\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right) \end{aligned} \tag{2.40}$$

而且服从如下正态分布：

$$\begin{aligned}\hat{Y}_0 &\sim N(\mu_{\hat{Y}_0}, \sigma^2_{\hat{Y}_0}) \\ \hat{Y}_0 &\sim N\left(E(Y|X_0), \sigma^2\left(\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right)\right)\end{aligned} \tag{2.41}$$

进一步地，可以构造 t 统计量：

$$t_{\hat{Y}_0} = \frac{\hat{Y}_0 - E(Y|X_0)}{S_{\hat{Y}_0}} \sim t(n-2) \tag{2.42}$$

其中：

$$S_{\hat{Y}_0} = \sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right)} \tag{2.43}$$

给定显著性水平 α 的情况下，可以查表得到理论 t 值 $t_{1-\alpha/2}(n-2)$，从而可以计算得到均值预测的置信区间：

$$\hat{Y}_0 - t_{1-\alpha/2}(n-2) \cdot S_{\hat{Y}_0} \leq E(Y|X_0) \leq \hat{Y}_0 + t_{1-\alpha/2}(n-2) \cdot S_{\hat{Y}_0} \tag{2.44}$$

2.2.5.2 样本外个值预测

对于模型：

$$\begin{aligned}Y_i &= \beta_1 + \beta_2 X_i + u_i \\ u_i &\sim i.i.d\ N(0, \sigma^2) \\ Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i\end{aligned}$$

容易证明随机变量 Y_0 服从如下的正态分布：

$$\begin{aligned}Y_0 &\sim N(\mu_{Y_0}, \sigma^2_{Y_0}) \\ Y_0 &\sim N\left(\beta_1 + \beta_2 X_0, \sigma^2\right)\end{aligned} \tag{2.45}$$

因为已经证明了：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$\begin{aligned}\hat{Y}_0 &\sim N(\mu_{\hat{Y}_0}, \sigma^2_{\hat{Y}_0}) \\ \hat{Y}_0 &\sim N\left(E(Y|X_0), \sigma^2\left(\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right)\right)\end{aligned} \tag{2.46}$$

因此我们可以构造一个新的正态随机变量 $(Y_0 - \hat{Y}_0)$：

$$\begin{aligned}(Y_0 - \hat{Y}_0) &\sim N\left(\mu_{(Y_0-\hat{Y}_0)}, \sigma^2_{(Y_0-\hat{Y}_0)}\right) \\ (Y_0 - \hat{Y}_0) &\sim N\left(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right)\right)\end{aligned} \tag{2.47}$$

进一步地，可以构造 t 统计量：

$$t_{(Y_0-\hat{Y}_0)} = \frac{(Y_0 - \hat{Y}_0)}{S_{(Y_0-\hat{Y}_0)}} \sim t(n-2) \tag{2.48}$$

其中：

$$S_{(Y_0-\hat{Y}_0)} = \sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right)} \tag{2.49}$$

给定显著性水平 α 的情况下，可以查表得到理论 t 值 $t_{1-\alpha/2}(n-2)$，从而可以计算得到个值预测的置信区间：

$$\hat{Y}_0 - t_{1-\alpha/2}(n-2) \cdot S_{(Y_0-\hat{Y}_0)} \leq (Y_0|X_0) \leq \hat{Y}_0 + t_{1-\alpha/2}(n-2) \cdot S_{(Y_0-\hat{Y}_0)} \tag{2.50}$$

2.2.6 怎样看懂 EViews 回归分析报告结果?

表 2.2: 一元回归方程报告解读 1：标题信息区

row	content	mark
第 1 行	Dependent Variable: SPENDS	因变量
第 2 行	Method: Least Squares	方法
第 3 行	Date: Time:	日期、时间
第 4 行	Sample: 1 10	样本框

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 2.2: 一元回归方程报告解读 1：标题信息区

row	content	mark
第 5 行	Included observations: 10	样本数

表 2.3: 一元回归方程报告解读：回归结果区

row	column	content	mark	math
第 1 行	第 1 列	Variable	方程变量	
第 2 行	第 1 列	C	回归方程的截距	
第 3 行	第 1 列	INCOME	回归方程的变量 X	
第 1 行	第 2 列	Coefficient	系数 (回归参数的点估计值)	
第 2 行	第 2 列	…	截距系数	$\hat{\beta}_1$
第 3 行	第 2 列	…	斜率系数	$\hat{\beta}_2$
第 1 行	第 3 列	Std. Error	系数的样本标准差	
第 2 行	第 3 列	…	截距系数的样本标准差	$S_{\hat{\beta}_1}$
第 3 行	第 3 列	…	斜率系数的样本标准差	$S_{\hat{\beta}_2}$
第 1 行	第 4 列	t-Statistic	系数的样本 t 值	
第 2 行	第 4 列	…	截距系数的样本 t 值	$t^*_{\hat{\beta}_1}$
第 3 行	第 4 列	…	斜率系数的样本 t 值	$t^*_{\hat{\beta}_2}$
第 1 行	第 5 列	Prob.	系数的样本 t 值对应的概率值 p	

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 2.3: 一元回归方程报告解读：回归结果区

row	column	content	mark	math
第 2 行	第 5 列	...	截距系数的样本 t 值对应的概率值 p	$p(\alpha/2, t^*_{\hat{\beta}_1}, f)$
第 3 行	第 5 列	...	斜率系数的样本 t 值对应的概率值 p	$p(\alpha/2, t^*_{\hat{\beta}_2}, f)$

表 2.4: 一元回归方程报告解读 3：回归诊断区

row	column	content	mark	math
第 1 行	第 1 列	R-squared	判定系数	R^2
第 2 行	第 1 列	Adjusted R-squared	调整判定系数	$\bar{R}^2$
第 3 行	第 1 列	S.E. of regression	回归标准误	$\hat{\sigma}$
第 4 行	第 1 列	Sum squared resid	残差平方和	$RSS = \sum e_i^2$
第 5 行	第 1 列	Log likelihood	对数似然值	
第 6 行	第 1 列	F-statistic	回归方程的 F 统计量值	F^*
第 7 行	第 1 列	Prob(F-statistic)	F 统计量值对应的概率值 p	$p(\alpha, F^*, f_1, f_2)$
第 1 行	第 3 列	Mean dependent var	因变量的均值	$\bar{Y}$
第 2 行	第 3 列	S.D. dependent var	应变量的标准差	$S_{Y_i} = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{(n-1)}}$
第 3 行	第 3 列	Akaike info criterion	AIC 指数	
第 4 行	第 3 列	Schwarz criterion	SC 指数	

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 2.4: 一元回归方程报告解读 3：回归诊断区

row	column content	mark	math
第 5 行	第 3 列 Hannan-Quinn criter.	HQ 指数	
第 6 行	第 3 列 Durbin-Watson stat	DW 统计量	德宾沃森统计量

2.3 实验内容

（1）利用 eviews 菜单（Quick ⇒ Estimate Equation）进行线性回归模型分析，学会读懂回归报告结果。

（2）利用 EViews 命令代码编程，回归理论知识点，验证回归分析的全部细节步骤。

a. 计算直线回归方程的回归系数 $\hat{\beta}_1$、$\hat{\beta}_2$

b. 计算回归误差方差（$\hat{\sigma}^2$）其标准差 ($\hat{\sigma}$)

c. 计算回归截距系数的样本方差 ($S^2_{\hat{\beta}_1}$) 及其样本标准差 ($S_{\hat{\beta}_1}$)；计算回归斜率系数的样本方差 ($S^2_{\hat{\beta}_2}$) 及其样本标准差 ($S_{\hat{\beta}_2}$)

d. 计算平方和分解式（TSS、ESS、RSS）及其相对应的自由度（df_{TSS}，df_{ESS}，df_{RSS} ），并写出方差分析表（**ANOV** table）

e. 计算得到两个变量的样本相关系数 r，以及回归方程的判定系数 r^2，并比较相关系数与判定系数的关系（数量关系）

f. 计算回归系数的样本 t 值 ($t^*_{\hat{\beta}_2}$ 和 $t^*_{\hat{\beta}_1}$)

g. 用 eviews 命令计算得到理论 t 值（$t_{1-\alpha/2}(n-2)$，给定 $\alpha = 0.05$），并对回归系数显著性进行假设检验。

h. 计算回归方程整体显著性检验的的样本 F 值（F_*）

i. 用 eviews 命令计算得到理论 F 值（$F_{1-\alpha}(df_1, df_2) = F_{1-\alpha}(df_{ESS}, df_{RSS})$，给定 $\alpha = 0.05$ ），并对模型整体显著性进行假设检验

j. 给定样本外的 X_0 ，计算点预测值 $\hat{Y}_0$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

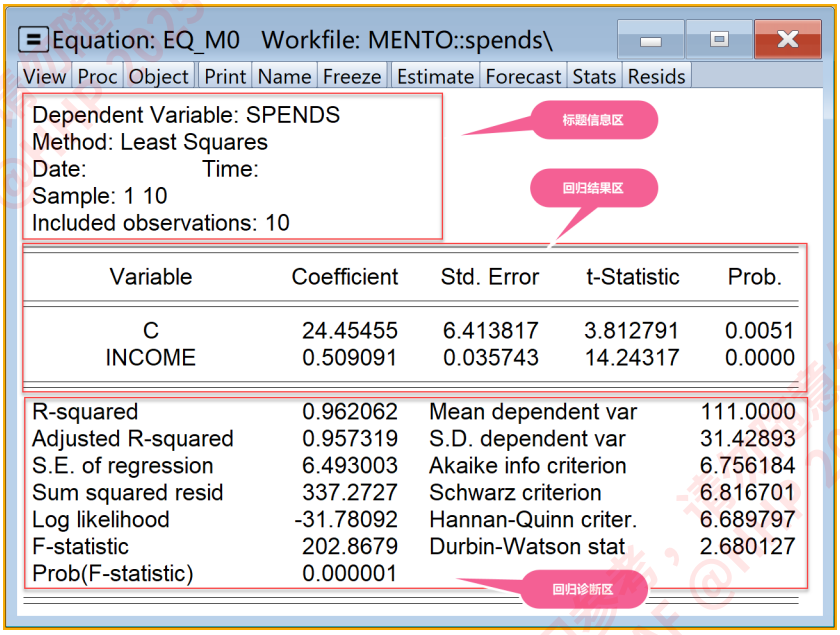


图 2.1: 看懂 EViews 回归分析报告

k. 计算均值预测下的样本标准差 $S_{(\hat{Y}_0)}$

l. 计算均值预测的置信区间 $\hat{Y}_0 - t_{1-\alpha/2}(n-2) \cdot S_{\hat{Y}_0} \leq E(Y \mid X = X_0) \leq \hat{Y}_0 + t_{1-\alpha/2}(n-2) \cdot S_{\hat{Y}_0}$

m. 计算个值预测下的样本标准差 $S_{(Y_0 - \hat{Y}_0)}$

n. 计算个值预测的置信区间 $\hat{Y}_0 - t_{\alpha/2}(n-2) \cdot S_{Y_0 - \hat{Y}_0} \leq E(Y \mid X = X_0) \leq \hat{Y}_0 + t_{\alpha/2}(n-2) \cdot S_{Y_0 - \hat{Y}_0}$

2.4 实验准备

2.4.1 实验软件

本次实验需要提前准备好如下软件：

- 统计分析软件 Eviews 9.0 版本及以上
- 公式编辑软件 Mathtype 6.0 版本及以上
- 写作编辑软件 Office Word/Excel 2010 版本及以上

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 浏览器软件 chrome 66.0 版本及以上或 360 极速浏览器 9.5 版本及以上

2.4.2 实验材料

蒙特卡洛模拟数据：表 2.5 给出给出了 10 个家庭在 spends 家庭收入，income 家庭支出等方面的数据。

表 2.5: 家庭收入和家庭支出 (n=10)

obs	spends	income
1	70	80
2	65	100
3	90	120
4	95	140
5	110	160
6	115	180
7	120	200
8	140	220
9	155	240
10	150	260

变量说明见表 2.6 ：

表 2.6: 变量定义及说明

variable	label
obs	样本编号
spends	家庭收入
income	家庭支出

上机数据说明：

- 所有同学的 X 数据 (家庭收入 *income*) 都一样！
- 每个同学的 Y 数据 (家庭支出 *spends*) 都不一样！

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 请下载数据表后，找到自己的两列，并进行 Excel 预处理 (以便导入到 Eviews!)

如果消费和收入之间可以构造如下一元线性回归模型：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \quad \text{(SRM)} \tag{2.51}$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad \text{(SRF)} \tag{2.52}$$

2.4.3 实验规则

Eviews 操作会得到各种计算结果，我们需要把计算结果以对象的形式保存在 Eviews 的工作文件内。因此我们提前设计好了各种对象保存命名规则，以便计算过程中的连续性、一致性调用。表 2.7 给出了记号、含义、对应英文词和示例。

表 2.7: 可能用到的 Eviews 对象的记号表达

remark	mean	word	example
upr	字母大写	uper	样本序列 Y_i，可记为 y_upr
lwr	字母小写	lower	离差序列 y_i，可记为 y_lwr
avr	均值	average	样本均值 $\bar{X}$，可记为 x_avr
sqr	平方	square	样本序列 Y_i^2，可记为 y_upr_sqr
hat	估计值	hat	$\hat{\beta}_2$ 和 $\hat{Y}_i$，可分别记为 b2_hat 和 y_upr_hat
str	星号	star	斜率系数的样本 t 值 $t^*_{\hat{\beta}_2}$，可记为 t_str_b2_hat
scl	标量	scalar	样本均值 $\bar{X}$ 作为标量，可记为 x_avr_scl
ser	序列	series	样本均值 X_i 可作为有 n 个数值的序列，可记为 x_avr_ser
sgm	方差参数	sigma	回归误差方差 $\hat{\sigma}^2$，可记为 sgm_hat_sqr
b	回归参数	beta	斜率参数 β_2，可记为 b2
mns	减号	minus	样本标准差 $S_{(\hat{Y}_0 - Y_0)}$，可记为 s_y0h_mns_y0
fst	预测	forecast	均值预测值 $E(Y\|X = X_0)$，可记为 fst_exp
exp	期望值	expect	均值预测值 $E(Y\|X = X_0)$，可记为 fst_ex
ind	个值	individual	个值预测值 $(Y_0\|X = X_0)_L$，可记为 fst_ind
lft	左界	left	均值区间预测的左界值 $E(Y\|X = X_0)_L$，可记为 y_exp_lft
rht	右界	right	均值区间预测的右界值 $E(Y\|X = X_0)_R$，可记为 y_exp_rht

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 2.8: 计算对象、表达式及 Eviews 命名

name_chn	cat_eng	math	name_eviews
自变量 X 序列	series	X_i	x_upr
因变量 Y 序列	series	Y_i	y_upr
X 均值标量	scalar	$\bar{X}$	x_avr_scl
Y 均值标量	scalar	$\bar{Y}$	y_avr_scl
X 均值序列	series	$\bar{X}_i$	x_avr_ser
Y 均值序列	series	$\bar{Y}_i$	y_avr_ser
样本数 n	scalar	n	n
X 平方序列	series	X_i^2	x_upr_sqr
Y 平方序列	series	Y_i^2	y_upr_sqr
XY 序列	series	X_iY_i	xy_upr
离差 x 序列	series	x_i	x_lwr
离差 y 序列	series	y_i	y_lwr
离差 x 平方序列	series	x_i^2	x_lwr_sqr
离差 y 平方序列	series	y_i^2	y_lwr_sqr
离差 xy 序列	series	x_iy_i	xy_lwr
斜率估计值	scalar	$\hat{\beta}_2$	b2_hat
截距估计值	scalar	$\hat{\beta}_1$	b1_hat
回归预测值 Y 帽序列	series	$\hat{Y}_i$	y_upr_hat
回归预测值的离差 y 帽序列	series	$\hat{y}_i$	y_lwr_hat
残差序列	series	e_i	ei
残差平方序列	series	e_i^2	ei_sqr
回归误差方差	scalar	$\hat{\sigma}^2$	sgm_hat_sqr
回归误差标准差	scalar	$\hat{\sigma}$	sgm_hat
斜率的样本方差	scalar	$S^2_{\hat{\beta}_2}$	ss_b2_hat
斜率的样本标准差	scalar	$S_{\hat{\beta}_2}$	s_b2_hat
截距的样本方差	scalar	$S^2_{\hat{\beta}_1}$	ss_b1_hat
截距的样本标准差	scalar	$S_{\hat{\beta}_1}$	s_b1_hat
总平方和	scalar	TSS	tss
残差平方和	scalar	RSS	rss
回归平方和	scalar	ESS	ess
总平方和的自由度	scalar	df_{TSS}	df_tss
残差平方和的自由度	scalar	df_{RSS}	df_rss

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 2.8: 计算对象、表达式及 Eviews 命名

name_chn	cat_eng	math	name_eviews
回归平方和的自由度	scalar	df_{ESS}	df_ess
简单相关系数	scalar	r	r
判定系数	scalar	R^2	r_sqr
斜率系数的 t 统计量	scalar	$t^*_{\hat{\beta}_2}$	t_str_b2_hat
截距系数的 t 统计量	scalar	$t^*_{\hat{\beta}_1}$	t_str_b1_hat
理论 t 值	scalar	$t_{1-\alpha/2}(n-k)$	t_value
回归方程的 F 统计量	scalar	F^*	f_str
理论 F 值	scalar	$F_{1-\alpha}(k-1, n-k)$	f_value
样本外数据	scalar	X_0	x0
样本外回归值	scalar	$\hat{Y}_0$	y0_hat
均值预测	scalar	$E(Y\|X=X_0)$	fst_exp
$\hat{Y}_0$ 的样本标准差	scalar	$S_{\hat{Y}_0}$	s_y0h
均值区间预测的左界	scalar	$E(Y\|X=X_0)_L$	y_exp_lft
均值区间预测的右界	scalar	$E(Y\|X=X_0)_R$	y_exp_rht
个值预测	scalar	$(Y_0\|X=X_0)$	fst_ind
$\hat{Y}_0 - Y_0$ 的样本标准差	scalar	$S_{(\hat{Y}_0-Y_0)}$	s_mns
个值区间预测的左界	scalar	$(Y_0\|X=X_0)_L$	y_ind_lft
个值区间预测的右界	scalar	$(Y_0\|X=X_0)_R$	y_ind_rht

2.5 主要实验步骤

2.5.1 新建工作文件并导入数据

Eviews 操作目标：构建工作文件，成功导入数据。

Eviews 操作思路：利用 EViews 代码创建工作文件并导入数据。

在命令视窗中依次输入并运行如下 EViews 代码：

```
'创建工作文件（工作文件名=mento，子页命名=spend），无结构无日期，样本数为10
wfcreate(wf=mento,page=spend)  u 10
```

```
'导入外部数据，路径为d:\github\books\data\lab2-mento-demon.xlsx
import d:\github\books\data\lab2-mento-demon.xlsx
```

在工作文件视窗下，可以看到创建的工作文件和导入的数据，可以双击查看（见图 2.2 ）：

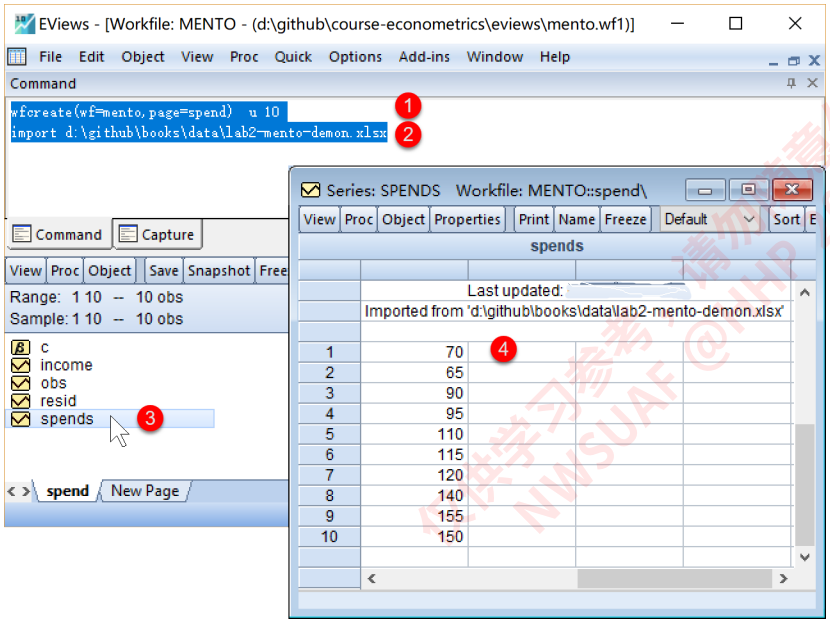


图 2.2: 创建工作文件并导入数据

提示

上述过程也可以通过菜单操作实现：

创建工作文件：

（1）EViews 主窗口上依次点击 ⇒ `File`⇒ `New`⇒ `Workfile`

（2）进行 workfile create 引导设置：

- workfile structure type: `unstructured/undatede`
- data range：默认设置
- workfile names(optional):
  - WF: `mento`
  - Page: `spends` （建议命名 **spends**）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

导入外部数据：EViews 主窗口上依次点击 ⇒ `File` ⇒ `Import` ⇒ `Import from file ...`

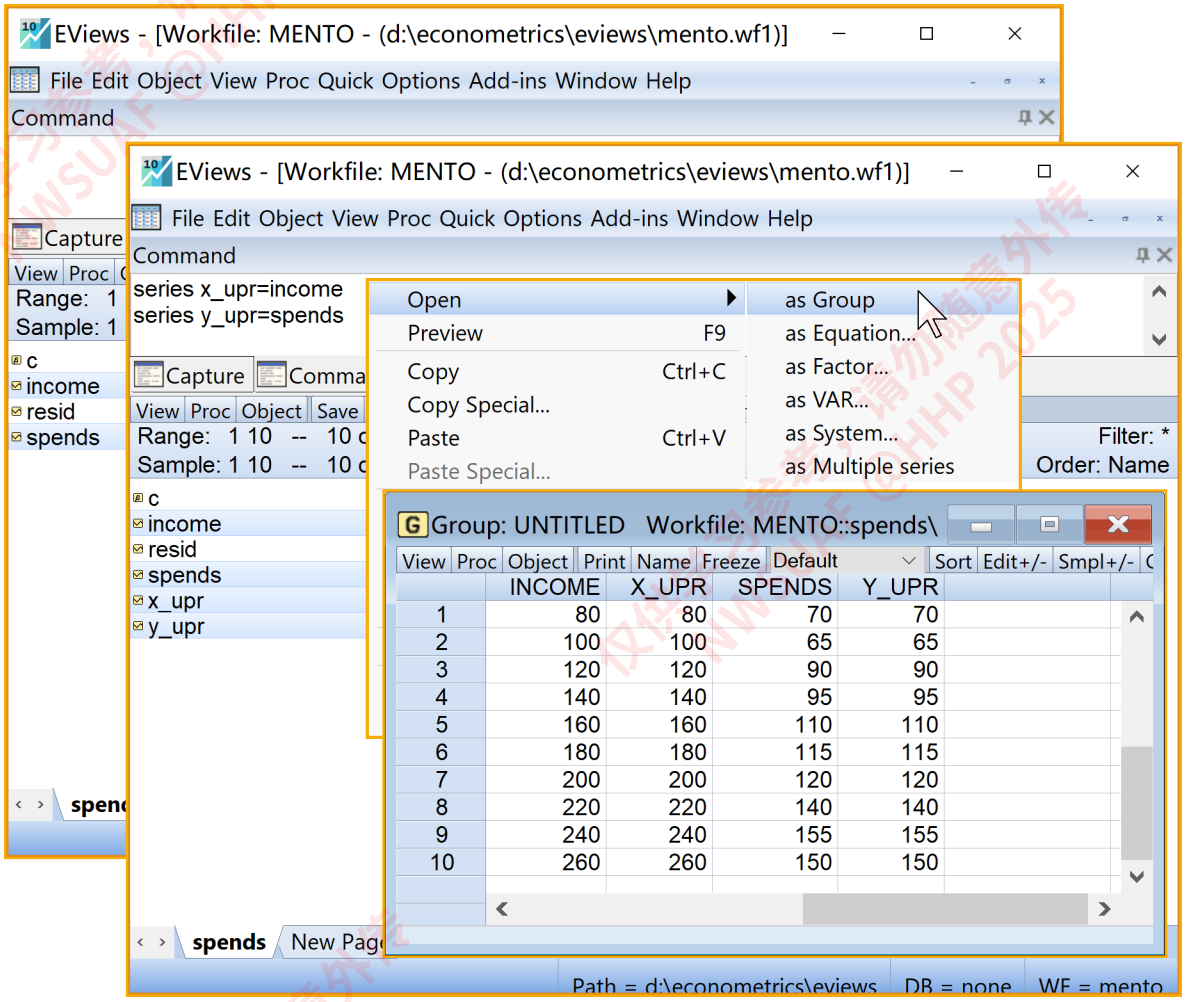


图 2.3: 导入数据的 Eviews 视窗

2.5.2 进行 Eviews 回归分析

Eviews 操作目标：得到回归方程，查看回归结果

Eviews 操作思路：构建回归方程对象。回归方程为：

$$spends_i = + \beta_1 + \beta_2 income_i + u_i \tag{2.53}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 生成线性回归方程对象eq_main
equation eq_m0.ls spends c income
```

在工作文件视窗下，可以看到如下新生成的方程对象，可以双击查看eq_m0（见图 2.4 ）：

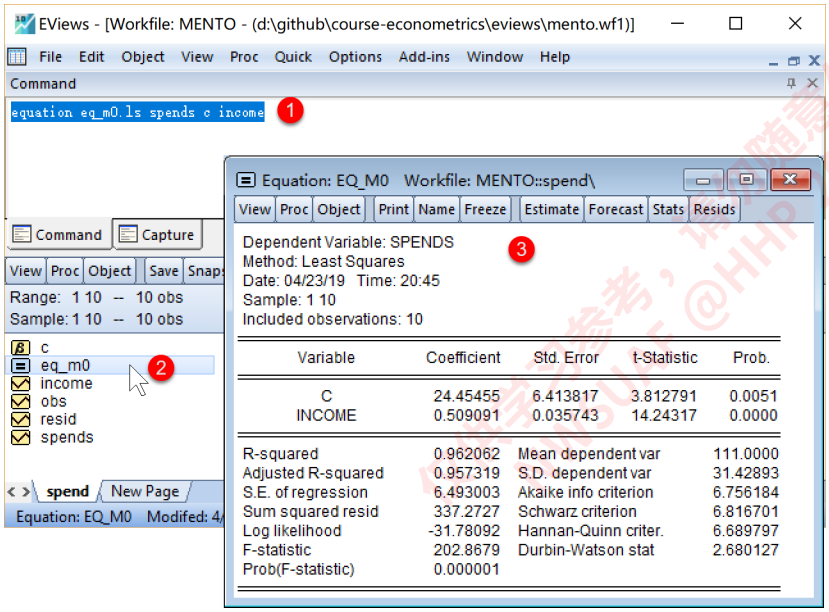


图 2.4: 生成回归方程对象

提示

上述过程也可以通过菜单操作实现：

（1）进入方程估计的引导界面：

- EViews 主窗口上依次点击点 `Quick` ⇒ `Estimation Equation`

（2）完成方程估计的引导设置。

- 设置方程。`Equation Estimation` ⇒ `Specification` ⇒ `Equation specification` 中依次输入变量 `spend c income`（注意变量之间的空格，以及截距 `c`）
- 选择估计方程。`Estimation settings` 中的 `Method` 下拉框 ⇒ 下拉选择 `LS - Least Squares (NLS and ARMA)`

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 完成设置，点击 OK

（3）命名并保存回归方程：

- 在未命名的方程对象UNTITLED 视窗下，点击菜单栏 Name ⇒ 输入命名 eq_m0（建议命名）⇒ 完成命名，点击 Ok

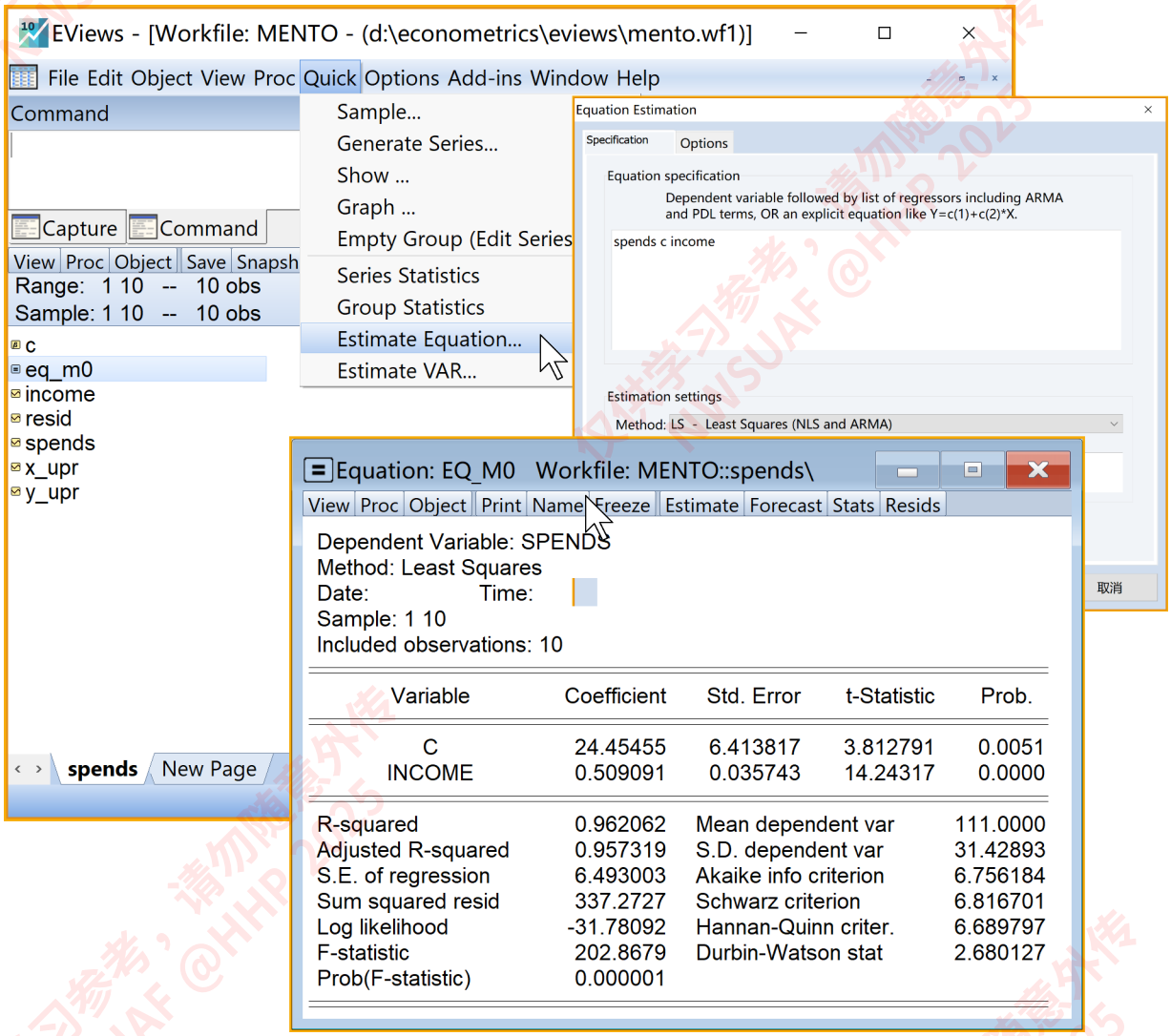


图 2.5: 主回归模型 Eviews 操作

回归方程结果见图 2.6 ：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

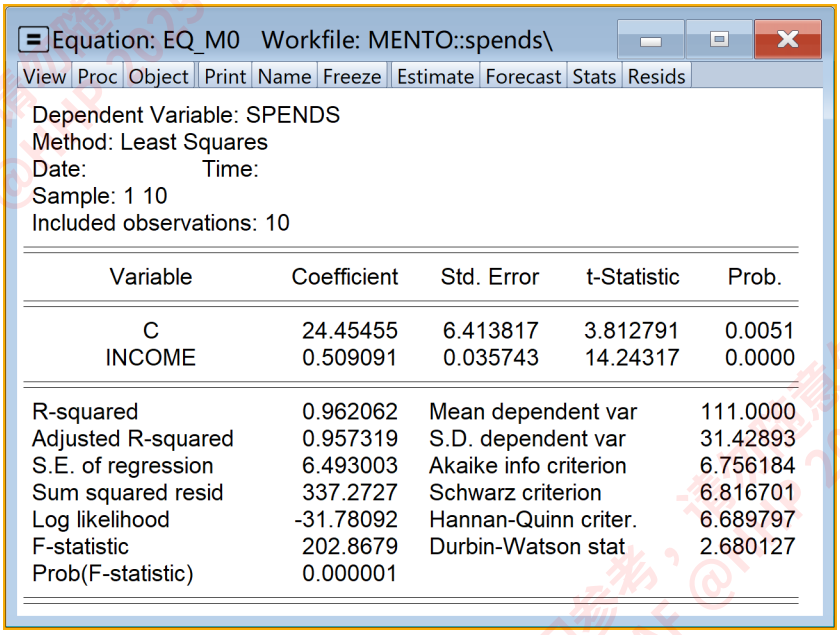


图 2.6: 主回归模型 Eviews 报告

家庭消费支出回归模型的简要回归报告如下：

$$\widehat{spends}_i = + \hat{\beta}_1 + \hat{\beta}_2 income_i \tag{2.54}$$

$$\begin{array}{lll} \widehat{spends} = & +24.45 & +0.51 income_i \\ (s) & (6.4138) & (0.0357) \\ (t) & (+3.81) & (+14.24) \end{array} \tag{2.55}$$

2.5.3 构建几个重要变量对象

Eviews 操作目标：构造几个重要 EViews 对象，便于后面分析使用

Eviews 操作思路：转换变量名；均值 $\bar{X}$ 和 $\bar{Y}$ 的标量对象；样本数 n 的标量对象；均值 $\bar{X}$ 和 $\bar{Y}$ 的序列对象

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$\bar{X} = \frac{\sum_{X_i}}{n}$$
$$\bar{Y} = \frac{\sum_{Y_i}}{n}$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 转换变量名
series x_upr=x                          '序列X
series y_upr=y                          '序列Y

'计算三个重要标量
scalar x_avr_scl=@mean(x_upr)           'X的均值(标量)
scalar y_avr_scl=@mean(y_upr)           'Y的均值(标量)
scalar n=@obs(x_upr)                    '样本数n(标量)

'计算均值序列
series x_avr_ser=@mean(x_upr)           'X的均值(序列)
series y_avr_ser=@mean(y_upr)           'Y的均值(序列)
```

在工作文件视窗下，可以看到如下新生成的对象，可以双击查看（见图 2.7 和图 2.8 ）：

- X_i 的序列对象x_upr
- Y_i 的序列对象y_upr
- 均值 $\bar{X}$ 的标量对象x_avr_scl
- 均值 $\bar{Y}$ 的标量对象y_avr_scl
- 样本数 n 的标量对象n
- 均值 $\bar{X}$ 的标量对象x_ser_scl
- 均值 $\bar{Y}$ 的标量对象y_ser_scl

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

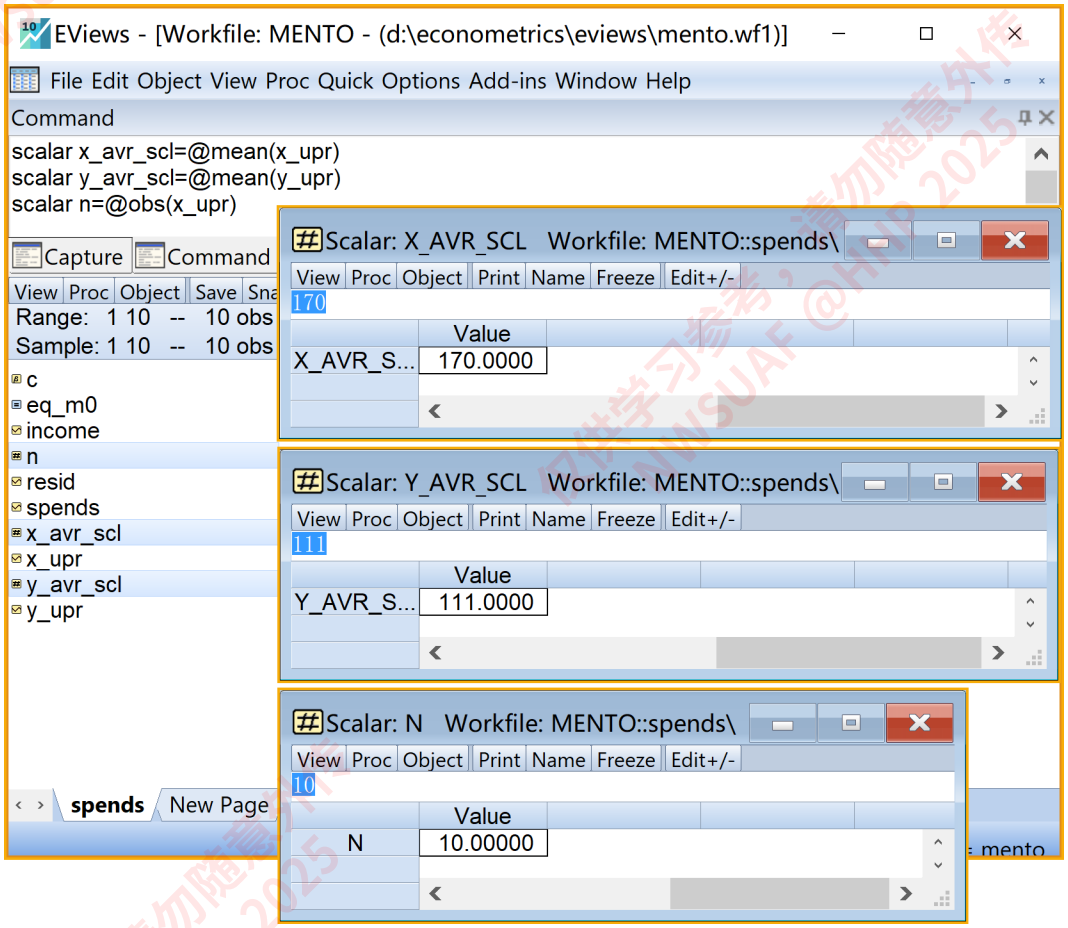


图 2.7: 构建几个重要变量对象 1

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

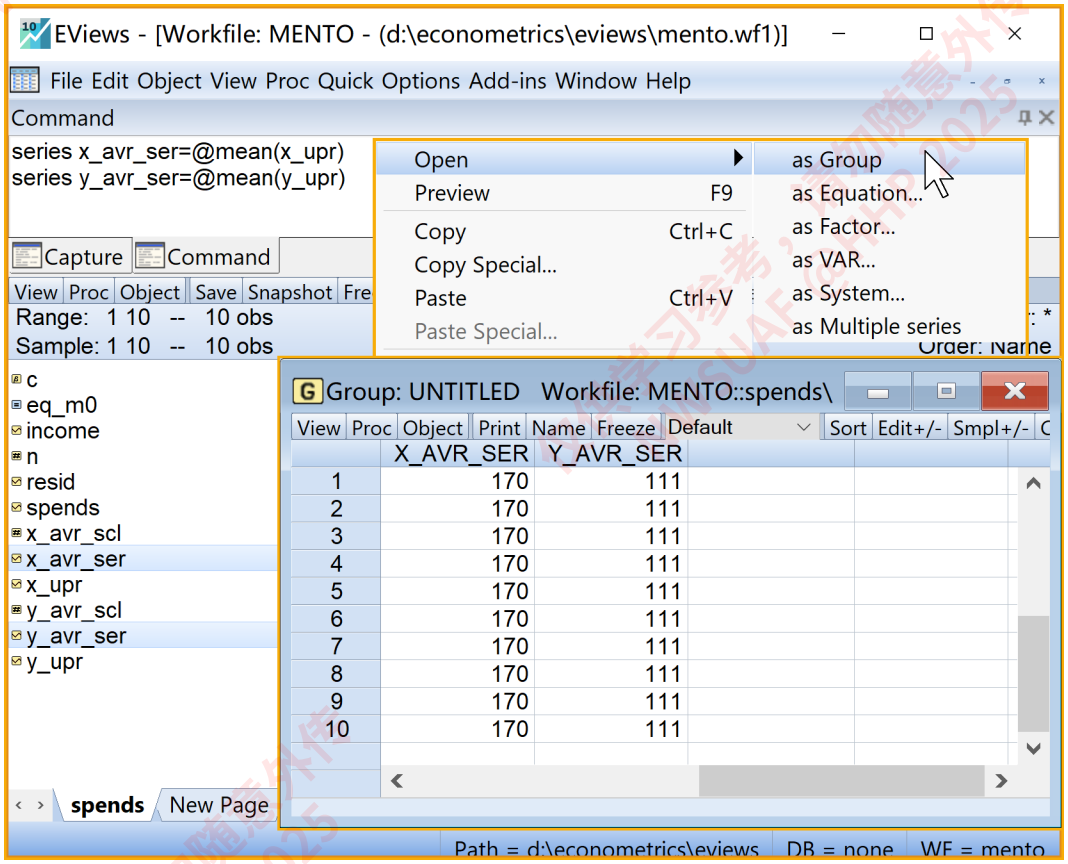


图 2.8: 构建几个重要变量对象 2

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

2.5.4 计算 FF 原序列 (5 个序列 series)

目标：得到 FF 原序列 (5 个序列 **series**)

思路：分别计算得到 $X_i; Y_i; X_i^2; Y_i^2; X_iY_i$ 五个序列，并以组对象的形式保存

在命令视窗中依次输入并运行如下 EViews 代码：

```
'计算FF，并以group形式打开
series x_upr_sqr=x_upr^2                '序列X^2
series y_upr_sqr=y_upr^2                '序列Y^2
series xy_upr= x_upr*y_upr              '序列XY
group ff_upr x_upr y_upr x_upr_sqr y_upr_sqr xy_upr     '构建FF序列组（group）
```

在工作文件视窗下，可以看到如下新生成的对象，可以双击查看（见图 2.9 ）：

- X_i^2 的序列对象x_upr_sqr
- Y_i^2 的序列对象y_upr_sqr
- X_iY_i 的序列对象xy_upr
- FF$(X_i; Y_i; X_i^2; Y_i^2, X_iY_i)$ 的组对象ff_upr

Eviews 操作：（Eviews 具体操作过程见图 2.9）

2.5.5 计算 ff 离差序列 (5 个序列 series)

目标：得到 FF 原序列 (5 个序列 **series**)

思路：分别计算得到 $(x_i; y_i; x_i^2; y_i^2, x_iy_i)$ 五个序列，并以组对象的形式保存。

$$
\begin{aligned}
x_i &= (X_i - \bar{X}) \\
y_i &= (Y_i - \bar{Y}) \\
x_i^2 &= (X_i - \bar{X})^2 \\
y_i^2 &= (Y_i - \bar{Y})^2 \\
x_iy_i &= (X_i - \bar{X})(Y_i - \bar{Y})
\end{aligned}
$$

在命令视窗中依次输入并运行如下 EViews 代码：

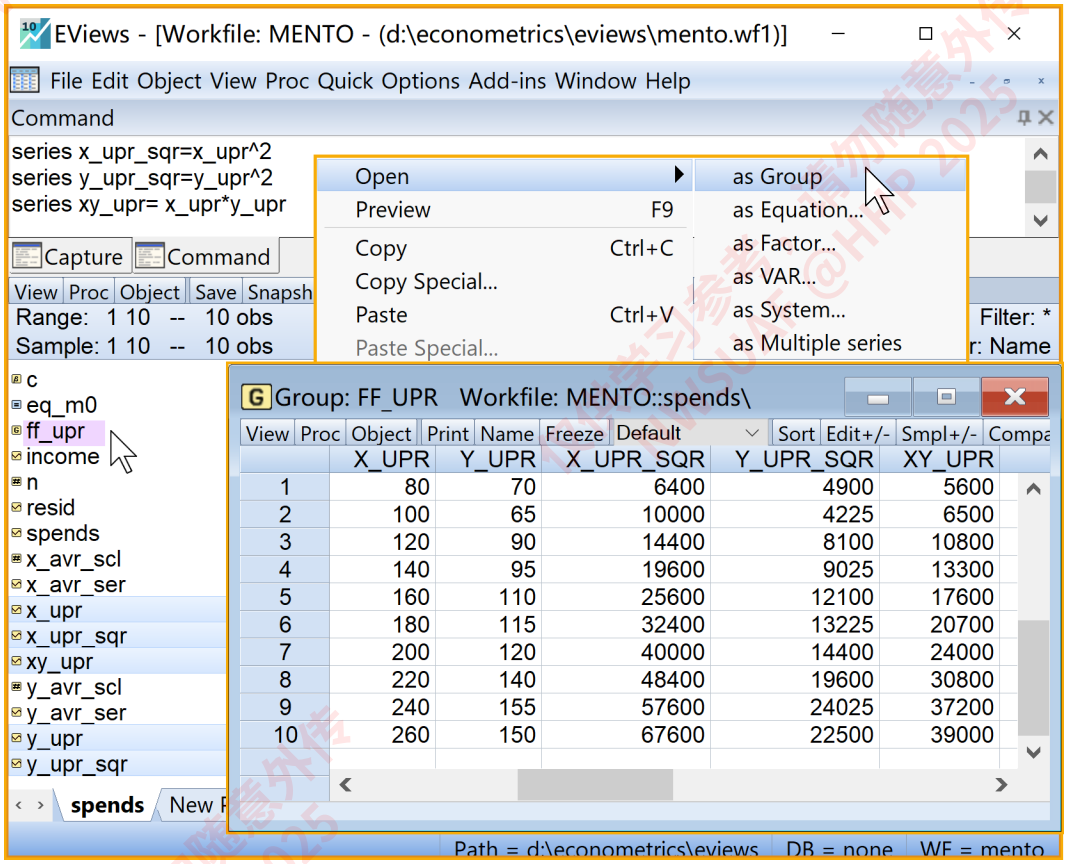


图 2.9: 统计量 FF（Favorite Five）X_i、Y_i、X_i^2、Y_i^2、X_iY_i

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
'计算ff离差series，并构建group
series x_lwr=x_upr-x_avr_ser          '离差序列x
series y_lwr= y_upr-y_avr_ser         '离差序列y
series x_lwr_sqr= x_lwr ^2            '离差序列x^2
series y_lwr_sqr =y_lwr ^2            '离差序列y^2
series xy_lwr=x_lwr*y_lwr             '离差序列xy
group ff_lwr x_lwr y_lwr x_lwr_sqr y_lwr_sqr xy_lwr '构建ff离差序列组（group）
```

在工作文件视窗下，可以看到如下新生成的对象，可以双击查看（见图 2.10 ）：

- x_i 离差的序列对象x_lwr
- y_i 离差的序列对象y_lwr
- x_i^2 离差的序列对象x_lwr_sqr
- y_i^2 离差的序列对象y_lwr_sqr
- $x_i y_i$ 离差的序列对象xy_lwr
- ff$(x_i; y_i; x_i^2; y_i^2, x_i y_i)$ 的组对象ff_lwr

2.5.6 计算回归系数的估计值:

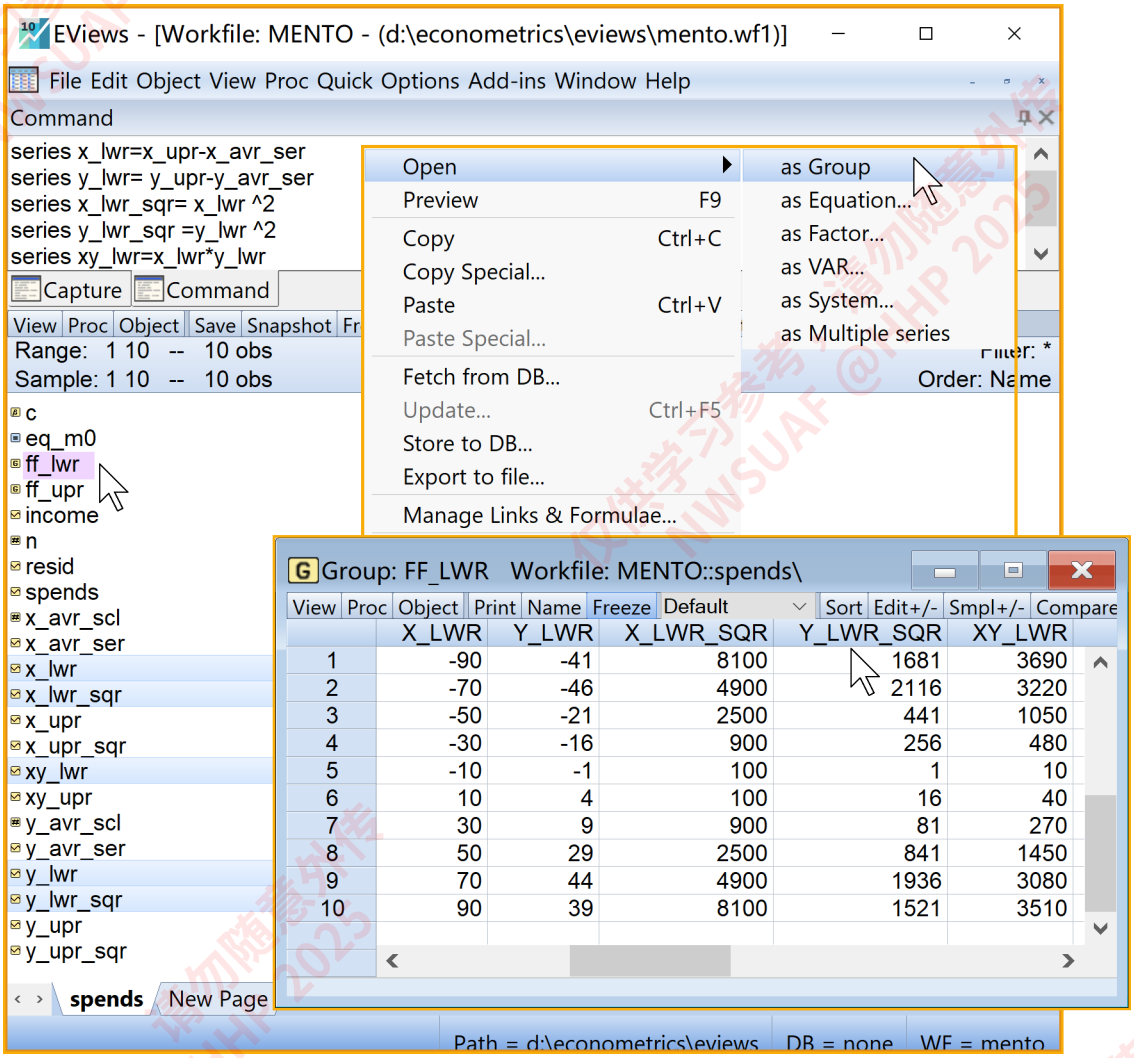
Eviews 操作目标：计算得到回归方程的回归系数（包含 2 个系数估计值）。

Eviews 操作思路：根据 FF 或 ff，利用理论公式，计算得出方程的斜率系数和截距系数。理论计算公式为：

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \quad \text{(b2\_hat)}$$
$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \quad \text{(b1\_hat)}$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 计算回归系数scalar: b2和b1
scalar b2_hat = @sum(xy_lwr)/@sum(x_lwr_sqr)  '斜率系数
scalar b1_hat =y_avr_scl-b2_hat *x_avr_scl    '截距系数
```



	X_LWR	Y_LWR	X_LWR_SQR	Y_LWR_SQR	XY_LWR
1	-90	-41	8100	1681	3690
2	-70	-46	4900	2116	3220
3	-50	-21	2500	441	1050
4	-30	-16	900	256	480
5	-10	-1	100	1	10
6	10	4	100	16	40
7	30	9	900	81	270
8	50	29	2500	841	1450
9	70	44	4900	1936	3080
10	90	39	8100	1521	3510

图 2.10: 统计量 ff（favorite five）x_i、y_i、x_i^2、y_i^2、$x_i y_i$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

在工作文件视窗下，可以看到如下新生成的标量对象，可以双击查看（见图 2.11 ）：

- 回归方程斜率系数 $\hat{\beta}_2$ 的标量对象b2_hat
- 回归方程截距系数 $\hat{\beta}_1$ 的标量对象b1_hat

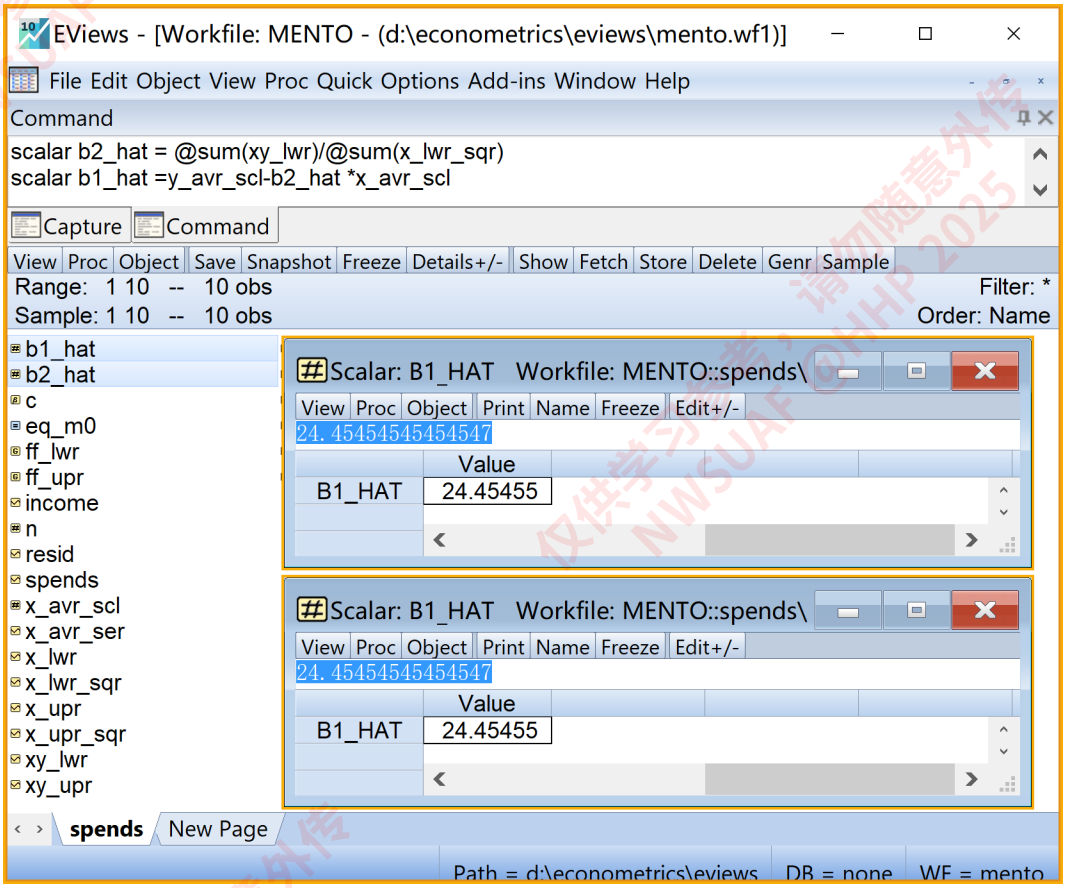


图 2.11: 回归方程的斜率系数

2.5.7 计算回归拟合值、拟合值的离差、回归残差和残差平方序列（series）：

目标：得到回归拟合值 $\hat{Y}_i$、拟合值的离差 $\hat{y}_i$、回归残差 e_i 和回归残差平方 e_i^2 等五个序列对象，用于后续分析

思路：借助回归方程和回归系数值，先计算回归拟合值及其离差，再计算残差和残差平方序列

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 * X_i$$
$$\hat{y}_i = \hat{Y}_i - \bar{Y}$$
$$e_i = Y_i - \hat{Y}_i$$
$$e_i^2 = (e_i)^2$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 计算回归预测值及其离差series
series y_upr_hat= b1_hat+ b2_hat*x_upr        '回归预测值
series y_lwr_hat= y_upr_hat-y_avr_ser        '回归预测的离差值

' 计算残差及残差平方series
series ei=y_upr-y_upr_hat                    '残差序列
series ei_sqr= ei^2                          '残差平方序列
```

在工作文件视窗下，可以看到如下新生成的序列对象，可以双击查看（见图 2.12 和图 2.13 ）：

- 回归拟合值 $\hat{Y}_i$ 的序列对象y_upr_hat
- 回归拟合值离差 $\hat{y}_i$ 的序列对象y_lwr_hat
- 回归残差 e_i 的序列对象ei
- 回归残差平方 e_i^2 的序列对象ei_sqr

2.5.8 计算回归方程的误差方差及标准差

Eviews 操作目标：计算回归方程的误差方差 $\hat{\sigma}^2$ 及标准差 $\hat{\sigma}$（标量），与主回归结果进行核验。

Eviews 操作思路：利用残差序列对象，并参考如下公式

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$$
$$\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n-2}}$$

在命令视窗中依次输入并运行如下 EViews 代码：

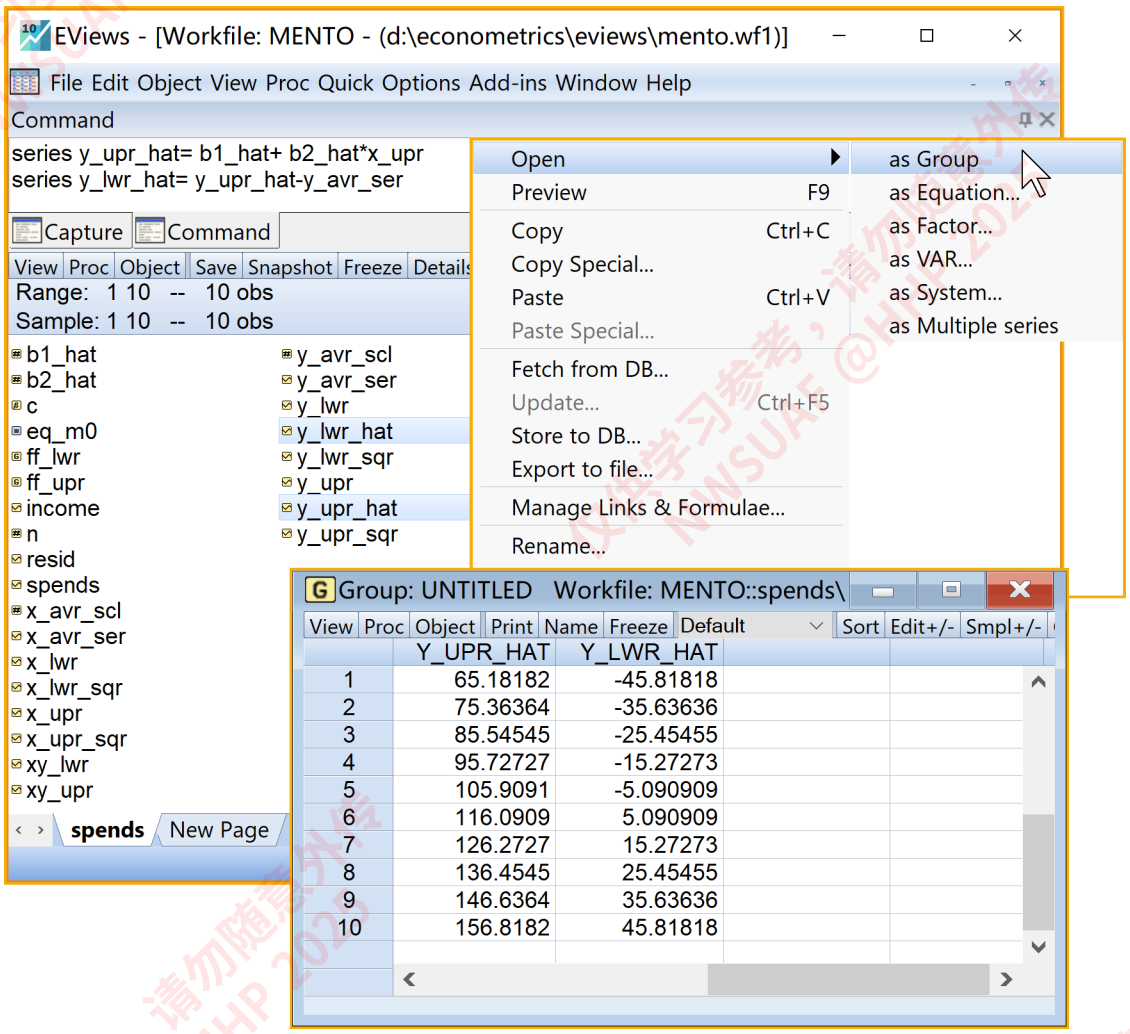


图 2.12: 回归预测值及其离差

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

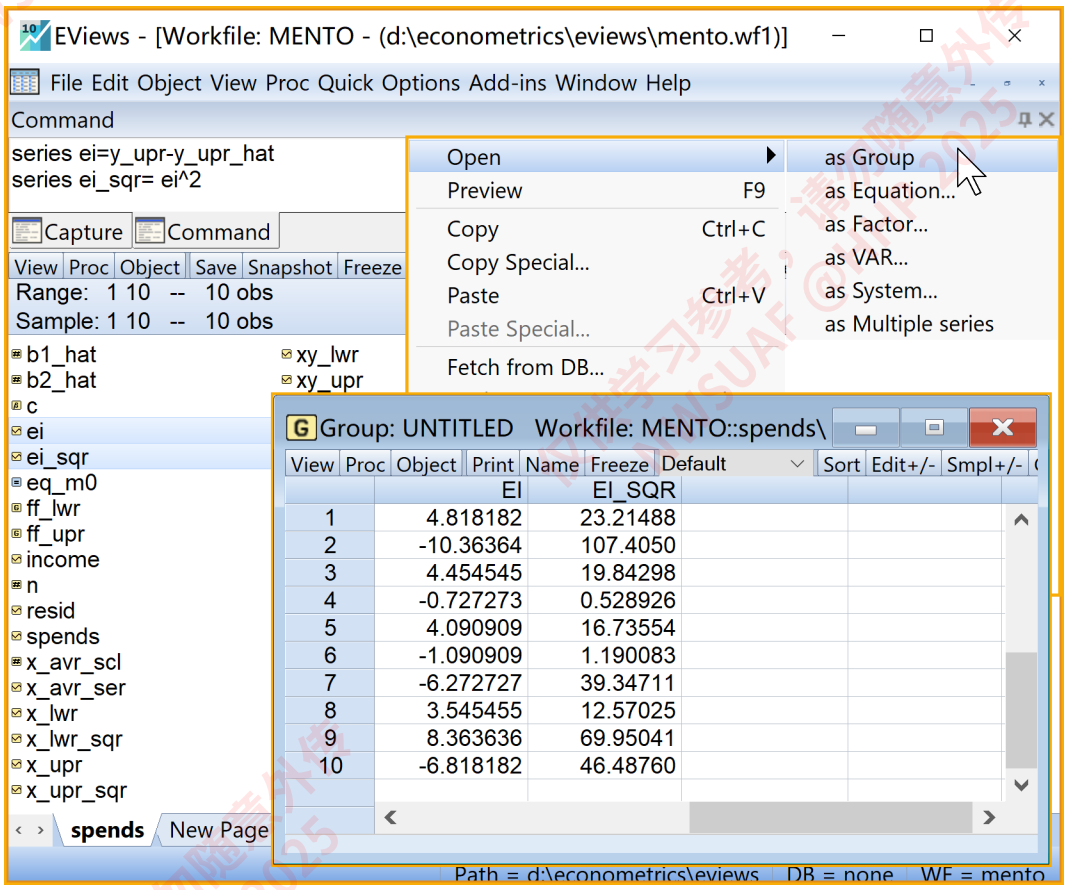


图 2.13: 残差 e_i 及残差平方 e_i^2 序列

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
' 计算回归误差方差及其标准差scalar
scalar sgm_hat_sqr=@sum(ei_sqr)/(n-2)        '方程的回归误差方差
scalar sgm_hat= sgm_hat_sqr^0.5              '方程的回归误差标准差
```

在工作文件视窗下，可以看到如下的标量对象，可以双击查看（见图 2.14 ）：

- 回归误差方差 $\hat{\sigma}^2$ 的标量对象sgm_hat_sqr
- 回归误差标准差 $\hat{\sigma}$ 的标量对象sgm_hat

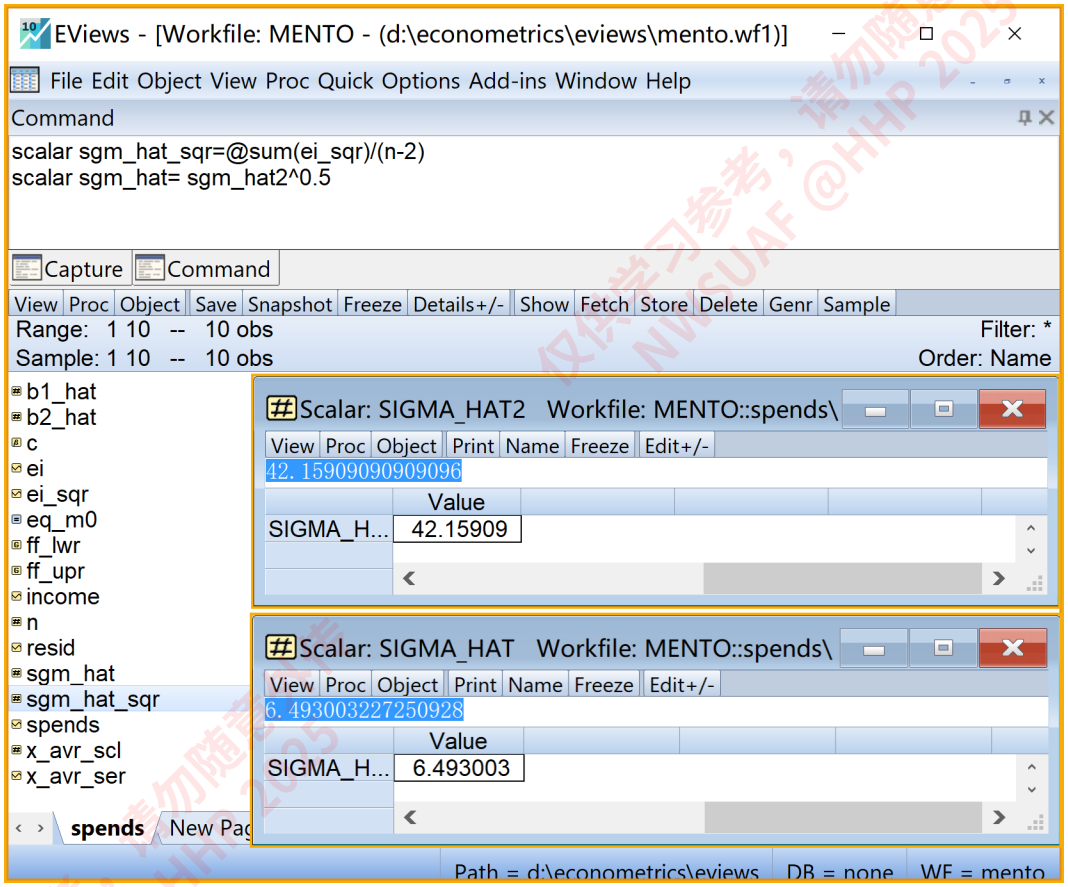


图 2.14: 回归误差方差及其标准差

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

2.5.9 计算回归系数的样本方差和标准差：

目标：分别得到斜率系数的样本方差和标准差（$S^2_{\hat{\beta}_2}$、$S^2_{\hat{\beta}_2}$），以及截距系数的样本方差和标准差（$S_{\hat{\beta}_1}$、$S_{\hat{\beta}_1}$）

思路：借助如下的计算公式：

$$
\begin{aligned}
S^2_{\hat{\beta}_2} &= \frac{\hat{\sigma}^2}{\sum x_i^2} \\
S_{\hat{\beta}_2} &= \sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}} \\
S^2_{\hat{\beta}_1} &= \frac{\hat{\sigma}^2}{\sum x_i^2} * \frac{\sum X_i^2}{n} \\
S_{\hat{\beta}_1} &= \sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2} * \frac{\sum X_i^2}{n}}
\end{aligned}
$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 计算回归系数的样本方差和标准差scalar
scalar ss_b2_hat=sgm_hat_sqr/@sum(x_lwr_sqr)          '斜率系数的样本方差
scalar ss_b1_hat=ss_b2_hat*(@sum(x_upr_sqr)/n)        '截距系数的样本方差
scalar s_b2_hat=ss_b2_hat^0.5                         '斜率系数的样本标准差
scalar s_b1_hat=ss_b1_hat^0.5                         '截距系数的样本标准差
```

在工作文件视窗下，可以看到如下 EViews 对象，可以双击查看（见图 2.15 ）：

- 斜率系数样本方差 $S^2_{\hat{\beta}_2}$ 的标量对象ss_b2_hat
- 截距系数样本方差 $S^2_{\hat{\beta}_1}$ 的标量对象ss_b1_hat
- 斜率系数样本标准差 $S_{\hat{\beta}_2}$ 的标量对象s_b2_hat
- 截距系数样本标准差 $S_{\hat{\beta}_1}$ 的标量对象s_b1_hat

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

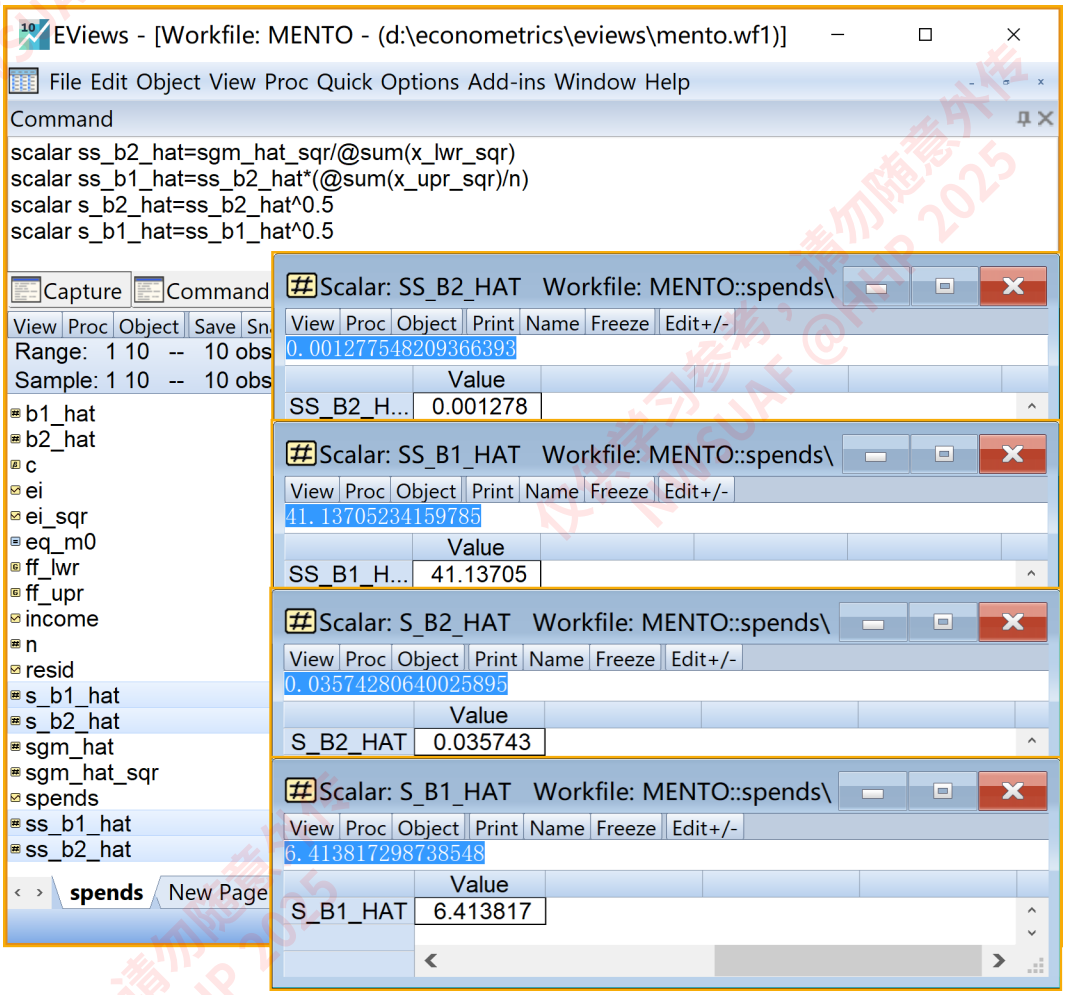


图 2.15: 回归系数的样本方差和标准差

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

2.5.10 进行平方和分解，计算 TSS、ESS 和 RSS，以及各自的自由度（标量）

Eviews 操作目标：进行平方和分解，得到方差分析表（ANOVA）。

Eviews 操作思路：掌握自由度的计算，利用理论公式

$$
\begin{aligned}
TSS &= \sum y_i^2 = \sum (Y_i - \bar{Y})^2 \\
RSS &= \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 \\
ESS &= \sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{Y})^2 \\
TSS &= RSS + ESS \\
df_{TSS} &= n - 1 \\
df_{RSS} &= n - k \\
df_{ESS} &= k - 1 \\
df_{TSS} &= df_{RSS} + df_{ESS}
\end{aligned}
$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 进行平方和分解并得到相应自由度(标量scalar)
scalar tss=@sum(y_lwr_sqr)      '总平方和TSS
scalar rss=@sum(ei_sqr)         '残差平方和RSS
scalar ess=tss-rss              '回归平方和ESS
scalar df_tss=n-1               ' TSS的自由度
scalar df_rss=n-2               ' RSS的自由度
scalar df_ess=1                 ' ESS的自由度
```

在工作文件视窗下，可以看到如下的标量对象，可以双击查看（见图 2.16 ）：

- 总平方和 TSS 的标量对象tss
- 剩余平方和 RSS 的标量对象rss
- 回归平方和 ESS 的标量对象ess
- 总平方和的自由度 df_{TSS} 的标量对象df_tss
- 剩余平方和的自由度 df_{RSS} 的标量对象df_rss
- 回归平方和的自由度 df_{ESS} 的标量对象df_ess

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

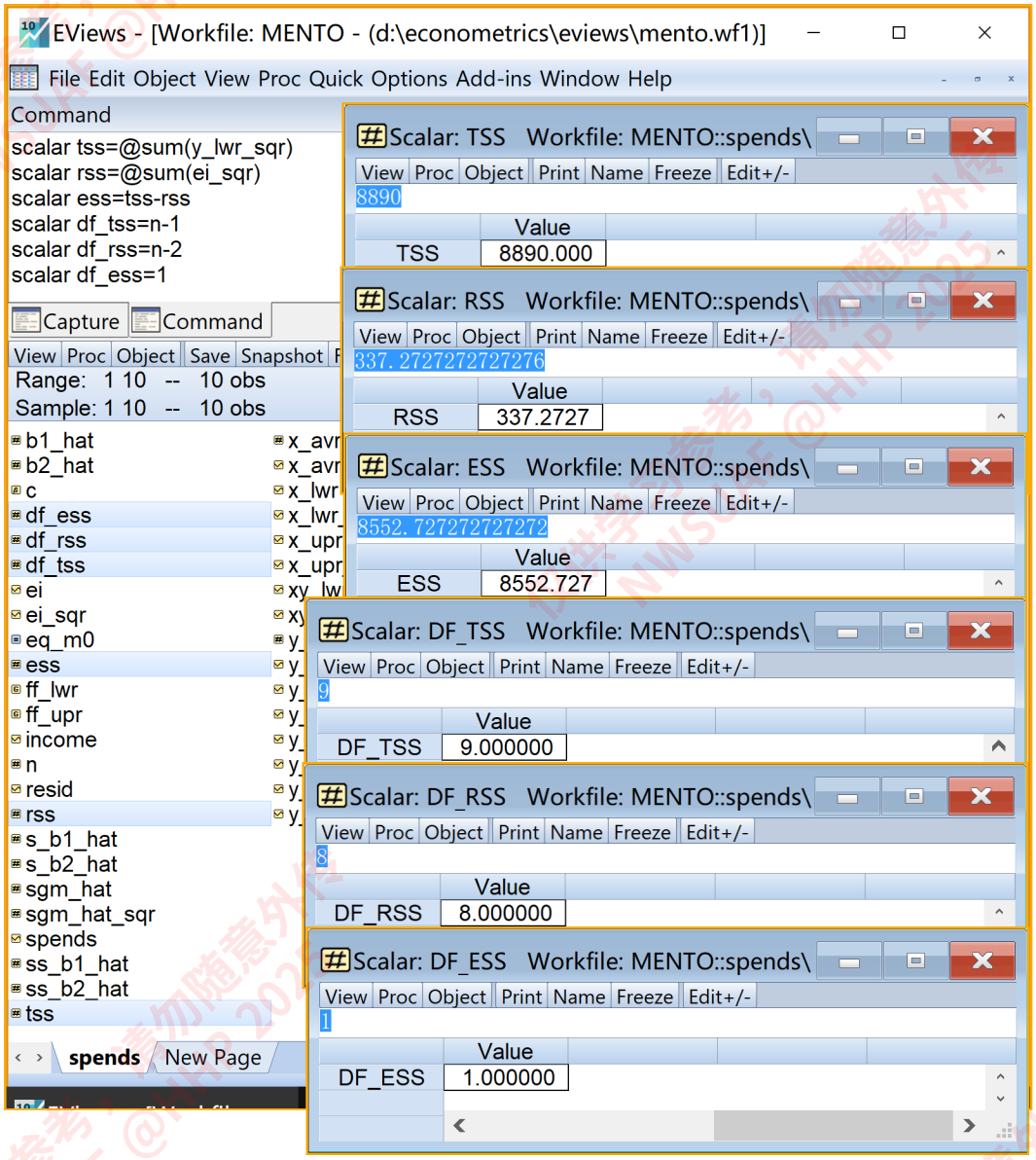


图 2.16: 平方和分解并得到相应自由度

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

2.5.11 计算相关系数 r 和判定系数 r^2(标量 scalar):

Eviews 操作目标：得到自变量 X_i 与因变量 Y_i 的相关系数，计算回归方程的判定系数。

Eviews 操作思路：借助方差分析表结果计算判定系数。利用理论公式

$$r = \frac{Cov(X,Y)}{S_X * S_Y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

$$r^2 = \frac{ESS}{TSS}$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 计算相关系数r和判定系数(标量scalar)
scalar r=@cor(x_upr,y_upr)        'X和Y变量的相关系数
scalar r_sqr=ess/tss              '回归方程的判定系数
```

在工作文件视窗下，可以看到如下的 EViews 对象，可以双击查看（见图 2.17 ）：

- 自变量 X_i 与因变量 Y_i 相关系数的标量对象r
- 回归方程判定系数 R^2 的标量对象r_sqr

2.5.12 回归系数的 t 检验

目标：检验各个回归系数是否显著

思路：根据理论的矩阵公式，计算样本 t 统计量；计算给定显著性水平下的理论 t 值；进行 t 假设检验。

回归系数的样本 t 统计量以及给定显著性水平下的理论 t 值的计算公式为：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

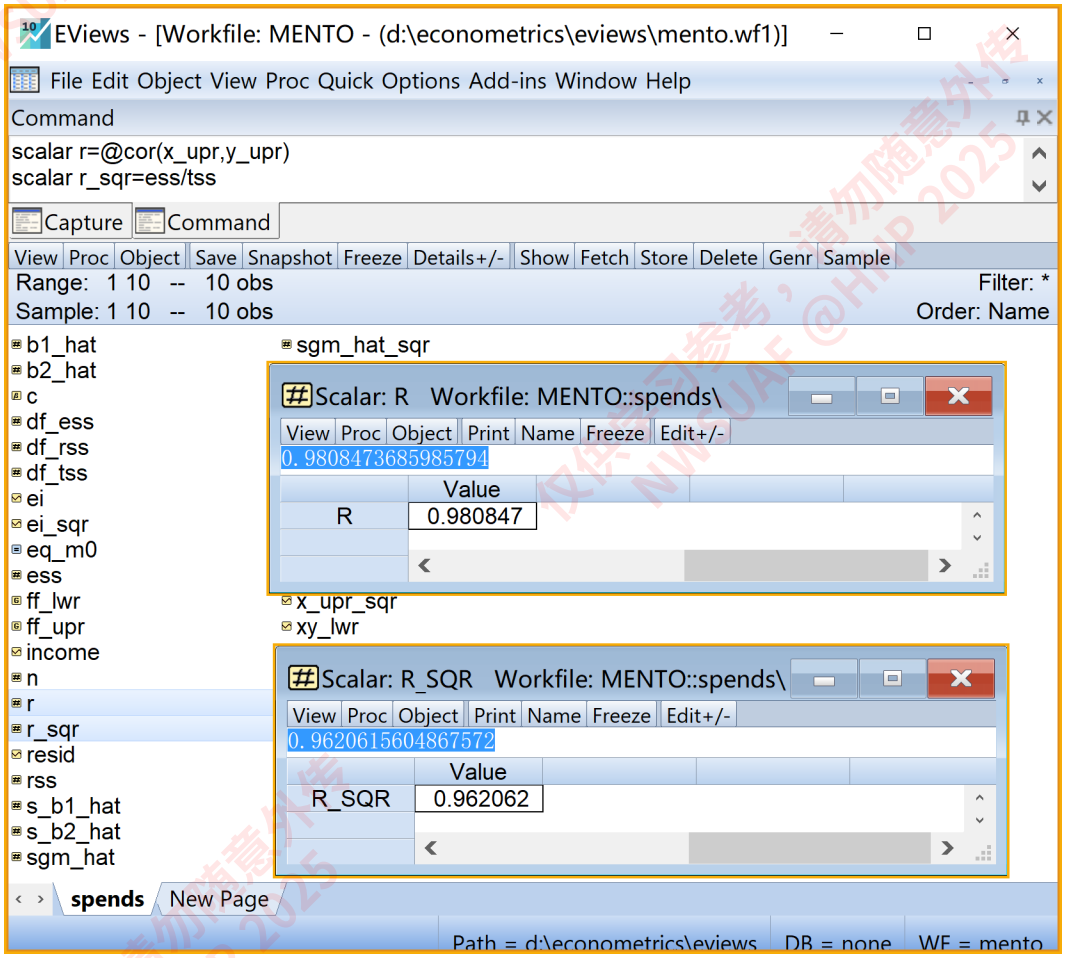


图 2.17: 相关系数和判定系数

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$t^*_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{S_{\hat{\beta}_2}}$$

$$t^*_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$$

$$t_{rht} = t_{(1-\alpha)/2}(df_{RSS}) = t_{0.975}(8)$$

$$t_{lft} = t_{\alpha/2}(df_{RSS}) = t_{0.025}(8)$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 回归系数的t检验
scalar t_str_b2_hat =b2_hat/s_b2_hat   '斜率的t统计量
scalar t_str_b1_hat =b1_hat/s_b1_hat   '截距的t统计量
scalar t_value =@qtdist(0.975,8)       'a=0.05下的t查表值（右侧正值）
scalar t_value2 =@qtdist(0.025,8)      'a=0.05下的t查表值（左侧负值）
```

在工作文件视窗下，可以看到如下的 EViews 对象，可以双击查看（见图 2.18 ）：

- 斜率系数样本 t 统计量 $t^*_{\beta_2}$ 的标量对象t_str_b2_hat
- 截距系数样本 t 统计量 $t^*_{\beta_1}$ 的标量对象t_str_b1_hat
- 给定 $\alpha = 0.05$ 水平下的理论 t 值（右侧正值）$t_{1-\alpha/2}(n-k)$ 的标量对象t_value
- 给定 $\alpha = 0.05$ 水平下的理论 t 值（左侧负值）$t_{\alpha/2}(n-k)$ 的标量对象t_value2

2.5.13 对回归方程的整体显著性进行 F 假设检验

目标：进行回归方程的整体显著性进行 F 假设检验

思路：根据理论公式，计算回归方程的样本 F 统计量；给定显著性水平下，计算得到理论 F 值；完成模型整体显著性 F 检验过程。

回归方程的样本 F 统计量，以及给定显著性水平下理论 F 值的计算公式为：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

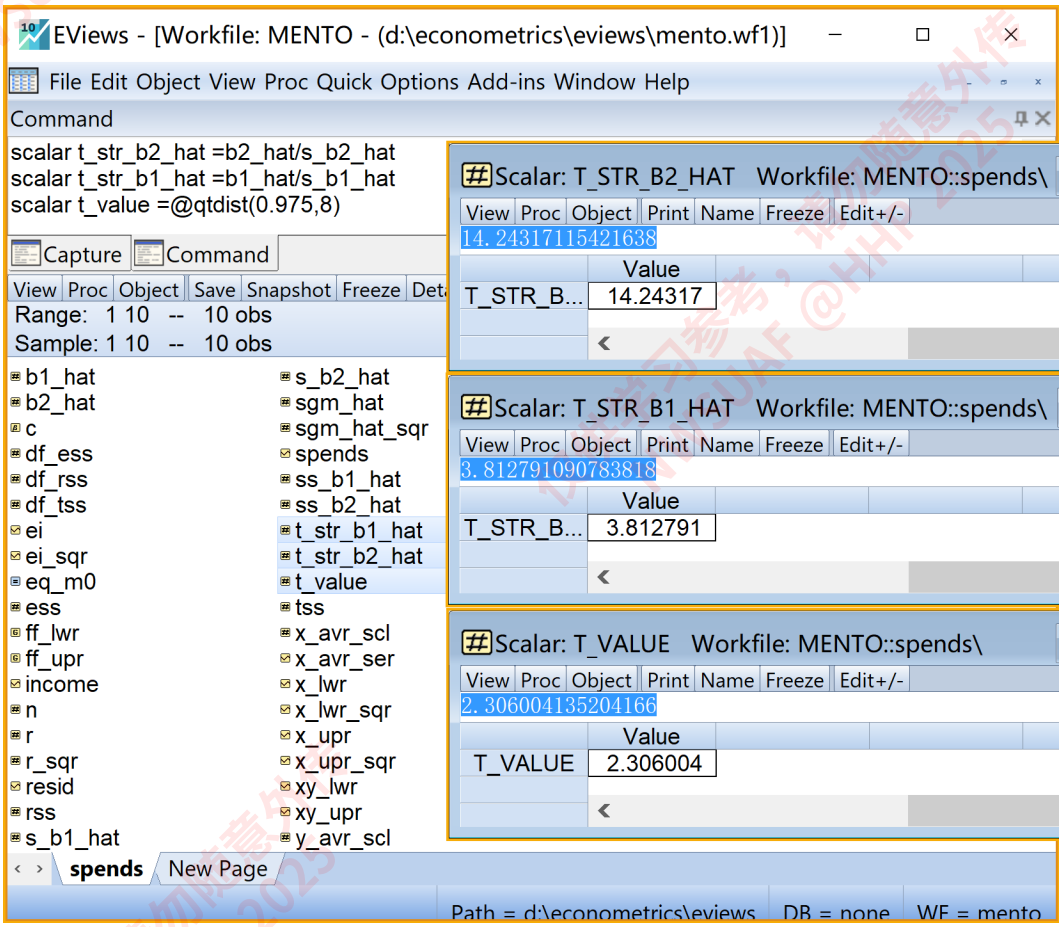


图 2.18: 回归系数的 t 检验

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$F^* = \frac{MSS_{ESS}}{MSS_{RSS}} = \frac{ESS/df_{ESS}}{RSS/df_{RSS}}$$
$$F_{rht} = F_{1-\alpha}(df_{ESS}, df_{RSS}) = F_{0.95}(1,8)$$
$$F_{lft} = F_{\alpha}(df_{ESS}, df_{RSS}) = F_{0.05}(1,8)$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 回归方程的F检验
scalar f_str =(ess/df_ess)/(rss/df_rss)    '回归方程的F统计量值
scalar f_value=@qfdist(0.95,df_ess,df_rss)        'a=0.05下的t查表值（右侧大值）
scalar f_value2=@qfdist(0.05,df_ess,df_rss)       'a=0.05下的t查表值（左侧小值）
```

在工作文件视窗下，可以看到如下的 EViews 对象，可以双击查看（见图 2.19 ）：

- 回归方程的样本 F 统计量 F^* 的标量对象f_str
- 给定 $\alpha = 0.05$ 水平下的理论 F 值（右侧大值）$F_{1-\alpha}(n-1, n-k)$ 的标量对象f_value
- 给定 $\alpha = 0.05$ 水平下的理论 F 值（左侧小值）$F_{\alpha}(n-1, n-k)$ 的标量对象f_value2

2.5.14 进行样本外的均值预测、个值预测，并计算置信区间

目标：样本外的均值预测、个值预测，计算得到置信区间（给定显著性水平）

思路：构建样本外 X_0；计算得到回归估计值 $\hat{Y}_0$；构造分别计算得到均值预测的 t 分布样本标准差 $S_{\hat{Y}_0}$ 和个值预测的 t 分布样本标准差 $S_{Y_0-\hat{Y}_0}$；给定显著性水平下，计算得到理论 t 值；利用公式分别计算均值预测和个值预测的置信区间。

相关参考计算公式为：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

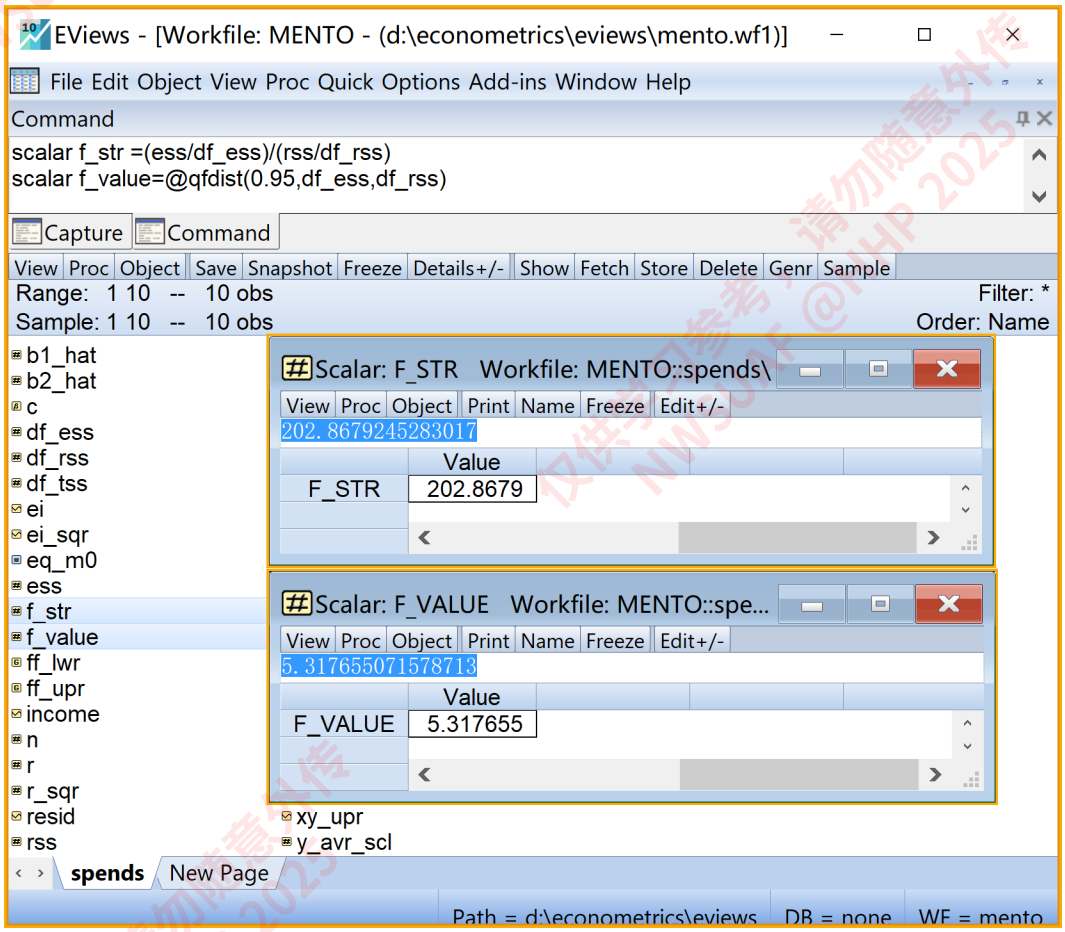


图 2.19: 回归方程的 F 检验

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$X_0 = 280$$

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 * X_0$$

$$S_{\hat{Y}_0} = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(X_0 - \bar{X})}{\sum x_i^2} \right)}$$

$$S_{Y_0 - \hat{Y}_0} = \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right)}$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{(n-k)}$$

$$E(Y|X = X_0)_{left} = \hat{Y}_0 - t_{1-\alpha/2}(df_{RSS}) * S_{\hat{Y}_0}$$

$$E(Y|X = X_0)_{right} = \hat{Y}_0 + t_{1-\alpha/2}(df_{RSS}) * S_{\hat{Y}_0}$$

$$(Y_0|X = X_0)_{left} = \hat{Y}_0 - t_{(1-\alpha/2)}(df_{RSS}) * S_{Y_0 - \hat{Y}_0}$$

$$(Y_0|X = X_0)_{right} = \hat{Y}_0 + t_{(1-\alpha/2)}(df_{RSS}) * S_{Y_0 - \hat{Y}_0}$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 计算样本外预测(标量scalar)
scalar x0 =280                                                          '样本外X值
scalar y0_hat =b1_hat+b2_hat*x0                                         '样本外估计Y值
scalar s_y0h=(sgm_hat_sqr*(1/n+(x0-x_avr_scl)^2/@sum(x_lwr_sqr)))^0.5   '均值预测的样本标准差
scalar s_mns = (sgm_hat_sqr*(1+1/n+(x0-x_avr_scl)^2/@sum(x_lwr_sqr) ))^0.5  '个值预测的样本标准差
scalar y_exp_lft= y0_hat-t_value*s_y0h      '均值预测的置信区间的左界值
scalar y_exp_rht= y0_hat+t_value*s_y0h      '均值预测的置信区间的右界值
scalar y_ind_lft= y0_hat-t_value*s_mns      '个值预测的置信区间的左界值
scalar y_ind_rht= y0_hat+t_value*s_mns      '个值预测的置信区间的右界值
```

在工作文件视窗下，可以看到如下的 EViews 对象，可以双击查看（见图 2.20 和图 2.21 ）：

- 样本外 $X_0 = 280$ 的标量对象x0
- 样本外估计 $\hat{Y}_0$ 值的标量对象y0_hat
- 均值预测的样本标准差 $S_{\hat{Y}_0}$ 的标量对象s_y0h
- 个值预测的样本标准差 $S_{Y_0 - \hat{Y}_0}$ 的标量对象s_y0h_mns_y0

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 均值预测的置信区间的左界值 $\hat{Y}_0 - t_{1-\alpha/2}(n-2) \cdot S_{\hat{Y}_0}$ 的标量对象y_exp_lft
- 均值预测的置信区间的右界值 $\hat{Y}_0 + t_{1-\alpha/2}(n-2) \cdot S_{\hat{Y}_0}$ 的标量对象y_exp_rht
- 个值预测的置信区间的左界值 $\hat{Y}_0 - t_{1-\alpha/2}(n-2) \cdot S_{Y_0-\hat{Y}_0}$ 的标量对象y_ind_lft
- 个值预测的置信区间的右界值 $\hat{Y}_0 + t_{1-\alpha/2}(n-2) \cdot S_{Y_0-\hat{Y}_0}$ 的标量对象y_ind_rht

