

计量经济学 Eviews 实验指导书

Lab 5 多重共线性的诊断和矫正

胡华平

2018/3/27

目录

1 实验目的及要求	3
2 实验原理	4
2.1 多重共线性诊断的经验方法:	4
2.2 多重共线性修正方法:	4
3 实验内容	5
3.1 实验方案设计	5
3.2 实验背景——郎利数据 (class data)	6
4 主要实验步骤	9
4.1 导入数据并进行预处理	9
4.2 采用最小二乘法建立主回归模型	9
4.3 借查模型是否存在多重共线性:	10
4.3.1 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。	10
4.3.2 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	10
4.3.3 辅助回归分析法:	15
4.3.4 回归系数方差分解法 (Coefficient Variance Decomposition) (选学)	24
4.4 对存在多重共线性的模型进行修正	27
4.4.1 简单剔除变量法	27
4.4.2 主成分法 (Principal Components) (自学)	34

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。

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。

2.1 多重共线性诊断的经验方法：

- a. 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）
- b. 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）
 - 散点图矩阵发现高度线性相关的数据分布模式
- c. 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
- d. 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

2.2 多重共线性修正方法：

- a. 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
- b. 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)； $p \in [0.05, 0.01]$ (非常显著)； $p \leq 0.01$ (极其显著)
 - t^* 值判别法：2t 法则
- c. 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
- d. 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
- e. 拯救多重共线性的其他方法
 - 脊回归 (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

3 实验内容

3.1 实验方案设计

1. 采用最小二乘法建立主回归模型
2. 借查模型是否存在多重共线性：
 - a. 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）
 - b. 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）
 - 散点图矩阵发现高度线性相关的数据分布模式
 - c. 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
 - d. 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：
 - a. 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
 - b. 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
 - c. 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
 - d. 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
 - e. 拯救多重共线性的其他方法
 - 脊回归 (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

表 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
1962	70551	1169	554894	4007	2827	130081	16

3.2 实验背景——郎利数据 (class data)

就业情况的郎利数据：表1给出美国 1947-1961 年间就业情况及主要影响因素的数据表。

变量说明见表2：

请考虑如下样本回归模型：

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \hat{\beta}_3 X_{3t} + \hat{\beta}_4 X_{4t} + \hat{\beta}_5 X_{5t} + \hat{\beta}_6 X_{6t} + e_t \quad (1)$$

表 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

请回答如下问题：

- 根据回归模型(1)，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：

- 利用 Eviews 对样本回归模型(1)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：

- 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）：

- 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：

- 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：

- 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：

- 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 3：主回归和辅助回归的判定系数计算

回归类型	模型设置	判定系数 $R^2 =$
主回归	Y 对全部 X_i	
辅助回归	X_1 对其他全部 X_i	
辅助回归	X_2 对其他全部 X_i	
辅助回归	X_3 对其他全部 X_i	
辅助回归	X_4 对其他全部 X_i	
辅助回归	X_5 对其他全部 X_i	
辅助回归	X_6 对其他全部 X_i	

- 请按下列要求完成下表和提问：

- 利用上表结果，计算方差膨胀因子和容忍度指标；
- 利用 Eviews 菜单，对主模型(1)操作，得到方差膨胀因子（截图并列出结果到下表中）
- 根据上述结果，请你得出关于多重共线性的初步结论。

答：

表 4：方差膨胀因子 VIF 和容忍度 TOL 的计算与比较

辅助模型设置	容忍度 TOL_j	辅助模型 VIF_j	主模型 VIF
X ₁ 对其他全部 X_i			
X ₂ 对其他全部 X_i			
X ₃ 对其他全部 X_i			
X ₄ 对其他全部 X_i			
X ₅ 对其他全部 X_i			
X ₆ 对其他全部 X_i			

f. 按下列要求计算病态数 (K) 和病态指数 (CI)：

- 利用 Eviews 对 7 个自变量进行主成分 (principal components) 分析 (将结果截图过来)
- 得到特征值 (eigenvalues) 向量 (将结果截图过来)
- 计算病态数 (K) 和病态指数 (CI) (写出计算公式和计算结果)
- 根据上述结果, 请你得出关于多重共线性的初步结论。

答：

3. 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：

- b. 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：

- c. 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!!】

4 主要实验步骤

4.1 导入数据并进行预处理

- 目标:
- 思路:
- 新建 Eviews 工作文件 (workfile)
 - 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
 - Eviews 菜单操作:
 - a. 依次操作: File⇒New⇒Workfile
 - b. 进行 workfile create 引导设置:
 - * workfile structure type: unstructured/undated
 - * data range: 16
 - * workfile names(optional):
 - WF: longley (建议命名)
 - Page: employee (建议命名)
 - 导入数据
 - 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)
 - 菜单操作 (Excel 和 Eviews):
 - a. Excel 找到数据。Excel 表格中仅保留自己需要的数据 (Year, Y, X1, X2, X3, X4, X5, X6)
 - b. Eviews 导入数据。File⇒Import⇒Import From File:d:/econometrics/data/Lab5-longley

4.2 采用最小二乘法建立主回归模型

- 目标:
- 思路:
- 提示: 主回归模型为

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \hat{\beta}_3 X_{3t} + \hat{\beta}_4 X_{4t} + \hat{\beta}_5 X_{5t} + \hat{\beta}_6 X_{6t} + e_t$$

- Eviews 菜单操作:
 - 1) 依次选择 ⇒Quick⇒Estimation Equation
 - 2) 引导设置 Equation Estimation⇒specification
 - a. Equation specification: 输入命令 Y c X1 X2 X3 X4 X5 X6
 - 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

表 5: t 检验不显著的斜率系数

	t value	Pr(> t)
X1	0.1774	0.8631
X2	-1.0695	0.3127
X5	-0.2261	0.8262

主回归分析结果见图1:

4.3 借查模型是否存在多重共线性:

4.3.1 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。

- 目标：观察 t^* 检验，判定系数 R^2 , F^* 检验的关系
- 思路：
- 提示：
 - 主回归分析报告的 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）
- 分析结论：根据主回归报告（见2），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
 - $F^* = 330.2853$, 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。

4.3.2 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

- 目标：观察 t^* 检验，判定系数 R^2 , F^* 检验的关系
- 思路：
- 提示：
 - 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) :
 - a. 依次选择 X 变量: X1, X2, X3, X4, X5, X6
 - b. 以组对象打开，鼠标右键: Open as group
 - c. 命名并保存组对象: name (建议命名为 group_x)
 - 2) 绘制散点图矩阵
 - a. 进入 group 视窗：双击打开组对象 group_x
 - b. 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
 - c. 点击完成：OK