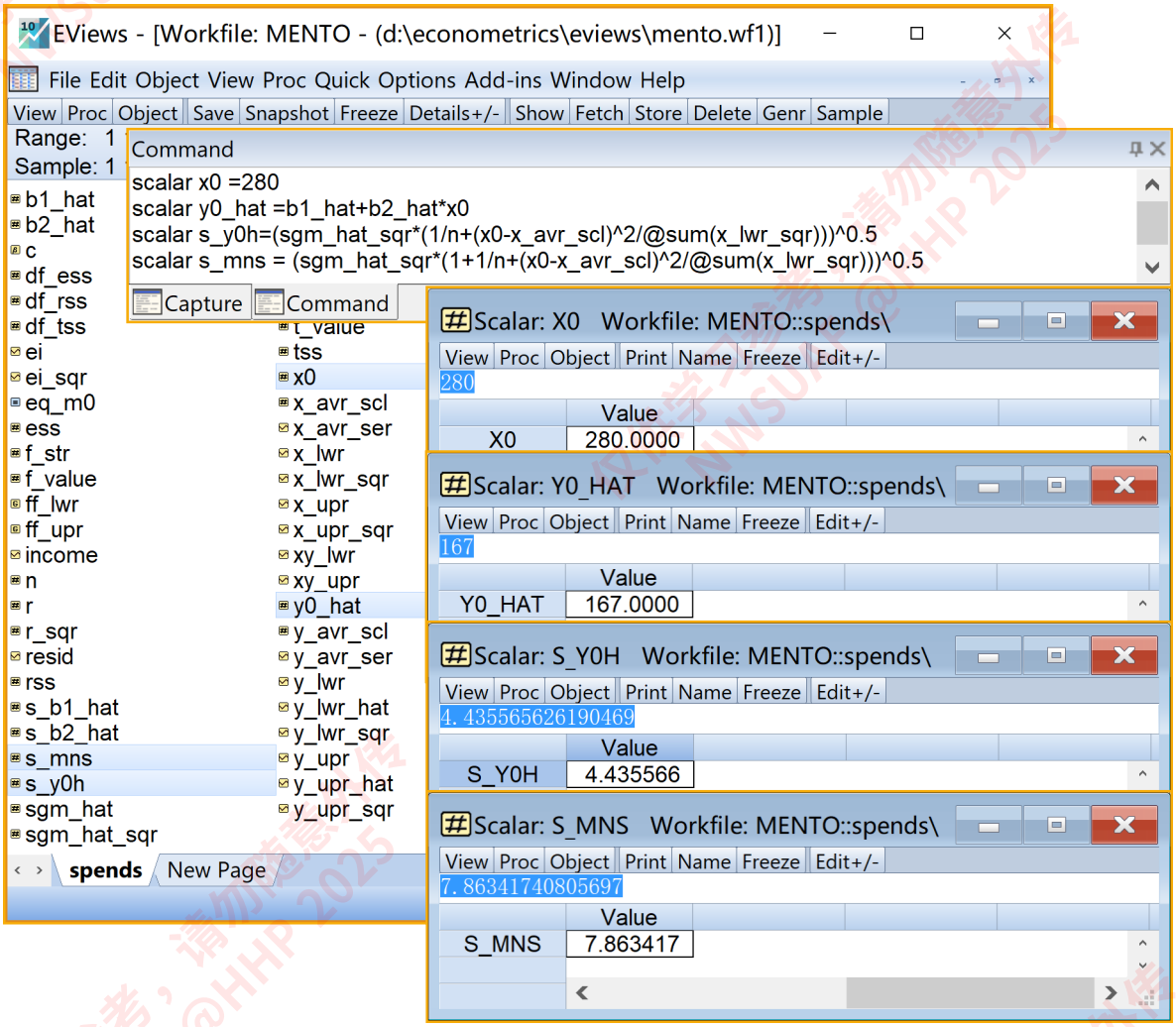


图 2.20: 均值和个值的样本外预测：准备工作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

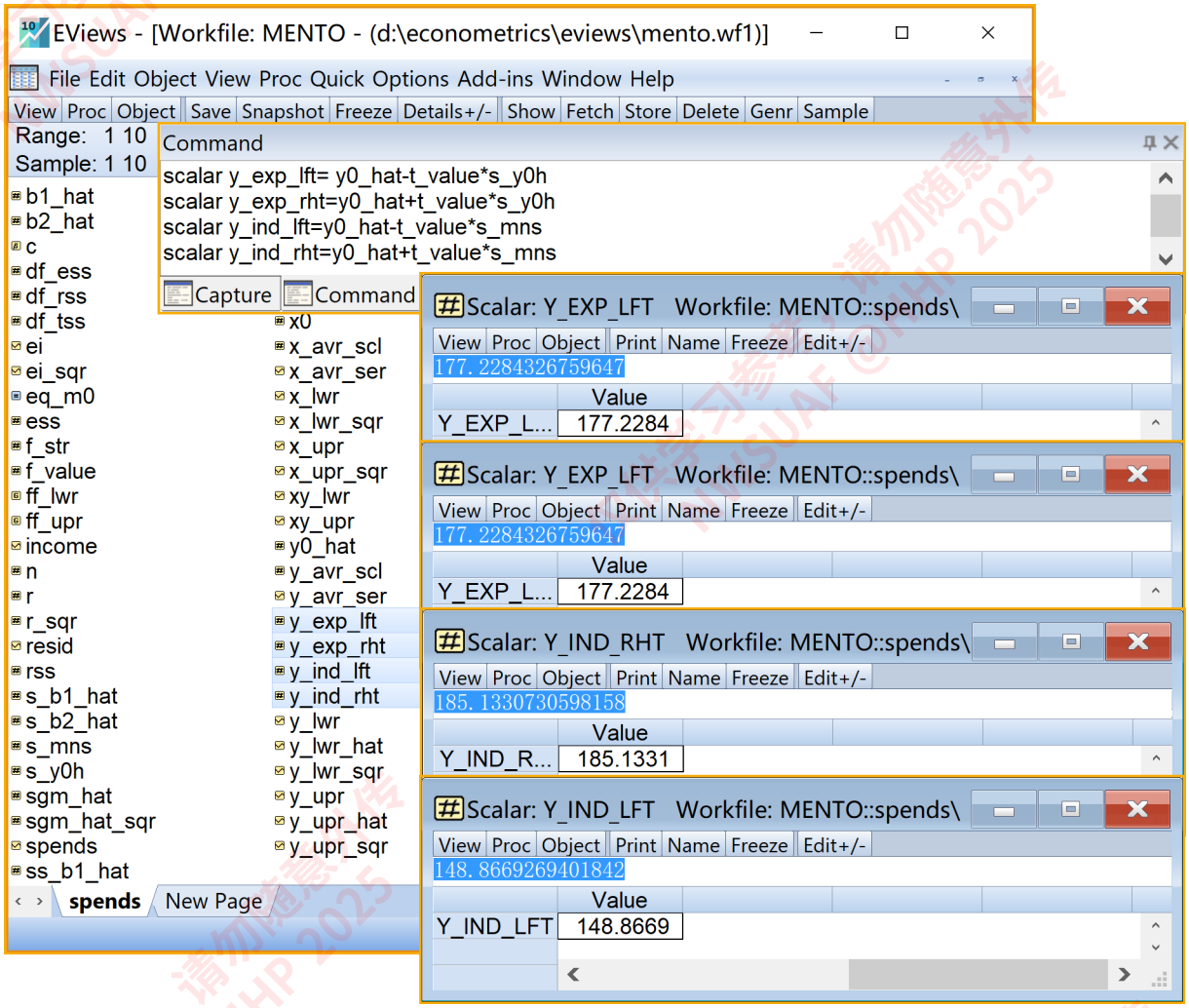


图 2.21: 均值和个值的样本外预测：区间预测

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

2.6 附录：prg 源代码

实际操作中，在 EViews 命令视窗中逐条输入代码，既容易出错，又不便于维护这些代码，还不能进行代码的重复使用（在第一章 中已经论述）。

因此，读者可以创建一个.prg 编程文件，并在其中编写 EViews 代码，进行管理、维护、运行和分析。下面代码按本章主要实验步骤编写，读者可以用于本章的 EViews 编程参考，进行实验练习。

```
'=================================================================================================
'说明：以下为EViews编程文件mento.prg的代码
'将展示第二章中“蒙特卡洛模拟数据案例”主要分析步骤的“批量式命令驱动”实现方法(::
'其中，符号'起始的行，为注释行，其他为EViews命令行。
'=================================================================================================

'创建工作文件（工作文件名=mento，子页命名=spend），无结构无日期，样本数为10
wfcreate(wf=mento,page=spend)  u 10

'导入外部数据，路径为d:\github\books\data\lab1-basic-development.xlsx
import d:\github\books\data\lab2-mento-demon.xlsx

' 命令行1：生成线性回归方程对象eq_main
equation eq_main.ls y c x

' 转换变量名
series x_upr=x                          '序列X
series y_upr=y                          '序列Y

'计算三个重要标量
scalar x_avr_scl=@mean(x_upr)          'X的均值(标量)
scalar y_avr_scl=@mean(y_upr)          'Y的均值(标量)
scalar n=@obs(x_upr)                   '样本数n(标量)

'计算均值序列
series x_avr_ser=@mean(x_upr)
series y_avr_ser=@mean(y_upr)
series xy_upr= x_upr*y_upr
```

```
'计算FF，并以group形式打开
series x_upr_sqr=x_upr^2                '序列X^2
series y_upr_sqr=y_upr^2                '序列Y^2
series xy_upr= x_upr*y_upr              '序列XY
group ff_upr x_upr y_upr x_upr_sqr y_upr_sqr xy_upr     '构建FF序列组（group）

'计算ff离差series，并构建group
series x_lwr=x_upr-x_avr_ser            '离差序列x
series y_lwr= y_upr-y_avr_ser           '离差序列y
series x_lwr_sqr= x_lwr ^2              '离差序列x^2
series y_lwr_sqr =y_lwr ^2              '离差序列y^2
series xy_lwr=x_lwr*y_lwr               '离差序列xy
group ff_lwr x_lwr y_lwr x_lwr_sqr y_lwr_sqr xy_lwr     '构建ff离差序列组（group）

' 计算回归系数scalar: b2和b1
scalar b2_hat = @sum(xy_lwr)/@sum(x_lwr_sqr)  '斜率系数
scalar b1_hat =y_avr_scl-b2_hat *x_avr_scl    '截距系数

' 计算回归预测值及其离差series
series y_upr_hat= b1_hat+ b2_hat*x_upr      '回归预测值
series y_lwr_hat= y_upr_hat-y_avr_ser       '回归预测的离差值

' 计算残差及残差平方series
series ei=y_upr-y_upr_hat                   '残差序列
series ei_sqr= ei^2                         '残差平方序列

' 计算回归误差方差及其标准差scalar
scalar sgm_hat_sqr=@sum(ei_sqr)/(n-2)       '方程的回归误差方差
scalar sgm_hat= sgm_hat_sqr^0.5             '方程的回归误差标准差

' 计算回归系数的样本方差和标准差scalar
scalar ss_b2_hat=sgm_hat_sqr/@sum(x_lwr_sqr)        '斜率系数的样本方差
scalar ss_b1_hat=ss_b2_hat*(@sum(x_upr_sqr)/n)      '截距系数的样本方差
scalar s_b2_hat=ss_b2_hat^0.5                       '斜率系数的样本标准差
```

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
scalar s_b1_hat=ss_b1_hat^0.5                          '截距系数的样本标准差

' 进行平方和分解并得到相应自由度(标量scalar)
scalar tss=@sum(y_lwr_sqr)      '总平方和TSS
scalar rss=@sum(ei_sqr)         '残差平方和RSS
scalar ess=tss-rss              '回归平方和ESS
scalar df_tss=n-1                              ' TSS的自由度
scalar df_rss=n-2                              ' RSS的自由度
scalar df_ess=1                 'ESS的自由度

' 计算相关系数r和判定系数(标量scalar)
scalar r=@cor(x_upr,y_upr)       'X和Y变量的相关系数
scalar r_sqr=ess/tss             '回归方程的判定系数

' 回归系数的t检验
scalar t_str_b2_hat =b2_hat/s_b2_hat  '斜率的t统计量
scalar t_str_b1_hat =b1_hat/s_b1_hat  '截距的t统计量
scalar t_value =@qtdist(0.975,8)      'a=0.05下的t查表值（右侧正值）
scalar t_value2 =@qtdist(0.025,8)     'a=0.05下的t查表值（左侧负值）

' 回归方程的F检验
scalar f_str =(ess/df_ess)/(rss/df_rss)   '回归方程的F统计量值
scalar f_value=@qfdist(0.95,df_ess,df_rss)       'a=0.05下的t查表值（右侧大值）
scalar f_value2=@qfdist(0.05,df_ess,df_rss)      'a=0.05下的t查表值（左侧小值）

' 计算样本外预测(标量scalar)
scalar x0 =280                                                          '样本外X值
scalar y0_hat =b1_hat+b2_hat*x0                                         '样本外估计Y值
scalar s_y0h=(sgm_hat_sqr*(1/n+(x0-x_avr_scl)^2/@sum(x_lwr_sqr)))^0.5   '均值预测的样本标
scalar s_mns = (sgm_hat_sqr*(1+1/n+(x0-x_avr_scl)^2/@sum(x_lwr_sqr) ))^0.5    '个值预测的样本
scalar y_exp_lft= y0_hat-t_value*s_y0h                                  '均值预测的置信区
scalar y_exp_rht= y0_hat+t_value*s_y0h                                  '均值预测的置信区
scalar y_ind_lft= y0_hat-t_value*s_mns                                  '个值预测的置信区
scalar y_ind_rht= y0_hat+t_value*s_mns                                  '个值预测的置信区
```

' ==========

2.7 作业题

蒙特卡洛模拟数据：表 2.9 给出给出了 7 个家庭在收入和支出方面的数据。

表 2.9: 家庭收入和家庭支出 (n=10)

obs	income	spends
1	80	55
2	100	88
3	120	90
4	140	80
5	160	118
6	180	120
7	200	145
8	220	135
9	240	145
10	260	175

变量说明见表 2.10 ：

表 2.10: 变量定义及说明

variable	label
obs	样本编号
spends	家庭收入
income	家庭支出

如果消费和收入之间可以构造如下一元线性回归模型：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \quad \text{SRM}$$
$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad \text{SRF}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

请在 Eviews 中对模型进行计算（手动计算），要求回答如下问题：

运用 eviews 菜单（Quick ⇒ Estimate Equation），对上述模型进行回归分析，分别列出主要的回归报告，并比较与你在 a 题中手动计算的结果是否一致。

利用 eviews 将（X，Y）数据绘制散点图（scatter），并对残差序列绘制线图 (line plot)。

1. 将数据导入 Eviews，并计算出 FF 序列 $\bar{X}$、$\bar{Y}$、X^2、Y^2、XY，以及 ff 序列 x、y、x^2、y^2、xy，并以 Group 形式保存并截图过来。

2. 根据所给数据，计算直线回归方程的回归系数 $\hat{\beta}_1$、$\hat{\beta}_2$（写出公式、主要计算过程和结果）。

3. 分别写出样本回归方程（SRF）和样本回归模型（SRM）（根据计算结果写出实际方程和模型）。

4. 计算回归误差方差（$\hat{\sigma}^2$）其标准差 (以及 $\hat{\sigma}$)。（写出公式、主要计算过程和结果）。

5. 计算得到回归截距系数的样本方差 ($S^2_{\hat{\beta}_1}$) 及其样本标准差 ($S_{\hat{\beta}_1}$)；计算回归斜率系数的样本方差 ($S^2_{\hat{\beta}_2}$) 及其样本标准差 ($S_{\hat{\beta}_2}$)。（写出公式及主要计算结果）。

6. 计算得到两个模型的平方和分解式（TSS、ESS、RSS）及其相对应的自由度（df_{TSS}，df_{ESS}，df_{RSS}），并写出方差分析表（**ANOV** table）。

7. 计算得到两个变量的样本相关系数 r，以及回归方程的判定系数 r^2，并比较相关系数与判定系数的关系（数量关系）。（写出公式及主要计算过程）。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

8. 计算两个回归系数的样本 t 值 ($t^*_{\hat{\beta}_2}$ 和 $t^*_{\hat{\beta}_1}$)。

9. 用 eviews 命令计算得到理论 t 值（$t_{1-\alpha/2}(n-2)$，给定 $\alpha = 0.05$ ），并对回归系数显著性进行假设检验。

10. 计算回归方程整体显著性检验的的样本 F 值（F_*)。

11. 用 eviews 命令计算得到理论 F 值（$F_{1-\alpha}(df_1, df_2) = F_{1-\alpha}(df_{ESS}, df_{RSS})$，给定 $\alpha = 0.05$ ），并对模型整体显著性进行假设检验。

12. 给定样本外的 $X_0 = 280$ ，计算点预测值 $\hat{Y}_0$ 。

13. 计算均值预测下的样本标准差 $S_{(\hat{Y}_0)}$。

14. 计算均值预测的置信区间 $\hat{Y}_0 - t_{1-\alpha/2}(n-2) \cdot S_{\hat{Y}_0} \leq E(Y \mid X = X_0) \leq \hat{Y}_0 + t_{1-\alpha/2}(n-2) \cdot S_{\hat{Y}_0}$。

a. 均值预测 $E(Y|X = X_0) =$

b. 均值区间预测的左界 $E(Y|X = X_0)_L =$

c. 均值区间预测的右界 $E(Y|X = X_0)_L =$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

15. 计算个值预测下的样本标准差 $S_{(Y_0-\hat{Y}_0)}$。

16. 计算个值预测的置信区间 $\hat{Y}_0 - t_{\alpha/2}(n-2) \cdot S_{Y_0-\hat{Y}_0} \leq E(Y \mid X = X_0) \leq \hat{Y}_0 + t_{\alpha/2}(n-2) \cdot S_{Y_0-\hat{Y}_0}$。

a. 个值预测 $(Y_0|X = X_0)$

b. 个值区间预测的左界 $(Y_0|X = X_0)_L$

c. 个值区间预测的右界 $(Y_0|X = X_0)_R$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

3 模型函数形式

仅供学习参考，请勿随意外传
NWSUAF @HHP 2025

3.1 实验目的及要求

- **目的**：掌握几种模型函数形式的特征，理解函数形式选择的原理。
- **要求**：在老师指导下能用 Eviews 软件进行各种形式模型的变换与处理，包括普通线性模型、过原点模型、标准化处理模型、双对数模型、半对数模型、倒数模型，得到正确的分析结果；能运用合适的计算公式，得到正确的斜率和弹性计算值。

3.2 实验原理

- 无论是一元线性回归还是多元线性回归，模型正确设置的一个重要前提就是能够正确地选择合适的函数形式，或者需要对变量进行变换处理（标准化、取对数等）。同样的实证案例，使用普通线性模型、过原点模型、标准化处理模型、双对数模型、半对数模型、倒数模型需要视案例具体情形而定，还需要结合经济学理论，通过多次尝试方才能够找到相对满意的具体模型形式。
- 此外，斜率和弹性，都是具有数学和经济学含义的重要概念，不同函数形式下实现对二者的正确计算，是得出有价值的模型估计结论的重要途径。

3.2.1 模型设置为有截距还是无截距?

定义 3.1 (过原点回归模型). 也称为无截距模型，是指没有截距项的一类线性回归模型。一般地，过原点总体回归模型（PRM）和样本回归模型（SRM）分别记为：

$$Y_i = \beta_2 X_i + u_i \tag{3.1}$$

$$Y_i = \hat{\beta}_2 X_i + e_i \tag{3.2}$$

实际模型设置中，究竟要采用经典的有截距模型，还是采用无截距模型，需要注意的是：

- 尽管模型含有截距项，但若该项的出现是统计上不显著的（即统计上等于零），则从任何实际方面考虑，都可认为这个结果是一个过原点回归模型。
- 如果在模型中确实有截距，而我们却执意拟合一个过原点回归，我们就犯了设定错误（specification error）。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 对于有截距项的模型来说，总有 $\sum e_i = 0$，但对于无截距项的模型，$\sum e_i = 0$ 则不一定成立，只有 $\sum e_i X_i = 0$。
- 对于有截距项的模型来说，总有判定系数 $r^2 \geq 0$；但是，对无截距模型来说 r^2 有时可能出现负值。

3.2.2 变量单位变换下怎么选择模型?

定义 3.2 (尺度因子). 也称为单位变换因子，是指对一个变量的尺度或单位进行变换的权数，一般用 $\omega(\omega \in \mathcal{R}, \omega \neq 0)$ 表示。一般地，总体回归模型（PRM）和样本回归模型（SRM）分别记为：

在同一个变量，可能有不同计量单位取值时，我们可以得到如下备选模型：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \tag{3.3}$$

$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^* + e_i^* \tag{3.4}$$

其中：$Y_i^* = \omega_1 Y_i$；$X_i^* = \omega_2 X_i$

此时，采用最小二乘法进行参数估计，容易证明两个模型的关系如下：

$$\hat{\beta}_2^* = \frac{\omega_1}{\omega_2}\hat{\beta}_2 \tag{3.5}$$

$$\hat{\beta}_1^* = \omega_1 \hat{\beta}_1 \tag{3.6}$$

$$\hat{\sigma}^{*^2} = \omega_1^2 \hat{\sigma}^2 \tag{3.7}$$

$$var(\beta_2^*) = \left(\frac{\omega_1}{\omega_2}\right)^2 var(\hat{\beta}_2) \tag{3.8}$$

$$var(\beta_2^*) = (\omega_1)^2 var(\hat{\beta}_1) \tag{3.9}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$r^2_{(X,Y)} = r^2_{(X^*,Y^*)} \tag{3.10}$$

模型对比，得出如下主要结论：

- 若 $\omega_1 = \omega_2$，即两个变量的尺度因子相等时：则斜率系数及其标准误不受尺度因子的影响，但截距及其标准误却放大或缩小至 ω_1 倍
- 若变量 X 的尺度不变（$\omega_2 = 1$），而变量 Y 的尺度因子（ω_1）变化：则斜率和截距系数以及它们各自的标准误都要乘以同样的因子 ω_1
- 若变量 Y 的尺度因子不变（$\omega_1 = 1$），而变量 X 的尺度（ω_2）变化：则斜率系数及其标准误都要乘以 $\frac{1}{\omega_2}$，而截距系数及其标准误不变

3.2.3 如何选择模型的函数形式?

重点讨论以下的三种回归模型：

- 双对数模型
- 半对数模型
- 倒数模型

定义 3.3 (双对数模型)**.** 是指自变量和因变量都以对数形式出现的一类线性回归模型（double-log model）。一般地，总体回归模型（PRM）和样本回归模型（SRM）分别记为：

$$ln(Y_i) = \beta_1 + \beta_2 ln(X_i) + u_i \tag{3.11}$$

$$ln(Y_i) = \hat{\beta}_1 + \hat{\beta}_2 ln(X_i) + e_i \tag{3.12}$$

定义 3.4 (半对数模型)**.** 是指自变量和因变量的其中之一以对数形式出现的一类线性回归模型（semi-log model）。一般地，可以分为：

（1）线性到对数模型（linear-log model）：只有因变量 Y 以对数形式出现。其总体回归模型（PRM）和样本回归模型（SRM）分别记为：

$$ln(Y_i) = \beta_1 + \beta_2 X_i + u_i \tag{3.13}$$

$$ln(Y_i) = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \tag{3.14}$$

（2）对数到线性模型（linear-log model）：只有自变量 X 以对数形式出现。其总体回归模型（PRM）和样本回归模型（SRM）分别记为：

$$Y_i = \beta_1 + \beta_2 ln(X_i) + u_i \tag{3.15}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 ln(X_i) + e_i \tag{3.16}$$

定义 3.5 (倒数模型)**.** 是指自变量以倒数形式出现的一类线性回归模型（reciprocal model）。一般地，总体回归模型（PRM）和样本回归模型（SRM）分别记为：

$$Y_i = \beta_1 + \beta_2 \frac{1}{X_i} + u_i \tag{3.17}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 \frac{1}{X_i} + e_i \tag{3.18}$$

对于上述三类模型形式的比较和选择，需要一些技巧和经验：

- 模型背后的理论（如菲利普斯曲线）可能给出了一个特定的函数形式。
- 最好能求出回归子相对回归元的变化率（即斜率 slope）和回归子对回归元的弹性（elasticity）。
- 所选模型的系数应该满足一定的先验预期。
- 有时多个模型都能相当不错地拟合一个给定的数据集。
- 通常不应该过分强调拟合优度 r^2 这个指标
- 在有些情形中，确定一个特定的函数形式不是那么容易，此时我们或许可以使用所谓的博克斯-考克斯变换 (Box-Cox transformations)

3.3 实验内容

- 进行不同模型形式的 Eviews 操作，包括过原点模型、尺度变换模型、双对数模型、半对数模型、倒数模型。
- 学会计算不同模型形式下 Y 对 X 的斜率和弹性公式。

3.4 实验准备

3.4.1 实验软件

本次实验需要提前准备好如下软件：

- 统计分析软件 Eviews 9.0 版本及以上
- 公式编辑软件 Mathtype 6.0 版本及以上
- 写作编辑软件 Office Word/Excel 2010 版本及以上
- 浏览器软件 chrome 66.0 版本及以上或 360 极速浏览器 9.5 版本及以上

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

3.4.2 实验材料

英国家庭食物支出：表 3.1 给出了 1980-1982 年间英国家庭支出调查中 1 519 个家庭在 Y 家庭总支出 (10 英镑)，X 食物支出 (10 英镑) 等方面的数据。数据只包括住在伦敦市区和市郊有 1~2 个子女的家庭，样本不包括自我雇佣和退休家庭。

表 3.1: 英国家庭食物支出 (n=1519)

obs	Y	X
1	50	21
2	90	34
3	180	35
4	80	36
5	90	30
1515	90	37
1516	70	20
1517	100	33
1518	130	79
1519	140	26

变量说明见表 3.2：

表 3.2: 变量定义及说明

variable	label
obs	家庭编号 ID
Y	家庭总支出 (10 英镑)
X	食物支出 (10 英镑)

3.4.3 实验规则

（1）图形对象（Graph）的命名规则：

- X 对 Y 的散点图（scatter）命名为scatter_classic

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- $log(X)$ 对 $log(Y)$ 的散点图（scatter）命名为scatter_loglog
- $1/X$ 对 Y 的散点图（scatter）命名为scatter_reclog

（2）方程对象（Equation）的命名规则：

- 线性模型 M1 的回归方程命名为m1
- 过原点模型 M2 的回归方程命名为m2
- 双对数模型 M3 的回归方程命名为m3
- 线性到对数模型 M4 的回归方程命名为m4
- 对数到线性模型 M5 的回归方程命名为m5
- 倒数模型 M6 的回归方程命名为m6
- 对数倒数模型 M7 的回归方程命名为m7

（3）向量对象（Vector）的命名规则。7 个模型中计算得到的回归系数 $(\hat{\beta}_1, \hat{\beta}_2)$ 以向量对象（Vector）形式保存。

- 线性模型 M1 的回归方程的回归系数coef1
- 过原点模型 M2 的回归方程的回归系数coef2
- 双对数模型 M3 的回归方程的回归系数coef3
- 线性到对数模型 M4 的回归方程的回归系数coef4
- 对数到线性模型 M5 的回归方程的回归系数coef5
- 倒数模型 M6 的回归方程的回归系数coef6
- 对数倒数模型 M7 的回归方程的回归系数coef7

（4）标量对象（Scalar）的命名规则。

（a）7 个模型中计算得到的斜率 $\frac{dY}{dX}$ 以标量标量对象（Scalar）形式保存。

- 线性模型 M1 的回归方程中，Y 对 X 的斜率slp1
- 过原点模型 M2 的回归方程中，Y 对 X 的斜率slp2

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 双对数模型 M3 的回归方程中，Y 对 X 的斜率slp3
- 线性到对数模型 M4 的回归方程中，Y 对 X 的斜率slp4
- 对数到线性模型 M5 的回归方程中，Y 对 X 的斜率slp5
- 倒数模型 M6 的回归方程中，Y 对 X 的斜率slp6
- 对数倒数模型 M7 的回归方程中，Y 对 X 的斜率slp7

（b）7 个模型中计算得到的点弹性 $\frac{dY}{dX} \cdot \frac{X_i}{Y_i}$ 以标量对象（Scalar）形式保存。

- 线性模型 M1 的回归方程中，Y 对 X 的点弹性els_point1
- 过原点模型 M2 的回归方程中，Y 对 X 的点弹性els_point2
- 双对数模型 M3 的回归方程中，Y 对 X 的点弹性els_point3
- 线性到对数模型 M4 的回归方程中，Y 对 X 的点弹性els_point4
- 对数到线性模型 M5 的回归方程中，Y 对 X 的点弹性els_point5
- 倒数模型 M6 的回归方程中，Y 对 X 的点弹性els_point6
- 对数倒数模型 M7 的回归方程中，Y 对 X 的点弹性els_point7

（c）7 个模型中计算得到的平均弹性 $\frac{dY}{dX} \cdot \frac{\bar{X}}{\bar{Y}}$ 以标量对象（Scalar）形式保存。

- 线性模型 M1 的回归方程中，Y 对 X 的平均弹性els_avr1
- 过原点模型 M2 的回归方程中，Y 对 X 的平均弹性els_avr2
- 双对数模型 M3 的回归方程中，Y 对 X 的平均弹性els_avr3
- 线性到对数模型 M4 的回归方程中，Y 对 X 的平均弹性els_avr4
- 对数到线性模型 M5 的回归方程中，Y 对 X 的平均弹性els_avr5
- 倒数模型 M6 的回归方程中，Y 对 X 的平均弹性els_avr6
- 对数倒数模型 M7 的回归方程中，Y 对 X 的平均弹性els_avr7

（5）表格对象（Table）的命名规则：

- 全部结算结果以表格形式保存，并命名为 tab_show

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

3.5 主要实验步骤

3.5.1 新建工作文件并导入数据

- Eviews 操作目标：构建工作文件，成功导入数据
- Eviews 操作思路：利用 EViews 代码创建工作文件并导入数据。

在命令视窗中依次输入并运行如下 EViews 代码：

```
'创建工作文件（工作文件名=food，子页命名=expenditure），无结构无日期，样本数为1519
wfcreate(wf=food,page=expenditure)  u 1519

'导入外部数据，路径为d:\github\books\data\Lab3-family-spends.xlsx
import d:\github\books\data\Lab3-family-spends.xlsx
```

在工作文件视窗下，可以看到创建的工作文件和导入的数据，可以双击查看（见图 3.1 ）：

> 提示
>
> 上述过程也可以通过菜单操作实现：
>
> - 创建工作文件：EViews 主窗口上依次点击 ⇒ `File`⇒ `New`⇒ `Workfile`
> - 导入外部数据：EViews 主窗口上依次点击 ⇒ `File` ⇒ `Import` ⇒ `Import from file` ...

3.5.2 绘制三个散点图（Y 表示食物支出，X 表示家庭总支出）

- Eviews 操作目标：分析散点图形态，初步了解 Y 与 X 的关系
- Eviews 操作思路：初步尝试绘制三种散点图，分别是：X 与 Y 的散点图；log(X) 与 log(Y) 的散点图；1/X 与 log(Y) 的散点图。

在命令视窗中依次输入并运行如下 EViews 代码：

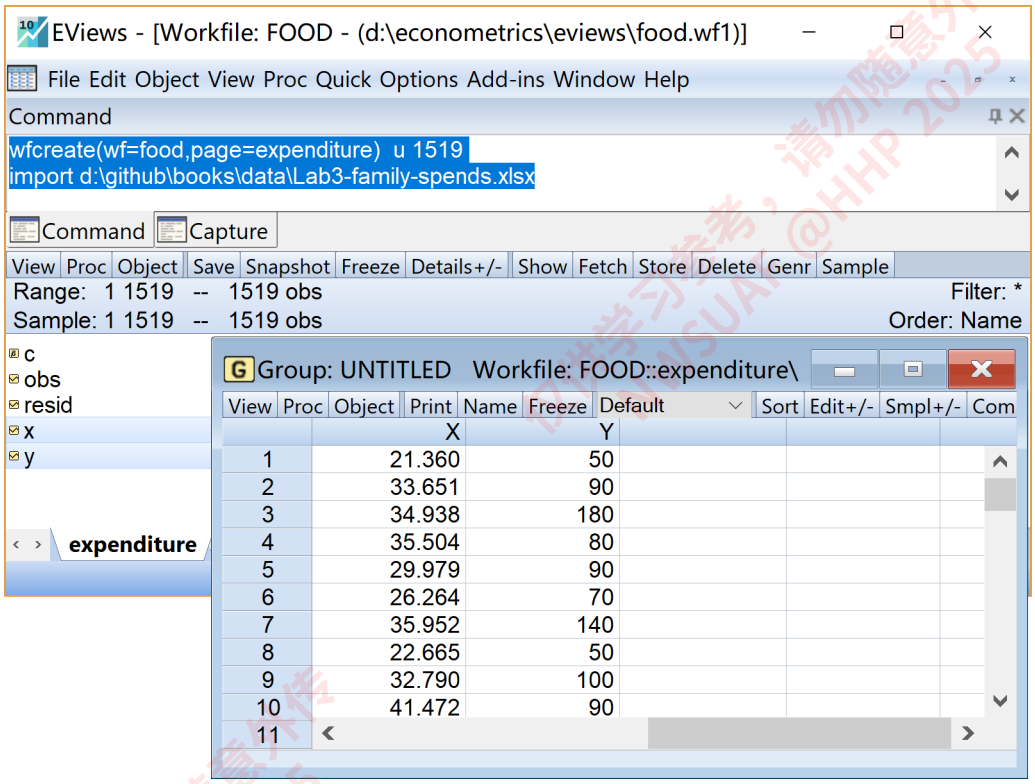


图 3.1: 创建工作文件并导入数据

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
'绘制散点图
graph scatter_classic.scat x y
graph scatter_loglog.scat log(x) log(y)
graph scatter_reclog.scat 1/x log(y)
```

在工作文件视窗下，可以看到如下新生成的图形对象，可以双击查看（见图 3.2 ）：

- X 与 Y 的散点图scatter_classic
- $log(X)$ 与 $log(Y)$ 的散点图scatter_loglog
- $1/X$ 与 $log(Y)$ 的散点图scatter_reclog

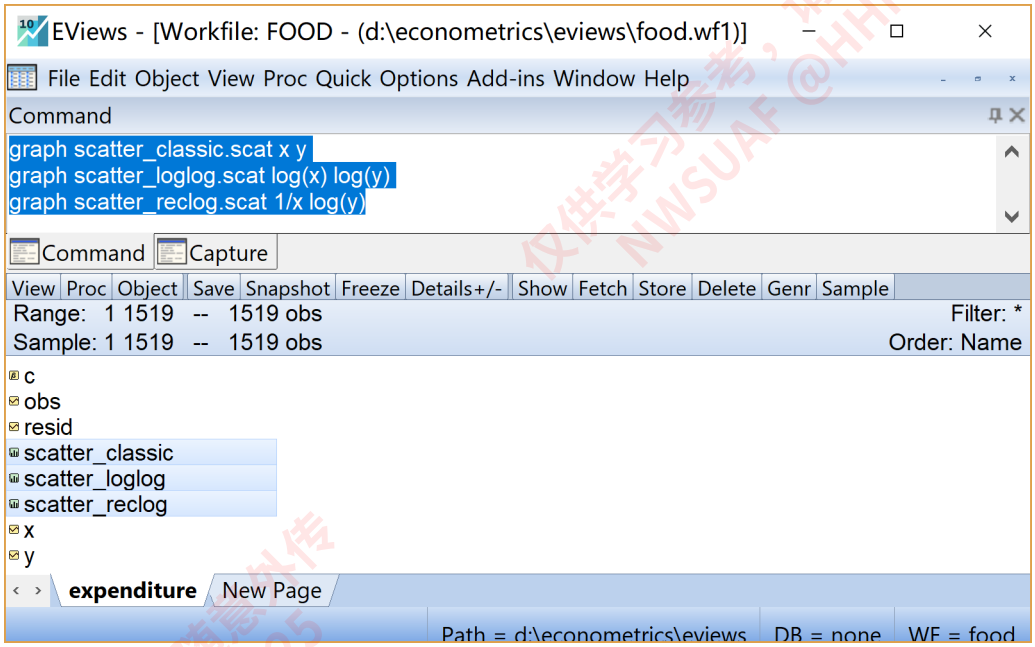
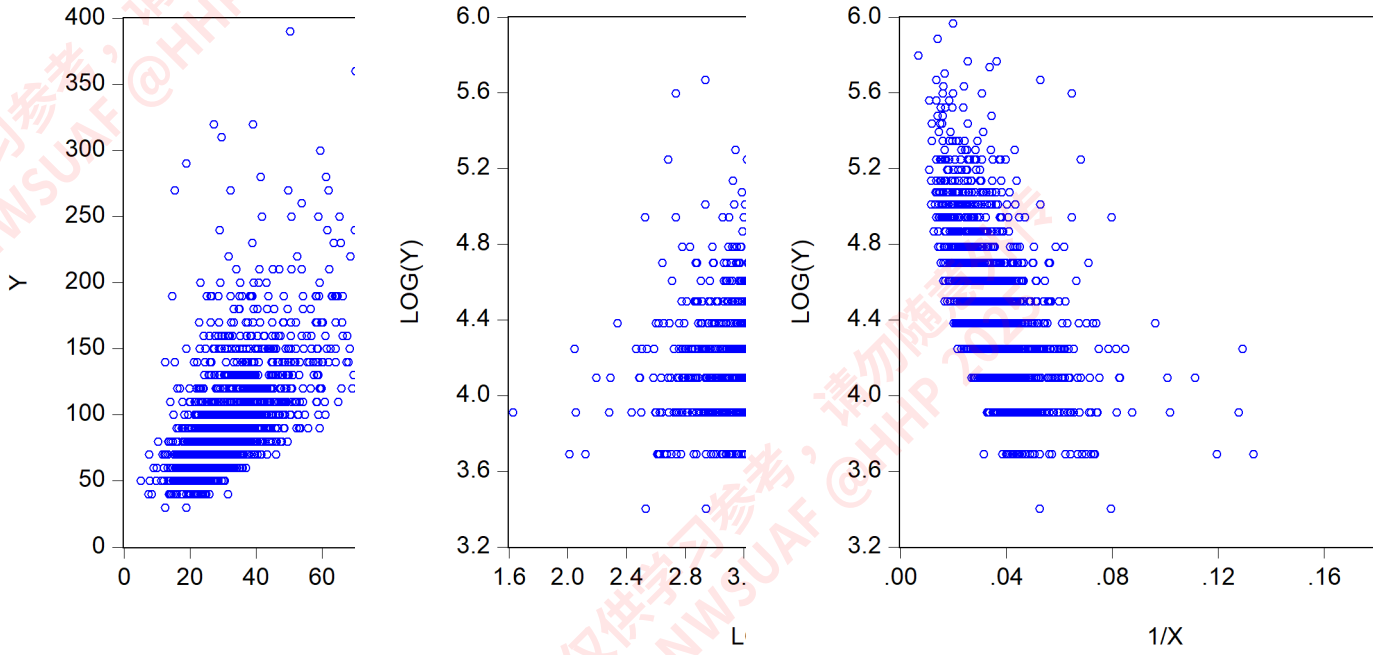


图 3.2: 创建工作文件并导入数据

> **提示**
>
> 上述过程也可以通过菜单操作实现：
>
> - EViews 主窗口中，依次点击菜单栏 ⇒ `Quick` ⇒ `Graph`
> - 图形引导窗口中，依次点击选择 ⇒ `Graph type`（图类型）⇒ `General`（通用）⇒ `Basic graph`（基本图形）⇒ `Scatter`（散点图）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

三类散点图对比见图 3.3：



(a) a1 (b) a2 (c) a3

图 3.3: 散点图对比

3.5.3 对 7 类模型进行 Eviews 回归分析

- Eviews 操作目标：得到回归方程，查看回归结果
- Eviews 操作思路：构建回归方程对象

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 生成7类线性回归模型的方程对象
equation m1.ls y c x                '经典模型m1
equation m2.ls y x                  '过原点模型m2
equation m3.ls log(y) c log(x)      '双对数模型m3
equation m4.ls log(y) c x           '对数线性模型m4
```

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
equation m5.ls y c log(x)          '线性对数模型m5
equation m6.ls y c 1/x             '倒数模型m6
equation m7.ls log(y) c 1/x        '对数倒数模型m7
```

在工作文件视窗下，可以看到如下新生成的方程对象，可以双击查看（见图 3.4 ）：

- 线性模型 M1 的回归方程🟰m1
- 过原点模型 M2 的回归方程🟰m2
- 双对数模型 M3 的回归方程🟰m3
- 线性到对数模型 M4 的回归方程🟰m4
- 对数到线性模型 M5 的回归方程🟰m5
- 倒数模型 M6 的回归方程🟰m6
- 对数倒数模型 M7 的回归方程🟰m7

7 类回归方程结果对比分析见图 3.5 ：

经典线性回归模型 m1 的简要回归报告如下：

$$\widehat{Y}_i = + \hat{\beta}_1 + \hat{\beta}_2 X_i \tag{3.19}$$

$$\begin{array}{lll} \widehat{Y} = & + 28.48 & + 2.13 X_i \\ (s) & (2.5259) & (0.0717) \\ (t) & (+11.27) & (+29.67) \end{array} \tag{3.20}$$

过原点模型 m2 的简要回归报告如下：

$$\widehat{Y}_i = + \hat{\beta}_1 \tag{3.21}$$

$$\begin{array}{ll} \widehat{Y} = & + 2.88 X_i \\ (s) & (0.0261) \\ (t) & (+111) \end{array} \tag{3.22}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

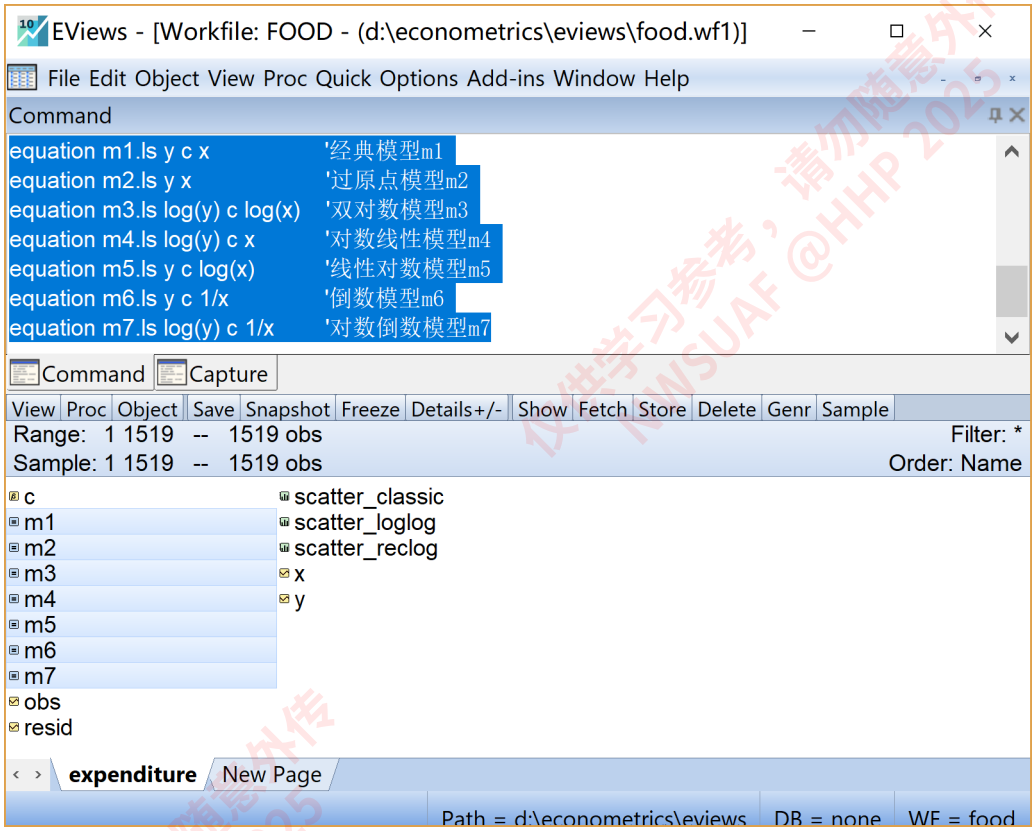
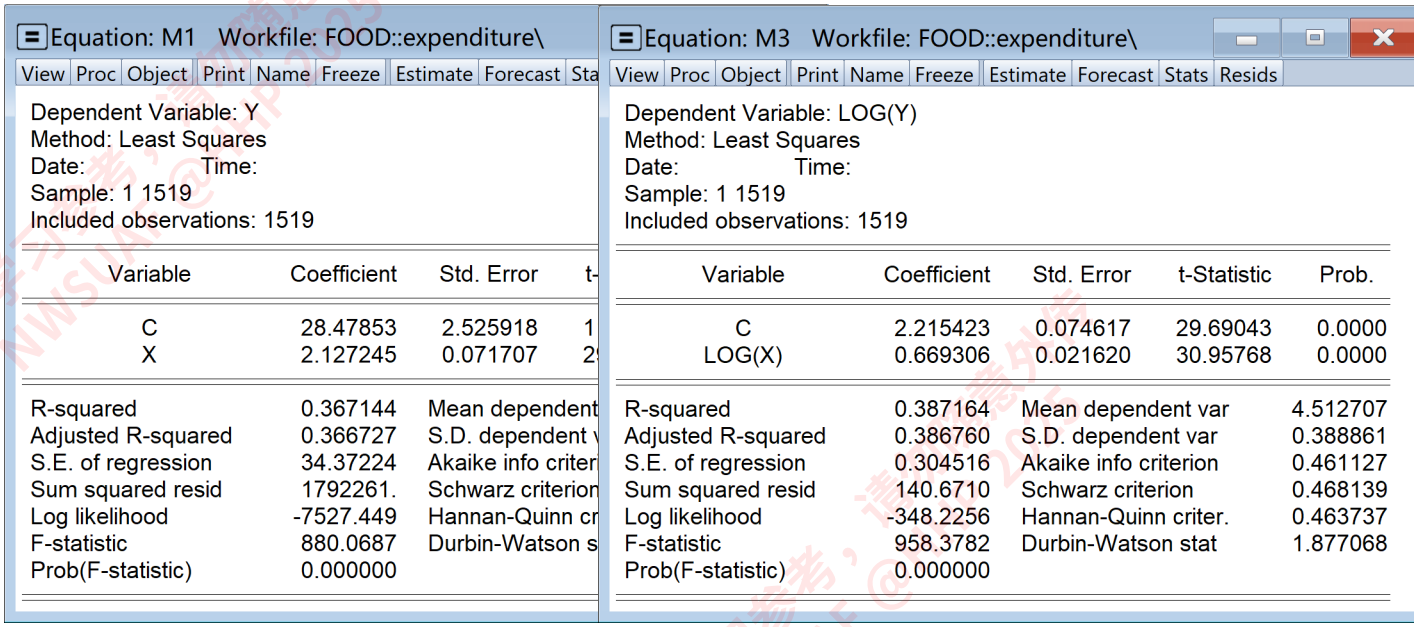


图 3.4: 生成 7 个回归方程对象

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025



Equation: M1 Workfile: FOOD::expenditure\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Sta

Dependent Variable: Y
Method: Least Squares
Date: Time:
Sample: 1 1519
Included observations: 1519

Variable	Coefficient	Std. Error	t-
C	28.47853	2.525918	1
X	2.127245	0.071707	2

R-squared	0.367144	Mean dependent	
Adjusted R-squared	0.366727	S.D. dependent v	
S.E. of regression	34.37224	Akaike info criter	
Sum squared resid	1792261.	Schwarz criterion	
Log likelihood	-7527.449	Hannan-Quinn cr	
F-statistic	880.0687	Durbin-Watson s	
Prob(F-statistic)	0.000000		

Equation: M3 Workfile: FOOD::expenditure\

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

Dependent Variable: LOG(Y)
Method: Least Squares
Date: Time:
Sample: 1 1519
Included observations: 1519

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.215423	0.074617	29.69043	0.0000
LOG(X)	0.669306	0.021620	30.95768	0.0000

R-squared	0.387164	Mean dependent var	4.512707
Adjusted R-squared	0.386760	S.D. dependent var	0.388861
S.E. of regression	0.304516	Akaike info criterion	0.461127
Sum squared resid	140.6710	Schwarz criterion	0.468139
Log likelihood	-348.2256	Hannan-Quinn criter.	0.463737
F-statistic	958.3782	Durbin-Watson stat	1.877068
Prob(F-statistic)	0.000000		

(a) a1 (b) a2

图 3.5: 经典模型 m1 和双对数模型 m3 的对比

双对数回归模型 m3 的简要回归报告如下：

$$\widehat{log(Y)}_i = + \hat{\beta}_1 + \hat{\beta}_2 log(X)_i \tag{3.23}$$

$$\begin{aligned} \widehat{log(Y)} = & \quad + 2.22 \quad + 0.67 ln(X)_i \\ (s) & \quad (0.0746) \quad (0.0216) \\ (t) & \quad (+29.69) \quad (+30.96) \end{aligned} \tag{3.24}$$

对数线性回归模型 m4 的简要回归报告如下：

$$\widehat{log(Y)}_i = + \hat{\beta}_1 + \hat{\beta}_2 X_i \tag{3.25}$$

$$\begin{aligned} \widehat{log(Y)} = & \quad + 3.86 \quad + 0.02 X_i \\ (s) & \quad (0.0222) \quad (0.0006) \\ (t) & \quad (+173.48) \quad (+31.48) \end{aligned} \tag{3.26}$$

线性对数回归模型 m5 的简要回归报告如下：

$$\widehat{Y}_i = + \hat{\beta}_1 + \hat{\beta}_2 log(X)_i \tag{3.27}$$

$$\begin{array}{llll} \widehat{Y} = & -133.81 & +67.74 ln(X)_i \\ (s) & (8.7212) & (2.5269) \\ (t) & (-15.34) & (+26.81) \end{array} \tag{3.28}$$

线性倒数回归模型 m6 的简要回归报告如下：

$$\widehat{Y}_i = + \hat{\beta}_1 + \hat{\beta}_2 \frac{1}{X}_i \tag{3.29}$$

$$\begin{array}{llll} \widehat{Y} = & +148.38 & -1437.00(1/X_i) \\ (s) & (2.5948) & (69.5030) \\ (t) & (+57.18) & (-20.68) \end{array} \tag{3.30}$$

对数倒数回归模型 m7 的简要回归报告如下：

$$\widehat{log(Y)}_i = + \hat{\beta}_1 + \hat{\beta}_2 \frac{1}{X}_i \tag{3.31}$$

$$\begin{array}{llll} \widehat{log(Y)} = & +5.03 & -14.95(1/X_i) \\ (s) & (0.0222) & (0.5954) \\ (t) & (+226.29) & (-25.12) \end{array} \tag{3.32}$$

3.5.4 提取各类回归方程的回归系数

- Eviews 操作目标：得到各类回归的回归系数
- Eviews 操作思路：用 EViews 代码提取各类回归方程中的回归系数

在命令视窗中依次输入并运行如下 EViews 代码：

```
'得到7个线性回归模型的回归系数(向量)
vector(2) coef1=m1.@coefs
vector(1) coef2=m2.@coefs
vector(2) coef3=m3.@coefs
vector(2) coef4=m4.@coefs
vector(2) coef5=m5.@coefs
vector(2) coef6=m6.@coefs
vector(2) coef7=m7.@coefs
```

在工作文件视窗下，可以看到如下新生成的向量对象，可以双击查看（见图 3.6 ）：

- 线性模型 M1 的回归方程的回归系数coef1
- 过原点模型 M2 的回归方程的回归系数coef2
- 双对数模型 M3 的回归方程的回归系数coef3
- 线性到对数模型 M4 的回归方程的回归系数coef4
- 对数到线性模型 M5 的回归方程的回归系数coef5
- 倒数模型 M6 的回归方程的回归系数coef6
- 对数倒数模型 M7 的回归方程的回归系数coef7

3.5.5 给出指定的样本数据点 (X_0, Y_0)

- Eviews 操作目标：给出指定的样本数据点，便于后续计算使用
- Eviews 操作思路：把样本数据点录入 EViews 工作视窗

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 给定样本数据点(X0, Y0)
scalar x0=100
scalar Y0=30
```

在工作文件视窗下，可以看到如下新生成的标量对象，可以双击查看（见图 3.7 ）：

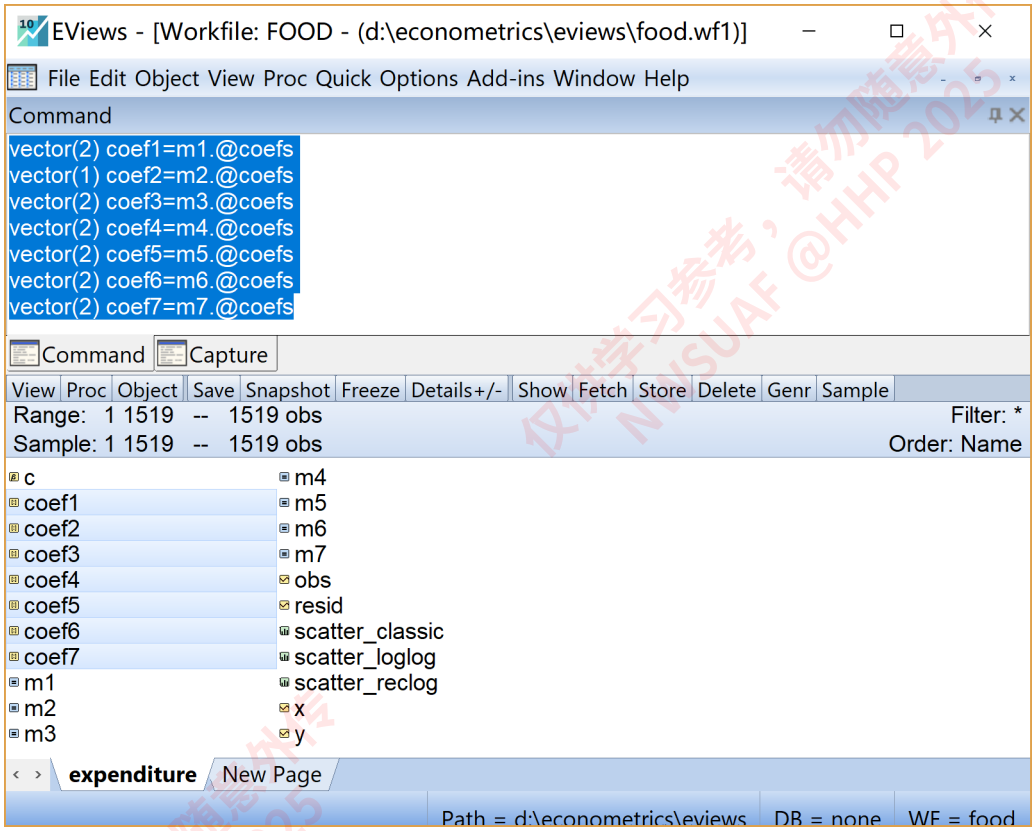


图 3.6: 提取得到 7 个回归方程的回归系数

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 样本点数据 X_0x0
- 样本点数据 Y_0y0

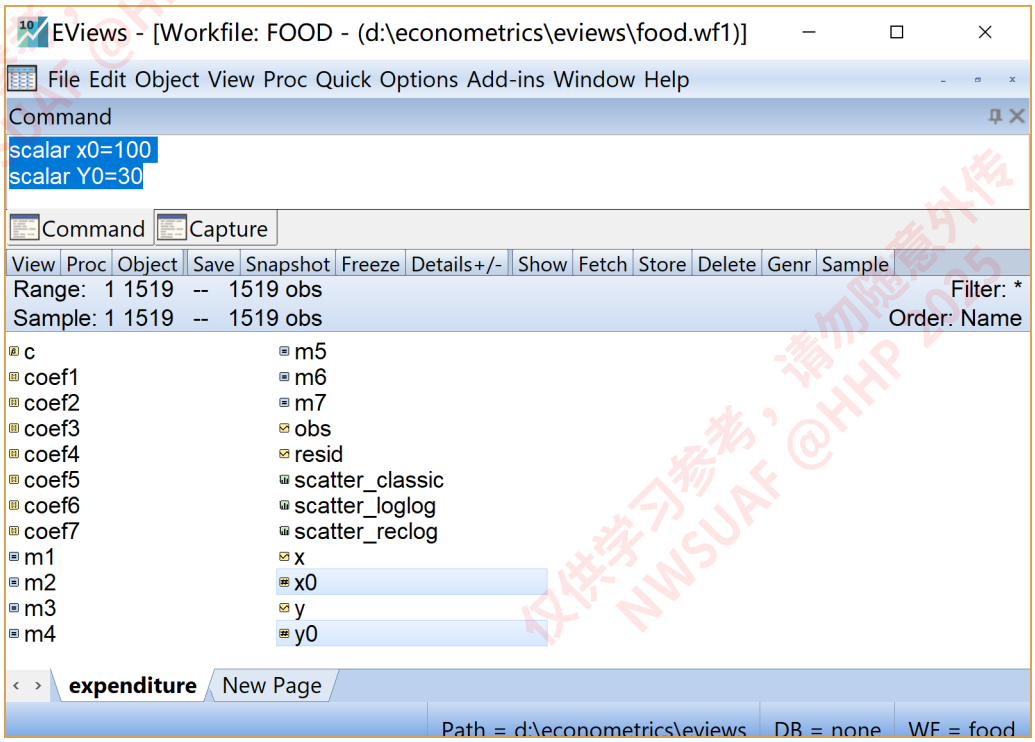


图 3.7: 提取得到 7 个回归方程的回归系数

3.5.6 得到 7 个线性模型中 Y 对 X 在 (X_0, Y_0) 处的点斜率

- Eviews 操作目标：得到各类模型的点斜率
- Eviews 操作思路：根据理论公式，计算各类模型的点斜率

在命令视窗中依次输入并运行如下 EViews 代码：

```
'得到7个线性回归模型中y对x在(X0,Y0)处点斜率
scalar slp1= coef1(2)
scalar slp2= coef2(1)
scalar slp3= coef3(2)*x0/y0
scalar slp4= coef4(2)*y0
```

```
scalar slp5= coef5(2)/x0
scalar slp6= coef6(2)/x0^2
scalar slp7= coef7(2)*y0/x0^2
```

在工作文件视窗下，可以看到如下新生成的标量对象，可以双击查看（见图 3.8 ）：

- 线性模型 M1 的回归方程中，Y 对 X 的斜率slp1
- 过原点模型 M2 的回归方程中，Y 对 X 的斜率slp2
- 双对数模型 M3 的回归方程中，Y 对 X 的斜率slp3
- 线性到对数模型 M4 的回归方程中，Y 对 X 的斜率slp4
- 对数到线性模型 M5 的回归方程中，Y 对 X 的斜率slp5
- 倒数模型 M6 的回归方程中，Y 对 X 的斜率slp6
- 对数倒数模型 M7 的回归方程中，Y 对 X 的斜率slp7

3.5.7 得到 7 个线性模型中 Y 对 X 在 (X_0, Y_0) 处的点弹性

- Eviews 操作目标：得到各类模型的点弹性
- Eviews 操作思路：根据理论公式，计算各类模型的点弹性

在命令视窗中依次输入并运行如下 EViews 代码：

```
'得到7个线性回归模型中y对x在(X0,Y0)处的点弹性
scalar els_point1= coef1(2)*x0/y0
scalar els_point2= coef2(1)*x0/y0
scalar els_point3= coef3(2)
scalar els_point4= coef4(2)*x0
scalar els_point5= coef5(2)/y0
scalar els_point6= -coef6(2)/(x0*y0)
scalar els_point7= coef7(2)/x0
```

在工作文件视窗下，可以看到如下新生成的标量对象，可以双击查看（见图 3.9 ）：

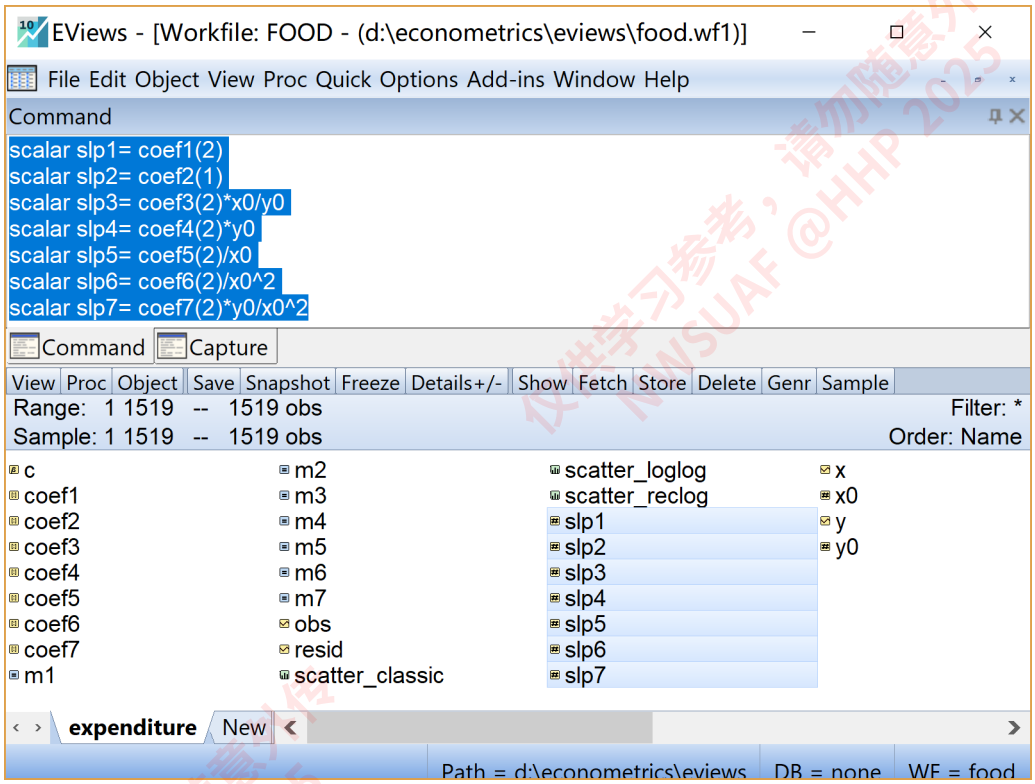


图 3.8: 计算得到 7 个回归方程中 Y 对 X 的斜率

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 线性模型 M1 的回归方程中，Y 对 X 的点弹性els_point1
- 过原点模型 M2 的回归方程中，Y 对 X 的点弹性els_point2
- 双对数模型 M3 的回归方程中，Y 对 X 的点弹性els_point3
- 线性到对数模型 M4 的回归方程中，Y 对 X 的点弹性els_point4
- 对数到线性模型 M5 的回归方程中，Y 对 X 的点弹性els_point5
- 倒数模型 M6 的回归方程中，Y 对 X 的点弹性els_point6
- 对数倒数模型 M7 的回归方程中，Y 对 X 的点弹性els_point7

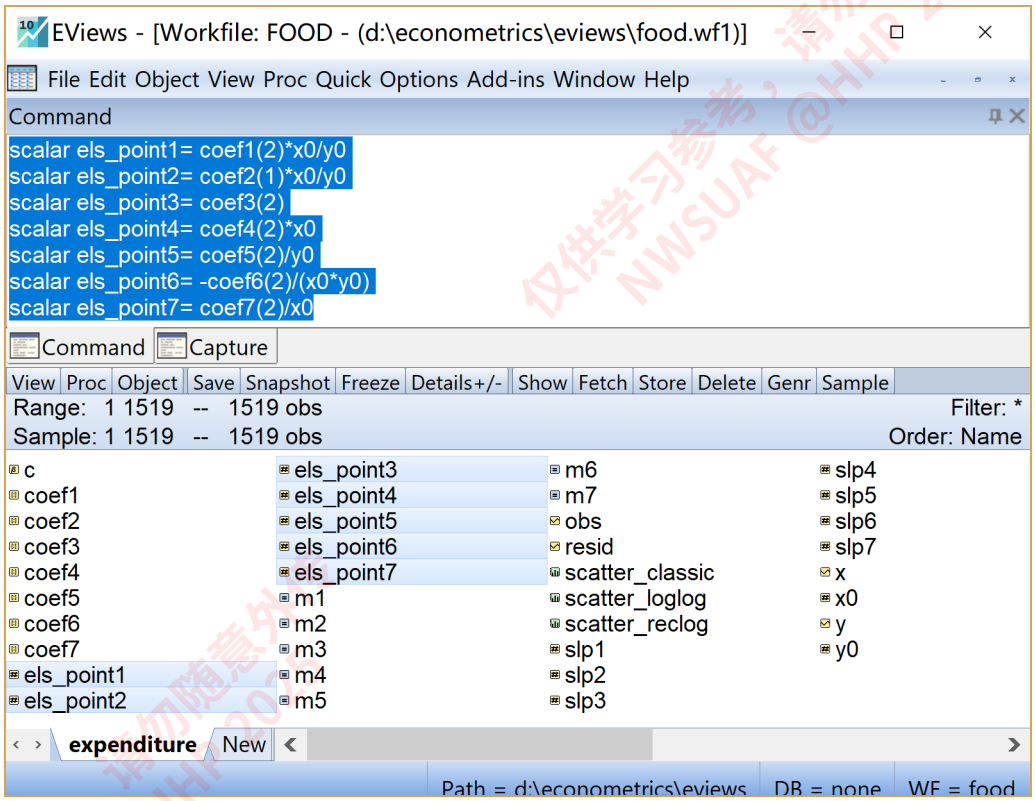


图 3.9: 计算得到 7 个回归方程中 Y 对 X 的斜率

3.5.8 得到 7 个线性模型中 Y 对 X 的平均弹性

- Eviews 操作目标：得到各类模型的平均弹性

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- Eviews 操作思路：根据理论公式，计算各类模型的平均弹性

在命令视窗中依次输入并运行如下 EViews 代码：

```
'得到7个线性回归模型中$Y$对$X$的平均弹性
scalar els_avr1= coef1(2)*@mean(x)/@mean(y)
scalar els_avr2= coef2(1)*@mean(x)/@mean(y)
scalar els_avr3= coef3(2)
scalar els_avr4= coef4(2)*@mean(x)
scalar els_avr5= coef5(2)/@mean(y)
scalar els_avr6= -coef6(2)/(@mean(x)*@mean(y))
scalar els_avr7= coef7(2)/@mean(x)
```

在工作文件视窗下，可以看到如下新生成的标量对象，可以双击查看（见图 3.10 ）：

- 线性模型 M1 的回归方程中，Y 对 X 的平均弹性els_avr1
- 过原点模型 M2 的回归方程中，Y 对 X 的平均弹性els_avr2
- 双对数模型 M3 的回归方程中，Y 对 X 的平均弹性els_avr3
- 线性到对数模型 M4 的回归方程中，Y 对 X 的平均弹性els_avr4
- 对数到线性模型 M5 的回归方程中，Y 对 X 的平均弹性els_avr5
- 倒数模型 M6 的回归方程中，Y 对 X 的平均弹性els_avr6
- 对数倒数模型 M7 的回归方程中，Y 对 X 的平均弹性els_avr7

3.5.9 把所有计算结果制作成表格

- Eviews 操作目标：将所有计算结果制作成表格，便于数据对比
- Eviews 操作思路：用 EViews 代码操作表格对象

在命令视窗中依次输入并运行如下 EViews 代码：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

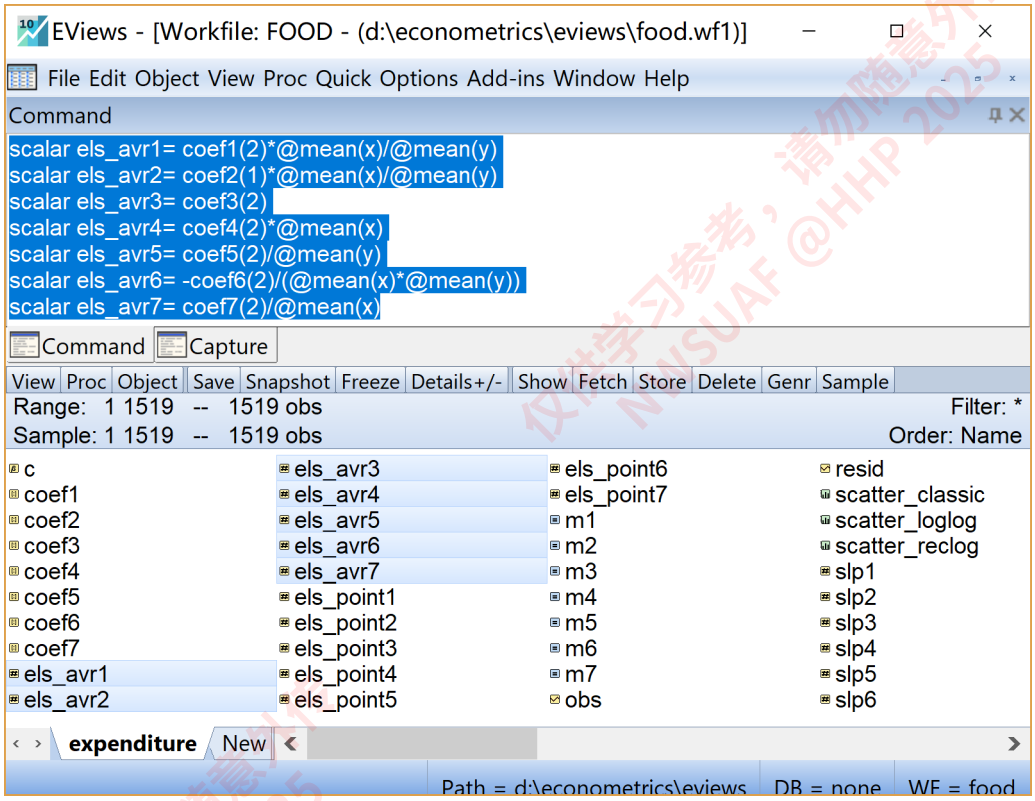


图 3.10: 计算得到 7 个回归方程中 Y 对 X 的斜率

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
' 把所有结果制作成表格
table(8,4) tab_show      '构造8行4列的表格
tab_show(1,1)="model"    '第1列存放模型代号
tab_show(2,1)="m1"
tab_show(3,1)="m2"
tab_show(4,1)="m3"
tab_show(5,1)="m4"
tab_show(6,1)="m5"
tab_show(7,1)="m6"
tab_show(8,1)="m7"

tab_show(1,2)="slop"     '第2列存放$Y$对$X$的点斜率
tab_show(2,2)=slp1
tab_show(3,2)=slp2
tab_show(4,2)=slp3
tab_show(5,2)=slp4
tab_show(6,2)=slp5
tab_show(7,2)=slp6
tab_show(8,2)=slp7

tab_show(1,3)="els_point"   '第3列存放$Y$对$X$的点弹性
tab_show(2,3)=els_point1
tab_show(3,3)=els_point2
tab_show(4,3)=els_point3
tab_show(5,3)=els_point4
tab_show(6,3)=els_point5
tab_show(7,3)=els_point6
tab_show(8,3)=els_point7

tab_show(1,4)="els_avr"     '第4列存放$Y$对$X$的平均弹性
tab_show(2,4)=els_avr1
tab_show(3,4)=els_avr2
tab_show(4,4)=els_avr3
tab_show(5,4)=els_avr4
tab_show(6,4)=els_avr5
```

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
tab_show(7,4)=els_avr6
tab_show(8,4)=els_avr7
```

在工作文件视窗下，可以看到如下新生成的表格对象tab_show，可以双击查看（见图 3.11 ）。

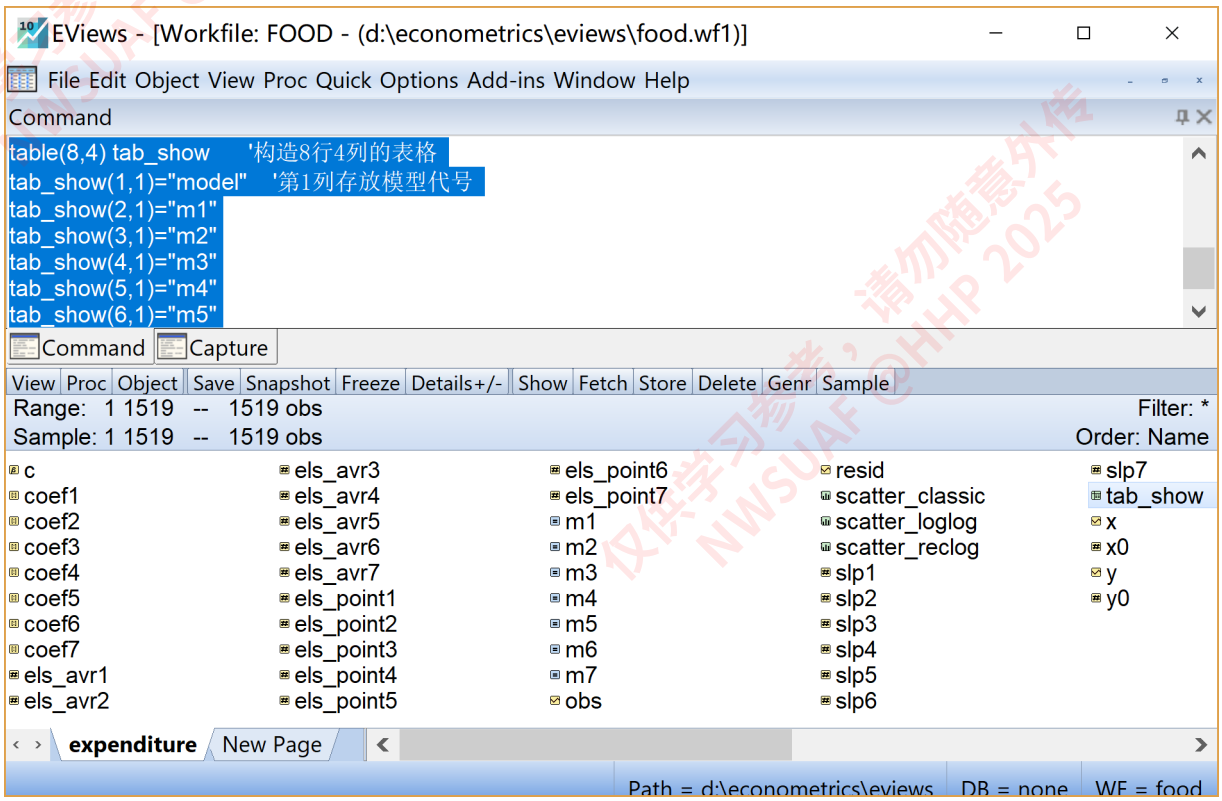


图 3.11: 计算结果对比分析表

7 类回归方程计算结果对比分析见图 3.12 ：

3.5.10 对倒数模型 M_6 进行 Eviews 回归分析并提取回归系数

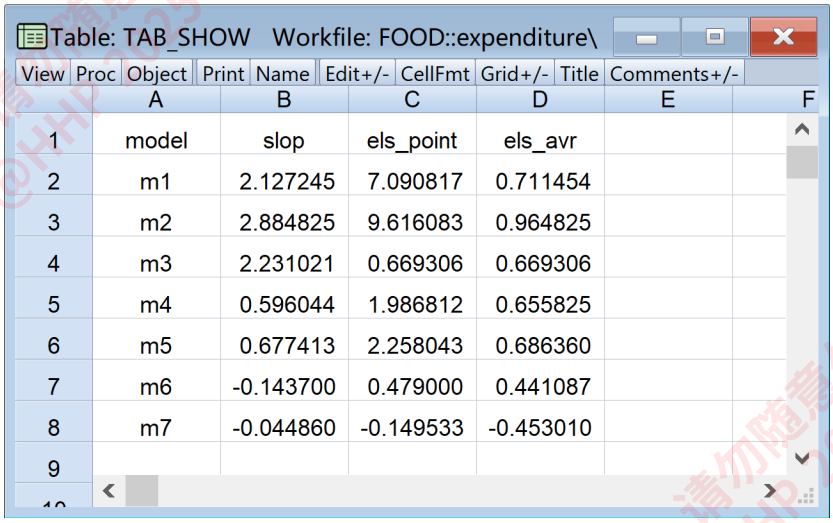
Eviews 操作目标：得到 Y_i 相对于 $1/X_i$ 的 Eviews 回归报告，提取回归系数

Eviews 操作思路：

（1）利用常规流程获得 Eviews 回归分析：

- 菜单操作：Quick–>estimate equation–> `y c 1/x`

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025



	A	B	C	D	E	F
1	model	slop	els_point	els_avr		
2	m1	2.127245	7.090817	0.711454		
3	m2	2.884825	9.616083	0.964825		
4	m3	2.231021	0.669306	0.669306		
5	m4	0.596044	1.986812	0.655825		
6	m5	0.677413	2.258043	0.686360		
7	m6	-0.143700	0.479000	0.441087		
8	m7	-0.044860	-0.149533	-0.453010		
9						

图 3.12: 7 类回归方程计算结果对比

（2）保存好分析报告（建议命名为 m6）：

- 菜单操作：Name–>m6

（3）提取报告中的回归系数（建议命名为 `coef6`）：

- 命令操作：`coef coef6=c`

Eviews 操提示：两个 Eviews 模型内置对象（C 和resid）

（1）C 和resid 都是 Eviews 模型内置对象，一旦建立 workfile 就会系统产生，用户不能对它进行删除或重命名（delete or rename）操作。

- C 属于**系数**对象（coef object），这类对象主要用于表示系数列向量（coefficient column vector）。C 是用来装载回归模型的系数 $\hat{\beta}_1$ 和 $\hat{\beta}_2$……
- resid 属于**序列**对象（series object），这类对象主要用于表示序列（series）。resid 是用来装载回归模型的残差 e_i

（2）C 和resid 是“临时容器”。它们只会装载最近一次 Eviews 建模分析（Estimate Equation）时的回归系数 $\hat{\beta}_i$ 和回归残差 e_i。一旦用户进行了新的 Eviews 建模分析（Estimate Equation）操作，它们就立即会被最新回归建模的系数和残差信息所“更新”。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

（3）如果用户要创建多个回归方程，又想保留每个回归方程的回归系数 $\hat{\beta}_i$ 和残差 e_i，可以通过下面两种方法将结果从“临时容器”中提取出来，并保存到指定的对象中去：

提示

a）键鼠操作法。对C 或rsid 右键拷贝（copy），然后在窗口区粘贴（paste），并进行重命名，保存。
b）命令操作法。

- 在命令窗口中输入 Eviews 命令 `coef c01=c`，即可得到当前回归模型系数C 的复制品 `c01`（用户可以自己定义对象名称）。
- 在命令窗口中输入 Eviews 命令 `series resid01=resid`，即可得到当前回归模型系数resid 的复制品 `resid01`（用户可以自己定义对象名称）。

（4）“临时容器”C 和rsid 按照 Eviews 建模先后次序不断被“更新”信息（分别是回归系数和残差序列）。“更新”基本原则是“依次占据对象的空间位置”。

（5）回归方程的系数 $\hat{\beta}_i$ 会依次占据C 对象的第 1 个单元格、第 2 个单元格、……，回归系数个数占据空间位置的多少因模型方程的不同而不同。尤其要注意的情形是：上一次操作回归模型的回归系数多，而最近一次操作回归模型的回归系数少。

（6）回归方程的残差resid 在同一个工作文件（workfile）下，用户进行各类模型操作的样本数大多保持相同，因此，最近依次模型操作一般都会“完全更新”上一次模型操作的残差序列。

3.5.11 倒数模型 M_6 情形下，计算 Y 相对于 X 的点斜率 $(X_0 = 100, Y_0 = 30)$ 点斜率 (slope)

Eviews 操作目标：得到 Y 相对于 X 的点斜率 $(X_0 = 100, Y_0 = 30)$ 点斜率 (slope)

- 总体回归模型（PRM）：$Y_i = \beta_1 + \beta_2 / X_i + u_i$
- 点斜率计算公式：β_2 / X_i^2

Eviews 操作思路：

- 提取 `coef6` 的第二个值（也即 $\hat{\beta}_2$），利用斜率公式计算得到 Y 相对于 X 的点斜率（建议命名为 slp6）：命令操作：`scalar slp6= -coef6(2)/100^2`

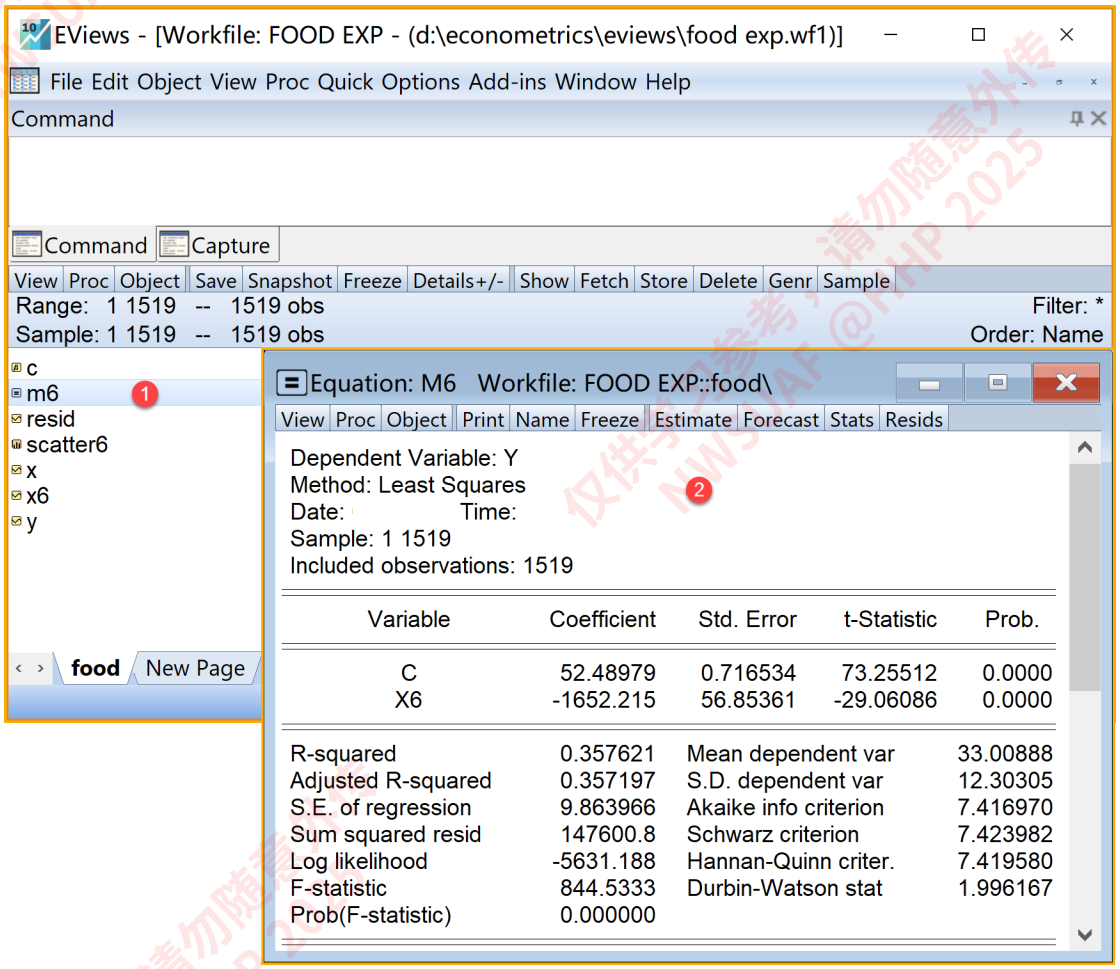


图 3.13: 构造倒数模型的回归方程

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

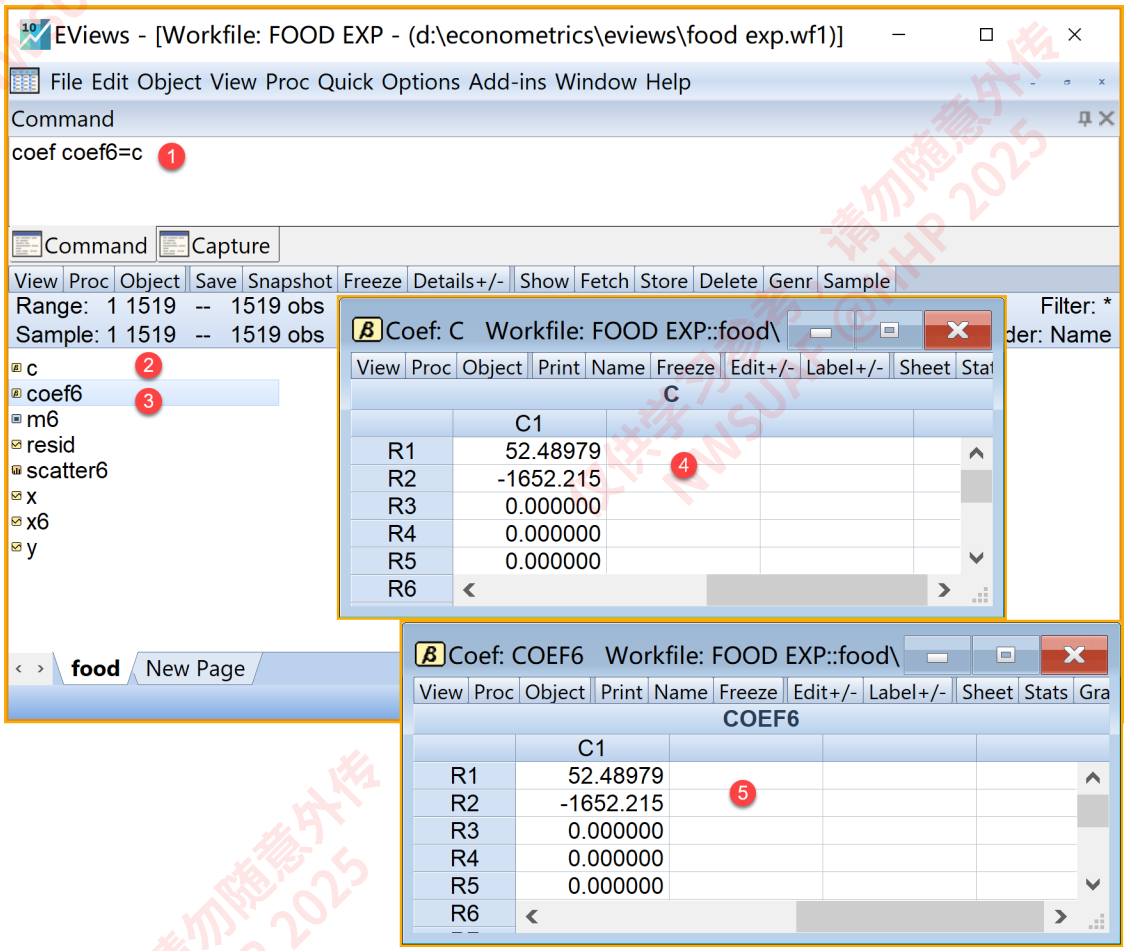


图 3.14: 提取回归方程的系数

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

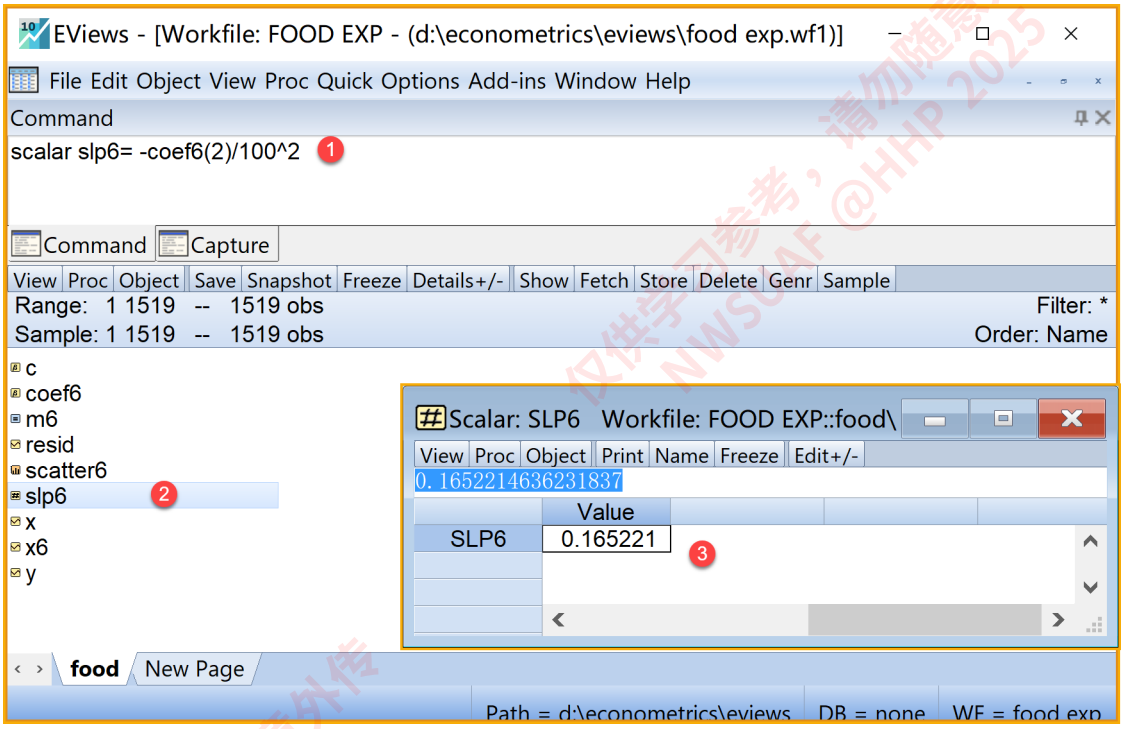


图 3.15: 得到 Y 对 X 的点斜率

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

3.5.12 倒数模型 M_6 情形下，计算 Y 相对于 X 的平均弹性 (elasticity)

Eviews 操作目标：得到 Y 相对于 X 的平均弹性 (elasticity)

- 总体回归模型（PRM）：$Y_i = \beta_1 + \beta_2/X_i + u_i$
- 平均弹性计算公式：$-\beta_2/(\bar{X}\bar{Y}$

Eviews 操作思路：

（1）求出标量 $\bar{X}$ 和 $\bar{Y}$（建议分别命名为 x_mean 和 y_mean）：

- 命令操作：`scalar x_mean=@mean(x)`
- 命令操作：`scalar y_mean=@mean(y)`

（2）提取 `coef6` 的第二个值（也即 $\hat{\beta}_2$），利用平均弹性公式计算得到 Y 相对于 X 的平均弹性（建议命名为 els6）：

- 命令操作：`scalar els6= -coef6(2)/(x_mean*y_mean)`

3.6 附录：prg 源代码

```
'=================================================================================================
'说明：以下为EViews编程文件foodexp.prg的代码
'将展示第三章中“英国家庭食物支出案例”主要分析步骤的“批量式命令驱动”实现方法(::
'其中，符号'起始的行，为注释行，其他为EViews命令行。
'=================================================================================================

'创建工作文件（工作文件名=food，子页命名=expenditure），无结构无日期，样本数为1519
wfcreate(wf=food,page=expenditure)  u 1519

'导入外部数据，路径为d:\github\books\data\Lab3-family-spends.xlsx
import d:\github\books\data\Lab3-family-spends.xlsx

' 给定样本数据点(X0, Y0)
scalar x0=100
```

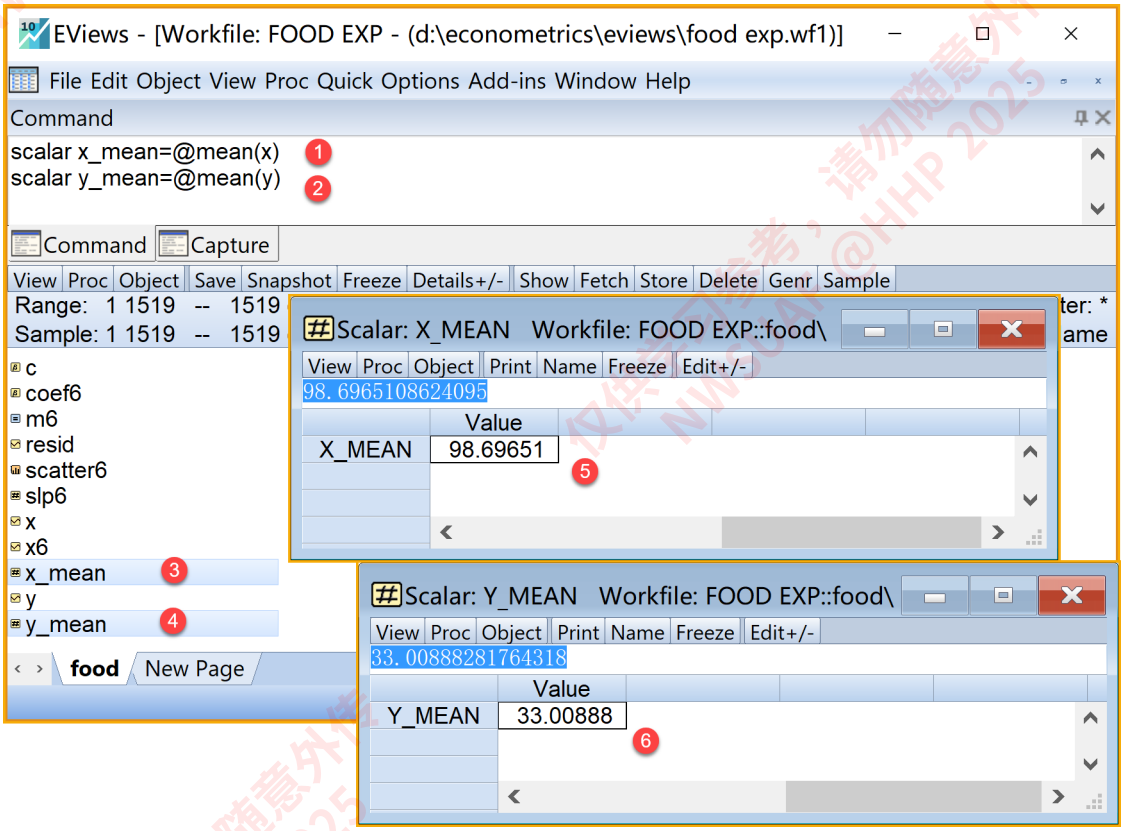


图 3.16: 得到 Y 和 X 的均值

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

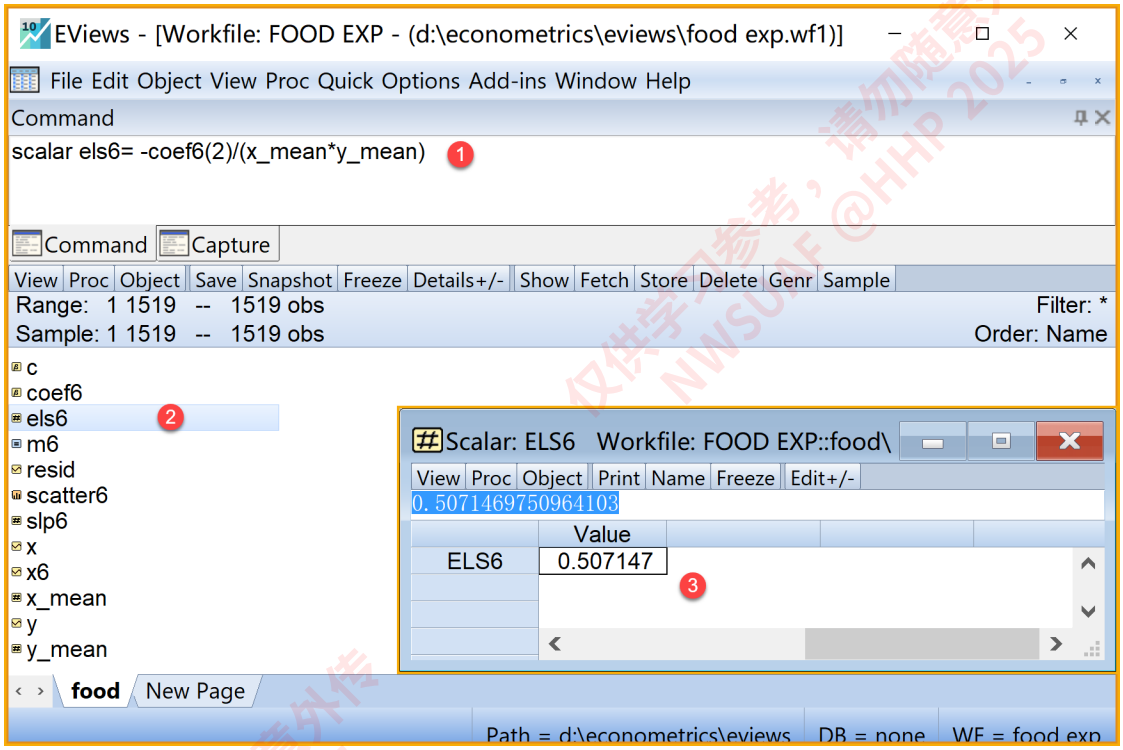


图 3.17: 得到 Y 对 X 的平均弹性

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
scalar Y0=30

' 生成7类线性回归模型的方程对象
equation m1.ls y c x                '经典模型m1
equation m2.ls y x                  '过原点模型m2
equation m3.ls log(y) c log(x)      '双对数模型m3
equation m4.ls log(y) c x           '对数线性模型m4
equation m5.ls y c log(x)           '线性对数模型m5
equation m6.ls y c 1/x              '倒数模型m6
equation m7.ls log(y) c 1/x         '对数倒数模型m7


'得到7个线性回归模型的回归系数(向量)
vector(2) coef1=m1.@coefs
vector(1) coef2=m2.@coefs
vector(2) coef3=m3.@coefs
vector(2) coef4=m4.@coefs
vector(2) coef5=m5.@coefs
vector(2) coef6=m6.@coefs
vector(2) coef7=m7.@coefs


'得到7个线性回归模型中y对x在(X0,Y0)处点斜率
scalar slp1= coef1(2)
scalar slp2= coef2(1)
scalar slp3= coef3(2)*x0/y0
scalar slp4= coef4(2)*y0
scalar slp5= coef5(2)/x0
scalar slp6= coef6(2)/x0^2
scalar slp7= coef7(2)*y0/x0^2

'得到7个线性回归模型中y对x在(X0,Y0)处的点弹性
scalar els_point1= coef1(2)*x0/y0
scalar els_point2= coef2(1)*x0/y0
scalar els_point3= coef3(2)
```


仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
scalar els_point4= coef4(2)*x0
scalar els_point5= coef5(2)/y0
scalar els_point6= -coef6(2)/(x0*y0)
scalar els_point7= coef7(2)/x0

'得到7个线性回归模型中$Y$对$X$的平均弹性
scalar els_avr1= coef1(2)*@mean(x)/@mean(y)
scalar els_avr2= coef2(1)*@mean(x)/@mean(y)
scalar els_avr3= coef3(2)
scalar els_avr4= coef4(2)*@mean(x)
scalar els_avr5= coef5(2)/@mean(y)
scalar els_avr6= -coef6(2)/(@mean(x)*@mean(y))
scalar els_avr7= coef7(2)/@mean(x)

' 把所有结果制作成表格
table(8,4) tab_show                '构造8行4列的表格

tab_show(1,1)="model"     '第1列存放模型代号
tab_show(2,1)="m1"
tab_show(3,1)="m2"
tab_show(4,1)="m3"
tab_show(5,1)="m4"
tab_show(6,1)="m5"
tab_show(7,1)="m6"
tab_show(8,1)="m7"

tab_show(1,2)="slop"     '第2列存放$Y$对$X$的点斜率
tab_show(2,2)=slp1
tab_show(3,2)=slp2
tab_show(4,2)=slp3
tab_show(5,2)=slp4
tab_show(6,2)=slp5
tab_show(7,2)=slp6
tab_show(8,2)=slp7
```

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
tab_show(1,3)="els_point"    '第3列存放$Y$对$X$的点弹性
tab_show(2,3)=els_point1
tab_show(3,3)=els_point2
tab_show(4,3)=els_point3
tab_show(5,3)=els_point4
tab_show(6,3)=els_point5
tab_show(7,3)=els_point6
tab_show(8,3)=els_point7

tab_show(1,4)="els_avr"      '第4列存放$Y$对$X$的平均弹性
tab_show(2,4)=els_avr1
tab_show(3,4)=els_avr2
tab_show(4,4)=els_avr3
tab_show(5,4)=els_avr4
tab_show(6,4)=els_avr5
tab_show(7,4)=els_avr6
tab_show(8,4)=els_avr7


' ================================================================================
```

3.7 作业题

家庭食物支出：表 3.3 给出了 1980-1982 年间英国家庭支出调查中 1 519 个家庭在 Y 家庭总支出 (10 英镑)，X 食物支出 (10 英镑) 等方面的数据。数据只包括住在伦敦市区和市郊有 1~2 个子女的家庭，样本不包括自我雇佣和退休家庭。

表 3.3: 家庭食物支出 (n=1516)

X1	X	Ydata1	Ydata2	Ydata3	Ydata4	Ydata5	Ydata6
id	X	2015014495	2016010317	2016011222	2016013009	2016014336	2016014344
name	X	刘琳	王雪明	韩双瑞	任畅	黄艺婕	高泽川

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 3.3: 家庭食物支出 (n=1516)

X1	X	Ydata1	Ydata2	Ydata3	Ydata4	Ydata5	Ydata6
class	X	保险 1601	保险 1601	保险 1601	保险 1601	保险 1601	保险 1601
n1	21.36	50.78	51.06	50.62	50.67	47.49	49.79
n2	33.651000000000	89.06	88.76	88.94	88.49	89.11	89.43
n3	34.938000000000	180.24	181.67	181.26	178.85	179.53	180.55
n1513	54.131	189.09	188.46	190.37	192.73	189.75	191.04
n1514	24.872	40.08	40.58	39.79	41.09	39.81	40
n1515	36.954000000000	89.06	89.47	87.99	88.63	89.08	90.76
n1516	19.509	70.650000000000	68.06	70.05	70.06	72.680000000000	69.97

变量说明见表 3.4：

表 3.4: 变量定义及说明

variable	label
obs	家庭编号 ID
Y	家庭总支出 (10 英镑)
X	食物支出 (10 英镑)

根据上述资料请回答如下问题：

a. 利用家庭总支出 (Y) 与食物支出 (X) 数据，通过对表**??** 中概括的各类模型，对变量进行相应变换，并分别作出散点图（七个图）。（提示：分别把图拷贝过来）

b. 利用家庭总支出 (Y) 与食物支出 (X) 数据，对表**??** 中概括的各类模型进行回归拟合（七个模型）？（提示：分别把分析报告截图复制过来）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

c. 利用 (b) 的分析结果，分别计算各模型的平均弹性 (els) 以及在点 $(X_0 = 100, Y_0 = 30)$ 处的斜率 (slp)。

d. 基于 (a) 和 (b) 中得到的结果，你认为哪个模型看来比较适当？

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

4 多元线性回归

仅供学习参考，请勿随意外传
NWSUAF @HHP 2025

4.1 实验目的及要求

目的：掌握多元线性回归模型的估计、检验。

要求：在老师指导下完成多元线性回归模型的建立、估计、统计检验，得到正确的分析结果；能运用矩阵方法实现前述操作。

4.2 实验原理

当多元线性回归模型在满足线性模型古典假设的前提下，最小二乘估计结果具有无偏性、有效性等性质，在此基础上进一步对估计所得的模型进行经济意义检验及统计检验。

4.2.1 基本模型和重要概念的矩阵表达

对于如下的 k 变量线性回归模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad \text{(PRM)} \tag{4.1}$$

如果样本数为 n，则可以将上述 PRM 模型表达为矩阵形式：

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{4.2}$$

进一步地，可以得到精简化的 PRM 矩阵形式：

$$\begin{matrix} \mathbf{y} = & \mathbf{X} & + & \mathbf{u} \\ (n \times 1) & (n \times k)(k \times 1) & & (n \times 1) \end{matrix} \tag{4.3}$$

其中：

- **向量**（默认为列向量）用加粗体的小写字母表达
- **矩阵**用大写粗体字母表达

- 矩阵或向量的维度需要注意标明

进一步地，我们可以用矩阵方法表达正态经典线性回归模型假设（N-CLRM）：

- 多元回归情形下，自变量 X 间无完全共线性。可记为 $\rho(\mathbf{X}) = k$，也即矩阵 $\mathbf{X}$ 为列满秩
- 随机干扰项期望为 0。可记为 $E(\mathbf{u}) = \mathbf{0}$
- 随机干扰项同方差且无自相关。可记为 $E(\mathbf{uu}') = \sigma^2\mathbf{I}$
- 在正态性假设下，关于随机干扰项的全部假设可以记为 $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$

随机干扰项的方差协方差矩阵为：

$$
\begin{aligned}
var - cov(\mathbf{u}) &= E(\mathbf{uu}') \\
&= E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \\
&= E\begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \vdots & \vdots & \vdots & \vdots \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{bmatrix} \\
&= \begin{bmatrix} E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n) \\ \vdots & \vdots & \vdots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2) \end{bmatrix}
\end{aligned}
$$

如果满足 N-CLRM 假设，则随机干扰项的方差协方差矩阵进一步可以写成：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{aligned}
var-cov(\mathbf{u}) &= E(\mathbf{u}\mathbf{u}') \\
&= \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_n^2 \end{bmatrix} && \leftarrow (E(u_i) = 0) \\
&= \begin{bmatrix} \sigma^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma^2 \end{bmatrix} && \leftarrow (var(u_i) = \sigma^2) \\
&= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} && \leftarrow (cov(u_i, u_j) = 0, i \neq j) \\
&= \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \\
&= \sigma^2 \mathbf{I}
\end{aligned}
$$

4.2.2 OLS 估计及 BLUE 性质的矩阵表达

给定如下的样本回归模型（SRM）：

$$
Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i \quad \text{(SRM)} \tag{4.4}
$$

最小二乘方法将求解最小化过程：

$$
\begin{aligned}
Q &= \sum e_i^2 \\
&= \mathbf{e}'\mathbf{e} \\
&= (\mathbf{y} - \mathbf{X}\hat{})'(\mathbf{y} - \mathbf{X}\hat{}) \\
&= \mathbf{y}'\mathbf{y} - 2\hat{}'\mathbf{X}'\mathbf{y} + \hat{}'\mathbf{X}'\mathbf{X}\hat{}
\end{aligned} \tag{4.5}
$$

进一步可以得到：

$$
\begin{aligned}
\frac{\partial Q}{\partial} &= 0 \\
\frac{\partial(\mathbf{y}'\mathbf{y} - 2\,'\mathbf{X}'\mathbf{y} + \hat{}\,'\mathbf{X}'\mathbf{X}\,)}{\partial} &= 0 \\
-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{} &= 0 \\
-\mathbf{X}'\mathbf{y} + \mathbf{X}'\mathbf{X}\hat{} &= 0 \\
\mathbf{X}'\mathbf{X}\hat{} &= \mathbf{X}'\mathbf{y}
\end{aligned}
$$

如果矩阵 $\mathbf{X}'\mathbf{X}$ 的逆矩阵存在，则两边同时左乘 $(\mathbf{X}'\mathbf{X})^{-1}$，得到 OLS 估计量：

$$
\hat{} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{4.6}
$$

对于回归系数的 OLS 估计量$\hat{}$，进一步讨论其方差和协方差矩阵 $var-cov(\ )$，一般记为：

$$
\begin{aligned}
var-cov(\ ) &= E\left(\left(\hat{\beta} - E(\hat{\beta})\right)\left(\hat{\beta} - E(\hat{\beta})\right)'\right) \\
&= \begin{bmatrix}
var(\hat{\beta}_1) & cov(\hat{\beta}_1, \hat{\beta}_2) & \cdots & cov(\hat{\beta}_1, \hat{\beta}_k) \\
cov(\hat{\beta}_2, \hat{\beta}_1) & var(\hat{\beta}_2) & \cdots & cov(\hat{\beta}_2, \hat{\beta}_k) \\
\vdots & \vdots & \vdots & \vdots \\
cov(\hat{\beta}_k, \hat{\beta}_1) & cov(\hat{\beta}_k, \hat{\beta}_2) & \cdots & var(\hat{\beta}_k)
\end{bmatrix}
\end{aligned}
$$

如果满足 N-CLRM 假设，则回归系数的 OLS 估计量$\hat{}$ 的方差和协方差矩阵 $var-cov(\ )$ 可以进一步可以写成：

$$
\begin{aligned}
var-cov(\ ) &= \mathbf{E}\left(\hat{} - \mathbf{E}(\hat{})\right)\left(\hat{} - \mathbf{E}(\hat{})\right)'\right) \\
&= \mathbf{E}\left(\hat{} - \ )\left(\hat{} - \ )'\right) \\
&= \mathbf{E}\left(((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right)'\right) \\
&= \mathbf{E}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned} \tag{4.7}
$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

此外，很用以证明 OLS 方法下，利用样本回归模型式 4.4 得到的估计量 $\hat{\sigma}^2$，是对总体回归模型式 4.1 参数 σ^2 的无偏估计，也即：

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\sum e_i^2}{n-k} = \frac{\mathbf{e}'\mathbf{e}}{n-k} \\ E(\hat{\sigma}^2) &= \sigma^2 \end{aligned} \tag{4.8}$$

那么，可以很快得到回归系数的 OLS 估计量$\hat{}$ 的**样本**方差和协方差矩阵 $S_{ij}^2(\hat{})$

$$\begin{aligned} S_{ij}^2(\hat{}) &= \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \\ &= \frac{\mathbf{e}'\mathbf{e}}{n-k}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \tag{4.9}$$

下面我们将证明高斯-马尔可夫定理 (Gauss-Markov Theorem)：在正态经典线性回归模型假设（N-CLRM）下，采用普通最小二乘法（OLS），得到的估计量 $\hat{\beta}$，是真实参数 β 最优的、线性的、无偏估计量（BLUE）。记为：

$$\xrightarrow[\text{N-CLRM}]{\text{OLS}} \hat{} \xrightarrow{\text{BLUE}}$$

线性性. 因为模型参数的 OLS 估计为：

$$\hat{} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

又因为矩阵 $\mathbf{X}$ 为列满秩，也即 $\rho(\mathbf{X}) = k$，所以$\hat{}$ 关于 $\mathbf{y}$ 是线性的。 □

无偏性. 根据模型参数 OLS 估计，容易得到如下过程：

$$\begin{aligned} \hat{} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X} \ + \mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \ + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \ + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \end{aligned}$$

进一步可证明

$$
\begin{aligned}
E(\hat{\ }) &= E(\ + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}) \\
&= \mathbf{E}(\ ) + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}(\mathbf{u}) \\
&=
\end{aligned}
$$

因此，$\hat{\ }$ 是参数 $\ $ 的无偏估计量得证。 □

方差最小/最优性. 假设存在用其他方法估计的线性无偏估计量 *，则要求 $\mathbf{C}$ 满足如下条件：

$$
\mathbf{CX} = 0
$$

从而保证如下式子成立：

$$
\begin{aligned}
^* &= \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}\right)\mathbf{y} \\
&= \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}\right)(\mathbf{X}\ + \mathbf{u}) \\
&= \ + \mathbf{CX}\ + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{Cu} \\
&= \ + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{Cu}
\end{aligned}
$$

进一步得到：

$$
^* - \ = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{Cu}
$$

根据方差定义，有：

$$
\begin{aligned}
var - cov(^*) &= \mathbf{E}\left((^* - \ )(^* - \ )'\right) \\
&= \mathbf{E}\left(\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{Cu}\right)\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{Cu}\right)'\right) \\
&= \ ^2(\mathbf{X}'\mathbf{X})^{-1} + \ ^2\mathbf{CC}' \\
&= var - cov(\hat{\ }) + \ ^2\mathbf{CC}'
\end{aligned}
$$

其中，我们可以证明 $^2\mathbf{CC}'$ 是**半正定矩阵**，矩阵对角线元素 ≥ 0，因此有：

$$
var - cov(^*) \geq var - cov(\hat{\ })
$$

从而表明 N-CLRM 假设下，OLS 方法估计得到的$\hat{\ }$，方差最小。 □

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

4.2.3 平方和分解与拟合优度的矩阵表达

对于多元回归模型：

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad \text{(PRM)} \\
Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i \quad \text{(SRM)} \\
\hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} \quad \text{(SRF)}
\end{aligned}
$$

通过对 Y_i 的变异及其来源的分解，可以得到：

$$(Y_i - \bar{Y}_i) = (\hat{Y}_i - \bar{Y}_i) + (Y_i - \bar{Y}_i) \tag{4.10}$$

$$y_i = \hat{y}_i + e_i \tag{4.11}$$

$$\sum y_i^2 = \sum \hat{y}_i^{\,2} + \sum e_i^2 \tag{4.12}$$

$$TSS = ESS + RSS \tag{4.13}$$

其中 TSS 表示总离差平方和，ESS 表示回归平法和，RSS 表示残差平方和。它们分别可以用矩阵表达为：

$$TSS = \mathbf{y}'\mathbf{y} - n\bar{Y}^2 \tag{4.14}$$

$$RSS = \mathbf{e}\mathbf{e}' = \mathbf{y}\mathbf{y}' - \hat{}'\mathbf{X}'\mathbf{y} \tag{4.15}$$

$$ESS = \hat{}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2 \tag{4.16}$$

进一步地，可以得到方差分析表（ANOVA）：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 4.1: k 变量线性回归模型的方差分析表（ANOVA）

平方和	代数公式	矩阵公式	自由度 df	均方和 MSS
总平方和 TSS	$\sum y_i^2$	$\mathbf{y}'\mathbf{y} - n\bar{Y}^2$	$df_{TSS} = n-1$	$MSS_{TSS} = TSS/df_{TSS}$
残差平方和 RSS	$\sum e_i^2$	$\mathbf{y}\mathbf{y}' - \hat{}'\mathbf{X}'\mathbf{y}$	$df_{RSS} = n-k$	$MSS_{RSS} = RSS/df_{RSS}$
回归平法和 ESS	$\sum \hat{y}_i^2$	$\hat{}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2$	$df_{ESS} = k-1$	$MSS_{ESS} = TSS/df_{ESS}$

根据拟合优度的定义定义 2.10 ，判定系数 R^2 的矩阵计算公式为：

$$\begin{aligned} R^2 &= \frac{ESS}{TSS} \\ &= \frac{\hat{}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2} \end{aligned} \tag{4.17}$$

4.2.4 回归系数显著性检验（t 检验）的矩阵方法实现

根据回归系数显著性检验的定义定义 2.11 ，利用矩阵方法实现 t 检验的过程如下：

对于多元回归模型

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{4.18}$$

$$\mathbf{y} = \mathbf{X} \ + \mathbf{u} \quad \text{(PRM)} \tag{4.19}$$

$$\mathbf{y} = \mathbf{X} \hat{} + \mathbf{e} \quad \text{(SRM)} \tag{4.20}$$

在 N-CLRM 假设下，采用 OLS 估计方法，可以证明：

$$\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \tag{4.21}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$\hat{} \sim N\left(, \sigma^2 \mathbf{X}'\mathbf{X}^{-1} \right) \tag{4.22}$$

从而可以构造 t 统计量

$$\mathbf{t} = \frac{\hat{} - }{\mathbf{S}} \sim \mathbf{t}(\mathbf{n}-\mathbf{k}) \tag{4.23}$$

对于总体回归模型式 4.18 的任一参数 $_{\mathbf{j}}, j \in (1, 2, \cdots, k)$ 提出假设：

$$_{\mathbf{j}} : \begin{cases} H_0 : _{\mathbf{j}} = 0 \\ H_1 : _{\mathbf{j}} \neq 0 \end{cases}$$

根据原假设 H_0，可以得到：

$$\mathbf{t}^* = \frac{\hat{}}{\sqrt{\mathbf{S_{ij}^2}(\hat{}_{\mathbf{kk}})}} \tag{4.24}$$

其中 $\mathbf{S_{ij}^2}(\hat{}_{\mathbf{kk}})$ 表示，由$\hat{}$ 的**样本**方差和协方差矩阵 $S_{ij}^2(\hat{})$ 的对角线元素组成的列向量，即

$$S_{ij}^2(\hat{\beta}_{kk}) = [s^2_{\hat{\beta}_1}, s^2_{\hat{\beta}_2}, \cdots, s^2_{\hat{\beta}_k}]'$$

若给定显著性水平 α 和自由度 $(n-k)$，很快可以得到 t 分布的查表 t 值，也即 $t_{(1-\alpha/2)}(n-k)$。然后比较样本 t 统计量 $\mathbf{t}^*$ 与理论 t 分布查的表 t 值 $(t_{(1-\alpha/2)}(n-2))$ 的关系。根据如下法则做出参数 β_2 的显著性检验结论：

- 如果列向量 $\mathbf{t}^*$ 的第 k 个元素 $t^*_{\hat{\beta}_k} > t_{(1-\alpha/2)}(n-2)$，则表明参数 β_k 的 t 检验在 α 水平下是**显著**的，也即显著地拒绝 $H_0 : \beta_k = 0$，从而接受 $H_1 : \beta_k \neq 0$。
- 如果列向量 $\mathbf{t}^*$ 的第 k 个元素 $t^*_{\hat{\beta}_k} \leq t_{(1-\alpha/2)}(n-2)$，则表明参数 β_k 的 t 检验在 α 水平下是**不显著**的，也即不能显著地拒绝 $H_0 : \beta_k = 0$，从而只能暂时接受 $H_0 : \beta_2 = 0$。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

4.2.4.1 模型整体显著性检验（F 检验）的矩阵方法实现

对于多元回归模型

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad \text{(U-PRM)} \tag{4.25}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i \quad \text{(U-SRM)} \tag{4.26}$$

$$\mathbf{y} = \mathbf{X} \ + \mathbf{u} \quad \text{(PRM)} \tag{4.27}$$

$$\mathbf{y} = \mathbf{X} \ + \mathbf{e} \quad \text{(SRM)} \tag{4.28}$$

我们称总体回归模型式 4.25 和对应的样本回归模型式 4.26 为为**无约束模型**（unrestricted model）。

对于总体回归模型式 4.25 的斜率参数 $_{\mathbf{j}}, j \in (2, \cdots, k)$ 提出如下**联合假设**（joint hypothesis）：

$$_{\mathbf{j}} : \begin{cases} H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0 \\ H_1 : \beta_j \text{ not all 0, for } j \in (2, \cdots, k) \end{cases}$$

在原假设 $H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0$ 下，我们可以得到如下模型：

$$Y_i = \beta_1 + u_i \quad \text{(R-PRM)} \tag{4.29}$$

$$Y_i = \hat{\beta}_1 + e_i \quad \text{(R-SRM)} \tag{4.30}$$

此时，我们称总体回归模型式 4.29 和对应的样本回归模型式 4.30 为**受约束模型**（restricted model）。

在备择假设 $H_1 : \beta_j$ 不全为 0，$j \in (2, \cdots, k)$ 下，我们可以得到该假设下的一种特殊回归模型[1]（如 $\beta_j \neq 0, j \in (2, \cdots, k)$），也即无约束总体回归模型式 4.25 和无约束样本回归模型式 4.26 。

[1]实际上在假设 $\{H_1 : \beta_j$ 不全为 0，$j \in (2, \cdots, k)\}$ 下，回归模型的形式有很多种，但对于这里所要说明的 F 检验而言，这些不同的模型形式已经不再重要

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

定义 4.1 (受约束模型)**.** 一般也称为参数约束回归模型（restricted model），是指总体参数满足某种约束条件的一类回归模型。

定义 4.2 (无约束模型)**.** 一般也称为参数无约束回归模型（unrestricted model），是指总体参数没有被指定满足某种约束条件的一类回归模型。

根据回归系数显著性检验的定义，利用矩阵方法实现 F 检验的过程如下：

在 N-CLRM 假设下，采用 OLS 估计方法，容易证明：

对于无约束总体回归模型式 4.25 有

$$\begin{aligned} u_i &\sim i.i.d\ N(0,\sigma^2) \\ Y_i &\sim i.i.d\ N(\beta_1+\beta_2X_i+\cdots+\beta_kX_i,\sigma^2) \\ RSS_U &= \sum (Y_i-\hat{Y}_i)^2 \sim \chi^2(n-k) \end{aligned} \tag{4.31}$$

对于受约束总体回归模型式 4.29 有

$$\begin{aligned} u_i &\sim i.i.d\ N(0,\sigma^2) \\ Y_i &\sim i.i.d\ N(\beta_1,\sigma^2) \\ RSS_R &= \sum (Y_i-\hat{Y}_i)^2 \sim \chi^2(n-1) \end{aligned} \tag{4.32}$$

然后我们可以构造得到一个 F 统计量:

$$\begin{aligned} F^* &= \frac{(RSS_R-RSS_U)/(k-1)}{RSS_U/(n-k)} \\ &= \frac{ESS_U/df_{ESS_U}}{RSS_U/df_{RSS_U}} \\ &\sim F(df_{ESS_U},df_{RSS_U}) \end{aligned} \tag{4.33}$$

若给定显著性水平 α 和样本数 (n)，很快可以得到 F 分布的查表 F 值，也即 $F_{(1-\alpha)}(k-1,n-k)$，然后比较其与样本 F 统计量 F^* 的关系。

根据如下法则做出总体回归模型整体显著性检验结论：

- 如果 $F^* > F_{(1-\alpha)}(k-1, n-k)$，则表明总体回归模型的 F 检验在 α 水平下是**显著**的，也即显著地拒绝 $H_0: \beta_2 = \beta_3 = \cdots = \beta_k = 0$，从而接受 $H_1: \beta_j$ 不全为 0，$j \in (2, \cdots, k)$，认为模型式 4.25 整体统计上是有意义的！
- 如果 $F^* \leq F_{(1-\alpha)}(k-1, n-k)$，则表明总体回归模型的 F 检验在 α 水平下是**不显著**的，也即不能显著地拒绝 $H_0: \beta_2 = \beta_3 = \cdots = \beta_k = 0$，从而只能暂时接受 $H_0: \beta_2 = 0$，认为模型式 4.25 整体在统计上是无意义的！

具体第，计算 F^* 的矩阵公式为

$$F^* = \frac{ESS_U/df_{ESS_U}}{RSS_U/df_{RSS_U}} = \frac{\left(\hat{\beta}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2\right)/(k-1)}{\left(\mathbf{y}\mathbf{y}' - \hat{\beta}'\mathbf{X}'\mathbf{y}\right)/(n-k)}$$

此外，我们还可以通过拟合优度 R^2，计算得到 F^*

$$F^* = \frac{ESS_U/df_{ESS_U}}{RSS_U/df_{RSS_U}} = \frac{R_U^2/(k-1)}{\left(1-R_U^2\right)/(n-k)} \tag{4.34}$$

4.2.5 样本外预测的矩阵方法实现

根据一元线性回归样本外预测（节 @ref(out-sample-forecast)）的知识内容，下面将用矩阵方法实现样本外均值预测 $\mathbf{E}(\mathbf{Y_0}|\mathbf{X_0})$ 和样本外个值预测 $(\mathbf{Y_0}|\mathbf{X_0})$。其中，给定样本外数据 $\mathbf{X_0} = [1, X_{20}, X_{30}, \cdots, X_{k0}]'$（列向量）。

对于多元回归模型

$$\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \\
Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i \\
\mathbf{y} &= \mathbf{X} + \mathbf{u} \\
\mathbf{y} &= \mathbf{X}\hat{} + \mathbf{e} \\
\hat{\mathbf{y}} &= \mathbf{X}\hat{}
\end{aligned} \tag{4.35}$$

对于样本外均值预测 $\mathbf{E}(\mathbf{Y_0}|\mathbf{X_0})$，矩阵实现步骤如下：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{aligned}
E(\hat{Y}_0) &= E(\mathbf{X_0}\hat{\ }) = \mathbf{X_0}\ = E(\mathbf{Y_0}) \\
var(\hat{Y}_0) &= E(\mathbf{X_0}\hat{\ } - \mathbf{X_0}\ )^2 \\
&= E(\mathbf{X_0}(\hat{\ } - \ )(\hat{\ } - \ )'\mathbf{X_0'}) \\
&= E\mathbf{X_0}\left((\hat{\ } - \ )(\hat{\ } - \ )'\right)\mathbf{X_0'} \\
&= \sigma^2\mathbf{X_0}\left(\mathbf{X'X}\right)^{-\mathbf{1}}\mathbf{X_0'}
\end{aligned}
$$

$$
S^2(\hat{Y}_0) = \hat{\sigma}^2\mathbf{X_0}(\mathbf{X'X})^{-\mathbf{1}}\mathbf{X_0'}
$$

因此 $\hat{\mathbf{Y}}_\mathbf{0}$ 服从如下正态分布：

$$
\begin{aligned}
&\hat{Y}_0 \sim N(\mu_{\hat{Y}_0}, \sigma^2_{\hat{Y}_0}) \\
&\hat{Y}_0 \sim N\left(E(Y_0|X_0), \sigma^2\mathbf{X_0}(\mathbf{X'X})^{-\mathbf{1}}\mathbf{X_0'}\right)
\end{aligned}
\tag{4.36}
$$

因此可以构造 t 统计量：

$$
t_{\hat{Y}_0} = \frac{\hat{Y}_0 - E(Y|X_0)}{S_{\hat{Y}_0}} \quad \sim t(n-k) \tag{4.37}
$$

其中：

$$
\mathbf{S_{\hat{Y}_o}} = \sqrt{\hat{\sigma}^2 X_0 (X'X)^{-1} X_0'} \tag{4.38}
$$

$$
\hat{\sigma}^2 = \frac{\mathbf{ee'}}{(n-k)} \tag{4.39}
$$

给定显著性水平 α 的情况下，可以查表得到理论 t 值 $t_{1-\alpha/2}(n-k)$，从而可以计算得到均值预测的置信区间：

$$
\hat{Y}_0 - t_{1-\alpha/2}(n-2)\cdot S_{\hat{Y}_0} \le E(Y|X_0) \le \hat{Y}_0 + t_{1-\alpha/2}(n-2)\cdot S_{\hat{Y}_0} \tag{4.40}
$$

对于多元线性回归模型式 4.35 ，样本外个值预测 $(\mathbf{Y_0}|\mathbf{X_0})$ 的矩阵实现步骤如下：

因为有

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$e_0 = Y_0 - \hat{Y}_0 \tag{4.41}$$

所以 e_0 的期望为：

$$\begin{aligned} E(e_0) &= E(Y_0 - \hat{Y}_0) \\ &= E(\mathbf{X_0} + u_0 - \mathbf{X_0}\hat{} \,) \\ &= E\left(u_0 - \mathbf{X_0}\hat{} \, - \,\right) \\ &= E\left(u_0 - \mathbf{X_0}(\mathbf{X'X})^{-1}\mathbf{X'u}\right) \\ &= 0 \end{aligned}$$

同时，e_0 的方差为：

$$\begin{aligned} var(e_0) &= E(Y_0 - \hat{Y}_0)^2 \\ &= E(e_0^2) \\ &= E\left(u_0 - \mathbf{X_0}(\mathbf{X'X})^{-1}\mathbf{X'u}\right)^2 \\ &= \sigma^2\left(1 + \mathbf{X_0}(\mathbf{X'X})^{-1}\mathbf{X_0'}\right) \end{aligned}$$

进一步地，e_0 服从如下正态分布：

$$\begin{aligned} e_0 &\sim N(\mu_{e_0}, \sigma_{e_0}^2) \\ e_0 &\sim N\left(0, \sigma^2\left(1 + \mathbf{X_0}(\mathbf{X'X})^{-1}\mathbf{X_0'}\right)\right) \end{aligned} \tag{4.42}$$

因此可以构造 t 统计量：

$$t_{e_0} = \frac{\hat{Y}_0 - Y_0}{S_{e_0}} \sim t(n-k) \tag{4.43}$$

其中：

$$S_{Y_0 - \hat{Y}_0} = S_{e_0} = \sqrt{\hat{\sigma}^2\left(1 + X_0(X'X)^{-1}X_0'\right)} \tag{4.44}$$

$$\hat{\sigma}^2 = \frac{\mathbf{ee'}}{(n-k)} \tag{4.45}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

给定显著性水平 α 的情况下，可以查表得到理论 t 值 $t_{1-\alpha/2}(n-k)$，从而可以计算得到均值预测的置信区间：

$$\hat{Y}_0 - t_{1-\alpha/2}(n-2)\cdot S_{Y_0-\hat{Y}_0} \leq (Y_0|X_0) \leq \hat{Y}_0 + t_{1-\alpha/2}(n-2)\cdot S_{Y_0-\hat{Y}_0} \tag{4.46}$$

4.3 实验内容

在 Eviews 中运用**矩阵方法**，计算如下步骤:

- 计算直线回归方程的回归系数向量 $(\hat{\ })$，并写出样本回归模型（SRM）。
- 计算回归误差方差 $(\hat{\sigma}^2)$ 和回归误差标准差 $(\hat{\sigma})$。
- 计算回归系数的样本方差协方差矩阵 $(\widehat{var}\_\widehat{cov})$。
- 得出回归系数的样本标准差向量 $(S_{\hat{\beta}})$。
- 进行平方和分解，计算 TSS、ESS 和 RSS。
- 计算判定系数 R^2，调整判定系数 $(\hat{R}^2)$。
- 计算样本 t 统计量 $(\mathbf{t}^*)$，并进行 t 假设检验。
- 对回归方程的进行样本外均值预测 $E(Y \mid X = X_0)$
- 对回归方程的进行样本外个值预测 $(Y_0 \mid X = X_0)$

4.4 实验准备

4.4.1 实验软件

本次实验需要提前准备好如下软件：

- 统计分析软件 Eviews 9.0 版本及以上
- 公式编辑软件 Mathtype 6.0 版本及以上
- 写作编辑软件 Office Word/Excel 2010 版本及以上

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 浏览器软件 chrome 66.0 版本及以上或 360 极速浏览器 9.5 版本及以上

4.4.2 实验材料

玫瑰的需求：表 4.2 给出美国底特律市区对玫瑰的季度需求数据。

表 4.2: 玫瑰的需求 (n=16)

YEAR	Q	X2	X3	X4	X5
1971	11484	2.3	3.5	158	1
1971	9348	2.5	2.8	173	2
1972	8429	3.1	4.1	165	3
1972	10079	2.9	3.6	173	4
1972	9240	2.7	3.2	178	5
1972	8862	2.8	3.7	199	6
1973	6216	3.6	3.8	186	7
1973	8253	3.2	3.5	189	8
1973	8038	2.6	3.1	180	9
1973	7476	2.9	3.2	183	10
1974	5911	3.8	3.6	182	11
1974	7950	3.6	3.6	185	12
1974	6134	2.8	2.9	184	13
1974	5868	3.0	3.1	188	14
1975	3160	4.2	3.6	176	15
1975	5872	3.7	3.5	188	16

变量说明见表 4.3 ：

表 4.3: 变量定义及说明

variable	label
YEAR	年份. 季度
Q	玫瑰销售量 (打)
X2	玫瑰批发价格 (/)||$X3$| (/打)
X4	家庭可支配收入 ($/周)

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 4.3: 变量定义及说明

variable	label
X5	时间趋势

请考虑如下两个需求函数：

$$Y_t = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_5 X_{5t} + e_{1t} \tag{4.47}$$

$$ln(Y_t) = \hat{\beta}_1 + \hat{\beta}_2 ln(X_{2t}) + \hat{\beta}_3 ln(X_{3t}) + \hat{\beta}_4 ln(X_{4t}) + \hat{\beta}_5 X_{5t} + e_{2t} \tag{4.48}$$

4.4.3 实验规则

本次实验将利用矩阵方法进行计算，我们对计算过程中的 Eviews 对象做如下命名约定：

表 4.4: 计算对象、表达式及 Eviews 命名

name_chn	cat_eng	math	name_eviews
序列 Y	series	Y	q
组 X	group	X	xg
矩阵 **y**	matrix	$\mathbf{y}$	y
矩阵 **X**	matrix	$\mathbf{X}$	x
矩阵 $(\mathbf{X}'\mathbf{X})$	matrix	$(\mathbf{X}'\mathbf{X})$	xtx
矩阵 $(\mathbf{X}'\mathbf{X})^{-1}$	matrix	$(\mathbf{X}'\mathbf{X})^{-1}$	xtxi
矩阵 $\mathbf{X}'\mathbf{y}$	matrix	$\mathbf{X}'\mathbf{y}$	xty
矩阵^	matrix	$\hat{}$	beta_hat
回归误差方差	scalar	$\hat{\sigma}^2$	sigma2_hat
回归误差标准差	scalar	$\hat{\sigma}$	sigma_hat
$\hat{\beta}$ 样本方差协方差矩阵	matrix	$\text{var}-\text{cov}(\hat{})$	s2_varcov_beta_hat
$\hat{\beta}$ 样本方差矩阵	matrix	$\mathbf{S^2}$	s2_beta_hat
$\hat{\beta}$ 样本标准差矩阵	matrix	$\mathbf{S}$	s_beta_hat
均值修正值	scalar	$n\bar{Y}^2$	mean_adj
总平方和	scalar	TSS	tss

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 4.4: 计算对象、表达式及 Eviews 命名

name_chn	cat_eng	math	name_eviews
残差平方和	scalar	RSS	rss
回归平方和	scalar	ESS	ess
判定系数	scalar	R^2	r2
调整判定系数	scalar	$\bar{R}^2$	r2_adj
矩阵 t 统计量	matrix	$\mathbf{t}^*$	t_str_beta_hat
理论 t 值	scalar	$t_{1-\alpha/2}(n-k)$	t_value
F 统计量	scalar	F^*	f_str
理论 F 值	scalar	$F_{1-\alpha}(k-1,n-k)$	f_value
样本外 X0	matrix	X_0	x0
样本外回归值 $\hat{Y}_0$	scalar	$\hat{Y}_0$	y0_hat
均值预测	scalar	$E(Y \mid X=X_0)$	forecast_exp
$\hat{Y}_0$ 的样本标准差	scalar	$S_{\hat{Y}_0}$	s_y0h
均值区间预测的左界	scalar	$E(Y \mid X=X_0)_L$	y_exp_lft
均值区间预测的右界	scalar	$E(Y \mid X=X_0)_R$	y_exp_rht
个值预测	scalar	$(Y_0 \mid X=X_0)$	forecast_ind
$(\hat{Y}_0-Y_0)$ 的样本标准差	scalar	$S_{(\hat{Y}_0-Y_0)}$	s_y0h_mns_y0
个值区间预测的左界	scalar	$(Y_0 \mid X=X_0)_L$	y_ind_lft
个值区间预测的右界	scalar	$(Y_0 \mid X=X_0)_R$	y_ind_rht

4.5 主要实验步骤——以对数模型为例 4.58

4.5.1 新建工作文件并导入数据

- Eviews 操作目标：构建工作文件，成功导入数据
- Eviews 操作思路：利用 EViews 代码创建工作文件并导入数据。

在命令视窗中依次输入并运行如下 EViews 代码：

```
'创建工作文件（工作文件名=rose，子页命名=sale），无结构无日期，样本数为16
wfcreate(wf=rose,page=sale)  u 16
```

```
'导入外部数据，路径为d:\github\books\data\Lab3-family-spends.xlsx
import d:\github\books\data\lab4-rose-demand-origin.xlsx
```

在工作文件视窗下，可以看到创建的工作文件和导入的数据，可以双击查看（见图 4.1 ）：

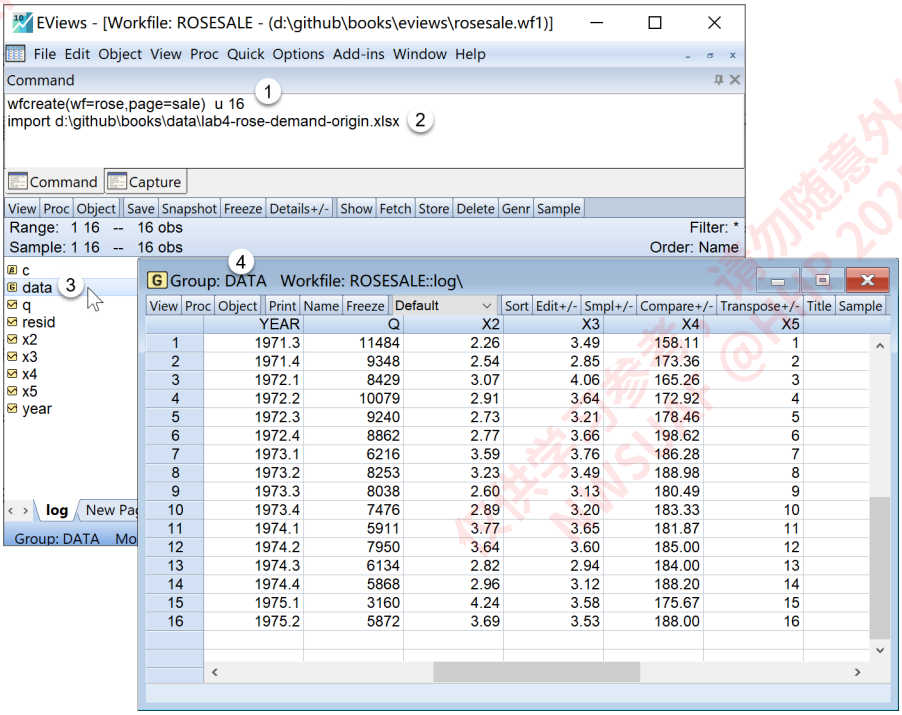


图 4.1: 创建工作文件并导入数据

提示

上述过程也可以通过菜单操作实现：

（1）创建工作文件：

（a）EViews 主窗口上依次点击 ⇒ `File`⇒ `New`⇒ `Workfile`

（b）进行 workfile create 引导设置：

- workfile structure type: `unstructured/undatede`
- data range：
- workfile names(optional):
  - WF: `rose`

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

– Page: `sale_log` （建议命名 **sale_log**）

（2）导入外部数据：

- EViews 主窗口上依次点击 ⇒ `File` ⇒ `Import` ⇒ `Import from file ...`

4.5.2 进行对数模型的 Eviews 回归分析

- Eviews 操作目标：得到回归方程，查看回归结果
- Eviews 操作思路：构建回归方程对象。回归模型为：

$$\begin{aligned} log(Q)_i = &+ \beta_1 + \beta_2 log(X2)_i \\ &+ \beta_3 log(X3)_i + \beta_4 log(X4)_i \\ &+ \beta_5 X5_i + u_i \end{aligned} \tag{4.49}$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
'生成线性回归模型的方程对象
equation eq_log.ls log(q) c log(x2) log(x3) log(x4) x5    '对数模型
```

在工作文件视窗下，可以看到如下新生成的方程对象，可以双击查看eq_log（见图 4.2 ）：

提示

上述过程也可以通过菜单操作实现：
（1）进入方程估计的引导界面。

- EViews 主窗口上依次点击点 `Quick` ⇒ `Estimation Equation`

（2）完成方程估计的引导设置。

- 设置方程。`Equation Estimation` ⇒ `Specification` ⇒ `Equation specification` 中依次输入变量 `log(q) c log(x2) log(x3) log(x4) x5`（注意变量之间的空格，以及截距 `c`）
- 选择估计方程。`Estimation settings` 中的 `Method` 下拉框 ⇒ 下拉选择 `LS - Least Squares (NLS and ARMA)`

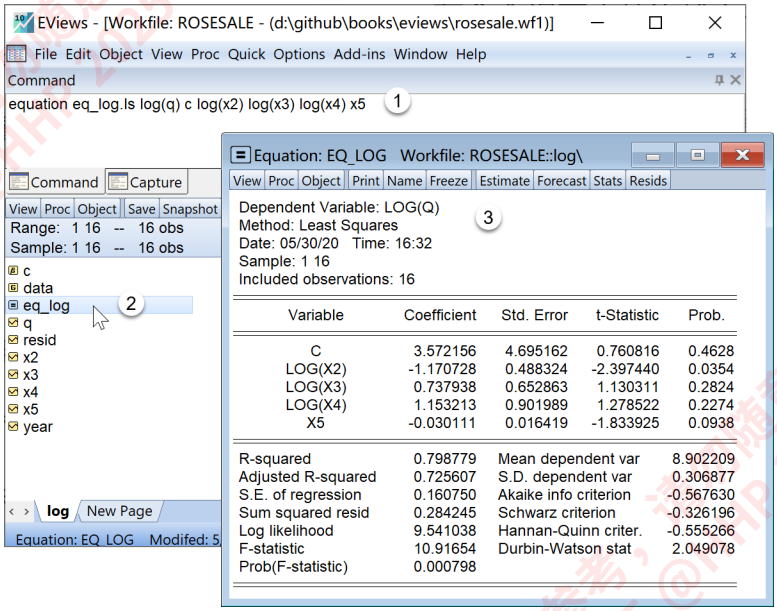


图 4.2: 生成回归方程对象

- 完成设置，点击 OK

（3）命名并保存回归方程。

- 在未命名的方程对象UNTITLED 视窗下，点击菜单栏 Name ⇒ 输入命名 eq_log（建议命名）⇒ 完成命名，点击 Ok

回归方程结果见图 4.3 ：

玫瑰销售的对数模型的简要回归报告如下：

$$\begin{aligned}\widehat{log(Q)}_i = &+ \hat{\beta}_1 + \hat{\beta}_2 log(X2)_i \\ &+ \hat{\beta}_3 log(X3)_i + \hat{\beta}_4 log(X4)_i \\ &+ \hat{\beta}_5 X5_i\end{aligned} \tag{4.50}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

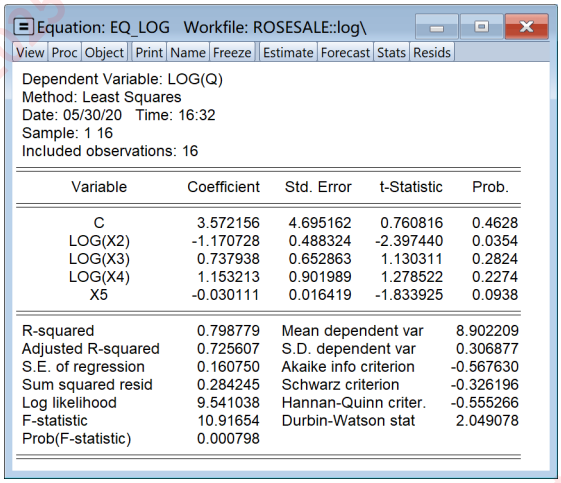


图 4.3: 玫瑰销售的对数模型回归结果

$$
\begin{array}{llll}
\widehat{log(Q)} = & +3.57 & -1.17ln(X2)_i & +0.74ln(X3)_i \\
(s) & (4.6952) & (0.4883) & (0.6529) \\
(t) & (+0.76) & (-2.40) & (+1.13) \\
(cont.) & +1.15ln(X4)_i & -0.03X5_i & \\
(s) & (0.9020) & (0.0164) & \\
(t) & (+1.28) & (-1.83) & \\
(over) & n = 16 & \hat{\sigma} = 0.1607 & \\
(fit) & R^2 = 0.7988 & \bar{R}^2 = 0.7256 & \\
(Ftest) & F^* = 10.92 & p = 0.0008 &
\end{array} \tag{4.51}
$$

4.5.3 构建几个重要变量对象

- Eviews 操作目标：构造几个重要 EViews 对象，便于后面分析使用
- Eviews 操作思路：样本数 n，构造常数序列，构造组对象

在命令视窗中依次输入并运行如下 EViews 代码：

```
'构造几个重要变量对象
scalar n=@obs(x2)                    '样本数n(标量)
```

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
series cst=1                          '新建元素全为1的序列对象（用于构造矩阵X）
group xg cst log(x2) log(x3) log(x4) x5     '构造为group，便于观察
```

在工作文件视窗下，可以看到如下新生成的对象，可以双击查看（见图 4.4 ）：

- 样本数 n 的标量对象n
- 元素全为 1 的常数序列对象cst
- 组对象 xgxg

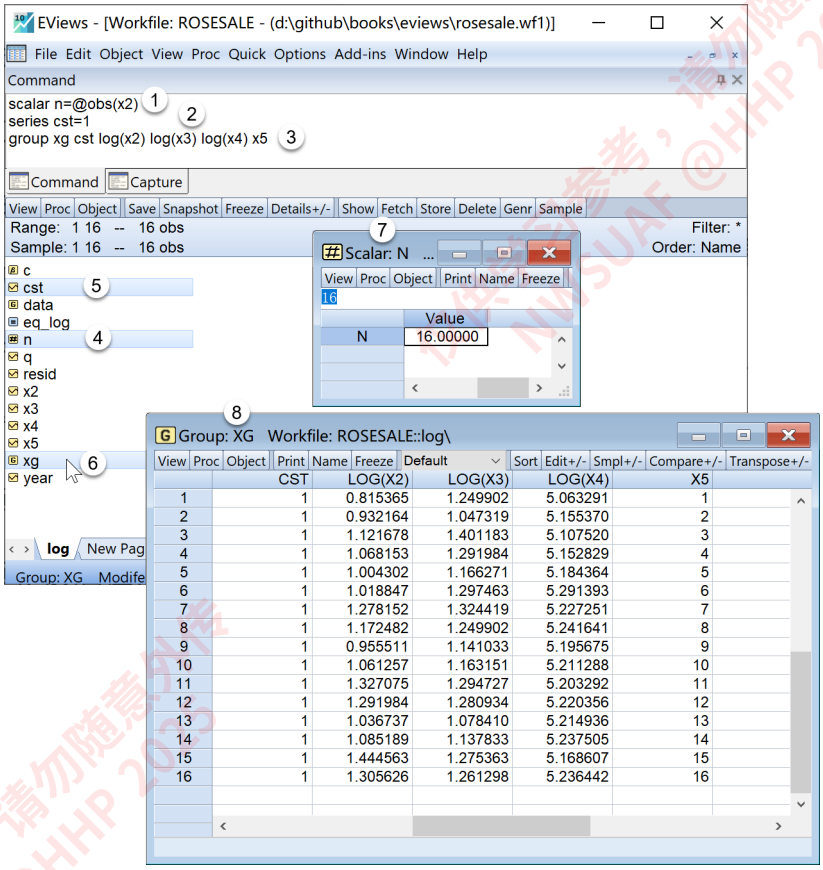


图 4.4: 构建几个重要变量对象

4.5.4 构造 X 矩阵和 Y 矩阵对象

- Eviews 操作目标：构造 X 矩阵和 Y 矩阵对象，便于后面分析使用

- Eviews 操作思路：把组对象转换成 X 矩阵；利用 log() 函数构造 Y 矩阵

在命令视窗中依次输入并运行如下 EViews 代码：

```
'构造X矩阵和Y矩阵对象。
matrix x=xg                    '转换为X矩阵对象
matrix y=log(q)                '构造Y矩阵对象
```

在工作文件视窗下，可以看到如下新生成的对象，可以双击查看（见图 4.5 ）：

- 矩阵 **X** 的矩阵对象x
- 矩阵 **y** 的矩阵对象y

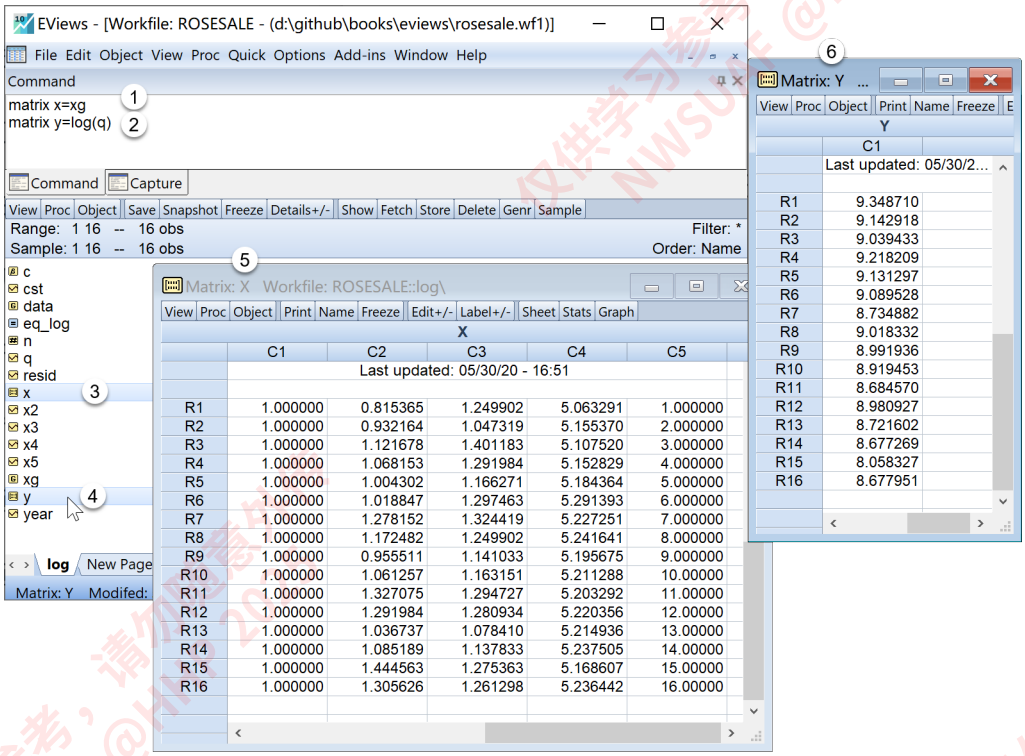


图 4.5: 构造 X 矩阵和 Y 矩阵对象

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

4.5.5 计算回归方程的回归系数向量

- Eviews 操作目标：计算得到回归方程的回归系数向量（包含 5 个系数估计值）。
- Eviews 操作思路：利用理论公式，先计算得到几个重要矩阵（后面分析还要用到），最后利用矩阵运算计算得出回归系数向量。理论计算公式为：

$$\hat{} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
'计算回归方程的回归系数向量
matrix xtx=@transpose(x)*x   '得到重要矩阵X'X
matrix xtxi=@inverse(xtx)    '得到重要矩阵(X'X)^(-1)
matrix xty=@transpose(x)*y   '得到重要矩阵X'y
matrix beta_hat=xtxi*xty     '得到回归系数矩阵
```

在工作文件视窗下，可以看到如下新生成的矩阵对象，可以双击查看（见图 4.6 ）：

- 重要矩阵 $\mathbf{X}'\mathbf{X}$ 的矩阵对象xtx
- 重要矩阵 $(\mathbf{X}'\mathbf{X})^{-1}$ 的矩阵对象xtxi
- 重要矩阵 $(\mathbf{X}'\mathbf{y})$ 的矩阵对象xty
- 回归系数向量$\hat{} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ 的矩阵对象beta_hat

4.5.6 计算回归方程的误差方差及标准差

- Eviews 操作目标：计算回归方程的误差方差及标准差（标量），与主回归结果进行核验。
- Eviews 操作思路：利用理论公式

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

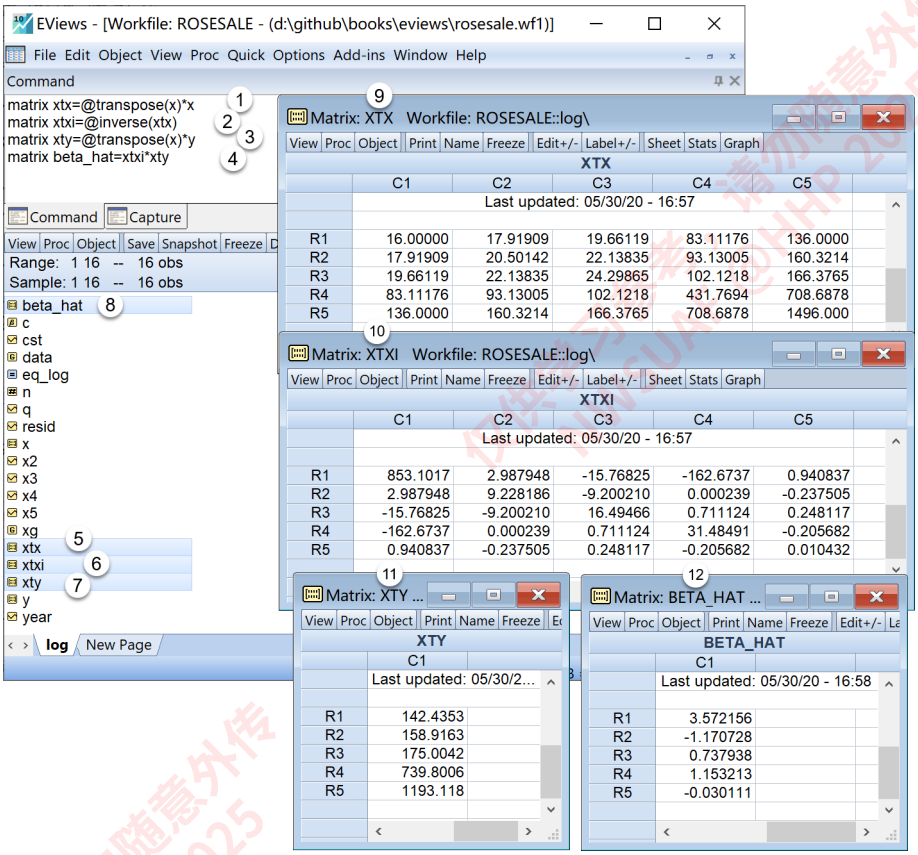


图 4.6: 计算回归系数向量

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

回归误差方差 ($\hat{\sigma}^2$) 和回归误差标准差 ($\hat{\sigma}$) 的理论计算公式分别为：

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k} = \frac{\mathbf{y}'\mathbf{y} - \hat{}'\mathbf{X}'\mathbf{y}}{n-k}$$
$$\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n-k}} = \sqrt{\frac{\mathbf{y}\mathbf{y}' - \hat{}'\mathbf{X}'\mathbf{y}}{n-k}}$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
'计算回归误差方差及标准差
scalar sgm_hat_sqr=1/(n-5)*(@transpose(y)*y-@transpose(beta_hat)*xty)    '回归误差方差
scalar sgm_hat=@sqr(sgm_hat_sqr)
```

在工作文件视窗下，可以看到如下的标量对象，可以双击查看（见图 4.7 ）：

- 回归误差方差 $\hat{\sigma}^2$ 的标量对象sgm_hat_sqr
- 回归误差标准差 $\hat{\sigma}$ 的标量对象sgm_hat

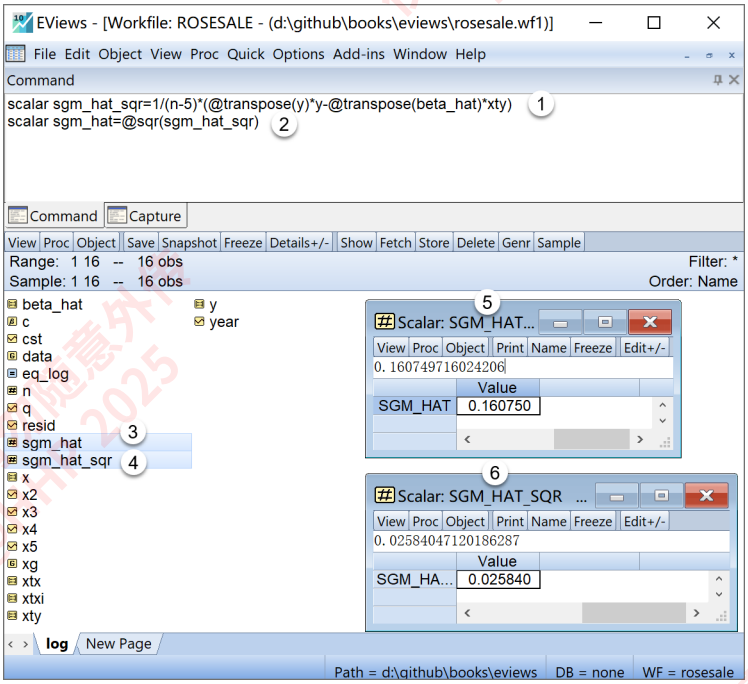


图 4.7: 计算回归方程的误差方差及标准差

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

4.5.7 计算回归系数的方差协方差矩阵、系数的样本方差和标准差（列向量）

- Eviews 操作目标：计算回归方程的方差协方差矩阵，得到系数的样本方差和标准差。
- Eviews 操作思路：先得到方差协方差矩阵，再提取对角线元素，利用理论公式

$$\widehat{var}\_\widehat{cov}(\hat{}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

$$\mathbf{S^2}$$

$$\mathbf{S} = \sqrt{S^2_{\hat{\beta}}}$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
'计算回归系数的方差协方差矩阵、系数的样本方差和标准差(列向量)
matrix s2_varcov_beta_hat=sgm_hat_sqr*xtxi        '回归系数的样本方差协方差矩阵
matrix s2_beta_hat=@getmaindiagonal(s2_varcov_beta_hat)  '回归系数的样本方差
matrix s_beta_hat=@sqr(s2_beta_hat)                      '回归系数的样本标准差
```

在工作文件视窗下，可以看到如下 EViews 对象，可以双击查看（见图 4.8 ）：

- 回归系数的样本方差协方差矩阵 $\widehat{var}\_\widehat{cov}(\hat{})$ 的矩阵对象s2_varcov_beta_hat
- 回归系数的样本方差 $\mathbf{S^2}$ 的矩阵对象s2_beta_hat
- 回归系数的样本标准差 $\mathbf{S}$ 的矩阵对象s_beta_hat

4.5.8 进行平方和分解，计算 TSS、ESS 和 RSS，以及各自的自由度（标量）

- Eviews 操作目标：进行平方和分解，得到方差分析表（ANOVA）。
- Eviews 操作思路：掌握自由度的计算，利用理论公式

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

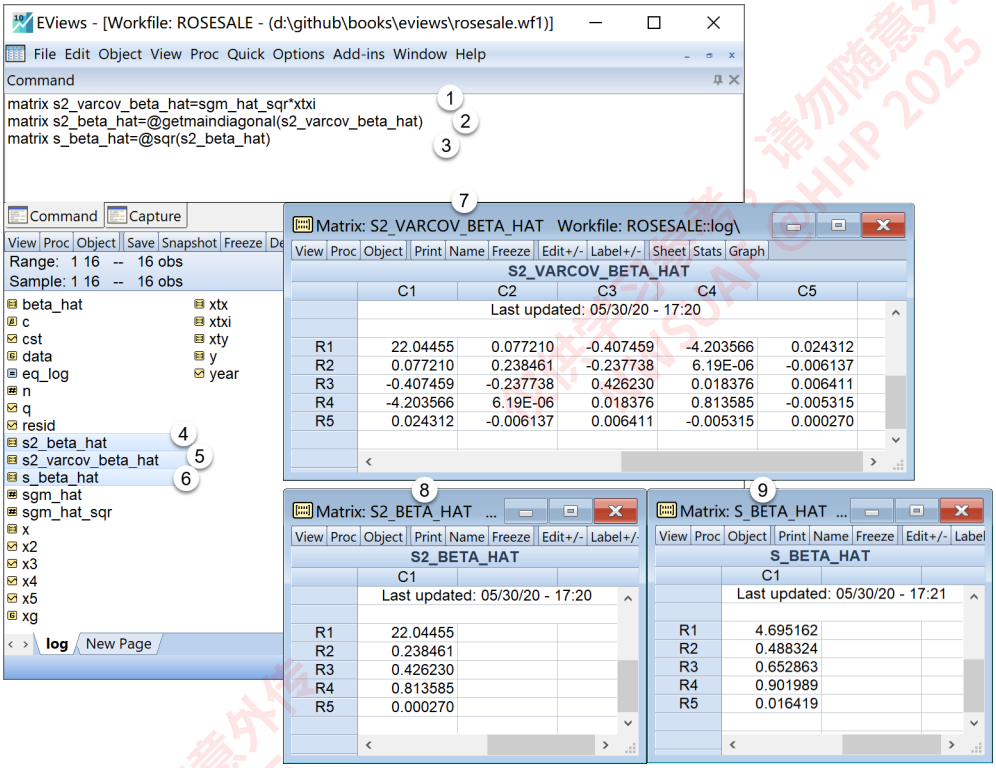


图 4.8: 计算回归系数的方差协方差矩阵、系数的样本方差和标准差

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{aligned}
& n\bar{Y}^2 \\
TSS &= \mathbf{y}'\mathbf{y} - n\bar{Y}^2 \\
RSS &= \mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{y} - \hat{}'\mathbf{X}'\mathbf{y} \\
ESS &= \hat{\mathbf{y}}'\hat{\mathbf{y}} = \hat{}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2 \\
df_{TSS} &= n-1 \\
df_{RSS} &= n-k \\
df_{ESS} &= k-1
\end{aligned}
$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
'进行平方和分解，计算TSS、ESS和RSS，以及各自的自由度（标量）
scalar mean_adj=16*(@mean(log(q)))^2            '均值修正值
scalar tss=@transpose(y)*y-mean_adj        '总平方和TSS
scalar rss=@transpose(y)*y-@transpose(beta_hat)*xty    '剩余平方和RSS
scalar ess=@transpose(beta_hat)*xty-mean_adj          '回归平方和ESS
scalar df_tss=n-1     ' TSS的自由度
scalar df_rss=n-5     'RSS的自由度
scalar df_ess=4       'ESS的自由度
```

在工作文件视窗下，可以看到如下的标量对象，可以双击查看（见图 4.9 ）：

- 总平方和 TSS 的标量对象tss
- 残差平方和 RSS 的标量对象rss
- 回归平方和 ESS 的标量对象ess
- 总平方和的自由度 df_{TSS} 的标量对象df_tss
- 残差平方和的自由度 df_{RSS} 的标量对象df_rss
- 回归平方和的自由度 df_{ESS} 的标量对象df_ess

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

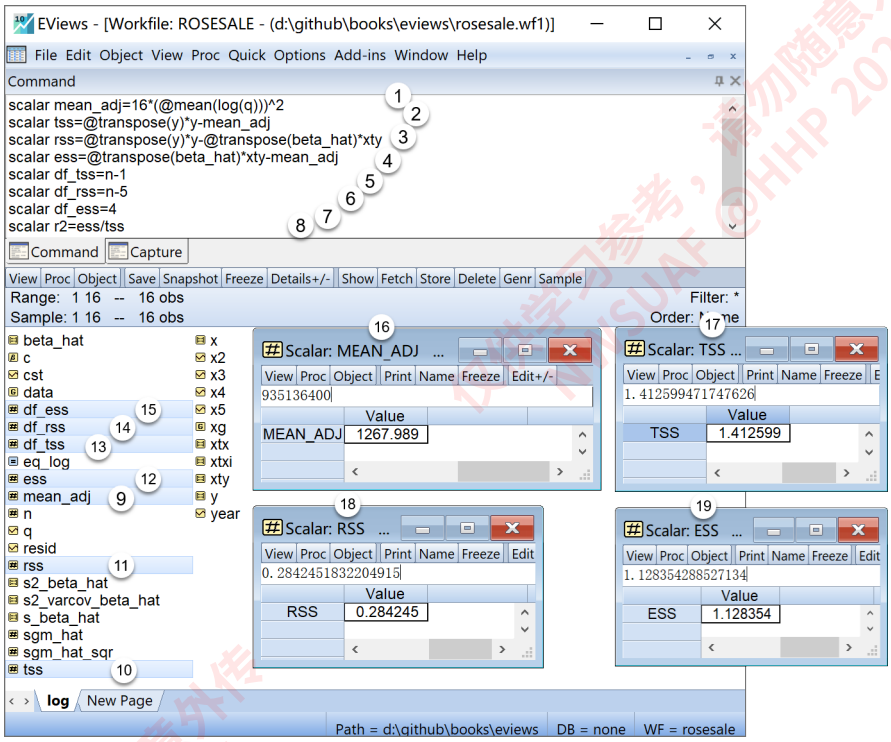


图 4.9: 对回归方程进行平方和分解

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

4.5.9 计算自变量的相关系数表格、回归方程的判定系数和调整判定系数

- Eviews 操作目标：得到自变量的相关系数表，计算方程的判定系数和调整判定系数。
- Eviews 操作思路：构建自变量组对象，得到相关系数表；利用方差分析表结果计算判定系数和调整判定系数。利用理论公式

$$
\begin{aligned}
R^2 &= \frac{ESS}{TSS} \\
&= \frac{\hat{}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2} \\
\bar{R}^2 &= 1 - \frac{RSS/f_{RSS}}{TSS/f_{TSS}} \\
&= 1 - \frac{\mathbf{y}'\mathbf{y} - \hat{} \mathbf{X}'\mathbf{y}/n-k}{(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)/n-1}
\end{aligned}
$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
'计算自变量的相关系数表格、回归方程的判定系数和调整判定系数
group varx x2 x3 x4 x5                ' 构建只含X的group
freeze(tab_cor) varx.cor      '把group的相关系数矩阵表视图保存为表格
scalar r2=ess/tss                '回归方程的判定系数
scalar r2_adj=1-(rss/df_rss)/(tss/df_tss)       '回归方程的调整判定系数
```

在工作文件视窗下，可以看到如下的 EViews 对象，可以双击查看（见图 4.10 ）：

- 只含回归元变量（X2、X3、X4、X5）的组对象varx
- 回归元变量间相关系数的表格对象tab_cor
- 回归方程判定系数 R^2 的标量对象r2
- 回归方程调整判定系数 $\bar{R}^2$ 的标量对象r2_adj

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

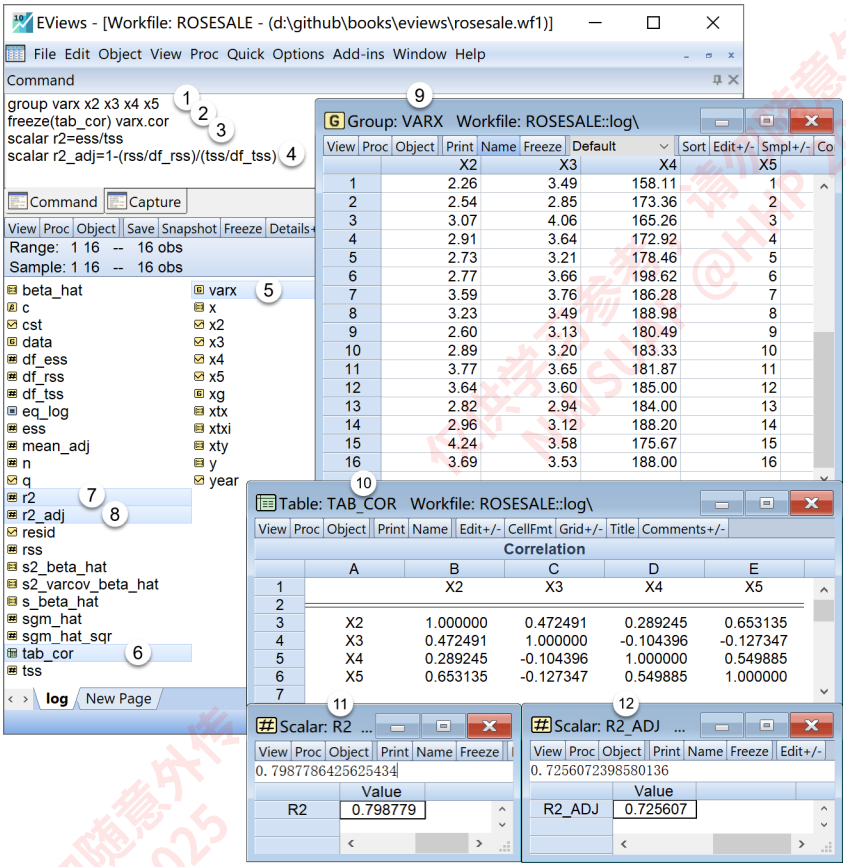


图 4.10: 相关系数表、回归方程的判定系数和调整判定系数

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

4.5.10 对回归方程的回归系数进行显著性 t 检验

- 目标：检验各个回归系数是否显著
- 思路：根据理论的矩阵公式，计算样本 t 统计量；计算给定显著性水平下的理论 t 值；进行 t 假设检验。

回归系数的样本 t 统计量以及给定显著性水平下的理论 t 值的计算公式为：

$$\mathbf{t}^* = \hat{\mathbf{S}}$$

$$t_{1-\alpha/2}(n-k) = t_{0.975}(11)$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
'对回归方程的回归系数进行显著性t检验
matrix t_str_beta_hat=@ediv(beta_hat,s_beta_hat)    '计算得到回归系数的样本t统计量
scalar t_value=@qtdist(0.975,df_rss)                '给定 =0.05 水平下的理论t值（右侧正值
```

在工作文件视窗下，可以看到如下的 EViews 对象，可以双击查看（见图 4.11 ）：

- 回归系数的样本 t 统计量 $\mathbf{t}^*$ 的矩阵对象t_str_beta_hat
- 给定 $\alpha = 0.05$ 水平下的理论 t 值（右侧正值）$t_{1-\alpha/2}(n-k)$ 的标量对象t_value

4.5.11 对回归方程的整体显著性进行 F 假设检验

- 目标：进行回归方程的整体显著性进行 F 假设检验
- 思路：根据理论的矩阵公式，计算样本 F 统计量；给定显著性水平下，计算得到理论 F 值；完成模型整体显著性 F 检验过程。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

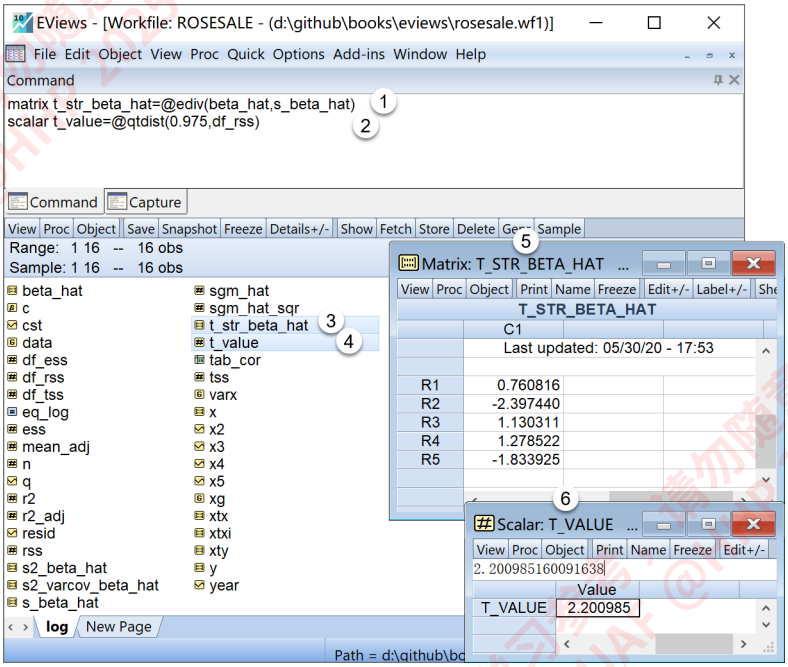


图 4.11: 对回归方程的回归系数进行显著性 t 检验

回归方程的样本 F 统计量，以及给定显著性水平下理论 F 值的计算公式为：

$$
\begin{aligned}
F^* &= \frac{ESS/f_{ESS}}{RSS/f_{RSS}} = \frac{MSS_{ESS}}{MSS_{RSS}} \\
&= \frac{(\hat{\mathbf{X}}'\mathbf{y} - n\bar{Y}^2)/k-1}{(\mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y})/n-k} \\
F_{1-\alpha}(k-1, n-k) &= F_{0.95}(4, 11)
\end{aligned}
$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 对回归方程的整体显著性进行F假设检验
scalar f_str=(ess/df_ess)/(rss/df_rss)                    '计算回归方程的样本F统计量
scalar f_value=@qfdist(0.95,df_ess,df_rss)      '计算给定 =0.05 水平下的查表的理论F值
```

在工作文件视窗下，可以看到如下的 EViews 对象，可以双击查看（见图 4.12 ）：

- 回归方程的样本 F 统计量 F^* 的标量对象f_str
- 给定 $\alpha = 0.05$ 水平下的理论 F 值（右侧大值）$F_{1-\alpha}(n-1, n-k)$ 的标量对象f_value

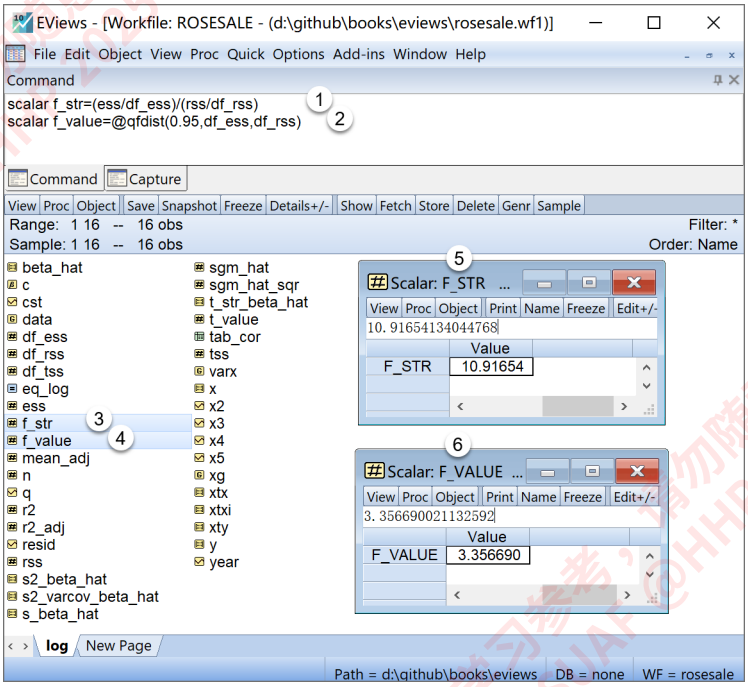


图 4.12: 对回归方程的整体显著性进行 F 假设检验

4.5.12 进行样本外的均值预测、个值预测，并计算置信区间

- 目标：样本外的均值预测、个值预测，计算得到置信区间（给定显著性水平）
- 思路：构建样本外 $\mathbf{X_0}$ 矩阵；计算得到回归估计值 $\hat{Y}_0$；构造分别计算得到均值预测的 t 分布样本标准差 $\mathbf{S}_{\hat{\mathbf{Y}}_\mathbf{0}}$ 和个值预测的 t 分布样本标准差 $\mathbf{S}_{\mathbf{Y_0}-\hat{\mathbf{Y}}_\mathbf{0}}$；给定显著性水平下，计算得到理论 t 值；利用公式分别计算均值预测和个值预测的置信区间。

相关参考计算公式为：

$$\hat{\mathbf{Y}}_\mathbf{0} = \mathbf{X}_\mathbf{0}\hat{}$$

$$\mathbf{S}_{\hat{\mathbf{Y}}_\mathbf{0}} = \sqrt{\hat{\sigma}^2 X_0 (X'X)^{-1} X_0'} \tag{4.52}$$

$$S_{Y_0-\hat{Y}_0} = S_{e_0} = \sqrt{\hat{\sigma}^2 \left(1 + X_0 (X'X)^{-1} X_0'\right)} \tag{4.53}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$\hat{\sigma}^2 = \frac{\mathbf{e}\mathbf{e}'}{(n-k)} \tag{4.54}$$

$$\hat{Y}_0 - t_{1-\alpha/2}(n-2)\cdot S_{\hat{Y}_0} \le E(Y|X_0) \le \hat{Y}_0 + t_{1-\alpha/2}(n-2)\cdot S_{\hat{Y}_0} \tag{4.55}$$

$$\hat{Y}_0 - t_{1-\alpha/2}(n-2)\cdot S_{Y_0-\hat{Y}_0} \le (Y_0|X_0) \le \hat{Y}_0 + t_{1-\alpha/2}(n-2)\cdot S_{Y_0-\hat{Y}_0} \tag{4.56}$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
' 进行样本外的均值预测、个值预测，并计算置信区间
matrix(1,5) x0                                          '产生1行*5列的空矩阵
x0.fill(b=r) 1,log(20),log(4),log(4),200                '样本外X值为(x2=log(20), x3=log(4), x4=log(4), x5
scalar y0_hat=x0*beta_hat                                       '样本外估计Y值
scalar s_y0h=@sqr(sgm_hat_sqr*x0*xtxi*@transpose(x0))           '均值预测的样本标准差
scalar s_y0h_mns_y0=@sqr(sgm_hat_sqr*(1+x0*xtxi*@transpose(x0))) '个值预测的样本标准差
scalar y_exp_lft=y0_hat-t_value*s_y0h                           '均值预测的置信区间的左界值
scalar y_exp_rht=y0_hat+t_value*s_y0h                           '均值预测的置信区间的右界值
scalar y_ind_lft=y0_hat-t_value*s_y0h_mns_y0                    '个值预测的置信区间的左界值
scalar y_ind_rht=y0_hat+t_value*s_y0h_mns_y0                    '个值预测的置信区间的右界值
```

在工作文件视窗下，可以看到如下的 EViews 对象，可以双击查看（见图 4.13 ）：

- 样本外 $\mathbf{X_0}$ 矩阵 (元素为 1,x2=log(20), x3=log(4), x4=log(4), x5=200) 的矩阵对象x0
- 样本外估计 $\hat{Y}_0$ 值的标量对象y0_hat
- 均值预测的样本标准差 $S_{\hat{Y}_0}$ 的标量对象s_y0h
- 个值预测的样本标准差 $S_{Y_0-\hat{Y}_0}$ 的标量对象s_y0h_mns_y0
- 均值预测的置信区间的左界值 $\hat{Y}_0 - t_{1-\alpha/2}(n-2)\cdot S_{\hat{Y}_0}$ 的标量对象y_exp_lft
- 均值预测的置信区间的右界值 $\hat{Y}_0 + t_{1-\alpha/2}(n-2)\cdot S_{\hat{Y}_0}$ 的标量对象y_exp_rht
- 个值预测的置信区间的左界值 $\hat{Y}_0 - t_{1-\alpha/2}(n-2)\cdot S_{Y_0-\hat{Y}_0}$ 的标量对象y_ind_lft
- 个值预测的置信区间的右界值 $\hat{Y}_0 + t_{1-\alpha/2}(n-2)\cdot S_{Y_0-\hat{Y}_0}$ 的标量对象y_ind_rht

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

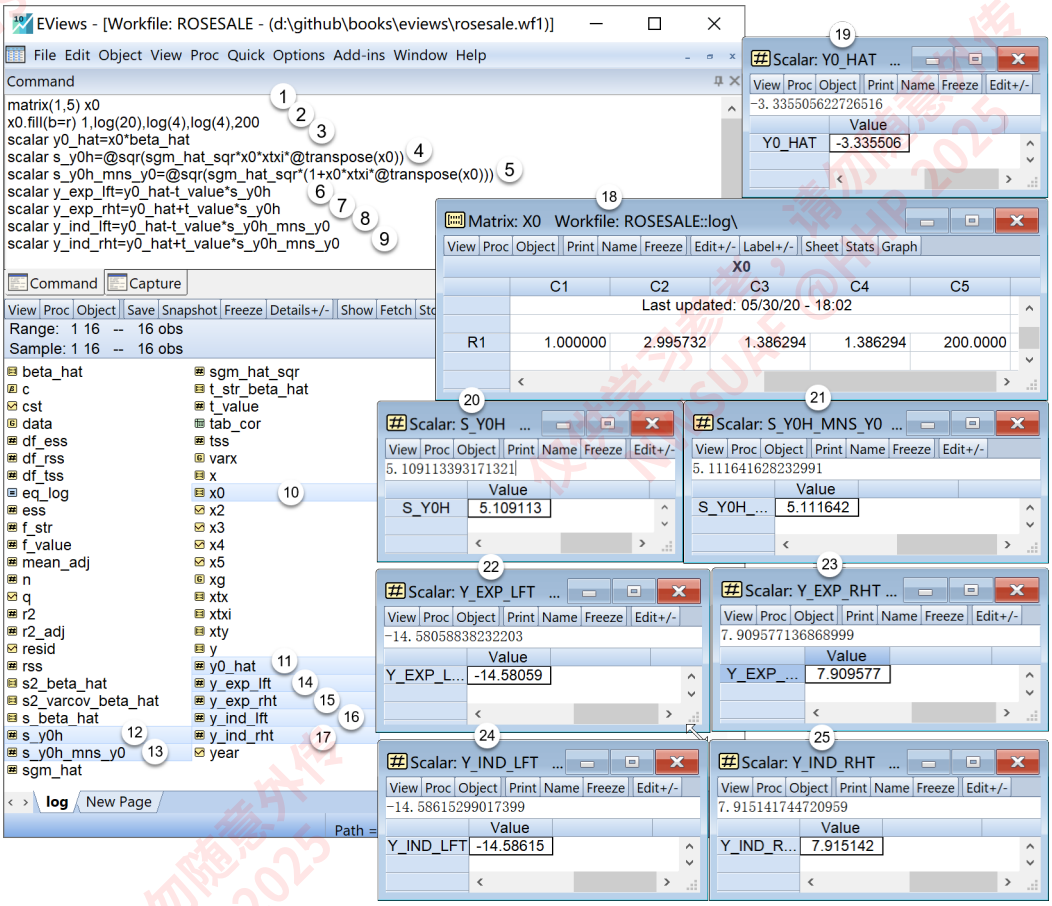


图 4.13: 进行样本外的均值预测、个值预测，并计算置信区间

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

4.6 附录：prg 源代码

实际操作中，在 EViews 命令视窗中逐条输入代码，既容易出错，又不便于维护这些代码，还不能进行代码的重复使用（在第一章的节 @ref(command-system) 中已经论述）。

因此，读者可以创建一个.prg 编程文件，并在其中编写 EViews 代码，进行管理、维护、运行和分析。下面代码按本章主要实验步骤编写，读者可以用于本章的 EViews 编程参考，进行实验练习。

```
'=================================================================================================
'说明：以下为EViews编程文件rosesale.prg的代码
'将展示第三章中“英国家庭食物支出案例”主要分析步骤的“批量式命令驱动”实现方法(::
'其中，符号'起始的行，为注释行，其他为EViews命令行。
'=================================================================================================

'创建工作文件（工作文件名=rose，子页命名=sale），无结构无日期，样本数为16
wfcreate(wf=rose,page=sale)  u 16

'导入外部数据，路径为d:\github\books\data\Lab3-family-spends.xlsx
import d:\github\books\data\lab4-rose-demand-origin.xlsx

'生成线性回归模型的方程对象
equation eq_main.ls log(q) c log(x2) log(x3) log(x4) x5    '对数模型

'构造几个重要变量对象
scalar n=@obs(x2)                  '样本数n(标量)
series cst=1                       '新建元素全为1的序列对象（用于构造矩阵X）
group xg cst log(x2) log(x3) log(x4) x5    '构造为group，便于观察

'构造X矩阵和Y矩阵对象。
matrix x=xg                    '转换为X矩阵对象
matrix y=log(q)                '构造Y矩阵对象

'计算回归方程的回归系数向量
matrix xtx=@transpose(x)*x  '得到重要矩阵X'X
matrix xtxi=@inverse(xtx)   '得到重要矩阵(X'X)^(-1)
matrix xty=@transpose(x)*y  '得到重要矩阵X'y
```

```
matrix beta_hat=xtxi*xty    '得到回归系数矩阵

'计算回归误差方差及标准差
scalar sgm_hat_sqr=1/(n-5)*(@transpose(y)*y-@transpose(beta_hat)*xty)    '回归误差方差
scalar sgm_hat=@sqr(sgm_hat_sqr)                                          '回归误差标准差

'计算回归系数的方差协方差矩阵、系数的样本方差和标准差(列向量)
matrix s2_varcov_beta_hat=sgm_hat_sqr*xtxi                '回归系数的方差协方差矩阵
matrix s2_beta_hat=@getmaindiagonal(s2_varcov_beta_hat)   '回归系数的样本方差
matrix s_beta_hat=@sqr(s2_beta_hat)                       '回归系数的样本标准差

'进行平方和分解，计算TSS、ESS和RSS，以及各自的自由度（标量）
scalar mean_adj=16*(@mean(q))^2                           '均值修正值
scalar tss=@transpose(y)*y-mean_adj                       '总平方和TSS
scalar rss=@transpose(y)*y-@transpose(beta_hat)*xty       '剩余平方和RSS
scalar ess=@transpose(beta_hat)*xty-mean_adj              '回归平方和ESS
scalar df_tss=n-1
scalar df_rss=n-5                                         'RSS的自由度
scalar df_ess=4                                           'ESS的自由度

'计算自变量的相关系数表格、回归方程的判定系数和调整判定系数
group varx x2 x3 x4 x5              ' 构建只含X的group
freeze(tab_cor) varx.cor     '把group的相关系数矩阵表视图保存为表格
scalar r2=ess/tss             '回归方程的判定系数
scalar r2_adj=1-(rss/df_rss)/(tss/df_tss)      '回归方程的调整判定系数

'对回归方程的回归系数进行显著性t检验
matrix t_str_beta_hat=@ediv(beta_hat,s_beta_hat)   '计算得到回归系数的样本t统计量
scalar t_value=@qtdist(0.975,df_rss)               '给定 =0.05 水平下的理论t值（右侧正值）

' 对回归方程的整体显著性进行F假设检验
scalar f_str=(ess/4)/(rss/11)                    '计算回归方程的样本F统计量
scalar t_value=@qfdist(0.95,df_ess,df_rss)       '计算给定 =0.05 水平下的查表的理论F值

' 进行样本外的均值预测、个值预测，并计算置信区间
```

```
matrix(1,5) x0                                          '产生1行*5列的空矩阵
x0.fill(b=r) 1,log(20),log(4),log(4),200                '样本外X值为(x2=log(20), x3=log(4), x4=log(4), x5
scalar y0_hat=x0*beta_hat                                                  '样本外估计Y值
scalar s_y0h=@sqr(sgm_hat_sqr*x0*xtxi*@transpose(x0))                      '均值预测的样本标准差
scalar s_y0h_mns_y0=@sqr(sgm_hat_sqr*(1+x0*xtxi*@transpose(x0)))           '个值预测的样本标准差
scalar y_exp_lft=y0_hat-t_value*s_y0h                                      '均值预测的置信区间的左界值
scalar y_exp_rht=y0_hat+t_value*s_y0h                                      '均值预测的置信区间的右界值
scalar y_ind_lft=y0_hat-t_value*s_y0h_mns_y0                               '个值预测的置信区间的左界值
scalar y_ind_rht=y0_hat+t_value*s_y0h_mns_y0                               '个值预测的置信区间的右界值


' ========================================================================
```

4.7 实验作业

玫瑰的需求：表 4.5 给出美国底特律市区对玫瑰的季度需求数据。

表 4.5: 玫瑰的需求 (n=13)

X1	YEAR	X2	X3	X4	X5	Ydata1	Ydata2	Ydata3	Ydata4
id	YEAR	X2	X3	X4	X5	2015014495	2016010317	2016011222	2016013000
name	YEAR	X2	X3	X4	X5	刘琳	王雪明	韩双瑞	任畅
class	YEAR	X2	X3	X4	X5	保险 1601	保险 1601	保险 1601	保险 1601
n1	1971.3	2.259999999999999	3.09	158.1100000000001	1	11483.44	11483.77	11485.56	11484.07
n2	1971.4	2.54	2.85	173.36	2	9348.049999999999	9348.382	9350.049999999999	9347.51
n11	1974.1	3.77	3.65	181.87	11	5910.48	5910.71	5911.4	5910.45
n12	1974.2	3.64	3.6	185	12	7949.9	7949.92	7951.02	7950.71
n13	1974.3	2.82	2.94	184	13	6134.28	6134.14	6133.92	6136.16

变量说明见表 4.6 ：

表 4.6: 变量定义及说明

variable	label
YEAR	年份. 季度
Q	玫瑰销售量 (打)
X2	玫瑰批发价格 (/)||$X3$| (/打)
X4	家庭可支配收入 ($/周)
X5	时间趋势

请考虑如下两个需求函数：

$$Y_t = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_5 X_{5t} + e_{1t} \tag{4.57}$$

$$ln(Y_t) = \hat{\beta}_1 + \hat{\beta}_2 ln(X_{2t}) + \hat{\beta}_3 ln(X_{3t}) + \hat{\beta}_4 ln(X_{4t}) + \hat{\beta}_5 X_{5t} + e_{2t} \tag{4.58}$$

请回答如下问题:

（1）关于线性模型式 4.57 ，运用菜单操作，得到回归分析报告。

（2）关于线性模型式 4.57 ，在 Eviews 中运用矩阵方法，计算如下步骤：

a. 计算直线回归方程的回归系数向量 ( )，并写出样本回归模型（SRM_2）。

b. 计算回归误差方差 ($\hat{\sigma}^2$) 和回归误差标准差 ($\hat{\sigma}$)。

c. 计算回归系数的样本方差协方差矩阵 ($\widehat{var}\_\widehat{cov}$( ))。

d. 得出回归系数的样本标准差向量 ($\mathbf{S}$)。

e. 进行平方和分解，计算 TSS、ESS 和 RSS。

f. 计算判定系数 R^2，调整判定系数 ($\hat{R}^2$)。

g. 计算样本 t 统计量 ($\mathbf{t}^*$)，并进行 t 假设检验。

h. 对回归方程的整体显著性进行 F 假设检验。

i. 对回归方程的进行样本外均值预测 $E(Y \mid X = X_0)$。

j. 对回归方程的进行样本外个值预测 $(Y_0 \mid X = X_0)$。

（3）关于对数线性模型式 4.58 ，运用菜单操作，得到回归分析报告。

（4）关于对数线性模型式 4.58 ，在 Eviews 中运用矩阵方法，计算如下步骤：

a. 算直线回归方程的回归系数向量 $(\hat{\ })$，并写出样本回归模型（SRM_2）

b. 计算回归误差方差 $(\hat{\sigma}^2)$ 和回归误差标准差 $(\hat{\sigma})$。

c. 计算回归系数的样本方差协方差矩阵 $(\widehat{var}\_\widehat{cov}(\hat{\ }))$。

d. 得出回归系数的样本标准差向量 $(\mathbf{S})$。

e. 进行平方和分解，计算 TSS、ESS 和 RSS。

f. 计算判定系数 R^2，调整判定系数 $(\hat{R}^2)$。

g. 计算样本 t 统计量 $(\mathbf{t}^*)$，并进行 t 假设检验。

h. 对回归方程的整体显著性进行 F 假设检验。

i. 对回归方程的进行样本外均值预测 $E(Y \mid X = X_0)$。

j. 对回归方程的进行样本外个值预测 $(Y_0 \mid X = X_0)$。

（5）根据对数模型特征，可知 $\hat{\beta}_2$、$\hat{\beta}_3$ 和 $\hat{\beta}_4$ 分别为玫瑰需求的自价格弹性，交叉价格弹性和收入弹性。它们的先验符号是什么？你的结果同先验预期相符吗？

（6）根据你的分析，你会选择哪个模型 (如果可选)？为什么？

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

（7）仅考虑对数设定形式模型式 4.58 ：

a. 所估计的需求自价格弹性 (即对玫瑰价格的弹性) 是什么?

b. 它是统计显著的吗?

c. 如果是，它是否在统计上异于 1?（此题为选作）

d. 理论上，你对 $\hat{\beta}_3$ 和 $\hat{\beta}_4$ 的预期符号是什么?eviews 结果和这些预期相符吗?

e. 如果 $\hat{\beta}_3$ 和 $\hat{\beta}_4$ 的系数在统计意义上不显著，可能是什么原因?

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

5 多重共线性问题

仅供学习参考，请勿随意外传
NWSUAF @HHP 2025

5.1 实验目的及要求

- **目的**：掌握多重共线性的检验与处理方法。
- **要求**：在老师指导下完成计量经济模型的多重共线性检验，并对存在多重共线性的模型进行修正，最终得到正确的分析结果。

5.2 实验原理

线性回归模型的解释变量不满足相互独立的基本假设前提下，如果模型的解释变量存在多重共线性，将导致最小二乘法得到的模型参数估计量非有效且方差变大，参数估计量经济含义不合理等。

5.2.1 什么是模型多重共线性?

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad \text{(PRM)} \tag{5.1}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i \quad \text{(SRM)} \tag{5.2}$$

定义 5.1 (多重共线性)**.** 在多元线性回归模型中，各解释变量 $\{X_2, X_3, \cdots, X_k\}$ 之间有交互相关，但又非完全相关的现象。正式地：

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} + \cdots + \lambda_k X_{ki} + v_i = 0 \tag{5.3}$$

其中，v_i 为随机误差项。

我们称总体回归模型式 5.1 存在**多重共线性**，此时 $\lambda_1; \lambda_2; \cdots; \lambda_k$ 不全为 0，且 $v_i \neq 0$。

我们称总体回归模型式 5.1 存在**完全多重共线性**，此时 $\lambda_1; \lambda_2; \cdots; \lambda_k$ 不全为 0，且 $v_i = 0$。

如果模型出现多重共线性问题，在 N-CLRM 假设下，OLS 估计量仍然是最优线性无偏估计量（BLUE）：

- 只要不是**完全共线性**，在近似多重共线性的情形下，OLS 估计量仍然是无偏的
- 只要不是**完全共线性**，在近似多重共线性的情形下，OLS 估计量的方差一定是小的

- 多重共线性本质上是一种样本现象。即使总体中 X 变量间不存在共线性，由于抽样方法或小样本问题，也可能带来多重共线性问题

但是如果模型出现多重共线性问题，则往往会带来如下一些理论后果：

- 更大的方差和协方差，估计精度大大下降。（尽管 OLS 估计仍是 BLUE）
- 参数估计置信区间变宽，系数的检验倾向于不显著。（即回归系数的 t 检验更倾向接受原假设 H0，认为系数为零）
- 系数的 t 值倾向于统计上不显著，但判定系数 R^2 却会很高
- 0LS 估计量及其标准误对数据的微小变化非常敏感

5.2.2 如何诊断模型多重共线性问题?

为了诊断样本回归模型式 5.2 是否存在多重共线性问题，并判明其严重程度，我们一般可以采用如下的经验方法：

（1）主回归方程分析报告结果异常。根据样本回归模型式 5.2 ，如果出现如下情形，我们认为回归方程结果异常，很有可能模型存在多重共线性问题：

- 主回归分析报告的 R^2 值高（大于 0.8）
- 分析报告 F^* 检验显著
- 但不显著的 t^* 检验较多（多于回归系数个数的一半及以上）

（2）图形分析发现自变量之间存在明显相关关系。利用样本数据，我们可以分析得到自变量之间的**相关系数矩阵**和**矩阵散点图**，如果出现如下情形，我们认为回归方程结果异常，很有可能模型存在多重共线性问题：

- 相关系数矩阵发现高度线性相关（相关系数大于 0.8）
- 散点图矩阵发现高度线性相关的数据分布模式

（3）分析检查辅助回归方程是否异常。首先构建主回归方程，然后分别构建回归元之间的辅助回归方程。如果出现如下情形，我们认为回归方程结果异常，很有可能模型存在多重共线性问题：

- 辅助回归方程的判定系数 R_j^2 大于主回归方程的判定系数 R^2

- 辅助回归方程的方差膨胀因子：$VIF_j \in [10, 100]$ 表明中度多重共线性；$VIF_j \geq 100$ 表明严重多重共线性
- 辅助回归方程的的容忍度：$TOL_j \in [0.01, 0.1]$ 表明中度多重共线性；$TOL_j \leq 0.01$ 表明严重多重共线性

（4）利用回归系数方差分解法（Coefficient Variance Decomposition）。通过计算特征值 (eigenvalues)，进而得到病态数 (K_j) 和方差分解比率 VDP_j，最后做出多重共线性的诊断结论：

- 若发现至少一个病态数 $K_j \leq 0.001$，则表明存在严重多重共线性；
- 观察病态数最小时所对应的方差分解比率，如果有多个斜率系数的 $VDP_j \geq 0.5$，则表明存在严重的多重共线性

定义 5.2 (多重共线性辅助回归方程)**.** 为侦察多元回归模型是否存在多重共线性，构建关于自变量之间的一类线性回归方程。正式地，对于样本回归模型式 5.2 ，我们可以构建得到如下多重共线性辅助回归方程：

$$\begin{aligned} X_{2i} &= \hat{\alpha}_1 + \cdots + \hat{\alpha}_j X_{ji} + \cdots + \hat{\alpha}_k X_{ki} + \epsilon_{2i} \\ &\vdots \\ X_{ji} &= \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \cdots + \hat{\alpha}_k X_{ki} + \epsilon_{ji} \\ &\vdots \\ X_{ki} &= \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \cdots + \hat{\alpha}_j X_{ji} + \cdots + \epsilon_{ki} \end{aligned} \tag{5.4}$$

对于样本回归模型式 5.2 ，在普通最小二乘法下，我们可以证明：

$$S^2_{\hat{\beta}_j} = \frac{\hat{\sigma}^2}{(n-1)S^2_{X_j}} \cdot \frac{1}{1 - R^2_j} \tag{5.5}$$

其中 R^2_j 为辅助回归方程式 5.4 的判定系数。$S^2_{X_j}$ 为变量 X_j 的样本方差。

定义 5.3 (方差膨胀因子和容忍度)**.** 用来度量在最小二乘法下，多元线性回归模型的多重共线性问题严重程度的两个指标。

具体地，**方差膨胀因子**表达为单独一个变量对模型估计方差贡献的比率，一般记为 VIF（variance-inflation factor），并正式定义为：

$$VIF_j = \frac{1}{1 - R_j^2} \tag{5.6}$$

容忍度是方差膨胀因子的倒数，一般记为 TOL（tolerrance），并正式定义为：

$$TOl_j = 1 - R_j^2 \tag{5.7}$$

根据多元回归分析的矩阵方法实现（第 @ref(matrix-regression) 章），对于多元回归模型

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{5.8}$$

$$\mathbf{y} = \mathbf{X} \ + \mathbf{u} \quad \text{(PRM)} \tag{5.9}$$

$$\mathbf{y} = \mathbf{X\hat{} } \ + \mathbf{e} \quad \text{(SRM)} \tag{5.10}$$

可以得到 OLS 下参数估计的方差协方差矩阵：

$$var - cov(\hat{} \ ) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 \mathbf{V}\mathbf{D}^{-\mathbf{1}}\mathbf{V}' \tag{5.11}$$

其中，$\mathbf{D}$ 是含有矩阵 $(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}$ 的**特征值**（eigenvalues）E_m（$m \in 1, 2, \cdots, k$）的一个对角矩阵，而 $\mathbf{V}$ 是由相应**特征向量**构成的一个矩阵。

定义 5.4 (病态数和方差分解比率)**.** 用来诊断多元线性回归模型的多重共线性问题严重程度的一个指标。

具体地，**病态数**（condition number）采用微分的方法，考察引入一个变量对（多重共线性）模型结果恶化情形出现的相对改变数，一般记为 K，并正式定义为：

$$K_j = \frac{min(E_m)}{E_j} \tag{5.12}$$

方差分解比率（variance-decomposition proportion），一般记为 VDP_{ji}，并正式定义为：

$$\phi_{ij} = \frac{v_{ij}^2}{E_j}$$
$$VDP_{ji} = \frac{\phi_{ij}}{\phi_i} \tag{5.13}$$

其中 v_{ij} 为矩阵 $\mathbf{V}$ 的第 (i, j) 个元素。

5.2.3 如何修正模型多重共线性问题?