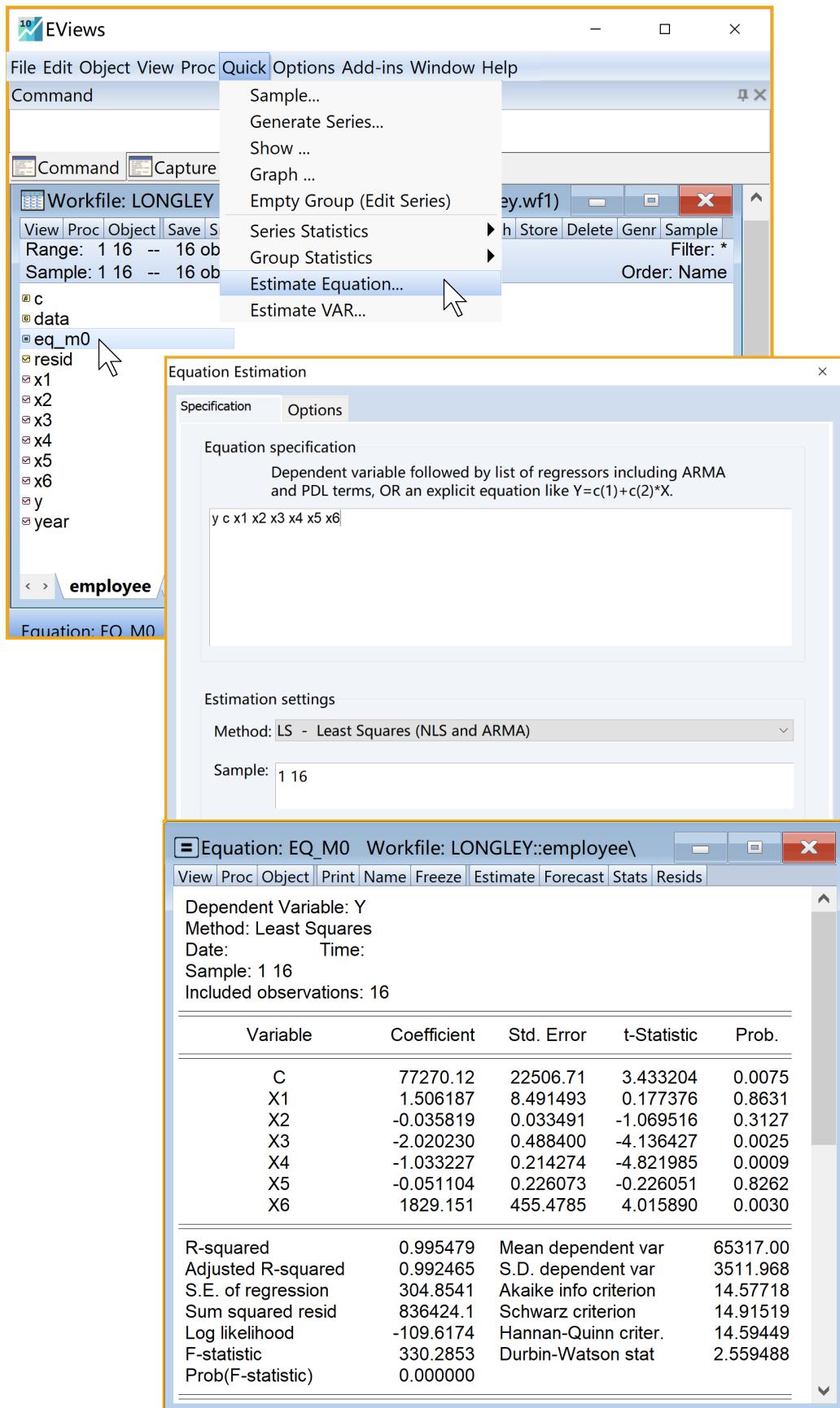


图 1: 主回归模型 Eviews 建模操作

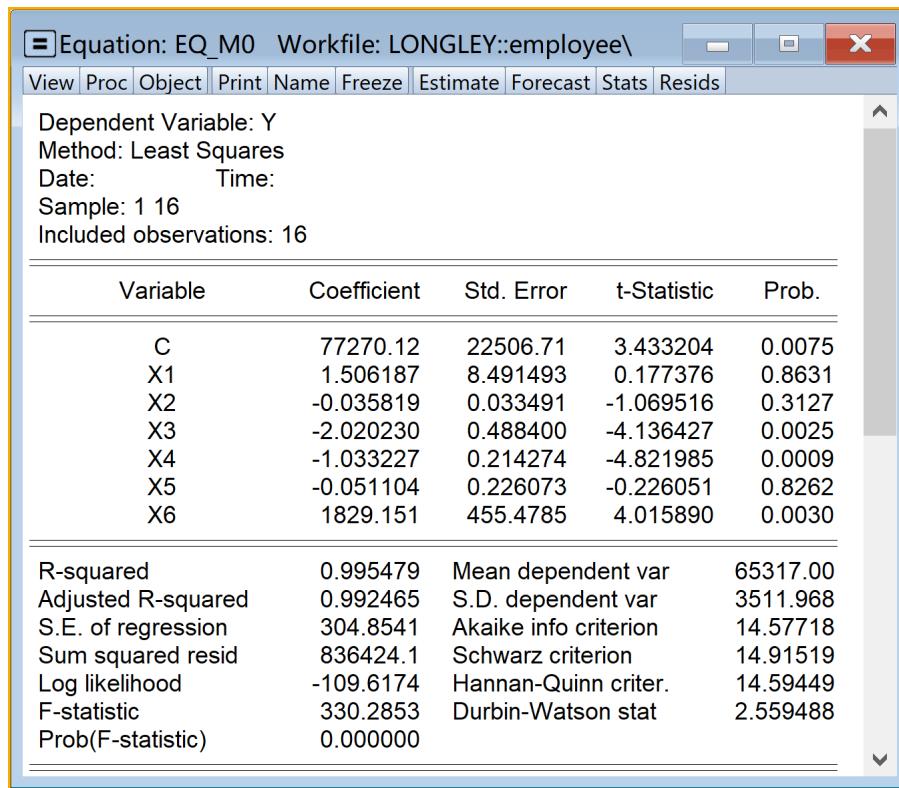


图 2: 主回归模型 Eviews 报告

- d. 命名并保存绘图 (graph) 对象 : (建议命名为 scatter_matrix)
 - e. 查看结果: 双击 scatter_matrix (见图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: \Rightarrow View \Rightarrow 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 tab_cor
 - 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4)
- 分析结论根据相关系数矩阵表 (图 @ref:(fig:fig-cor) 和散点矩阵图 (图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 X2: $r_{2,1}=0.9916$

X1 VS X5: $r_{5,1}=0.9792$

X1 VS X6: $r_{6,1}=0.9911$

The screenshot displays the EViews interface with several windows open:

- Command Window:** Shows the command history and current workfile details: "Workfile: LONGLEY - (d:\econometrics\ewviews\longley.wf1)".
- Object Explorer:** Lists project objects including data series (C, data, eq_m0, group_x, resid, scatter_matrix, tab_cor), variables (x1-x6, y), and a group named "employee".
- Covariance Analysis Dialog:** A modal dialog for "Covariance Analysis" with "Method: Ordinary". It includes sections for Statistics (Correlation checked), Partial analysis (Series or groups for conditioning optional), Options (Weighting: None), and Sample (1 16). The "Balanced sample (listwise deletion)" checkbox is selected.
- Correlation Matrix:** A spreadsheet titled "Correlation" showing the correlation coefficients between variables X1 through X6. The matrix is as follows:

	X1	X2	X3	X4	X5	X6
X1	1.000000	0.991589	0.620633	0.464744	0.979163	0.991149
X2	0.991589	1.000000	0.604261	0.446437	0.991090	0.995273
X3	0.620633	0.604261	1.000000	-0.177421	0.686552	0.668257
X4	0.464744	0.446437	-0.177421	1.000000	0.364416	0.417245
X5	0.979163	0.991090	0.686552	0.364416	1.000000	0.993953
X6	0.991149	0.995273	0.668257	0.417245	0.993953	1.000000

- Table View:** A spreadsheet titled "Table: TAB_COR" showing the correlation matrix for variables A through G. The matrix is as follows:

	A	B	C	D	E	F	G
1		X1	X2	X3	X4	X5	X6
2							
3	X1	1.000000	0.991589	0.620633	0.464744	0.979163	0.991149
4	X2	0.991589	1.000000	0.604261	0.446437	0.991090	0.995273
5	X3	0.620633	0.604261	1.000000	-0.177421	0.686552	0.668257
6	X4	0.464744	0.446437	-0.177421	1.000000	0.364416	0.417245
7	X5	0.979163	0.991090	0.686552	0.364416	1.000000	0.993953
8	X6	0.991149	0.995273	0.668257	0.417245	0.993953	1.000000

图 3: 自变量 X 间相关系数矩阵的 Eviews 操作

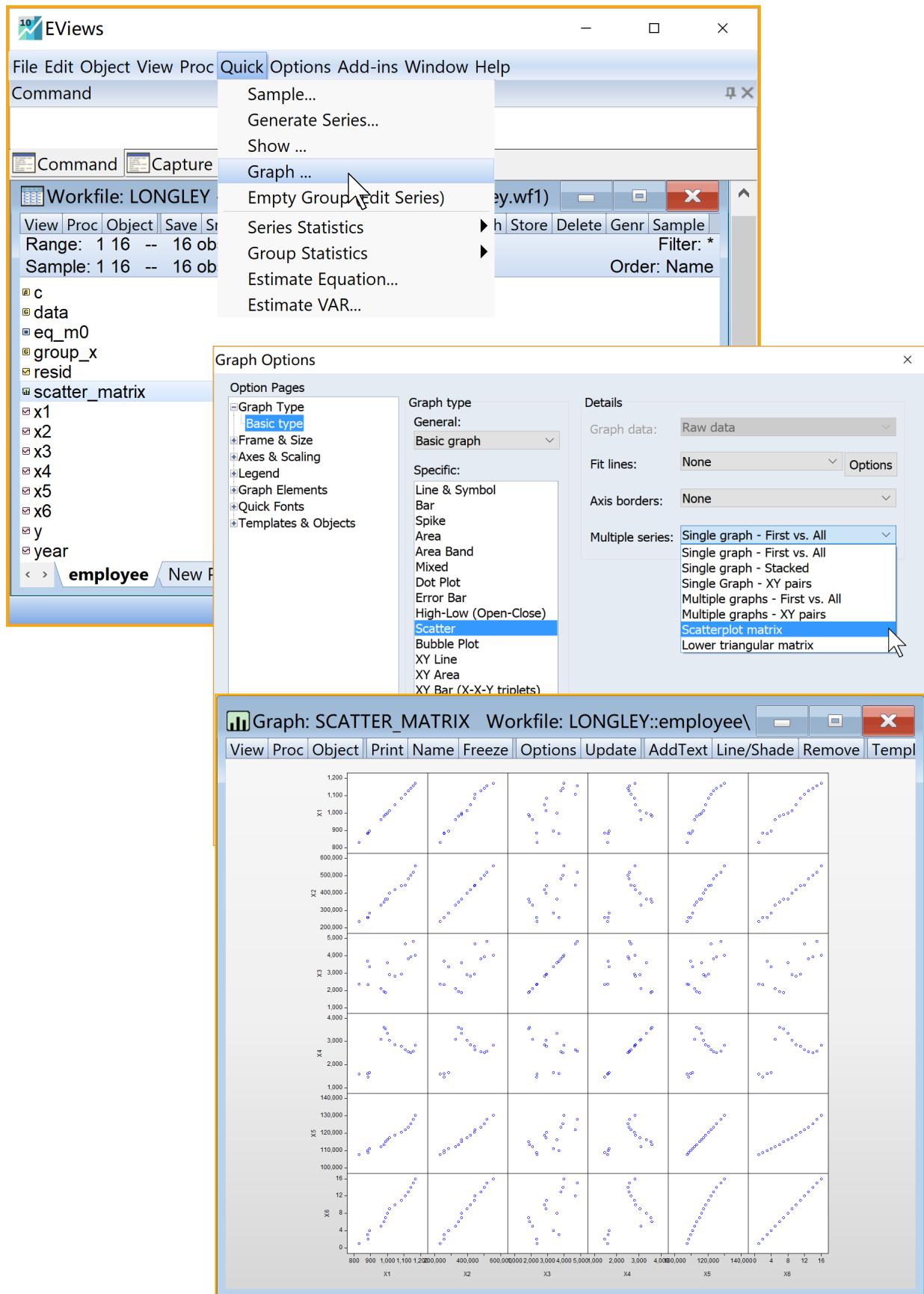


图 4: 自变量 X 间散点矩阵图的 Eviews 操作

X2 VS X5: $r_{5,2}=0.9911$

X2 VS X6: $r_{6,2}=0.9953$

X5 VS X6: $r_{6,5}=0.994$

4.3.3 辅助回归分析法:

- 定义:

- 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。

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \hat{\beta}_3 X_{3t} + \hat{\beta}_4 X_{4t} + \hat{\beta}_5 X_{5t} + \hat{\beta}_6 X_{6t} + e_t \quad (\text{Mo}) \quad (2)$$

$$X_{1t} = \hat{\alpha}_0 + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_5 X_{5t} + \hat{\alpha}_6 X_{6t} + e_t \quad (\text{A1}) \quad (3)$$

$$X_{2t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_5 X_{5t} + \hat{\alpha}_6 X_{6t} + e_t \quad (\text{A2}) \quad (4)$$

$$X_{3t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_5 X_{5t} + \hat{\alpha}_6 X_{6t} + e_t \quad (\text{A3}) \quad (5)$$

$$X_{4t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_5 X_{5t} + \hat{\alpha}_6 X_{6t} + e_t \quad (\text{A4}) \quad (6)$$

$$X_{5t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_6 X_{6t} + e_t \quad (\text{A5}) \quad (7)$$

$$X_{6t} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{4t} + \hat{\alpha}_5 X_{5t} + e_t \quad (\text{A6}) \quad (8)$$

4.3.3.1 构建 6 个辅助回归方程

- 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。
- 提示: 此处仅以 A1 (3) 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):
 - 1) 依次选择 \Rightarrow Quick \Rightarrow Estimation Equation

- 2) 引导设置 Equation Estimation \Rightarrow specification

- a. Equation specification: 输入命令 X1 c X2 X3 X4 X5 X6

- b. Estimation settings:

- Method: 下拉选择 LS - Least Squares (NLS and ARMA)

- Sample: (默认设置)

- c. 点击完成: OK

- d. 命名保存方程对象 \blacksquare : (建议命名为 eq_a1, 其他辅助方程依次命名为 eq_a2, ..., eq_a6)

- e. 查看结果: 双击 \blacksquare eq_a1

全部 6 个辅助回归方程的 Eviews 分析报告结果: A1 辅助模型见图 6; A2 辅助模型见图 7; A3 辅助模型见图 8; A4 辅助模型见图 9; A5 辅助模型见图 10; A6 辅助模型见图 11;