一旦发现模型存在比较严重的多重共线性问题，则需要对模型进行修正处理，具体方法可参考：

（1）简单剔除变量法：

- 依据经济学和实践经验观察，进行变量甄选或变量变换
- 变量变换法，进行变量处理。具体又包括差分变换法、比率变换法

（2）逐步回归法：包括前向逐步回归 (forward stepwise) 和后向逐步回归 (backward stepwise)

- p 值判别法：$p \in [0.1, 0.05)$(比较显著)；$p \in [0.05, 0.01)$(比较显著)；$p \leq 0.01$(极其显著)
- t^* 值判别法：2t 法则

（3）补充新数据（有时候有用!）。由于多重共线性是一个样本特性，故有可能在关于同样变量的另一样本中共线性没有第一个样本那么严重

（4）多项式回归模型中离差形式或正交多项式 (orthogonal polynomials) 以降低共线性的影响。多项式回归模型的一个特点是解释变量以不同的幂出现，从而容易导致多重共线性

（5）矫正多重共线性问题的其他方法

- 脊回归或岭回归 (ridge regression)，常被用来” 解决” 多重共线性问题。
- 主成分分析法。先根据主成分分析确定主成分个数（看累积解释百分比），再用主成分得分（scoring）序列进行回归分析。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

5.3 实验内容

（1）采用最小二乘法建立主回归模型。

（2）侦查模型是否存在多重共线性。

- 观察主回归方程分析报告：建立主回归方程，分析回归报告结果。主回归分析报告的 R^2 值高（大于 0.8），F^* 检验显著，但不显著的 t^* 检验较多（多于回归系数个数的一半及以上）
- 矩阵相关系数和矩阵散点图：绘制回归元之间的相关系数矩阵和散点图矩阵。相关系数矩阵发现高度线性相关（相关系数大于 0.8），以及散点图矩阵发现高度线性相关的数据分布模式。
- 分析辅助回归方程：首先构建主回归方程，然后分别构建回归元之间的辅助回归方程。辅助回归方程的判定系数 R_j^2 大于主回归方程的判定系数 R^2。辅助回归方程的方差膨胀因子 $VIF_j \in [10, 100]$ 表明中度多重共线性，$VIF_j \geq 100$ 表明严重多重共线性。辅助回归方程的的容忍度 $TOL_j \in [0.01, 0.1]$ 表明中度多重共线性，$TOL_j \leq 0.01$ 表明严重多重共线性。
- 主成分分析法 (principal components)：计算特征值 (eigenvalues)，进而得到病态数 (K) 和病态指数 $CI = \sqrt{k}$。病态数：$K \in [100, 1000]$ 表明中度多重共线性；$K \geq 1000$ 表明严重多重共线性。病态指数：$CI \in [10, 30]$ 表明中度多重共线性；$CI \geq 30$ 表明严重多重共线性

（3）根据上述对多重共线性的诊断，对模型进行合理修正。

- 简单剔除变量法：依据经济学和实践经验观察，进行变量甄选或变量变换
- 逐步回归法：包括前向逐步回归 (forward stepwise) 和后向逐步回归 (backward stepwise) 。采用 p 值判别法，若 $p \in [0.1, 0.05)$，比较显著；若 $p \in [0.05, 0.01)$，比较显著；若 $p \leq 0.01$，极其显著。也可以采用 t^* 值判别法，也即 2t 法则。
- 补充新数据（有时候有用！）。由于多重共线性是一个样本特性，故有可能在关于同样变量的另一样本中共线性没有第一个样本那么严重。
- 多项式回归模型中离差形式或正交多项式 (orthogonal polynomials) 以降低共线性的影响。多项式回归模型的一个特点是解释变量以不同的幂出现，从而容易导致多重共线性
- 拯救多重共线性的其他方法。包括脊回归 (ridge regression) 和主成分分析法。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

5.4 实验准备

5.4.1 实验软件

本次实验需要提前准备好如下软件：

- 统计分析软件 Eviews 9.0 版本及以上
- 公式编辑软件 Mathtype 6.0 版本及以上
- 写作编辑软件 Office Word/Excel 2010 版本及以上
- 浏览器软件 chrome 66.0 版本及以上或 360 极速浏览器 9.5 版本及以上

5.4.2 实验材料

就业情况的郎利数据：表 5.1 给出美国 1947-1961 年间就业情况及主要影响因素的数据表。

表 5.1: 美国劳动力市场的郎利调研数据 (n=16)

Year	Y	X1	X2	X3	X4	X5	X6
1947	60323	830	234289	2356	1590	107608	1
1948	61122	885	259426	2325	1456	108632	2
1949	60171	882	258054	3682	1616	109773	3
1950	61187	895	284599	3351	1650	110929	4
1951	63221	962	328975	2099	3099	112075	5
1952	63639	981	346999	1932	3594	113270	6
1953	64989	990	365385	1870	3547	115094	7
1954	63761	1000	363112	3578	3350	116219	8
1955	66019	1012	397469	2904	3048	117388	9
1956	67857	1046	419180	2822	2857	118734	10
1957	68169	1084	442769	2936	2798	120445	11
1958	66513	1108	444546	4681	2637	121950	12
1959	68655	1126	482704	3813	2552	123366	13
1960	69564	1142	502601	3931	2514	125368	14
1961	69331	1157	518173	4806	2572	127852	15

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

1962	70551	1169	554894	4007	2827	130081	16

变量说明见表 5.2 ：

表 5.2: 变量定义及说明

variable	label
Year	年份
Y	就业人数 (打)
X1	消费价格指数
X2	名义 GNP
X3	失业人数
X4	军队人数
X5	14 岁以上的非机构人口数
X6	时间趋势

5.4.3 实验规则

本实验将要求保留 Eviews 操作过程的相关结果，因此对 Eviews 对象命名规则设计如下：

（1）方程对象（Equation）的命名规则：

- 主回归方程对象保存命名为eq_m0
- 辅助回归方程对象依次保存命名为eq_a1、eq_a2、eq_a3、eq_a4、eq_a5、eq_a6
- 用经济学和实践经验删除变量来矫正模型，新方程保存并命名为eq_adj_man
- 用逐步回归方法来矫正模型，新方程保存并命名为eq_adj_step

（2）向量对象（Vector）的命名规则：

- 构造含有 6 个元素的向量对象，用于存放 6 个辅助回归方程的判定系数，并命名为r2
- 构造含有 6 个元素的向量对象，用于存放 6 个辅助回归方程计算得到的方差膨胀因子，并命名为vif
- 构造含有 6 个元素的向量对象，用于存放 6 个辅助回归方程计算得到的容忍度，并命名为tol

（4）表格对象（Table）的命名规则：

- 自变量的相关系数矩阵保存为表格对象，并命名为：tab_cor
- 回归系数方差分解结果保存为表格对象，并命名为：tab_cvd

仅供学习参考，请勿随意外传
NWSUAF @HHP 2025

5.5 主要实验步骤

5.5.1 新建工作文件并导入数据

- 目标：构建工作文件，成功导入数据
- 思路：利用 EViews 代码创建工作文件并导入数据。

Eviews 操作：

（1）新建 Eviews 工作文件（见 `@fig-load-wage` ）

> 提示
>
> 首先，依次操作：File⇒New⇒Workfile
> 然后，进行 workfile create 引导设置：
>
> - workfile structure type: `unstructured/undatede`
> - data range：16
> - workfile names(optional):
>   - WF: `longley`（建议命名）
>   - Page: `employee`（建议命名）

（2）导入外部 xlsx 数据

> 提示
>
> 首先，处理 Excel 数据。
> 然后，Eviews 导入数据。
>
> - `File ⇒ Import ⇒ Import From File: d:\github\books\data\Lab5-longley-short-origin.xlsx`

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

提示

通过运行 EViews 代码，也可以实现上述操作。在命令视窗中依次输入并运行如下 EViews 代码：

```
******
'创建工作文件（工作文件名=longley，子页命名=employee），无结构无日期，样本数为16
`wfcreate(wf=longley, page=employee)  u 16`

******
'导入外部数据，路径为`d:\\github\\books\\data\\Lab5-longley-short-origin.xlsx`
`import d:\\github\\books\\data\\Lab5-longley-short-origin.xlsx`
```

5.5.2 采用最小二乘法建立主回归模型

Eviews 操作目标：得到回归方程，查看回归结果

思路：构建回归方程对象。回归模型为：

$$\begin{aligned} Y_i = &+ \beta_1 + \beta_2 X1_i + \beta_3 X2_i + \beta_4 X3_i \\ &+ \beta_5 X4_i + \beta_6 X5_i + \beta_7 X6_i + u_i \end{aligned} \tag{5.14}$$

Eviews 操作：构建主回归方程

（1）进入方程估计的引导界面。

- EViews 主窗口上依次点击点 `Quick` ⇒ `Estimation Equation`

（2）完成方程估计的引导设置。

- 设置方程。`Equation Estimation` ⇒ `Specification` ⇒ `Equation specification` 中依次输入变量 `y c x1 x2 x3 x4 x5 x6`（注意变量之间的空格，以及截距 `c`）
- 选择估计方程。`Estimation settings` 中的 `Method` 下拉框 ⇒ 下拉选择 `LS - Least Squares (NLS and ARMA)`
- 完成设置，点击 `OK`

（3）命名并保存回归方程。

- 在未命名的方程对象UNTITLED 视窗下，点击菜单栏 Name ⇒ 输入命名 eq_m0（建议命名）⇒ 完成命名，点击 Ok

在工作文件视窗下，可以看到如下新生成的方程对象，可以双击查看eq_m0（见图 5.1 ）：

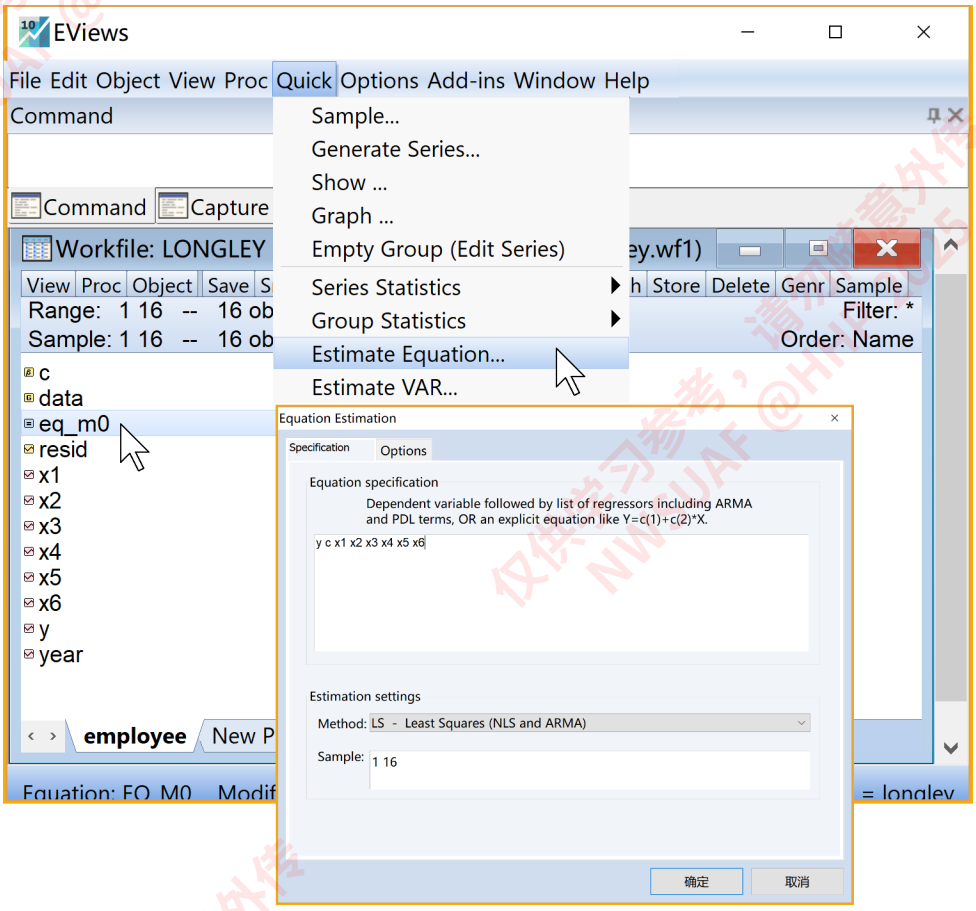


图 5.1: 主回归模型 Eviews 建模操作

回归方程结果见图 5.2 ：

> 提示
>
> 通过运行 EViews 代码，也可以实现上述操作。在命令视窗中依次输入并运行如下 EViews 代码：
>
> ```
> ****
> '生成线性回归模型的方程对象
> equation eq_m0.ls y c x1 x2 x3 x4 x5 x6
> ```

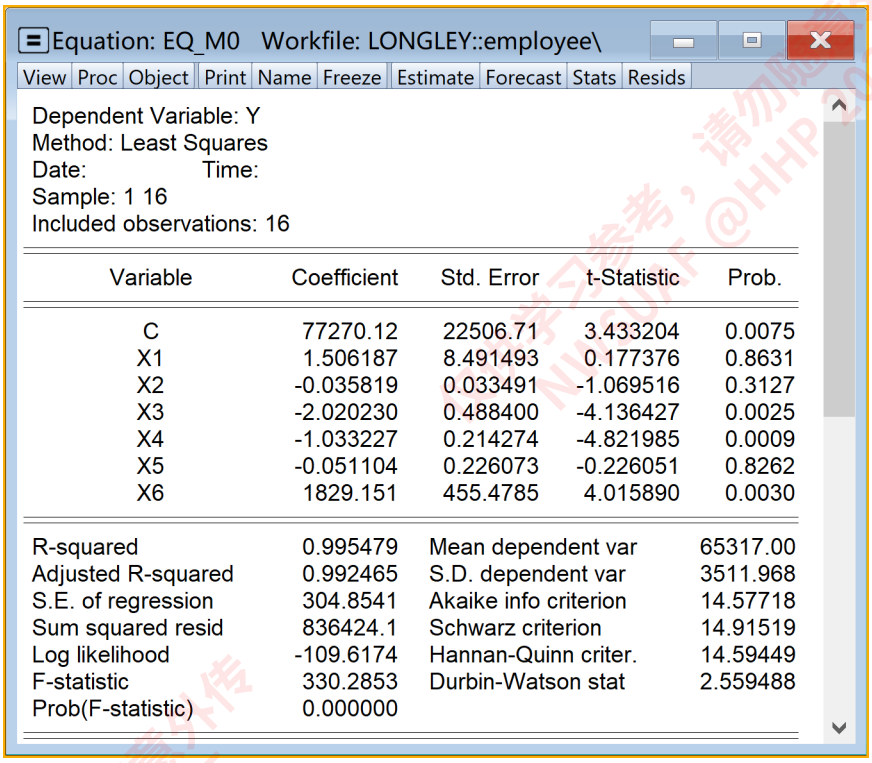


图 5.2: 郎利数据的回归结果

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

5.5.3 侦查多重共线性

5.5.3.1 主回归方程报告观察法

目标：观察 t^* 检验，判定系数 R^2，F^* 检验的关系

思路：观察主回归分析报告的相关指标，诊断的标准如下：

- R^2 值高（大于 0.8）
- F^* 检验显著
- 不显著的 t^* 检验较多（斜率系数个数的一半及以上）

郎利数据案例中，根据 EViews 回归结果（见图 5.2 ），可以整理成简要回归报告如下：

$$\begin{aligned}\widehat{Y}_i = &+ \hat{\beta}_1 + \hat{\beta}_2 X1_i + \hat{\beta}_3 X2_i + \hat{\beta}_4 X3_i \\ &+ \hat{\beta}_5 X4_i + \hat{\beta}_6 X5_i + \hat{\beta}_7 X6_i\end{aligned} \tag{5.15}$$

$$\begin{array}{llll}\widehat{Y} = & +77270.12 & +1.51X1_i & -0.04X2_i \quad -2.02X3_i \\ (s) & (22506.7070) & (8.4915) & (0.0335) \quad (0.4884) \\ (t) & (+3.43) & (+0.18) & (-1.07) \quad (-4.14) \\ (cont.) & -1.03X4_i & -0.05X5_i & +1829.15X6_i \\ (s) & (0.2143) & (0.2261) & (455.4785) \\ (t) & (-4.82) & (-0.23) & (+4.02) \\ (over) & n = 16 & \hat{\sigma} = 304.8541 & \\ (fit) & R^2 = 0.9955 & \bar{R}^2 = 0.9925 & \\ (Ftest) & F^* = 330.29 & p = 0.0000 &\end{array} \tag{5.16}$$

显然，在 $\alpha = 0.05$ 水平下 t 检验结论见表 5.3：

表 5.3: 主方程回归情况（$\alpha = 0.05$ 水平下 t 检验）

回归元	回归系数	标准误	t 统计量	p 值	显著性结论
(Intercept)	77270.115	22506.707	3.43	0.0075	
X1	1.506	8.492	0.18	0.8631	不显著

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 5.3: 主方程回归情况（$\alpha = 0.05$ 水平下 t 检验）

回归元	回归系数	标准误	t 统计量	p 值	显著性结论
X2	-0.036	0.034	-1.07	0.3127	不显著
X3	-2.020	0.488	-4.14	0.0025	
X4	-1.033	0.214	-4.82	0.0009	
X5	-0.051	0.226	-0.23	0.8262	不显著
X6	1829.151	455.478	4.02	0.0030	

根据主回归报告（见图 5.2 和方程式 5.15 ），下列证据将表明模型可能存在**严重**的多重共线性问题：

- 判定系数 $R^2 = 0.9955$，表明样本回归线拟合较好
- $F^* = 330.29$，对应的概率值 $p = 0.0000$，表明 F 检验极显著
- 斜率系数 t^* 检验不显著的有 3 个（主模型全部斜率系数共有 6 个），分别是 X1、X2、X5。

5.5.3.2 矩阵相关系数和矩阵散点图

目标：观察 t^* 检验，判定系数 R^2，F^* 检验的关系

思路：得到回归元之间的相关系数表，绘制它们的散点矩阵图。分析图表是否存在如下特征：

- 散点图矩阵发现高度线性相关的数据分布模式
- 相关系数矩阵发现高度线性相关（相关系数大于 0.8）

Eviews 菜单操作：

（1）构造 X 数据的组对象（group）：

- 依次选择 X 变量：X1，X2，X3，X4，X5，X6
- 以组对象打开，鼠标右键：Open as group
- 命名并保存组对象：name（**建议命名为** `group_x`）

（2）绘制散点图矩阵

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 进入 group 视窗：双击打开组对象 `group_x`
- 进入引导菜单：⇒ View ⇒ Graph
  - 选择绘图类型 (Graph type)：Scatter
  - 选择绘图细节 (Detail)：Multiple series ⇒ 下拉框选中 Scatterplot matrix
- 点击完成：OK
- 命名并保存绘图（graph）对象：（建议命名为 scatter_matrix）
- 查看结果：双击scatter_matrix （见图 5.3 ）

（3）制作得到相关系数矩阵表格（table）

- 进入 group 视窗：双击打开组（group）对象group_x
- 进入引导菜单：⇒ View ⇒ Covariance Analysis
  - 选择分析类型 (Statistics)：只勾选 Correlation
  - 其他设置细节：（默认设置）
  - 点击完成：OK
- 将上述组（group）对象group_x 另存为表格（table）对象
  - 另存为表格 (table) 对象：点击 Freeze
  - 命名并保存表格 (table) 对象：点击 name(建议为 tab_cor)
- 查看结果：双击tab_cor （见图 5.4 ）

提示

通过运行 EViews 代码，也可以实现上述操作。在命令视窗中依次输入并运行如下 EViews 代码：

```
****
`'多重共线性诊断：计算自变量的相关系数表格和散点矩阵图`
`group varx x1 x2 x3 x4 x5 x6          '构建只含X的group`
`freeze(tab_cor) varx.cor              '把group的相关系数矩阵表视图保存为表格`
`graph scatter_matrix.scatmat varx     '绘制散点矩阵图`
```

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

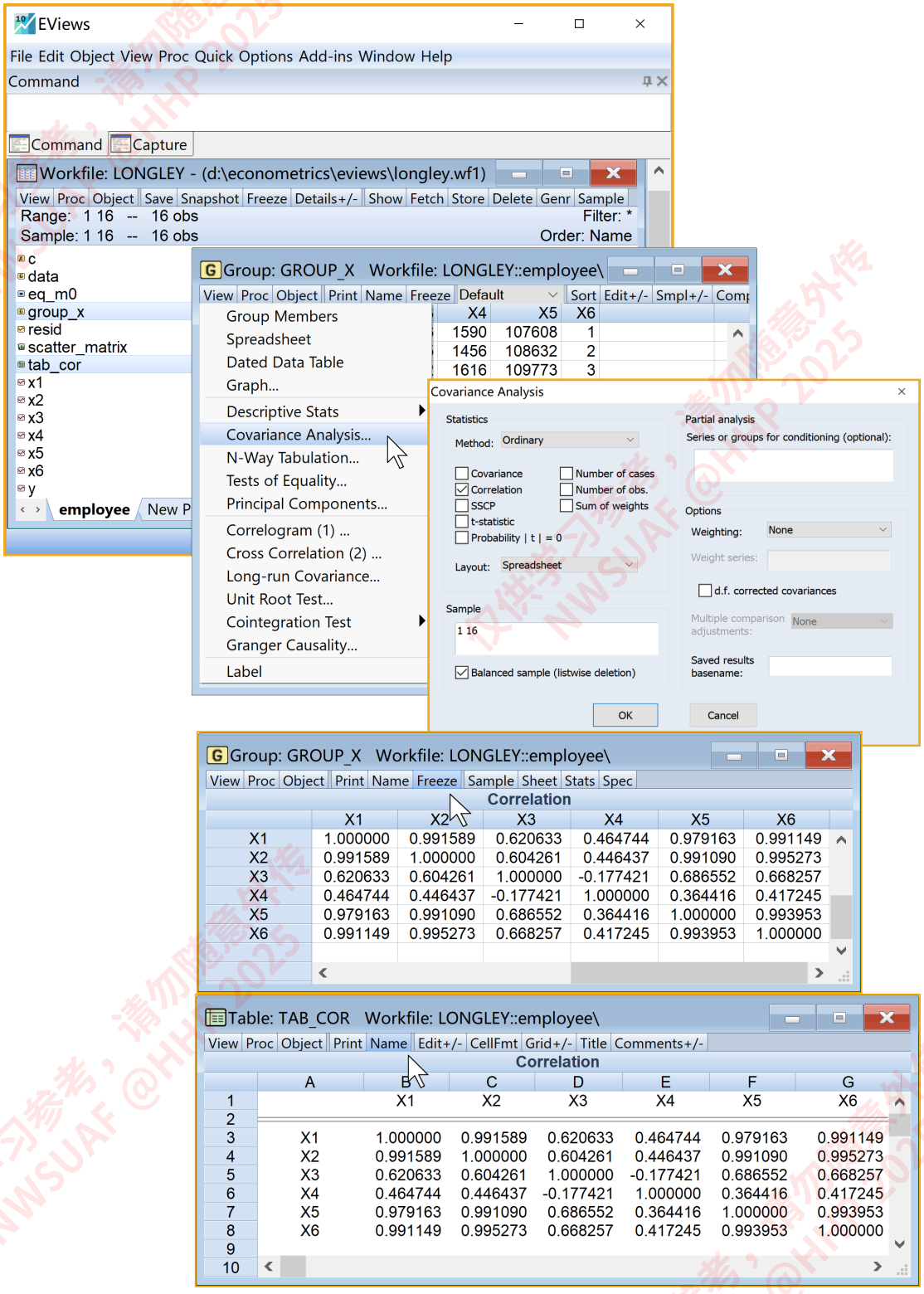


图 5.3: 自变量 X 间相关系数矩阵的 Eviews 操作

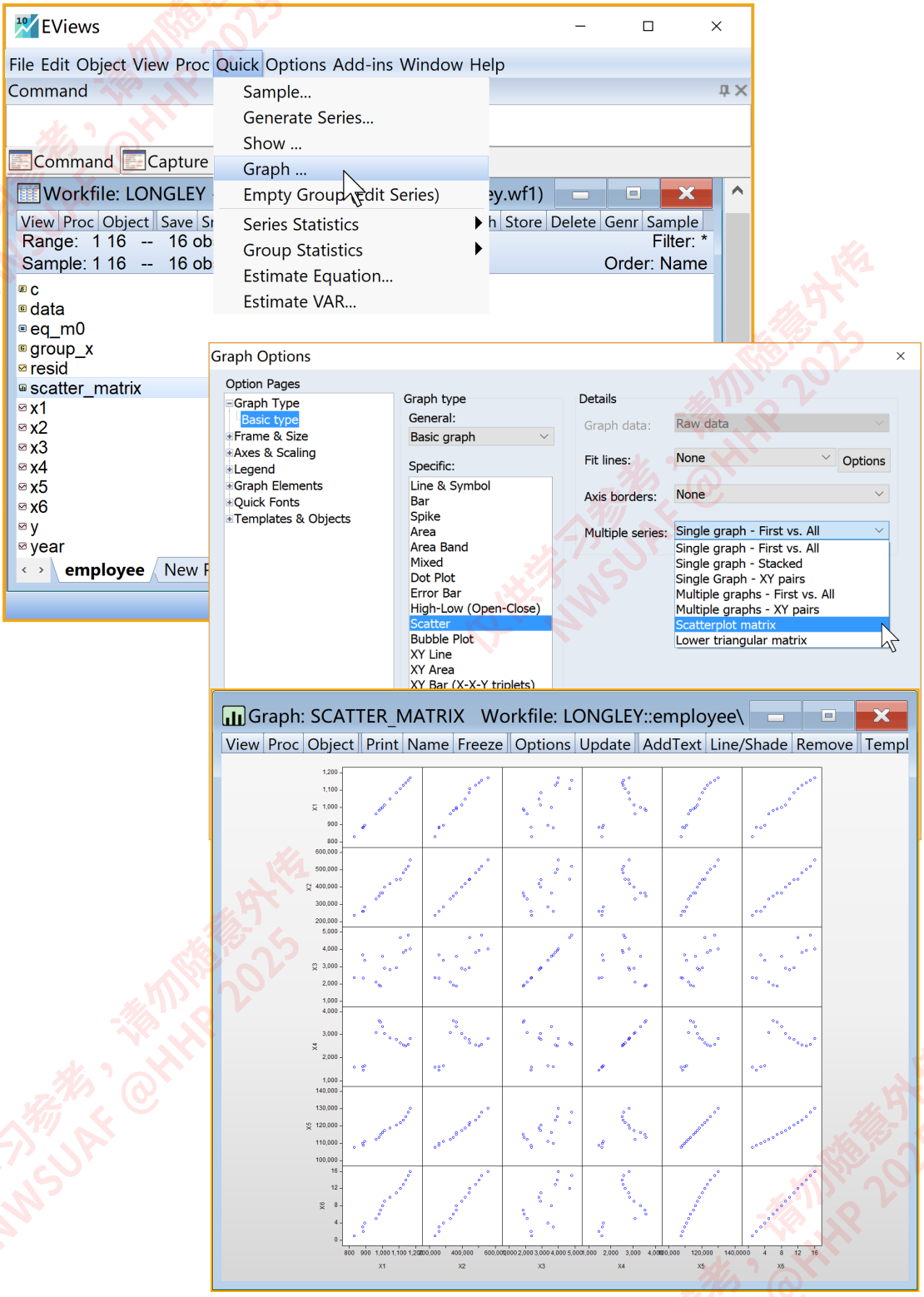


图 5.4: 自变量 X 间散点矩阵图的 Eviews 操作

表 5.4: 相关系数表

	X1	X2	X3	X4	X5	X6
X1	1.00					
X2	0.99	1.00				
X3	0.62	0.60	1.00			
X4	0.46	0.45	-0.18	1.00		
X5	0.98	0.99	0.69	0.36	1.00	
X6	0.99	1.00	0.67	0.42	0.99	1

根据相关系数矩阵表（图 5.3）和散点矩阵图（图 5.4 ），我们发现有 6 对 X 变量呈现明显线性相关关系，分别是：

- X1 VS X2：$r_{2,1}$=0.9916
- X1 VS X5：$r_{5,1}$=0.9792
- X1 VS X6：$r_{6,1}$=0.9911
- X2 VS X5：$r_{5,2}$=0.9911
- X2 VS X6：$r_{6,2}$=0.9953
- X5 VS X6：$r_{6,5}$=0.994

5.5.3.3 构建辅助回归方程

目标：构建并得到 6 个辅助回归方程及其 Eviews 报告。

思路：分别构建一个 X 变量对其他 X 变量的线性回归，共会得到 6 个辅助回归方程的 Eviews 报告。

定义 5.5 (主模型和辅助模型)**.** 主回归模型 (Main Model) 是指 Y 变量对全部 X 变量的线性回归。

辅助回归模型 (Auxiliary Model) 是指一个 X 变量对其他 X 变量的线性回归。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \hat{\beta}_3 X_{3t} + \hat{\beta}_4 X_{4t} + \hat{\beta}_5 X_{5t} + \hat{\beta}_6 X_{6t} + e_t \quad (\text{M0}) \tag{5.17}$$

$$X_{1t} = \hat{\alpha}_0 + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_5 X_{5t} + \hat{\alpha}_6 X_{6t} + e_t \quad (\text{A1}) \tag{5.18}$$

$$X_{2t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_5 X_{5t} + \hat{\alpha}_6 X_{6t} + e_t \quad (\text{A2}) \tag{5.19}$$

$$X_{3t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_5 X_{5t} + \hat{\alpha}_6 X_{6t} + e_t \quad (\text{A3}) \tag{5.20}$$

$$X_{4t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_5 X_{5t} + \hat{\alpha}_6 X_{6t} + e_t \quad (\text{A4}) \tag{5.21}$$

$$X_{5t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_6 X_{6t} + e_t \quad (\text{A5}) \tag{5.22}$$

$$X_{6t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_5 X_{5t} + e_t \quad (\text{A6}) \tag{5.23}$$

> ⚠ 警告
>
> **提示：**此处仅以 A1 式 5.18 为例，其他辅助模型类似操作。

Eviews 菜单操作（具体操作演示见图 5.5 ）：

（1）依次选择 ⇒ Quick ⇒ Estimation Equation

（2）引导设置 Equation Estimation ⇒ specification

- Equation specification：输入方程设置 `x1 c x2 x3 x4 x5 x6`
- Estimation settings：
  - Method: 下拉选择 LS - Least Squares (NLS and ARMA)
  - Sample: （默认设置）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

（3）点击完成：OK

（4）命名保存方程对象:（建议命名为 eq_a1，其他辅助回归方程依次命名为 eq_a2，…，eq_a6）

- 查看结果：双击eq_a1

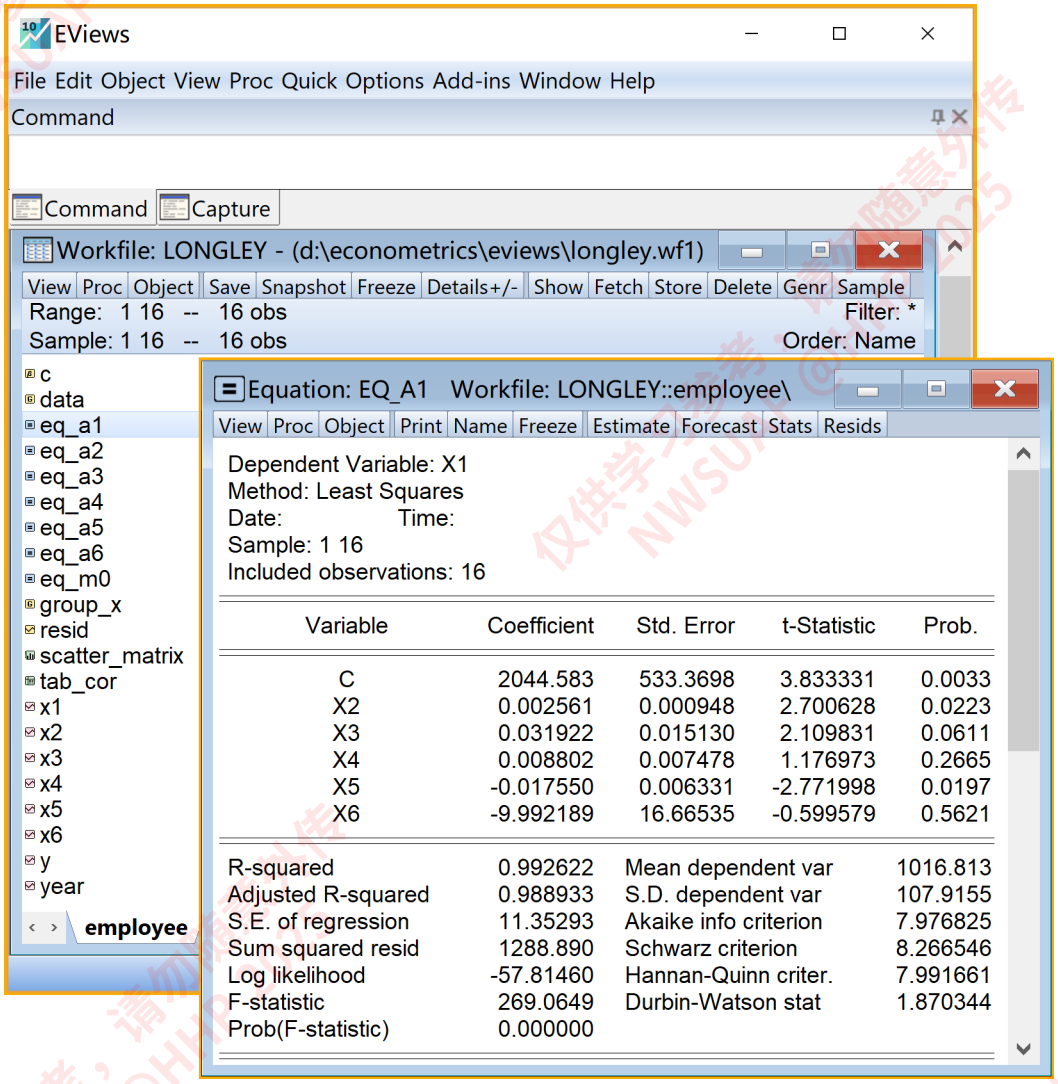
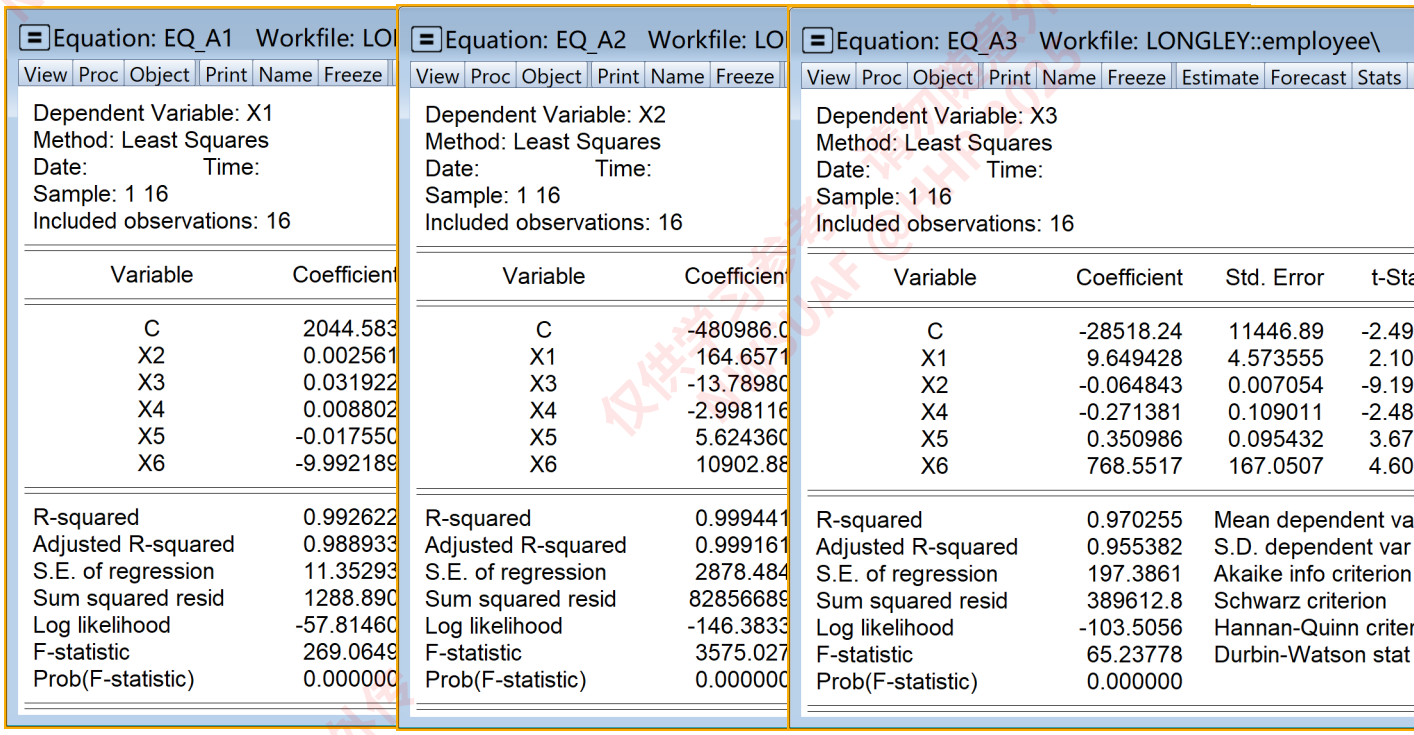


图 5.5: 辅助回归 Eviews 操作过程

全部 6 个辅助回归方程的 Eviews 分析报告结果：A1-A3 辅助模型见图 5.6 ；A4-A6 辅助模型见图 5.7 。



Equation: EQ_A1 Workfile: LOI

View | Proc | Object | Print | Name | Freeze

Dependent Variable: X1
Method: Least Squares
Date: Time:
Sample: 1 16
Included observations: 16

Variable	Coefficient
C	2044.583
X2	0.002561
X3	0.031922
X4	0.008802
X5	-0.017550
X6	-9.992189

R-squared	0.992622
Adjusted R-squared	0.988933
S.E. of regression	11.35293
Sum squared resid	1288.890
Log likelihood	-57.81460
F-statistic	269.0649
Prob(F-statistic)	0.000000

(a) A1

Equation: EQ_A2 Workfile: LOI

View | Proc | Object | Print | Name | Freeze

Dependent Variable: X2
Method: Least Squares
Date: Time:
Sample: 1 16
Included observations: 16

Variable	Coefficient
C	-480986.0
X1	164.6571
X3	-13.78980
X4	-2.998116
X5	5.624360
X6	10902.88

R-squared	0.999441
Adjusted R-squared	0.999161
S.E. of regression	2878.484
Sum squared resid	82856689
Log likelihood	-146.3833
F-statistic	3575.027
Prob(F-statistic)	0.000000

(b) A2

Equation: EQ_A3 Workfile: LONGLEY::employee\

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

Dependent Variable: X3
Method: Least Squares
Date: Time:
Sample: 1 16
Included observations: 16

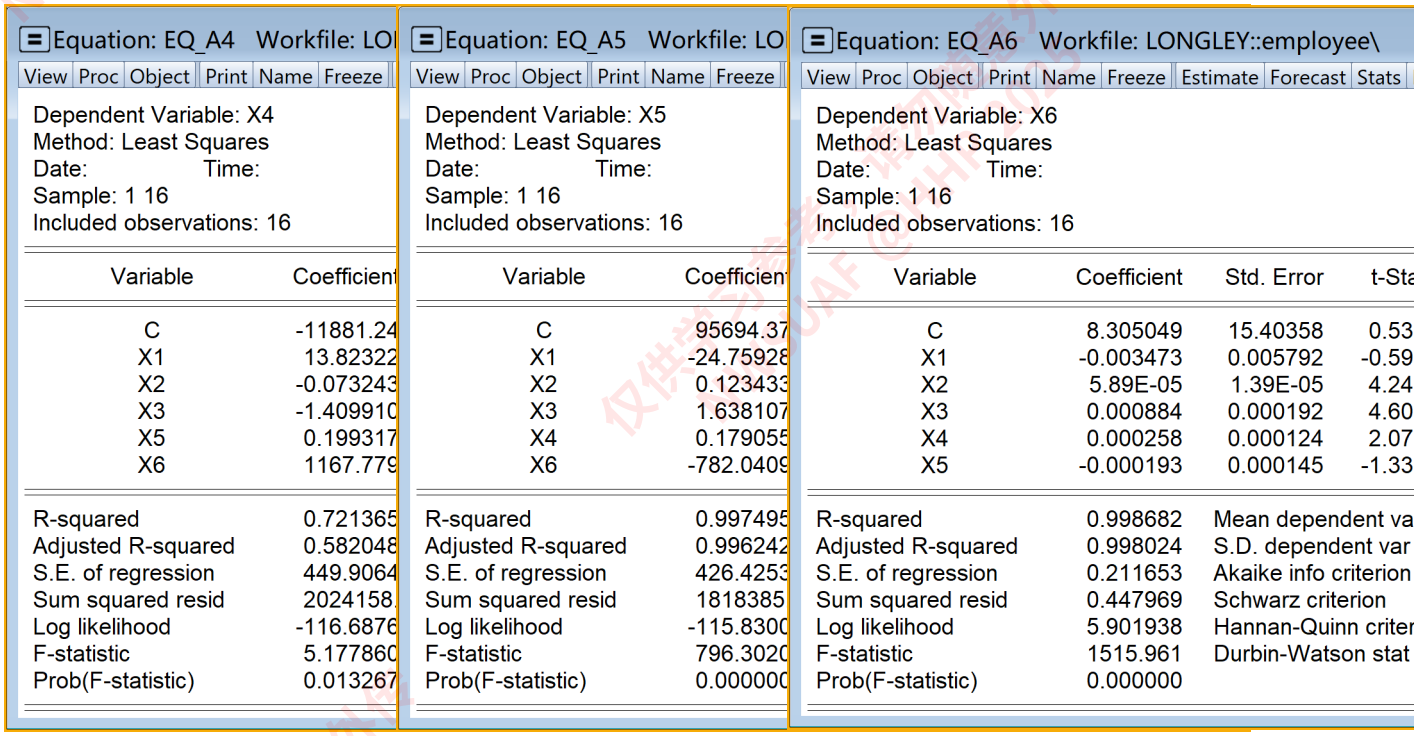
Variable	Coefficient	Std. Error	t-Sta
C	-28518.24	11446.89	-2.49
X1	9.649428	4.573555	2.10
X2	-0.064843	0.007054	-9.19
X4	-0.271381	0.109011	-2.48
X5	0.350986	0.095432	3.67
X6	768.5517	167.0507	4.60

R-squared	0.970255	Mean dependent va
Adjusted R-squared	0.955382	S.D. dependent var
S.E. of regression	197.3861	Akaike info criterion
Sum squared resid	389612.8	Schwarz criterion
Log likelihood	-103.5056	Hannan-Quinn criter
F-statistic	65.23778	Durbin-Watson stat
Prob(F-statistic)	0.000000	

(c) A3

图 5.6: 辅助回归方程 A1-A3

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025



Equation: EQ_A4 Workfile: LO

View Proc Object Print Name Freeze

Dependent Variable: X4
Method: Least Squares
Date: Time:
Sample: 1 16
Included observations: 16

Variable	Coefficient
C	-11881.24
X1	13.82322
X2	-0.073243
X3	-1.409910
X5	0.199317
X6	1167.779

R-squared	0.721365
Adjusted R-squared	0.582048
S.E. of regression	449.9064
Sum squared resid	2024158.
Log likelihood	-116.6876
F-statistic	5.177860
Prob(F-statistic)	0.013267

Equation: EQ_A5 Workfile: LO

View Proc Object Print Name Freeze

Dependent Variable: X5
Method: Least Squares
Date: Time:
Sample: 1 16
Included observations: 16

Variable	Coefficient
C	95694.37
X1	-24.75928
X2	0.123433
X3	1.638107
X4	0.179055
X6	-782.0409

R-squared	0.997495
Adjusted R-squared	0.996242
S.E. of regression	426.4253
Sum squared resid	1818385
Log likelihood	-115.8300
F-statistic	796.3020
Prob(F-statistic)	0.000000

Equation: EQ_A6 Workfile: LONGLEY::employee\

View Proc Object Print Name Freeze Estimate Forecast Stats

Dependent Variable: X6
Method: Least Squares
Date: Time:
Sample: 1 16
Included observations: 16

Variable	Coefficient	Std. Error	t-Sta
C	8.305049	15.40358	0.53
X1	-0.003473	0.005792	-0.59
X2	5.89E-05	1.39E-05	4.24
X3	0.000884	0.000192	4.60
X4	0.000258	0.000124	2.07
X5	-0.000193	0.000145	-1.33

R-squared	0.998682	Mean dependent va
Adjusted R-squared	0.998024	S.D. dependent var
S.E. of regression	0.211653	Akaike info criterion
Sum squared resid	0.447969	Schwarz criterion
Log likelihood	5.901938	Hannan-Quinn criter
F-statistic	1515.961	Durbin-Watson stat
Prob(F-statistic)	0.000000	

(a) A4 (b) A5 (c) A6

图 5.7: 辅助回归方程 A4-A6

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

提示

通过运行 EViews 代码，也可以实现上述操作。在命令视窗中依次输入并运行如下 EViews 代码：

```
****
''多重共线性诊断：构建辅助回归方程'
equation eq_a1.ls X1 c X2 X3 X4 X5 X6
equation eq_a2.ls X2 c X1 X3 X4 X5 X6
equation eq_a3.ls X3 c X1 X2 X4 X5 X6
equation eq_a4.ls X4 c X1 X2 X3 X5 X6
equation eq_a5.ls X5 c X1 X2 X3 X4 X6
equation eq_a6.ls X6 c X1 X2 X3 X4 X5
```

各个辅助回归方程的简要结果如下：

$$
\begin{array}{llllll}
\widehat{X1} = & +2044.58 & +0.00X2_i & +0.03X3_i & +0.01X4_i & \\
(s) & (533.3698) & (0.0009) & (0.0151) & (0.0075) & \\
(t) & (+3.83) & (+2.70) & (+2.11) & (+1.18) & \\
(cont.) & -0.02X5_i & -9.99X6_i & & & \\
(s) & (0.0063) & (16.6654) & & & \\
(t) & (-2.77) & (-0.60) & & & \\
(over) & n = 16 & \hat{\sigma} = 11.3529 & & & \\
(fit) & R^2 = 0.9926 & \bar{R}^2 = 0.9889 & & & \\
(Ftest) & F^* = 269.06 & p = 0.0000 & & &
\end{array}
\tag{5.24}
$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{array}{llll}
\widehat{X2} = & -480986.04 & +164.66X1_i & -13.79X3_i \quad -3.00X4_i \\
(s) & (148413.7872) & (60.9699) & (1.5002) \quad (1.7873) \\
(t) & (-3.24) & (+2.70) & (-9.19) \quad (-1.68) \\
(cont.) & +5.62X5_i & +10902.88X6_i & \\
(s) & (1.1804) & (2570.7562) & \\
(t) & (+4.76) & (+4.24) & \\
(over) & n = 16 & \hat{\sigma} = 2878.4838 & \\
(fit) & R^2 = 0.9994 & \bar{R}^2 = 0.9992 & \\
(Ftest) & F^* = 3575.03 & p = 0.0000 &
\end{array} \tag{5.25}
$$

$$
\begin{array}{llll}
\widehat{X3} = & -28518.24 & +9.65X1_i & -0.06X2_i \quad -0.27X4_i \\
(s) & (11446.8866) & (4.5736) & (0.0071) \quad (0.1090) \\
(t) & (-2.49) & (+2.11) & (-9.19) \quad (-2.49) \\
(cont.) & +0.35X5_i & +768.55X6_i & \\
(s) & (0.0954) & (167.0507) & \\
(t) & (+3.68) & (+4.60) & \\
(over) & n = 16 & \hat{\sigma} = 197.3861 & \\
(fit) & R^2 = 0.9703 & \bar{R}^2 = 0.9554 & \\
(Ftest) & F^* = 65.24 & p = 0.0000 &
\end{array} \tag{5.26}
$$

$$
\begin{array}{llll}
\widehat{X4} = & -11881.24 & +13.82X1_i & -0.07X2_i \quad -1.41X3_i \\
(s) & (33002.4231) & (11.7447) & (0.0437) \quad (0.5663) \\
(t) & (-0.36) & (+1.18) & (-1.68) \quad (-2.49) \\
(cont.) & +0.20X5_i & +1167.78X6_i & \\
(s) & (0.3276) & (561.6770) & \\
(t) & (+0.61) & (+2.08) & \\
(over) & n = 16 & \hat{\sigma} = 449.9064 & \\
(fit) & R^2 = 0.7214 & \bar{R}^2 = 0.5820 & \\
(Ftest) & F^* = 5.18 & p = 0.0133 &
\end{array} \tag{5.27}
$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{array}{llll}
\widehat{X5} = & +95694.37 & -24.76X1_i & +0.12X2_i \quad +1.64X3_i \\
(s) & (8682.0335) & (8.9319) & (0.0259) \quad (0.4454) \\
(t) & (+11.02) & (-2.77) & (+4.76) \quad (+3.68) \\
(cont.) & +0.18X4_i & -782.04X6_i & \\
(s) & (0.2943) & (587.1614) & \\
(t) & (+0.61) & (-1.33) & \\
(over) & n = 16 & \hat{\sigma} = 426.4253 & \\
(fit) & R^2 = 0.9975 & \bar{R}^2 = 0.9962 & \\
(Ftest) & F^* = 796.30 & p = 0.0000 &
\end{array} \tag{5.28}
$$

$$
\begin{array}{llll}
\widehat{X6} = & +8.31 & -0.00X1_i & +0.00X2_i \quad +0.00X3_i \\
(s) & (15.4036) & (0.0058) & (0.0000) \quad (0.0002) \\
(t) & (+0.54) & (-0.60) & (+4.24) \quad (+4.60) \\
(cont.) & +0.00X4_i & -0.00X5_i & \\
(s) & (0.0001) & (0.0001) & \\
(t) & (+2.08) & (-1.33) & \\
(over) & n = 16 & \hat{\sigma} = 0.2117 & \\
(fit) & R^2 = 0.9987 & \bar{R}^2 = 0.9980 & \\
(Ftest) & F^* = 1515.96 & p = 0.0000 &
\end{array} \tag{5.29}
$$

5.5.3.4 判定系数比较法

目标：分别 6 个辅助回归方程的判定系数 $R_j^2(j=1,\cdots,6)$，比较它们与主模型判定系数 R^2 的大小关系。

思路：分别提取辅助回归方程的 Eviews 报告中的判定系数。构造一个列向量（vector）对象（含 6 个元素），把 6 个辅助回归方程的判定系数依次放置其中。

> 提示
>
> **提示：**多重共线性问题的诊断依据是，辅助回归方程的判定系数 $R_j^2(j=1,\cdots,6)$ 大于主回归方程的判定系数 R^2

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

Eviews 操作：

（1）在命令视窗中依次输入并运行如下 EViews 代码：

```
'多重共线性诊断：辅助方程的判定系数比较
vector(6) r2
r2.fill(o=1) eq_a1.@r2   '提取辅助方程A1的判定系数
r2.fill(o=2) eq_a2.@r2   '提取辅助方程A2的判定系数
r2.fill(o=3) eq_a3.@r2   '提取辅助方程A3的判定系数
r2.fill(o=4) eq_a4.@r2   '提取辅助方程A4的判定系数
r2.fill(o=5) eq_a5.@r2   '提取辅助方程A5的判定系数
r2.fill(o=6) eq_a6.@r2   '提取辅助方程A6的判定系数
```

（2）在工作文件视窗下，可以看到如下新生成的对象，可以双击查看（见图 5.8 ）：包含六个辅助方程判定系数的向量对象r2

提示

（1）代码说明：

- 代码 `r2.fill(o=1)` 表示给列向量（vector）对象r2 的第 1 个元素（`o=1`）赋值（`.fill`）
- 代码 `eq_a1.@r2` 表示提取方程（equation）对象eq_a1 的判定系数 R^2（`.@r2`）

（2）参考资料：

- 向量对象操作命令`.fill` 的具体使用细节，请参看 Eviews在线帮助文档，网址为 http:// www.eviews.com/ help/ helpintro.html#page/ content%2Fvectorcmd-fill.html%23ww174507
- 方程结果提取命令 `@r2` 的具体使用细节，请参看 Eviews在线帮助文档，网址为：http:// www.eviews.com/ help/ helpintro.html#page/ content%2Fequationcmd-Equation.html%23ww178434
- 关于 EViews 方程对象的具体结果，内容请参看 Eviews在线帮助文档，网址为：http:// www.eviews.com/ help/ helpintro.html#page/ content/ Regress1-Equation_Output.html

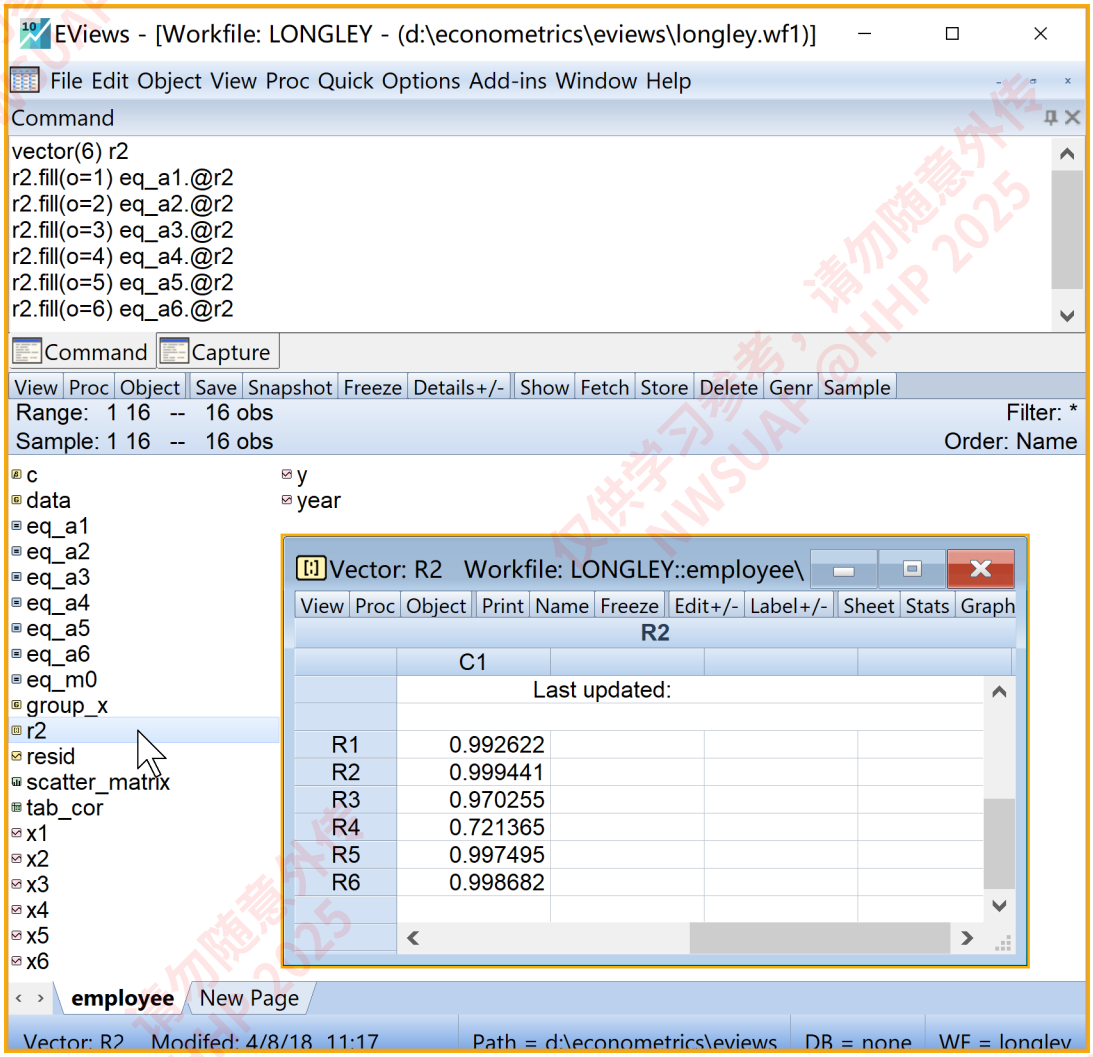


图 5.8: 辅助回归模型 A1~A6 的判定系数提取汇总操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

根据上述比较分析，判定系数比较法的结论初步认为模型式 5.17 可能存在严重的多重共线性问题理由如下：

表 5.5: 辅助方程的判定系数

辅助模型	模型形式	判定系数
A1	X1 ~ X2 + X3 + X4 + X5 + X6	0.99
A2	X2 ~ X1 + X3 + X4 + X5 + X6	1
A3	X3 ~ X1 + X2 + X4 + X5 + X6	0.97
A4	X4 ~ X1 + X2 + X3 + X5 + X6	0.72
A5	X5 ~ X1 + X2 + X3 + X4 + X6	1
A6	X6 ~ X1 + X2 + X3 + X4 + X5	1

5.5.3.5 方差膨胀因子（VIF）比较法

目标：分别得到 6 个辅助回归方程的方差膨胀因子 $VIF_j(j=1,\cdots,6)$，与参考值进行比较，得到相关结论。

思路：

- Eviews 手动计算：根据 6 个辅助回归方程的判定系数 R_j^2（见表 5.5 ），分别计算得到各自的方差膨胀因子 VIF_j。
- Eviews 方程提取：利用 Eviews 菜单功能 Coefficient Diagnostics ⇒ Variance Inflation Factors 一次性得到主回归模型的 Eviews 方程eq_m0 全部方差膨胀因子 $VIF_j(j=1,\cdots,6)$

> 提示
>
> **提示：**
>
> - 操作提示：构造一个列向量（vector）对象（含 6 个元素），用来装载 6 个辅助回归方程的方差膨胀因子 $VIF_j(j=1,\cdots,(6))$。
> - 诊断提示：辅助回归方程的方差膨胀因子中如果 $VIF_j \in [10,100]$ 表明中度多重共线性；如果 $VIF_j \geq 100$ 表明严重多重共线性
> - 公式提示：辅助回归方程方差膨胀因子的理论计算公式为

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$VIF_j = \frac{1}{1-R_j^2}, (j = 1, \cdots, 6) \tag{5.30}$$

（方法 1）Eviews 命令操作（手动计算实现，具体见图 5.9 ）：

在命令视窗中依次输入并运行如下 EViews 代码：

```
'多重共线性诊断：方差膨胀因子
vector(6) unit=1                    '为计算做准备
vector vif=@ediv(unit,(unit-r2))    '利用公式计算VIF
```

在工作文件视窗下，可以看到如下新生成的对象，可以双击查看（见图 5.9 ）：包含 6 个辅助回归方程各自方差膨胀因子 (VIF_i) 的向量对象vif

提示

（1）代码说明：

- 代码 `vector(6) unit=1` 表示生成一个包含 6 个元素全为 1、名为 `unit` 的列向量`unit`。
- 代码 `vector vif=\@ediv(unit,(unit-r2))` 表示生成一个名为 vif 列向量（vector）对象`vif`，并使该列向量等于另外两个向量之除，也即 `=\@ediv(unit,(unit-r2))`。
- 代码`\@ediv()` 为 Eviews 矩阵运算函数，用于两个向量元素间相除的计算。

（2）参考资料：

- `@ediv` 函数代码的具体使用细节请参看 Eviews在线帮助文档

（方法 2）Eviews 菜单操作（菜单操作实现，具体见图 5.10 ）：

- 打开主方程：双击方程（equation）对象`eq_m0`
- 进入功能菜单：
  - 选择分析菜单：⇒ `View` ⇒ `Coefficient Diagnostics` ⇒ `Variance Inflation Factors`
  - 另存为表格 (table) 对象：点击 `Freeze`

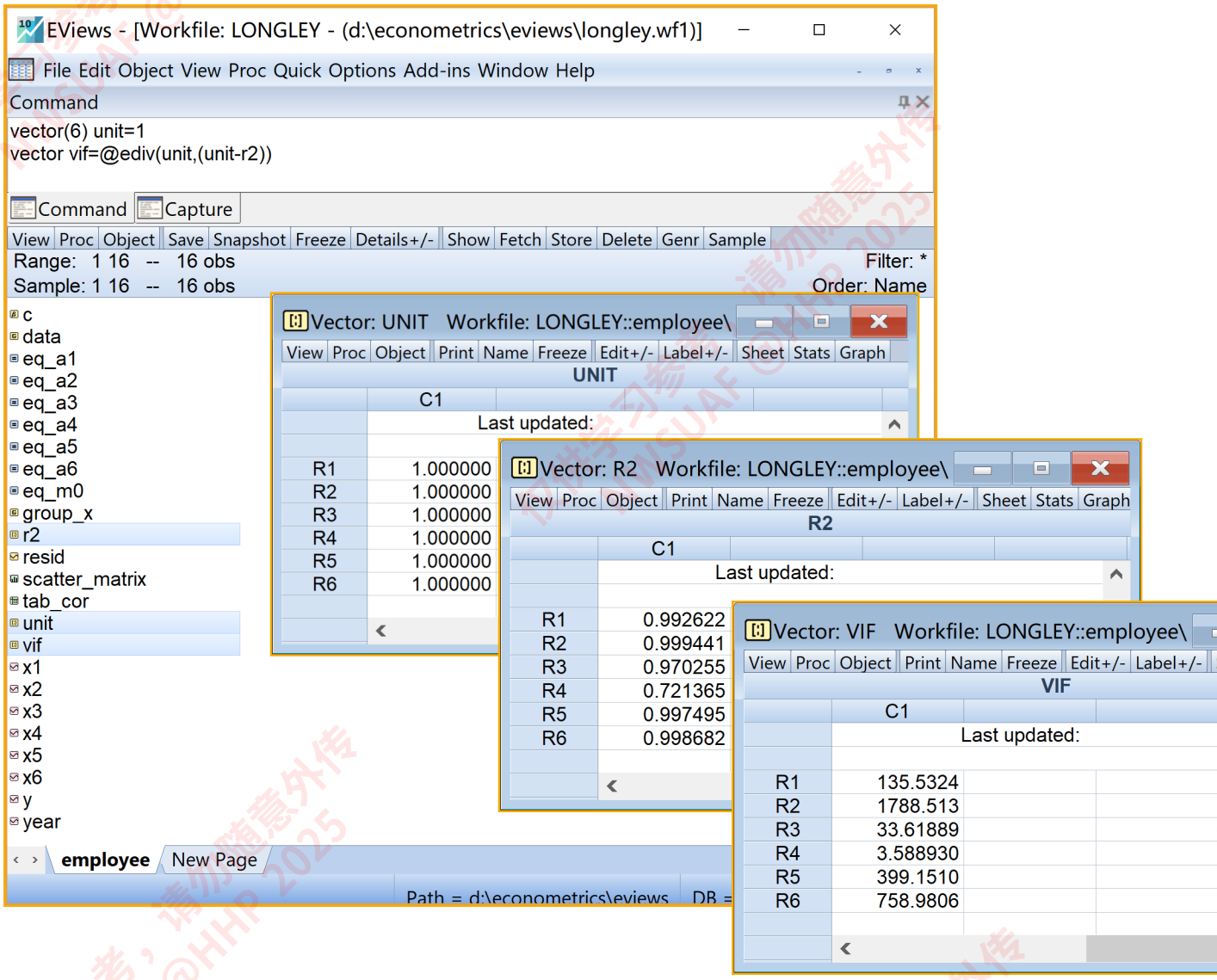


图 5.9: 辅助回归模型 A1~A6 的 VIF 手动提取操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 命名并保存表格 (table) 对象：点击 `name`(建议为 tab_vif)
- 查看结果：双击`tab_vif`

提示

（1）通过运行 EViews 代码，也可以实现上述操作。在命令视窗中依次输入并运行如下 EViews 代码：

```
****
`'直接冻结主方程的一个view（varinf）为VIF表格`
`freeze(tab_vif) eq_m0.varinf`
```

（2）代码说明：`.varinf` 为方程视图 (Equation View) 的 EViews 调用命令。
（3）参考资料：

- EViews 关于方程视图 (Equation View)`.varinf` 调用命令的在线说明
- EViews 关于方差膨胀因子的说明文档

多重共线性诊断结果。根据诊断经验标准，主回归方程 M0 的方差膨胀因子中，如果 $VIF_j \in [10, 100]$ 表明中度多重共线性；如果 $VIF_j \geq 100$ 表明严重多重共线性。

表 5.6: 主模型 M0 系数 VIF 诊断结果

自变量	VIF	>100	10~100
X1	135.5	很严重	...
X2	1788.5	很严重	...
X3	33.6	...	中等
X4	3.6	...	...
X5	399.1	很严重	...
X6	759.0	很严重	...

根据计算结果汇总（见表 5.6 ），可以认为主模型 M0 存在较严重的多重共线性问题。其中 VIF 值大于 100 的系数就包括 X1、X2、X5、X6。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

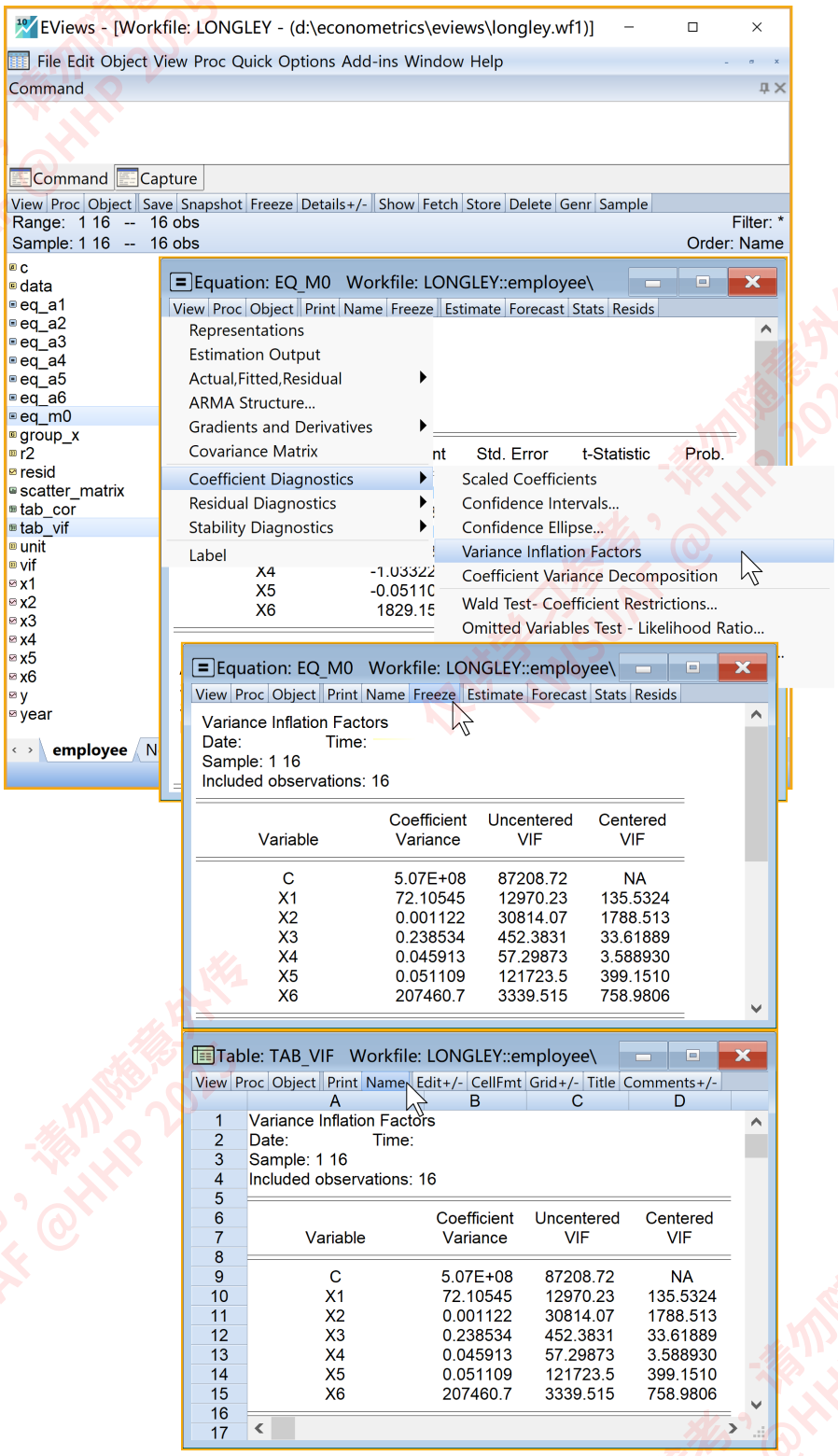


图 5.10: 辅助回归模型 A1~A6 的 VIF 系统提取操作

5.5.3.6 容忍度（TOL）比较法

目标：分别得到 6 个辅助回归方程的容忍度 $TOL_j(j=1,\cdots,6)$，与参考值进行比较，得到相关结论。

思路：根据 6 个辅助回归方程的判定系数 R_j^2（见表 5.5 ），分别计算得到各自的容忍度 TOL_j。

提示

提示：

- 操作提示：构造一个列向量（vector）对象（含 6 个元素），用来装载 6 个辅助回归方程的容忍度 $TOL_j(j=1,\cdots,(6))$
- 诊断提示：辅助回归方程的的容忍度如果 $TOL_j \in [0.01, 0.1]$ 表明中度多重共线性；如果 $TOL_j \leq 0.01$ 表明存在严重的多重共线性
- 公式提示：辅助回归方程容忍度的理论计算公式为

$$TOL_j = 1 - R_j^2 = \frac{1}{VIF_j}, (j=1,\cdots,6) \tag{5.31}$$

在命令视窗中依次输入并运行如下 EViews 代码：

```
'多重共线性诊断：容忍度
vector tol=unit-r2                    '利用公式计算TOL
```

在工作文件视窗下，可以看到如下新生成的对象，可以双击查看（见图 5.11 ）：

- 辅助方程容忍度 TOL_i 组成的向量对象tol

5.5.3.7 回归系数方差分解法

目标：利用 Eviews 的共线性诊断菜单，进行回归系数方差分解（Coefficient Variance Decomposition）。分析病态数（Condition Numbers，K）和方差分解比率（variance-decomposition proportions，VDP），并与参考值进行比较，得到相关结论

思路：特征值（Eigenvalue）；病态数（condition number，K）；方差分解比率（variance-decomposition proportions，VDP）

仅供学习参考，请勿随意外传 NWSUAF @HHR 2025

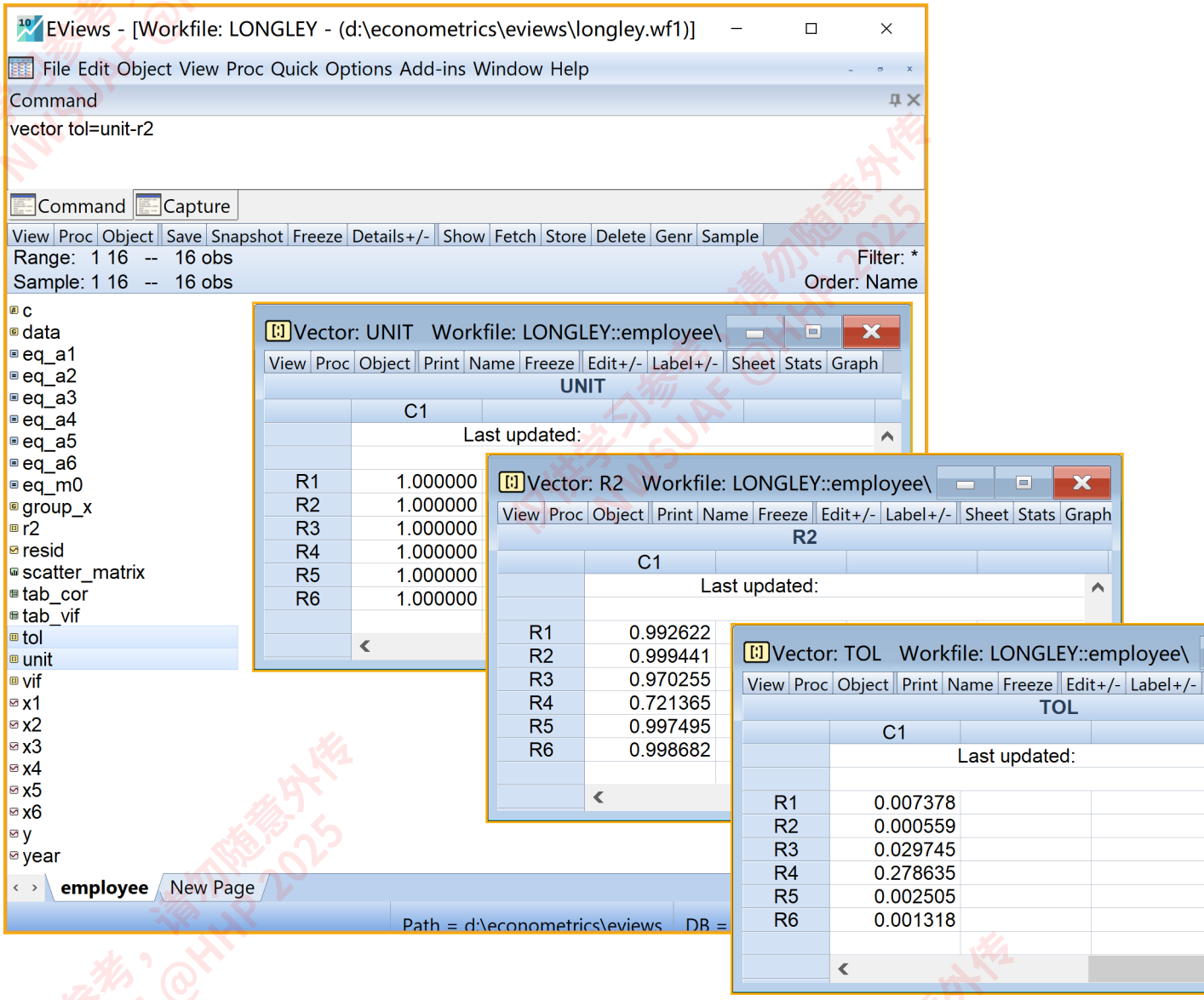


图 5.11: 辅助回归模型 A1~A6 的 TOL 手动提取操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

提示

提示：

- 操作提示：用 Eviews 的共线性诊断菜单 ⇒ View ⇒ Coefficient Diagnostics ⇒ Coefficient Variance Decomposition
- 诊断提示：若发现至少一个病态数 $K \leq (0.001$，则表明存在严重多重共线性；观察病态数最小时所对应的方差分解比率，如果有多个斜率系数的 $VDP \geq 0.5$ ，则表明它们存在严重的多重共线性

$$\widehat{var\_cov}(\ ) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

Eviews 操作（菜单操作实现，具体见图 5.12 ）：

（1）打开主方程：双击方程（equation）对象eq_m0

（2）进入功能菜单：

- 选择分析菜单：⇒ View ⇒ Coefficient Diagnostics ⇒ Coefficient Variance Decomposition
- 另存为表格 (table) 对象：点击 Freeze
- 命名并保存表格 (table) 对象：点击 name(建议为 tab_cvd)
- 查看结果：双击tab_cvd

提示

（1）通过运行 EViews 代码，也可以实现上述操作。在命令视窗中依次输入并运行如下 EViews 代码：

```
****
`'多重共线性诊断：回归系数方差分解法`
`freeze(tab_vcd) eq_m0.cvardecomp `
```

（2）代码说明：

- 系数方差分解表视图命令.cvardecomp 属于一种方程视图 (Equation View) 的 EViews 调用命令，可以得到方程的系数方差分解表。

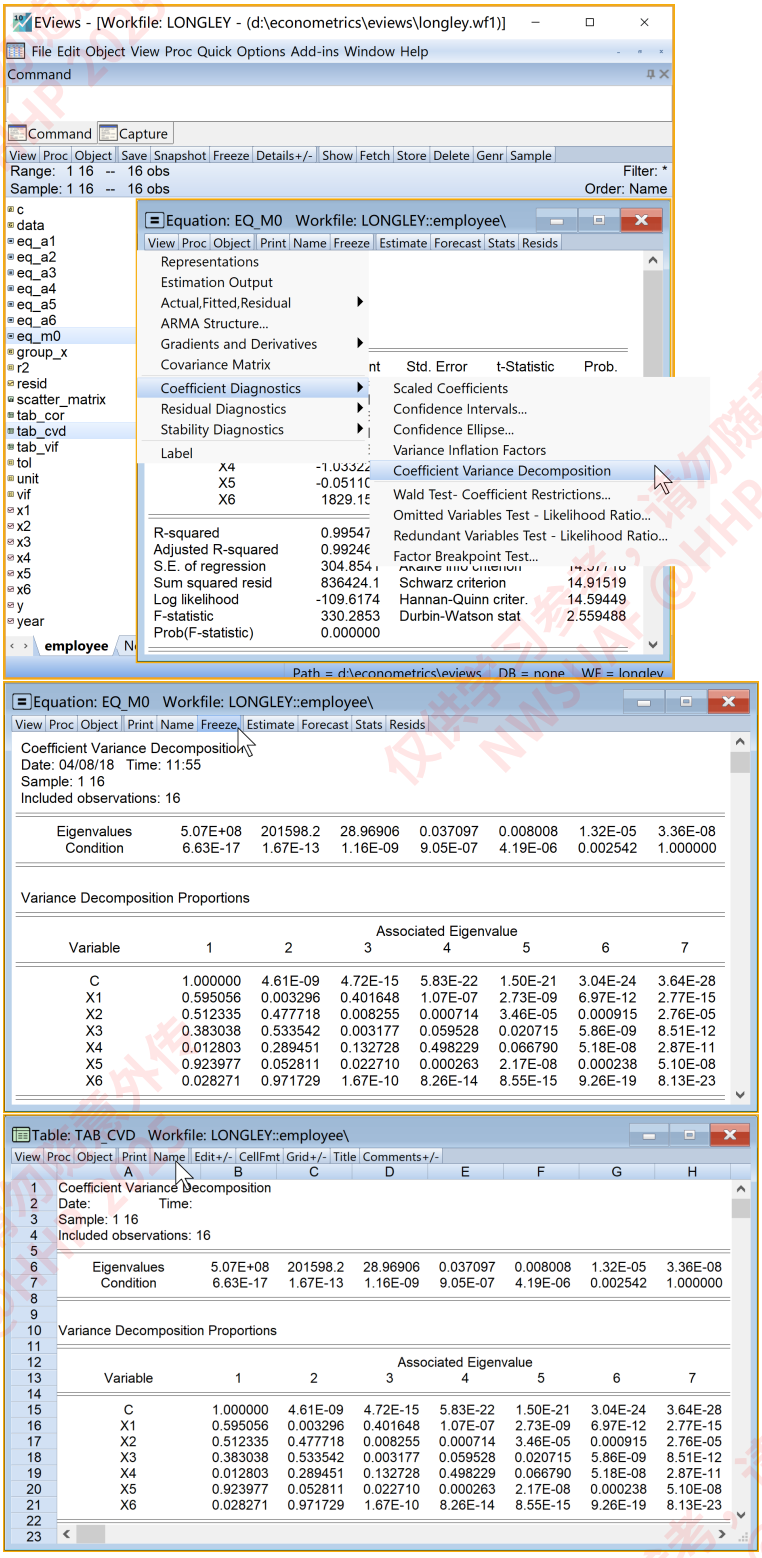


图 5.12: 主回归模型系数方差分解的操作过程

- 系数方差分解表视图命令`.cvardecomp` 命令的具体使用方法可参阅EViews 在线帮助文档

（3）参考资料：

- 系数方差分解诊断方法由 Belsley, Kuh and Welsch (BKW) 2004 提出，具体细节可以参考Eviews 帮助文档
- Eviews 软件中病态数的计算是基于矩阵 $(\mathbf{X}'\mathbf{X})^{-1}$，而不是基于矩阵 $\mathbf{X}$

多重共线性诊断分析，具体 Eviews 报告见图 5.13 。

Coefficient Variance Decomposition
Date: Time:
Sample: 1 16
Included observations: 16

Eigenvalues	5.07E+08	201598.2	28.96906	0.037097	0.008008	1.32E-05	3.36E-08
Condition	6.63E-17	1.67E-13	1.16E-09	9.05E-07	4.19E-06	0.002542	1.000000

Variance Decomposition Proportions

	Associated Eigenvalue						
Variable	1	2	3	4	5	6	7
C	1.000000	4.61E-09	4.72E-15	5.83E-22	1.50E-21	3.04E-24	3.64E-28
X1	0.595056	0.003296	0.401648	1.07E-07	2.73E-09	6.97E-12	2.77E-15
X2	0.512335	0.477718	0.008255	0.000714	3.46E-05	0.000915	2.76E-05
X3	0.383038	0.533542	0.003177	0.059528	0.020715	5.86E-09	8.51E-12
X4	0.012803	0.289451	0.132728	0.498229	0.066790	5.18E-08	2.87E-11
X5	0.923977	0.052811	0.022710	0.000263	2.17E-08	0.000238	5.10E-08
X6	0.028271	0.971729	1.67E-10	8.26E-14	8.55E-15	9.26E-19	8.13E-23

Eigenvectors

	Associated Eigenvalue						
Variable	1	2	3	4	5	6	7
C	0.999994	-0.003403	0.000287	2.82E-06	-9.73E-06	-1.08E-05	-2.34E-06
X1	-0.000291	-0.001086	0.999861	-0.014433	-0.004961	-0.006168	-0.002438
X2	1.07E-06	5.16E-05	-0.000565	0.004645	-0.002202	0.278758	-0.960347
X3	1.34E-05	0.000795	-0.005115	-0.618683	0.785518	-0.010288	-0.007777
X4	1.08E-06	0.000257	-0.014504	-0.785263	-0.618815	-0.013425	-0.006268
X5	-9.66E-06	-0.000116	-0.006330	0.019049	-0.000372	-0.960193	-0.278617
X6	-0.003403	-0.999993	-0.001094	-0.000679	0.000471	0.000121	-2.24E-05

图 5.13: 主回归模型系数方差分解的 Eviews 报告

5.5.4 多重共线性模型矫正

5.5.4.1 经济学和实践观察法

目标：根据某种考虑或规则，删除特定变量，重新估计回归模型，得到相关结论。

思路：面对严重的共线性，最简单的方法就是去掉某些变量，但剔除变量会导致设定误差。实际中需要权衡利弊。

> 提示
>
> 理论提示：酌情删除：经济学和实践观察法。利用先验信息（成为研究领域的专家！）酌情删除特定变量，减弱模型的多重共线性问题。那怎样才能获得先验信息呢？它往往源自经验研究工作或者有关基础理论。
> 操作提示：
>
> - 改用真实 GNP，不用名义 GNP（X2）：将名义 GNP（X2）除以价格指数 CPI（X1）
> - 留下 14 岁以上非机构人口数（X5），去掉时间趋势（X6）：14 岁以上非机构人口数随时间不断增长，它与时间趋势变量高度相；而且时间趋势变量还和很多其他变量高度相关。
> - 去掉失业人数变量（X3）：可能失业率是劳动力市场状况的一个更好的度量指标，但我们没有这方面的数据，而失业人数也没有充分的理由包括进来。

Eviews 操作（菜单操作实现，见图 5.15 ）：

（1）变量变换，生成实际 GNP（= 名义 GNP/CPI）（见图 5.15 ）

- 命令视窗（Command）输入命令：`series x_rgnp=X2/X1`
- 运行命令：命令行中按 Enter 键
- 查看计算结果

（2）引导设置 Equation Estimation⇒specification

- Equation specification：输入命令 Y c x_rgnp X3 X4 X5
- Estimation settings：
  - Method: 下拉选择 `LS - Least Squares (NLS and ARMA)`

- Sample: **默认设置**

- 点击 `OK`

（3）模型命名：建议为 `eq_adj_man`

酌情法矫正的分析结果如图：

> 提示
>
> 通过运行 EViews 代码，也可以实现上述操作。在命令视窗中依次输入并运行如下 EViews 代码：
>
> ```
> ****
> `'多重共线性矫正：经济学和实践观察法`
> `series x_rgnp=x2/x1                                    ' 构造新变量`
> `equation eq_adj_man.ls y c x2/x1 x3 x4 x5              '回归新方程`
> ```

利用酌情删除法的矫正多重共线性问题后的简要回归报告如下：

$$\begin{aligned}\widehat{Y}_i = &+ \hat{\beta}_1 + \hat{\beta}_2 \frac{X2}{X1})_i + \hat{\beta}_3 X3_i + \hat{\beta}_4 X4_i \\ &+ \hat{\beta}_5 X5_i\end{aligned} \tag{5.32}$$

$$\begin{array}{llll}\widehat{Y} = & +47497.87 & +73.95X2/X1_i & -0.54X3_i \\ (s) & (14656.2227) & (21.7092) & (0.3180) \\ (t) & (+3.24) & (+3.41) & (-1.69) \\ (cont.) & -0.73X4_i & -0.05X5_i & \\ (s) & (0.3010) & (0.1966) & \\ (t) & (-2.42) & (-0.27) & \\ (over) & n=16 & \hat{\sigma}=498.5883 & \\ (fit) & R^2=0.9852 & \bar{R}^2=0.9798 & \\ (Ftest) & F^*=183.31 & p=0.0000 & \end{array} \tag{5.33}$$

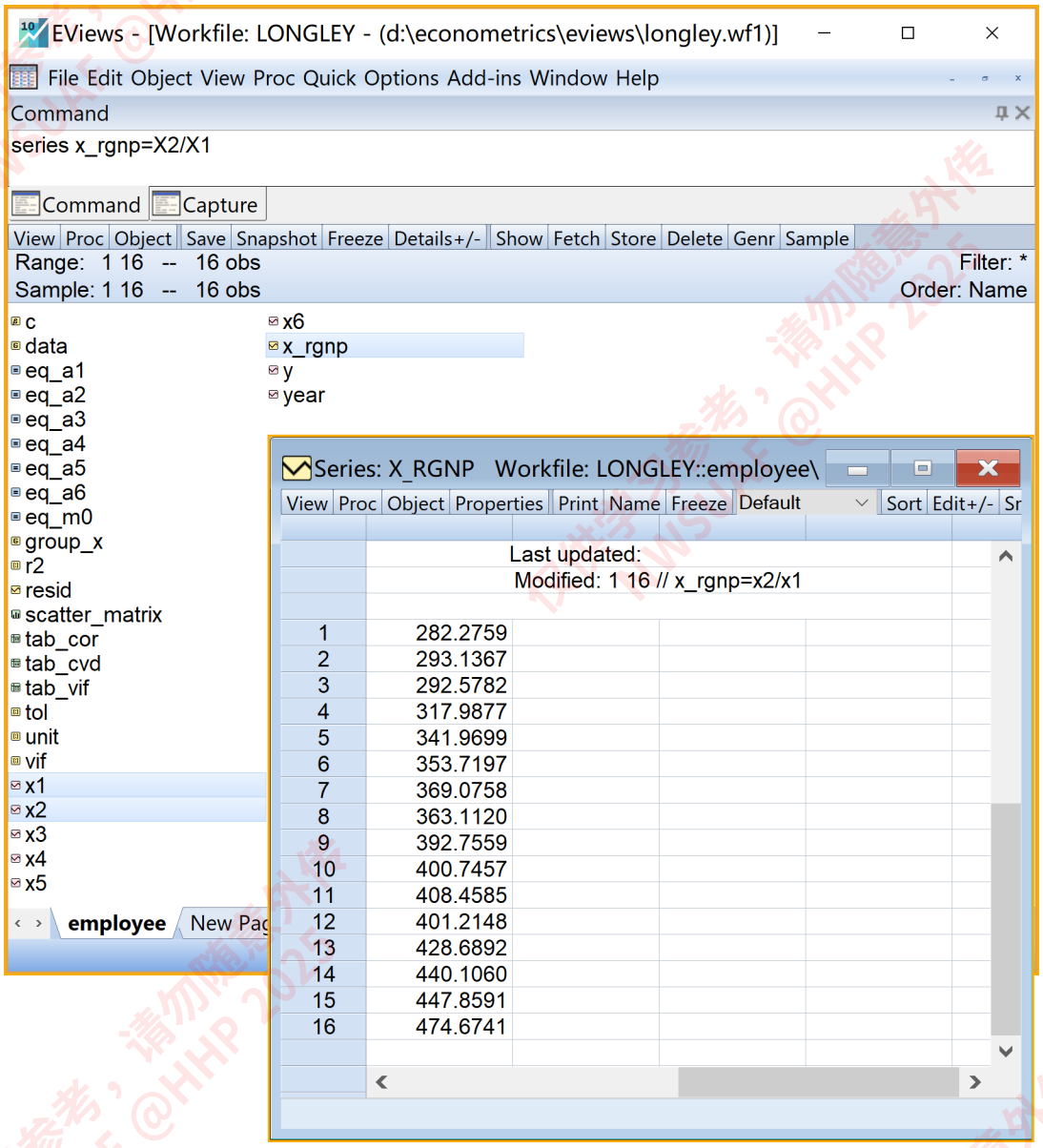


图 5.14: 多重共线性模型的变量变换操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

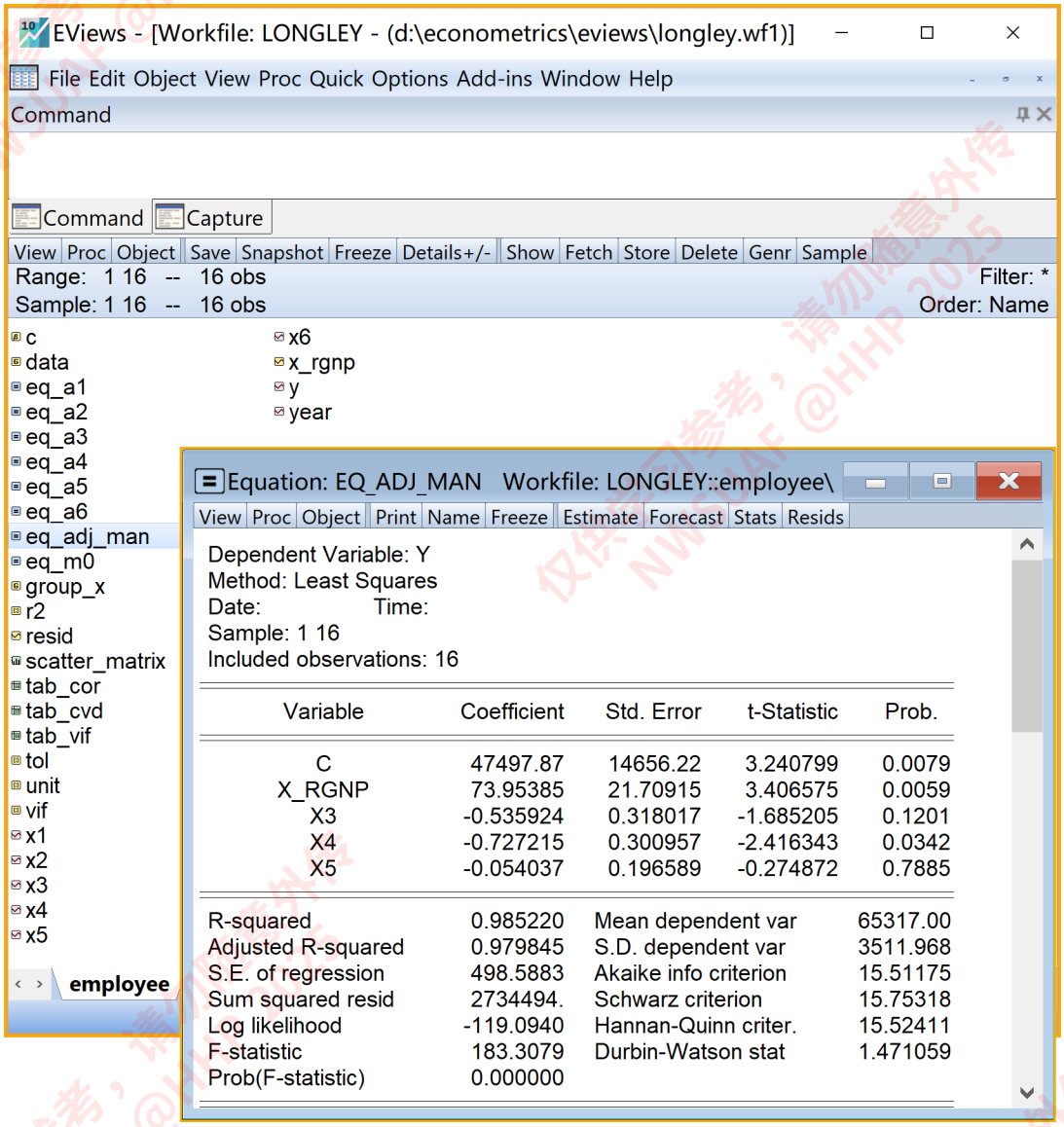


图 5.15: 多重共线性问题模型的矫正：酌情删除法

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

5.5.4.2 变量变换法

目标：根据某种考虑或规则，删除特定变量，重新估计回归模型，得到相关结论。

思路：面对严重的共线性，最简单的方法就是去掉某些变量，但剔除变量会导致设定误差。实际中需要权衡利弊。

理论提示 1：一阶差分法 (first difference form) 巧妙删除变量。

模型中两个解释变量 $X_{k,i}$ 和 $X_{w,i}$ 可能导致高度多重共线性，但是分别对二者进行一阶差分，再进行回归建模，新模型可能的多重共线性问题很可能大大缓解！

具体变换如下:

$$Y_t = \beta_1 + \beta_2 X_{2,t} + \beta_3 X_{3,t} + u_t \quad \text{(原模型)} \tag{5.34}$$

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + u_{t-1} \quad \text{(滞后 1 阶变量模型)} \tag{5.35}$$

$$Y_t - Y_{t-1} = \beta_2 (X_{2,t} - X_{2,t-1}) + \beta_3 (X_{3,t} - X_{3,t-1}) + (u_t - u_{t-1}) \quad \text{(一阶差分模型)} \tag{5.36}$$

$$Y_t^* = \beta_2 X_{2,t}^* + \beta_3 X_{3,t}^* + v_t \quad \text{(精简化模型)} \tag{5.37}$$

需要注意的是，“按下葫芦浮起瓢”，治疗比疾病更糟糕？差分变换 Y_{t-1} 减少了自由度；同时 $v_t = (u_t - u_{t-1})$ 可能带来异方差问题。

理论提示 2：比率变换法 (ratio transformation) 巧妙删除变量。

模型中两个解释变量 $X_{k,i}$ 和 $X_{w,i}$ 可能导致高度多重共线性，如果可以用其中的一个变量同时对模型其他变量进行比率变换，而且如果变换后的所有变量还能具有经济学含义，那么理论上将至少消掉一个回归元，从而大大缓解甚至消除多重共线性问题！

具体变换如下:

消费支出决定案例：Y_t 为以真实价格表示的消费支出，$X_{2,t}$ 表示 GDP，$X_{3,t}$ 表示总人口。

$$Y_t = \beta_1 + \beta_2 X_{2,t} + \beta_3 X_{3,t} + u_t \quad \text{(原模型)} \tag{5.38}$$

$$\frac{Y_t}{X_{3,t}} = \frac{\beta_1}{X_{3,t}} + \beta_2 \frac{X_{2,t}}{X_{3,t}} + \frac{u_t}{X_{3,t}} \quad \text{(比率变换模型)} \tag{5.39}$$

$$Y_t^* = \beta_1^* + \beta_2^* X_{2,t}^* + v_t \quad \text{(精简化模型)} \tag{5.40}$$

同样需要注意的是，“按下葫芦浮起瓢”，治疗比疾病更糟糕？$v_t = \frac{u_t}{X_{3,t}}$ 可能带来异方差问题。

5.5.4.3 逐步最小二乘回归法

目标：根据某种考虑或规则，删除特定变量，重新估计回归模型，得到相关结论。

思路：面对严重的共线性，最简单的方法就是去掉某些变量，但剔除变量会导致设定误差。实际中需要权衡利弊。

理论提示：逐步最小二乘回归法（Stepwise Least Squares Regression）通过多个统计标准，可以自动判断模型该引入还是删除某些自变量 X。这些统计标准主要包括分析引入新变量对回归平方和 ESS 的贡献大小，及 F 检验等。

- 前向逐步回归法（Stepwise-Forwards），是从一个简化模型（很少 X 变量）开始，再逐步引入新的 X 变量，直至达到某个统计标准（主要是 p 值标准）
- 后向逐步回归法（Stepwise-Backwards），是从一个完全模型（全部 X 变量）开始，对模型逐步删除某些 X 变量，直至剩余变量都达到某个统计标准（主要是 p 值标准）

Eviews 操作（以后向逐步回归法为例，见图 5.16 ）：

（1）依次选择 ⇒ Quick ⇒ Estimation Equation

（2）引导设置 Equation Estimation ⇒ Specification

- 输入因变量和选择一直保留的自变量（Dependent variable followed by list of always included regressors）：Y （此处如果仅填 Y 变量，则任何 X 变量都没有强制一定要留在模型中）
- 输入自变量（List of search regressors）：Y c X1 X2 X3 X4 X5 X6
- 估计方法（Estimation settings）：

- Method：下拉选择 `STEPLS - stepwise Least Squares`
- Sample：默认设置

（3）引导设置 Equation Estimation ⇒ Options

- 方法设置（Selection Method）：
  - 下拉选择 Stepwise
  - 点击选择 Backwards
- 标准设置（Stopping Criteria）：
  - 点击选择 p-value
  - 填写（p-value forwards）：0.05
  - 填写（p-value backwards）：0.05
  - 其他：默认设置
- 权重设置（Weights）：默认设置
- 步数设置（Maximum steps）：默认设置
- 完成设置：点击 `OK`

（4）模型命名：建议为 `eq_adj_step`

（5）查看分析报告（见图 5.17 ）

提示

（1）通过运行 EViews 代码，也可以实现上述操作。在命令视窗中依次输入并运行如下 EViews 代码：

```
****
`'多重共线性矫正：逐步最小二乘回归法（后向逐步回归，p值标准为0.05）`
`equation eq_adj_step.stepls(back,btol=0.05) Y @ c X1 X2 X3 X4 X5 X6 `
```

（2）代码说明：

- 回归方法命令`.stepls(back,btol=0.05)`，可以引用逐步回归分析方法。
- 逐步回归法是方程对象（Equation）的一种分析方法（method），可以参看 EViews 的方程方法（Equation Methods）列表说明.

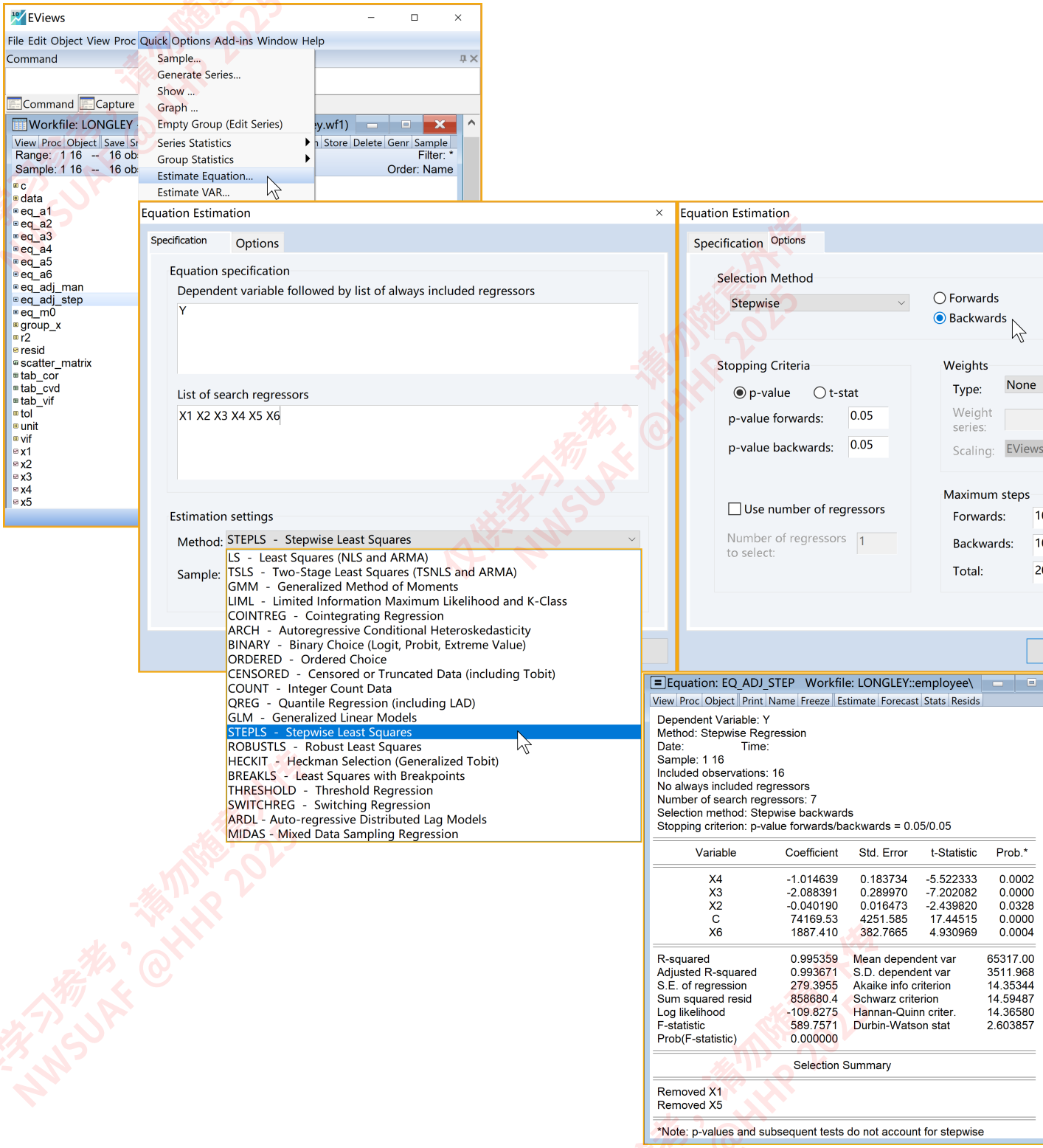


图 5.16: 多重共线性模型的矫正：后向逐步回归法

Dependent Variable: Y
Method: Stepwise Regression
Date: Time:
Sample: 1 16
Included observations: 16
No always included regressors
Number of search regressors: 7
Selection method: Stepwise backwards
Stopping criterion: p-value forwards/backwards = 0.05/0.05

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
X4	-1.014639	0.183734	-5.522333	0.0002
X3	-2.088391	0.289970	-7.202082	0.0000
X2	-0.040190	0.016473	-2.439820	0.0328
C	74169.53	4251.585	17.44515	0.0000
X6	1887.410	382.7665	4.930969	0.0004

R-squared	0.995359	Mean dependent var	65317.00
Adjusted R-squared	0.993671	S.D. dependent var	3511.968
S.E. of regression	279.3955	Akaike info criterion	14.35344
Sum squared resid	858680.4	Schwarz criterion	14.59487
Log likelihood	-109.8275	Hannan-Quinn criter.	14.36580
F-statistic	589.7571	Durbin-Watson stat	2.603857
Prob(F-statistic)	0.000000		

Selection Summary

Removed X1
Removed X5

*Note: p-values and subsequent tests do not account for stepwise selection.