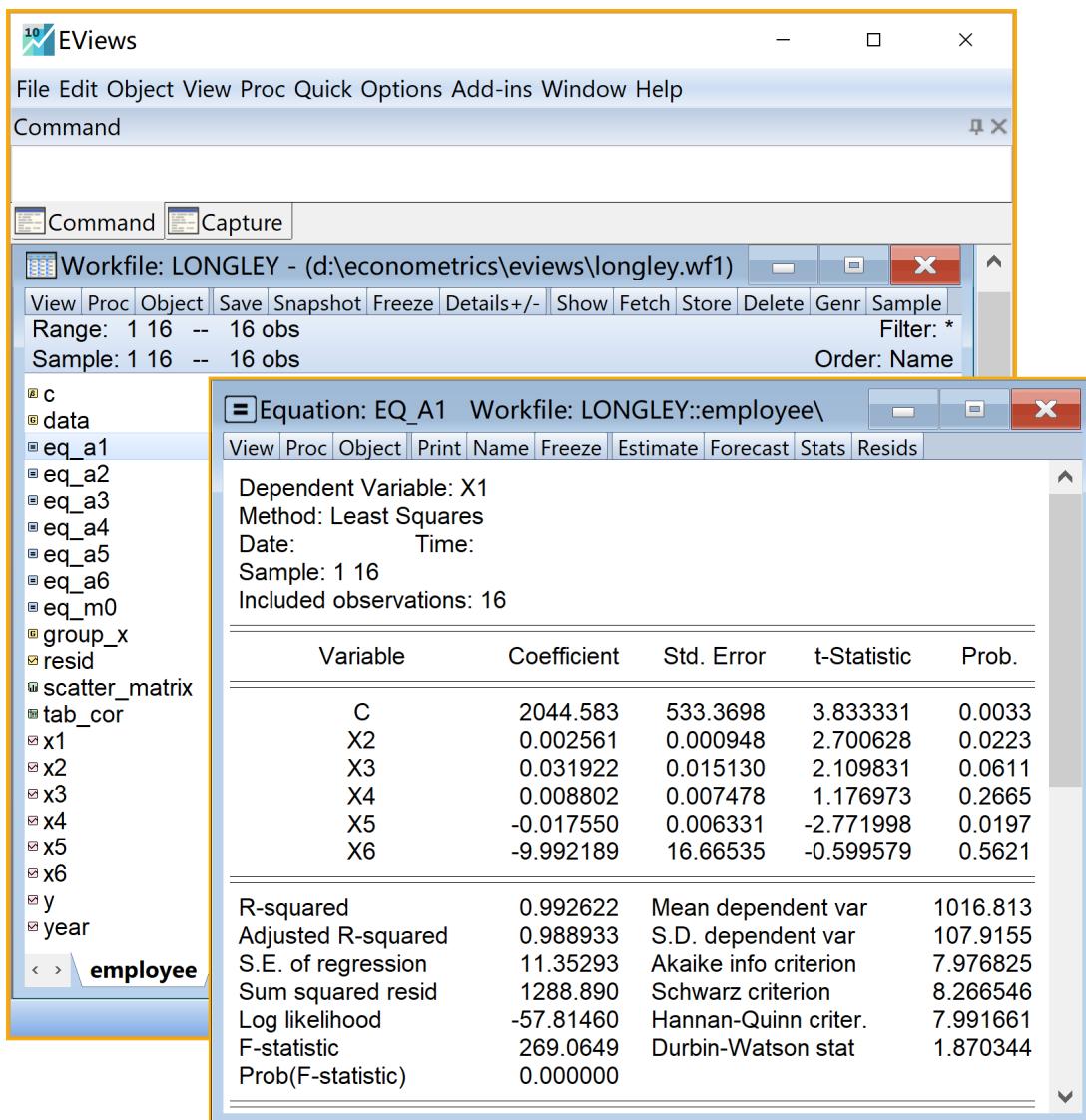


图 5: 辅助回归 Eviews 操作过程

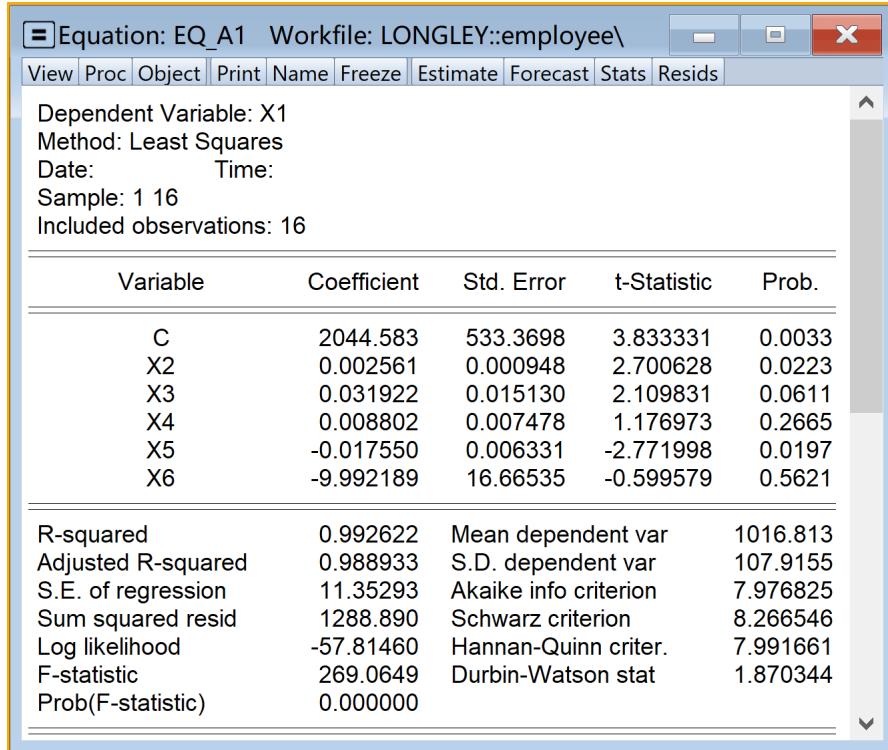


图 6: 辅助回归模型 A1 的 Eviews 报告

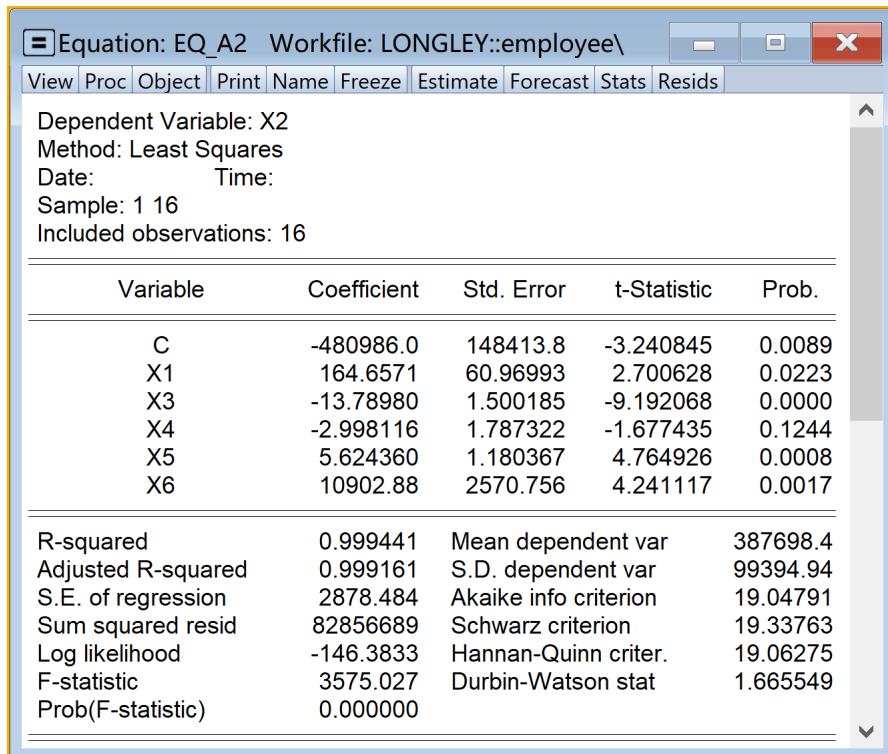


图 7: 辅助回归模型 A2 的 Eviews 报告

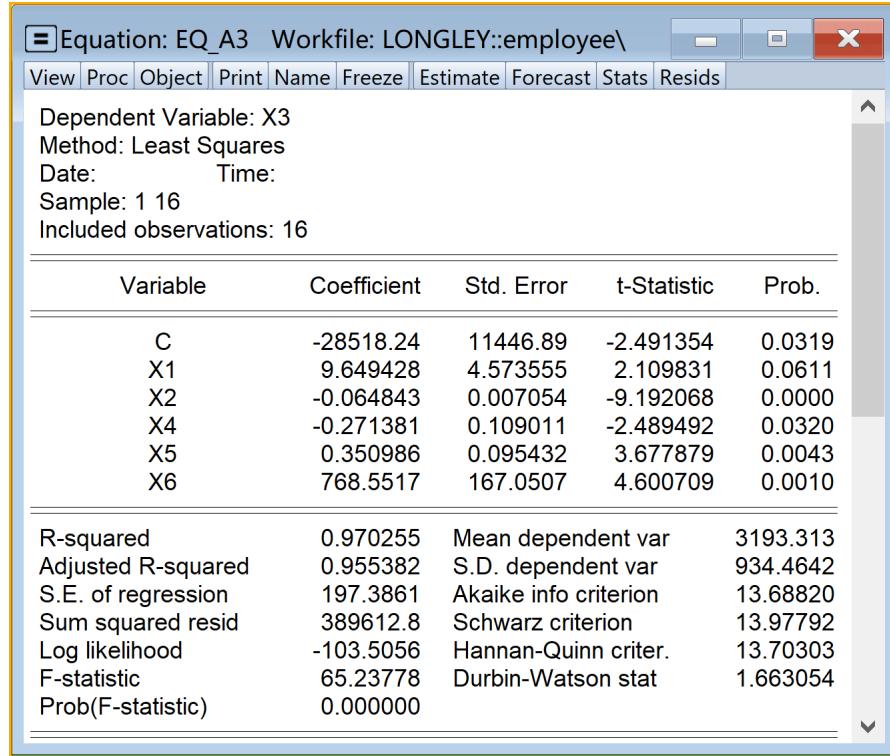


图 8: 辅助回归模型 A3 的 Eviews 报告

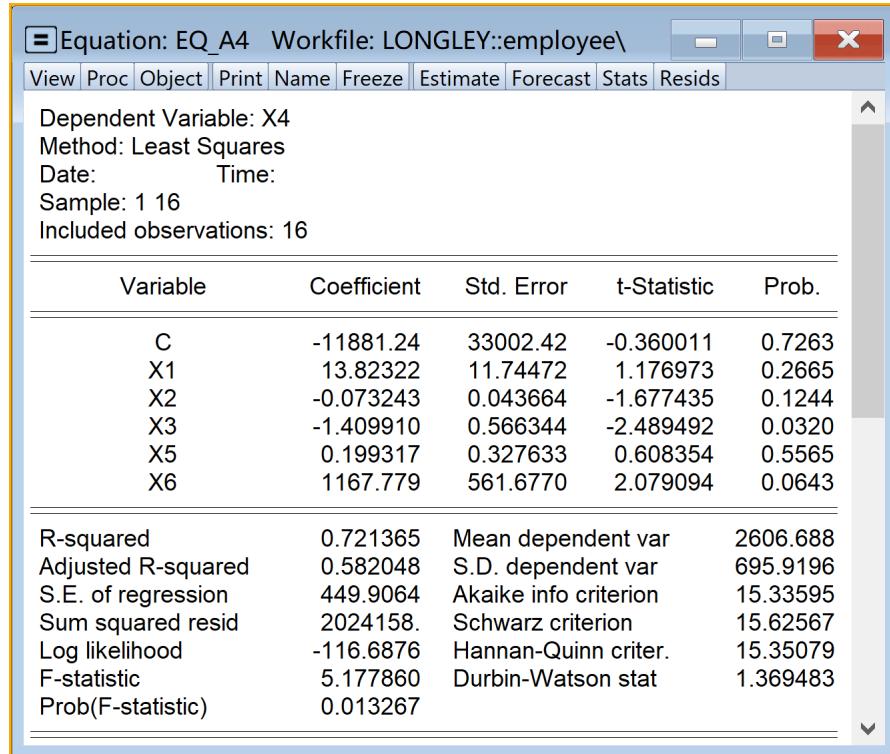


图 9: 辅助回归模型 A4 的 Eviews 报告

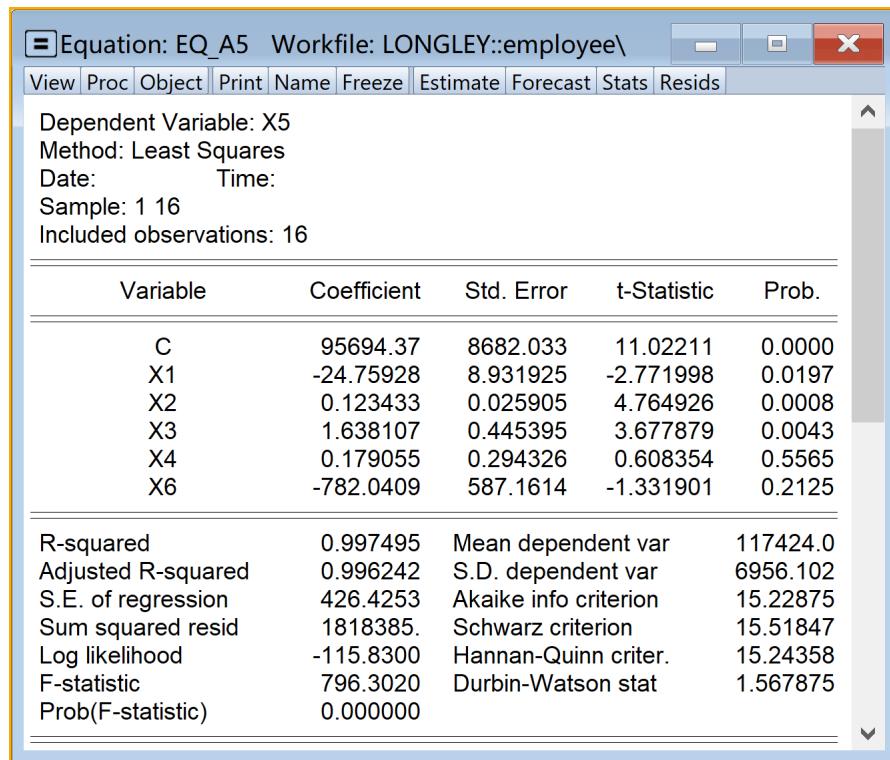


图 10: 辅助回归模型 A5 的 Eviews 报告

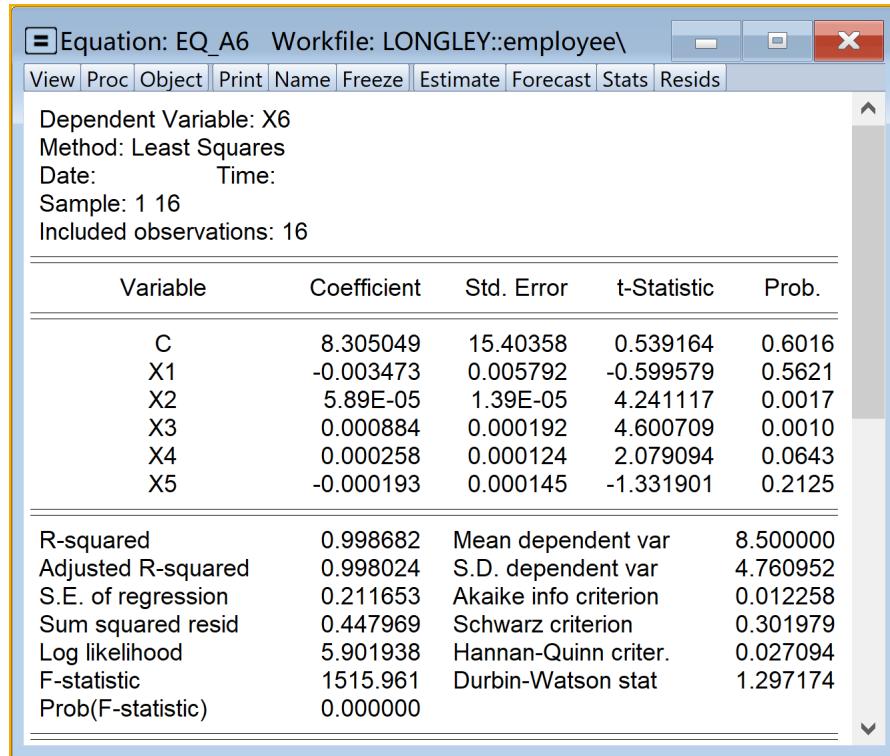


图 11: 辅助回归模型 A6 的 Eviews 报告

4.3.3.2 判定系数比较法

- 目标：分别 6 个辅助回归方程的判定系数 $R_j^2(j = 1, \dots, 6)$ ，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。
- 提示：
 - 操作提示：构造一个列向量（vector）对象 r2 （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。
 - 诊断提示：辅助回归方程的判定系数 $R_j^2(j = 1, \dots, 6)$ 大于主回归方程的判定系数 R^2
- Eviews 菜单操作（见图12）：
 - 1) 构建一个列向量（vector）对象 r2 （含 6 个元素），建议命名为 $r2$
 - a. 命令视窗（Command）输入命令：vector(6) r2
 - b. 运行命令：命令行中按 Enter 键
 - 2) 提取 6 个辅助回归方程的判定系数，并依次放置到列向量（vector）对象 r2 中。在命令视窗（Command）中依次输入并运行下列命令：
 - a. 提取并放置辅助模型 A1 方程(3)的 R_1^2 : r2.fill(o=1) eq_a1.@r2
 - b. 提取并放置辅助模型 A2 方程(4)的 R_2^2 : r2.fill(o=2) eq_a2.@r2
 - c. 提取并放置辅助模型 A3 方程(5)的 R_3^2 : r2.fill(o=3) eq_a3.@r2
 - d. 提取并放置辅助模型 A4 方程(6)的 R_4^2 : r2.fill(o=4) eq_a4.@r2
 - e. 提取并放置辅助模型 A5 方程(7)的 R_5^2 : r2.fill(o=5) eq_a5.@r2
 - f. 提取并放置辅助模型 A6 方程(8)的 R_6^2 : r2.fill(o=6) eq_a6.@r2
 - g. 查看结果：双击 r2
 - 3) 说明（Eviews 代码行的解读¹）：
 - a. 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)
 - b. 代码 $\text{eq_a1}@r2$ 表示提取方程（equation）对象 eq_a1 的判定系数 R^2 (.@r2)