图 5.17: 多重共线性模型的后向逐步回归法矫正报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 回归方法命令`.stepls(back,btol=0.05)`的具体使用方法，可参阅EViews 在线帮助文档

（3）参考资料：

- 逐步回归分析方法，可参看 Eviews在线说明文档

逐步回归法矫正后的简要回归报告如下：

$$\begin{aligned}\widehat{Y}_i = &+ \hat{\beta}_1 + \hat{\beta}_2 X2_i + \hat{\beta}_3 X3_i \\ &+ \hat{\beta}_4 X4_i + \hat{\beta}_5 X5_i\end{aligned} \tag{5.41}$$

$$\begin{array}{llll}\widehat{Y} = & +74169.53 & -0.04X2_i & -2.09X3_i \\ (s) & (4251.5849) & (0.0165) & (0.2900) \\ (t) & (+17.45) & (-2.44) & (-7.20) \\ (cont.) & -1.01X4_i & +1887.41X6_i & \\ (s) & (0.1837) & (382.7665) & \\ (t) & (-5.52) & (+4.93) & \\ (over) & n = 16 & \hat{\sigma} = 279.3955 & \\ (fit) & R^2 = 0.9954 & \bar{R}^2 = 0.9937 & \\ (Ftest) & F^* = 589.76 & p = 0.0000 & \end{array} \tag{5.42}$$

5.5.4.4 主成分法（自学）

5.6 附录：prg 源代码

实际操作中，在 EViews 命令视窗中逐条输入代码，既容易出错，又不便于维护这些代码，还不能进行代码的重复使用（在第一章的节 1.8 中已经论述）。

因此，读者可以创建一个.prg 编程文件，并在其中编写 EViews 代码，进行管理、维护、运行和分析。下面代码按本章主要实验步骤编写，读者可以用于本章的 EViews 编程参考，进行实验练习。

```
'=========================================================================================================
'说明：以下为EViews编程文件longley.prg的代码
'将展示第五章中“郎利就业数据案例”主要分析步骤的“批量式命令驱动”实现方法(::
'其中，符号'起始的行，为注释行，其他为EViews命令行。
'=========================================================================================================

'创建工作文件（工作文件名=longley，子页命名=employee），无结构无日期，样本数为16
wfcreate(wf=longley,page=employee)  u 16

'导入外部数据，路径为d:\github\books\data\Lab5-longley-short-origin.xlsx
import d:\github\books\data\Lab5-longley-short-origin.xlsx

'生成线性回归模型的方程对象
equation eq_m0.ls Y c x1 x2 x3 x4 x5 x6   '回归方程
scalar r2_m0=eq_m0.@r2

'多重共线性诊断：计算自变量的相关系数表格和散点矩阵图
group varx x1 x2 x3 x4 x5 x6          '构建只含X的group
freeze(tab_cor) varx.cor              '把group的相关系数矩阵表视图保存为表格
graph scatter_matrix.scatmat varx     '绘制散点矩阵图

'多重共线性诊断：构建辅助回归方程
equation eq_a1.ls X1 c X2 X3 X4 X5 X6
equation eq_a2.ls X2 c X1 X3 X4 X5 X6
equation eq_a3.ls X3 c X1 X2 X4 X5 X6
equation eq_a4.ls X4 c X1 X2 X3 X5 X6
equation eq_a5.ls X5 c X1 X2 X3 X4 X6
equation eq_a6.ls X6 c X1 X2 X3 X4 X5

'多重共线性诊断：辅助方程的判定系数比较
vector(6) r2
r2.fill(o=1) eq_a1.@r2   '提取辅助方程A1的判定系数
r2.fill(o=2) eq_a2.@r2   '提取辅助方程A2的判定系数
r2.fill(o=3) eq_a3.@r2   '提取辅助方程A3的判定系数
r2.fill(o=4) eq_a4.@r2   '提取辅助方程A4的判定系数
```

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
r2.fill(o=5) eq_a5.@r2   '提取辅助方程A5的判定系数
r2.fill(o=6) eq_a6.@r2   '提取辅助方程A6的判定系数

'多重共线性诊断：方差膨胀因子
vector(6) unit=1                    '为计算做准备
vector vif=@ediv(unit,(unit-r2))    '利用公式计算VIF
freeze(tab_vif) eq_m0.varinf        '直接冻结主方程的一个view（varinf）为VIF表格

'多重共线性诊断：容忍度
vector tol=unit-r2                  '利用公式计算TOL


'多重共线性诊断：回归系数方差分解法
freeze(tab_vcd) eq_m0.cvardecomp    '

'多重共线性矫正：经济学和实践观察法
series x_rgnp=x2/x1                                 ' 构造新变量
equation eq_adj_man.ls y c x_rgnp x3 x4 x5          '回归新方程

'多重共线性矫正：逐步最小二乘回归法（后向逐步回归，p值标准为0.05）
equation eq_adj_step.stepls(back,btol=0.05) Y @ c X1 X2 X3 X4 X5 X6

' ================================================================================
```

5.7 实验作业

就业情况的郎利数据：表 5.1 给出美国 1947-1961 年间就业情况及主要影响因素的数据表。

表 5.7: 美国劳动力市场的郎利调研数据 (n=47)

X1	x1	x2	x3	x4	x5	x6	Ydata1	Ydata2	Ydata3
id	x1	x2	x3	x4	x5	x6	2015014495	2016010317	2016011222
name	x1	x2	x3	x4	x5	x6	刘琳	王雪明	韩双瑞
class	x1	x2	x3	x4	x5	x6	保险 1601	保险 1601	保险 1601

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 5.7: 美国劳动力市场的郎利调研数据 (n=47)

X1	x1	x2	x3	x4	x5	x6	Ydata1	Ydata2	Ydata3
n1	83	509300	3740	2552	120287	1	64630.6	64630.1	64630.33
n2	84	529500	3852	2514	121836	2	65777.74000000000	65776.64999999999	65776.23
n42	400	9855900	5692	1405	218061	42	136890.85999999999	136889.85	136891.1
n43	410	10171600	6801	1412	220800	43	136932.07	136933.24	136931.37
n44	417	10500200	8378	1425	223532	44	136483.95000000001	136485.51	136486.79

变量说明见表 5.8 ：

表 5.8: 变量定义及说明

variable	label
Year	年份
Y	就业人数 (打)
X1	消费价格指数
X2	名义 GNP
X3	失业人数
X4	军队人数
X5	14 岁以上的非机构人口数
X6	时间趋势

请考虑如下样本回归模型：

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \hat{\beta}_3 X_{3t} + \hat{\beta}_4 X_{4t} + \hat{\beta}_5 X_{5t} + \hat{\beta}_6 X_{6t} + e_t \tag{5.43}$$

请回答如下问题:

（1）根据回归模型式 5.43 ，写出总体回归模型（PRM），并对参数的理论预期（符号、大小、关系）进行说明。

答：

PRM：

参数预期：

（2）利用 Eviews 对样本回归模型式 5.43 进行回归分析（将报告截图过来，并写出相应的简要报告形式——三行式或四行式）。参数估计结果符合你的理论预期么？

答：

报告截图：

简要报告（三行式或四行式）：

参数估计结果是否符合你前面的理论预期：

（3）数据中存在多重共线性的证据吗?（请以此按照下列方法进行诊断）：

a. 根据上题中主回归报告结果，观察 t 检验，判定系数，F 检验的关系，请你得出关于多重共线性的初步结论。
答：

b. 利用 Eviews 绘制矩阵散点图 (matrix scatter)（截图过来），请你得出关于多重共线性的初步结论。

答：

a. 利用 Eviews 得到简单相关系数矩阵（截图过来），请你得出关于多重共线性的初步结论。

答:

a. 利用 Eviews 进行辅助回归诊断（完成下表，并将 8 个辅助回归结果依次截图过来），请你得出关于多重共线性的初步结论。

答：

表 5.9: 主回归和辅助回归的判定系数计算

辅助模型	模型形式	判定系数
Main	Y ~ X1 + X2 + X3 + X4 + X5 + X6	
A1	X1 ~ X2 + X3 + X4 + X5 + X6	
A2	X2 ~ X1 + X3 + X4 + X5 + X6	
A3	X3 ~ X1 + X2 + X4 + X5 + X6	
A4	X4 ~ X1 + X2 + X3 + X5 + X6	

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 5.9: 主回归和辅助回归的判定系数计算

辅助模型	模型形式	判定系数
A5	X5 ~ X1 + X2 + X3 + X4 + X6	
A6	X6 ~ X1 + X2 + X3 + X4 + X5	

（4）请按下列要求完成下表和提问：

a. 利用上表（见表 5.9）结果，计算方差膨胀因子和容忍度指标；

b. 利用 Eviews 菜单，对主模型式 5.43 操作，得到方差膨胀因子 (截图并列出结果到下表中)

c. 根据上述结果，请你得出关于多重共线性的初步结论。

表 5.10: 方差膨胀因子 VIF 和容忍度 TOL 的计算与比较

辅助模型	模型形式	判定系数	TOL_j	VIF_j	VIF_j
A1	X1 ~ X2 + X3 + X4 + X5 + X6				
A2	X2 ~ X1 + X3 + X4 + X5 + X6				
A3	X3 ~ X1 + X2 + X4 + X5 + X6				
A4	X4 ~ X1 + X2 + X3 + X5 + X6				
A5	X5 ~ X1 + X2 + X3 + X4 + X6				
A6	X6 ~ X1 + X2 + X3 + X4 + X5				

（5）计算病态数（Condition Numbers，K）和方差分解比率（variance-decomposition proportions，VDP。并根据分析结果，得出关于多重共线性的初步结论。

答：

（6）若存在多重共线性问题，那你会采用什么补救措施 (如果有的话)?

a. 简单剔除变量法——经济学和实践观察法。请写明剔除变量的理由，并将纠正后模型的 Eviews 结果截图过来，新模型减弱了多重共线性问题吗？

答：

a. 简单剔除变量法——逐步回归法。将逐步回归后模型的 Eviews 结果截图过来，新模型减弱了多重共线性问题吗？

答：

a. 主成分法 (principal components)。将因子回归模型的 Eviews 结果截图过来，新模型减弱了多重共线性问题吗？（此题选作）

答：

【本次实验题目完毕啦!!】

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

6 异方差问题

仅供学习参考，请勿随意外传
NWSUAF @HHP 2025

6.1 实验目的及要求

目的：掌握异方差问题的检验与处理方法。

要求：在老师指导下完成计量经济模型的异方差检验，并对存在异方差的模型进行修正，最终得到正确的分析结果。

6.2 实验原理

对于不同的样本点，随机误差项的方差不再是常数，则认为模型出现了异方差问题。

异方差的实质表现为随机误差项的方差随着解释变量（引起异方差的解释变量）观测值的变化而变化。

对于出现异方差的原模型主要采用校正其异方差，再对校正后的模型采用普通最小二乘法估计。

6.2.1 什么是模型异方差问题?

对于模型

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad \text{(PRM)} \tag{6.1}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i \quad \text{(SRM)} \tag{6.2}$$

$$\mathbf{y} = \mathbf{X} \ + \mathbf{u} \quad \text{(PRM-matrix)} \tag{6.3}$$

$$\mathbf{y} = \mathbf{X} \ + \mathbf{e} \quad \text{(SRM-matrix)} \tag{6.4}$$

在正态经典线性回归模型假设（N-CLRM）下，随机干扰项 u_i 假设为同方差：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{aligned}
var-cov(\mathbf{u}) &= E(\mathbf{u}\mathbf{u}') \\
&= \begin{bmatrix} E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n) \\ \vdots & \vdots & \vdots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2) \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_n^2 \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \\
&= \sigma^2\mathbf{I}
\end{aligned}
$$

定义 6.1 (异方差)**.** 如果回归模型不再满足 N-CLRM 假设中关于随机干扰项是同方差的条件，那么我们称模型存在异方差（heteroskedasticity）问题。正式地，对于回归模型式 6.1 ，如果随机干扰项 u_i 表现为：

$$
\begin{aligned}
var-cov(\mathbf{u}) &= E(\mathbf{u}\mathbf{u}') \\
&= \begin{bmatrix} E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n) \\ \vdots & \vdots & \vdots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2) \end{bmatrix} && \leftarrow (E(u_i) \equiv 0) \\
&= \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_n^2 \end{bmatrix} && \leftarrow (var(u_i) \equiv \sigma_i^2; cov(u_i, u_j) \equiv \sigma_{ij}^2, i \neq j) \\
&= \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} && \leftarrow (cov(u_i, u_j) \equiv 0, i \neq j)
\end{aligned}
$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

则称回归模型式 6.1 存在异方差问题。

异方差问题的来源很多，包括但不限于：

- 回归模型的设定偏误，比如忽略了重要的解释变量
- 数据采集技术的不同和发展
- 样本数据异常值（outliers）的出现
- 截面数据往往更容易出现异方差问题
- 一个或多个回归元的分布呈偏态 (skewness)

值得注意的是：

- 为了分析的简单和方便，上述定义中随机干扰项 u_i 表现为异方差（$var(u_i) = \sigma_i^2$），但仍旧保证均值为 0（$E(u_i) = 0$），协方差为 0（$cov(u_i, u_j) = 0, i \neq j$）
- 同方差是一种理想化情形（$var(u_i) = \sigma^2$），但异方差（$var(u_i) = \sigma_i^2$）则是更为普遍的存在，而且异方差的具体表现形式有多种多样！
- 只要出现了**任何一种**异方差（$var(u_i) = \sigma_i^2$）形式，都认为模型式 6.1 存在异方差问题！
- 其他假定不变，同方差性假定不成立时，OLS 估计量通常不再是最优线性无偏估计量（BLUE）。此时，OLS 估计量仍然是线性的和无偏的（因这两条性质都与方差无关），但是，不再是“最优的”或“有效的”，也即还是线性无偏估计量（LUE）。

6.2.2 如何诊断模型异方差问题?

异方差诊断是下一步模型矫正的基础，诊断工作的要点有：

（1）根据异方差定义定义 6.1 ，异方差问题刻画的是总体回归模型式 6.1 中随机干扰项 u_i 的方差为异方差。然而总体回归模型式 6.1 中随机干扰项 u_i 的行为不能直接观测得到，我们只好间接观测对应样本回归模型式 6.2 中残差 e_i，通过观察残差 e_i 的表现来间接认识随机干扰项 u_i 的特征。因此，下面的异方差问题诊断方法，始终围绕样本回归模型中残差 e_i 的表现特征来展开。

（2）异方差问题在理论上可以有很多种形式[1]，因此在诊断之前我们只能先猜测模型可能有哪一种（或哪几种）异方差形式。而且不同的异方差诊断工具只能识别特定的异方差形式，所以实践操作中可

[1]记住“同方差只有一种情形，但异方差则可能有千万种”，类似于“喜剧的故事只有一个结局，但悲剧的故事则千差万别”！

能需要尝试多种诊断工具。一旦诊断出有任何一种（或多种）异方差模式，我们都认为模型定义 6.1 存在异方差问题。

（3）诊断工作不仅仅只是诊断得到**是否存在异方差问题**的结论，而且往往还需要诊断出**属于哪一种异方差形式**，因为后面的异方差问题的模型矫正中（如 WLS 矫正法）还需要利用这一有价值的信息！

6.2.3 图形识别检验法

通过观察残差 e_i 的表现来间接认识随机干扰项 u_i 的特征，最直观的是观察残差 e_i 的图形模式：

（1）残差描点图（dot plot）

- 残差 e_i 的描点图（dot plot）
- 标准化残差 $e_i^* = \frac{e_i}{S_{e_i}}$ 的描点图（dot plot）
- e_i^2 的描点图（dot plot）

（2）残差散点图（scatter plot）

- e_i^2 与 Y_i 的散点图（scatter plot）
- e_i^2 与 Y_i^2 的散点图（scatter plot）
- e_i^2 分别与 $(X_{2i}, X_{3i}, \cdots, X_{ki})$ 的散点图（scatter plot）
- e_i^2 分别与 $(X_{2i}^2, X_{3i}^2, \cdots, X_{ki}^2)$ 的散点图（scatter plot）

图形识别检验法的判断准则是，一旦发现存在我们所知的系统模式，则认为模型很可能存在异方差问题。例如：

- 残差 e_i^2 的描点图（dot plot），如果表现为线性增加的图形模式，则表明随机干扰项的方差单调增加。
- e_i^2 与 Y_i 的散点图（scatter plot），如果表现为近似直线模式，则表明随机干扰项的方差单调增加。
- e_i^2 与 X_{3i}^2 的散点图（scatter plot），如果表现为近似下开口抛物线模式，则表明随机干扰项的方差先增加，后减小
- …

6.2.4 正式检验方法：Park 检验法

Park 检验方法针对的异方差情形是：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad \text{(PRM)} \tag{6.5}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i \quad \text{(SRM)} \tag{6.6}$$

$$var(u_i) \equiv \sigma_i^2 = e^{(\mathbf{Z_t'})} \quad \text{(type-Park)} \tag{6.7}$$

其中，e 表示自然底数，也即 e=2.71828182845905， 表示参数，$\mathbf{Z_t}$ 表示由独立变量组成的向量，一般由主回归方程的回归元组成，但又不是必须由它们组成。

Park 检验方法的基本步骤是：

（1）先对样本回归模型式 6.6 进行回归分析，得到残差序列 e_i 和 e_i^2

（2）再对辅助回归模型式 6.8 进行回归分析

$$ln(e_i^2) = \hat{\alpha}_1 + \hat{\alpha}_2 ln(X_{2i}) + \cdots + \hat{\alpha}_k ln(X_{ki}) + v_i \quad \text{(Auxiliary-Park)} \tag{6.8}$$

（3）根据辅助回归方程结果得到异方差诊断结论

- 如果诊断辅助方程式 6.28 的 F 检验**不显著**（对应的概率值 P>0.1），则表明主模型式 6.5 是同方差
- 如果诊断辅助方程式 6.28 的 F 检验**显著**（对应的概率值 P<0.1），则表明主模型式 6.5 是异方差。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

6.2.5 正式检验方法：Glejser 检验法

Glejser 检验方法针对的异方差情形是：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad \text{(PRM)} \tag{6.9}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i \quad \text{(SRM)} \tag{6.10}$$

$$\begin{aligned} var(u_i) &\equiv \sigma_i^2 \\ \sigma_i^2 &= (\sigma^2 + \mathbf{Z}_\mathbf{t}' \,)^\gamma \quad \text{(type-Glejser)} \\ \gamma &= 1, 2 \end{aligned} \tag{6.11}$$

其中， 表示参数，$\mathbf{Z_t}$ 表示由独立变量组成的向量，一般由主回归方程的回归元组成，但又不是必须由它们组成。

Glejser 检验方法的基本步骤是：

（1）先对样本回归模型式 6.10 进行回归分析，得到残差序列 e_i 和 $|e_i|$

（2）再对如下多种形式的辅助回归模型式 6.8 进行回归分析[2]

$$|e_i| = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \cdots + \hat{\alpha}_m X_{mi} + v_i \quad \text{(Glejser1)} \tag{6.12}$$

$$|e_i| = \hat{\alpha}_1 + \hat{\alpha}_2 \sqrt{X_{2i}} + \cdots + \hat{\alpha}_m \sqrt{X_{mi}} + v_i \quad \text{(Glejser2)} \tag{6.13}$$

$$|e_i| = \hat{\alpha}_1 + \hat{\alpha}_2 \frac{1}{X_{2i}} + \cdots + \hat{\alpha}_m \frac{1}{X_{mi}} + v_i \quad \text{(Glejser3)} \tag{6.14}$$

$$|e_i| = \hat{\alpha}_1 + \hat{\alpha}_2 \frac{1}{\sqrt{X_{2i}}} + \cdots + \hat{\alpha}_m \frac{1}{\sqrt{X_{mi}}} + v_i \quad \text{(Glejser4)} \tag{6.15}$$

$$|e_i| = \sqrt{\hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \cdots + \hat{\alpha}_m X_{mi}} + v_i \quad \text{(Glejser5)} \tag{6.16}$$

[2]可参看 pdf，或者参看。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$|e_i| = \sqrt{\hat{\alpha}_1 + \hat{\alpha}_2 X_{2i}^2 + \cdots + \hat{\alpha}_m X_{mi}^2} + v_i \quad \text{(Glejser6)} \tag{6.17}$$

（3）根据辅助回归方程结果得到异方差诊断结论

- 如果诊断辅助方程的 F 检验**不显著**（对应的概率值 P>0.1），则表明主模型式 6.9 是同方差。
- 如果诊断辅助方程的 F 检验**显著**（对应的概率值 P<0.1），则表明主模型式 6.9 是异方差。

6.2.6 正式检验方法：BPG 检验法

BPG 检验方法针对的异方差情形是：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad \text{(PRM)} \tag{6.18}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i \quad \text{(SRM)} \tag{6.19}$$

$$\begin{aligned} var(u_i) &\equiv \sigma_i^2 \\ &= \sigma^2 \cdot f(\mathbf{Z}_\mathbf{t}') \qquad \text{(type-BPG)} \\ &= \sigma^2 \cdot f\left(\alpha_1 + \alpha_2 X_{2i} + \cdots + \alpha_m X_{mi}\right) \end{aligned} \tag{6.20}$$

其中，$f(\cdot)$ 表示函数， 表示参数，$\mathbf{Z}_\mathbf{t}$ 表示由独立变量组成的向量，一般由主回归方程的回归元组成，但又不是必须由它们组成，$m \in (1, k]$。

BPG 检验方法的基本步骤是：

（1）先对样本回归模型式 6.19 进行回归分析，得到残差序列 e_i 和 e_i^2

（2）再对辅助回归模型式 6.32 进行回归分析

$$e_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \cdots + \hat{\alpha}_m X_{mi} + v_i \quad \text{(Auxiliary-BPG)} \tag{6.21}$$

（3）计算并得到 ESS(解释的平方和)，进一步定义有

$$\chi^{2^*} = \frac{ESS^*}{2} \sim \chi^2(m-1)$$

（4）根据辅助回归方程结果得到异方差诊断结论

- 如果诊断辅助方程式 6.32 的 χ^2 检验**不显著**（对应的概率值 P>0.1），则表明主模型式 6.18 是同方差
- 如果诊断辅助方程式 6.32 的 χ^2 检验**显著**（对应的概率值 P<0.1），则表明主模型式 6.18 是异方差。

6.2.7 正式检验方法：White 检验法

White 检验方法针对的异方差情形是：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad \text{(PRM)} \tag{6.22}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i \quad \text{(SRM)} \tag{6.23}$$

$$\begin{aligned} var(u_i) &\equiv \sigma_i^2 \\ &= \sigma^2 \cdot f(\mathbf{Z}_\mathbf{t}') \\ &= \sigma^2 \cdot f\left(X_{pi}, X_{pi}^2, X_{pi}X_{qi}\right) \\ &p, q \in (1, k]; p \neq q \end{aligned} \quad \text{(type-White)} \tag{6.24}$$

其中，$f(\cdot)$ 表示函数， 表示参数，$\mathbf{Z}_\mathbf{t}$ 表示由独立变量组成的向量，一般由主回归方程的回归元、平方项或交叉项组成，但又不是必须由它们组成，$m \in (1, k]$。

White 检验方法的基本步骤是：

（1）先对样本回归模型式 6.23 进行回归分析，得到残差序列 e_i 和 e_i^2

（2）再对如下两种辅助回归模型进行回归分析，其中辅助回归模型式 6.25 只含有回归元及其平方项，辅助回归模型式 6.26 则还包含了回归元的交叉项。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$e_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \cdots + \hat{\alpha}_m X_{mi} + \hat{\lambda}_2 X_{2i}^2 + \cdots + \hat{\lambda}_m X_{mi}^2 + v_i \qquad \text{(square-White)} \tag{6.25}$$

$$\begin{aligned} e_i^2 = {} & \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \cdots + \hat{\alpha}_m X_{mi} + \hat{\lambda}_2 X_{2i}^2 + \cdots + \hat{\lambda}_m X_{mi}^2 \\ & + \hat{\delta}_2 X_{2i} X_{3i} + \cdots + \hat{\delta}_p X_{pi} X_{qi} + v_i \qquad \text{(cross-White)} \\ & p, q \in (1, k]; p \neq q \end{aligned} \tag{6.26}$$

（3）计算并得到辅助回归方程的判定系数 R^2(解释的平方和)，进一步定义有

$$nR^2 \sim \chi_{df}^2$$

（4）对于辅助回归方程式 6.25 ，诊断结论为：

- 辅助回归方程式 6.25 不包含交叉项，则纯粹用于诊断主模型是否含有异方差问题。
- 如果辅助回归方程式 6.25 的 χ^2 检验**不显著**（对应的概率值 P>0.1），则表明主模型式 6.22 是同方差。
- 如果辅助回归方程式 6.25 的 χ^2 检验**显著**（对应的概率值 P<0.1），则表明主模型式 6.22 是异方差。

（5）对于辅助回归方程式 6.26 ，诊断结论为：

- 辅助回归方程式 6.26 包含有交叉项，则用于诊断主模型是否同时含有异方差问题或模型设定偏误
- 如果辅助回归方程式 6.25 的 χ^2 检验**不显著**（对应的概率值 P>0.1），则表明主模型式 6.22 **不存在**异方差问题或模型设定偏误
- 如果辅助回归方程式 6.25 的 χ^2 检验**显著**（对应的概率值 P<0.1），则表明主模型式 6.22 **存在**异方差问题或模型设定偏误。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

6.2.8 如何矫正模型异方差问题?

如前所述，如果诊断出任何一种形式的异方差问题，我们都认为模型存在异方差问题。异方差问题较严重的话，则需要对模型进行校正，以尽量缓解异方差问题的影响。矫正模型异方差问题的思路主要有两条：

（1）使用稳健标准误（robust standard errors）来矫正异方差。这里我们将主要采用 White (1980) 提出的异方差一致性矫正法（heteroskedasticity consistent covariance）进行模型处理[3]。

（2）使用加权最小二乘法（WLS）矫正异方差问题。针对异方差问题的不同形式，可以借助权重处理的办法，尽可能地将异方差问题转换成同方差。加权最小二乘法（WLS）的操作，本质上是两阶段最小二乘方法（two-stage least squares）。先对原模型进行权重处理，时期转换为同方差模型，然后再用最小二乘法进行回归估计。

使用加权最小二乘法（WLS）矫正异方差问题，具体操作选择有：

（1）异方差情形：如果方差 σ_i^2 已知且等于样本方差 S^2

- 广义最小二乘法（WLS）处理：对主模型两边同时除以 S_i

（2）异方差情形：如果方差 σ_i^2 正比于 X_i^2

- 加权最小二乘法（WLS）处理：对主模型两边同时除以 X_i

（3）异方差情形：如果方差 σ_i^2 正比于 X_i

- 加权最小二乘法（WLS）处理：对主模型两边同时除以 $\sqrt{X_i}$

（4）异方差情形：如果方差 σ_i^2 正比于 $\hat{Y}_i^2$

- 加权最小二乘法（WLS）处理：对主模型两边同时除以 $\hat{Y}_i$

（5）异方差情形：如果方差 σ_i^2 未知

- 加权最小二乘法（WLS）处理：对主模型两边同时取对数 $ln()$

[3]可参看网页

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

6.3 实验内容

（1）采用最小二乘法建立主回归模型

（2）侦查模型是否存在异方差问题

- 非正式检验方法（图形检验法）：残差趋势图（dot plot）、残差散点图（scatter plot）
- 正式检验方法：Park 检验法、Glejser 检验法、BPG 检验法、White 检验法

（3）在发现存在异方差的基础上，对原模型进行异方差问题的处理：使用加权最小二乘法校正异方差；使用 White 校正法解决异方差。

6.4 实验准备

6.4.1 实验软件

本次实验需要提前准备好如下软件：

- 统计分析软件 Eviews 9.0 版本及以上
- 公式编辑软件 Mathtype 6.0 版本及以上
- 写作编辑软件 Office Word/Excel 2010 版本及以上
- 浏览器软件 chrome 66.0 版本及以上或 360 极速浏览器 9.5 版本及以上

6.4.2 实验材料

汽车油耗：表 6.1 给出给出了 81 辆汽车在 Y 单位油耗的行驶里程数（英里/加仑），X2 最高时速（英里/小时），X3 发动机马力，X4 汽车空间（立方英尺），X5 车身重量（百磅）等方面的数据。

表 6.1: 汽车单位油耗里程数据 (n=81)

obs	Y	X2	X3	X4	X5
1	65	96	49	89	18
2	56	97	55	92	20

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 6.1: 汽车单位油耗里程数据 (n=81)

obs	Y	X2	X3	X4	X5
3	56	97	55	92	20
4	49	105	70	92	20
5	46	96	53	92	20
77	18	165	322	50	45
78	17	140	238	115	45
79	17	147	263	50	45
80	17	157	295	119	45
81	13	130	236	107	55

变量说明见表 6.2 ：

表 6.2: 变量定义及说明

variable	label
obs	汽车品牌序号
Y	单位油耗的行驶里程数（英里/加仑）
X2	最高时速（英里/小时）
X3	发动机马力
X4	汽车空间（立方英尺）
X5	车身重量（百磅）

请考虑如下样本回归模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + e_i \tag{6.27}$$

6.4.3 实验规则

本实验将要求保留 Eviews 操作过程的相关结果，因此对 Eviews 对象命名规则设计如下：

（1）方程对象（Equation）的命名规则：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 主回归方程对象保存命名为eq_m0
- 异方差问题 WLS 矫正后的回归方程，根据具体矫正类型，酌情保存并命名为：
  - eq_adj_type1
  - eq_adj_type2
  - ...
- 异方差问题 White 一致标准误矫正后的回归方程，保存并命名为：eq_adj_white

（2）序列对象（Series）的命名规则：

- （主回归方程的）残差序列，保存并命名为ei
- （主回归方程的）残差平方序列，保存并命名为ei_sqr
- 各变量的平方序列，分别保存并命名为：
  - Y_sqr
  - X2_sqr
  - X3_sqr
  - X4_sqr
  - X5_sqr

（3）图形对象（Graph）的命名规则：

- 残差及残差平方的描点图（dot plot），绘制成一张图，保存并命名为并命名为：dot_resid
- 全部变量对残差的散点图（scatter），绘制成一张图，保存并命名为并命名为：scatter_ei

（4）表格对象（Table）的命名规则：

- 异方差的 Park 检验结果，另存为并命名为：tab_park
- 异方差的 Glejser 检验结果（前三类情形），分别另存为并命名为：
  - G1
  - G2
  - G3
- 异方差的 BPG 检验结果，另存为并命名为：tab_bpg

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 异方差的 White 检验结果，另存为并命名为：tab_white

6.5 主要实验步骤

6.5.1 导入数据并进行预处理

（1）新建 Eviews 工作文件（见图 6.1 ）。Eviews 菜单操作如下：

- 依次操作：File⇒New⇒Workfile
- 进行 workfile create 引导设置：

> 提示
>
> 提示：Excel 数据，每个同学的 Y 数据都不同，找到自己学号对应下的 Y
> 具体引导设置如下：
>
> - workfile structure type: `unstructured/undatede`
> - data range：81
> - workfile names(optional):
>   - WF: `car`（建议命名）
>   - Page: `mileage`（建议命名）

（2）Eviews 导入数据。菜单操作（Excel 和 Eviews）如下：

- Excel 找到数据。Excel 表格中仅保留自己需要的数据（obsYX2X3X4X5）
- Eviews 导入数据。File⇒Import⇒Import From File：`d:/ econometrics/ data/ Lab6-car.xlsx`

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

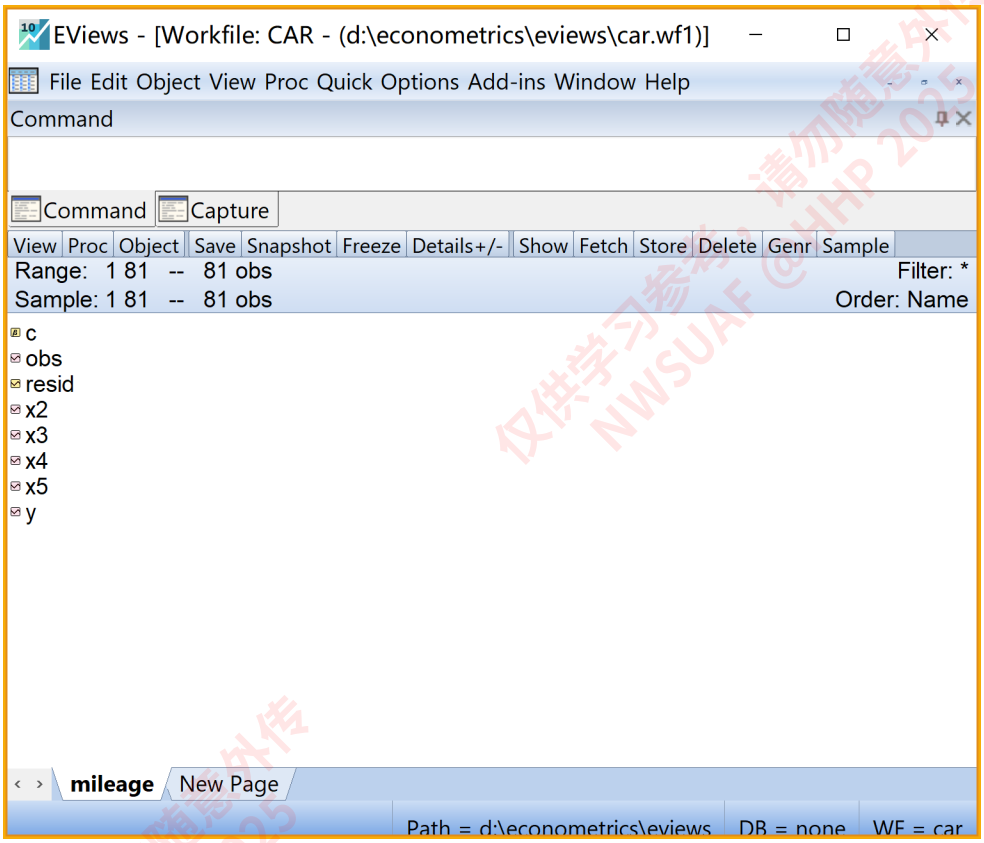


图 6.1: 导入数据的 Eviews 视窗

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

6.5.2 采用最小二乘法建立主回归模型

主回归模型为：

$$Y_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \hat{\beta}_5 X_{5i} + e_i$$

Eviews 菜单操作（见图 6.2 ）：

（1）依次选择 ⇒Quick⇒Estimation Equation

（2）引导设置 Equation Estimation⇒specification

- Equation specification：输入命令 Y c X2 X3 X4 X5
- Estimation settings：
  - Method: 下拉选择 `LS - Least Squares (NLS and ARMA)`
  - Sample: **默认设置**
- 点击 `OK`

（3）模型命名：建议为 `eq_m0`

主回归分析结果见图 6.3 ：

6.5.3 侦查模型是否存在异方差

6.5.3.1 初步观察法（观察主回归方程）

目标：观察主回归方程分析报告，分析回归报告结果，得出初步结论

思路：观察 t^* 检验，判定系数 R^2，F^* 检验的关系

提示：模型使用的数据是否是截面数据；主回归分析报告的 R^2 值；模型整体 F^* 检验结果；斜率系数的 t^* 检验结果。

分析结论：根据主回归报告（见图 6.3 ），表明模型可能存在**严重**的异方差问题。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

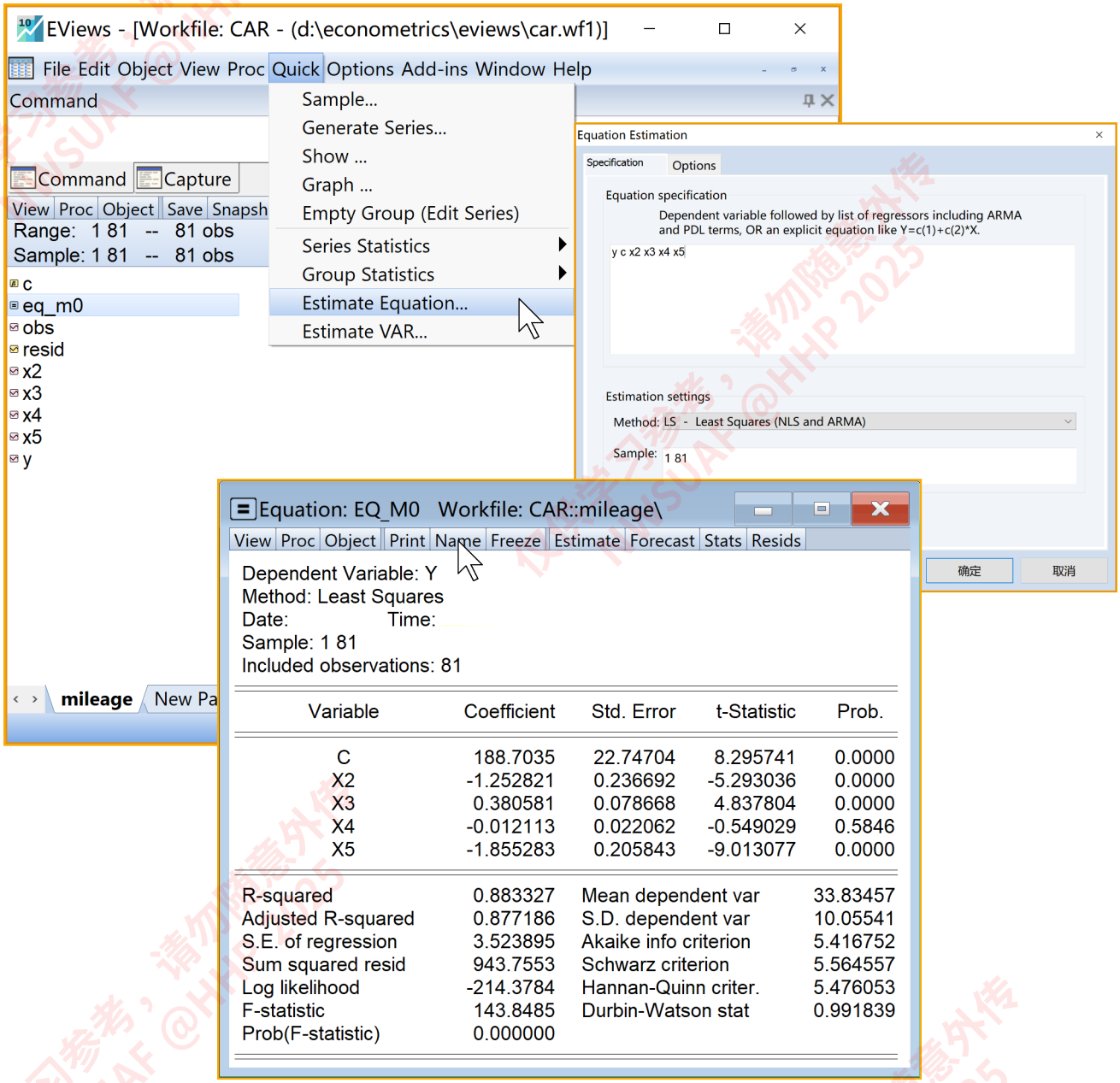


图 6.2: 主回归模型 Eviews 操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

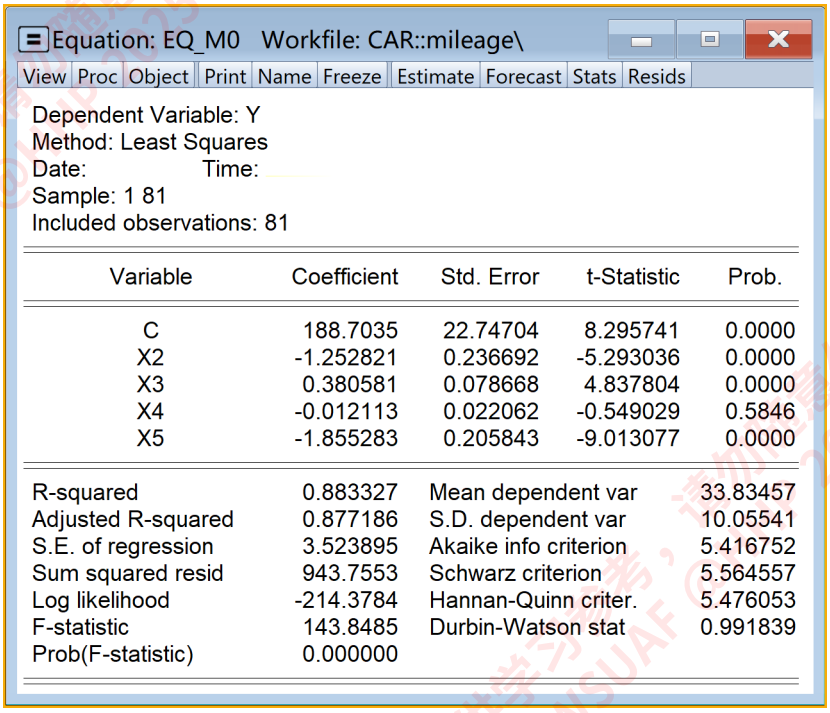


图 6.3: 主回归模型 Eviews 报告

6.5.3.2 非正式检验法（图示法）

目标：观察 e_i 或 e_i^2 的图形模式

思路：判定 e_i 或 e_i^2 与 i、X_i、X_i^2、Y_i、Y_i^2 等的图形关系

提示：

- 描点图（dot plot）是分析一个变量的图形模式。例如 e_i 或 e_i^2（做纵轴）相对于 i（做横轴）的图形关系
- 散点图（scatter plot）是分析两个变量之间的图形模式。例如 e_i 或 e_i^2（做纵轴），相对于 X_i 或 X_i^2 或 Y_i 或 Y_i^2 的图形关系。

Eviews 菜单操作：

（1）分别生成新序列 e_i、e_i^2、Y_i^2 和 X_i^2（见图 6.4 ）

- 生成残差 e_i^2 和 e_i^2 序列（建议分别命名为 ei 和 ei_sqr）

  - 命令视窗（Command）输入命令：`series ei=resid`
  - 命令视窗（Command）输入命令：`series ei_sqr=resid^2`
  - 运行命令：命令行中按 Enter 键

- 生成残差 Y_i^2 序列（建议命名为 Y_sqr）

  - 命令视窗（Command）输入命令：`series Y_sqr=Y^2`
  - 运行命令：命令行中按 Enter 键

a. 生成残差 X_i^2 序列（建议分别命名为 X2_sqr,X3_sqr,X4_sqr,X5_sqr）

- 命令视窗（Command）输入命令：`series X2_sqr=X2^2`
- 命令视窗（Command）输入命令：`series X3_sqr=X3^2`
- 命令视窗（Command）输入命令：`series X4_sqr=X4^2`
- 命令视窗（Command）输入命令：`series X5_sqr=X5^2`
- 运行命令：上述命令行中依次按 Enter 键

- 查看结果：

  - 双击`ei`
  - 双击`ei_sqr`
  - 双击`Y_sqr`
  - 双击`X2_sqr`
  - 双击`X3_sqr`
  - 双击`X4_sqr`
  - 双击`X5_sqr`

（2）绘制 e_i 和 e_i^2 序列的描点图（dot plot）（见图 6.5 ）

- 选择序列对象：键盘 Ctrl 键 + 依次单击选择序列`ei` 和`ei_sqr`
- 进入引导菜单：⇒ Quick ⇒ Graph

  - 选择绘图类型 (Graph type)：Dot plot
  - 选择绘图细节 (Detail)：⇒ Multiple series ⇒ 下拉选择 Multiple graphs

- 点击完成：OK
- 命名并保存绘图（graph）对象：建议命名为 dot_resid

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 查看结果：双击dot_resid（见图 6.6 ）

（3）绘制 e_i 序列对 $Y_i; X_{1i}; X_{2i}; X_{3i}; X_{4i}$ 的散点图（scatter plot）（见图 6.7 ）

- 选择序列对象：键盘 Ctrl 键 + 依次单击选择序列ei；Y；X2；X3；X4；X5
- 进入引导菜单：⇒ Quick ⇒ Graph
  - 选择绘图类型 (Graph type)：Scatter
  - 选择绘图细节 (Detail)：⇒ Multiple series ⇒ 下拉选择 Multiple graphs - First vs. all
- 点击完成：OK
- 命名并保存绘图（graph）对象：建议命名为 scatter_ei
- 查看结果：双击scatter_ei（见图 6.7 ）

（4）绘制 e_i^2 序列对 $Y_i; X_{1i}; X_{2i}; X_{3i}; X_{4i}$ 的散点图（scatter plot）（见图 6.8 ）

- 选择序列对象：键盘 Ctrl 键 + 依次单击选择序列ei_sqr；Y；X2；X3；X4；X5
- 进入引导菜单：⇒ Quick ⇒ Graph
  - 选择绘图类型 (Graph type)：Scatter
  - 选择绘图细节 (Detail)：⇒ Multiple series ⇒ 下拉选择 Multiple graphs - First vs. all
- 点击完成：OK
- 命名并保存绘图（graph）对象：建议命名为 scatter_ei_sqr
- 查看结果：双击scatter_ei_sqr（见图 6.8 ）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

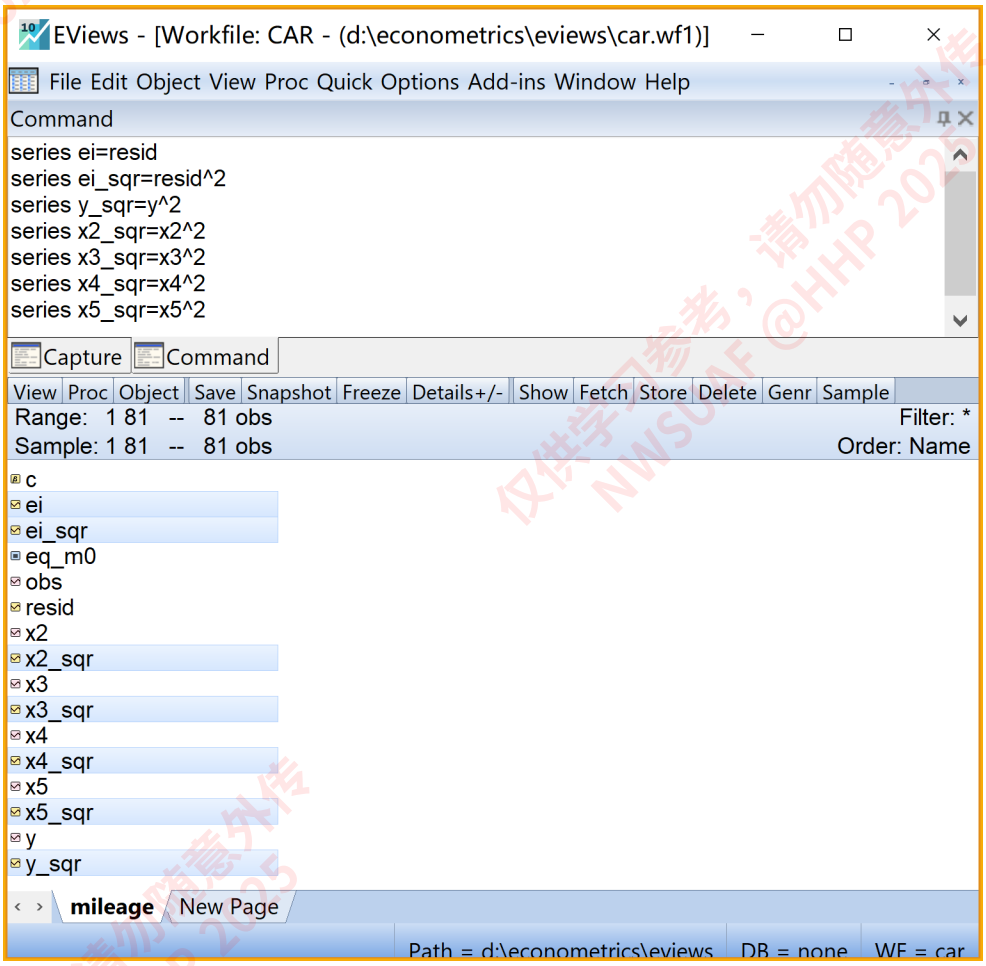


图 6.4: 生成相关变量

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

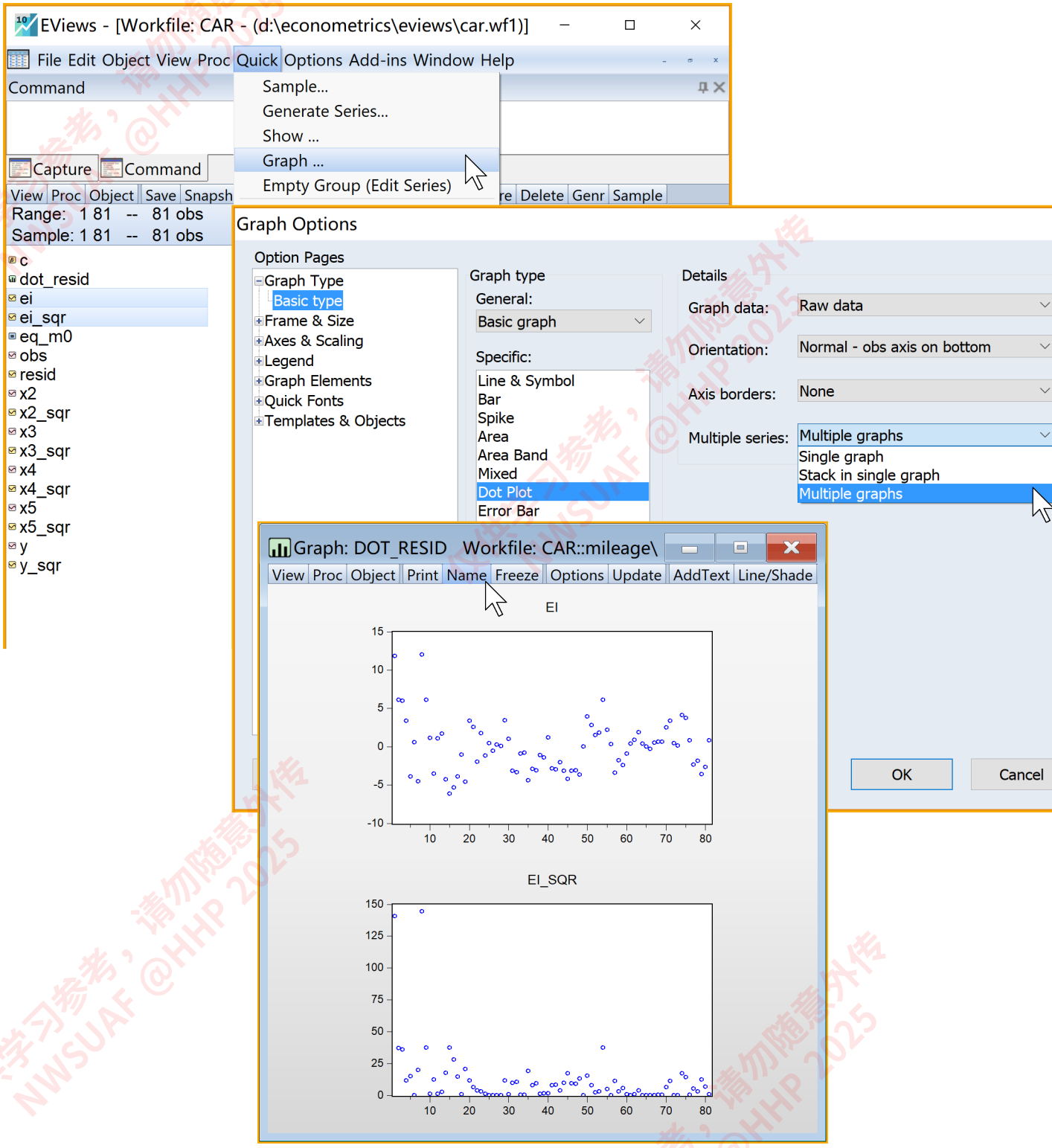


图 6.5: 残差及残差平方的描点图

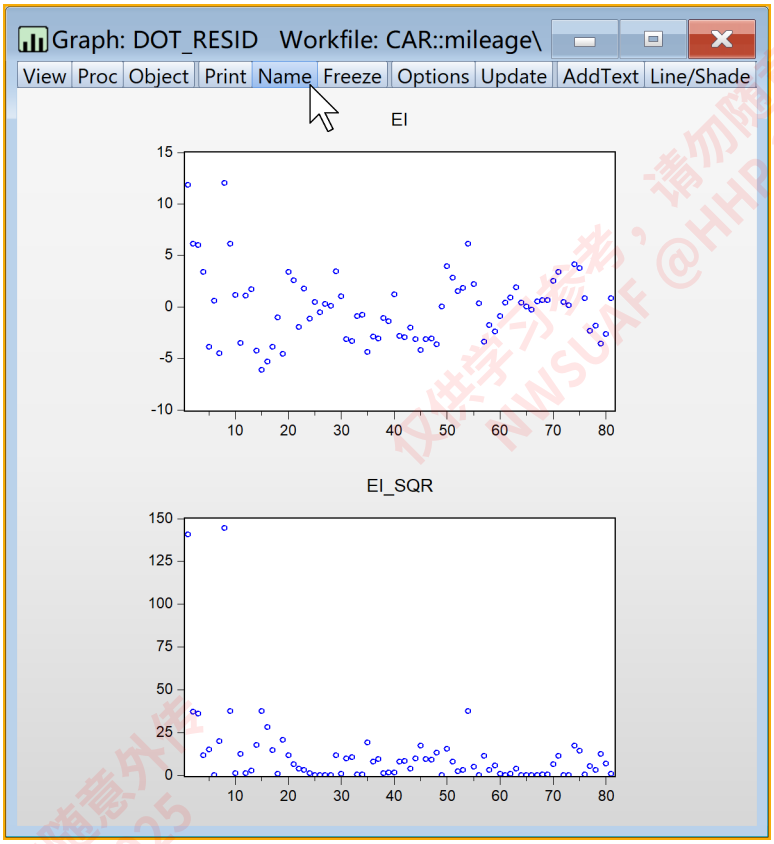


图 6.6: 残差及残差平方的描点图报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

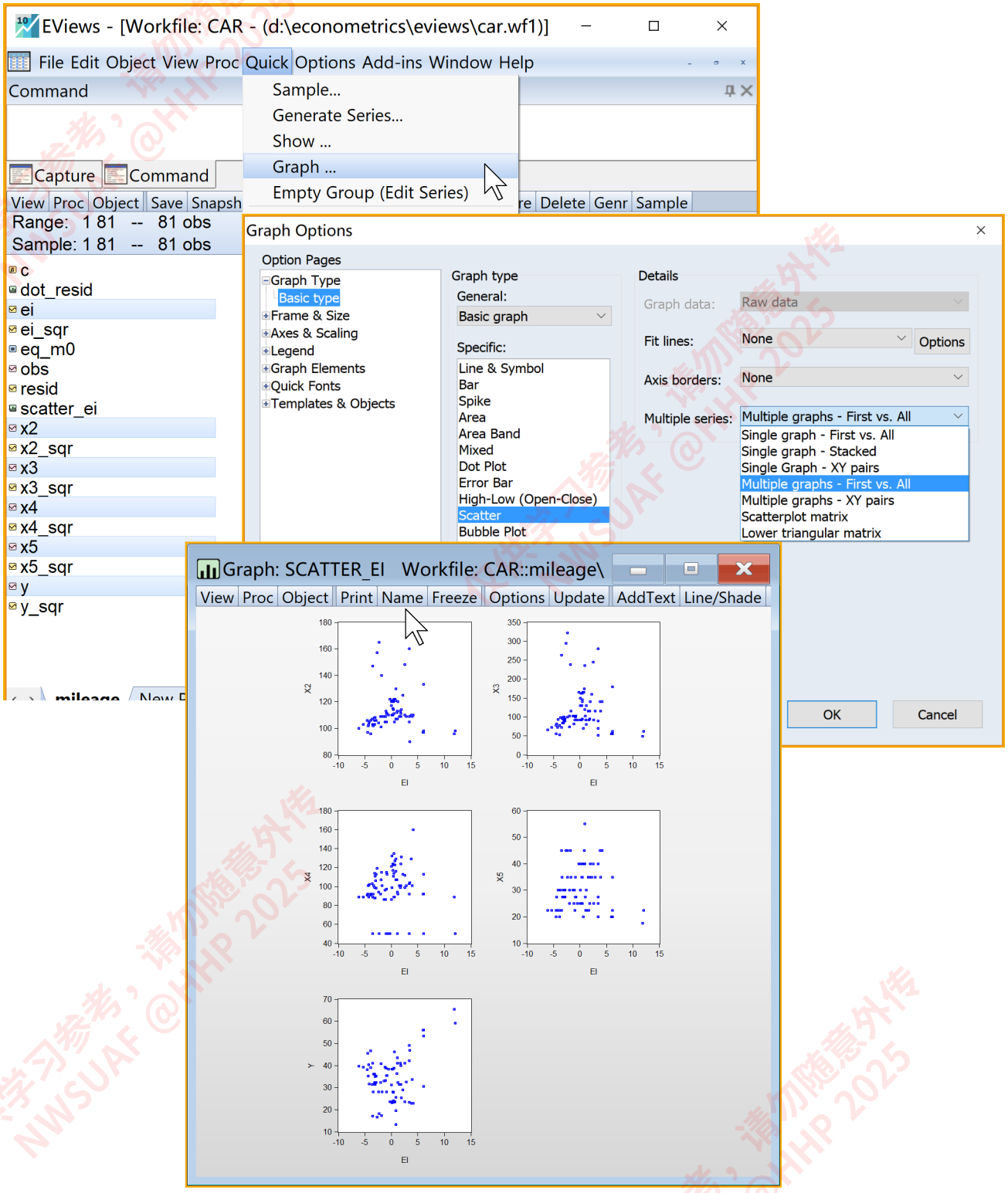


图 6.7: 残差与模型变量的散点图

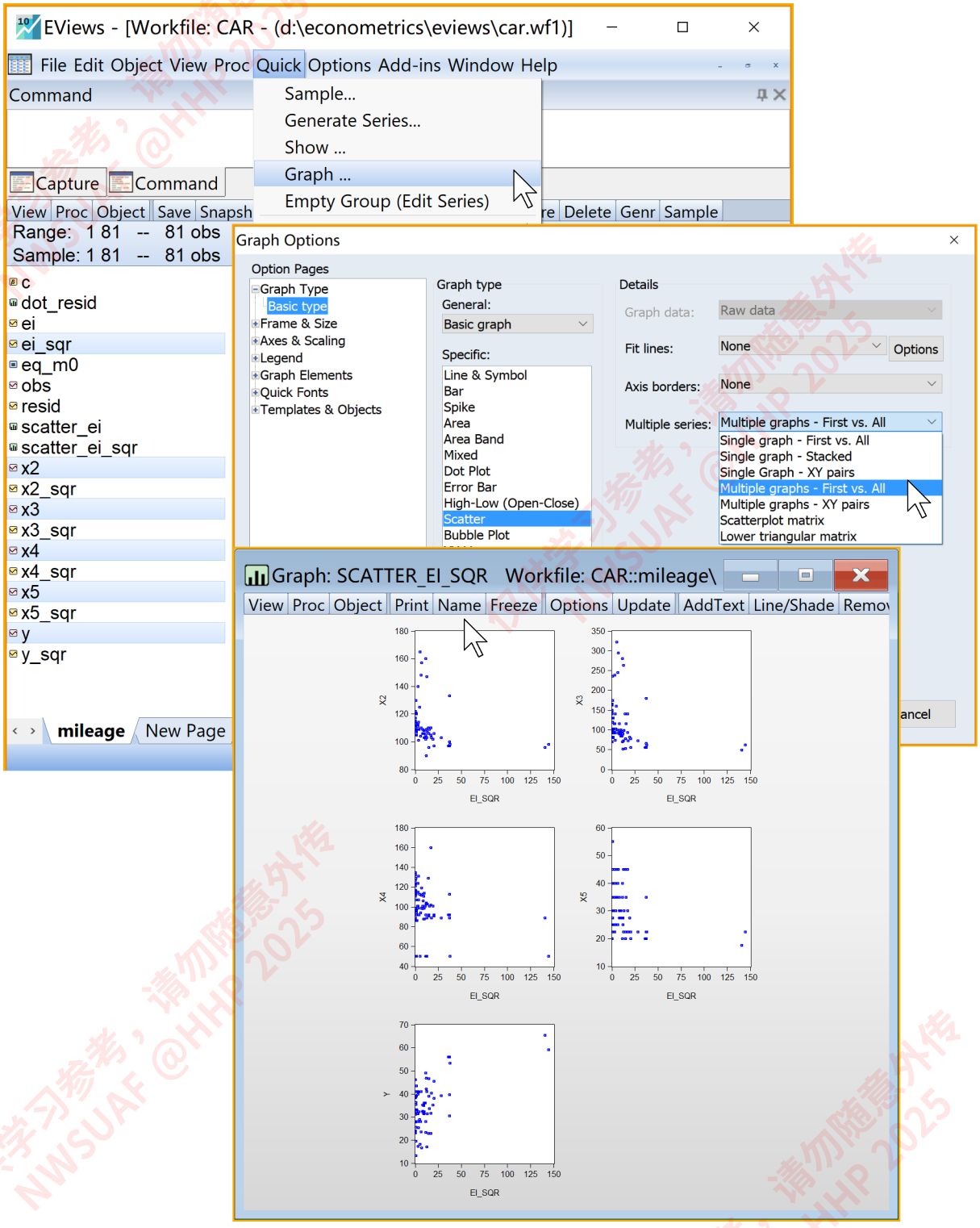


图 6.8: 残差平方与模型变量的散点图

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

6.5.3.3 正式检验法

目标：利用 Eviews 的异方差诊断菜单，分别对主回归模型式 6.27 进行进行异方差诊断。

思路：诊断方法包括 Park 检验、Glejser 检验、BPG 检验和 White 检验等。根据辅助诊断方程的理论假设，分析 Eviews 诊断报告，与相关参考标准进行比较，得到相关结论

定义 6.2 (主回归模型和辅助诊断模型)**.** 主回归模型 (Main Model) 是指 Y 变量对全部 X 变量的线性回归（如主模型式 6.27 ）

辅助诊断模型 (Auxiliary Model) 是指利用主回归模型的残差序列 e_i 对 X 变量进行特定的线性回归（具体的辅助诊断模型有多个）

提示

提示：

- 操作提示：用 Eviews 的异方差诊断菜单 ⇒ View ⇒ Residual Diagnostics ⇒ Heteroskedasticity Test
- 诊断提示：若发现假设检验结果不显著（视具体检验方法而不同）则表明主模型式 6.27 为同方差；否则就表明主模型式 6.27 为异方差
- 理论提示：你也可以自己根据实际情况设定个性化的辅助诊断方程

6.5.3.4 Park 检验法

诊断辅助方程：

$$ln(e_i^2) = \hat{\alpha}_1 + \hat{\alpha}_2 ln(X_{2i}) + \cdots + \hat{\alpha}_k ln(X_{ki}) + v_i \tag{6.28}$$

诊断标准：

- 如果诊断辅助方程式 6.28 的 F 检验**不显著**（对应的概率值 P>0.1），则表明主模型式 6.27 是同方差
- 如果诊断辅助方程式 6.28 的 F 检验**显著**（对应的概率值 P<0.1），则表明主模型式 6.27 是异方差。

提示

Eviews 说明：Eviews 没有 Park 异方差检验的诊断菜单，但可以通过选择 Harvey 异方差菜单，并修改相关变量设置来得到 Park 检验报告。

Eviews 操作（菜单操作实现，具体见图 6.9 ）。具体操作步骤如下：

（1）打开主方程：双击方程（equation）对象eq_m0

（2）进入引导菜单：⇒ View ⇒ Residual Diagnostics ⇒ Heteroskedasticity Test ⇒ Specification

- 设置诊断方法（Test type）：点击选择 Harvey
- 设置诊断方程（Regressors）：输入 c log(X2) log(X3) log(X4) log(X5)

（3）完成设置：点击 Ok

（4）命名并保存表格（table）对象

- 另存为表格（table）对象：点击 Freeze
- 命名并保存表格（talbe）对象：点击 name(建议为 tab_park)
- 查看结果：双击tab_park

具体 Eviews 报告见图 6.10 ：

6.5.3.5 Glejser 检验法

诊断辅助方程：

$$|e_i| = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \cdots + \hat{\alpha}_k X_{ki} + v_i \quad \text{G1} \tag{6.29}$$

$$|e_i| = \hat{\alpha}_1 + \hat{\alpha}_2 \sqrt{X_{2i}} + \cdots + \hat{\alpha}_k \sqrt{X_{ki}} + v_i \quad \text{G2} \tag{6.30}$$

$$|e_i| = \hat{\alpha}_1 + \hat{\alpha}_2 \frac{1}{X_{2i}} + \cdots + \hat{\alpha}_k \frac{1}{X_{ki}} + v_i \quad \text{G3} \tag{6.31}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

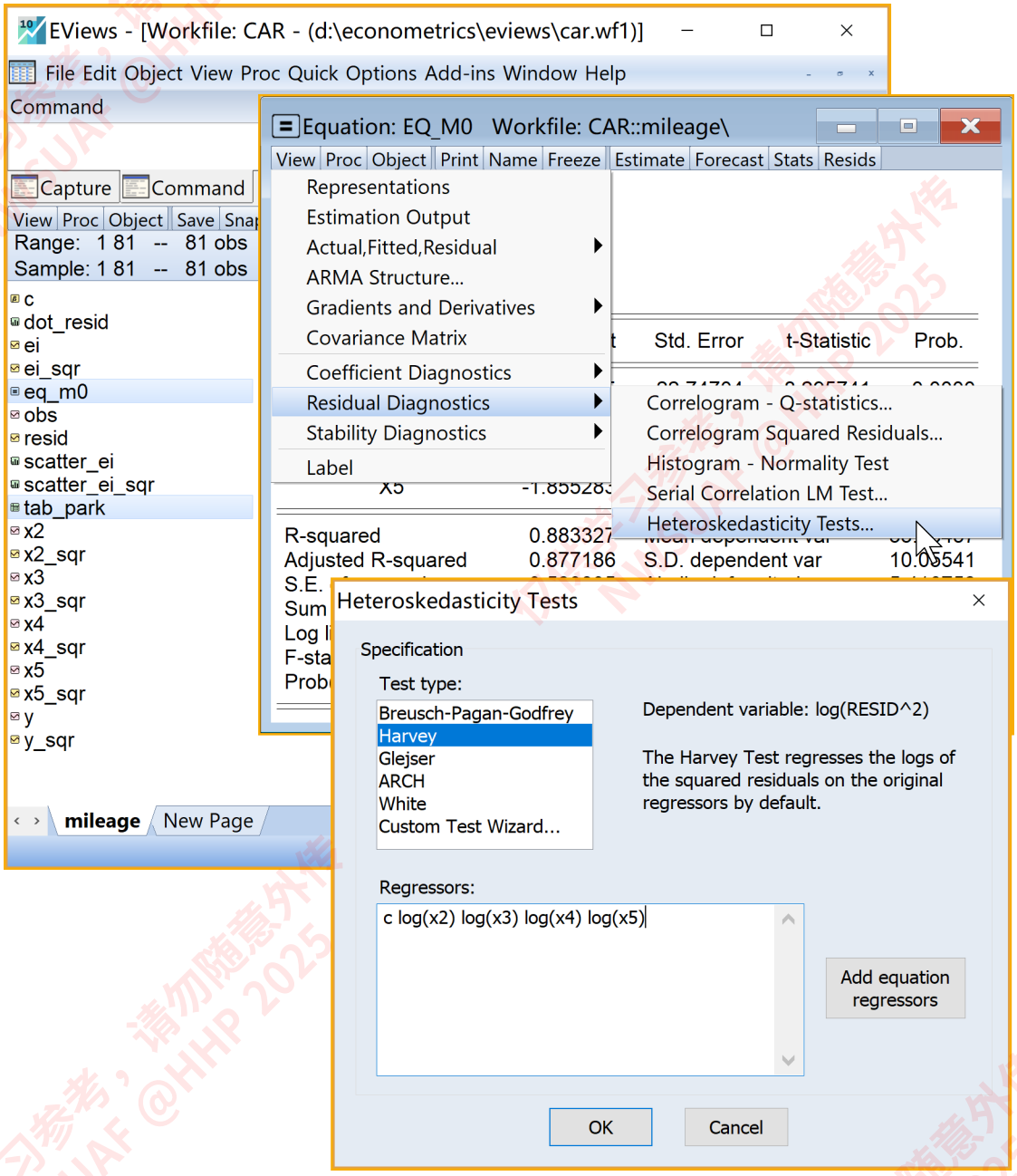


图 6.9: Park 异方差检验操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

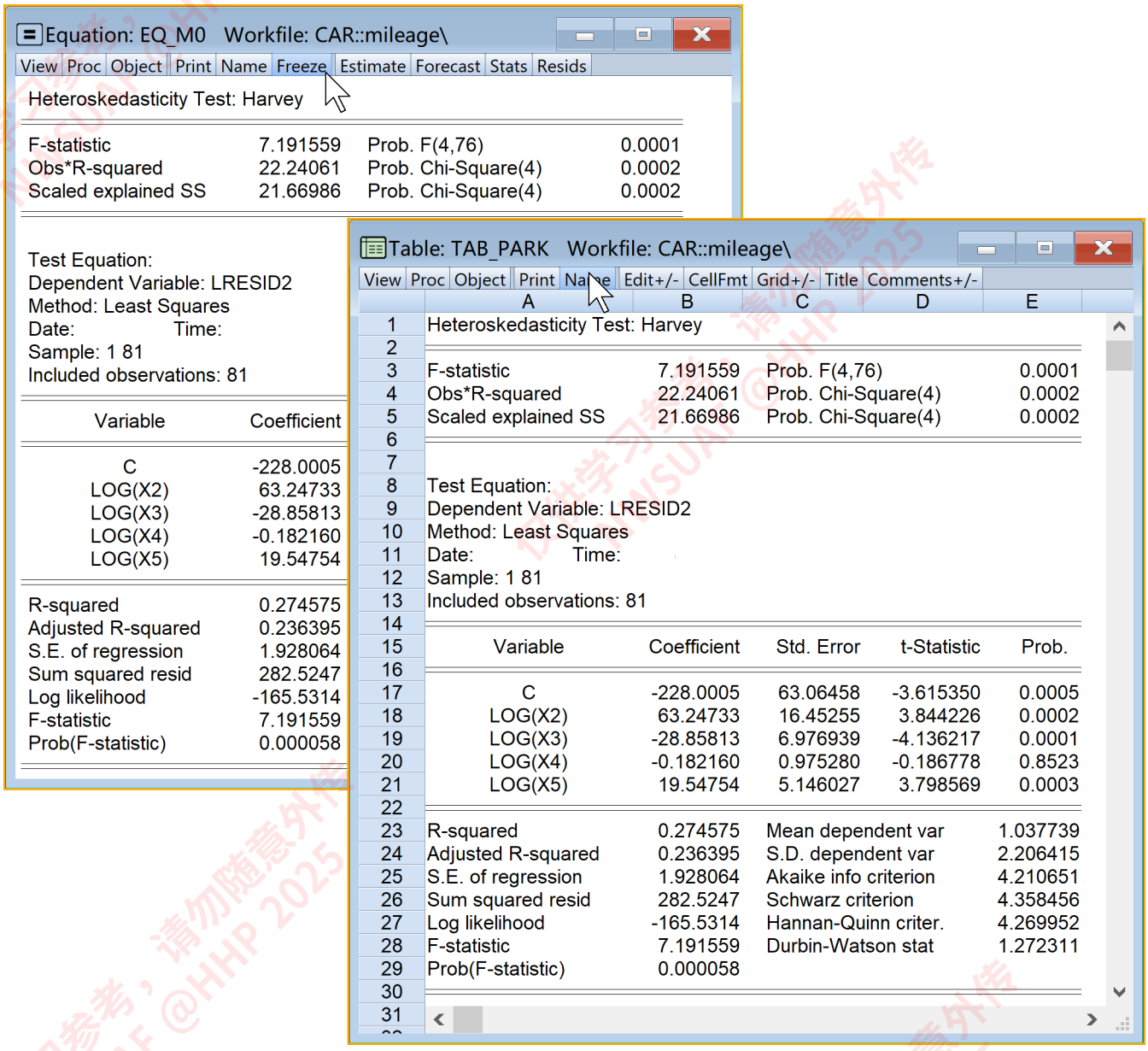


图 6.10: Park 异方差检验报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

诊断标准：

- 如果诊断辅助方程式 6.29 的 F 检验**不显著**（对应的概率值 P>0.1），则表明主模型式 6.27 是同方差
- 如果诊断辅助方程式 6.30 的 F 检验**显著**（对应的概率值 P<0.1），则表明主模型式 6.27 是异方差。

提示

注意事项：

- Glejser 辅助诊断方程式 6.30 如果 X_i 的取值有 0 值，则不能做 $\frac{1}{X_i}$ 的操作，否则计算机会报错（如 `Overflow`，见图 6.11 ）。
- Glejser 辅助诊断方程式 6.31 如果 X_i 的取值有负数值或 0 值，则不能做 $log(X_i)$ 的操作，否则计算机会报错（如 `Log of non positive number`，见图 6.12 ）。



图 6.11: Glejser 异方差检验操作报错（倒数情形）



图 6.12: Glejser 异方差检验操作报错（对数情形）

Eviews 操作 1（Glejser 辅助方程式 6.29 ，具体见图 6.13 ）：

（1）打开主方程：双击方程（equation）对象eq_m0

（2）进入引导菜单：⇒ View ⇒ Residual Diagnostics ⇒ Heteroskedasticity Test ⇒ Specification

- 设置诊断方法（Test type）：点击选择 Glejser
- 设置诊断方程（Regressors）：输入 c X2 c X3 c X4 c X5

（3）完成设置：点击 Ok

（4）命名并保存表格（table）对象

- 另存为表格（table）对象：点击 Freeze

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 命名并保存表格（table）对象：点击 name(建议为 tab_G1)
- 查看结果：双击tab_G1

具体 Eviews 报告见图 6.14 ：

Eviews 操作 2（Glejser 辅助方程式 6.30 ，具体见图 6.15 ）：

（1）打开主方程：双击方程（equation）对象eq_m0

（2）进入引导菜单：⇒ View ⇒ Residual Diagnostics ⇒ Heteroskedasticity Test ⇒ Specification

- 设置诊断方法（Test type）：点击选择 Glejser
- 设置诊断方程（Regressors）：输入 c X2^0.5 X3^0.5 X4^0.5 X5^0.5

（3）完成设置：点击 Ok

（4）命名并保存表格（table）对象

- 另存为表格（table）对象：点击 Freeze
- 命名并保存表格（table）对象：点击 name(建议为 tab_G2)
- 查看结果：双击tab_G2

具体 Eviews 报告见图 6.16 ：

Eviews 操作 3（Glejser 辅助方程式 6.31 ，具体见图 6.17 ）：

（1）打开主方程：双击方程（equation）对象eq_m0

（2）进入引导菜单：⇒ View ⇒ Residual Diagnostics ⇒ Heteroskedasticity Test ⇒ Specification

- 设置诊断方法（Test type）：点击选择 Glejser
- 设置诊断方程（Regressors）：输入 c 1/ X2 1/ X3 1/ X4 1/ X5

（3）完成设置：点击 Ok

（4）命名并保存表格（table）对象

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

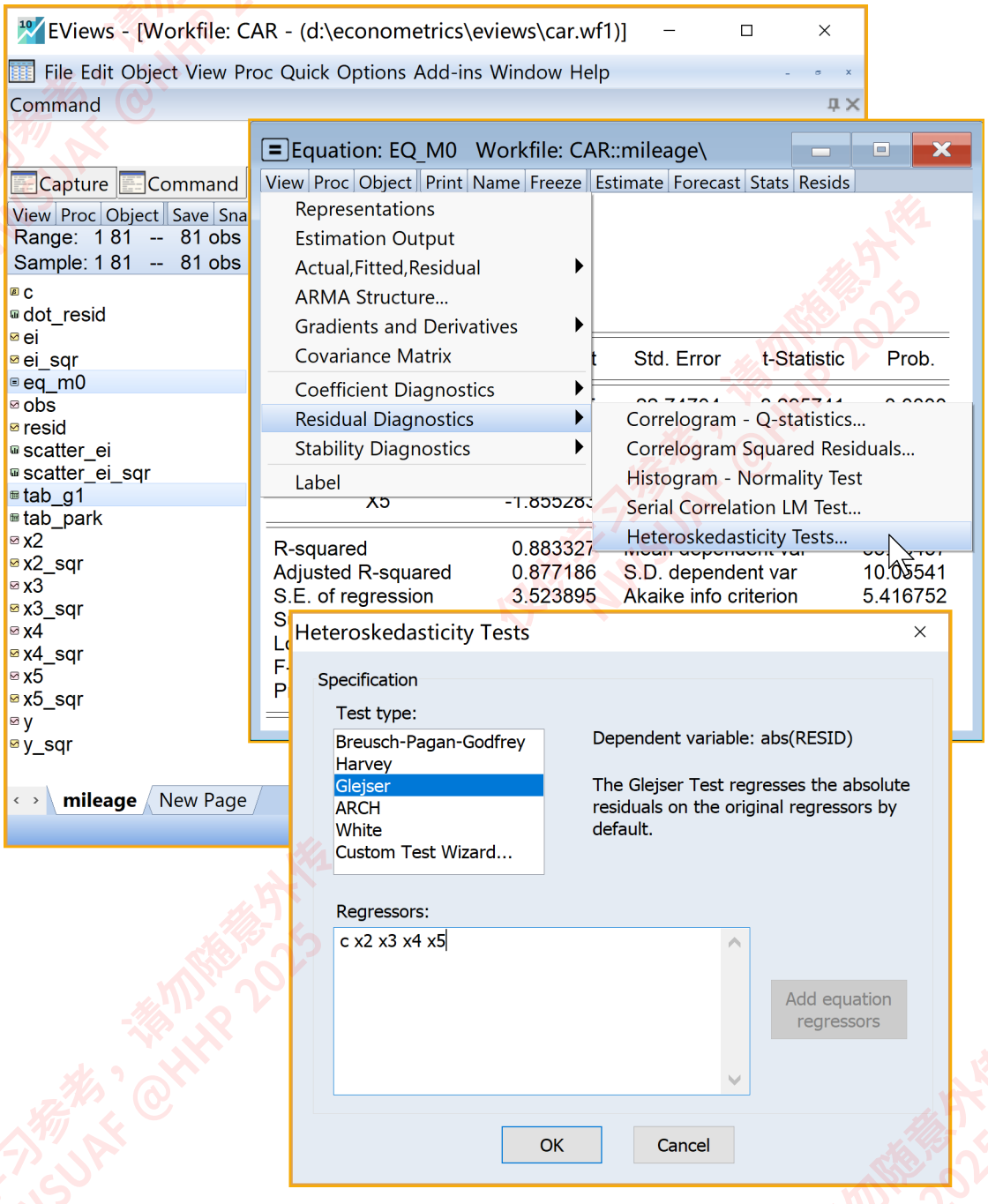


图 6.13: Glejser 异方差检验操作 1

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

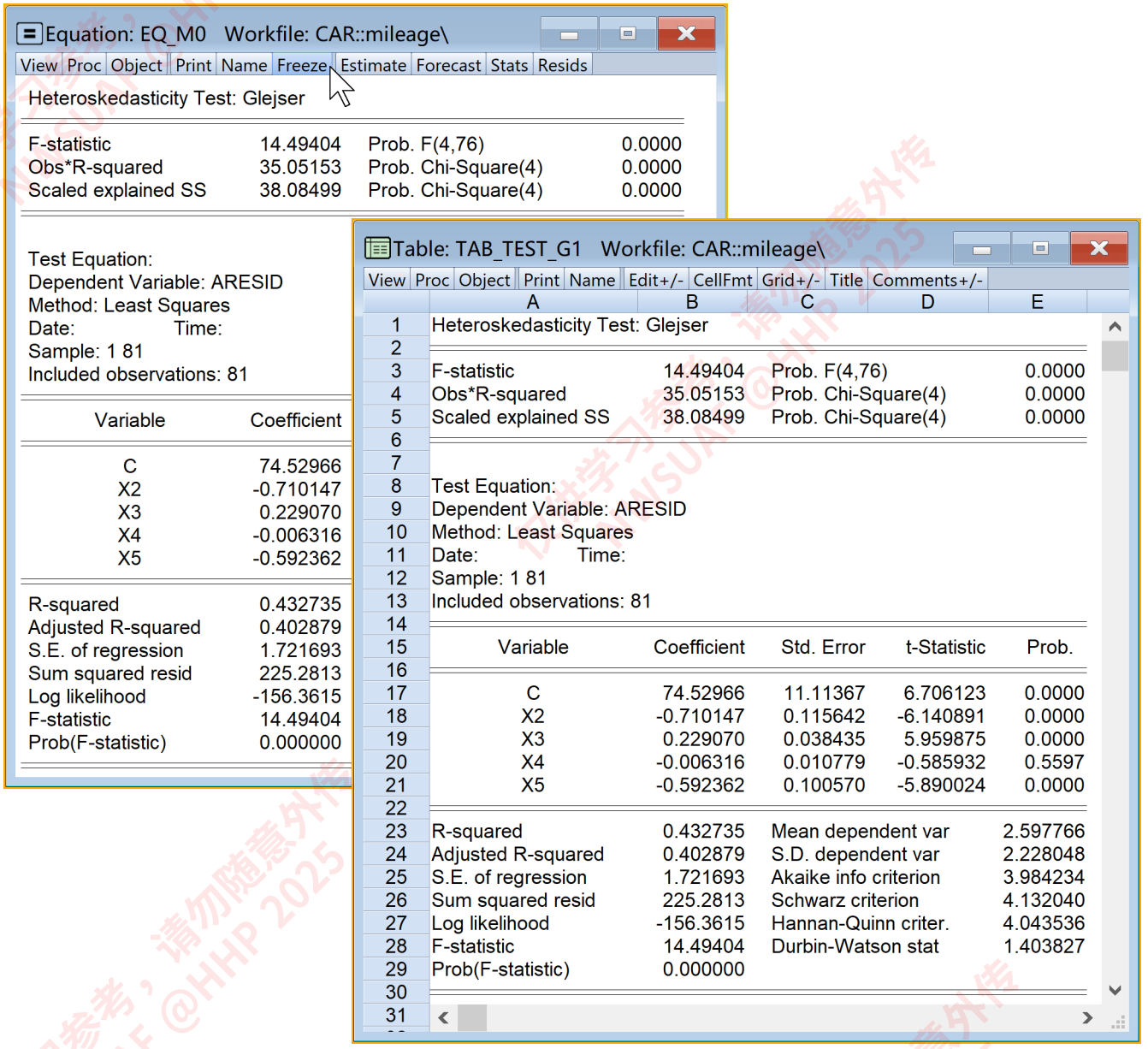


图 6.14: Glejser 异方差检验报告 1

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

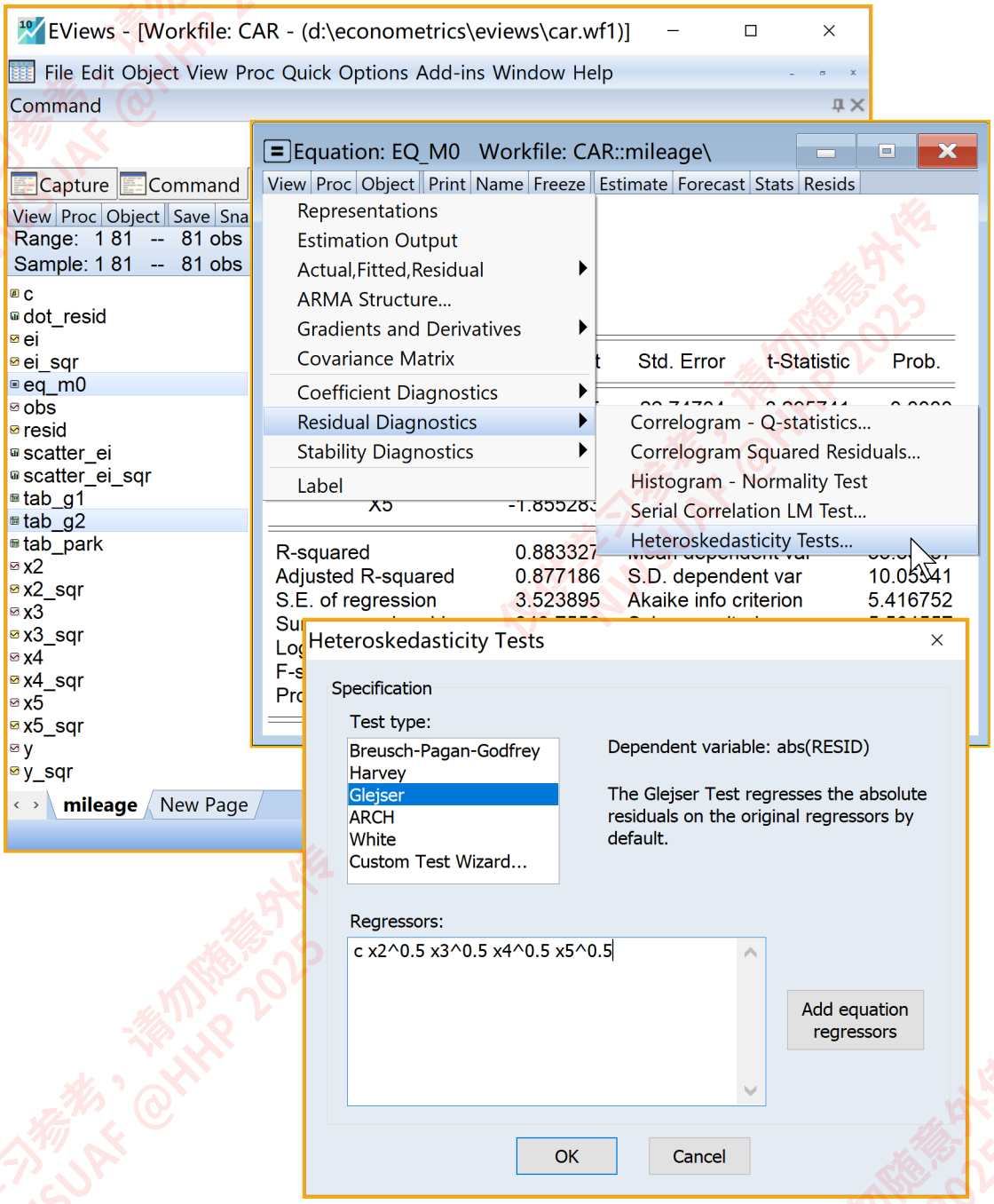


图 6.15: Glejser 异方差检验操作 2

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

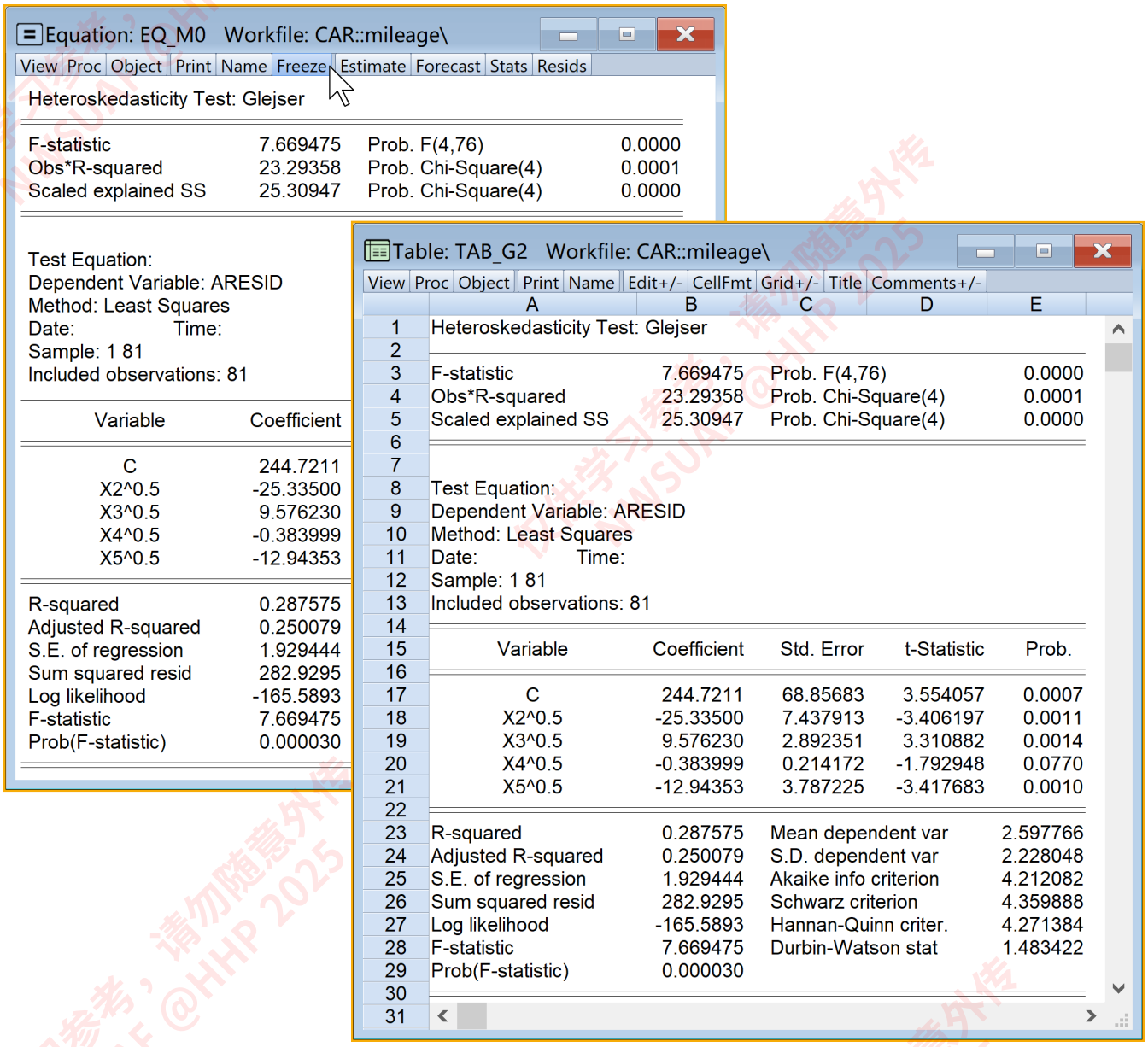


图 6.16: Glejser 异方差检验报告 2

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 另存为表格（table）对象：点击 Freeze
- 命名并保存表格（table）对象：点击 name(建议为 tab_G3)
- 查看结果：双击tab_G3

具体 Eviews 报告见图 6.18 ：

6.5.3.6 BPG 检验法

诊断辅助方程：

$$e_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \cdots + \hat{\alpha}_k X_{ki} + v_i \tag{6.32}$$

诊断标准：

- 如果诊断辅助方程式 6.32 的 χ^2 检验（`scaled explained SS`）**不显著**（对应的概率值 P>0.1），则表明主模型式 6.27 是同方差
- 如果诊断辅助方程式 6.32 的 χ^2 检验（`scaled explained SS`）**显著**（对应的概率值 P<0.1），则表明主模型式 6.27 是异方差。

Eviews 操作（菜单操作实现，具体见图 6.19 ）：

（1）打开主方程：双击方程（equation）对象eq_m0

（2）进入引导菜单：⇒ View ⇒ Residual Diagnostics ⇒ Heteroskedasticity Test ⇒ Specification

- 设置诊断方法（Test type）：点击选择 Breusch-Pagan-Godfrey
- 设置诊断方程（Regressors）：输入 c X2 X3 X4 X5

（3）完成设置：点击 Ok

（4）命名并保存表格（table）对象

- 另存为表格（table）对象：点击 Freeze
- 命名并保存表格（table）对象：点击 name(建议为 tab_bpg)

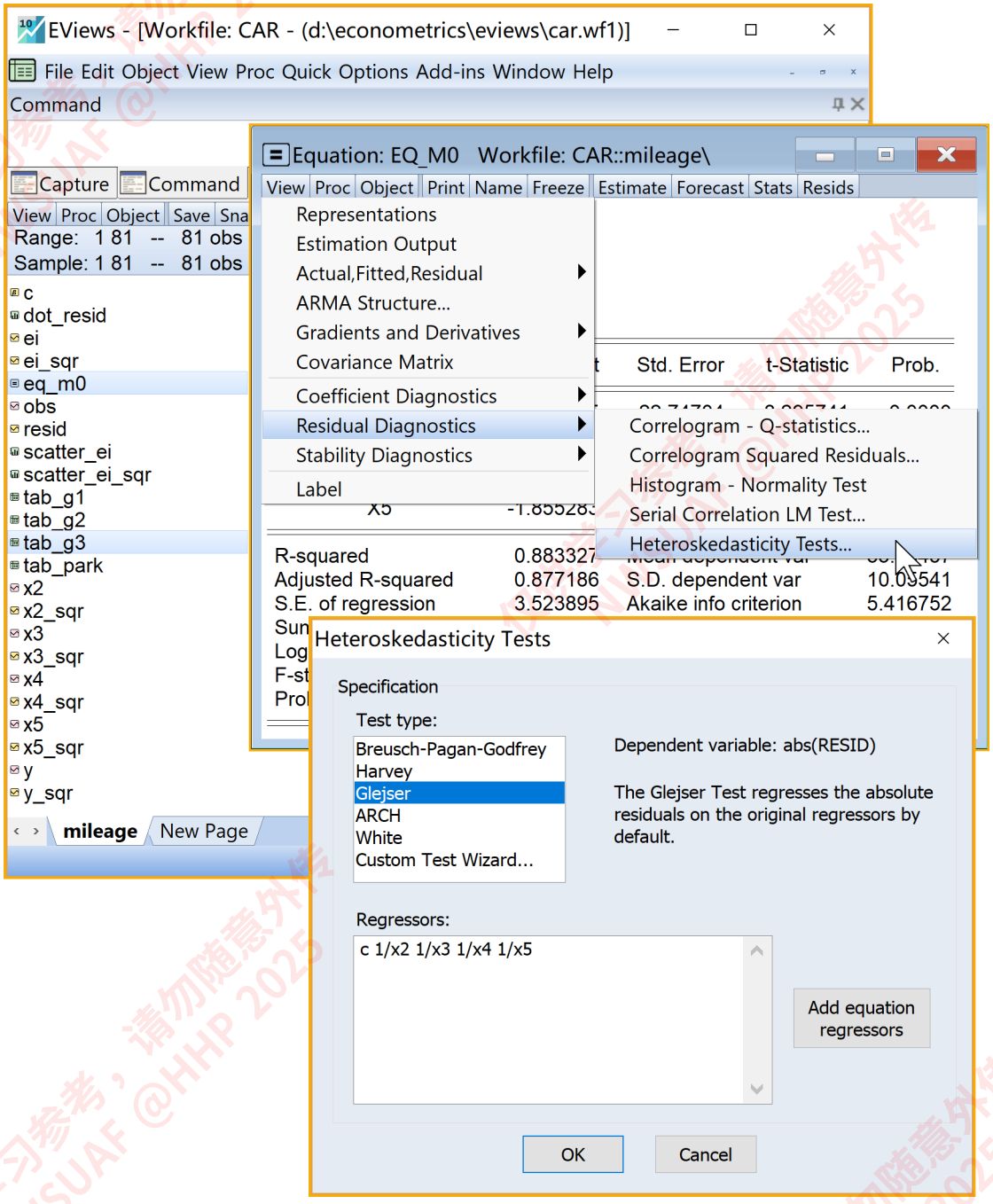


图 6.17: Glejser 异方差检验操作 3

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

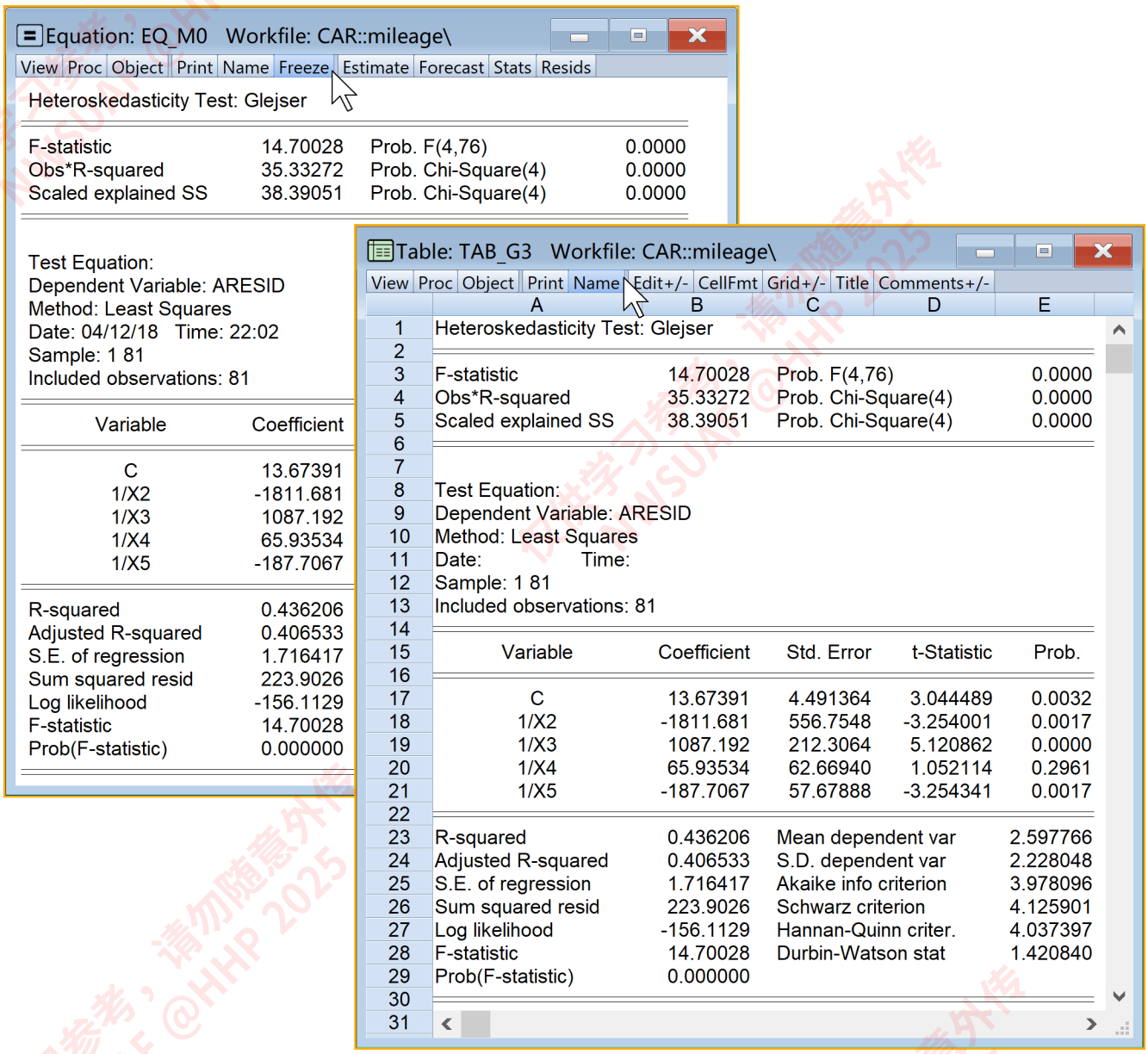


图 6.18: Glejser 异方差检验报告 3

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 查看结果：双击tab_bpg

具体 Eviews 报告见图 6.20 ：

6.5.3.7 White 检验法

诊断辅助方程：

$$Y_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + e_i \tag{6.33}$$

$$e_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \hat{\alpha}_3 X_{3i} + \hat{\alpha}_4 X_{2i}^2 + \hat{\alpha}_5 X_{3i}^2 + \hat{\alpha}_6 X_{2i} X_{3i} + v_i \tag{6.34}$$

诊断标准：

- 如果诊断辅助方程式 6.34 的 χ^2 检验（`scaled explained SS`）**不显著**（对应的概率值 P>0.1），则表明主模型式 6.33 是同方差
- 如果诊断辅助方程式 6.34 的 χ^2 检验（`scaled explained SS`）**显著**（对应的概率值 P<0.1），则表明主模型式 6.33 是异方差。

Eviews 操作（菜单操作实现，具体见图 6.21 ）：

（1）打开主方程：双击方程（equation）对象eq_m0

（2）进入引导菜单：⇒ View ⇒ Residual Diagnostics ⇒ Heteroskedasticity Test ⇒ Specification

- 设置诊断方法（Test type）：点击选择 White
- 设置诊断方程（下面两类方程自行选择其一）：
  - 交叉项方程：**勾选** Include White cross term
  - 非交叉项方程：**不勾选** Include White cross term

（3）完成设置：点击 Ok

（4）命名并保存表格（table）对象

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

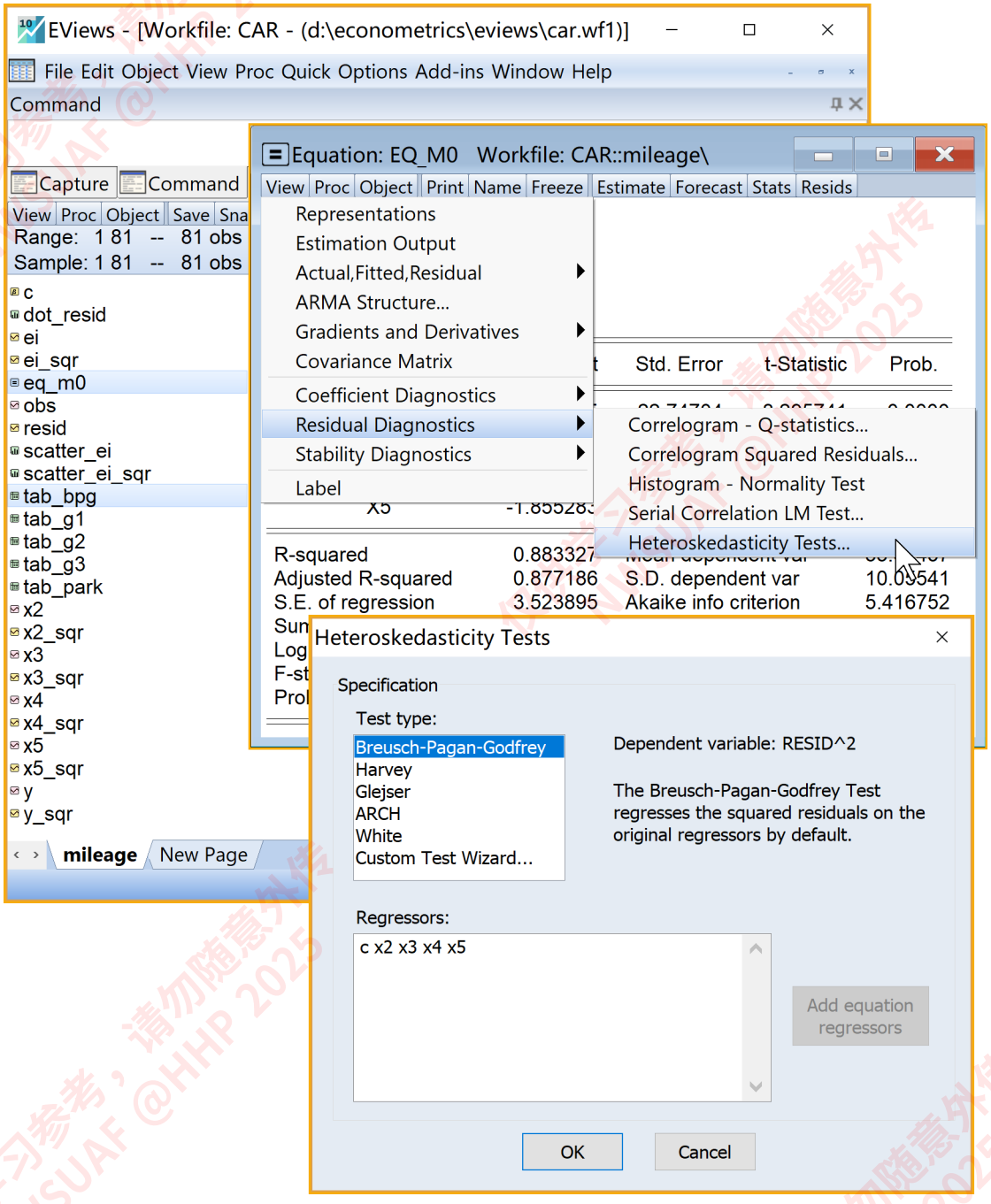


图 6.19: BPG 异方差检验操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

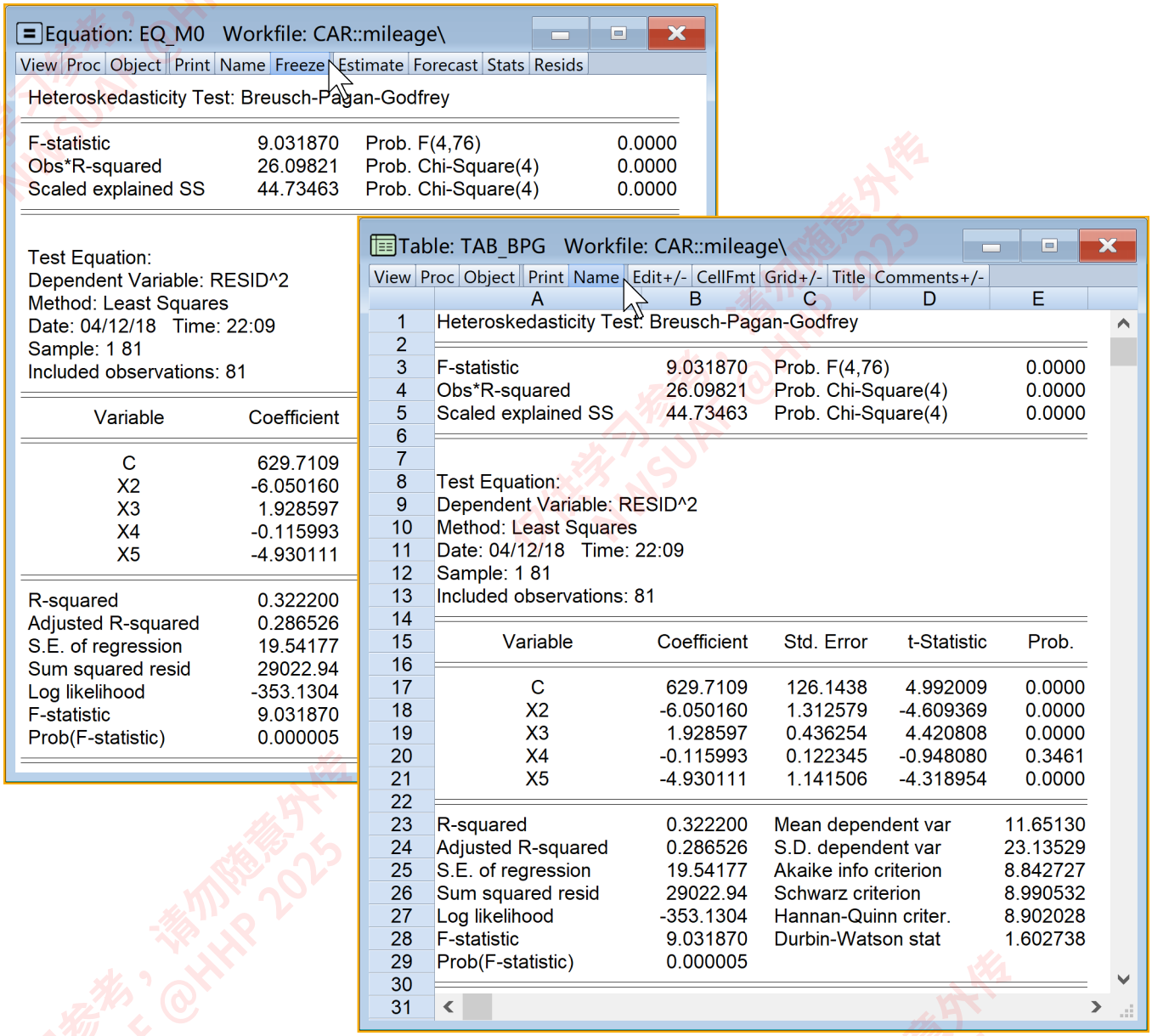


图 6.20: BPG 异方差检验报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 另存为表格（table）对象：点击 Freeze
- 命名并保存表格（table）对象：点击 name(建议为 tab_white)
- 查看结果：双击tab_white

具体 Eviews 报告见图 6.22 ：

6.5.4 使用加权最小二乘法（WLS）矫正异方差问题

6.5.4.1 方差 σ_i^2 已知且等于样本方差 S^2

如果主模型式 6.27 存在异方差问题，且假设方差正比于样本方差 S_i^2，则有：

$$var(u_i) = E(u_i^2) = \sigma_i^2 = \sigma^2 S_i^2 \tag{6.35}$$

对主模型式 6.27 两边同时除以 S_i，得到：

$$\frac{Y_i}{S_i} = \frac{\hat{\beta}_1}{S_i} + \hat{\beta}_2\frac{X_{2i}}{S_i} + \hat{\beta}_3\frac{X_{3i}}{S_i} + \hat{\beta}_4\frac{X_{4i}}{S_i} + \hat{\beta}_5\frac{X_{5i}}{S_i} + \frac{e_i}{S_i} \tag{6.36}$$

$$Y_i^* = \beta_1^* + \beta_2^* X_{2i}^* + \beta_3^* X_{3i}^* + \beta_4^* X_{4i}^* + \beta_5^* X_{5i}^* + v_i \tag{6.37}$$

6.5.4.2 方差 σ_i^2 正比于 X_i^2

如果主模型式 6.27 存在异方差问题，且假设方差正比于 X_i^2，则有：

$$var(u_i) = E(u_i^2) = \sigma_i^2 = \sigma^2 X_i^2 \tag{6.38}$$

对主模型式 6.27 两边同时除以 X_i，得到：

$$\frac{Y_i}{X_{2i}} = \frac{\hat{\beta}_1}{X_{2i}} + \hat{\beta}_2 + \hat{\beta}_3\frac{X_{3i}}{X_{2i}} + \hat{\beta}_4\frac{X_{4i}}{X_{2i}} + \hat{\beta}_5\frac{X_{5i}}{X_{2i}} + \frac{e_i}{X_{2i}} \tag{6.39}$$

$$Y_i^* = \beta_1^* + \beta_2^* X_{2i}^* + \beta_3^* X_{3i}^* + \beta_4^* X_{4i}^* + \beta_5^* X_{5i}^* + v_i \tag{6.40}$$

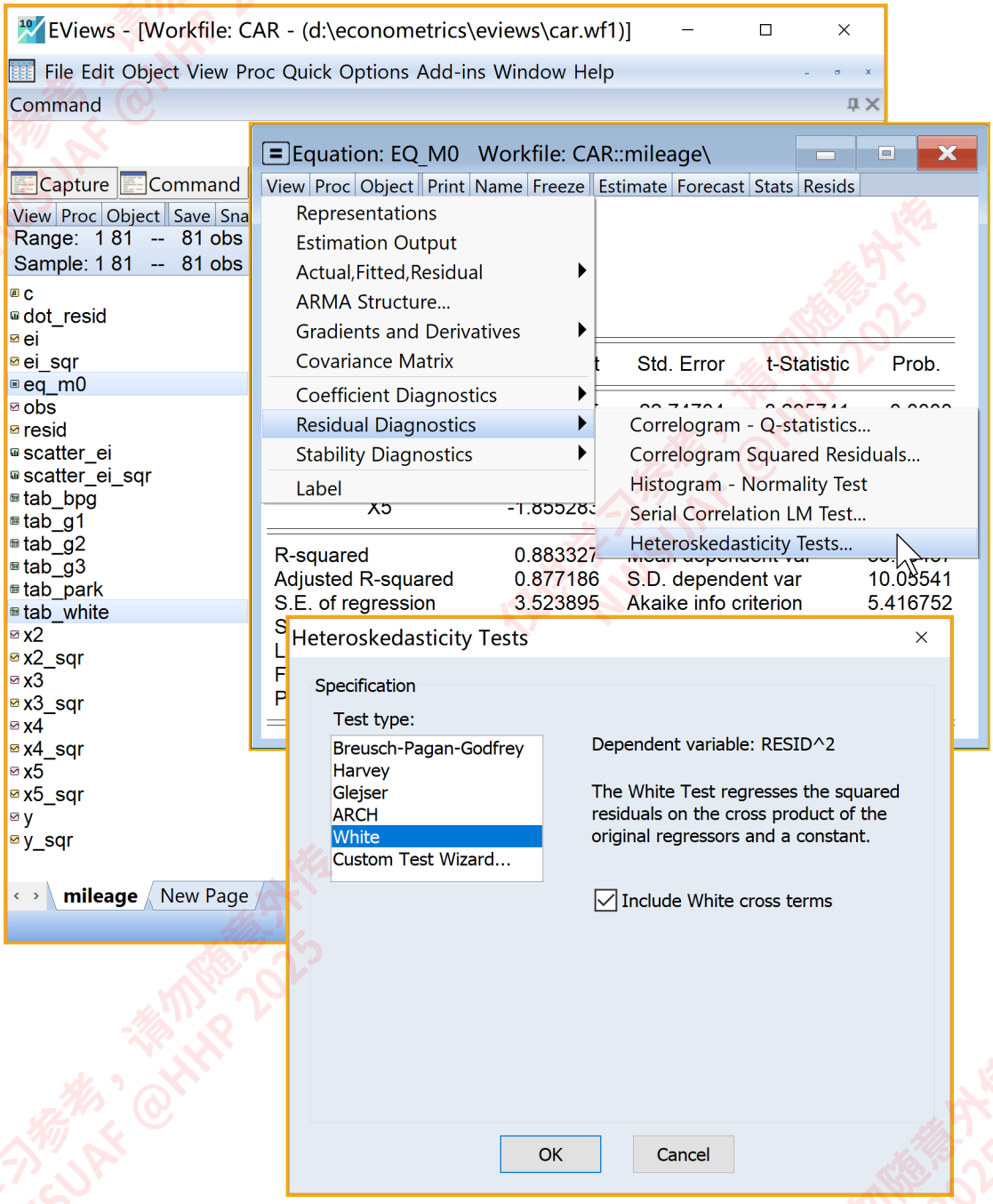


图 6.21: White 异方差检验报告操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

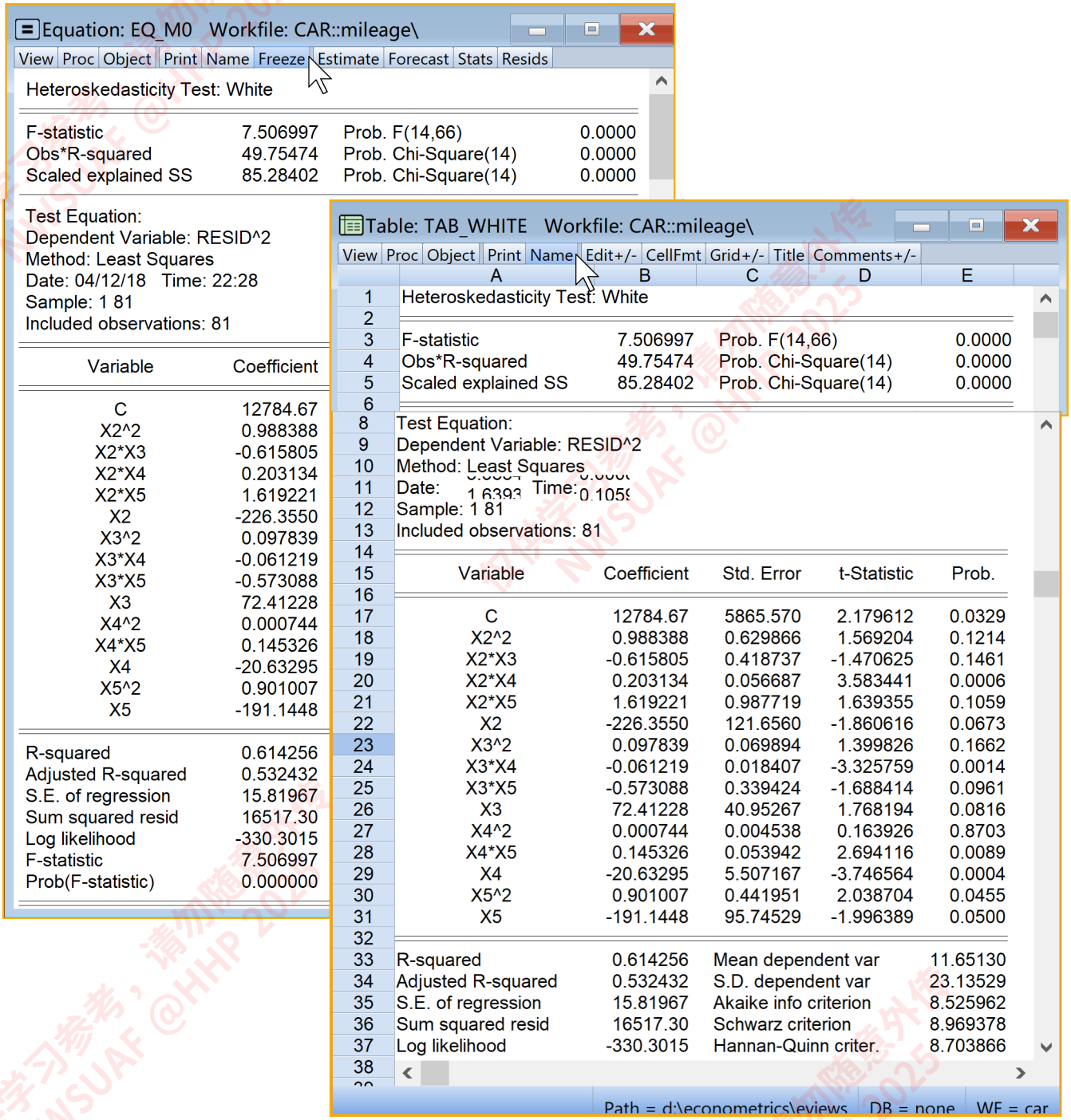


图 6.22: White 异方差检验报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

6.5.4.3 方差 σ_i^2 正比于 X_i

如果主模型式 6.27 存在异方差问题，且假设方差正比于 X_i，则有：

$$var(u_i) = E(u_i^2) = \sigma_i^2 = \sigma^2 X_i \tag{6.41}$$

对主模型式 6.27 两边同时除以 $\sqrt{X_i}$，得到：

$$\frac{Y_i}{\sqrt{X_i}} = \frac{\hat{\beta}_1}{\sqrt{X_i}} + \hat{\beta}_2\sqrt{X_i} + \hat{\beta}_3\frac{X_{3i}}{\sqrt{X_i}} + \hat{\beta}_4\frac{X_{4i}}{\sqrt{X_i}} + \hat{\beta}_5\frac{X_{5i}}{\sqrt{X_i}} + \frac{e_i}{\sqrt{X_i}} \tag{6.42}$$

$$Y_i^* = \beta_1^* + \beta_2^* X_{2i}^* + \beta_3^* X_{3i}^* + \beta_4^* X_{4i}^* + \beta_5^* X_{5i}^* + v_i \tag{6.43}$$

6.5.4.4 方差 σ_i^2 正比于 $\hat{Y}_i^{\,2}$

如果主模型式 6.27 存在异方差问题，且假设方差正比于 $\hat{Y}_i^{\,2}$，则有：

$$var(u_i) = E(u_i^2) = \sigma_i^2 = \sigma^2 \hat{Y}_i^{\,2} \tag{6.44}$$

对主模型式 6.27 两边同时除以 $\hat{Y}_i$，得到：

$$\frac{Y_i}{\hat{Y}_i} = \frac{\hat{\beta}_1}{\hat{Y}_i} + \hat{\beta}_2\frac{X_{2i}}{\hat{Y}_i} + \hat{\beta}_3\frac{X_{3i}}{\hat{Y}_i} + \hat{\beta}_4\frac{X_{4i}}{\hat{Y}_i} + \hat{\beta}_5\frac{X_{5i}}{\hat{Y}_i} + \frac{e_i}{\hat{Y}_i} \tag{6.45}$$

$$Y_i^* = \beta_1^* + \beta_2^* X_{2i}^* + \beta_3^* X_{3i}^* + \beta_4^* X_{4i}^* + \beta_5^* X_{5i}^* + v_i \tag{6.46}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

6.5.4.5 方差 σ_i^2 未知

如果主模型式 6.27 存在异方差问题，且假设方差正比于 $\hat{Y}_i^{\,2}$，则有：

$$var(u_i) = E(u_i^2) = \sigma_i^2 = \sigma^2 \hat{Y}_i^{\,2} \tag{6.47}$$

对主模型式 6.27 两边同时取对数 $ln()$，得到：

$$lnY_t = \hat{\beta}_1 + \hat{\beta}_2 lnX_{2i} + \hat{\beta}_l n3X_{3i} + \hat{\beta}_4 lnX_{4i} + \hat{\beta}_5 lnX_{5i} + e_i \tag{6.48}$$

6.5.5 使用 White 校正法矫正异方差问题

Eviews 操作（见图 6.23 ）

（1）依次选择 ⇒ Quick ⇒ Estimation Equation

（2）引导设置 Equation Estimation ⇒ Specification
- 方程设置（Equation Specification）：输入变量 Y c X2 X3 X4 X5

- 估计方法（Estimation settings）：
  - Method：选择 `LS - Least Squares(NLS and ARMA)`
  - Sample：默认设置

（3）引导设置 Equation Estimation ⇒ Options

- 系数协方差设置（Coefficient covariance）
  - 协方差方法（Coefficient method）：下拉选择 Huber-White
- 权重设置（Weights）：默认设置
- 最优化设置（Optimization）：默认设置
- 完成设置：点击 `OK`

（4）模型命名：建议为 `eq_adj_white`

查看分析报告（见图 6.24 ）：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

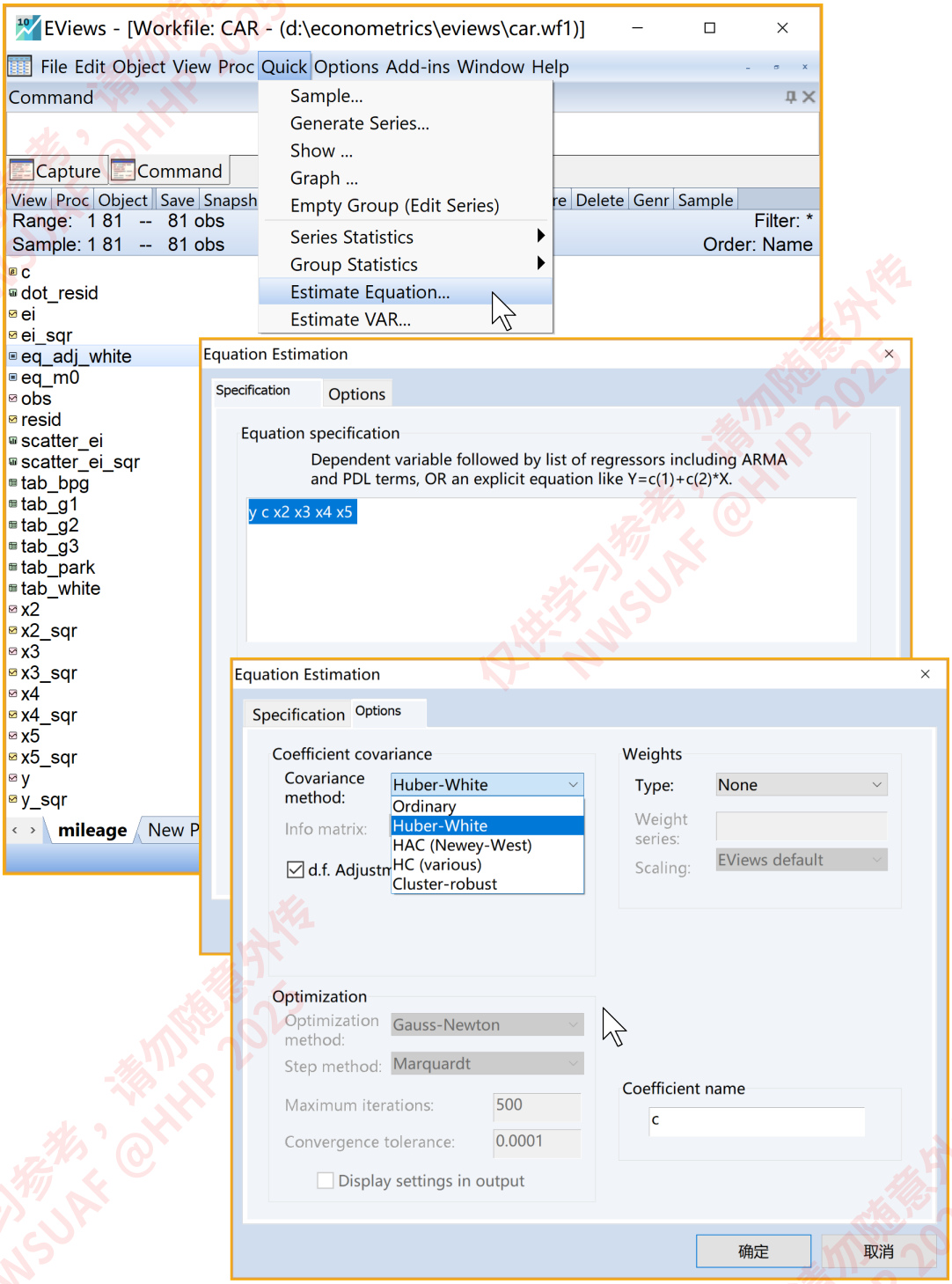


图 6.23: White 异方差矫正法的操作

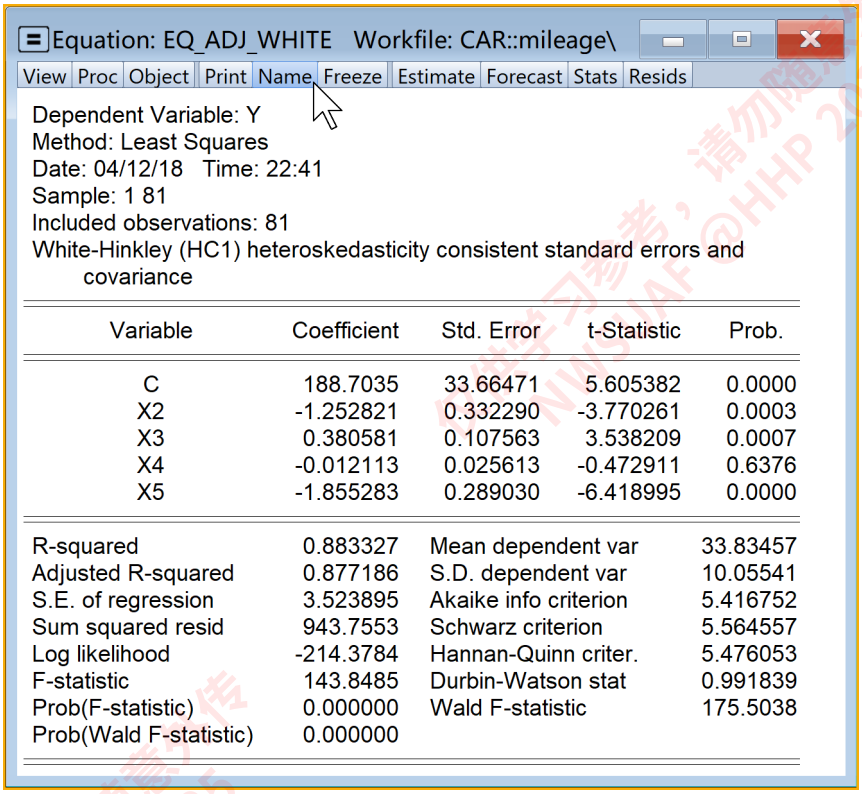


图 6.24: White 异方差矫正法的 Eviews 报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

6.6 作业题

财富 500 强企业高管薪水数据：表 6.3 给出给出了 447 辆汽车在 Y 高管薪水及分红，X2 担任 CEO 年数（不足 6 个月视为的 0），X3 总裁 CEO 的年龄，X4 企业的总销售收入，X5 企业的利润，X6 企业的总资产等方面的数据。

表 6.3: 财富 500 强企业高管薪水数据 (n=447)

obs	Y	X2	X3	X4	X5	X6
1	3030	7	61	161315	2956	257389
2	6050	0	51	144416	22071	237545
3	3571	11	63	139208	4430	49271
4	3300	6	60	100697	6370	92630
5	10000	18	63	100469	9296	355935
443	1866	10	59	2934	375	35800
444	906	6	59	2933	194	4986
445	2300	9	57	2910	183	2739
446	875	7	50	2905	132	5009
447	1758	5	62	2896	17	2855

变量说明见表 6.4 ：

表 6.4: 变量定义及说明

variable	label
obs	TOP500 公司序号
Y	高管薪水及分红
X2	担任 CEO 年数（不足 6 个月视为的 0）
X3	总裁 CEO 的年龄
X4	企业的总销售收入
X5	企业的利润
X6	企业的总资产

请考虑如下样本回归模型：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + e_i \tag{6.49}$$

请回答如下问题:

（1）根据回归模型式 6.49 ，写出总体回归模型（PRM），并对参数的理论预期（符号、大小、关系）进行说明。

（2）利用 Eviews 对样本回归模型式 6.49 进行回归分析（将报告截图过来，并写出相应的简要报告形式——三行式或四行式）。参数估计结果符合你的理论预期么？

（3）回归模型式 6.49 存在异方差问题的证据吗？请以此按照下列方法进行诊断，并分别得到分析结论（要求截图过来并进行简要说明）：

（a）非正式检验法（图解法）：

- 绘制 e_i 序列和 e_i^2 序列的描点图（dot plot），得到你的初步结论。
- 分别绘制 e_i 序列分别与 Y_i、$X_i (i = 2, \cdots, 6$ 序列的散点图（scatter plot），得到你的初步结论
- 分别绘制 e_i^2 与 Y_i^2、$X_i^2 (i = 2, \cdots, 6$ 序列的散点图（scatter plot），得到你的初步结论

（b）正式检验法

- 利用 Park 检验法，并得出你的初步结论（要求写出 park 诊断方程，并将 park 检验结果截图过来）
- 利用 Glejser 检验法，并得出你的初步结论 (请确定 X，并做三个类型的辅助回归。分别将 Glejser 检验结果截图过来）

$$|e_i| = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2i} + \cdots + \hat{\alpha}_k X_{ki} + v_i \qquad \text{G1}$$
$$|e_i| = \hat{\alpha}_1 + \hat{\alpha}_2 \sqrt{X_{2i}} + \cdots + \hat{\alpha}_k \sqrt{X_{ki}} + v_i \qquad \text{G2}$$
$$|e_i| = \hat{\alpha}_1 + \hat{\alpha}_2 \frac{1}{X_{2i}} + \cdots + \hat{\alpha}_k \frac{1}{X_{ki}} + v_i \qquad \text{G3}$$

- 利用 BPG 检验法，并得出你的初步结论（要求写出 BPG 诊断方程，并将 BPG 检验结果截图过来）
- 利用 White 检验法（注意交叉项的使用），并得出你的初步结论（要求写出 White 诊断方程，并将 White 检验结果截图过来）

（4）若发现存在异方差问题，你如何进行纠正主模型式 6.49 ？

（a）使用 White 校正法纠正异方差问题，并比较与主模型式 6.49 的差别。（要求分别截图两个回归方程的 Eviews 报告，进行对照分析并得到结论）

（b）使用加权最小二乘法纠正异方差问题（提醒：根据前述分析，几种处理方法选择一种合适的处理方法，进行模型矫正分析。请说明你选择这种处理办法的理由，并得到分析结论！）

（5）现在做以 $ln(Y_i)$ 为因变量的第二个模型。异方差性有所改善吗?（要求截图相关 Eviews 报告，并简要陈述理由）。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

7 自相关问题

仅供学习参考，请勿随意外传
NWSUAF @HHP 2025

7.1 实验目的及要求

目的：掌握自相关问题的检验与处理方法。

要求：在老师指导下完成计量经济模型的自相关检验，并对存在自相关的模型进行修正，最终得到正确的分析结果。

7.2 实验原理

使用时间序列数据建模时，我们往往会面临时间序列数据自相关的现实问题。如果存在自相关问题，则可能对模型及分析结论的科学性产生一定程度影响。

7.2.1 什么时候模型会出现自相关问题?

对于总体回归模型（PRM）：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \tag{7.1}$$

随机干扰项 u_t 可能与其滞后变量存在自相关现象，我们定义如下：

定义 7.1 (随机干扰项 p 阶自相关)**.** 对于给定的滞后阶数 $p(p \subset \mathcal{N})$ 下，随机干扰项 u_t 与其滞后变量 u_{t-p} 之间存在 $cov(u_t, u_{t-p}) \neq 0$。

在经典线性回归模型假设（CLRM）下，我们假设主模型式 7.1 随机误差项 u_t 是无自相关的，也即 u_t 与 u_{t-1} 或任意其他滞后变量 u_{t-p}，$p \in (1, 2, \cdots)$ 都不存在线性相关关系。如果这一假设不能得到保证，我们认为模型将出现自相关问题。

定义 7.2 (模型出现自相关问题)**.** 对于主回归模型，如果给定**任意的**滞后阶数 $p \in (1, 2, \cdots)$，只要随机干扰项 u_t 与其滞后变量 u_{t-p} 之间存在 $cov(u_t, u_{t-p}) \neq 0$ 的现象。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

此时，模型式 7.1 出现自相关问题也可以表达为方差-协方差的矩阵形式：

$$E(\mathbf{uu}'|\mathbf{X}) = \begin{bmatrix} E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n) \\ \cdots & \cdots & \cdots & \cdots \\ E(u_nu_1) & \cdots & \cdots & E(u_nu_n) \end{bmatrix} \tag{7.2}$$

$$E(\mathbf{uu}'|\mathbf{X}) = \begin{bmatrix} \sigma^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma^2 & \cdots & \sigma_{2n}^2 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1}^2 & \cdots & \cdots & \sigma^2 \end{bmatrix} \tag{7.3}$$

其中，方差-协方差矩阵（见公式式 7.2 ）的协方差部分（矩阵下三角）元素将不全为 0（见公式式 7.3 ）。

简而言之，只要的随机干扰项 u_t 出现了任何形式的 p 阶自相关，我们都称模型式 7.1 出现自相关问题。一种简单的情形就是马尔科夫 1 阶自相关：

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \\ \varepsilon_t &\sim \text{i.i.d} \quad N(0, \sigma^2) \end{aligned}$$

当然，模型自相关的情形还有其他更加多样和复杂的形态，例如：

$$\begin{aligned} u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t \\ u_t &= \rho_2 u_{t-2} + \varepsilon_t \\ &\cdots \\ \varepsilon_t &\sim \text{i.i.d} \quad N(0, \sigma^2) \end{aligned} \tag{7.4}$$

7.2.2 随机干扰项自相关有哪些表现形态?

对于马尔科夫 1 阶自相关情形（见公式式 7.4 ），如果自相关系数 $-1 < \rho < 0$，则随机干扰项 u_t 及其一阶滞后变量具有某种图形模式（见图 7.1）：其中 u_t 的时序图表现为正弦波形态，而 u_t 与 u_{t-1} 的散点图表现为正的线性相关关系。

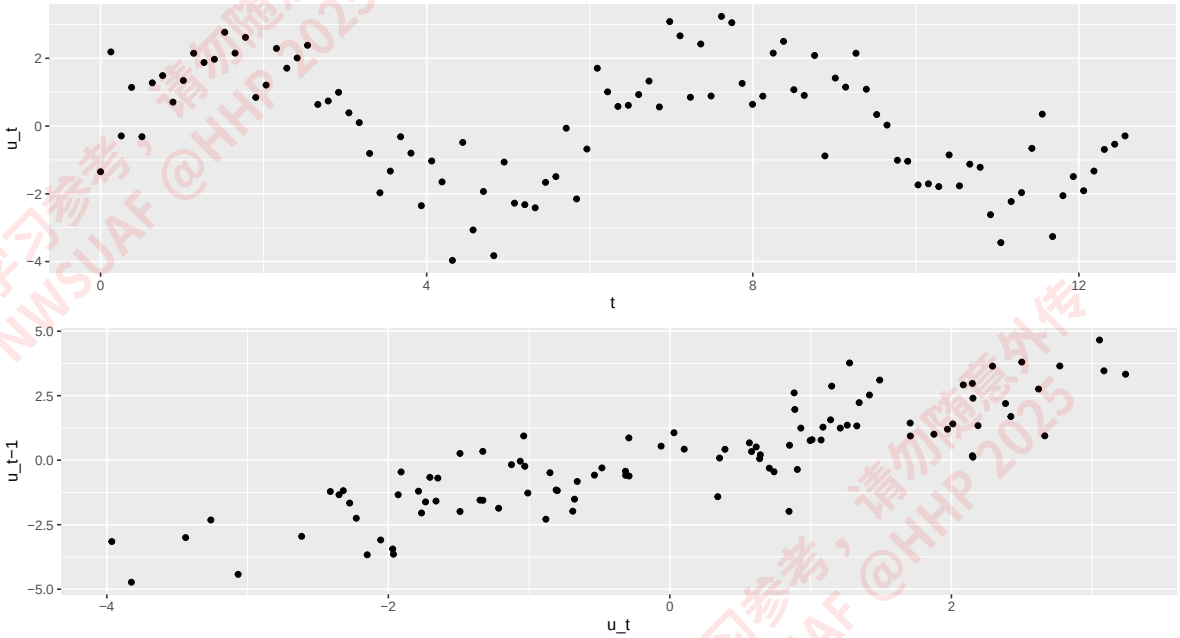


图 7.1: 随机干扰项自相关模式图

7.2.3 自相关问题模型会有什么后果?

模型出现自相关问题时时，主要的影响包括：

- 普通最小二乘法（OLS）估计量仍是线性的、无偏的和一致性的
- 普通最小二乘法（OLS）估计量将不再是有效的 (亦即最小方差)
- 模型参数的显著性检验失去意义
- 模型的预测将失效
- 模型的异方差检验流程也会变得不可靠

我们可以对马尔科夫一阶自相关情形做简要的证明：

马尔科夫 *1* 阶自相关下的理论后果. 已知模型存在马尔科夫一阶自相关情形，则有：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{aligned}
Y_t &= \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \\
u_t &= \rho u_{t-1} + \varepsilon_t \\
\varepsilon_t &\sim \text{i.i.d} \quad N(0, \sigma_\varepsilon^2)
\end{aligned}
$$

在经典线性回归模型假设下（CLRM），执意采用普通最小二乘法（OLS），我们容易可以得到斜率参数的点估计值 $\hat{\beta}_2$ 及其方差 $\text{var}(\hat{\beta}_2)$：

$$
\begin{aligned}
\hat{\beta}_2 &= \frac{\sum x_t y_t}{\sum x_t^2} \\
E(\hat{\beta}_2) &= \beta_2 \\
\text{var}(\hat{\beta}_2)_{CLRM}^{OLS} &= \frac{\sigma^2}{\sum x_t^2}
\end{aligned}
$$

反之，如果我们考虑到一阶自相关问题已经**违背**了经典线性回归模型假设（CLRM），但仍采用普通最小二乘法（OLS），则斜率参数的点估计值 $\hat{\beta}_2$ 及其方差 $\text{var}(\hat{\beta}_2)$ 的理论计算结果应该为：

$$
\begin{aligned}
\hat{\beta}_2 &= \frac{\sum x_t y_t}{\sum x_t^2} \\
E(\hat{\beta}_2) &= \beta_2 \\
\text{var}(\hat{\beta}_2)_{AR1}^{OLS} &= \frac{\sigma^2}{\sum x_t^2} + \frac{2\sigma^2}{\sum x_t^2} \\
&\quad \cdot \left[\rho \frac{\sum_{t=1}^{n-1} x_t x_{t+1}}{\sum_{t=1}^{n} x_t^2} + \rho^2 \frac{\sum_{t=1}^{n-2} x_t x_{t+2}}{\sum_{t=1}^{n} x_t^2} + \cdots + \rho^{n-1} \frac{x_1 x_n}{\sum_{t=1}^{n} x_t^2} \right] \\
&\simeq \frac{\sigma^2}{\sum x_t^2} \left(\frac{1 + r\rho}{1 - r\rho} \right)
\end{aligned}
$$

其中，r 为简单相关系数。因此，我们可以发现：

$$
\begin{aligned}
\text{var}(\hat{\beta}_2)_{AR1}^{OLS} &\simeq \frac{\sigma^2}{\sum x_t^2} \left(\frac{1 + r\rho}{1 - r\rho} \right) \\
&= \text{var}(\hat{\beta}_2)_{CLRM}^{OLS} \cdot \left(\frac{1 + r\rho}{1 - r\rho} \right)
\end{aligned}
$$

从而证明了 $\text{var}(\hat{\beta}_2)_{AR1}^{OLS}$ 与 $\text{var}(\hat{\beta}_2)_{CLRM}^{OLS}$ 是不相等的，也即表明当模型存在一阶自相关问题时，普通最小二乘法估计量仍是线性和无偏的，但却不再是方差最小的。[1] □

[1] 读者可以验证 $r = 0.6, \rho = 0.8$ 时，二者的方差大小关系

7.2.4 如何诊断模型存在自相关问题?

显然，模型如果存在自相关问题，应该努力侦察随机干扰项 u_t 及其滞后变量 $u_{t-1}, \cdots, u_{t-p}$ $(p \in \mathcal{N})$ 的行为模式。

对于总体回归模型存在一阶马尔科夫自回归情形：

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \qquad \text{(PRM-AR1)} \\ \varepsilon_t &\sim \text{i.i.d} \quad N(0, \sigma_\varepsilon^2) \end{aligned} \tag{7.5}$$

我们也有理由认为，样本回归模型（SRM）也会继承总体回归模型（PRM）的特征，从而有：

$$\begin{aligned} Y_t &= \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \cdots + \hat{\beta}_k X_{kt} + e_t \\ e_t &= \hat{\rho} e_{t-1} + v_t \qquad \text{(SRM-AR1)} \\ v_t &\sim \text{i.i.d} \quad N(0, \sigma_v^2) \end{aligned} \tag{7.6}$$

鉴于样本回归模型（SRM）的残差 e_t 及其滞后变量 $e_{t-1}, \cdots, e_{t-p}$ $(p \in \mathcal{N})$ 是可观测的，因此诊断模型是否存在自相关问题往往就转化为侦察残差 e_t 及其滞后变量 $e_{t-1}, \cdots, e_{t-p}$ $(p \in \mathcal{N})$ 的行为模式。总体来看，诊断模型是否存在自相关问题，主要包括两大类方法：

（1）图形观察法：

- 残差序列观察法（描点图法）：绘制 e_t 序列的描点图（dot plot）
- 残差序列观察法（描点图法）：确定滞后阶数并分别绘制 e_t 序列与 $e_{t-1}, e_{t-2}, \cdots$ 序列的散点图（scatter plot）

（2）定量分析法：

- 辅助回归法：构建残差 e_t 序列对 $e_{t-1}, e_{t-2}, \cdots$ 序列的辅助回归方程
- 自相关和偏相关分析法：Eviews 菜单操作对残差 e_t 序列进行自相关和偏相关分析
- Durbin-Watson 检验法：分析 Eviews 报告中的 D-W 统计量
- 拉格朗日检验法（LM-test）：Eviews 菜单操作进行布罗施-戈弗雷 (Breusch-Goldfrey) 的拉格朗日检验 (B-G LM test)

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

显然，模型无自相关是一种假设的理想状态。而事实是，模型只要存在任何一种情形的自相关模式，我们都称模型出现了自相关问题。因此，我们使用以上任一诊断方法，只要发现了自相关的明显证据，则都应该引起重视。

辅助回归方程诊断法：分析主回归模型式 7.8 ，然后或者通过图形观察，或者通过猜测和尝试，再对残差及其滞后变量进行辅助回归式 7.9 的建模分析：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \quad \text{(PRM)} \tag{7.7}$$

$$Y_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \cdots + \hat{\beta}_k X_{kt} + e_t \quad \text{(SRM)} \tag{7.8}$$

$$e_t = \hat{\rho}_1 e_{t-1} + \hat{\rho}_2 e_{t-2} + \cdots + \hat{\rho}_p e_{t-p} + \varepsilon_t \quad \text{(auxiliary equation)} \tag{7.9}$$

其中，$p = 1, 2, \cdots, t-1$。实际上，具体的辅助诊断模型式 7.9 往往需要多次尝试，由简单模型逐渐到扩展模型。为简单起见，我们可以先做 AR(1) 辅助方程（也即 e_t 对 e_{t-1} 的回归）或 AR(2) 辅助方程（也即 e_t 对 e_{t-1} 和 e_{t-2} 的辅助回归）。

辅助回归方程诊断法的诊断标准是：

- 如果诊断辅助方程式 7.9 的 F 检验**不显著**（对应的概率值 P>0.1），则表明主模型式 7.8 无自相关问题（注意是指表现为辅助方程形式的自相关问题）
- 如果诊断辅助方程式 7.9 的 F 检验**显著**（对应的概率值 P<0.1），则表明主模型式 7.8 存在自相关问题（注意是指表现为辅助方程形式的自相关问题）

7.2.5 如何矫正存在自相关问题模型?

对于存在自相关问题的模型，我们的最终目标应该是矫正并得到一个替代的新模型，使得该新模型能满足经典线性回归模型假设（CLRM），从而可以继续采用普通最小二乘法（OLS），最终估计并获得总体参数的最优线性无偏估计量（BLUE）。

自相关问题模型矫正的前提，首要工作是判明问题模型的自相关形式。它可以是简单的马尔科夫一阶自相关（见公式式 7.5 ），也可以是更复杂的高阶自相关形式。然后再采用针对性的方法，把自相关问题模型处理为无自相关问题的新模型。从模型矫正的操作方法来看，这些针对性的矫正方法可以分

为广义差分方程法（General difference equation，GDE ）[2]和一致标准误校正法（Heteroscedasticity and autocorrelation consistent standard error，HAC）[3]。当然，从实际矫正操作流程来看，矫正方法也可以划分为如下两大类：

（1）广义差分方程两步法（Two-steps）：

- 广义最小二乘法 (GLS)：一阶差分法变换
- 广义最小二乘法 (GLS)：基于残差辅助方程近似得到 ρ
- 广义最小二乘法 (GLS)：基于 D-W 统计量近似计算得到 ρ
- 广义最小二乘法 (GLS)：基迭代法近似计算得到 ρ

（2）标准误校正一步法（One-steps）：

- 一致标准误校正法（HAC）：尼威-威斯特 (Newey-West) 校正法

7.3 实验内容

（1）采用最小二乘法建立主回归模型

（2）自相关问题模型的侦察方法

- 残差序列观察法（描点图法）：绘制 e_t 序列的描点图（dot plot）
- 残差序列观察法（描点图法）：确定滞后阶数并分别绘制 e_t 序列与 $e_{t-1}, e_{t-2}, \cdots$ 序列的散点图（scatter plot）
- 辅助回归法：构建残差 e_t 序列对 $e_{t-1}, e_{t-2}, \cdots$ 序列的辅助回归方程
- 自相关和偏相关分析法：Eviews 菜单操作对残差 e_t 序列进行自相关和偏相关分析（**注意滞后阶数的选择**）
- Durbin-Watson 检验法：分析 Eviews 报告中的 D-W 统计量
- 拉格朗日检验法（LM-test）：Eviews 菜单操作进行布罗施-戈弗雷 (Breusch-Goldfrey) 的拉格朗日检验 (B-G LM test)

[2]本质上属于广义最小二乘法（General least square，GLS）

[3]这种方法是对异方差怀特一致性标准误矫正法（White heteroscedasticity consistent standard error）的扩展

（3）自相关问题模型的矫正方法：

- 广义最小二乘法 (GLS)：一阶差分法变换
- 广义最小二乘法 (GLS)：基于残差辅助方程近似得到 ρ
- 广义最小二乘法 (GLS)：基于 D-W 统计量近似计算得到 ρ
- 广义最小二乘法 (GLS)：基迭代法近似计算得到 ρ
- 一致标准误校正法（HAC）：尼威-威斯特 (Newey-West) 校正法

7.4 实验准备

7.4.1 实验软件

本次实验需要提前准备好如下软件：

- 统计分析软件 Eviews 9.0 版本及以上
- 公式编辑软件 Mathtype 6.0 版本及以上
- 写作编辑软件 Office Word/Excel 2010 版本及以上
- 浏览器软件 chrome 66.0 版本及以上或 360 极速浏览器 9.5 版本及以上

7.4.2 实验材料

商业部门的工资水平与生产率：表 7.1 给出给出了美国 46 个商业部门 Y 部门工资水平，X 部门生产率等方面的数据。

表 7.1: 美国部门工资数据 (n=46)

obs	Y	X
1960	61	49
1961	62	51
1962	65	53
1963	66	55

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 7.1: 美国部门工资数据 (n=46)

obs	Y	X
1964	68	57
2001	114	119
2002	116	124
2003	118	129
2004	119	133
2005	120	136

变量说明见表 7.2：

表 7.2: 变量定义及说明

variable	label
obs	年份
Y	部门工资水平
X	部门生产率

请考虑如下样本回归模型：

$$log(Y_t) = \hat{\beta}_1 + \hat{\beta}_2 log(X_{2t}) + e_t \tag{7.10}$$

7.4.3 实验规则

本实验将要求保留 Eviews 操作过程的相关结果，因此对 Eviews 对象命名规则设计如下：

（1）方程对象（Equation）的命名规则：

- 主回归方程对象保存命名为eq_m0
- 自相关问题诊断中的辅助回归方程，根据具体诊断类型，酌情保存并命名为：残差 AR1 辅助方程诊断法：eq_ar1_test；…
- 自相关问题 GLS 矫正后的回归方程，根据具体矫正类型，酌情保存并命名为：

提示

- 基于残差辅助方程结果进行 1 阶广义差分（GLS）矫正：eq_adj_ar1
- 基于 D-W 统计量结果进行广义差分（GLS）矫正：eq_adj_dw
- 自相关问题可行广义最小二乘法 (FGLS) 矫正后的回归方程，根据具体矫正类型，酌情保存并命名为：
- 基于科克伦-奥克特迭代法 (Cochrane-Orcutt iterative procedure)：eq_adj_co

- 自相关问题一致性标准误校正法（HAC），也即尼威-威斯特 (Newey-West) 校正法矫正后的回归方程，保存并命名为：eq_adj_nw

（2）序列对象（Series）的命名规则：

- （主回归方程的）残差序列，保存并命名为et
- （主回归方程的）标准化残差序列，保存并命名为et_sd
- （主回归方程的）倍数放大（或缩小）的残差序列（此处为放大 100 倍，具体应视案例而定），保存并命名为et_100

-（主回归方程的）残差滞后序列，酌情分别保存并命名为：et_l1；et_l2；…。

（3）图形对象（Graph）的命名规则：

- 残差 e_t、标准化残差 e_t^*、倍数放大残差 $\hat{e}_t$ 的描点图（dot plot），绘制成一张图，保存并命名为并命名为：dot_resid
- 残差 e_t 相对于残差一阶滞后变量 e_{t-1} 的散点图（scatter），保存并命名为并命名为：scatter_et

（4）表格对象（Table）的命名规则：

- 自相关问题的自相关和偏相关诊断法，另存并命名为：tab_corrl_test
- 自相关问题的拉格朗日诊断法（LM-test），另存并命名为：tab_LM_test

（5）标量对象（Scalar）的命名规则。自相关问题模型矫正的不同方法中，可能会计算得到 ρ 的近似 $\hat{\rho}$，具体命名视不同矫正方法：

- 基于残差辅助方程结果进行 1 阶广义差分（GLS）矫正：rho_ar1

- 基于 D-W 统计量结果进行广义差分（GLS）矫正：#rho_dw
- 基于科克伦-奥克特迭代法的可行广义最小二乘法 (FGLS) 矫正：#rho_co

7.5 主要实验步骤

7.5.1 导入数据并进行预处理

新建 Eviews 工作文件（见图 7.2）。Eviews 菜单操作如下：

- 依次操作：File⇒New⇒Workfile
- 进行 workfile create 引导设置：

> 提示
>
> 具体设置如下：
> a）workfile structure type: `unstructured/undatede`
> b）data range：46
> c）workfile names(optional):
>
> - WF: `wage`（建议命名）
> - Page: `corrl`（建议命名）

Eviews 导入数据。菜单操作（Excel 和 Eviews）步骤如下：

- 提示：Excel 数据，每个同学的 Y 数据都不同，找到自己学号对应下的 Y 数据（X 数据所有同学都一样）
- Excel 找到数据。Excel 表格中仅保留自己需要的数据（obsYX）
- Eviews 导入数据。File⇒Import⇒Import From File：`d:/ econometrics/ data/ Lab7-wage.xlsx`

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

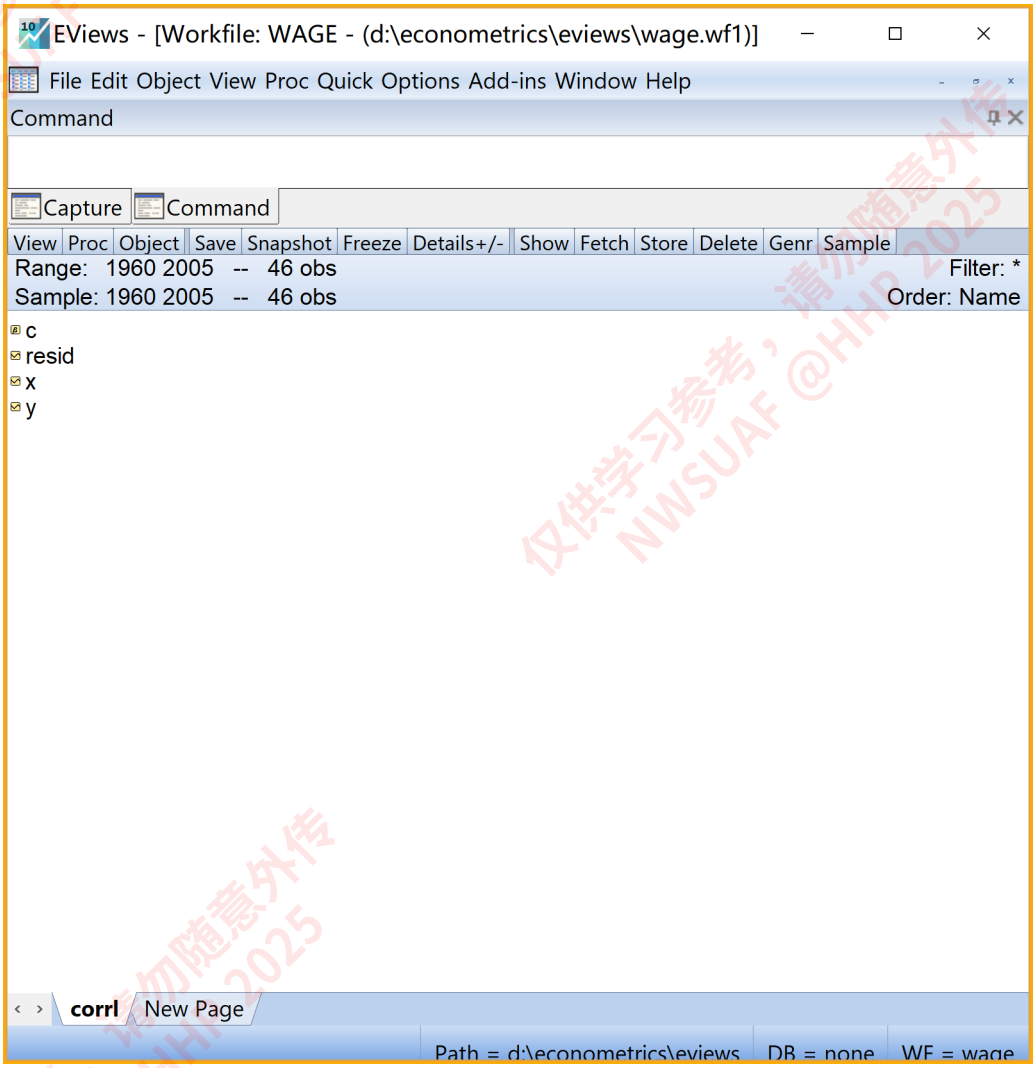


图 7.2: 导入数据的 Eviews 视窗

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

7.5.2 采用最小二乘法建立主回归模型

主回归模型为

$$log(Y_t) = \hat{\beta}_1 + \hat{\beta}_2 log(X_{2t}) + e_t$$

Eviews 菜单操作（见图 7.3）：

（1）依次选择 ⇒Quick⇒Estimation Equation

（2）引导设置 Equation Estimation⇒specification

> 提示
>
> 具体设置如下
> a）Equation specification：
>
> - 输入命令 log(Y) c log(X)
>
> b）Estimation settings：
>
> - Method: 下拉选择 `LS - Least Squares (NLS and ARMA)`
> - Sample: **默认设置**
>
> c）点击 `OK`

（3）模型命名：建议为 `eq_m0`

主回归分析结果见图 7.4：

7.5.3 侦查模型是否存在自相关问题

诊断方法包括残差观察法、辅助回归法、自相关和偏相关分析法、Durbin-Watson 检验法、拉格朗日检验法（LM-test）等。分析 Eviews 诊断报告，与相关参考标准进行比较，得到相关结论

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

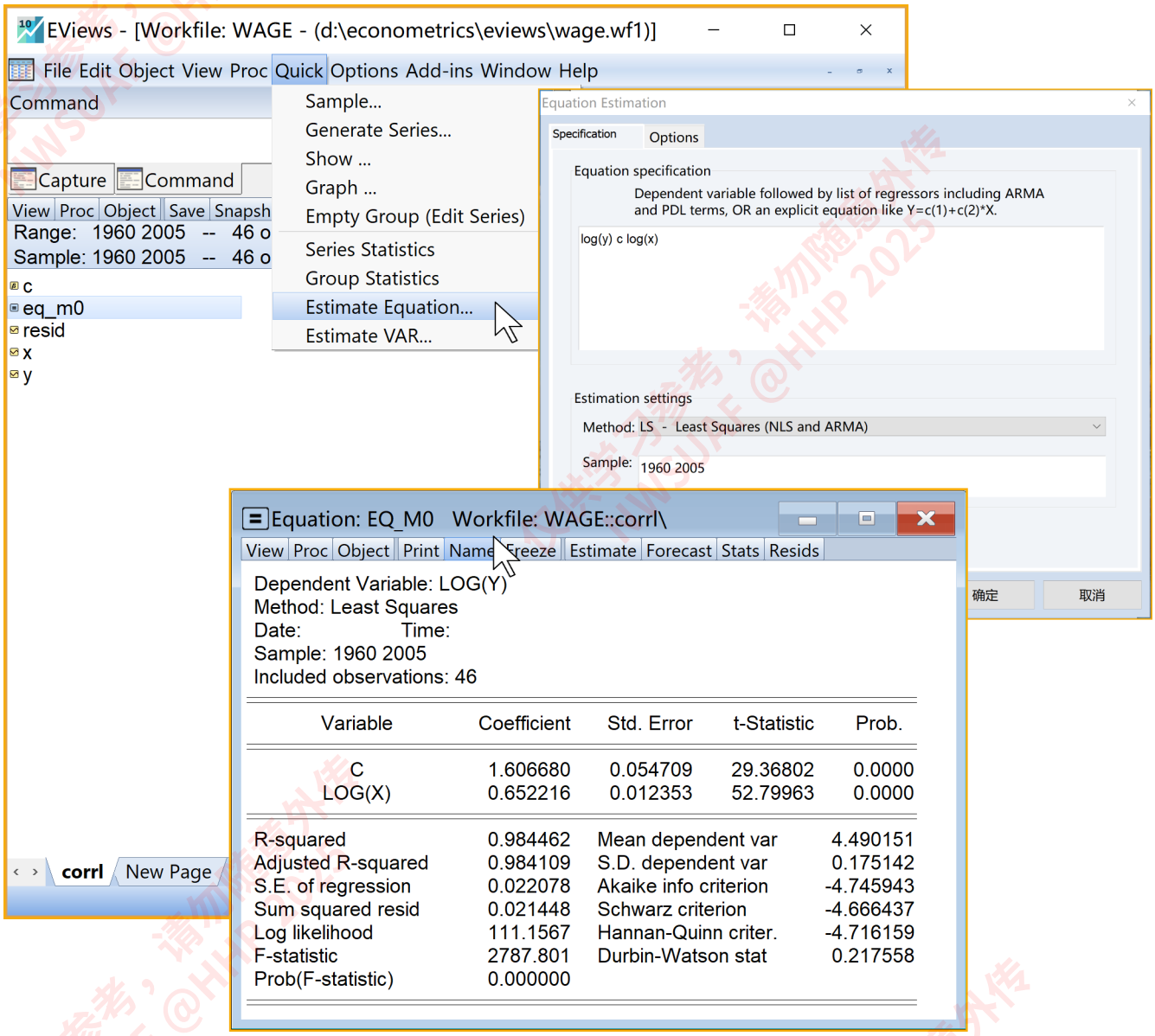


图 7.3: 主回归模型 Eviews 操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

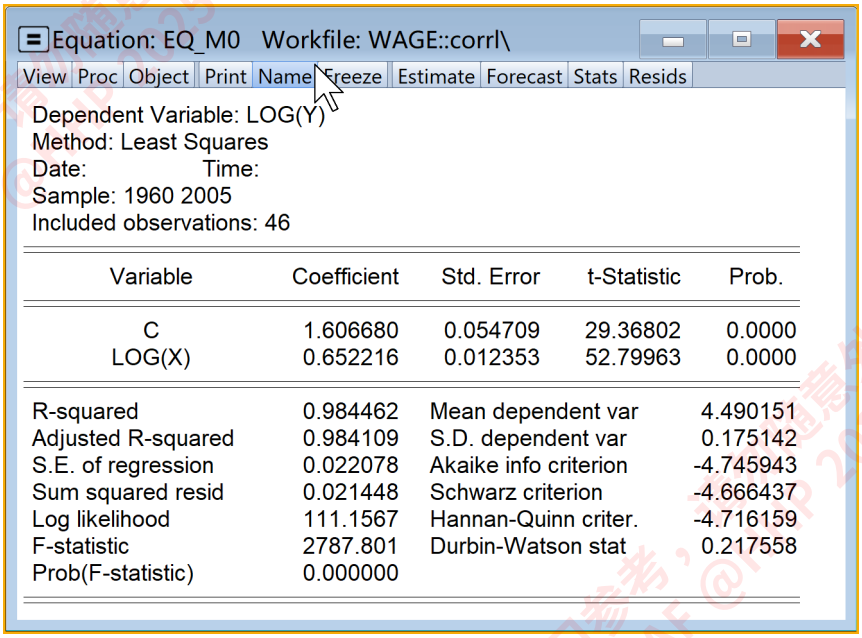


图 7.4: 主回归模型 Eviews 报告

7.5.3.1 残差观察法

目标：观察残差序列、标准化残差序列，以及残差滞后序列之间的图形模式和图形关系的图形模式

思路：判定 $e_t, e_t^*, e_{t-1}, \cdots, e_{t-s}$ 等的图形模式和关系

> 提示
>
> **提示：**
>
> - 描点图（dot plot）是分析一个变量的图形模式。例如残差 e_t 或标准化残差 e_t^*（做纵轴）相对于时间 t（做横轴）的图形关系
> - 散点图（scatter plot）是分析两个变量之间的图形模式。例如 e_t（做纵轴），相对于 e_{t-1} 或 e_{t-2} 或 e_{t-s} 的图形关系。

Eviews 菜单操作：

（1）分别生成残差序列 e_t、放大 100 倍的残差（请根据实际情况酌情考虑调整倍数，目的是便于与标准化残差做比较）、标准化残差序列 e_t^* 和滞后一阶变量 e_{t-1} 序列（见图 7.5）

- 命令视窗（Command）依次输入命令（建议分别命名为 `et`、`et_100`、`et_sd` 和 `et_l1`）