根据上述比较分析，判定系数比较法的结论初步认为模型(2)可能存在严重的多重共线性问题理由如下：

表 6: 主回归和辅助回归的判定系数 (分析结果)

回归类型	模型设置	判定系数 $R^2 =$
辅助回归	X1 对其他全部 X_i	0.9926
辅助回归	X2 对其他全部 X_i	0.9994
辅助回归	X3 对其他全部 X_i	0.9703
辅助回归	X4 对其他全部 X_i	0.7214
辅助回归	X5 对其他全部 X_i	0.9975
辅助回归	X6 其他全部 X_i	0.9987

¹具体细节请参看 Eviews 在线帮助文档，网址 http://www.eviews.com/help/helpintro.html#page/content/Regress1-Equation_Output.html

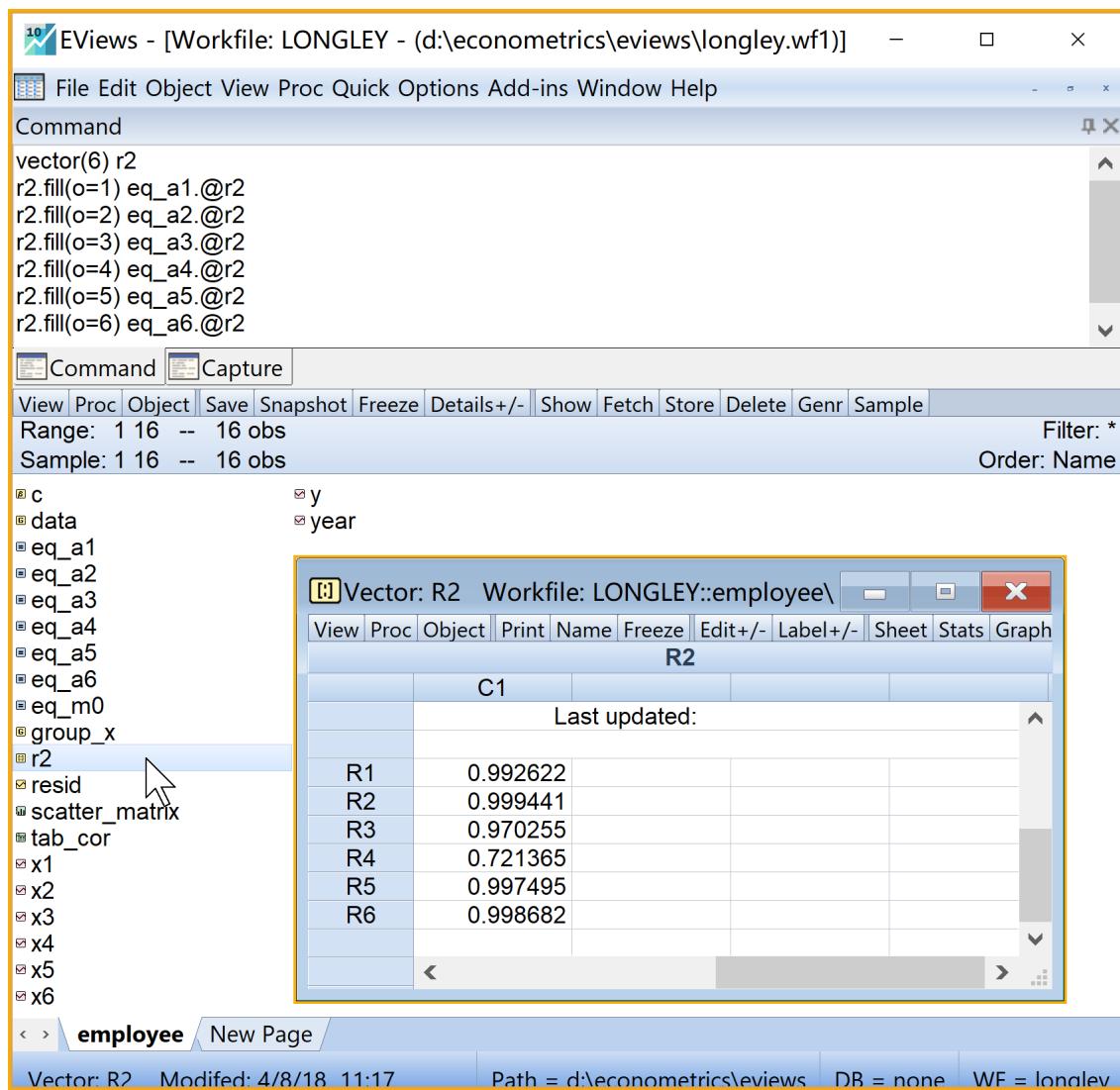


图 12: 辅助回归模型 A1 A6 的判定系数提取汇总操作

4.3.3.3 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 $VIF_j(j = 1, \dots, 6)$ ，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6），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 \Rightarrow Variance Inflation Factors 一次性得到主回归模型的 Eviews 方程 \Rightarrow eq_mo 全部方差膨胀因子 $VIF_j(j = 1, \dots, 6)$
 - 提示：
 - 操作提示：构造一个列向量（vector）对象 unit （含 6 个元素），用来装载 6 个辅助回归方程的方差膨胀因子 $VIF_j(j = 1, \dots, 6)$ 。
 - 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
- $$VIF_j = \frac{1}{1 - R_j^2}, (j = 1, \dots, 6) \quad (9)$$
- (方法 1) Eviews 操作（手动计算实现，具体见图13）：
 - 1) 计算准备。创建一个常数列向量（vector）对象 unit （含 6 个元素），使其元素全部为 1（建议将该常数列向量命名为 unit）
 - a. 命令视窗（Command）输入命令：vector(6) unit=1
 - b. 运行命令：命令行中按 Enter 键
 - c. 查看结果：双击 unit
 - 2) 计算 VIF_j 。根据理论计算公式(9)，对常数列向量 unit 和判定系数向量 r2 进行运算，得到方差膨胀因子向量 vif （建议将方差膨胀因子向量命名为 vif）
 - a. 命令视窗（Command）输入命令：vector vif=@ediv(unit, (unit-r2))
 - b. 运行命令：命令行中按 Enter 键
 - c. 查看结果：双击 vif
 - 3) 说明（Eviews 代码行的解读²）：
 - a. 代码 vector(6) vif=@ediv(unit, (unit-r2)) 表示生成一个名为 vif 的列向量（vector）对象 vif （vector vif），并使该列向量等于另外两个向量之除 ($=@ediv(unit, (unit-r2))$)
 - b. 代码 @ediv(unit, (unit-r2)) 表示用 Eviews 矩阵函数 @ediv(vector1, vector2) 两个向量进行元素相除的计算。