```
series et=resid
series et_100=100*resid
series et_sd=et/@stdev(et)
series et_l1=et(-1)
```

- 运行命令：命令行中按 Enter 键，并可以查看运行结果

> 提示
>
> 查看运行结果：
>
> - 双击et
> - 双击et_100
> - 双击et_sd
> - 双击et_l1

（2）绘制残差 e_t 和标准化残差 e_t^* 序列的描点图（dot plot）（见图 7.6）

- 选择序列对象：键盘 Ctrl 键 + 依次单击选择序列et 和et_sd
- 进入引导菜单：⇒ Quick ⇒ Graph

> 提示
>
> 具体设置细节：
>
> - 选择绘图类型 (Graph type)：Dot plot
> - 选择绘图细节 (Detail)：⇒ Multiple series ⇒ 下拉选择 Multiple graphs

- 点击完成：OK
- 命名并保存绘图（graph）对象：建议命名为 `dot_resid`
- 查看结果：双击dot_resid（见图 7.7）

（3）绘制 e_t 序列对 e_{t-1} 的散点图（scatter plot）（见图 7.8）

- 选择序列对象：键盘 Ctrl 键 + 依次单击选择序列et；et_l1
- 进入引导菜单：⇒ Quick ⇒ Graph

> 提示
>
> 具体设置细节：
>
> - 选择绘图类型 (Graph type)：Scatter
> - 选择绘图细节 (Detail)：**默认设置**

- 点击完成：OK
- 命名并保存绘图（graph）对象：建议命名为 scatter_et
- 查看结果：双击scatter_et（见图 7.8）

7.5.3.2 辅助回归法

目标：根据残差图模式，构建残差辅助回归方程，得到你的初步结论

思路：分析残差 e_t 序列对 $e_{t-1}, e_{t-2}, \cdots$ 序列的辅助回归方程式 7.9 的 Eviews 报告

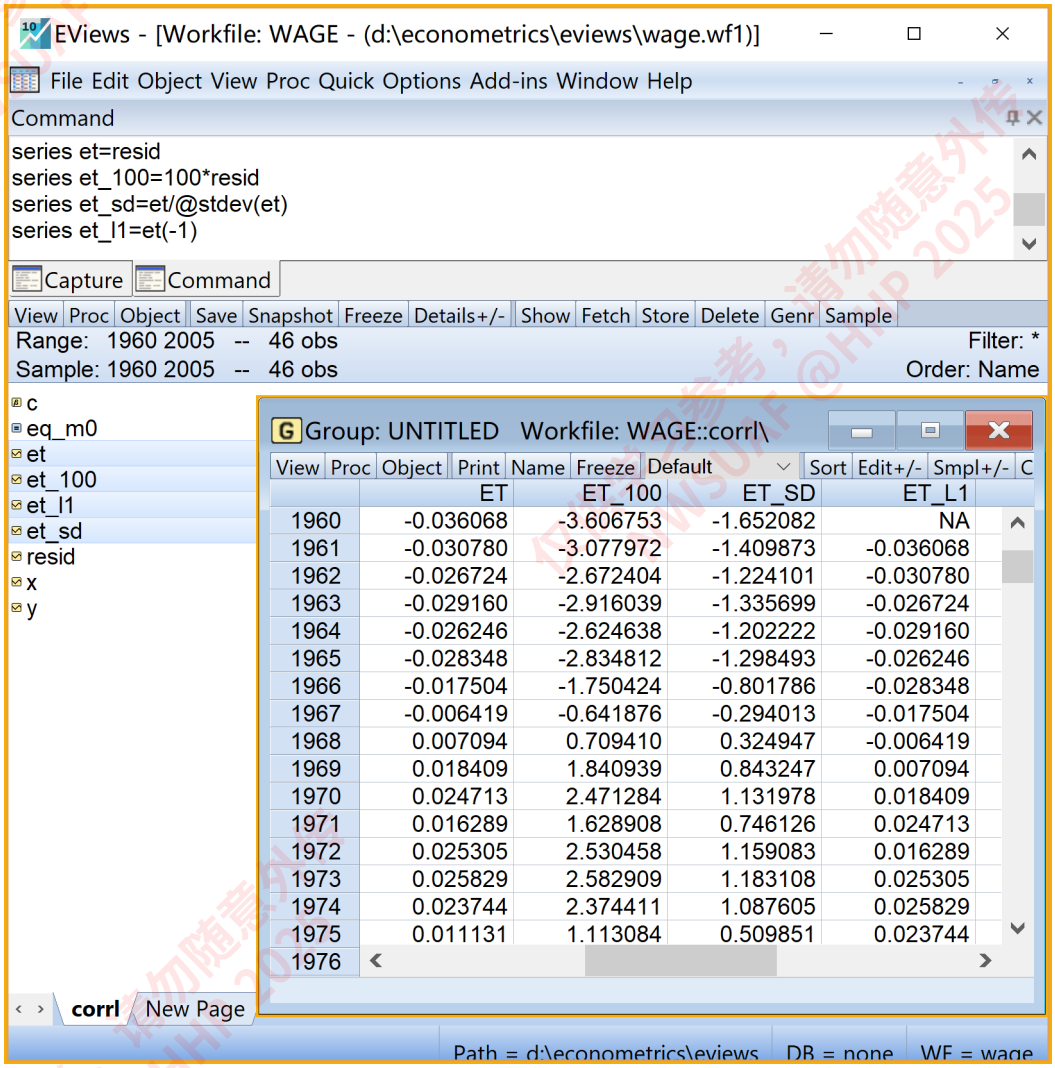
Eviews 操作（此处仅展示 AR(1) 辅助回归方程的情形，菜单操作实现具体见图 7.9）：

（1）依次选择 ⇒ Quick ⇒ Estimation Equation

（2）引导设置 Equation Estimation ⇒ specification
- Equation specification：输入命令 `et et_l1`

- Estimation settings：a）Method: 下拉选择 LS - Least Squares (NLS and ARMA)；b）Sample: （默认设置）
- 点击完成：OK
- 命名保存方程对象：（建议命名为 eq_ar1_test）
- 查看结果：双击eq_ar1_test

具体 Eviews 报告见图 7.10：



	ET	ET_100	ET_SD	ET_L1
1960	-0.036068	-3.606753	-1.652082	NA
1961	-0.030780	-3.077972	-1.409873	-0.036068
1962	-0.026724	-2.672404	-1.224101	-0.030780
1963	-0.029160	-2.916039	-1.335699	-0.026724
1964	-0.026246	-2.624638	-1.202222	-0.029160
1965	-0.028348	-2.834812	-1.298493	-0.026246
1966	-0.017504	-1.750424	-0.801786	-0.028348
1967	-0.006419	-0.641876	-0.294013	-0.017504
1968	0.007094	0.709410	0.324947	-0.006419
1969	0.018409	1.840939	0.843247	0.007094
1970	0.024713	2.471284	1.131978	0.018409
1971	0.016289	1.628908	0.746126	0.024713
1972	0.025305	2.530458	1.159083	0.016289
1973	0.025829	2.582909	1.183108	0.025305
1974	0.023744	2.374411	1.087605	0.025829
1975	0.011131	1.113084	0.509851	0.023744
1976				

图 7.5: 生成相关变量

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

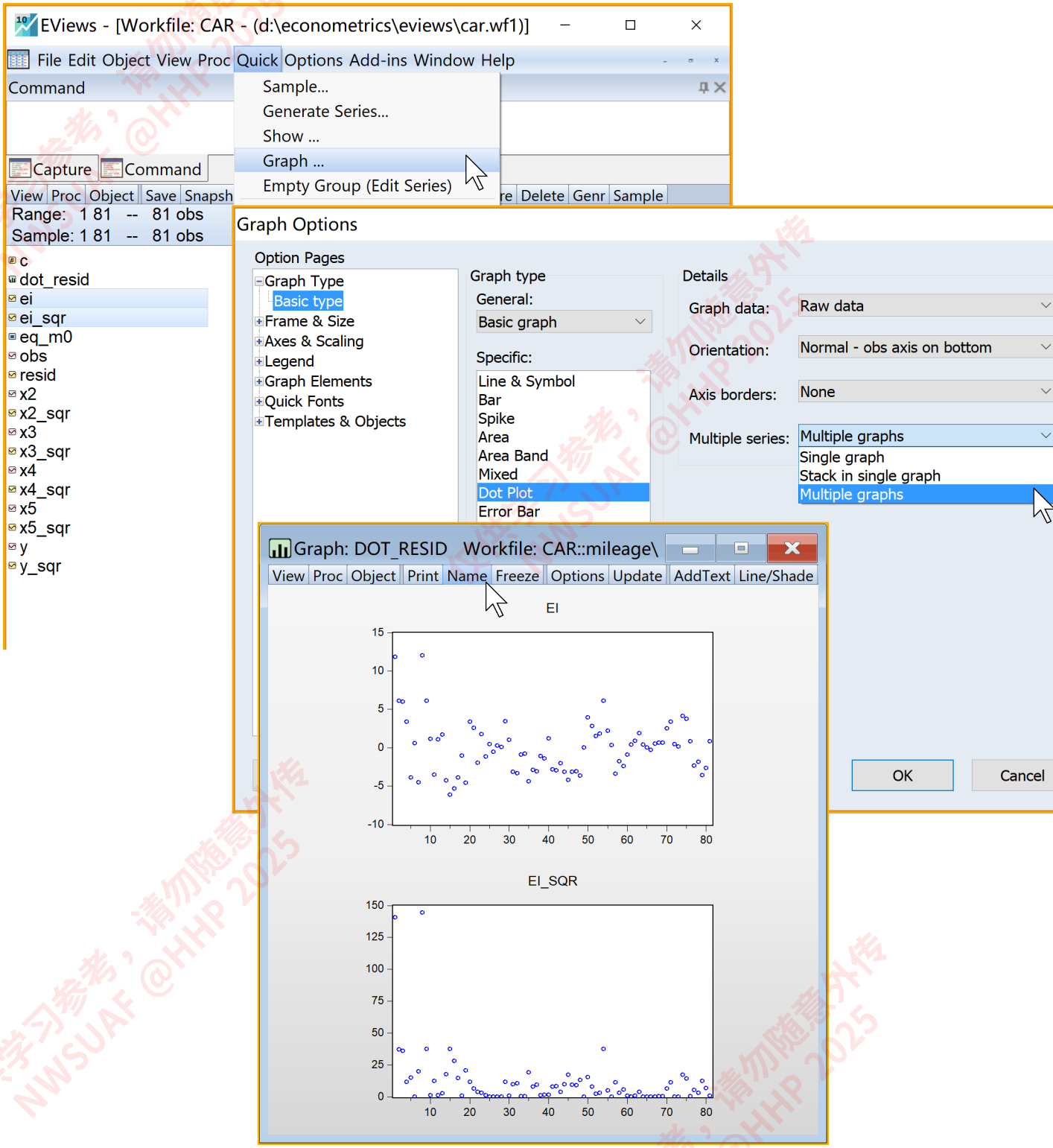


图 7.6: 残差及标准化残差的描点图



图 7.7: 残差及标准化残差的描点图报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

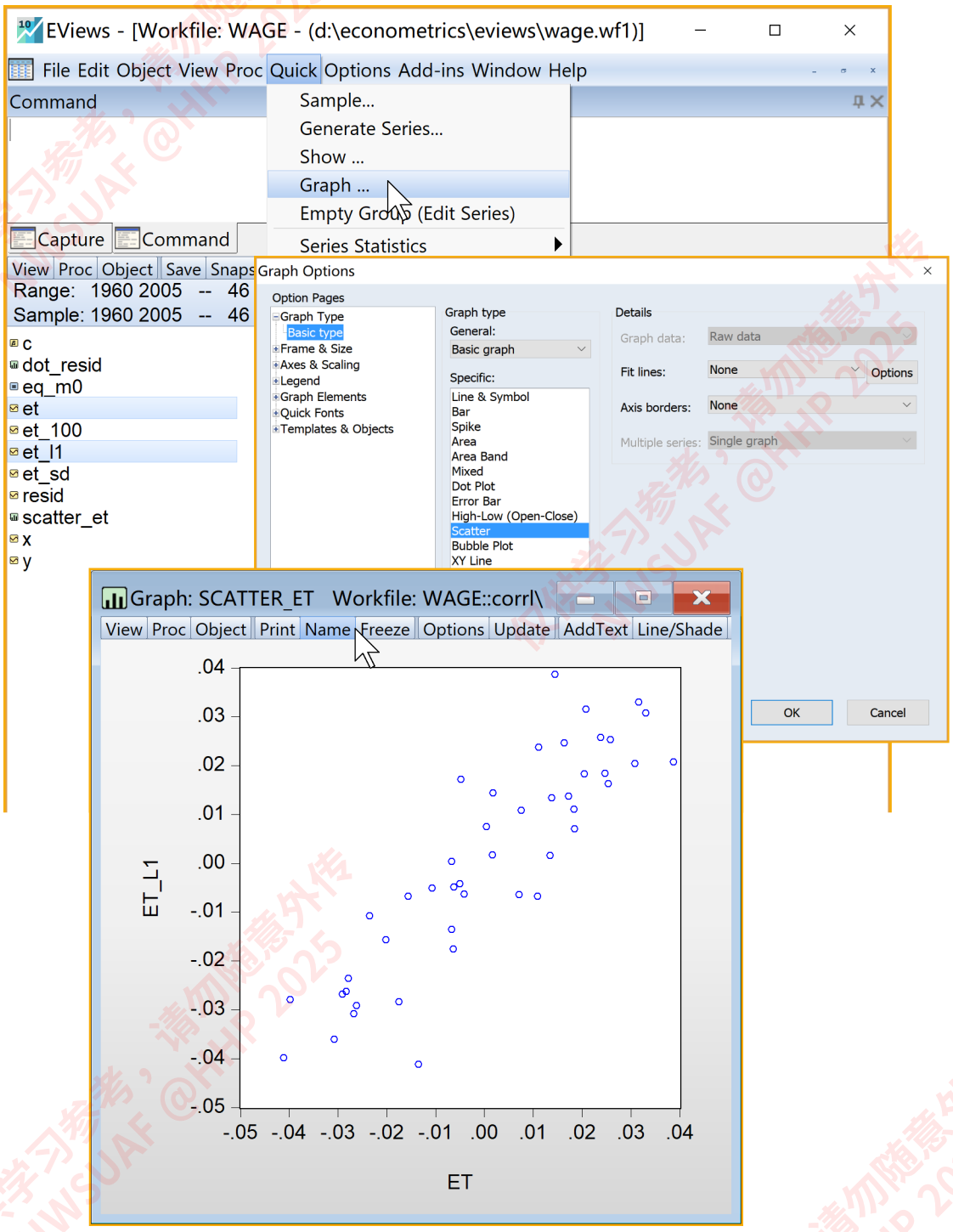


图 7.8: 残差与残差滞后变量的散点图

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

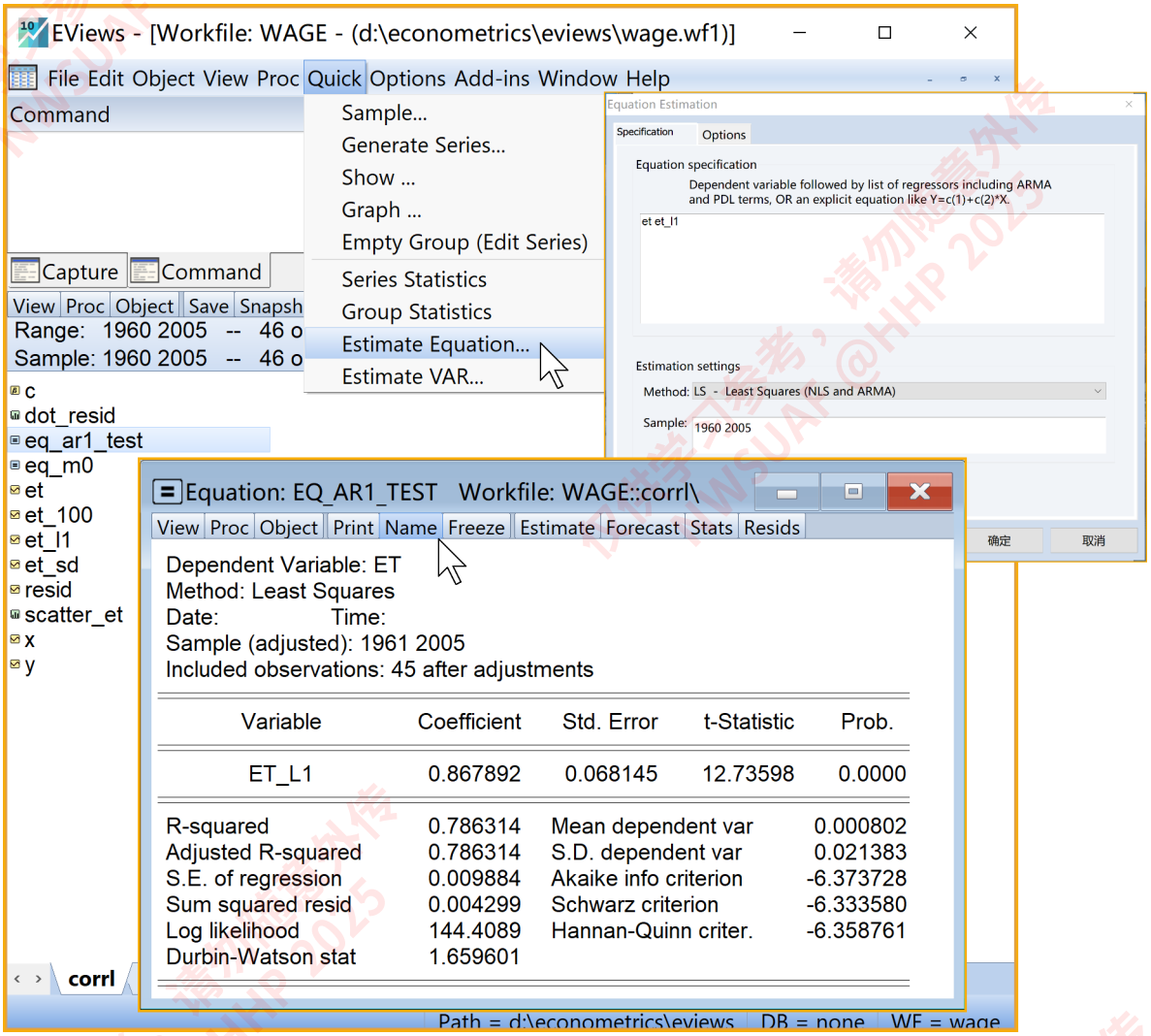


图 7.9: 残差辅助回归方程的 Eviews 诊断操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

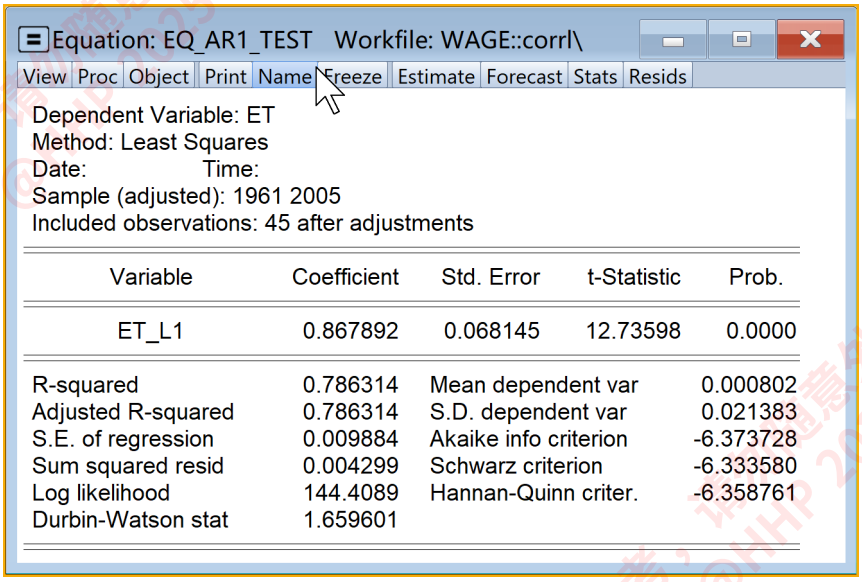


图 7.10: 残差辅助回归方程的诊断报告

7.5.3.3 自相关和偏相关分析法

目标：根据残差自相关和偏相关图表结果，得到你的初步结论

思路：分析残差 e_t 序列对 $e_{t-1}, e_{t-2}, \cdots$ 序列的自相关和偏相关 Eviews 报告（**注意滞后阶数的选择**）

Eviews 操作（菜单操作实现，具体见图 7.11）：

（1）打开残差序列：双击序列（Series）对象et

（2）进入引导菜单：⇒ View ⇒ Correlogram ⇒ Correlogram Specification

- 设置诊断序列（Correlogram of）：点击选择 Level
- 设置诊断阶数（Lags to include）：**默认设置**

（3）完成设置：点击 Ok

（4）命名并保存表格（table）对象

- 另存为表格（table）对象：点击 Freeze
- 命名并保存表格（talbe）对象：点击 name(建议为 tab_corrl_test)

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 查看结果：双击tab_corrl_test

具体 Eviews 报告见图 7.12：

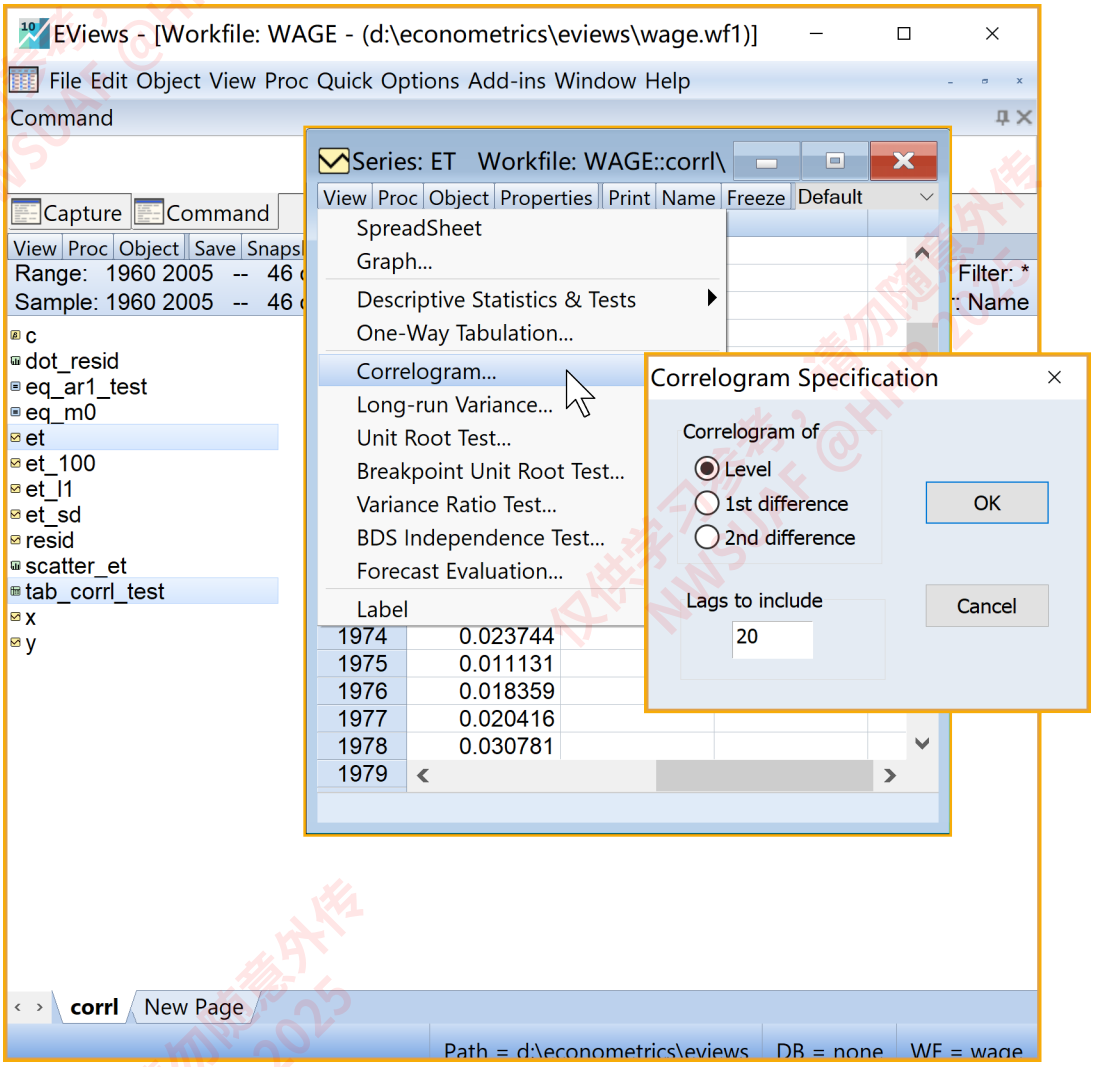


图 7.11: 残差自相关和偏相关分析的 Eviews 操作

7.5.3.4 Durbin-Watson 检验法

目标：观察主回归方程分析报告的 D-W 统计量，根据诊断标准，得出初步结论

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

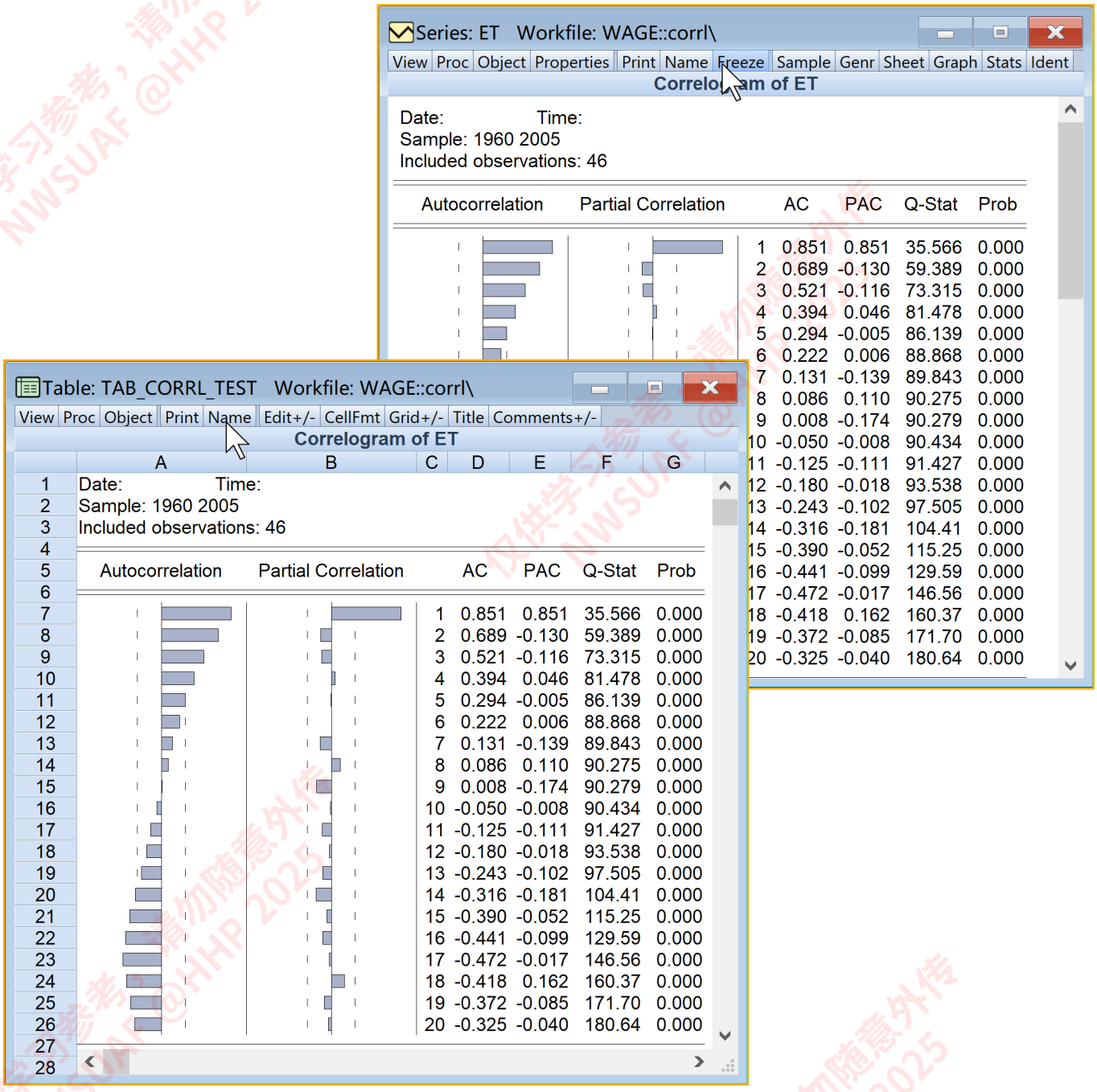


图 7.12: 残差自相关和偏相关的 Eviews 报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

思路：得到 Durbin-Watson 的 d 统计量，并查表得到理论 $\chi^2(n,k,\alpha)$ 分布标准值 d_L 和 d_U，进行比较

理论提示：

- Durbin-Watson 统计量服从 $\chi^2(n,k,\alpha)$分布。具体可以参看 Eviews 在线帮助文档网址
- d_L 和 d_U 的理论值使用 bootstrap 方法仿真计算得到，与 (n,k,α) 有关
- d_L 和 d_U 的理论值可以查表。理论查表值可以参考在线文档

判定准则：

- 如果 $0<d<d_L$，则表明主模型式 7.10 可能存在的一阶正自相关问题
- 如果 $4-d_L<d<4$，则表明主模型式 7.10 可能存在的一阶负自相关问题

分析结论：根据主回归报告（见图 7.4），Durbin-Watson 的 d 统计量为 $d=0.22$。查表可知在当 $(n=' rn', k=3, \alpha=0.05)$ 时，$d_L=1.475, d_U=1.566$，表明 $0<d<d_L$，因此认为主模型式 7.10 可能存在的一阶正自相关问题。

7.5.3.5 拉格朗日检验法（LM-test）

目标：对主回归方程进行布罗施-戈弗雷 (Breusch-Goldfrey) 的拉格朗日检验 (B-G LM test)，根据诊断标准，得出初步结论

思路：利用 Eviews 菜单操作 Residual Diagnostics ⇒ Serial Correlation LM Test

理论提示：

- 布罗施-戈弗雷 (Breusch-Goldfrey) 的拉格朗日检验 (B-G LM test) 可以检验高阶自相关的情形（见模型式 7.11 ）
- 布罗施-戈弗雷 (Breusch-Goldfrey) 的拉格朗日检验 (B-G LM test) 将计算得到一个服从 χ^2 分布的 LM 统计量，具体为 $LM=(n-p)R^2$ χ^2_p

$$u_t=\rho_1 e_{t-1}+\rho_2 e_{t-2}+\cdots+\rho_p e_{t-p}+\varepsilon_t \tag{7.11}$$

诊断标准：

- 如果 LM 检验不显著，也即 $LM < \chi^2(p,\alpha)$，则表明主模型式 7.10 不存在的 p 阶自相关问题
- 如果 LM 检验显著，也即 $LM > \chi^2(p,\alpha)$，则表明主模型式 7.10 存在的 p 阶自相关问题

Eviews 的菜单操作实现，具体见图 7.13)：

（1）打开主方程：双击方程（equation）对象eq_m0

（2）进入引导菜单：⇒ View ⇒ Residual Diagnostics ⇒ Serial Correlation LM Test

（3）设置检验阶数（Lag Specification）: 输入滞后阶数 1（**酌情尝试**）

（4）完成设置：点击 Ok

（5）命名并保存表格（table）对象

- 另存为表格（table）对象：点击 Freeze
- 命名并保存表格（table）对象：点击 name(建议为 tab_LM_test)
- 查看结果：双击tab_LM_test

具体 Eviews 报告见图 7.14：

7.5.4 自相关问题模型的矫正

自相关问题模型的矫正主要包括三类方法，分别是广义最小二乘法 (GLS)、可行广义最小二乘法 (FGLS)、一致性标准误校正法（HAC）。其中，广义最小二乘法 (GLS) 又包括了广义差分法变换、基于残差辅助方程近似估计法、以及基于 D-W 统计量近似计算法。下面我们将依次进行操作说明。

7.5.4.1 广义差分法变换（自相关系数已知）

目标：对主回归方程进行合适的广义差分变换，使得变换以后的新模型不再有自相关问题

思路：如果主模型随机干扰项的自相关系数已知（见模型式 7.13 ），则可以直接用差分变换得到新模型（见模型式 7.15 ），容易证明新模型将不再有自相关问题（见模型式 7.16 ）。

理论提示：如下将展示一阶自相情形下的广义差分变换的理论过程

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t \quad \text{(PRM)} \tag{7.12}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

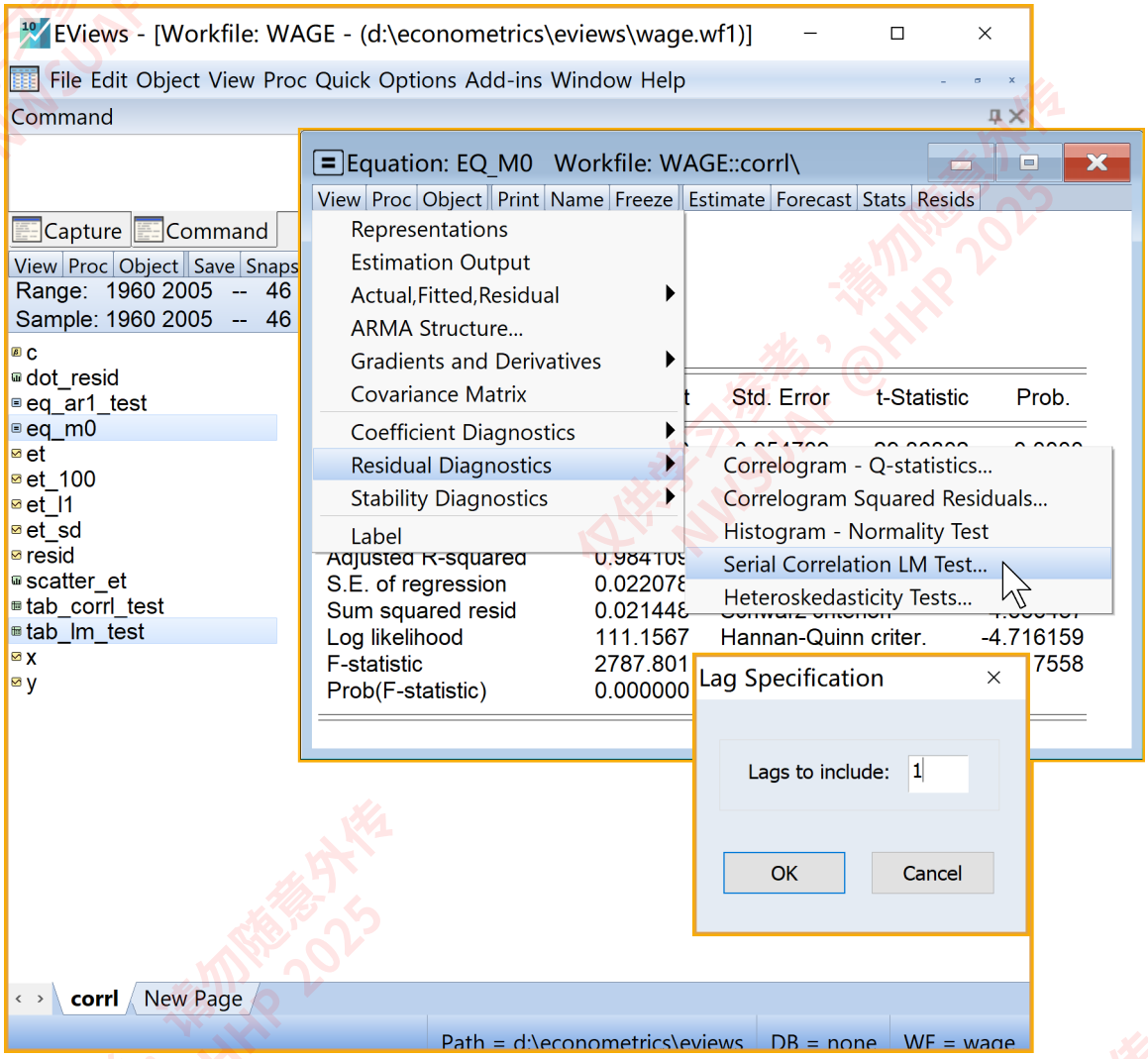


图 7.13: 拉格朗日自相关检验的 Eviews 操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

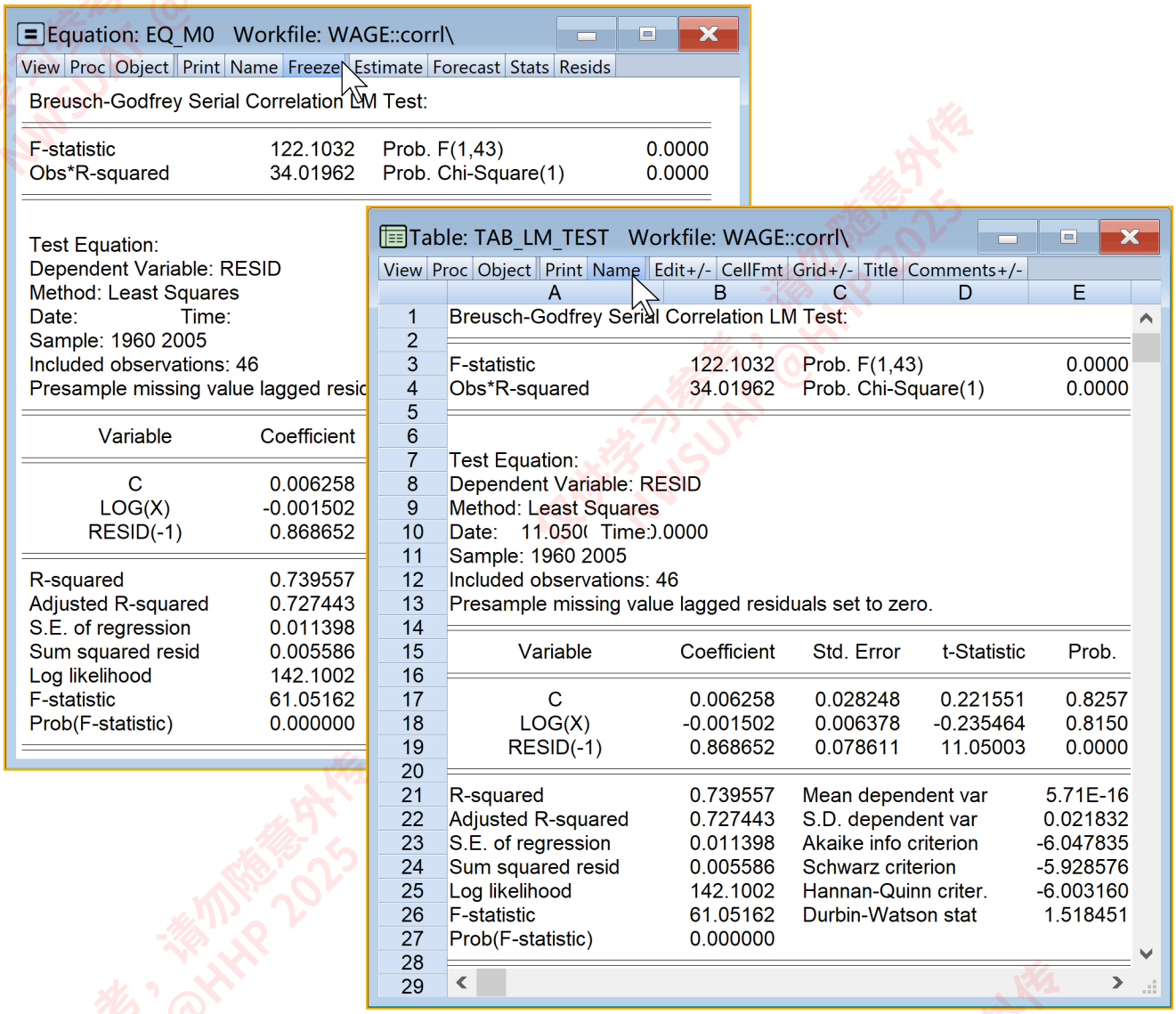


图 7.14: 拉格朗日自相关检验的 Eviews 报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$u_t = \rho u_{t-1} + \varepsilon_t \quad \text{(AR1)} \tag{7.13}$$

$$\rho Y_{t-1} = \rho\beta_1 + \beta_2 \rho X_{2t-1} + \rho u_{t-1} \quad \text{(Lag 1 Model)} \tag{7.14}$$

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_{2t} - \rho X_{2t-1}) + (u_t - \rho u_{t-1}) \quad \Delta\text{(Model)} \tag{7.15}$$

$$Y_t^* = \beta_1^* + \beta_2^* X_{2t}^* + \varepsilon_t \quad \text{(Adjusted Model)} \tag{7.16}$$

其中，AR(1) 模型式 7.13 中的 $\varepsilon_t \sim i.i.d \;\; N(0, \sigma_\varepsilon^2)$。

7.5.4.2 基于残差辅助方程近似估计得到 ρ（自相关系数未知）

目标：对主回归方程进行合适的广义差分变换，使得变换以后的新模型不再有自相关问题

思路：如果主模型随机干扰项的自相关系数未知（见模型式 7.18 ），则可以直接用基于残差辅助方程估计得到 $\hat{\rho}$，再根据 $\rho \simeq \hat{\rho}$ 用广义差分变换得到新模型（见模型式 7.22 ），容易证明新模型将不再有自相关问题（见模型式 7.23 ）。

理论提示：如下将展示一阶自相关情形下的广义差分变换的理论过程

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t \quad \text{(PRM)} \tag{7.17}$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad \text{(AR1)} \tag{7.18}$$

$$Y_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + e_t \quad \text{SRM} \tag{7.19}$$

$$e_t = \hat{\rho} e_{t-1} + v_t \quad \text{(Auxiliary Model)} \tag{7.20}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$\begin{aligned} \rho &\simeq \hat{\rho} \\ \rho Y_{t-1} &= \rho\beta_1 + \beta_2\rho X_{2t-1} + \rho u_{t-1} \quad \text{(Lag 1 Model)} \end{aligned} \tag{7.21}$$

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_{2t} - \rho X_{2t-1}) + (u_t - \rho u_{t-1}) \quad \Delta 1 \text{ (Model)} \tag{7.22}$$

$$Y_t^* = \beta_1^* + \beta_2^* X_{2t}^* + \varepsilon_t \quad \text{(Adjusted Model)} \tag{7.23}$$

其中，AR(1) 模型式 7.18 中的 $\varepsilon_t \sim i.i.d \;\; N(0, \sigma_\varepsilon^2)$。

Eviews 菜单操作：

（1）提取辅助回归方程的回归系数。构建一个标量（Scalar）对象rho_ar1，并将辅助回归方程的回归系数赋值给这个标量对象 rho_ar1。具体过程参看图 7.15

- 命令视窗（Command）输入命令:`scalar rho_ar1=@round(10000*eq_ar1_test.@coefs(1))/10000`
- 运行命令：命令行中按 Enter 键
- 查看结果：双击rho_ar1

（2）进行一阶广义差分变换，并估计新模型。依次选择 ⇒Quick⇒Estimation Equation⇒specification。具体过程参看图 7.16

- Equation specification:输入命令 `(log(y)-rho_ar1*log(y(-1))) c (log(x)-rho_ar1*log(x(-1)))`
- Estimation settings：具体设置如下

> 提示
>
> 具体设置：
>
> - Method: 下拉选择 `LS - Least Squares (NLS and ARMA)`
> - Sample: **默认设置**

- 点击 `OK`

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 模型命名：建议为 `eq_adj_ar1`
- 查看结果：双击eq_adj_ar1

（3）说明（Eviews 代码行的解读。

- 具体细节请参看 Eviews 在线帮助文档。
- 代码 `scalar rho_ar1=@round(10000*eq_ar1_test.@coefs(1))/10000` 表示给创建一个标量（Scalar）对象rho_ar1, 并把辅助模型式 7.9 的回归系数的第一个值提取出来，并赋值给这个名为 rho_ar1 的标量对象，同时保留四位小数。
- 代码 `eq_ar1_test.@coefs(1)` 表示提取方程（equation）对象eq_ar1_test 的回归系数的第一个值。
- 代码 `@round(10000*( ))/10000` 表示对数据保留 4 为小数。

7.5.4.3 基于 D-W 统计量近似计算得到 ρ（自相关系数未知）

目标：对主回归方程进行合适的广义差分变换，使得变换以后的新模型不再有自相关问题

思路：如果主模型随机干扰项的自相关系数未知（见模型式 7.25 ），则可以基于 Durbin-Waston 检验的 d 统计量计算得到 $\hat{\rho}$（见模型式 7.26 和模型式 7.27 ），再根据 $\rho \simeq \hat{\rho}$ 用广义差分变换得到新模型（见模型式 7.28 ），容易证明新模型将不再有自相关问题（见模型式 7.29 ）。

理论提示：如下将展示一阶自相情形下的广义差分变换的理论过程

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t \quad \text{(PRM)} \tag{7.24}$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad \text{(AR1)} \tag{7.25}$$

$$d \simeq 2(1-\hat{\rho}) \quad \text{(Durbin-Waston)} \tag{7.26}$$

$$\hat{\rho} \simeq 1 - d/2 \tag{7.27}$$

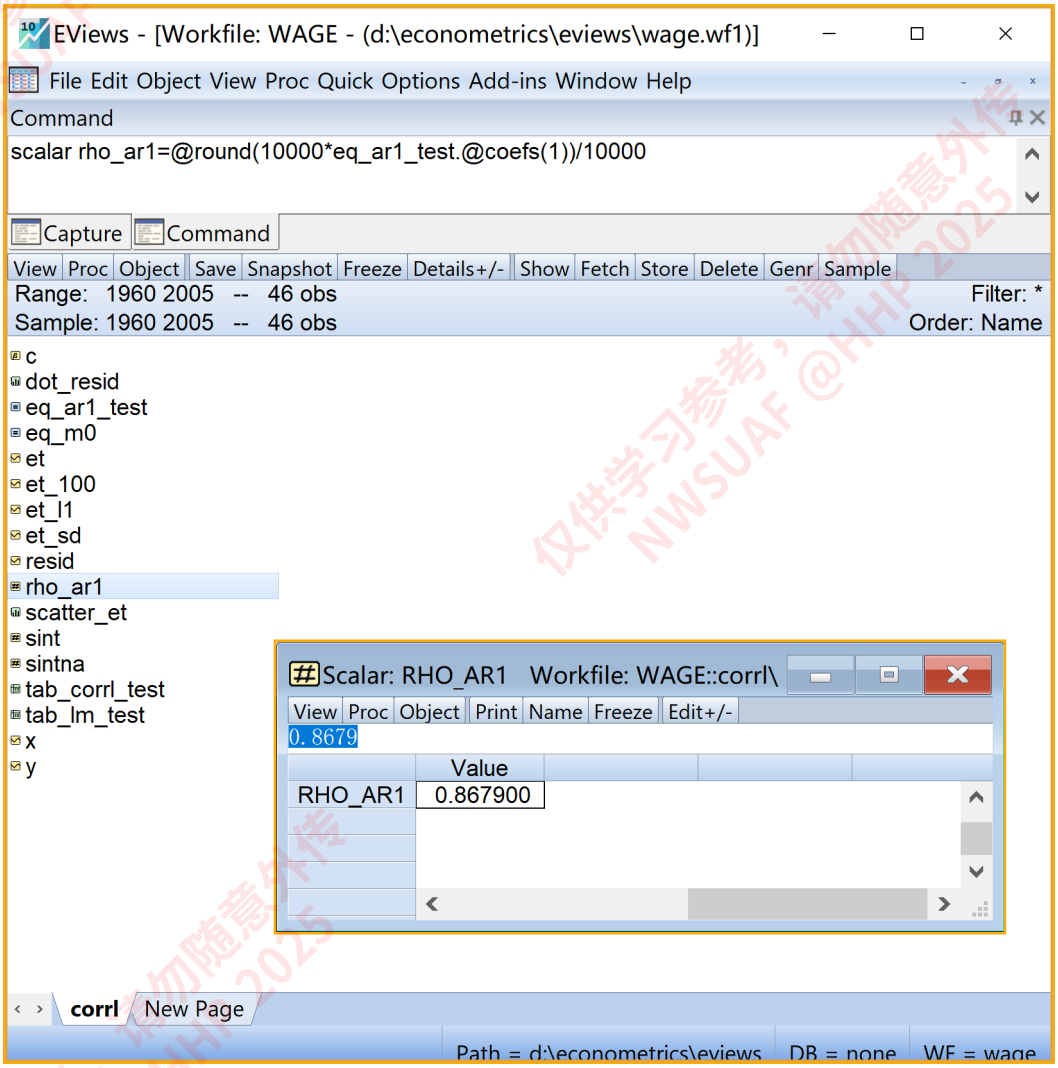


图 7.15: 基于残差辅助方程近似计算自相关系数

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

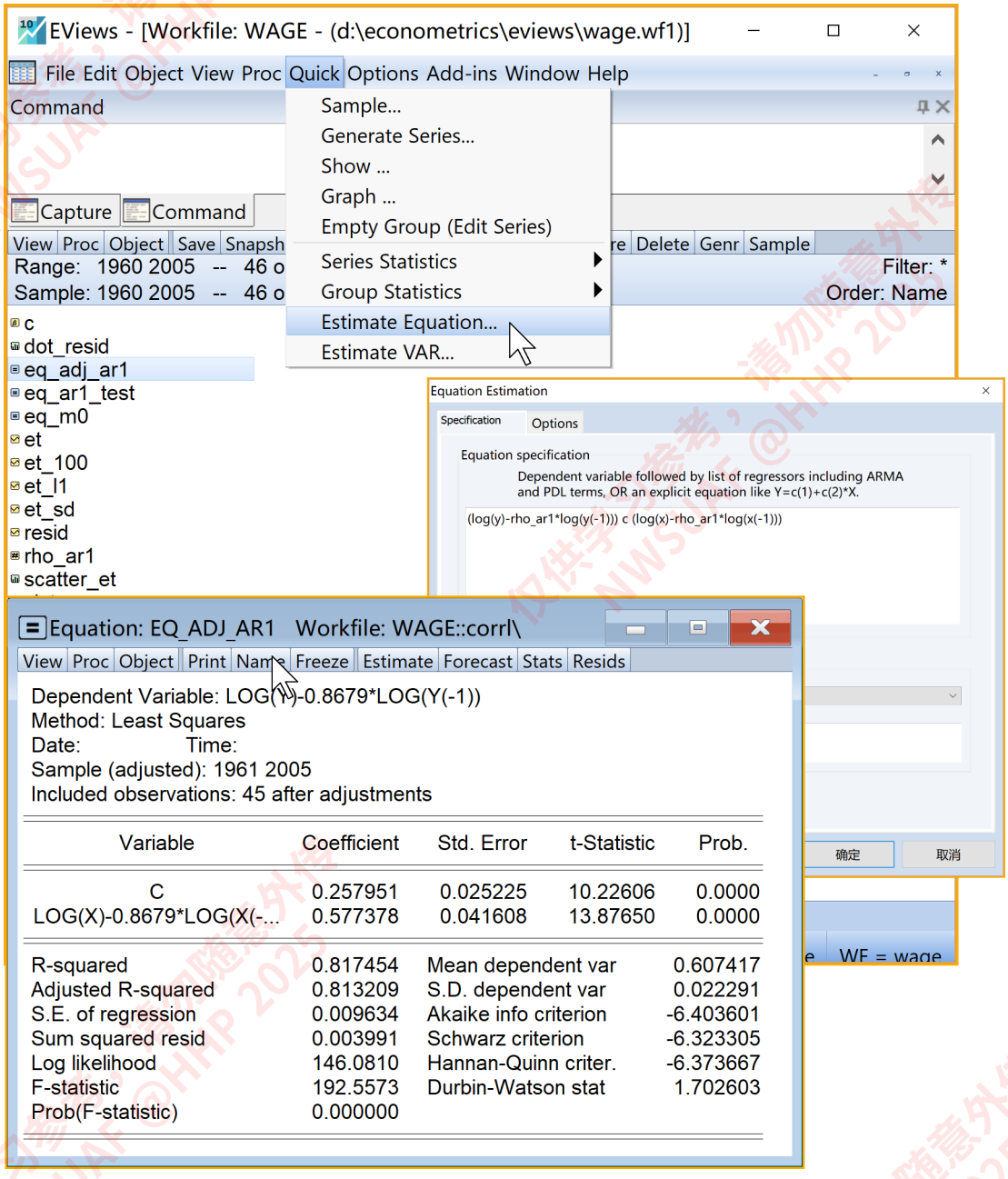


图 7.16: 基于残差辅助方程的广义差分矫正操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

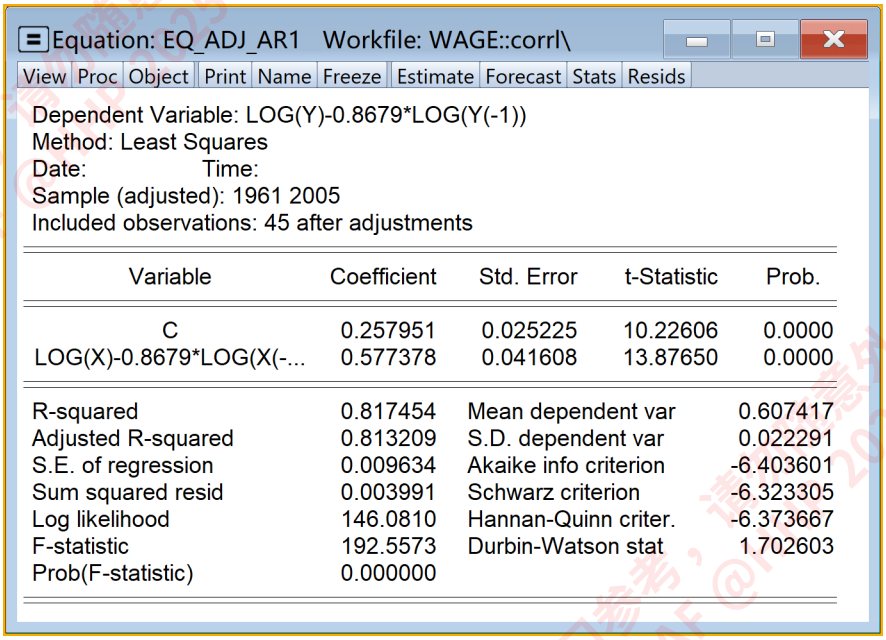


图 7.17: 基于残差辅助方程的广义差分矫正报告

$$\rho \simeq \hat{\rho}$$

$$\rho Y_{t-1} = rho\beta_1 + \beta_2 \rho X_{2t-1} + \rho u_{t-1} \quad \text{(Lag 1 Model)}$$

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_{2t} - \rho X_{2t-1}) + (u_t - \rho u_{t-1}) \quad \Delta 1 \text{ (Model)} \tag{7.28}$$

$$Y_t^* = \beta_1^* + \beta_2^* X_{2t}^* + \varepsilon_t \quad \text{(Adjusted Model)} \tag{7.29}$$

其中，AR(1) 模型式 7.25 中的 $\varepsilon_t \sim i.i.d \;\; N(0, \sigma_\varepsilon^2)$。

Eviews 菜单操作：

（1）提取并得到 Durbin-Waston 统计量。构建一个标量（Scalar）对象，建议命名为 dw。具体过程参看图 7.18

- 命令视窗（Command）输入命令：`scalar dw=eq_m0.@dw`
- 运行命令：命令行中按 Enter 键

- 查看结果：双击rho_dw

（2）计算得到自相关系数的近似值 $\rho \simeq \hat{\rho} = 1 - d/2$。构建一个标量（Scalar）对象，建议命名为 rho_dw。具体过程参看图 7.18 。

- 命令视窗（Command）输入命令：`scalar rho_dw=@round(10000*(1-dw/2))/10000`
- 运行命令：命令行中按 Enter 键
- 查看结果：双击rho_dw

（3）进行一阶广义差分变换，并估计新模型。依次选择 ⇒Quick⇒Estimation Equation⇒specification。具体过程参看图 7.19

- Equation specification:输入命令 `(log(y)-rho_dw*log(y(-1))) c (log(x)-rho_dw*log(x(-1)))`
- Estimation settings：设置如下

> 提示
>
> 具体设置：
>
> - Method: 下拉选择 `LS - Least Squares (NLS and ARMA)`
> - Sample: **默认设置**

- 点击 `OK`
- 模型命名：建议为 `eq_adj_dw`
- 查看结果：双击eq_adj_dw

（4）Eviews 代码行的解读：

- 代码 `scalar d=eq_m0.@dw` 表示给创建一个标量（Scalar）对象dw, 并把主模型式 7.10 的 Durbin-Waston 统计量值提取出来，并赋值给这个名为 dw 的标量对象。
- 代码 `scalar rho_dw=@round(10000*(1- dw/ 2))/ 10000` 计算得到标量（Scalar）对象rho_dw，并对数据取 4 位小数。

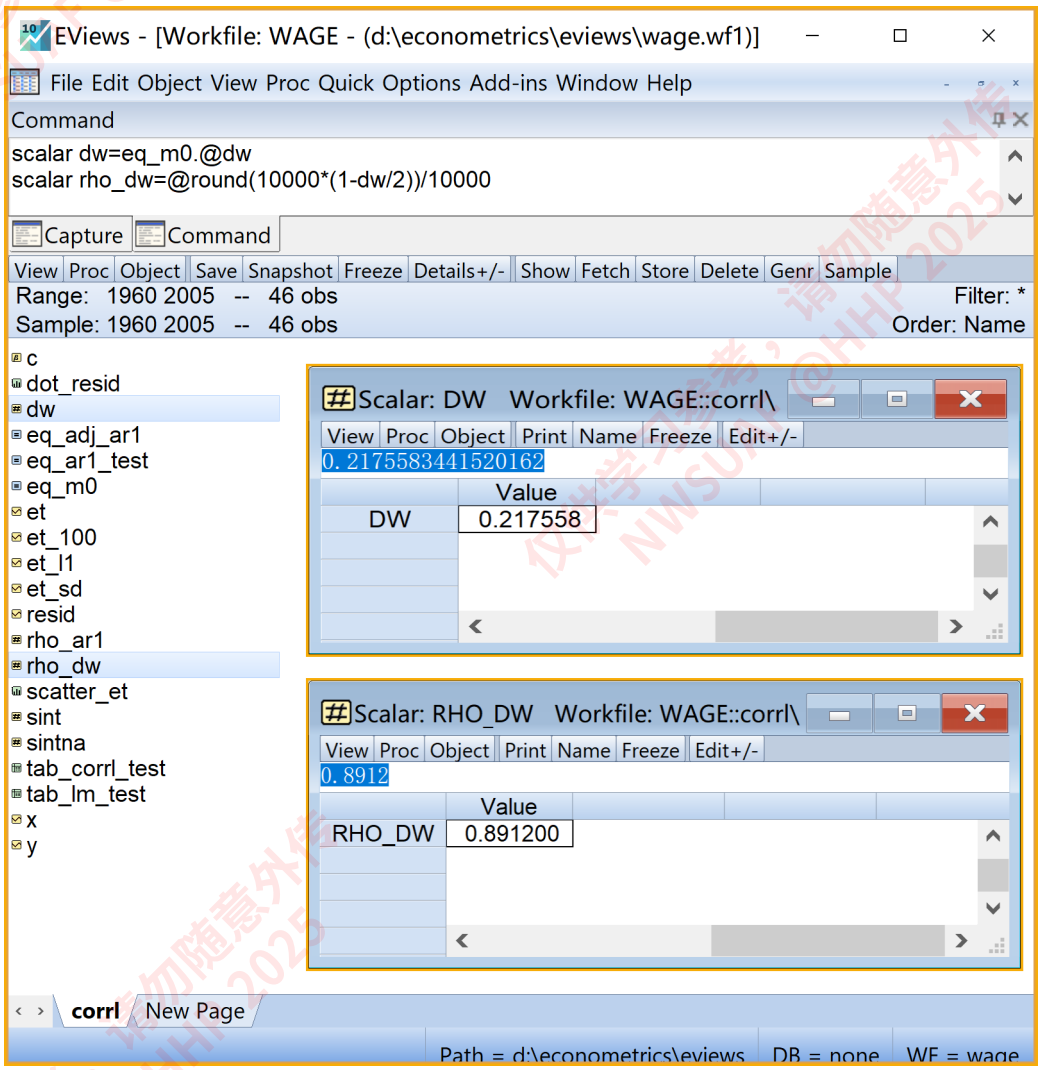


图 7.18: 基于 Durbin-Waston 统计量近似计算自相关系数

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

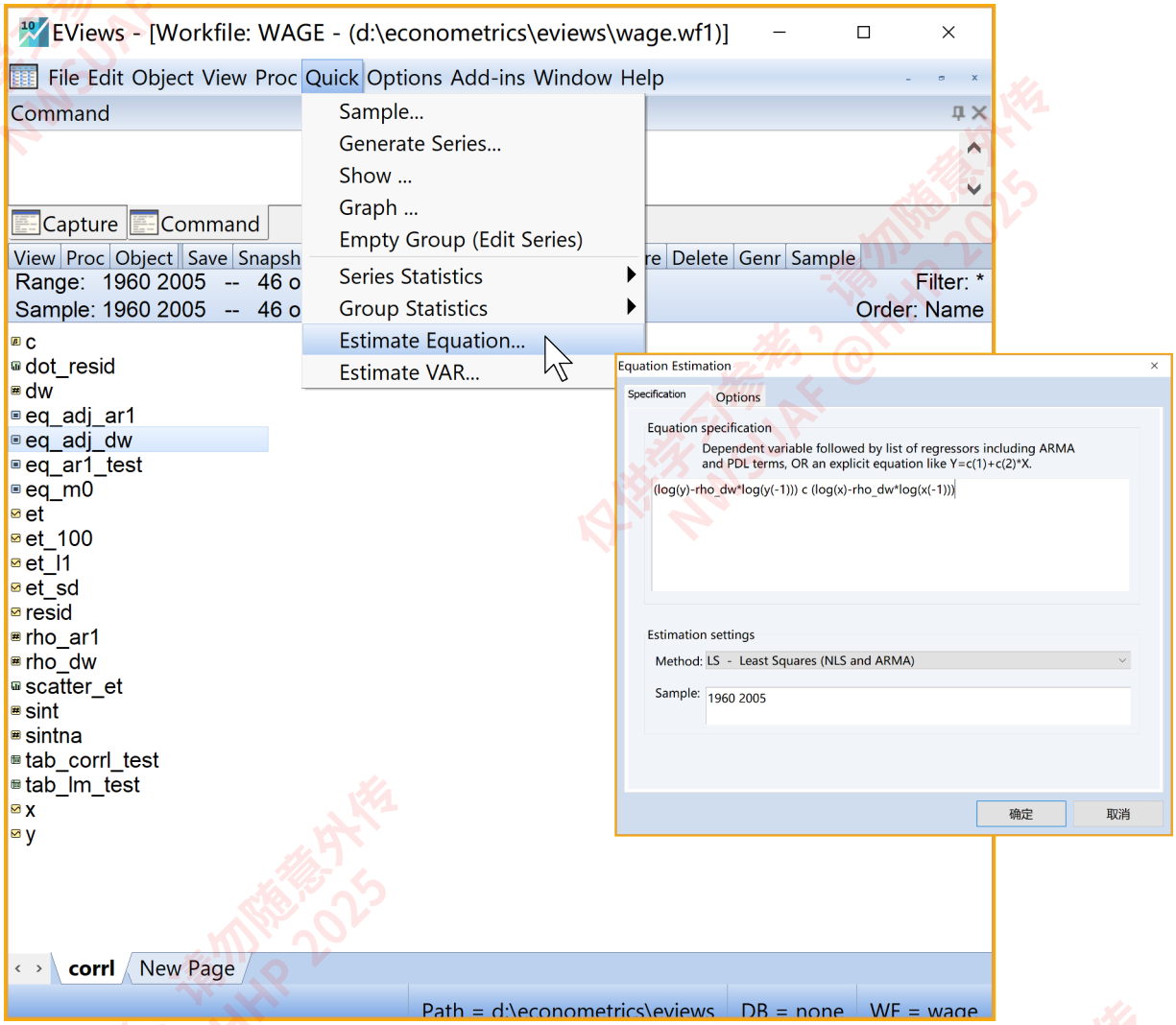


图 7.19: 基于 Durbin-Waston 统计量的广义差分模型矫正操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

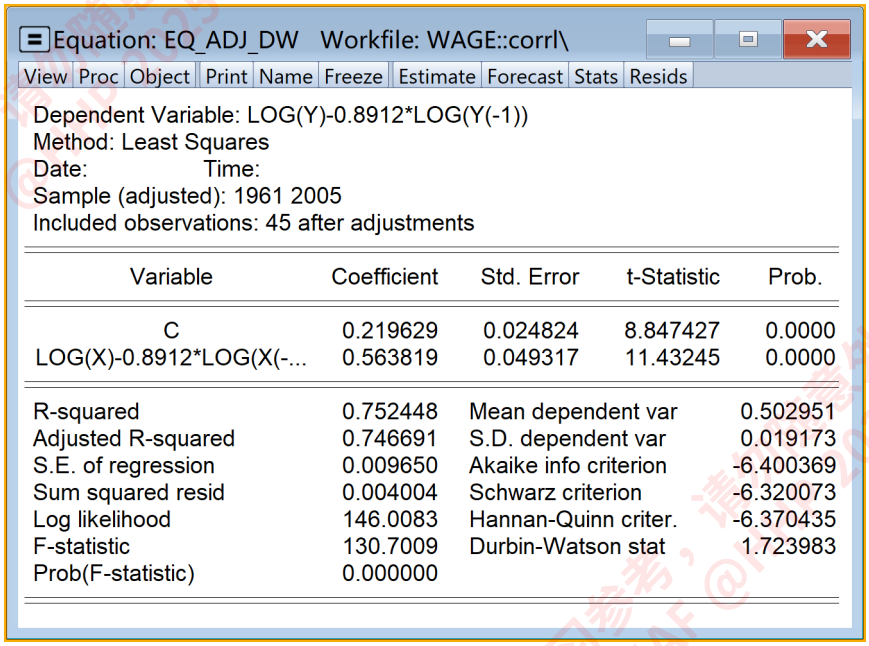


图 7.20: 基于 Durbin-Waston 统计量的广义差分模型矫正报告

7.5.4.4 基迭代法近似计算得到 ρ（自相关系数未知）

目标：基迭代法近似计算的方法被称为可行广义最小二乘法 (FGLS)。对主回归方程进行合适的广义差分变换，使得变换以后的新模型不再有自相关问题

思路：

- 如果主模型随机干扰项的自相关系数未知，而且存在高阶自相关情形，则可以使用基于迭代的可行广义最小二乘法 (FGLS) 计算得到 $\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_p \quad p \in (1, 2, \cdots)$，再根据 $\rho \simeq \hat{\rho}$ 用广义差分变换得到新模型（见模型式 7.28 ），容易证明新模型将不再有自相关问题（见模型式 7.29 ）。
- 这些迭代方法主要包括**科克伦-奥克特迭代法** (Cochrane-Orcutt iterative procedure)；科克伦-奥克特两步法 (Cochrane-Orcutt two-step procedure)；德宾两步法 (Durbin two-step procedure)；希尔德雷思-卢扫描或搜寻程序 (Hildreth-Lu scanning or search procedure) 等

理论提示：如下将展示**科克伦-奥克特迭代法**下对二阶自相关（$AR(p), p = 2$）情形下的广义差分变换的理论过程

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t \quad \text{(PRM)} \tag{7.30}$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t \quad \text{(AR2)} \tag{7.31}$$

…Cochrane-Orcutt iterative procedure

$$\rho_p \simeq \hat{\rho}_p \tag{7.32}$$

$$\rho_1 Y_{t-1} = \rho_1 \beta_1 + \beta_2 \rho_1 X_{2t-1} + \rho_1 u_{t-1} \qquad \text{Lag 1 Model}$$

$$\rho_2 Y_{t-2} = \rho_2 \beta_1 + \beta_2 \rho_2 X_{2t-2} + \rho_2 u_{t-2} \quad \text{Lag 2 Model} \tag{7.33}$$

$$\begin{aligned}(Y_t - \rho_1 Y_{t-1} - \rho_2 Y_{t-2}) &= \beta_1(1 - \rho_1 - \rho_2) \\ &+ \beta_2(X_{2t} - \rho_1 X_{2t-1} - \rho_2 X_{2t-2}) \quad (\Delta 2 \quad \text{Model}) \\ &+ (u_t - \rho_1 u_{t-1} - \rho_2 u_{t-2})\end{aligned} \tag{7.34}$$

$$Y_t^* = \beta_1^* + \beta_2^* X_{2t}^* + \varepsilon_t \quad \text{(Adjusted Model)} \tag{7.35}$$

Eviews 菜单操作（此处展示的是 AR(1) 的自相关情形，见图 7.21）：

（1）依次选择 ⇒Quick⇒Estimation Equation

（2）引导设置 Equation Estimation⇒specification

a. Equation specification：输入命令 log(Y) c log(X) AR(1)

- Estimation settings：具体设置如下

`Method: 下拉选择`LS - Least Squares (NLS and ARMA)`

`Sample: **默认设置**`

- 点击 `OK`

（3）模型命名：建议为 `eq_adj_CO`

- 主回归分析结果见图 7.22：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

（4）提取并得到**科克伦-奥克特迭代法**下近似的残差自相关系数 ρ。构建一个标量（Scalar）对象，建议命名为 rho_co。具体过程参看图 7.23

- 命令视窗（Command）输入命令：`scalar rho_co=eq_adj_co.@coefs(3)`
- 运行命令：命令行中按 Enter 键
- 查看结果：双击rho_co

（5）Eviews 代码行的解读：

- 代码 `scalar rho_co=eq_adj_co.@coefs(3)` 表示给创建一个标量（Scalar）对象rho_co，并把方程（Equation）对象eq_adj_co 的第三个回归系数提取出来，赋值给这个标量对象rho_co。

7.5.4.5 尼威-威斯特 (Newey-West) 一致性标准误矫正法

目标：直接用尼威-威斯特 (Newey-West) 一致性标准误矫正流程方法，构建回归分析模型，此时模型的自相关问题将会有所缓解。

思路：利用 Eviews 菜单操作，进行基于尼威-威斯特 (Newey-West) 一致性标准误矫正程序的建模分析。

理论提示：（数学表达和证明过程略）

- 异方差-自相关一致性标准误（heteroscedasticity-autocorralation consistent standard errors，HAC）也被简称为尼威-威斯特一致性标准误 (Newey-West consistent standard errors)
- 尼威-威斯特 (Newey-West) 一致性标准误矫正程序或菜单，在主流的统计软件里都会配置
- 尼威-威斯特 (Newey-West) 一致性标准误矫正程序，严格意义上对于大样本数据是有效的，因此不太适合于小样本数据的情形。

Eviews 操作（见图 7.24）

（1）依次选择 ⇒ Quick ⇒ Estimation Equation

（2）引导设置 Equation Estimation ⇒ Specification
- 方程设置（Equation Specification）：输入变量 log(Y) c log(X)

- 估计方法（Estimation settings）：具体如下

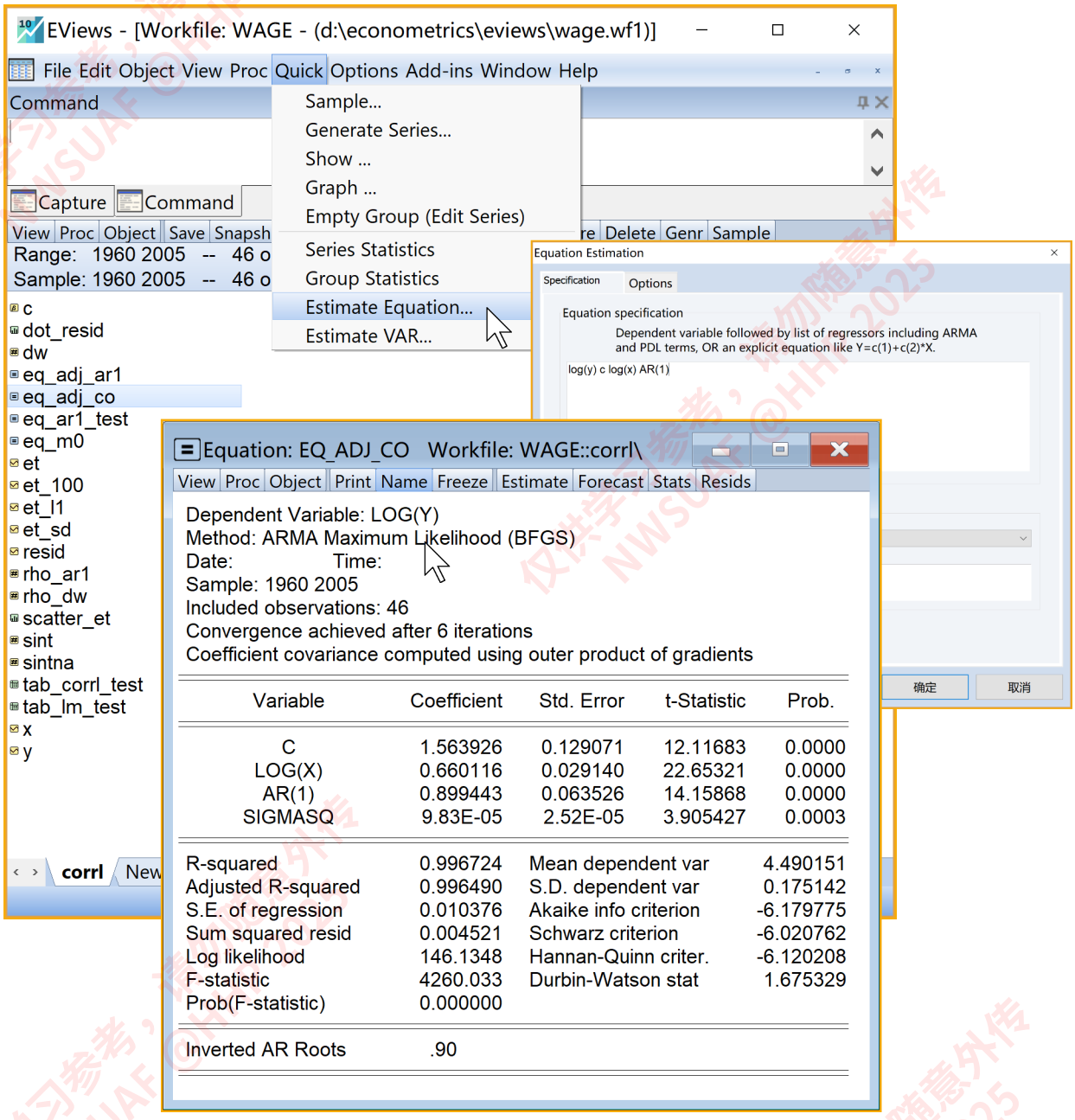


图 7.21: 基于科克伦-奥克特迭代法的 FGLS 模型矫正 Eviews 操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

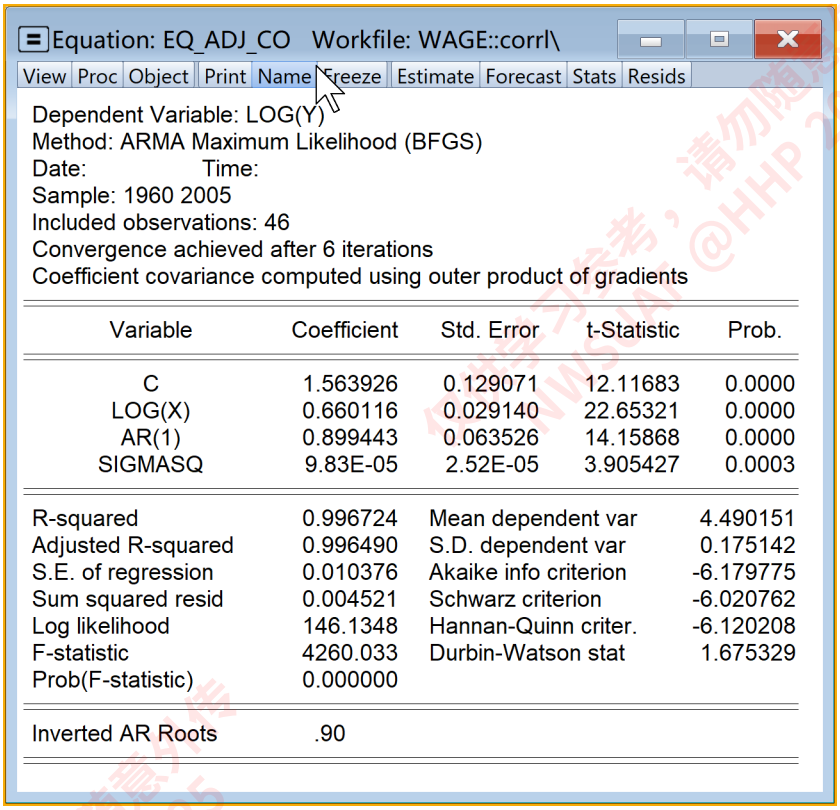


图 7.22: 基于科克伦-奥克特迭代法的 FGLS 模型矫正 Eviews 报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

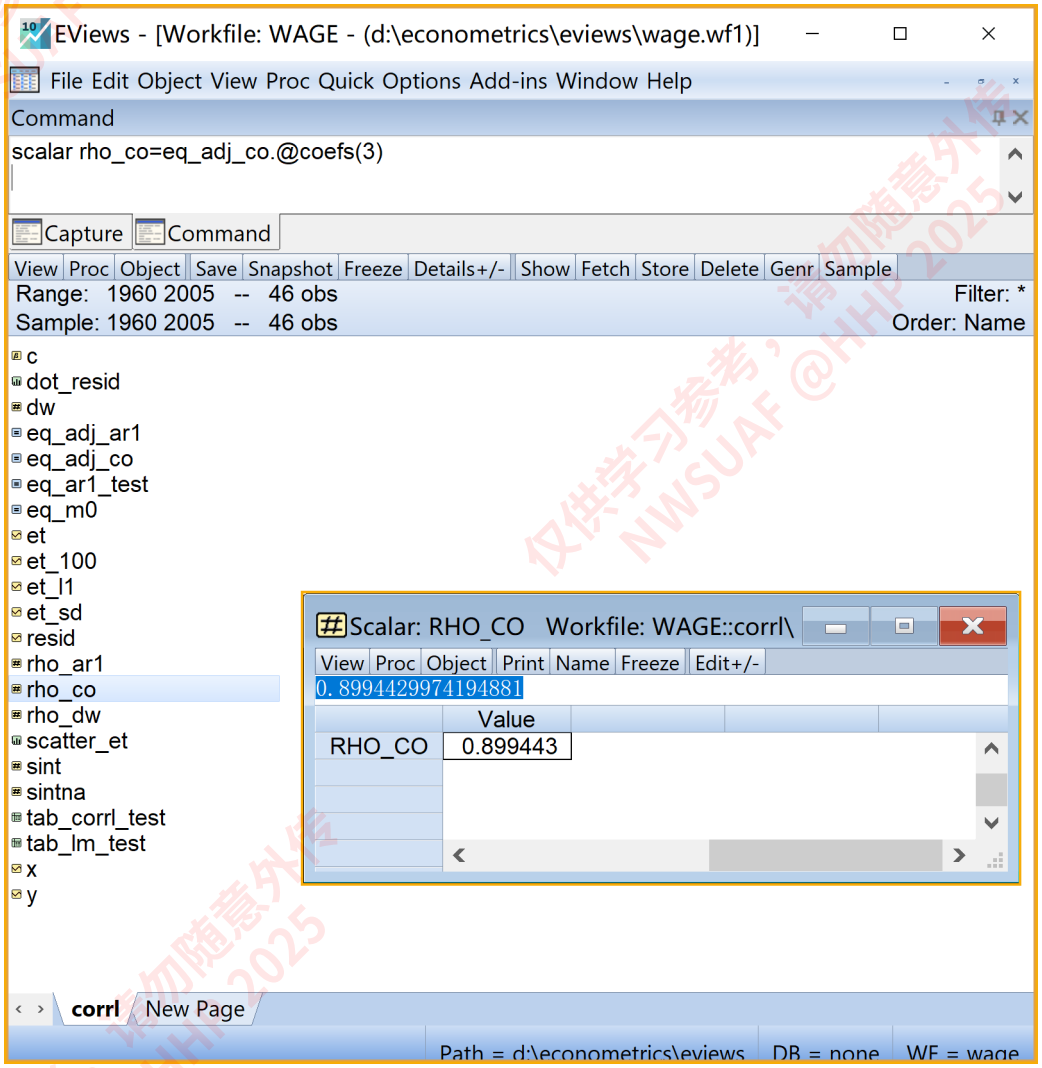


图 7.23: 基于科克伦-奥克特迭代法近似计算的自相关系数

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

* Method: 选择`LS - Least Squares(NLS and ARMA)`
* Sample: **默认设置**

（3）引导设置 Equation Estimation ⇒ Options

- 系数协方差设置（Coefficient covariance）：具体如下

* 协方差方法（Coefficient method）：下拉选择HAC (Newey-West)

- 权重设置（Weights）：默认设置
- 最优化设置（Optimization）：默认设置
- 完成设置：点击 OK

（4）模型命名：建议为 eq_adj_NW

（5）查看分析报告（见图 7.25）

7.6 作业题：存货案例

存货案例：表 7.3 给出给出了 41 年的 Y 存货（百万美元），X2 销售额（百万美元）等方面数据。

表 7.3: 存货与销售额 (n=41)

Year	Y	X2
1950	84646	46486
1951	90560	50229
1952	98145	53501
1953	101599	52805
1954	102567	55906
1986	423082	326227
1987	408226	334616
1988	439821	359081
1989	479106	394615
1990	509902	411663

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

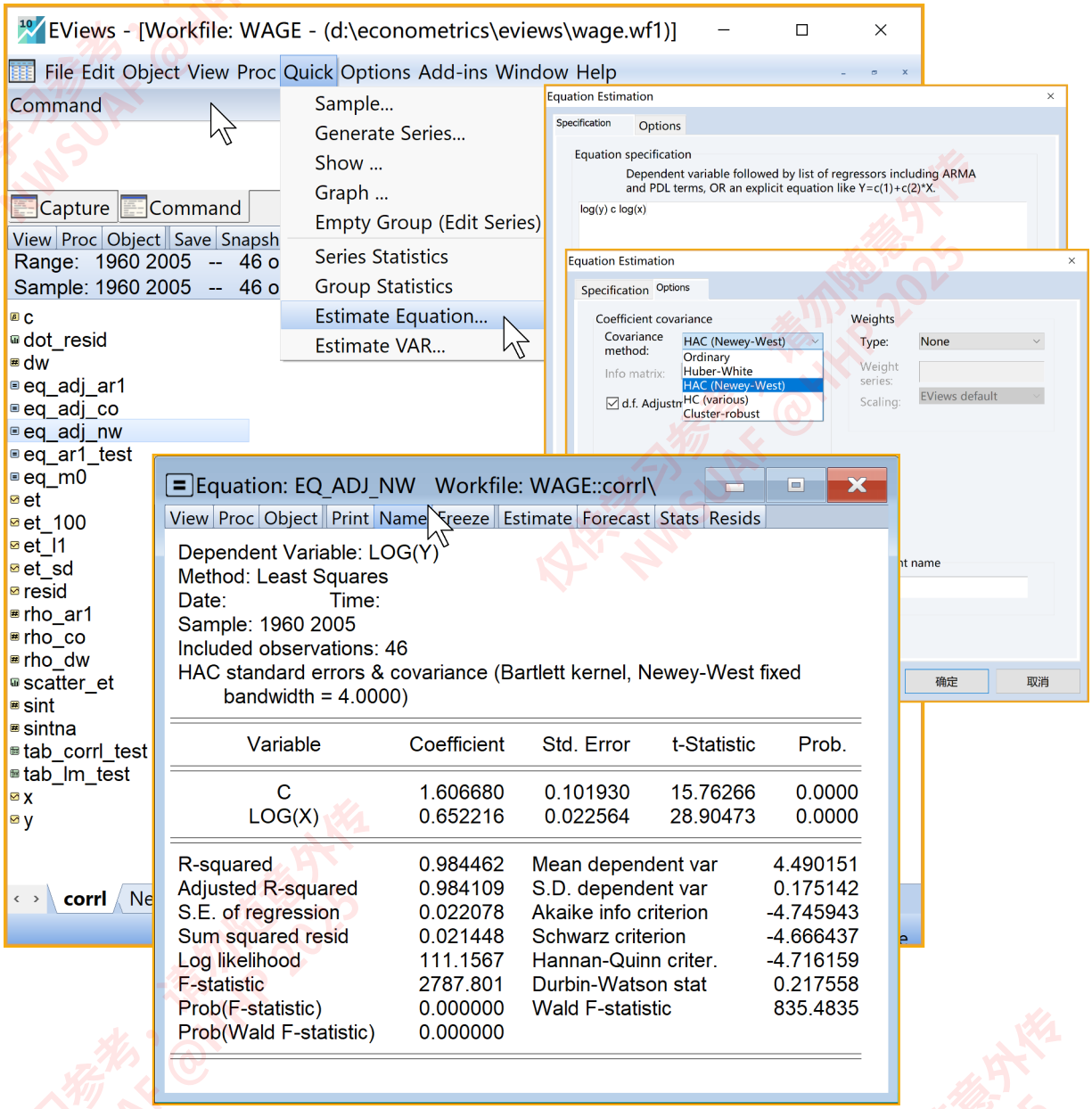


图 7.24: 尼威-威斯特 (Newey-West) 矫正法的操作

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

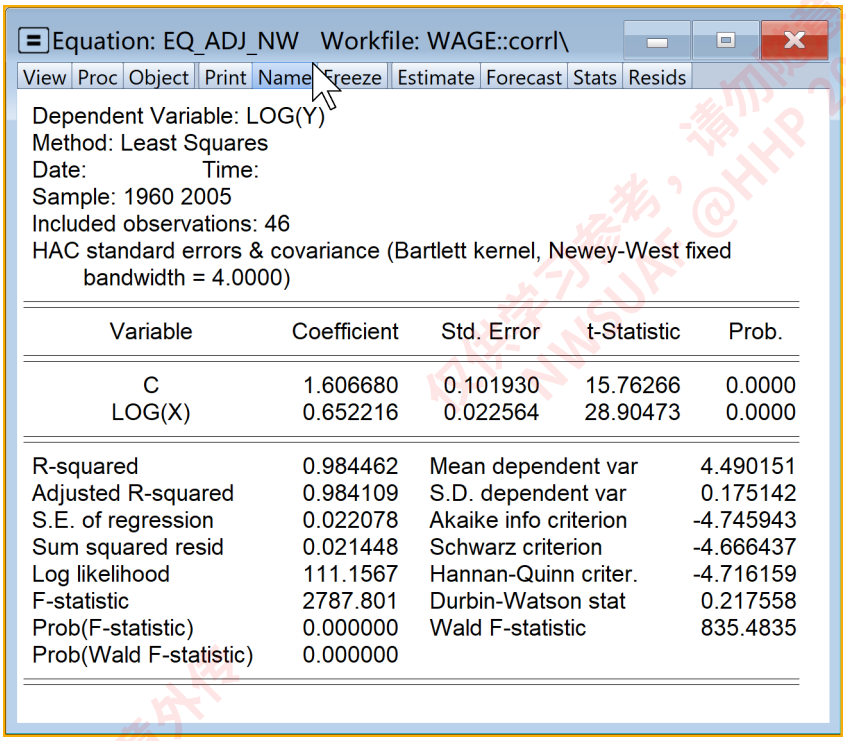


图 7.25: 尼威-威斯特 (Newey-West) 矫正法的 Eviews 报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

变量说明见表 7.4：

表 7.4: 变量定义及说明

variable	label
Year	年份
Y	存货（百万美元）
X2	销售额（百万美元）

请考虑如下样本回归模型：

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t \quad \text{(classic)} \tag{7.36}$$

$$ln(Y_t) = \beta_1 + \beta_2 ln(X_{2t}) + u_t \quad \text{(log)} \tag{7.37}$$

请回答如下问题:

（1）根据总体回归模型式 7.36 ，请对模型参数的理论预期（符号、大小、关系）进行说明。

（2）利用 Eviews 对模型式 7.36 进行回归分析（将报告截图过来，并写出相应的简要报告形式——三行式或四行式）。参数估计结果符合你的理论预期么？

（3）回归模型式 7.36 存在自相关问题的证据吗？请按照下列方法分别进行诊断，并分别得到分析结论（要求截图过来并进行简要说明）：

$$e_t = \hat{\rho}_1 e_{t-1} + \hat{\rho}_2 e_{t-2} + \cdots + v_t$$

a. 残差序列观察法（描点图法）：绘制 e_t 序列的描点图（dot plot），得到你的初步结论。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

b. 残差序列观察法（描点图法）：根据前述观察，确定滞后阶数并分别绘制 e_t 序列与 $e_{t-1}, e_{t-2}, \cdots$ 序列的散点图（scatter plot），得到你的初步结论

b. 辅助回归法：构建残差 e_t 序列对 $e_{t-1}, e_{t-2}, \cdots$ 序列的辅助回归方程式 7.9 ，得到你的初步结论

c. 自相关和偏相关分析法：Eviews 菜单操作对残差 e_t 序列进行自相关和偏相关分析（**注意滞后阶数的选择**）

d. Durbin-Watson 检验法：分析 Eviews 报告中的 D-W 统计量

e. 拉格朗日检验法（LM-test）：Eviews 菜单操作进行布罗施-戈弗雷 (Breusch-Goldfrey) 的拉格朗日检验 (B-G LM test)

（4）若发现回归模型式 7.36 存在自相关问题，你如何对该模型进行矫正？请按照下列方法分别进行诊断，并分别得到分析结论（要求截图过来并进行简要说明）：

a. 广义最小二乘法 (GLS)：一阶差分法变换

b. 广义最小二乘法 (GLS)：基于残差辅助方程式 7.9 中估计出来的 $\hat{\rho}$ 近似得到 ρ

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

c. 广义最小二乘法 (GLS)：基于 D-W 统计量近似计算得到 ρ

d. 广义最小二乘法 (GLS)：基迭代法近似计算得到 ρ

e. 一致标准误校正法（HAC）：尼威-威斯特 (Newey-West) 校正法

f. 对数变换法：对数模型式 7.37 进行 Eviews 回归分析

（5）如果模型式 7.36 与上面各种方法修正的结果相差较大，你认为哪一个方法要相对较好，为什么？

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

8 虚拟变量回归

仅供学习参考，请勿随意外传
NWSUAF @HHP 2025

8.1 实验目的及要求

- **目的**：掌握虚拟变量模型的设置和分析方法。
- **要求**：熟悉虚拟变量的设置方法；理解定性变量和定量变量模型的内涵；熟练加法模型和乘法模型的运用原理。

8.2 实验原理

计量经济学建模分析中，我们常常需要把一些**定性变量（Qualitative variables）**（如性别、地区、党派等）作为自变量放入回归模型中。从**变量层次（Variable Scale）**来看，这些变量没有具体的取值，只有特定属性类别。例如，性别变量的具体取值往往为男或女。显然，诸如此类的变量如果直接放到线性回归模型中，将会产生一系列的参数估计、模型解释等问题。

表 8.1: 变量类型和变量尺度

staff	salary	sale	score	race
Zhang	25	140	2	yellow
Tom	35	195	3	white
Lee	27	184	1	yellow
Jeff	42	256	5	white
Kevin	38	207	4	black

区域经理年薪案例中（见表 8.1 ），公司有五名区域经理，分别负责不同的国际市场。定量变量 *salary* 表示区域经理的年薪（万元）；变量 *sale* 表示负责市场的销售额；变量 *score* 表示客户对区域经理的评价（1 表示很不满意，2 表示不满意，3 表示一般，4 表示很满意，5 表示非常满意）；变量 *race* 表示区域市场主要消费群体的肤色（yellow 表示黄色消费群体、white 表示白色消费群体，black 表示黑色消费群体）。

定义 8.1 (变量类型)**.** 定量变量（Quantitative variable）一般也称为连续变量，是由测量或计数、统计所得到的量，可以通过数值表达，并具有直接的数值含义。

定性变量（Qualitative variable）一般也称为分类变量，主要用于区分事物性质差异，往往用语义类别表达，没有直接的数值含义。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

定义 8.2 (变量尺度). 变量尺度（Variable scale）刻画的是变量的数值含义或数值关系。它将意味着在数值含义和关系上，变量是有层次级别的差异性。根据变量层级不同，具体可以分为由低到高的 4 个层级：

名义尺度（nominal scale）变量：这类变量只用于属性分类，不具备任何数值含义或数值关系，也即不能加、减、乘、除，也不能比较大小。**序数尺度（order scale）变量**：这类变量具备很少的数值含义或数值关系，它可以比较大小，但不能进行加、减、乘、除。**区间尺度（interval scale）变量**：这类变量具备一定的数值含义或数值关系，它可以比较大小，也可以进行加、减，但不能进行乘、除。**比率尺度（ratio scale）变量**：这类变量具备最多的数值含义或数值关系，它可以比较大小，也可以进行加、减、乘、除。

根据以上定义，**区域经理年薪案例**中，可以认为年薪 *salary*、销售额 *sale*，以及客户评价 *score* 为定量变量，消费群体主要肤色 *race* 为定性变量。其中，从变量的度量尺度来看：

- 年薪 *salary* 和销售额 *sale* 两个变量为**比率尺度**变量
- 客户评价 *score* 变量为**序数尺度**变量
- 消费群体主要肤色 *race* 为**名义尺度**变量

8.2.1 如何把定性变量转换为虚拟变量?

一个定性变量的不同数据取值，称为该定性变量的属性。定性变量的任一属性，都可以设置为一个**虚拟变量**。实际上，我们可以用一套**虚拟变量体系**来完全表达一个定性变量。然后按照一定的规则构建**虚拟变量回归模型**，从而避免参数估计、模型解释等问题的出现。

定义 8.3 (虚拟变量). 对于某定性变量的任一特定属性，可以构造出一个虚拟变量（记为 D），使得该虚拟变量能够表达这一属性。同时，给该虚拟变量 D 赋值为 1，记为具备这一属性；给该虚拟变量赋值为 0，记为不具备该属性。正式地，假设定性变量 X 具有 m 个属性 $a_1, a_2, \cdots, a_m$，对于任意属性 $k, (k \in 1, 2, \cdots, m)$，可以定义如下的虚拟变量 D_k：

$$D_k = \begin{cases} 1, & \text{if } a_k \\ 0, & \text{if not } a_k \end{cases}$$

定义 8.4 (虚拟变量体系). 完整表达某个定性变量全部信息的一组虚拟变量。正式地，假设定性变量 X 具有 m 个属性 $a_1, a_2, \cdots, a_m$，可以用如下一组虚拟变量 $D_1, \cdots, D_k, \cdots, D_m$ 完全表达该定性变量：

$$X\{a_1,a_2,\cdots,a_m\}\Rightarrow\begin{cases}D_1=\begin{cases}1, & \text{if } a_1\\ 0, & \text{if not } a_1\end{cases}\\ \vdots\\ D_k=\begin{cases}1, & \text{if } a_k\\ 0, & \text{if not } a_k\end{cases}\\ \vdots\\ D_m=\begin{cases}1, & \text{if } a_m\\ 0, & \text{if not } a_m\end{cases}\end{cases}$$

区域经理年薪案例中，定性变量肤色（$race$）具有 3 个属性（$m=3$），具体为 $X\{a_1=yellow, a_2=white, a_3=black\}$，则可以构造出如下的虚拟变量体系[1]。

$$race\{a_1=\text{yellow},a_2=\text{white},a_3=\text{black}\}$$
$$\Longrightarrow\begin{cases}race\_yellow=\begin{cases}1, & \text{yellow}\\ 0, & \text{not yellow}\end{cases}\\ race\_white=\begin{cases}1, & \text{white}\\ 0, & \text{not white}\end{cases}\\ race\_black=\begin{cases}1, & \text{black}\\ 0, & \text{not black}\end{cases}\end{cases}$$

实际数据操作中，一般需要对定性变量 $race$ 进行重新编码（recode），生成三个对应的虚拟变量（见表 8.2 ）。

表 8.2: 把定性变量转变为虚拟变量体系

staff	race	race_black	race_white	race_yellow
Zhang	yellow	0	0	1
Tom	white	0	1	0
Lee	yellow	0	0	1
Jeff	white	0	1	0
Kevin	black	1	0	0

[1]一个定性变量如果有 m 个属性，那么可以用 m 个虚拟变量**完全表达**该定性变量，也可以用 $(m-1)$ 个虚拟变量**充分表达**该定性变量。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

表 8.2: 把定性变量转变为虚拟变量体系

staff	race	race_black	race_white	race_yellow

8.2.2 如何理解虚拟变量回归模型?

区域经理薪水案例中，如果不区分变量类型和特征，建议如下的回归模型（见方程式 8.1 ），则回归分析结果将会带来严重的问题。

$$salary = \alpha_1 + \alpha_2 race + \alpha_3 sale + u \tag{8.1}$$

事实上，应该将上述模型转换为**虚拟变量回归模型**（如方程式 8.2 ）。这类模型往往又被称为**方差分析模型**，其模型本质上是对不同 “类型” 或不同 “分组” 的均值水平（期望水平）进行比较和检验。

$$salary = \beta_1 + \beta_2 race\_yellow + \beta_3 race\_white + \beta_4 sale + u \tag{8.2}$$

很显然，在总体回归模型式 8.2 下，可以得到所有 3 类 “分组” 情形下的期望年薪水平：

$$\begin{aligned}
&E(Y|race_yellow = 1, race_white = 0, sale) = \beta_1 + \beta_2 + \beta_4 sale \quad \text{(market yellow)} \\
&E(Y|race_yellow = 0, race_white = 1, sale) = \beta_1 + \beta_3 + \beta_4 sale \quad \text{(market white)} \\
&E(Y|race_yellow = 0, race_white = 0, sale) = \beta_1 + \beta_4 sale \quad \text{(market black)}
\end{aligned} \tag{8.3}$$

对于模型式 8.2 ，我们称其为**有截距**的含有虚拟变量的、加法形式的回归模型。显然，虚拟变量 $race\_black$ 没有进入模型式 8.2 中，且模型式 8.2 设置有截距项 β_1。在这种设定下，我们称：

- **黑色 (black)** 为模型式 8.2 的**基础组**
- **黄色 (yellow)** 和**白色 (white)** 分别为模型式 8.2 的**比较组**。

同时，我们称：

- 有序变量 $rank$ 为**协变量 (covariates)** 或**控制变量 (control variable)**
- β_1 为**截距系数**，代表基础组的期望水平

- β_2, β_3 为**极差系数**，代表的是比较组与基础组期望水平的差距
- β_4 为**平行斜率系数**，代表协变量的影响效应

8.2.3 虚拟变量回归模型有哪些形式?

一个线性回归模型，只要回归元中包含了虚拟变量，这种模型就被称为**虚拟变量回归模型**，也可以称为**方差分析模型**（Analysis of variance, ANOVA）[2]。

根据回归元包含定量变量和虚拟变量的数量关系，可以将虚拟变量回归模型分为：

- 只含有虚拟变量的回归模型：全部解释变量都是由虚拟变量构成
- 同时含有虚拟变量和定量变量的回归模型：解释变量同时含有虚拟变量和定量变量

根据虚拟变量引入模型方式的不同，可以划分为：

- 加法模型：虚拟变量以独立项的形式出现在方程中
- 乘法模型：虚拟变量以交叉项的形式出现在方程中
- 混合模型：虚拟变量以独立项和/或交叉项的形式出现在方程中[3]。包括完全混合模型，以及部分混合模型。

根据虚拟变量模型是否参照**基础组**，可以划分为[4]：

- 有截距模型：此时模型解释中将有明确的**基础组**，其他组可以直接与之参照对比。
- 无截距模型：此时模型解释中将没有明确的**基础组**，各组间将不直接参照对比。

根据模型中的因变量 Y 是否取对数，可以划分为[5]：

- 经典线性模型：因变量为 Y
- 半对数模型：因变量为 $ln(Y)$

---

[2] **方差分析模型**（Analysis of variance, ANOVA）常用来分析定量化的因变量 Y 与定性回归元或虚拟变量之间的统计显著性关系。一般是通过比较不同类别或不同组的均值差，例如采用 t 检验可以判断两组均值是否有显著的差异

[3] 有时候模型设置中，某个虚拟变量体系（用来表达某个定性变量）的独立项可以完全不出现在方程中（也即没有它们的加法形式），而却可以出现它们与其他变量的交叉项（也即可以出现它们与其他变量的乘法形式）。

[4] 如果理论要求与**基础组**对比，则理论模型必须设置为有截距回归模型；否则，理论模型需要设置无截距回归模型。

[5] 半对数或对数模型将蕴含着弹性和斜率的经济学含义，在解释虚拟变量回归模型中往往很有现实意义。

根据虚拟变量模型应用情景的不同，可以划分为：

- 截面数据虚拟变量回归模型：此时虚拟变量用于表达回归元为定性变量的情形
- 时间序列季节虚拟变量回归模型：此时虚拟变量用于表达季节周期（具体请参看节 8.5.6）
- 分段线性虚拟变量回归模型：此时虚拟变量用于表达**阀值**分段（具体请参看节 8.5.7）

对于具体的实证分析案例，我们往往需要根据变量的属性和特征，构建不同类型的虚拟变量回归模型，比较不同模型的回归分析结果，甄选并得到其中相对理想的模型。显然，不同类型的虚拟变量模型设置，具有不同的经济学含义。甚至回归方程系数解读的直观性，模型构建意图表达的直接性等，也存在较大差异，都需要对各种备选的、可行的模型进行反复测试和甄选。

例如，仅是考虑**基础组**的有截距模型，可能用到的各类备选组合模型至少包括（具体回归方程设置见节 8.5.4 和节 8.5.5 ）：

- 只含有虚拟变量的、加法形式的**经典回归模型**
- 只含有虚拟变量的、加法形式的**半对数回归模型**
- 只含有虚拟变量的、乘法形式的**经典回归模型**
- 只含有虚拟变量的、乘法形式的**半对数回归模型**
- …
- 同时含有虚拟变量和定量变量的、加法形式的**经典回归模型**
- 同时含有虚拟变量和定量变量的、加法形式的**半对数回归模型**
- 同时含有虚拟变量和定量变量的、乘法形式的**经典回归模型**
- 同时含有虚拟变量和定量变量的、乘法形式的**半对数回归模型**
- …

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

8.3 实验内容

（1）将定性变量都处理成完全虚拟变量体系（也即定性变量有 m 个属性，则设置 m 个虚拟变量）

（2）分析只含有虚拟变量的回归模型（有截距）

a. 只含有虚拟变量的、加法形式的经典回归模型

b. 只含有虚拟变量的、加法形式的半对数回归模型

c. 只含有虚拟变量的、乘法形式的经典回归模型

d. 只含有虚拟变量的、乘法形式的半对数回归模型

（3）分析同时含有虚拟变量和定量变量的回归模型（有截距）：

a. 同时含有虚拟变量和定量变量的、加法形式的经典回归模型

b. 同时含有虚拟变量和定量变量的、加法形式的半对数回归模型

c. 同时含有虚拟变量和定量变量的、乘法形式的经典回归模型

d. 同时含有虚拟变量和定量变量的、乘法形式的半对数回归模型

（4）分析时间序列季节虚拟变量模型

（5）分析分段线性回归模型（piecewise linear regression）

8.4 实验准备

8.4.1 实验软件

本次实验需要提前准备好如下软件：

- 统计分析软件 Eviews 9.0 版本及以上
- 公式编辑软件 Mathtype 6.0 版本及以上
- 写作编辑软件 Office Word/Excel 2010 版本及以上
- 浏览器软件 chrome 66.0 版本及以上或 360 极速浏览器 9.5 版本及以上

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

8.4.2 实验材料

印度工人工资：本次实验将使用表 8.3 给出的 114 位印度工人在 wage 工人工资，age 年龄，edu 教育水平，dpt 合同类型，sex 性别等方面的数据。

表 8.3: 印度工人工资 (n=114)

obs	wage	age	edu	dpt	sex
1	117	26	pri	per	f
2	375	42	pri	per	f
3	175	33	pri	per	f
4	100	33	pri	per	f
5	162	30	pri	per	f
110	25	18	ill	tem	m
111	25	11	ill	tem	m
112	75	45	ill	tem	m
113	54	14	ill	tem	m
114	50	26	ill	tem	m

变量说明见表 8.4：

表 8.4: 变量定义及说明

variable	label	remark
obs	工人编号	序号 (observations)
wage	工人工资	美元/周 ($/week)
age	年龄	岁 (year)
edu	教育水平	ill= 文盲 (illiteracy)；pri= 初等教育 (primary)；mid= 中等教育 (middle)；hig= 高等教育 (higher)
dpt	合同类型	tem= 短期合同 (temporary)；per= 长期合同 (permanent)
sex	性别	f= 女 (female)；m= 男 (male)

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

8.4.3 实验规则

8.4.3.1 模型及参数命名规则

本次实验将对不同组合下的多个模型进行比较分析，我们对这些不同模型及参数做如下约定：

- 不同虚拟变量体系命名，依次以希腊字母体系 $\{\alpha_1, \alpha_2, \cdots\}; \{\beta_1, \beta_2, \cdots\}; \{\gamma_1, \gamma_2, \cdots\}; \cdots$ 表达
- 定量变量参数的命名，依次以希腊字母 $\delta_1; \delta_2; \cdots$ 表达
- 虚拟变量交叉项的参数命名，依次以希腊字母 $\lambda_1; \lambda_2; \cdots$ 表达

例如，同时含有虚拟变量和定量变量的、乘法形式的、经典线性的、完全变量引入回归模型（eq_both_pro_ord_full）

$$wage_i = \alpha_1 + \alpha_2 edu_{D2,i} + \alpha_3 edu_{D3,i} + \alpha_4 edu_{D4,i} + \beta_2 dpt_{D2,i} + \gamma_2 sex_{D2,i} + \delta age_i + \lambda_2 edu_{D2,i} * dpt_{D2,i} + \lambda_3 edu_{D3,i} * dpt_{D2,i} + \lambda_4 edu_{D4,i} * dpt_{D2,i} + u_i$$

8.4.3.2 Eviews 对象命名规则

因为案例模型涉及到多个定性变量和定量变量，同时还考虑对不同组合下的多个模型进行比较分析，为了记忆和分析方便，我们约定如下的 Eviews 对象命名规则：

（1）虚拟变量序列对象（Series）的命名规则：

- 定性变量的属性命名：确定**基础组**，例如本例中的基础组为 {**文盲，短期合同，女性**}（也即 {**illiteracy,temporary,female**}），并将各个定性变量的**基础组属性**都命名为 D_1 虚拟变量，其他比较组属性则依次命名为 $D_2, D_3, \cdots$。
- 虚拟变量体系命名：对于一个定性变量，应完全表达为虚拟变量体系，且以定性变量名 + 虚拟变量记号的方式命名。

  例如:教育水平定性变量 $edu(m = 4)$,应设置为 4 个虚拟变量定性edu_D1、edu_D2、edu_D3、edu_D4。

（2）方程对象（Equation）的命名规则：

- `only`，表示只含有虚拟变量的回归方程

- `both`，表示同时含有虚拟变量和定量变量的回归方程
- `plus`，表示加法形式（plus）的回归方程
- `prod`，表示乘法形式（product）的回归方程
- `ord`，表示因变量为普通形态 Y 的回归方程
- `log`，表示因变量为对数形态 $ln(Y)$ 的回归方程
- `full`，表示一个定性变量的虚拟变量与另一个变量的全部交叉项，以及该定性变量的虚拟变量的独立项，全部进入回归方程。

  例如：有截距模型下（考虑基础组），教育水平定性变量 $edu(m=4)$ 与工作类型定性变量 $dpt(m=2)$ 的交叉项 $edu_{D2}*dpt_{D2}$、$edu_{D3}*dpt_{D2}$ 和 $edu_{D4}*dpt_{D2}$，以及独立项 edu_{D2}、edu_{D3} 和 edu_{D4}，全部进入回归方程。

- `part`，表示一个定性变量的虚拟变量与另一个变量的全部交叉项进入回归方程，但是该定性变量的虚拟变量的独立项不进入回归方程。例如，有截距模型下（考虑基础组），教育水平定性变量 $edu(m=4)$ 与工作类型定性变量 $dpt(m=2)$ 的交叉项 $edu_{D2}*dpt_{D2}$、$edu_{D3}*dpt_{D2}$ 和 $edu_{D4}*dpt_{D2}$ 进入回归方程，但其独立项 edu_{D2}、edu_{D3} 和 edu_{D4} 并不引入回归方程中。

  例如：方程对象eq_only_plus_log_full，代表**只含有虚拟变量的、加法形式的、对数化的、全部变量进入的**回归方程。

8.5 主要实验步骤

8.5.1 导入数据并进行预处理

目标：构建工作文件并正确导入数据

思路：数据结构为 `unstructured/undatede`

新建 Eviews 工作文件（见图 8.1 ）

Eviews 菜单操作：

（1）依次操作：File⇒New⇒Workfile

（2）进行 workfile create 引导设置：

- workfile structure type: `unstructured/undatede`
- data range：114
- workfile names(optional): 建议命名如下

- WF: `worker`（**建议命名**）

- Page: `indian`（**建议命名**）

Eviews 导入数据的菜单操作（Excel 和 Eviews）步骤如下：

> 注意
>
> **提示**：Excel 数据，每个同学的 Y 数据都不同，找到自己学号对应下的 Y 数据（X 数据所有同学都一样）

- Excel 找到数据。Excel 表格中仅保留自己需要的数据（obswageageedudptsex）
- Eviews 导入数据。File ⇒ Import ⇒ Import From File：`d:/econometrics/data/Lab8-indian-wage.xlsx`

8.5.2 把定性变量设置成虚拟变量体系

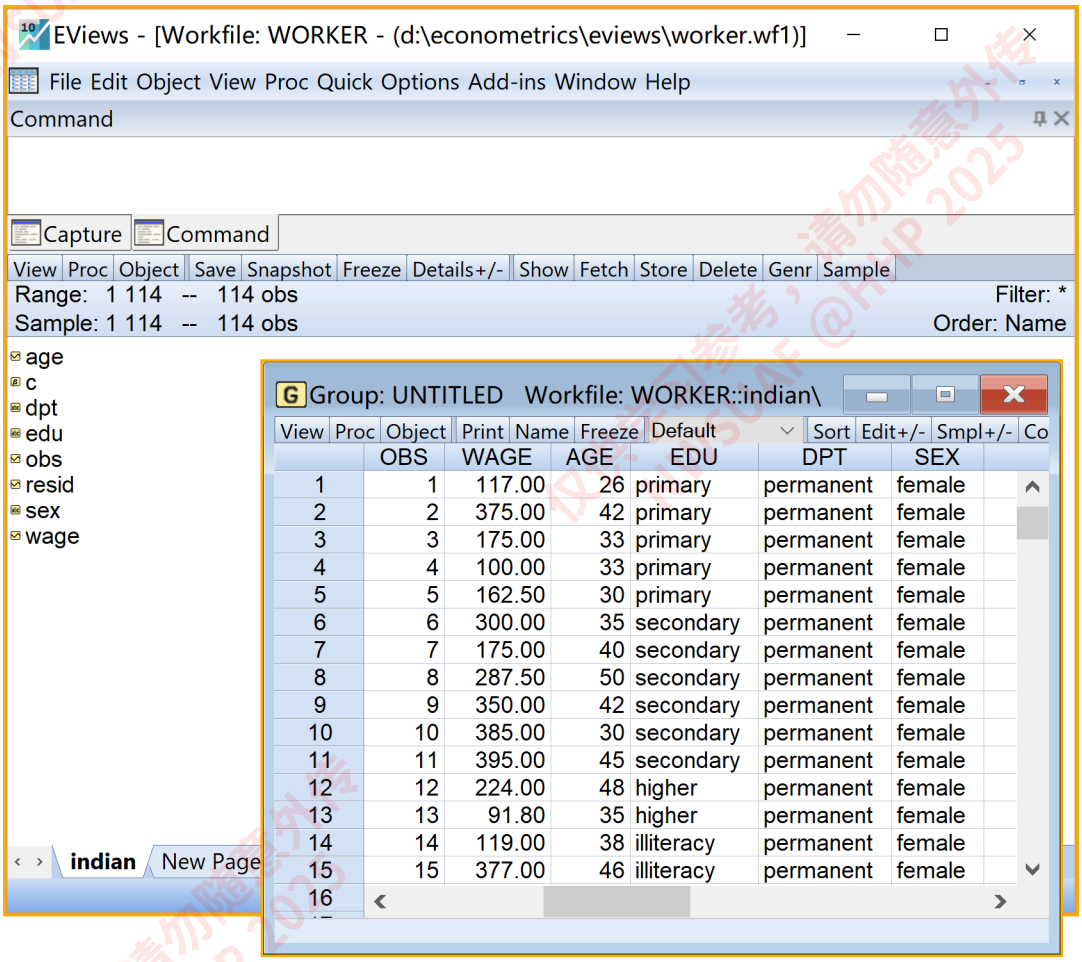
目标：学会用一套虚拟变量体系来完整表达一个定性变量

思路：按照完备、互斥的法则设置虚拟变量；如果要设置有截距模型，应统筹、优先考虑基础组的虚拟变量设置。

Eviews 操作：

（1）把定性变量**教育水平** edu（m=4）设置成虚拟变量体系。（具体操作见图 8.2）

- 命令视窗（Command）依次输入如下命令（建议分别命名为 edu_ill、edu_pri、edu_mid 和 edu_hig）



	OBS	WAGE	AGE	EDU	DPT	SEX
1	1	117.00	26	primary	permanent	female
2	2	375.00	42	primary	permanent	female
3	3	175.00	33	primary	permanent	female
4	4	100.00	33	primary	permanent	female
5	5	162.50	30	primary	permanent	female
6	6	300.00	35	secondary	permanent	female
7	7	175.00	40	secondary	permanent	female
8	8	287.50	50	secondary	permanent	female
9	9	350.00	42	secondary	permanent	female
10	10	385.00	30	secondary	permanent	female
11	11	395.00	45	secondary	permanent	female
12	12	224.00	48	higher	permanent	female
13	13	91.80	35	higher	permanent	female
14	14	119.00	38	illiteracy	permanent	female
15	15	377.00	46	illiteracy	permanent	female
16						

图 8.1: 导入数据的 Eviews 视窗

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

```
- `series edu_ill=@recode(edu="ill",1,0)'`
- `series edu_pri=@recode(edu="pri",1,0)'`
- `series edu_mid=@recode(edu="mid",1,0)`
- `series edu_hig=@recode(edu="hig",1,0)'`
```

- 运行命令：命令行中按 Enter 键
- 查看结果（以组 group 的形式查看）：具体步骤如下

> 提示
>
> - 按住键盘 Ctrl+ 依次点击edu、edu_ill、edu_pri、edu_mid、edu_hig
> - 点击鼠标右键 ⇒ Open ⇒ as Group

（2）把定性变量**合同类型** dpt（m=2）设置成虚拟变量体系。（具体操作见图 8.3）

- 命令视窗（Command）依次输入如下命令（建议分别命名为 dpt_tem 和 dpt_per）

```
`series dpt_tem=@recode(dpt="tem",1,0)'`
`series dpt_per=@recode(dpt="per",1,0)`
```

- 运行命令：命令行中按 Enter 键
- 查看结果（以组 group 的形式查看）：具体步骤如下

> 提示
>
> - 按住键盘 Ctrl+ 依次点击dpt、dpt_tem、dpt_per
> - 点击鼠标右键 ⇒ Open ⇒ as Group

（3）把定性变量**性别** sex（m=2）设置成虚拟变量体系。（具体操作见图 8.4）

- 命令视窗（Command）依次输入如下命令（建议分别命名为 sex_f 和 sex_m）

```
series sex_f=@recode(sex="f",1,0)
series sex_m=@recode(sex="m",1,0)
```

- 运行命令：命令行中按 Enter 键
- 查看结果（以组 group 的形式查看）：具体步骤如下

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

提示

- 按住键盘 Ctrl+ 依次点击sex、sex_f、sex_m
- 点击鼠标右键 ⇒ Open ⇒ as Group

（4）Eviews 代码行的解读

- 代码 `series edu_ill=@recode(edu="ill",1,0)` 表示给创建一个序列（Series）对象edu_ill，并对定性变量对象edu 进行重新编码处理（recode），并把重新编码处理后的数值赋值给序列（Series）对象edu_ill。
- 代码 `@recode(edu="ill",1,0)` 表示对定性变量对象edu 进行重新编码处理。具体做法是，如果 `edu` 的取值为 `ill`，则相应赋值为 1，或者就相应赋值为 0。

8.5.3 操作解读

实际操作中，我们首先要对定性变量进行重新编码，设置成各自的虚拟变量体系。Eviews 中对定性变量重新编码为虚拟变量的代码函数为 `@recode()`。我们可以事先将一个定性变量**完全地**进行虚拟变量编码[6]。也就是说，如果一个定性变量有 m 个属性，我们可以直接设置 m 个虚拟变量。

此外，便于后续多个模型的分析甄选，我们还应该进一步统一设计虚拟变量的名称、命名的顺序等。例如，假设后续的备选模型中将基础组设定为 {文盲，临时工，女性}（也即 {illiteracy，temporary，female}）[7]。则可以将全部定性变量的基础组属性 {illiteracy，temporary，female} 分别设置为虚拟变量 `edu_ill`（见表 8.5）、`dpt_tem`（见表 8.6）和 `sex_f`（见表 8.7）。

表 8.5: 用虚拟变量系统完全表达定性变量 edu (n=114)

	obs	edu	edu_ill	edu_pri	edu_mid	edu_hig
1	1	pri	0	1	0	0
2	2	pri	0	1	0	0
6	6	mid	0	0	1	0
7	7	mid	0	0	1	0
92	92	hig	0	0	0	1

[6]此时我们可以完全不用关心模型是否有截距（意味着是否有对照比较的**基础组**）

[7]理论上，基础组如何选择并不会从根本上改变模型的实际经济学意义，只是一旦选定一个基础组，也就意味着确定了一个相互比较的“基础参照系”。

表 8.5: 用虚拟变量系统完全表达定性变量 edu (n=114)

	obs	edu	edu_ill	edu_pri	edu_mid	edu_hig
93	93	hig	0	0	0	1
113	113	ill	1	0	0	0
114	114	ill	1	0	0	0

表 8.6: 用虚拟变量系统完全表达定性变量 dpt (n=114)

	obs	dpt	dpt_tem	dpt_per
1	1	per	0	1
2	2	per	0	1
3	3	per	0	1
4	4	per	0	1
111	111	tem	1	0
112	112	tem	1	0
113	113	tem	1	0
114	114	tem	1	0

表 8.7: 用虚拟变量系统完全表达定性变量 sex (n=114)

	obs	sex	sex_f	sex_m
1	1	f	1	0
2	2	f	1	0
3	3	f	1	0
4	4	f	1	0
111	111	m	0	1
112	112	m	0	1
113	113	m	0	1
114	114	m	0	1

8.5.4 仅含有虚拟变量

目标：把定性变量的虚拟变量以独立项的形式引入模型方程，解释回归报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

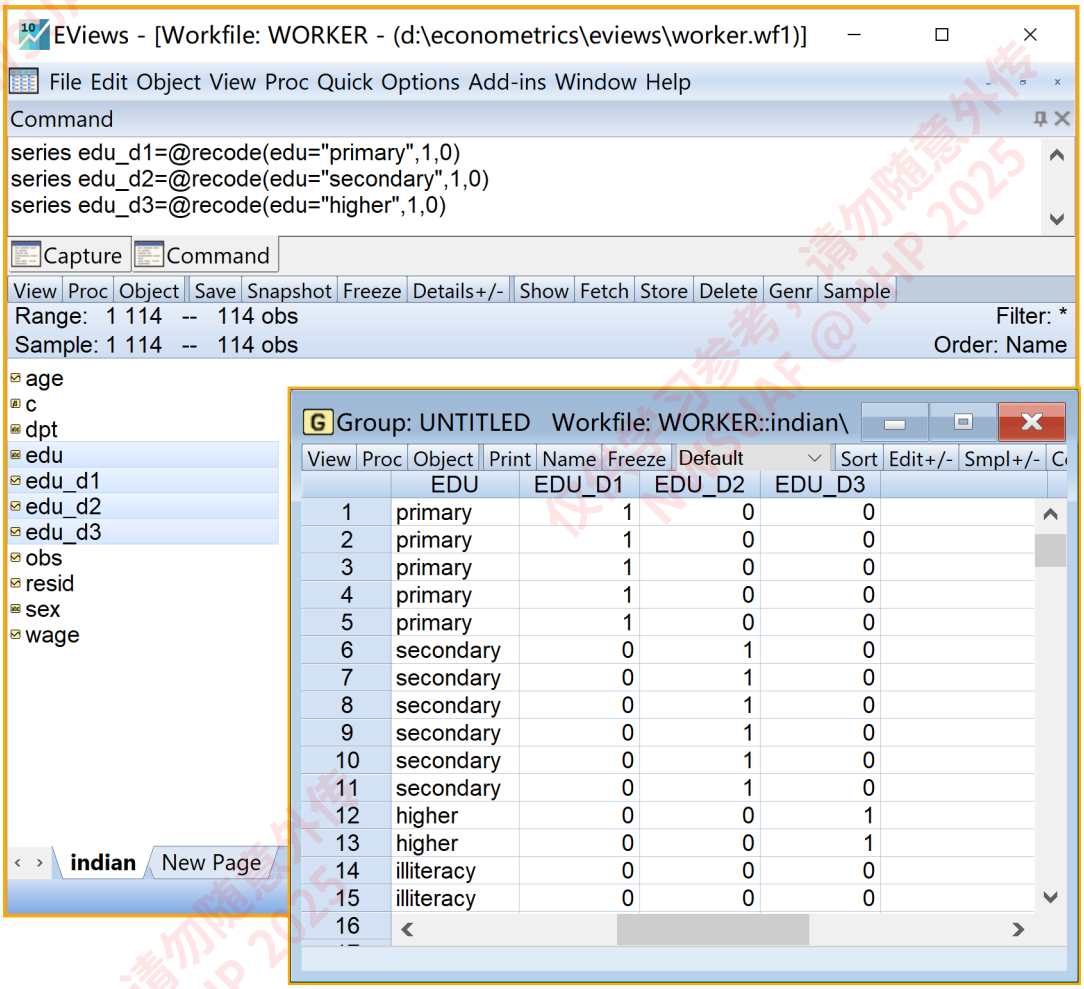


图 8.2: 定性变量 edu 用虚拟变量体系表达

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

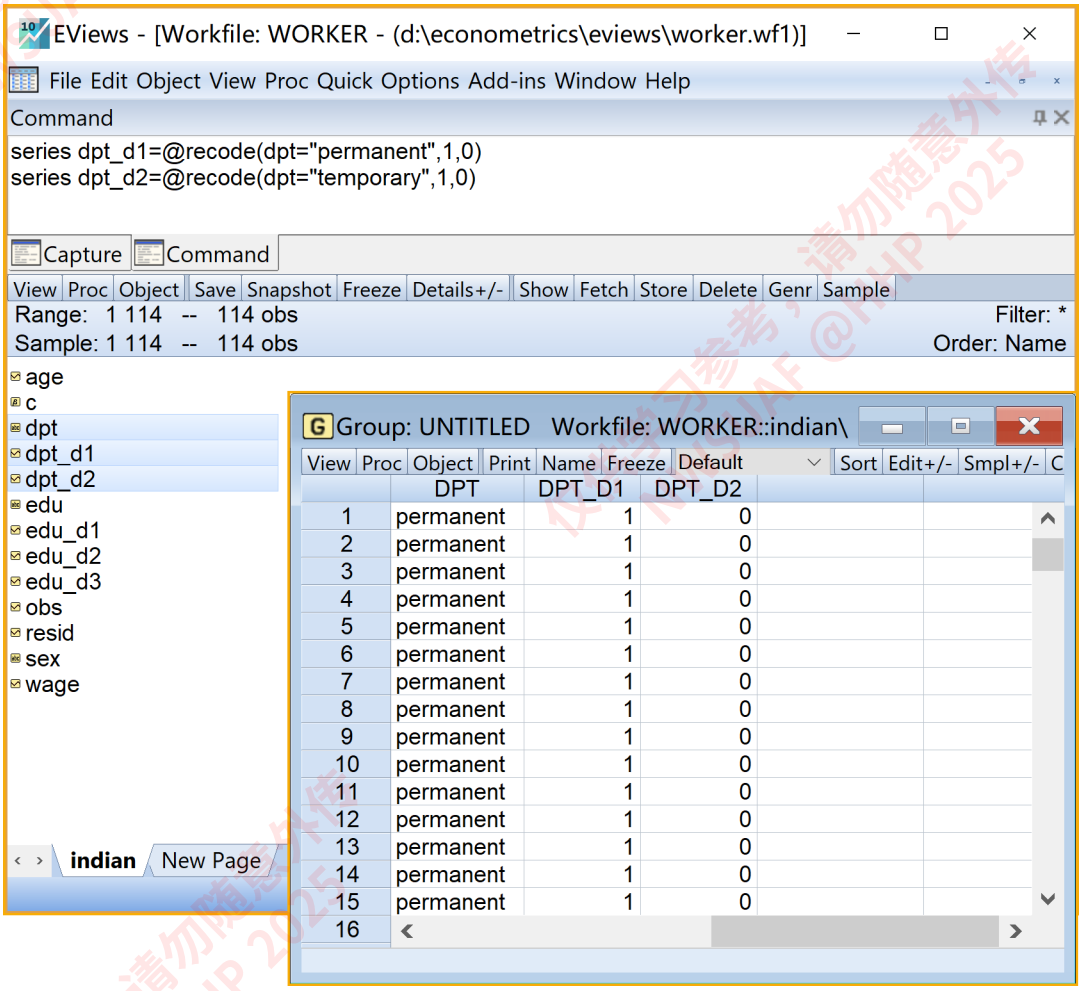


图 8.3: 定性变量 dpt 用虚拟变量体系表达

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

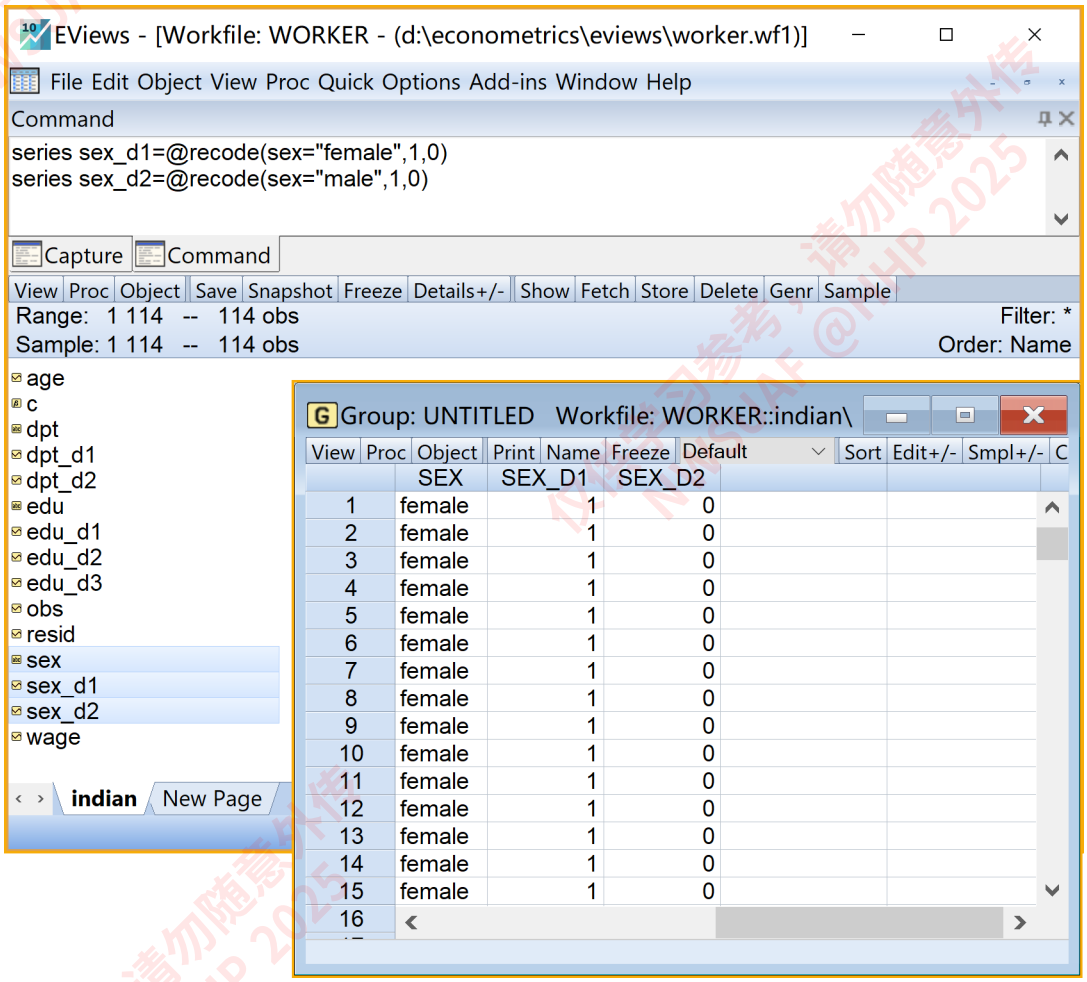


图 8.4: 定性变量 sex 用虚拟变量体系表达

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

思路：确定**基础组**，设置总体回归模型（PRM），进行 OLS 估计，得到 Eviews 分析报告

8.5.4.1 加法-经典模型

只含有虚拟变量的、加法形式的**经典回归模型**见方程式 8.4

$$\begin{aligned} wage_i = &+ \beta_1 + \beta_2 edu\_pri_i + \beta_3 edu\_mid_i \\ &+ \beta_4 edu\_hig_i + \beta_5 dpt\_per_i + \beta_6 sex\_m_i + u_i \end{aligned} \tag{8.4}$$

只含有虚拟变量的、加法形式的**经典回归模型**见方程式 8.4 ，菜单操作实现具体见图 8.5。具体步骤如下：

（1）确定基础组为 [**文盲 & 短期合同 & 女性**]，则如下虚拟变量将**不进入**回归模型

- ☑`edu_ill`
- ☑`dpt_tem`
- ☑`sex_f`

（2）设置回归模型。进入引导设置 Equation Estimation ⇒ specification

- Equation specification：输入命令 `wage c edu_pri edu_mid edu_hig dpt_per sex_m`
- Estimation settings：具体设置如下

> 提示
>
> - Method: 下拉选择 LS - Least Squares (NLS and ARMA)
> - Sample: （默认设置）

- 点击完成：OK
- 命名保存方程对象🟰：（建议命名为 `eq_only_plus`）
- 查看结果：双击🟰`eq_only_plus`

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

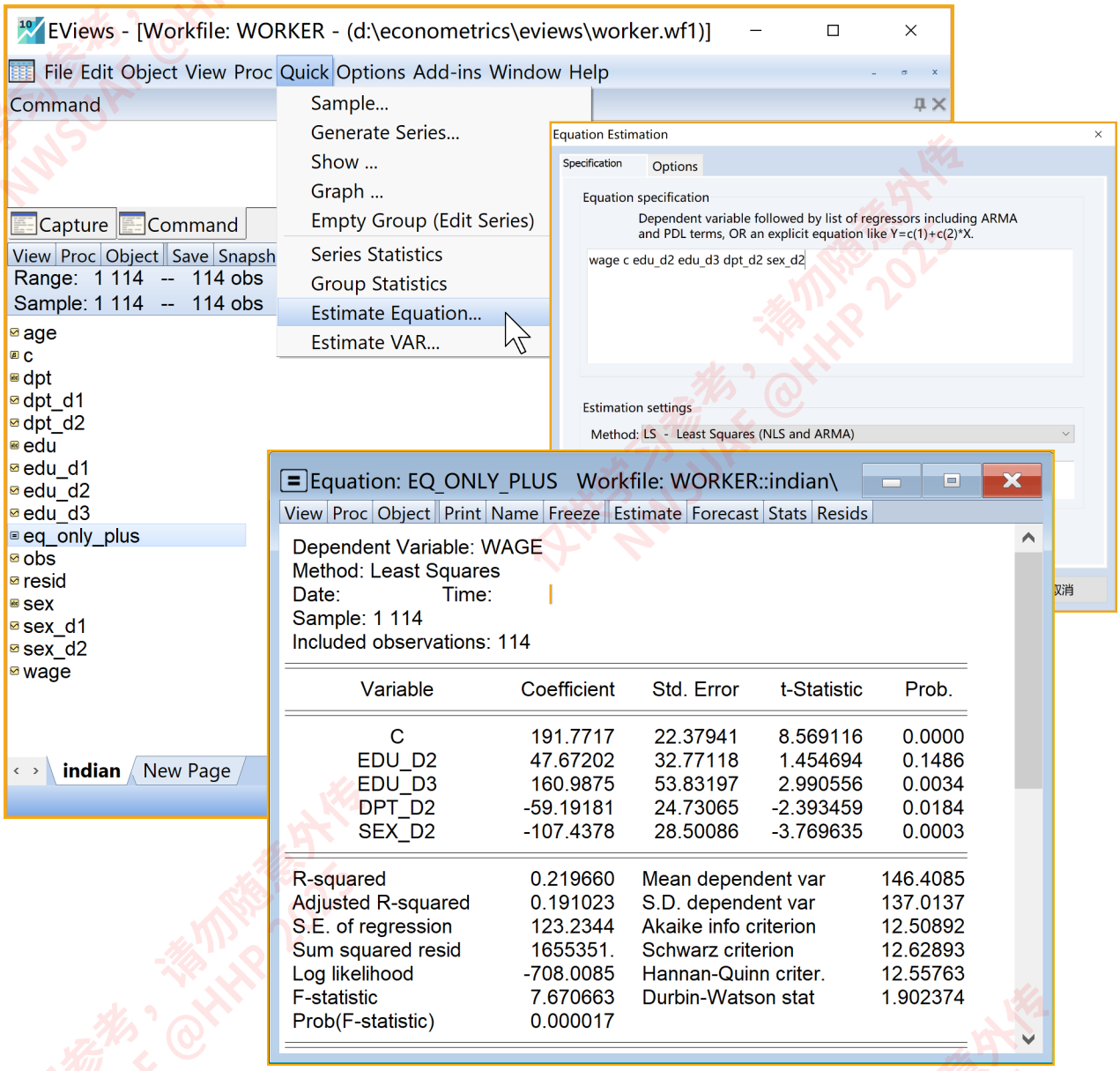


图 8.5: 只含虚拟变量的、加法形式的经典线性回归模型 Eviews 实现

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

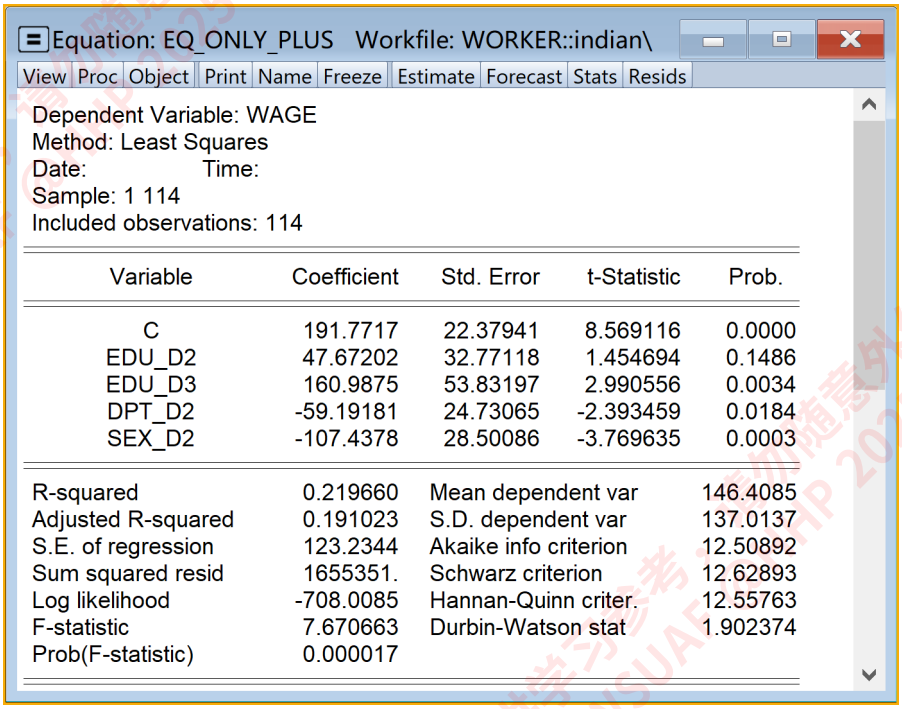


图 8.6: 只含虚拟变量的、加法形式的经典线性回归模型 Eviews 报告

具体 Eviews 报告见 @ref(fig:only-plus-report)：

只含虚拟变量的、加法形式的经典回归模型 Eviews 结果简要报告如下：

$$\begin{aligned} wage_i = &+ \beta_1 + \beta_2 edu\_pri_i + \beta_3 edu\_mid_i + \beta_4 edu\_hig_i \\ &+ \beta_5 dpt\_per_i + \beta_6 sex\_m_i + u_i \end{aligned} \tag{8.5}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{array}{llll}
\widehat{wage} = & +130.36 & +10.47 edu\_pri_i & +49.71 edu\_mid_i \\
(s) & (17.7609) & (33.4850) & (33.5470) \\
(t) & (+7.34) & (+0.31) & (+1.48) \\
(cont.) & +162.59 edu\_hig_i & +59.37 dpt\_per_i & -106.49 sex\_m_i \\
(s) & (54.2989) & (24.8399) & (28.7784) \\
(t) & (+2.99) & (+2.39) & (-3.70) \\
(over) & n = 114 & \hat{\sigma} = 123.7476 & \\
(fit) & R^2 = 0.2204 & \bar{R}^2 = 0.1843 & \\
(Ftest) & F^* = 6.11 & p = 0.0001 &
\end{array}
\tag{8.6}
$$

此时，根据总体回归模型式 8.5 和样本回归方程式 8.6 我们可以得到：

（1）基础组（**文盲 & 短期合同 & 女性**），也即 (**illiteracy & temporary & female**) 的期望工资收入见式 8.7，拟合的期望工资收入见式 8.8：

$$
\begin{aligned}
& E(wage|edu\_pri = 0; edu\_mid = 0; edu\_hig = 0; dpt\_per = 0; sex\_m = 0) \\
=& + \beta_1(1) + \beta_2(0) + \beta_3(0) + \beta_4(0) \\
& + \beta_5(0) + \beta_6(0) \\
=& \beta_1
\end{aligned}
\tag{8.7}
$$

$$
\begin{aligned}
& (\widehat{wage}|edu\_pri = 0; edu\_mid = 0; edu\_hig = 0; dpt\_per = 0; sex\_m = 0) \\
=& + \hat{\beta}_1(1) + \hat{\beta}_2(0) + \hat{\beta}_3(0) + \hat{\beta}_4(0) \\
& + \hat{\beta}_5(0) + \hat{\beta}_6(0) \\
=& + [130.36] \cdot (1) + [10.47] \cdot (0) + [49.71] \cdot (0) + [162.59] \cdot (0) \\
& + [59.37] \cdot (0) - [106.49] \cdot (0) \\
=& 130.36
\end{aligned}
\tag{8.8}
$$

（2）比较组 1（**文盲 & 短期合同 & 男性**），也即 (**illiteracy & temporary & male**) 的期望工资收入见式 8.9，拟合的期望工资收入见式 8.10：

$$
\begin{aligned}
&E(wage|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=0;sex\_m=1)\\
=&+\beta_1(1)+\beta_2(0)+\beta_3(0)+\beta_4(0)\\
&+\beta_5(0)+\beta_6(1)\\
=&\beta_1+\beta_6
\end{aligned}
\tag{8.9}
$$

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=0;sex\_m=1)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)+\hat{\beta}_4(0)\\
&+\hat{\beta}_5(0)+\hat{\beta}_6(1)\\
=&+[130.36]\cdot(1)+[10.47]\cdot(0)+[49.71]\cdot(0)+[162.59]\cdot(0)\\
&+[59.37]\cdot(0)-[106.49]\cdot(1)\\
=&23.87
\end{aligned}
\tag{8.10}
$$

（3）比较组 2（**高等教育 & 长期合同 & 男性**），也即 (**higher & permanent & male**) 的期望工资收入见式 8.11，拟合的期望工资收入见式 8.12：

$$
\begin{aligned}
&E(wage|edu\_pri=0;edu\_mid=0;edu\_hig=1;dpt\_per=1;sex\_m=1)\\
=&+\beta_1(1)+\beta_2(0)+\beta_3(0)+\beta_4(1)\\
&+\beta_5(1)+\beta_6(1)\\
=&\beta_1+\beta_4+\beta_5+\beta_6
\end{aligned}
\tag{8.11}
$$

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=1;dpt\_per=1;sex\_m=1)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)+\hat{\beta}_4(1)\\
&+\hat{\beta}_5(1)+\hat{\beta}_6(1)\\
=&+[130.36]\cdot(1)+[10.47]\cdot(0)+[49.71]\cdot(0)+[162.59]\cdot(1)\\
&+[59.37]\cdot(1)-[106.49]\cdot(1)\\
=&245.82
\end{aligned}
\tag{8.12}
$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

8.5.4.2 加法-半对数模型

只含有虚拟变量的、加法形式的**半对数回归模型**见方程式 8.13

$$\begin{aligned} log(wage)_i = &+ \beta_1 + \beta_2 edu\_pri_i + \beta_3 edu\_mid_i + \beta_4 edu\_hig_i \\ &+ \beta_5 dpt\_per_i + \beta_6 sex\_m_i + u_i \end{aligned} \tag{8.13}$$

只含有虚拟变量的、加法形式的**半对数回归模型**见方程式 8.13 ，菜单操作实现具体见图 8.7。具体步骤如下：

（1）确定基础组为 [**文盲 & 短期合同 & 女性**]，则如下虚拟变量将**不进入**回归模型

- `edu_ill`
- `dpt_tem`
- `sex_f`

（2）设置回归模型。进入引导设置 Equation Estimation ⇒ specification

- Equation specification：输入命令 `log(wage) c edu_pri edu_mid edu_hig dpt_per sex_m`
- Estimation settings。
  - Method: 下拉选择 LS - Least Squares (NLS and ARMA)
  - Sample: （默认设置）
- 点击完成：OK
- 命名保存方程对象：（建议命名为 `eq_only_plus_log`）
- 查看结果：双击`eq_only_plus_log`

具体 Eviews 报告见 @ref(fig:only-plus-log-report)：

作为对比，只含虚拟变量的、加法形式的半对数回归模型 Eviews 结果简要报告如下：

$$\begin{aligned} log(wage)_i = &+ \beta_1 + \beta_2 edu\_pri_i + \beta_3 edu\_mid_i + \beta_4 edu\_hig_i \\ &+ \beta_5 dpt\_per_i + \beta_6 sex\_m_i + u_i \end{aligned} \tag{8.14}$$

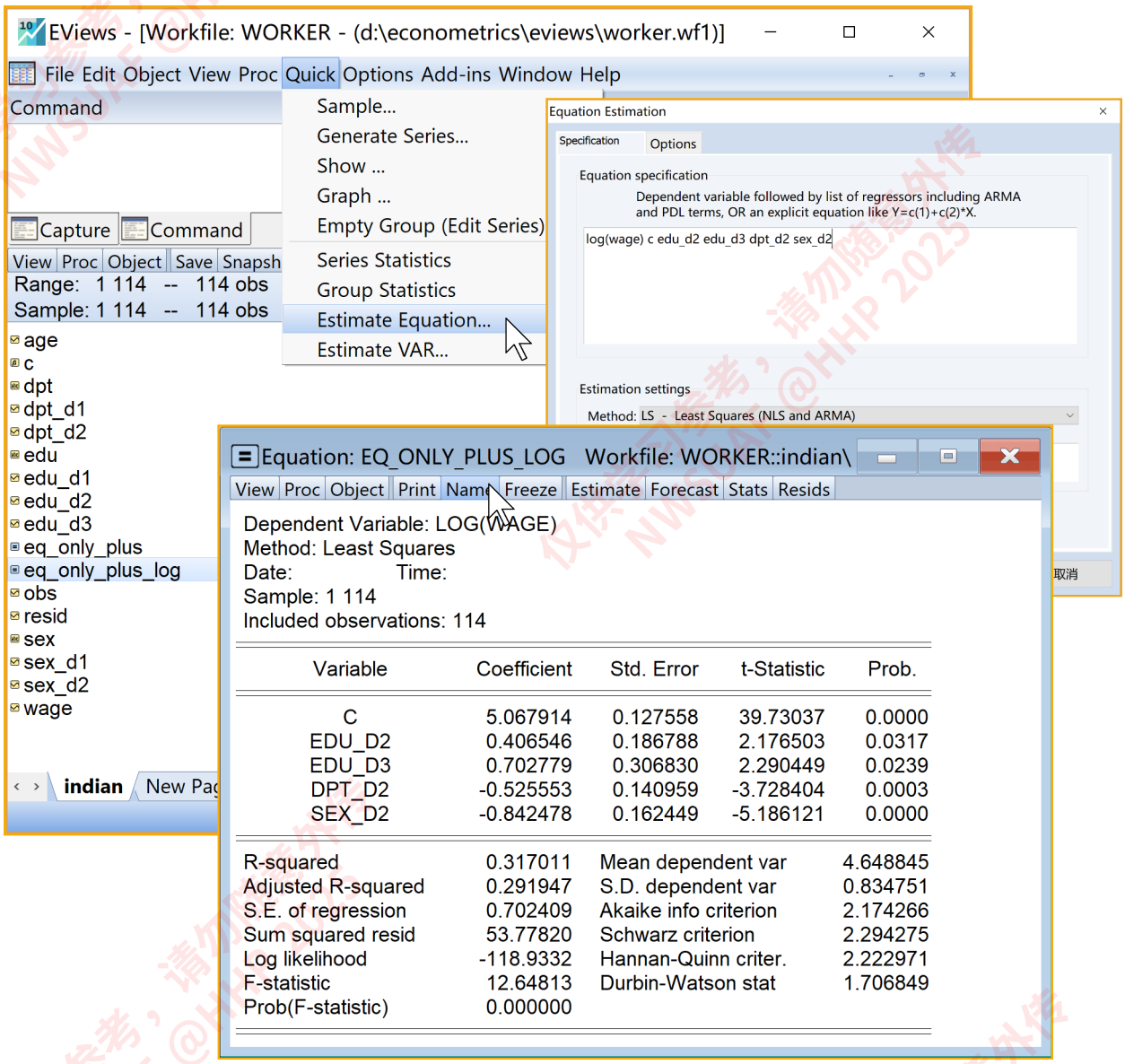
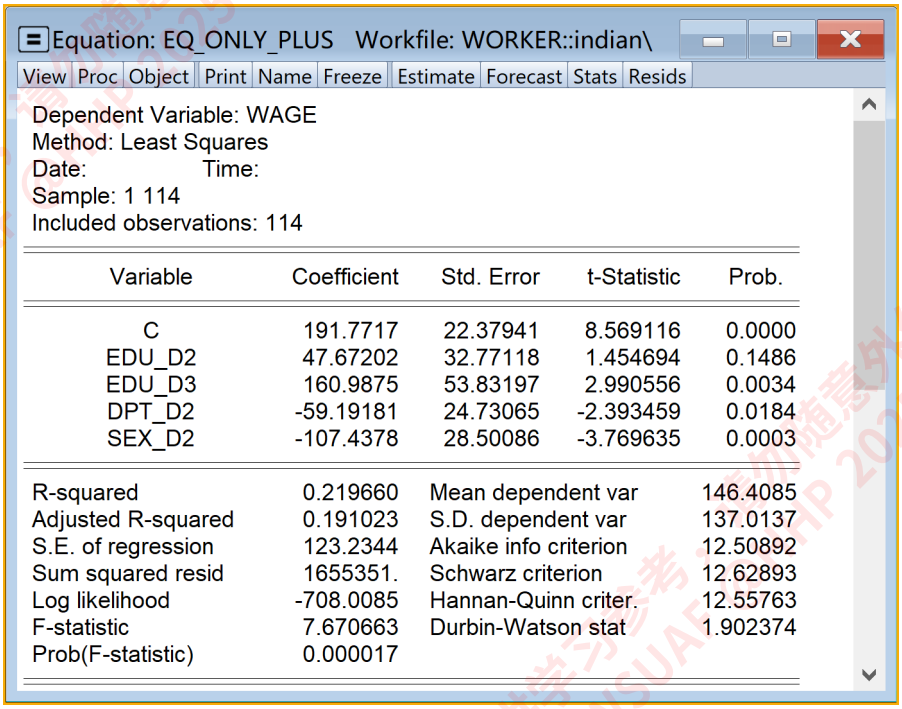


图 8.7: 只含虚拟变量的、加法形式的半对数线性回归模型 Eviews 实现

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025



Equation: EQ_ONLY_PLUS Workfile: WORKER::indian\

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

Dependent Variable: WAGE
Method: Least Squares
Date: Time:
Sample: 1 114
Included observations: 114

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	191.7717	22.37941	8.569116	0.0000
EDU_D2	47.67202	32.77118	1.454694	0.1486
EDU_D3	160.9875	53.83197	2.990556	0.0034
DPT_D2	-59.19181	24.73065	-2.393459	0.0184
SEX_D2	-107.4378	28.50086	-3.769635	0.0003

R-squared	0.219660	Mean dependent var	146.4085
Adjusted R-squared	0.191023	S.D. dependent var	137.0137
S.E. of regression	123.2344	Akaike info criterion	12.50892
Sum squared resid	1655351.	Schwarz criterion	12.62893
Log likelihood	-708.0085	Hannan-Quinn criter.	12.55763
F-statistic	7.670663	Durbin-Watson stat	1.902374
Prob(F-statistic)	0.000017		

图 8.8: 只含虚拟变量的、加法形式的半对数线性回归模型 Eviews 报告

$$
\begin{array}{llll}
\widehat{log(wage)} = & +4.53 & +0.06edu\_pri_i & +0.42edu\_mid_i \\
(s) & (0.1012) & (0.1909) & (0.1912) \\
(t) & (+44.75) & (+0.30) & (+2.19) \\
(cont.) & +0.71edu\_hig_i & +0.53dpt\_per_i & -0.84sex\_m_i \\
(s) & (0.3095) & (0.1416) & (0.1640) \\
(t) & (+2.30) & (+3.72) & (-5.10) \\
(over) & n = 114 & \hat{\sigma} = 0.7054 & \\
(fit) & R^2 = 0.3176 & \bar{R}^2 = 0.2860 & \\
(Ftest) & F^* = 10.05 & p = 0.0000 &
\end{array}
\tag{8.15}
$$

8.5.4.3 乘法-经典模型（完全）

只含有虚拟变量的、乘法形式的**经典回归模型（完全交叉项）**见方程式 8.16 ；

$$
\begin{aligned}
wage_i = &+ \beta_1 + \beta_2 sex\_m_i + \beta_3 edu\_pri_i \\
&+ \beta_4 edu\_mid_i + \beta_5 edu\_hig_i + \beta_6 dpt\_per_i \\
&+ \beta_7 edu\_pri \times dpt\_per_i + \beta_8 edu\_mid \times dpt\_per_i + \beta_9 edu\_hig \times dpt\_per_i + u_i
\end{aligned} \tag{8.16}
$$

只含有虚拟变量的、乘法形式的**经典回归模型**见方程式 8.16 ，菜单操作实现具体见图 8.9：

（1）确定基础组为 [**文盲 & 短期合同 & 女性**]，则如下虚拟变量将**不进入**回归模型

- ☑`edu_d1`
- ☑`dpt_d1`
- ☑`sex_d1`

（2）设置回归模型。进入引导设置 Equation Estimation ⇒ specification

- Equation specification：输入命令 `wage c edu_d2 edu_d3 edu_d4 dpt_d2 sex_d2 edu_d2*dpt_d2 edu_d3*dpt_d2 edu_d4*dpt_d2`
- Estimation settings：
  - Method: 下拉选择 LS - Least Squares (NLS and ARMA)
  - Sample: （默认设置）
- 点击完成：OK
- 命名保存方程对象🟰：（建议命名为 `eq_only_prod`）
- 查看结果：双击🟰`eq_only_prod`

具体 Eviews 报告见 @ref(fig:only-prod-report)：

只含虚拟变量的、乘法形式的经典回归模型（全部变量进入，见式 8.17）Eviews 结果简要报告如下（见式 8.18）：

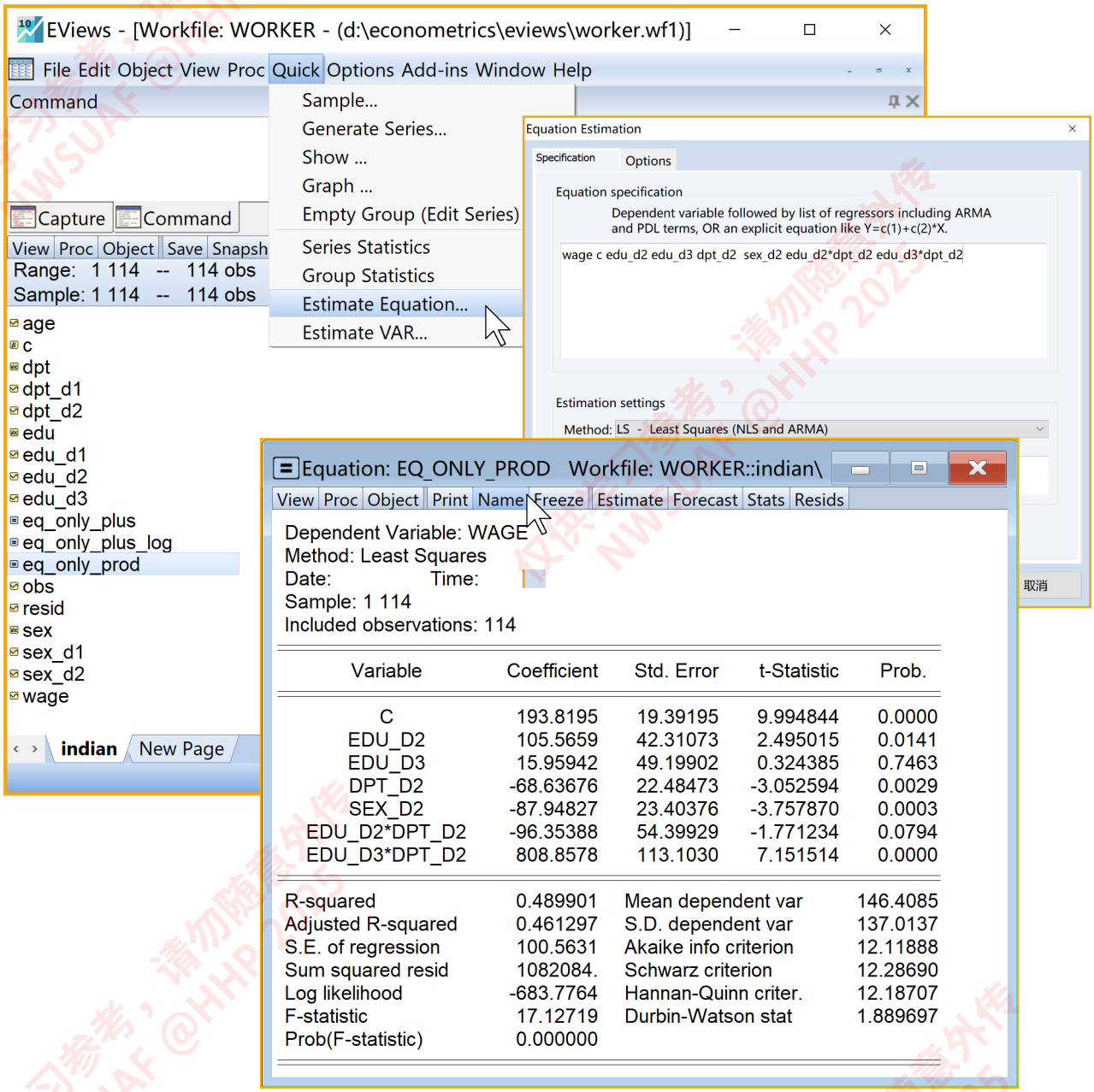


图 8.9: 只含虚拟变量的、乘法形式的经典线性回归模型 Eviews 实现

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

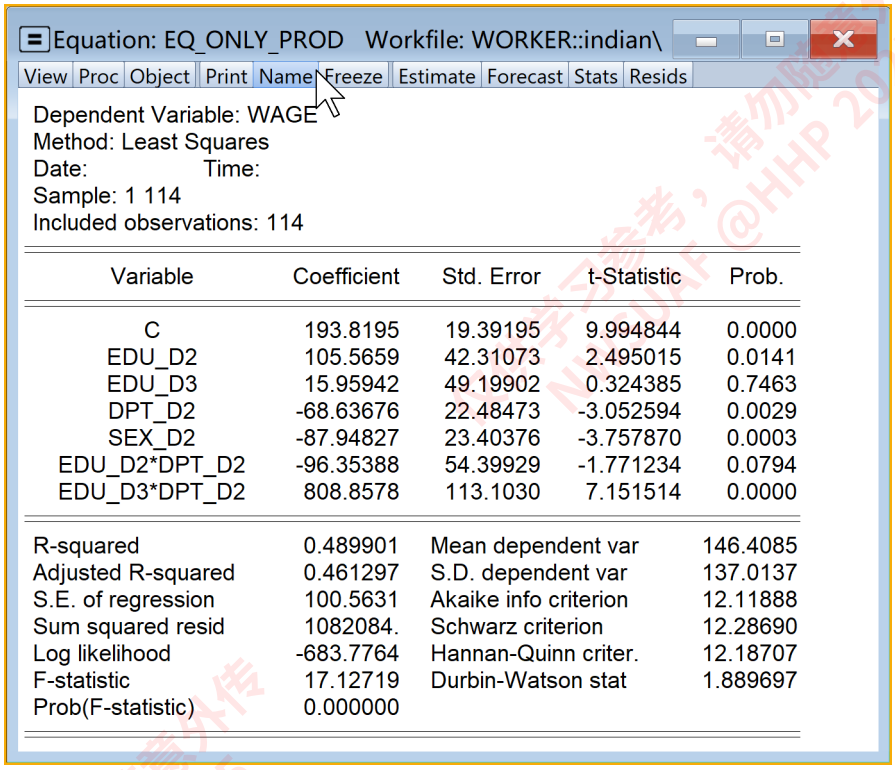


图 8.10: 只含虚拟变量的、乘法形式的经典线性回归模型 Eviews 报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{aligned}
wage_i = &+ \beta_1 + \beta_2 sex\_m_i + \beta_3 edu\_pri_i \\
&+ \beta_4 edu\_mid_i + \beta_5 edu\_hig_i + \beta_6 dpt\_per_i \\
&+ \beta_7 edu\_pri \times dpt\_per_i + \beta_8 edu\_mid \times dpt\_per_i + \beta_9 edu\_hig \times dpt\_per_i + u_i
\end{aligned}
\tag{8.17}
$$

$$
\begin{array}{llll}
\widehat{wage} = & +120.70 & -88.19 sex\_m_i & +23.03 edu\_pri_i \\
(s) & (15.2625) & (23.8515) & (32.6264) \\
(t) & (+7.91) & (-3.70) & (+0.71) \\
(cont.) & +13.72 edu\_mid_i & +829.30 edu\_hig_i & +74.78 dpt\_per_i \\
(s) & (35.2241) & (102.4020) & (25.0404) \\
(t) & (+0.39) & (+8.10) & (+2.99) \\
(cont.) & -32.62 edu\_pri:dpt\_per_i & +90.22 edu\_mid:dpt\_per_i & -814.86 edu\_hig:dpt\_per_i \\
(s) & (59.7485) & (55.7896) & (114.2653) \\
(t) & (-0.55) & (+1.62) & (-7.13) \\
(over) & n = 114 & \hat{\sigma} = 101.2582 & \\
(fit) & R^2 = 0.4925 & \bar{R}^2 = 0.4538 & \\
(Ftest) & F^* = 12.74 & p = 0.0000 &
\end{array}
\tag{8.18}
$$

此时，根据总体回归模型式 8.17 和样本回归方程式 8.18 我们可以得到：

（1）基础组（**文盲 & 短期合同 & 女性**），也即 (**illiteracy & temporary & female**) 的期望工资收入见式 8.19，拟合的期望工资收入见式 8.20：

$$
\begin{aligned}
&E(wage|edu\_pri = 0; edu\_mid = 0; edu\_hig = 0; dpt\_per = 0; sex\_m = 0) \\
=&+ \beta_1(1) + \beta_2(0) + \beta_3(0) \\
&+ \beta_4(0) + \beta_5(0) + \beta_6(0) \\
&+ \beta_7(0)\cdot(0) + \beta_8(0)\cdot(0) + \beta_9(0)\cdot(0) \\
=&\beta_1
\end{aligned}
\tag{8.19}
$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=0;sex\_m=0)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)\\
&+\hat{\beta}_4(0)+\hat{\beta}_5(0)+\hat{\beta}_6(0)\\
&+\hat{\beta}_7(0)\cdot(0)+\hat{\beta}_8(0)\cdot(0)+\hat{\beta}_9(0)\cdot(0)\\
=&+[120.70]\cdot(1)-[88.19]\cdot(0)+[23.03]\cdot(0)\\
&+[13.72]\cdot(0)+[829.30]\cdot(0)+[74.78]\cdot(0)\\
&-[32.62]\cdot(0)\cdot(0)+[90.22]\cdot(0)\cdot(0)-[814.86]\cdot(0)\cdot(0)\\
=&120.70
\end{aligned}
\tag{8.20}
$$

（2）比较组 1（**文盲 & 长期合同 & 女性**），也即 (**illiteracy & temporary & male**) 的期望工资收入见式 8.21，拟合的期望工资收入见式 8.22：

$$
\begin{aligned}
&E(wage|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=1;sex\_m=0)\\
=&+\beta_1(1)+\beta_2(0)+\beta_3(0)\\
&+\beta_4(0)+\beta_5(0)+\beta_6(1)\\
&+\beta_7(0)\cdot(1)+\beta_8(0)\cdot(1)+\beta_9(0)\cdot(1)\\
=&\beta_1+\beta_6
\end{aligned}
\tag{8.21}
$$

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=1;sex\_m=0)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)\\
&+\hat{\beta}_4(0)+\hat{\beta}_5(0)+\hat{\beta}_6(1)\\
&+\hat{\beta}_7(0)\cdot(1)+\hat{\beta}_8(0)\cdot(1)+\hat{\beta}_9(0)\cdot(1)\\
=&+[120.70]\cdot(1)-[88.19]\cdot(0)+[23.03]\cdot(0)\\
&+[13.72]\cdot(0)+[829.30]\cdot(0)+[74.78]\cdot(1)\\
&-[32.62]\cdot(0)\cdot(1)+[90.22]\cdot(0)\cdot(1)-[814.86]\cdot(0)\cdot(1)\\
=&195.48
\end{aligned}
\tag{8.22}
$$

（3）比较组 2（**高等教育 & 长期合同 & 女性**），也即 (**higher & permanent & male**) 的期望工资收入见式 8.23，拟合的期望工资收入见式 8.24：

$$
\begin{aligned}
&E(wage|edu\_pri=0;edu\_mid=0;edu\_hig=1;dpt\_per=1;sex\_m=0)\\
=&+\beta_1(1)+\beta_2(0)+\beta_3(0)\\
&+\beta_4(0)+\beta_5(1)+\beta_6(1)\\
&+\beta_7(0)\cdot(1)+\beta_8(0)\cdot(1)+\beta_9(1)\cdot(1)\\
=&\beta_1+\beta_5+\beta_6+\beta_9
\end{aligned}
\tag{8.23}
$$

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=1;dpt\_per=1;sex\_m=0)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)\\
&+\hat{\beta}_4(0)+\hat{\beta}_5(1)+\hat{\beta}_6(1)\\
&+\hat{\beta}_7(0)\cdot(1)+\hat{\beta}_8(0)\cdot(1)+\hat{\beta}_9(1)\cdot(1)\\
=&+[120.70]\cdot(1)-[88.19]\cdot(0)+[23.03]\cdot(0)\\
&+[13.72]\cdot(0)+[829.30]\cdot(1)+[74.78]\cdot(1)\\
&-[32.62]\cdot(0)\cdot(1)+[90.22]\cdot(0)\cdot(1)-[814.86]\cdot(1)\cdot(1)\\
=&209.92
\end{aligned}
\tag{8.24}
$$

8.5.4.4 乘法-经典模型（部分交叉）

只含有虚拟变量的、乘法形式的**经典回归模型（部分交叉项）**见方程式 8.25 ：

$$
\begin{aligned}
wage_i=&+\beta_1+\beta_2 sex\_m_i+\beta_3 dpt\_per_i+\beta_4 dpt\_per\times edu\_pri_i\\
&+\beta_5 dpt\_per\times edu\_mid_i+\beta_6 dpt\_per\times edu\_hig_i+u_i
\end{aligned}
\tag{8.25}
$$

只含虚拟变量的、乘法形式的经典回归模型（部分变量进入）Eviews 结果简要报告如下：

$$
\begin{aligned}
wage_i=&+\beta_1+\beta_2 sex\_m_i+\beta_3 dpt\_per_i+\beta_4 dpt\_per\times edu\_pri_i\\
&+\beta_5 dpt\_per\times edu\_mid_i+\beta_6 dpt\_per\times edu\_hig_i+u_i
\end{aligned}
\tag{8.26}
$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{array}{llll}
\widehat{wage} = & +139.60 & -97.26 sex\_m_i & +58.78 dpt\_per_i \quad -12.48 dpt\_per : edu\_pr \\
(s) & (15.9406) & (29.8761) & (29.8391) \quad (63.0830) \\
(t) & (+8.76) & (-3.26) & (+1.97) \quad (-0.20) \\
(cont.) & +102.33 dpt\_per : edu\_mid_i & +16.98 dpt\_per : edu\_hig_i & \\
(s) & (54.6840) & (62.9130) & \\
(t) & (+1.87) & (+0.27) & \\
(over) & n = 114 & \hat{\sigma} = 127.2802 & \\
(fit) & R^2 = 0.1752 & \bar{R}^2 = 0.1370 & \\
(Ftest) & F^* = 4.59 & p = 0.0008 &
\end{array}
\tag{8.27}
$$

8.5.4.5 乘法-半对数模型（部分交叉）

只含有虚拟变量的、乘法形式的**半对数回归模型（部分交叉项）**见方程式 8.28

$$
\begin{aligned}
log(wage)_i = & + \beta_1 + \beta_2 sex\_m_i + \beta_3 dpt\_per_i + \beta_4 dpt\_per \times edu\_pri_i \\
& + \beta_5 dpt\_per \times edu\_mid_i + \beta_6 dpt\_per \times edu\_hig_i + u_i
\end{aligned}
\tag{8.28}
$$

只含有虚拟变量的、乘法形式的**半对数回归模型**见方程式 8.28 ，菜单操作实现具体见图 8.11：

（1）确定基础组为 [**文盲 & 短期合同 & 女性**]，则如下虚拟变量将**不进入**回归模型

- ☑`edu_d1`
- ☑`dpt_d1`
- ☑`sex_d1`

（2）设置回归模型。进入引导设置 Equation Estimation ⇒ specification

- Equation specification：输入命令 `log(wage) c dpt_d2 sex_d2 edu_d2*dpt_d2 edu_d3*dpt_d2 edu_d4*dpt_d2`
- Estimation settings：
  - Method: 下拉选择 LS - Least Squares (NLS and ARMA)

    - Sample: （默认设置）

- 点击完成：OK
- 命名保存方程对象：（建议命名为 `eq_only_prod_log_part`）
- 查看结果：双击`eq_only_prod_log_part`

具体 Eviews 报告见 @ref(fig:only-prod-log-part-report)：

Equation: EQ_ONLY_PROD_LOG_PART Workfile: WORKER::indian\

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

Dependent Variable: LOG(WAGE)
Method: Least Squares
Date: Time:
Sample: 1 114
Included observations: 114

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.206461	0.116388	44.73376	0.0000
DPT_D2	-0.673343	0.139977	-4.810390	0.0000
SEX_D2	-0.797415	0.158877	-5.019071	0.0000
EDU_D2*DPT_D2	0.252461	0.237452	1.063209	0.2900
EDU_D3*DPT_D2	2.323345	0.700912	3.314745	0.0012

R-squared	0.332218	Mean dependent var	4.648845
Adjusted R-squared	0.307712	S.D. dependent var	0.834751
S.E. of regression	0.694545	Akaike info criterion	2.151749
Sum squared resid	52.58079	Schwarz criterion	2.271758
Log likelihood	-117.6497	Hannan-Quinn criter.	2.200454
F-statistic	13.55672	Durbin-Watson stat	1.640313
Prob(F-statistic)	0.000000		

图 8.12: 只含虚拟变量的、乘法形式的半对数线性回归模型 Eviews 报告

只含虚拟变量的、乘法形式的半对数回归模型（部分变量进入）Eviews 结果简要报告如下：

$$\begin{aligned} log(wage)_i = &+ \beta_1 + \beta_2 sex\_m_i + \beta_3 dpt\_per_i + \beta_4 dpt\_per \times edu\_pri_i \\ &+ \beta_5 dpt\_per \times edu\_mid_i + \beta_6 dpt\_per \times edu\_hig_i + u_i \end{aligned} \tag{8.29}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

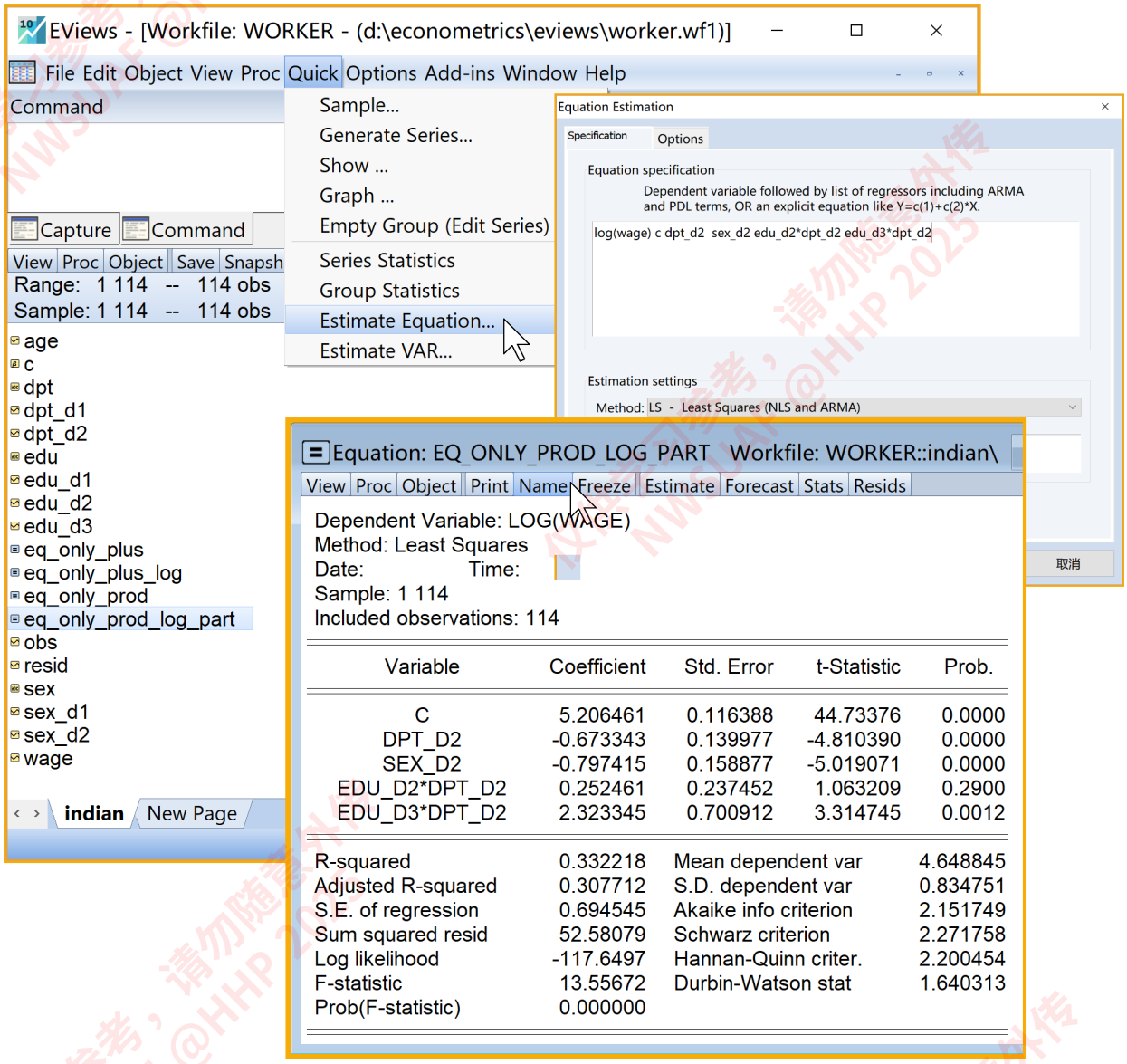


图 8.11: 只含虚拟变量的、乘法形式的半对数线性回归模型 Eviews 实现

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{array}{llll}
\widehat{log(wage)} = & +4.61 & -0.82 sex\_m_i & +0.45 dpt\_per_i \quad +0.06 dpt\_per:edu\_pri \\
(s) & (0.0899) & (0.1685) & (0.1683) \quad (0.3559) \\
(t) & (+51.21) & (-4.89) & (+2.67) \quad (+0.16) \\
(cont.) & +0.64 dpt\_per:edu\_mid_i & +0.38 dpt\_per:edu\_hig_i & \\
(s) & (0.3085) & (0.3549) & \\
(t) & (+2.09) & (+1.08) & \\
(over) & n = 114 & \hat{\sigma} = 0.7181 & \\
(fit) & R^2 = 0.2928 & \bar{R}^2 = 0.2600 & \\
(Ftest) & F^* = 8.94 & p = 0.0000 &
\end{array}
\tag{8.30}
$$

8.5.5 同时含有虚拟变量和定量变量

8.5.5.1 加法-经典模型

同时含有虚拟变量和定量变量的、加法形式的**经典回归模型**见方程式 8.31

$$
\begin{aligned}
wage_i = & + \beta_1 + \beta_2 edu\_pri_i + \beta_3 edu\_mid_i + \beta_4 edu\_hig_i \\
& + \beta_5 dpt\_per_i + \beta_6 sex\_m_i + \beta_7 age_i + u_i
\end{aligned}
\tag{8.31}
$$

同时含有虚拟变量和定量变量的、加法形式的**经典回归模型**见方程式 8.31 ，菜单操作实现具体见图 8.13）：

（1）确定基础组为 [**文盲 & 短期合同 & 女性**]，则如下虚拟变量将**不进入**回归模型

- edu_d1
- dpt_d1
- sex_d1

（2）设置回归模型。进入引导设置 Equation Estimation ⇒ specification

- Equation specification：输入命令 `wage c edu_d2 edu_d3 edu_d4 dpt_d2 sex_d2 age`

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- Estimation settings：

  具体设置为：- Method: 下拉选择 LS - Least Squares (NLS and ARMA) - Sample:（默认设置）

- 点击完成：OK
- 命名保存方程对象🟰：（建议命名为 `eq_both_plus`）
- 查看结果：双击🟰`eq_both_plus`

具体 Eviews 报告见 @ref(fig:both-plus-report)：

同时含虚拟变量和定量变量的、加法形式的经典回归模型 Eviews 结果简要报告如下：

$$\begin{aligned} wage_i = &+ \beta_1 + \beta_2 edu\_pri_i + \beta_3 edu\_mid_i + \beta_4 edu\_hig_i \\ &+ \beta_5 dpt\_per_i + \beta_6 sex\_m_i + \beta_7 age_i + u_i \end{aligned} \tag{8.32}$$

$$\begin{array}{llllll} \widehat{wage} = & +6.79 & +23.96 edu\_pri_i & +61.59 edu\_mid_i & +150.49 edu\_hig_i & \\ (s) & (31.8931) & (30.9789) & (31.0035) & (50.0725) & \\ (t) & (+0.21) & (+0.77) & (+1.99) & (+3.01) & \\ (cont.) & +31.16 dpt\_per_i & -83.20 sex\_m_i & +3.99 age_i & & \\ (s) & (23.7120) & (26.9981) & (0.8835) & & \\ (t) & (+1.31) & (-3.08) & (+4.51) & & \\ (over) & n = 114 & \hat{\sigma} = 113.9518 & & & \\ (fit) & R^2 = 0.3450 & \bar{R}^2 = 0.3083 & & & \\ (Ftest) & F^* = 9.39 & p = 0.0000 & & & \end{array} \tag{8.33}$$

此时，根据总体回归模型式 8.32 和样本回归方程式 8.33 我们可以得到：

（1）基础组（**文盲 & 短期合同 & 女性**），也即 (**illiteracy & temporary & female**)，并给定年龄为 30 岁，其期望工资收入见式 8.34，拟合的期望工资收入见式 8.35：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

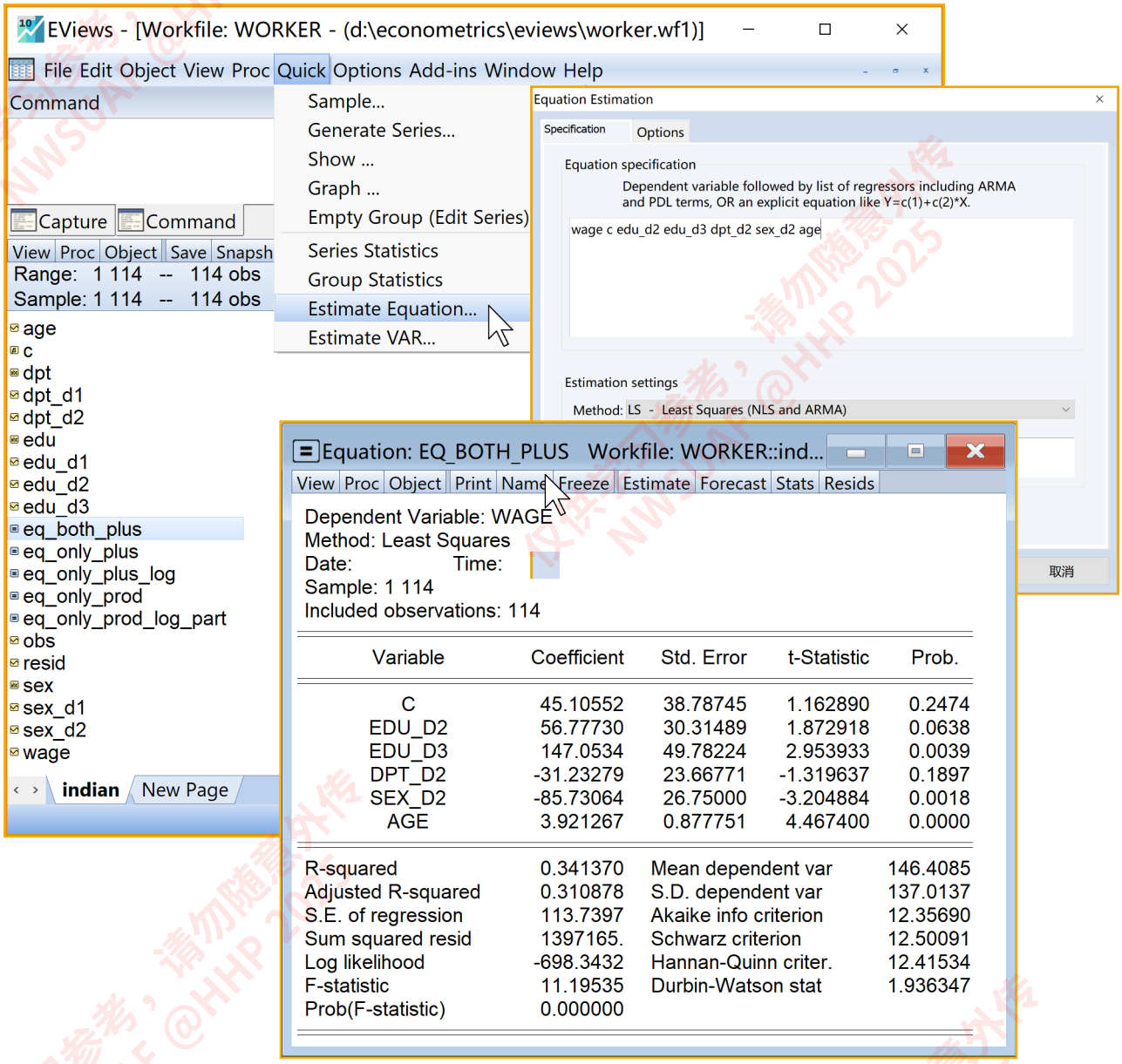


图 8.13: 同时含虚拟变量和定量变量的、加法形式的经典线性回归模型 Eviews 实现

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

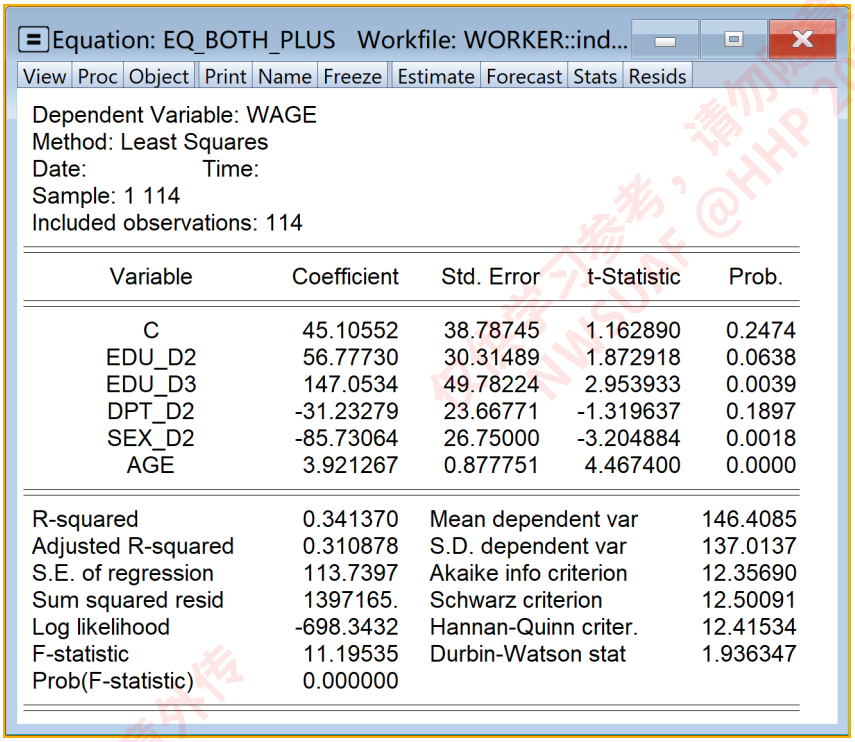


图 8.14: 同时含虚拟变量和定量变量的、加法形式的经典线性回归模型 Eviews 报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{aligned}
&E(wage|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=0;sex\_m=0;age=30)\\
=&+\beta_1(1)+\beta_2(0)+\beta_3(0)+\beta_4(0)\\
&+\beta_5(0)+\beta_6(0)+\beta_7(30.00)\\
=&\beta_1+30.00\beta_7
\end{aligned}
\tag{8.34}
$$

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=0;sex\_m=0;age=30)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)+\hat{\beta}_4(0)\\
&+\hat{\beta}_5(0)+\hat{\beta}_6(0)+\hat{\beta}_7(30.00)\\
=&+[6.79]\cdot(1)+[23.96]\cdot(0)+[61.59]\cdot(0)+[150.49]\cdot(0)\\
&+[31.16]\cdot(0)-[83.20]\cdot(0)+[3.99]\cdot(30.00)\\
=&126.41
\end{aligned}
\tag{8.35}
$$

（2）比较组 1（**文盲 & 短期合同 & 男性**），也即（**illiteracy & temporary & male**），并给定年龄为 30 岁，其期望工资收入见式 8.36，拟合的期望工资收入见式 8.37：

$$
\begin{aligned}
&E(wage|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=0;sex\_m=1;age=30)\\
=&+\beta_1(1)+\beta_2(0)+\beta_3(0)+\beta_4(0)\\
&+\beta_5(0)+\beta_6(1)+\beta_7(30.00)\\
=&\beta_1+\beta_6+30.00\beta_7
\end{aligned}
\tag{8.36}
$$

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=0;sex\_m=1;age=30)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)+\hat{\beta}_4(0)\\
&+\hat{\beta}_5(0)+\hat{\beta}_6(1)+\hat{\beta}_7(30.00)\\
=&+[6.79]\cdot(1)+[23.96]\cdot(0)+[61.59]\cdot(0)+[150.49]\cdot(0)\\
&+[31.16]\cdot(0)-[83.20]\cdot(1)+[3.99]\cdot(30.00)\\
=&43.21
\end{aligned}
\tag{8.37}
$$

（2）比较组 2（**高等教育 & 长期合同 & 男性**），也即（**higher & permanent & male**），且给定年龄为 30 岁，其期望工资收入见式 8.38，拟合的期望工资收入见式 8.39：

$$
\begin{aligned}
&E(wage|edu\_pri=0;edu\_mid=0;edu\_hig=1;dpt\_per=1;sex\_m=1;age=30)\\
=&+\beta_1(1)+\beta_2(0)+\beta_3(0)+\beta_4(1)\\
&+\beta_5(1)+\beta_6(1)+\beta_7(30.00)\\
=&\beta_1+\beta_4+\beta_5+\beta_6+30.00\beta_7
\end{aligned}
\tag{8.38}
$$

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=1;dpt\_per=1;sex\_m=1;age=30)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)+\hat{\beta}_4(1)\\
&+\hat{\beta}_5(1)+\hat{\beta}_6(1)+\hat{\beta}_7(30.00)\\
=&+[6.79]\cdot(1)+[23.96]\cdot(0)+[61.59]\cdot(0)+[150.49]\cdot(1)\\
&+[31.16]\cdot(1)-[83.20]\cdot(1)+[3.99]\cdot(30.00)\\
=&224.86
\end{aligned}
\tag{8.39}
$$

8.5.5.2 加法-半对数模型

同时含有虚拟变量和定量变量的、加法形式的**半对数回归模型**见方程式 8.40

$$
\begin{aligned}
log(wage)_i=&+\beta_1+\beta_2 edu\_pri_i+\beta_3 edu\_mid_i+\beta_4 edu\_hig_i\\
&+\beta_5 dpt\_per_i+\beta_6 sex\_m_i+\beta_7 age_i+u_i
\end{aligned}
\tag{8.40}
$$

同时含有虚拟变量和定量变量的、加法形式的**半对数回归模型**见方程式 8.40 ，菜单操作实现具体见图 8.15：

（1）确定基础组为 [**文盲 & 短期合同 & 女性**]，则如下虚拟变量将**不进入**回归模型 a. ☑`edu_d1` b. ☑`dpt_d1` c. ☑`sex_d1`

（2）设置回归模型。进入引导设置 Equation Estimation ⇒ specification

- Equation specification：输入命令 `log(wage) c edu_d2 edu_d3 edu_d4 dpt_d2 sex_d2 age`
- Estimation settings：

  具体设置为：- Method: 下拉选择 LS - Least Squares (NLS and ARMA) - Sample:（默认设置）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 点击完成：OK
- 命名保存方程对象：（建议命名为 `eq_both_plus_log`）
- 查看结果：双击`eq_both_plus_log`

具体 Eviews 报告见 @ref(fig:both-plus-log-report)：

同时含虚拟变量和定量变量的、加法形式的半对数回归模型 Eviews 结果简要报告如下：

$$\begin{aligned} log(wage)_i = &+ \beta_1 + \beta_2 edu\_pri_i + \beta_3 edu\_mid_i + \beta_4 edu\_hig_i \\ &+ \beta_5 dpt\_per_i + \beta_6 sex\_m_i + \beta_7 age_i + u_i \end{aligned} \tag{8.41}$$

$$\begin{array}{llll} \widehat{log(wage)} = & +3.62 & +0.16 edu\_pri_i & +0.51 edu\_mid_i \quad +0.62 edu\_hig_i \\ (s) & (0.1699) & (0.1651) & (0.1652) \quad (0.2668) \\ (t) & (+21.31) & (+0.95) & (+3.06) \quad (+2.33) \\ (cont.) & +0.32 dpt\_per_i & -0.67 sex\_m_i & +0.03 age_i \\ (s) & (0.1263) & (0.1438) & (0.0047) \\ (t) & (+2.53) & (-4.63) & (+6.23) \\ (over) & n = 114 & \hat{\sigma} = 0.6071 & \\ (fit) & R^2 = 0.4991 & \bar{R}^2 = 0.4710 & \\ (Ftest) & F^* = 17.77 & p = 0.0000 & \end{array} \tag{8.42}$$

8.5.5.3 乘法-经典模型（完全）

同时含有虚拟变量和定量变量的、乘法形式的**经典回归（完全）模型**见方程式 8.43

$$\begin{aligned} wage_i = &+ \beta_1 + \beta_2 sex\_m_i + \beta_3 edu\_pri_i + \beta_4 edu\_mid_i \\ &+ \beta_5 edu\_hig_i + \beta_6 dpt\_per_i + \beta_7 edu\_pri \times dpt\_per_i + \beta_8 edu\_mid \times dpt\_per_i \\ &+ \beta_9 edu\_hig \times dpt\_per_i + \beta_{10} age_i + u_i \end{aligned} \tag{8.43}$$

同时含有虚拟变量和定量变量的、乘法形式的**经典回归模型**见方程式 8.43 ，菜单操作实现具体见图 8.17：

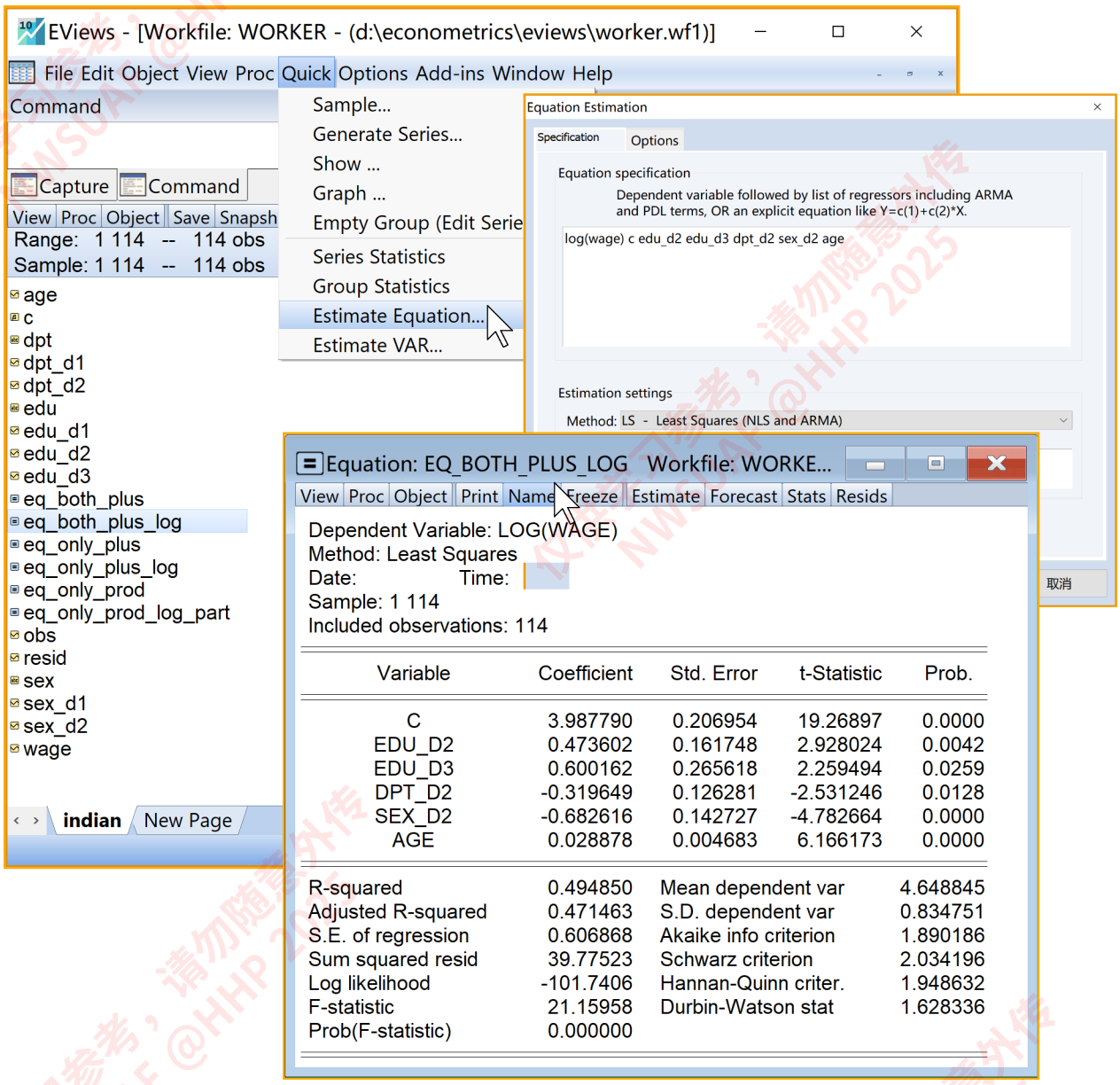


图 8.15: 同时含虚拟变量和定量变量的、加法形式的半对数线性回归模型 Eviews 实现

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

Equation: EQ_BOTH_PLUS_LOG Workfile: WORKE...

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

Dependent Variable: LOG(WAGE)
Method: Least Squares
Date: Time:
Sample: 1 114
Included observations: 114

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.987790	0.206954	19.26897	0.0000
EDU_D2	0.473602	0.161748	2.928024	0.0042
EDU_D3	0.600162	0.265618	2.259494	0.0259
DPT_D2	-0.319649	0.126281	-2.531246	0.0128
SEX_D2	-0.682616	0.142727	-4.782664	0.0000
AGE	0.028878	0.004683	6.166173	0.0000

R-squared	0.494850	Mean dependent var	4.648845
Adjusted R-squared	0.471463	S.D. dependent var	0.834751
S.E. of regression	0.606868	Akaike info criterion	1.890186
Sum squared resid	39.77523	Schwarz criterion	2.034196
Log likelihood	-101.7406	Hannan-Quinn criter.	1.948632
F-statistic	21.15958	Durbin-Watson stat	1.628336
Prob(F-statistic)	0.000000		

图 8.16: 同时含虚拟变量和定量变量的、加法形式的半对数线性回归模型 Eviews 报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

（1）确定基础组为 [**文盲 & 短期合同 & 女性**]，则如下虚拟变量将**不进入**回归模型 a. ☑`edu_d1` b. ☑`dpt_d1` c. ☑`sex_d1`

（2）设置回归模型。进入引导设置 Equation Estimation ⇒ specification

- Equation specification：输入命令 `wage c edu_d2 edu_d3 edu_d4 dpt_d2 sex_d2 age edu_d2*dpt_d2 edu_d3*dpt_d2 edu_d4*dpt_d2`
- Estimation settings：

  具体设置为：- Method: 下拉选择 LS - Least Squares (NLS and ARMA) - Sample:（默认设置）

- 点击完成：OK
- 命名保存方程对象🟰：（建议命名为 `eq_both_prod`）
- 查看结果：双击🟰`eq_both_prod`

具体 Eviews 报告见 @ref(fig:both-prod-report)：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

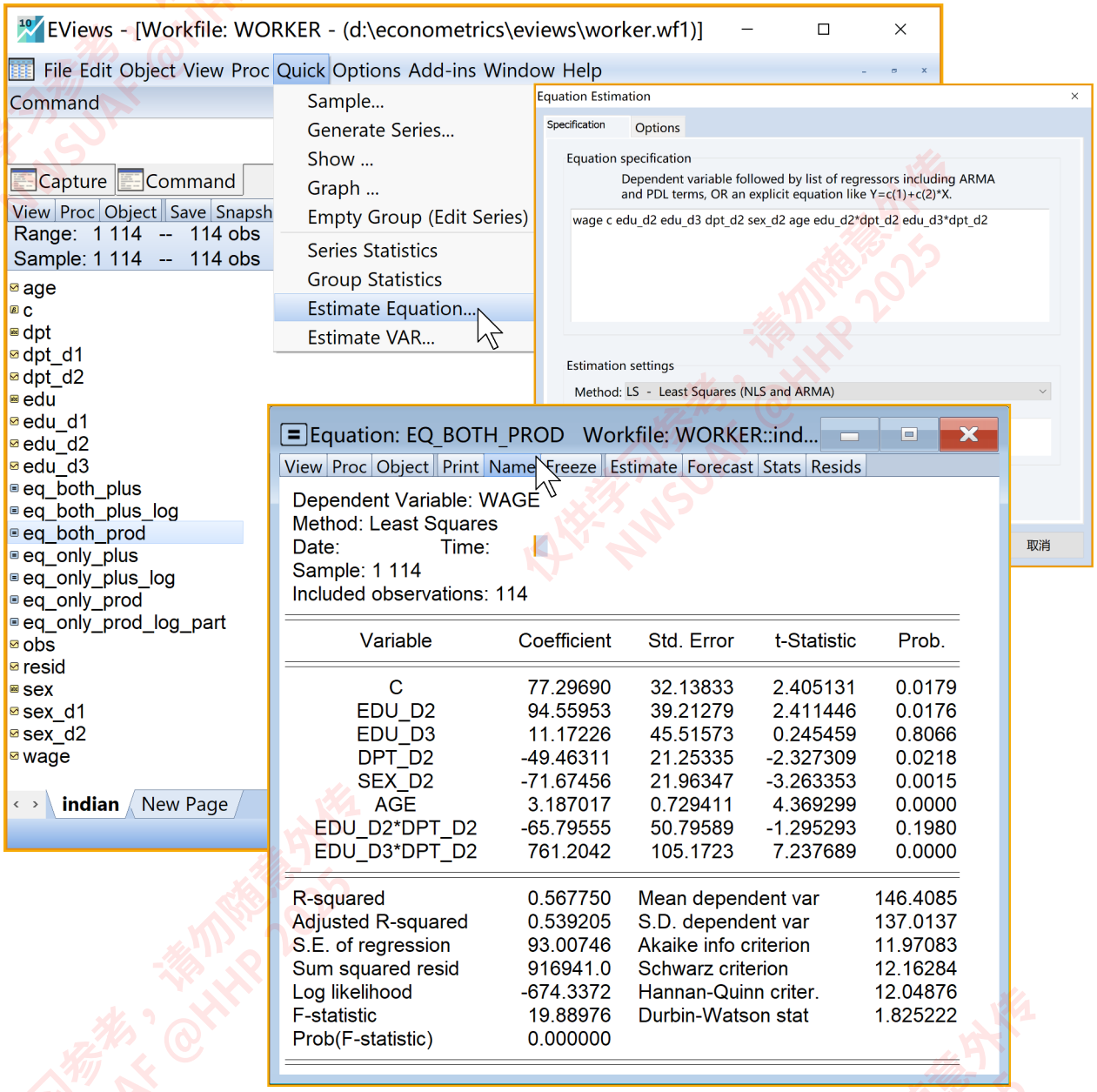


图 8.17: 同时含虚拟变量和定量变量的、乘法形式的经典线性回归模型 Eviews 实现

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

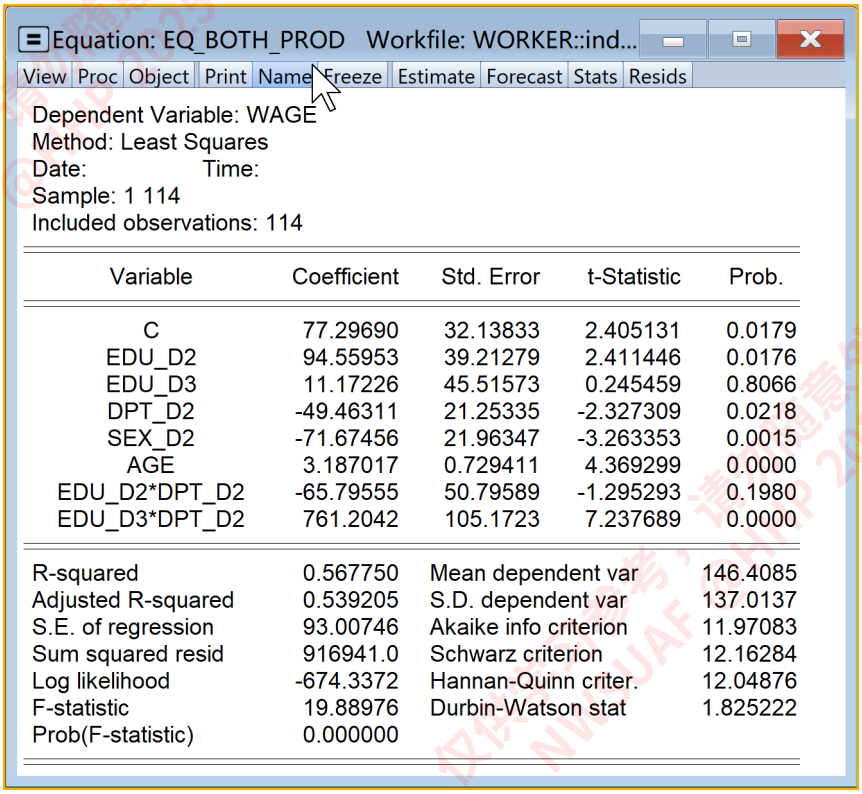


图 8.18: 同时含虚拟变量和定量变量的、乘法形式的经典线性回归模型 Eviews 报告

同时含虚拟变量和定量变量的、乘法形式的经典回归模型（见式 8.44）（全部变量进入）Eviews 结果简要报告（见式 8.45）如下：

$$\begin{aligned} wage_i = &+ \beta_1 + \beta_2 sex\_m_i + \beta_3 edu\_pri_i + \beta_4 edu\_mid_i \\ &+ \beta_5 edu\_hig_i + \beta_6 dpt\_per_i + \beta_7 edu\_pri \times dpt\_per_i + \beta_8 edu\_mid \times dpt\_per_i \\ &+ \beta_9 edu\_hig \times dpt\_per_i + \beta_{10} age_i + u_i \end{aligned} \tag{8.44}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{array}{llll}
\widehat{wage} = & +19.43 & -70.34 sex\_m_i & +32.29 edu\_pri_i \quad +35.59 edu\_mid_i \\
(s) & (26.9248) & (22.3647) & (30.1607) \quad (32.8591) \\
(t) & (+0.72) & (-3.15) & (+1.07) \quad (+1.08) \\
(cont.) & +777.93 edu\_hig_i & +54.46 dpt\_per_i & +3.25 age_i \quad -26.81 edu\_pri : dpt\_pe \\
(s) & (95.1488) & (23.5466) & (0.7361) \quad (55.1151) \\
(t) & (+8.18) & (+2.31) & (+4.41) \quad (-0.49) \\
(cont.) & +59.80 edu\_mid : dpt\_per_i & -766.28 edu\_hig : dpt\_per_i & \\
(s) & (51.9083) & (105.9477) & \\
(t) & (+1.15) & (-7.23) & \\
(over) & n = 114 & \hat{\sigma} = 93.3791 & \\
(fit) & R^2 = 0.5725 & \bar{R}^2 = 0.5355 & \\
(Ftest) & F^* = 15.48 & p = 0.0000 &
\end{array}
\tag{8.45}
$$

此时，根据上述总体回归模型式 8.44 和样本回归方程式 8.45 我们可以得到：

（1）基础组为：**文盲 & 短期合同 & 女性** (也即 **illiteracy & temporary & female**)，并给定年龄为 30 岁（$age = 30$），则其期望工资收入见式 8.46，拟合的期望工资收入见式 8.47：

$$
\begin{aligned}
& E(wage|edu\_pri = 0; edu\_mid = 0; edu\_hig = 0; dpt\_per = 0; sex\_m = 0; age = 30) \\
= & + \beta_1(1) + \beta_2(0) + \beta_3(0) + \beta_4(0) \\
& + \beta_5(0) + \beta_6(0) + \beta_7(0) \cdot (0) + \beta_8(0) \cdot (0) \\
& + \beta_9(0) \cdot (0) + \beta_{10}(30.00) \\
= & \beta_1 + 30.00\beta_{10}
\end{aligned}
\tag{8.46}
$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=0;sex\_m=0;age=30)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)+\hat{\beta}_4(0)\\
&+\hat{\beta}_5(0)+\hat{\beta}_6(0)+\hat{\beta}_7(0)\cdot(0)+\hat{\beta}_8(0)\cdot(0)\\
&+\hat{\beta}_9(0)\cdot(0)+\hat{\beta}_{10}(30.00)\\
=&+[19.43]\cdot(1)-[70.34]\cdot(0)+[32.29]\cdot(0)+[35.59]\cdot(0)\\
&+[777.93]\cdot(0)+[54.46]\cdot(0)-[26.81]\cdot(0)\cdot(0)+[59.80]\cdot(0)\cdot(0)\\
&-[766.28]\cdot(0)\cdot(0)+[3.25]\cdot(30.00)\\
=&116.86
\end{aligned}
\tag{8.47}
$$

（2）比较组 1 为：**文盲 & 长期合同 & 女性**也即 (**illiteracy & permanent & female**)，并给定年龄为 30 岁（$age=30$），则其期望工资收入见式 8.48，拟合的期望工资收入见式 8.49：

$$
\begin{aligned}
&E(wage|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=1;sex\_m=0;age=30)\\
=&+\beta_1(1)+\beta_2(0)+\beta_3(0)+\beta_4(0)\\
&+\beta_5(0)+\beta_6(1)+\beta_7(0)\cdot(1)+\beta_8(0)\cdot(1)\\
&+\beta_9(0)\cdot(1)+\beta_{10}(30.00)\\
=&\beta_1+\beta_6+30.00\beta_{10}
\end{aligned}
\tag{8.48}
$$

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=0;dpt\_per=1;sex\_m=0;age=30)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)+\hat{\beta}_4(0)\\
&+\hat{\beta}_5(0)+\hat{\beta}_6(1)+\hat{\beta}_7(0)\cdot(1)+\hat{\beta}_8(0)\cdot(1)\\
&+\hat{\beta}_9(0)\cdot(1)+\hat{\beta}_{10}(30.00)\\
=&+[19.43]\cdot(1)-[70.34]\cdot(0)+[32.29]\cdot(0)+[35.59]\cdot(0)\\
&+[777.93]\cdot(0)+[54.46]\cdot(1)-[26.81]\cdot(0)\cdot(1)+[59.80]\cdot(0)\cdot(1)\\
&-[766.28]\cdot(0)\cdot(1)+[3.25]\cdot(30.00)\\
=&171.32
\end{aligned}
\tag{8.49}
$$

（3）比较组 2 为：**高等教育 & 长期合同 & 女性**（也即 **higher & permanent & female**），并给定年龄为 30 岁（$age=30$）。则其期望工资收入见式 8.50，拟合的期望工资收入见式 8.51：

$$
\begin{aligned}
&E(wage|edu\_pri=0;edu\_mid=0;edu\_hig=1;dpt\_per=1;sex\_m=0;age=30)\\
=&+\beta_1(1)+\beta_2(0)+\beta_3(0)+\beta_4(0)\\
&+\beta_5(1)+\beta_6(1)+\beta_7(0)\cdot(1)+\beta_8(0)\cdot(1)\\
&+\beta_9(1)\cdot(1)+\beta_{10}(30.00)\\
=&\beta_1+\beta_5+\beta_6+\beta_9+30.00\beta_{10}
\end{aligned}
\tag{8.50}
$$

$$
\begin{aligned}
&(\widehat{wage}|edu\_pri=0;edu\_mid=0;edu\_hig=1;dpt\_per=1;sex\_m=0;age=30)\\
=&+\hat{\beta}_1(1)+\hat{\beta}_2(0)+\hat{\beta}_3(0)+\hat{\beta}_4(0)\\
&+\hat{\beta}_5(1)+\hat{\beta}_6(1)+\hat{\beta}_7(0)\cdot(1)+\hat{\beta}_8(0)\cdot(1)\\
&+\hat{\beta}_9(1)\cdot(1)+\hat{\beta}_{10}(30.00)\\
=&+[19.43]\cdot(1)-[70.34]\cdot(0)+[32.29]\cdot(0)+[35.59]\cdot(0)\\
&+[777.93]\cdot(1)+[54.46]\cdot(1)-[26.81]\cdot(0)\cdot(1)+[59.80]\cdot(0)\cdot(1)\\
&-[766.28]\cdot(1)\cdot(1)+[3.25]\cdot(30.00)\\
=&182.97
\end{aligned}
\tag{8.51}
$$

8.5.5.4 乘法-经典模型（部分）

同时含有虚拟变量和定量变量的、乘法形式的**经典回归（部分）模型**见方程式 8.52

$$
\begin{aligned}
wage_i=&+\beta_1+\beta_2 sex\_m_i+\beta_3 edu\_pri\times dpt\_per_i+\beta_4 edu\_mid\times dpt\_per_i\\
&+\beta_5 edu\_hig\times dpt\_per_i+\beta_6 dpt\_per_i+\beta_7 age_i+u_i
\end{aligned}
\tag{8.52}
$$

同时含虚拟变量和定量变量的、乘法形式的经典回归模型（部分变量进入）Eviews 结果简要报告（见式 8.53）如下：

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

$$
\begin{array}{llll}
\widehat{wage} = & +23.30 & -76.57 sex\_m_i & +29.97 dpt\_per_i \quad +3.88 age_i \\
(s) & (31.0993) & (28.1834) & (28.5419) \quad (0.9127) \\
(t) & (+0.75) & (-2.72) & (+1.05) \quad (+4.25) \\
(cont.) & +5.31 edu\_pri : dpt\_per_i & +92.00 edu\_mid : dpt\_per_i & +13.83 edu\_hig : dpt\_per_i \\
(s) & (58.7653) & (50.8699) & (58.4628) \\
(t) & (+0.09) & (+1.81) & (+0.24) \\
(over) & n = 114 & \hat{\sigma} = 118.2676 & \\
(fit) & R^2 = 0.2945 & \bar{R}^2 = 0.2549 & \\
(Ftest) & F^* = 7.44 & p = 0.0000 &
\end{array}
\tag{8.53}
$$

8.5.5.5 乘法-半对数模型（部分）

同时含有虚拟变量和定量变量的、乘法形式的**半对数回归 (部分) 模型**见方程式 8.54

$$
\begin{aligned}
log(wage)_i = &+ \beta_1 + \beta_2 sex\_m_i + \beta_3 edu\_pri \times dpt\_per_i + \beta_4 edu\_mid \times dpt\_per_i \\
&+ \beta_5 edu\_hig \times dpt\_per_i + \beta_6 dpt\_per_i + \beta_7 age_i + u_i
\end{aligned}
\tag{8.54}
$$

同时含有虚拟变量和定量变量的、乘法形式的**半对数回归模型**见方程式 8.54 ，菜单操作实现具体见图 @ref(fig:both-prod-part-graph：

（1）确定基础组为 [**文盲 & 短期合同 & 女性**]，则如下虚拟变量将**不进入**回归模型 a. ☑`edu_d1` b. ☑`dpt_d1` c. ☑`sex_d1`

（2）设置回归模型。进入引导设置 Equation Estimation ⇒ specification

- Equation specification：输入命令 `log(wage) c dpt_d2 sex_d2 age edu_d2*dpt_d2 edu_d3*dpt_d2 edu_d4*dpt_d2`
- Estimation settings：

  具体设置为：- Method: 下拉选择 LS - Least Squares (NLS and ARMA) - Sample:（默认设置）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 点击完成：OK
- 命名保存方程对象=：（建议命名为 `eq_both_prod_log_part`）
- 查看结果：双击=`eq_both_prod_log_part`

具体 Eviews 报告见 @ref(fig:both-prod-part-log-report)：

同时含虚拟变量和定量变量的、乘法形式的半对数回归模型（部分变量进入）Eviews 结果简要报告（见式 8.55）如下：

$$
\begin{array}{llllll}
\widehat{log(wage)} = & +3.76 & -0.67 sex\_m_i & +0.24 dpt\_per_i & +0.03 age_i \\
(s) & (0.1655) & (0.1500) & (0.1519) & (0.0049) \\
(t) & (+22.73) & (-4.49) & (+1.58) & (+5.79) \\
(cont.) & +0.18 edu\_pri : dpt\_per_i & +0.57 edu\_mid : dpt\_per_i & +0.36 edu\_hig : dpt\_per_i & \\
(s) & (0.3128) & (0.2707) & (0.3112) & \\
(t) & (+0.59) & (+2.10) & (+1.16) & \\
(over) & n = 114 & \hat{\sigma} = 0.6295 & & \\
(fit) & R^2 = 0.4616 & \bar{R}^2 = 0.4314 & & \\
(Ftest) & F^* = 15.29 & p = 0.0000 & &
\end{array}
\tag{8.55}
$$

8.5.6 时间序列季节虚拟变量模型

目的：通过学习时间序列季节虚拟变量模型，进一步深入理解虚拟变量回归模型的多种形式及应用

思路：设置并引入季节虚拟变量，构建时间序列季节虚拟变量模型，Eviews 估计回归结果

定义 8.5 (季节模式和季节调整).

- 季节模式（seasonal pattern）：大多数时间序列经济变量，通常表现出来的季节性往复行为或现象。
- 季节调整（seasonal adjusted）：将时间序列经济变量的季节性变化成分去除，从而得到一个新的变量序列的处理过程。

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

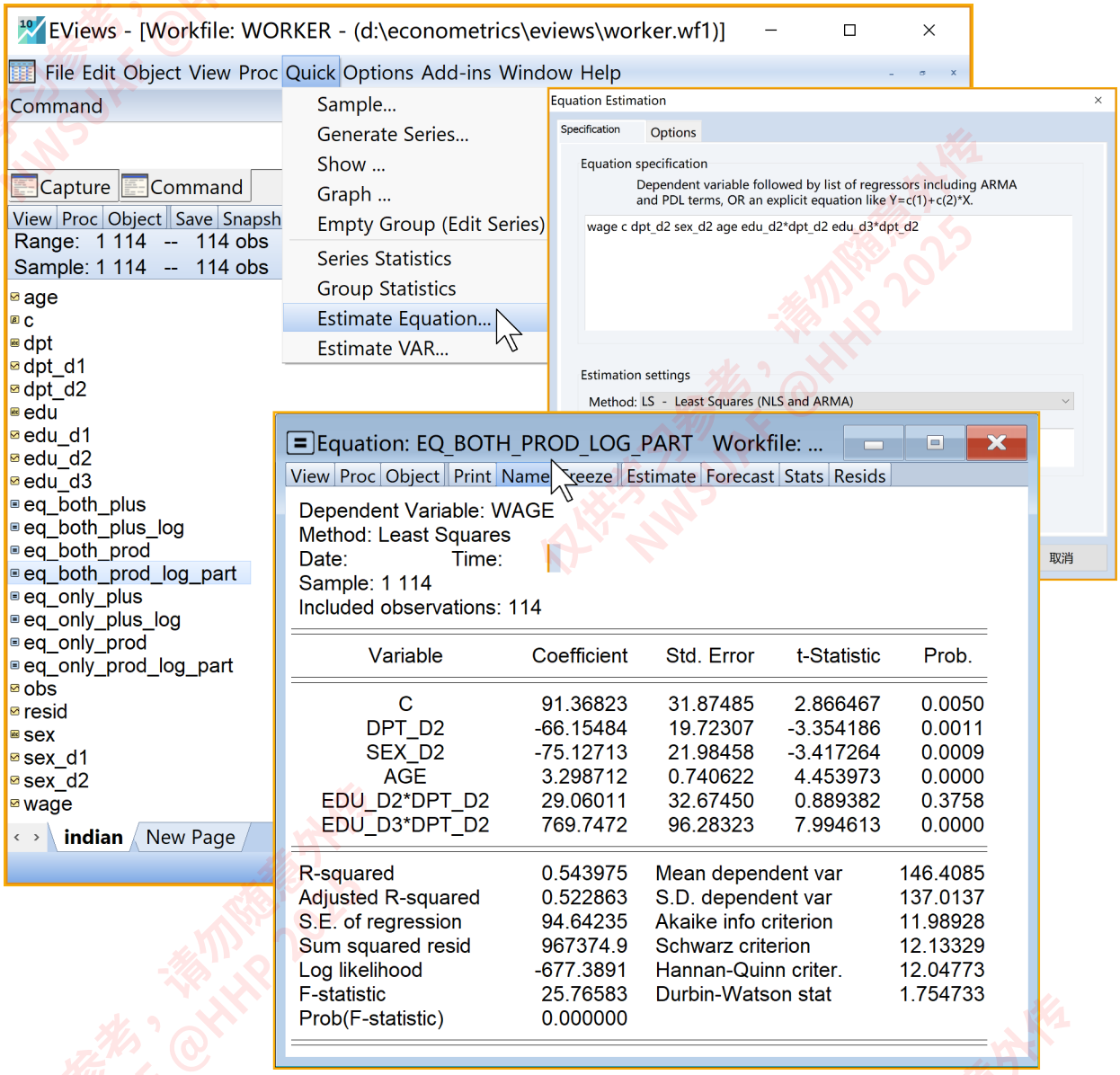


图 8.19: 同时含虚拟变量和定量变量的、乘法形式的半对数线性回归模型 Eviews 实现

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

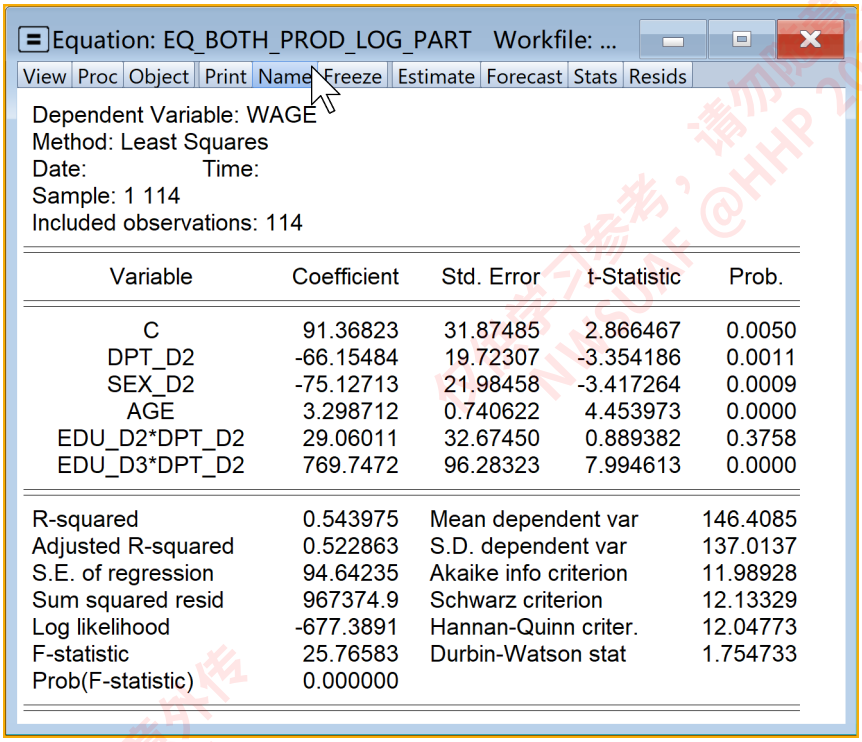


图 8.20: 同时含虚拟变量和定量变量的、乘法形式的半对数线性回归模型 Eviews 报告

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

事实上，一个时间序列经济变量往往同时存在四个成分，分别是季节成分（seasonal component）、周期成分（cyclical component）、趋势成分（trend component）和严格随机成分（strictly random component）。

根据基础组的有无，可以构建以第一季度为基础组的时间序列季节虚拟变量模型式 8.56 和无基础组的时间序列季节虚拟变量模型式 8.57 。其中，X_t 为定量变量。

$$Y_t = \beta_1 + \beta_2 X_t + \lambda_2 D2_t + \lambda_3 D3_t + \lambda_4 D4_t + u_t \tag{8.56}$$

$$Y_t = \beta_2 X_t + \lambda_1 D1_t + \lambda_2 D2_t + \lambda_3 D3_t + \lambda_4 D4_t + u_t \tag{8.57}$$

Eviews 操作（此处略）

8.5.7 分段线性回归模型

目的：通过学习分段式线性回归模型，进一步深入理解虚拟变量回归模型的多种形式及应用

思路：设置并引入虚拟变量，构建分段线性回归模型，Eviews 估计回归结果

定义 8.6 (分段线性回归).

- **分段**现象：在经济关系中，当解释变量 X 的值达到某一水平/阀值 X^* 之前，与被解释变量之间存在某种线性关系；当解释变量 X 的值达到或者超过水平/阀值 X^* 以后，与被解释变量的关系就会发生变化。因而总体看来，似乎被明显“分段”了。
- 分段线性回归模型（piecewise linear regression）：是指用虚拟变量估计不同水平/阀值的解释变量 X 对被解释变量 Y 的影响的一类线性回归模型。

一个阀值的分段线性回归模型式 8.58

$$Y_i = \beta_1 + \beta_2 X_i + \lambda(X_i - X^*)D_i + u_i \tag{8.58}$$

两个阀值的分段线性回归模型式 8.59

$$Y_i = \beta_1 + \beta_2 X_i + \lambda_1(X_i - X_1^*)D1_i + \lambda_2(X_i - X_2^*)D2_i + u_i \tag{8.59}$$

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

Eviews 操作（此处略）

8.6 作业题

印度工人工资：表 8.8 给出给出了 114 位印度工人在 wage 工人工资，age 年龄，edu 教育水平，dpt 合同类型，sex 性别等方面的数据。

表 8.8: 印度工人工资 (n=114)

obs	wage	age	edu	dpt	sex
1	117	26	primary	permanent	female
2	375	42	primary	permanent	female
3	175	33	primary	permanent	female
4	100	33	primary	permanent	female
5	162	30	primary	permanent	female
110	25	18	illiteracy	temporary	male
111	25	11	illiteracy	temporary	male
112	75	45	illiteracy	temporary	male
113	54	14	illiteracy	temporary	male
114	50	26	illiteracy	temporary	male

变量说明见表 8.9 ：

表 8.9: 变量定义及说明

variable	label	remark
obs	工人编号	序号
wage	工人工资	美元/周
age	年龄	岁
edu	教育水平	illiteracy= 文盲；primary= 初等教育；secondary= 中等教育；higher= 高等教育
dpt	合同类型	temporary= 短期合同；permanent= 长期合同
sex	性别	female= 女；male= 男

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

请回答如下问题:

（1）请对数据表 8.8 进行数据处理，将定性变量都处理成**完全**虚拟变量体系（也即定性变量有 m 个属性，则设置 m 个虚拟变量）。

要求：定性变量 edu 的虚拟变量设置为 edu_D1,edu_D2,…；定性变量 dpt 的虚拟变量设置为 dpt_D1,…；定性变量 sex 的虚拟变量设置为 sex_D1,…。虚拟变量的下标请与变量说明表 8.9 的保持一致。

（2）使用**全部变量**，并提出一个**加法形式**的有截距虚拟变量回归模型，用以预测工人工资。要求基础组设置为（**文盲 & 短期合同 & 女性**）的工人组。

a. 请写出加法形式的有截距虚拟变量回归模型（PRM）

b. 请利用你自己构建的模型，进行 eviews 分析，得到分析报告。

要求：将报告截图过来，并写出相应的简要报告形式——三行式或四行式）

c. 根据回归分析结果，请预测如下几类工人的工资水平：

（**文盲&短期合同&女性&50岁**）的工人
（**高等教育&长期合同&男性&50岁**）的工人
（**初等教育&短期合同&男性&30岁**）的工人
（**初等教育&长期合同&女性&30岁**）的工人

要求：分别写出理论表达式，以及数据分析的预测值）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

（3）使用**全部变量**，并提出一个**乘法形式**的有截距虚拟变量回归模型，用以预测工人工资。要求基础组设置为（**文盲 & 短期合同 & 女性**）的工人组，且要求教育程度 edu 的虚拟变量（edu_D1,edu_D2,…）与合同类型 dpt 的虚拟变量（dpt_D1,…）进行乘法交互。

a. 请写出该乘法形式的有截距虚拟变量回归模型（PRM）

b. 请利用你自己构建的模型，进行 eviews 分析，得到分析报告。

要求：将报告截图过来，并写出相应的简要报告形式——三行式或四行式）

c. 根据回归分析结果，请预测如下几类工人的工资水平：

- （**文盲&短期合同&女性&50岁**）的工人
- （**高等教育&长期合同&男性&50岁**）的工人
- （**初等教育&短期合同&男性&30岁**）的工人
- （**初等教育&长期合同&女性&30岁**）的工人

要求：分别写出理论表达式，以及数据分析的预测值）

（4）使用**全部变量**，对因变量 ln(wage) 提出一个**乘法形式**的有截距虚拟变量回归模型，用以预测工人工资。要求基础组设置为（**文盲 & 短期合同 & 女性**）的工人组，且要求性别 sex 的虚拟变量（sex_D1,…）与合同类型 dpt 的虚拟变量（dpt_D1,…）进行乘法交互。

a. 请写出该乘法形式的有截距虚拟变量回归模型（PRM）

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

b. 请利用你自己构建的模型，进行 eviews 分析，得到分析报告，并回答如下问题。

- 你认为这些新引入的交互项看上去有显著的交互影响吗?
- 男性工人与女性工人相比，收入存在明显差异吗？请给出你的理由。
- 长期合同的男性工人与长期合同的女性工人，收入存在明显差异吗？请给出你的理由。
- 短期合同的男性工人与长期合同的女性工人，收入存在明显差异吗？请给出你的理由。

要求：将报告截图过来，并写出相应的简要报告形式——三行式或四行式

c. 根据回归分析结果，请预测如下几类工人的工资水平：（要求分别写出理论表达式，以及数据分析的预测值）

- （**文盲 & 短期合同 & 女性 &50 岁**）的工人
- （**高等教育 & 长期合同 & 男性 &50 岁**）的工人
- （**初等教育 & 短期合同 & 男性 &30 岁**）的工人
- （**初等教育 & 长期合同 & 女性 &30 岁**）的工人

5. 使用**除教育 edu 变量之外**的其他全部变量，对因变量 ln(wage) 提出一个**乘法形式**的有截距虚拟变量回归模型，用以预测工人工资。要求基础组设置为（**文盲 & 短期合同 & 女性**）的工人组, 且要求教育程度 edu 的虚拟变量（edu_D1,edu_D2,···）与合同类型 dpt 的虚拟变量（dpt_D1,···）进行乘法交互。

a. 请写出该乘法形式的有截距虚拟变量回归模型（PRM）

- 你认为这些新引入的交互项看上去有显著的交互影响吗?
- 男性工人与女性工人相比，收入存在明显差异吗？请给出你的理由。
- 长期合同的男性工人与长期合同的女性工人，收入存在明显差异吗？请给出你的理由。
- 短期合同的男性工人与长期合同的女性工人，收入存在明显差异吗？请给出你的理由。

b. 请利用你自己构建的模型，进行 eviews 分析，得到分析报告，并回答如下问题。（要求将报告截图过来，并写出相应的简要报告形式——三行式或四行式）

- 你认为这些新引入的交互项看上去有显著的交互影响吗?

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

- 高等数育的工人与初等教育的工人，收入存在明显差异吗？请给出你的理由。
- [高等教育的长期合同工人] 与 [高等教育的短期合同工人]，收入存在明显差异吗？请给出你的理由。
- [高等教育的短期合同工人] 与 [初等教育的长期合同工人]，收入存在明显差异吗？请给出你的理由。

c. 根据回归分析结果，请预测如下几类工人的工资水平：

- （**文盲 & 短期合同 & 女性 &50 岁**）的工人
- （**高等教育 & 长期合同 & 男性 &50 岁**）的工人
- （**初等教育 & 短期合同 & 男性 &30 岁**）的工人
- （**初等教育 & 长期合同 & 女性 &30 岁**）的工人

要求：分别写出理论表达式，以及数据分析的预测值

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

实验课程指导

结合实验指导书，制作实验指导幻灯片。每章实验指导书 qmd 文件对应一个实验指导幻灯片 qmd 文件。

- 实验 1：EViews 入门。点击查看幻灯片 html 版本
- 实验 2：简单回归 `slide02-simple-regression.qmd`
- 实验 3：模型函数 `slide03-model-function.qmd`
- 实验 4：多元回归 `slide04-multiple-regression.qmd`
- 实验 5：多重共线性 `slide05-multilinearity.qmd`
- 实验 6：异方差性 `slide06-heteroskedasticity.qmd`
- 实验 7：自相关性 `slide07-autocorelation.qmd`
- 实验 8：虚拟变量模型 `slide08-dummy-model.qmd`

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025

参考文献

Hadley, Wickham. 2016. *Elegant Graphics for Data analysis (2nd)*. Use R! Springer.

IHS. 2017. *EViews 10 Command and Programming Reference*. Book. IHS Global Inc.

仅供学习参考，请勿随意外传 NWSUAF @HHP 2025