² 具体细节请参看 Eviews 在线帮助文档，网址 http://www.eviews.com/help/helpintro.html#page/content/matrixref-Matrix_Language_Reference.html

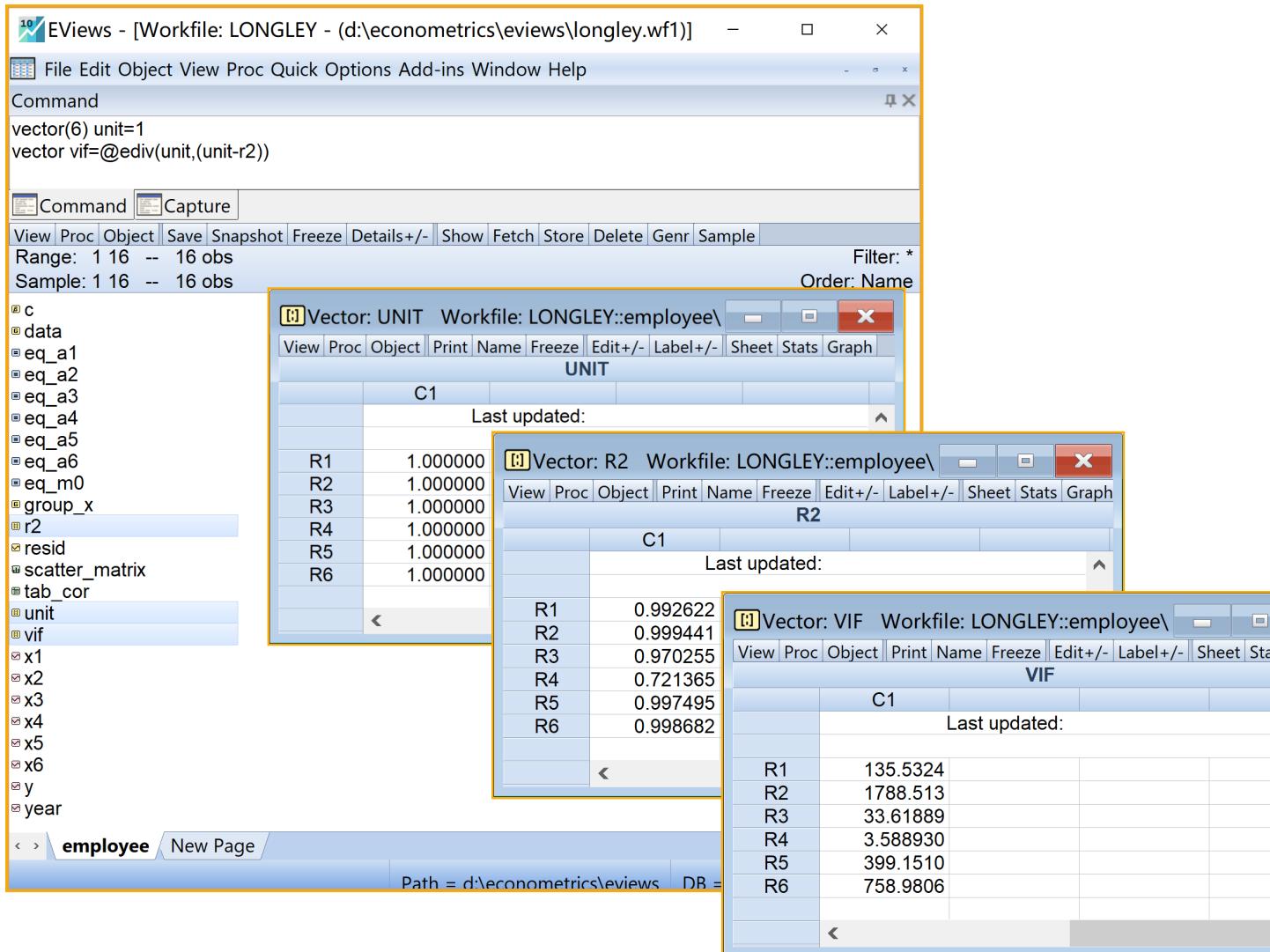


图 13: 辅助回归模型 A1 A6 的 VIF 手动提取操作

- (方法 2) Eviews 操作 (菜单操作实现, 具体见图14):
 - 1) 打开主方程: 双击方程 (equation) 对象 eq_mo
 - 2) 进入功能菜单:
 - 选择分析菜单: \Rightarrow View \Rightarrow Coefficient Diagnostics \Rightarrow Variance Inflation Factors
 - 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_vif)
 - 查看结果: 双击 tab_vif

4.3.3.4 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 $TOL_j(j = 1, \dots, 6)$, 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6), 分别计算得到各自的容忍度 TOL_j 。
- 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 $TOL_j(j = 1, \dots, 6)$
 - 诊断提示: 辅助回归方程的容忍度如果 $TOL_j \in [0.01, 0.1]$ 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

$$VIF_j = 1 - R_j^2 = \frac{1}{VIF_j}, (j = 1, \dots, 6) \quad (10)$$

- Eviews 操作 (手动计算, 具体见图15):
 - 1) 计算 TOL_j 。根据理论计算公式(10), 对常数列向量 unit 和判定系数向量 r2 进行运算, 得到容忍度向量 tol (建议将容忍度向量命名为 tol)
 - a. 命令视窗 (Command) 输入命令: vector tol=unit-r2
 - b. 运行命令: 命令行中按 Enter 键
 - c. 查看计算结果: 双击 tol

4.3.4 回归系数方差分解法 (Coefficient Variance Decomposition) (选学)

- 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)
- 提示:
 - 操作提示: 用 Eviews 的共线性诊断菜单 \Rightarrow View \Rightarrow Coefficient Diagnostics \Rightarrow Coefficient Variance Decomposition

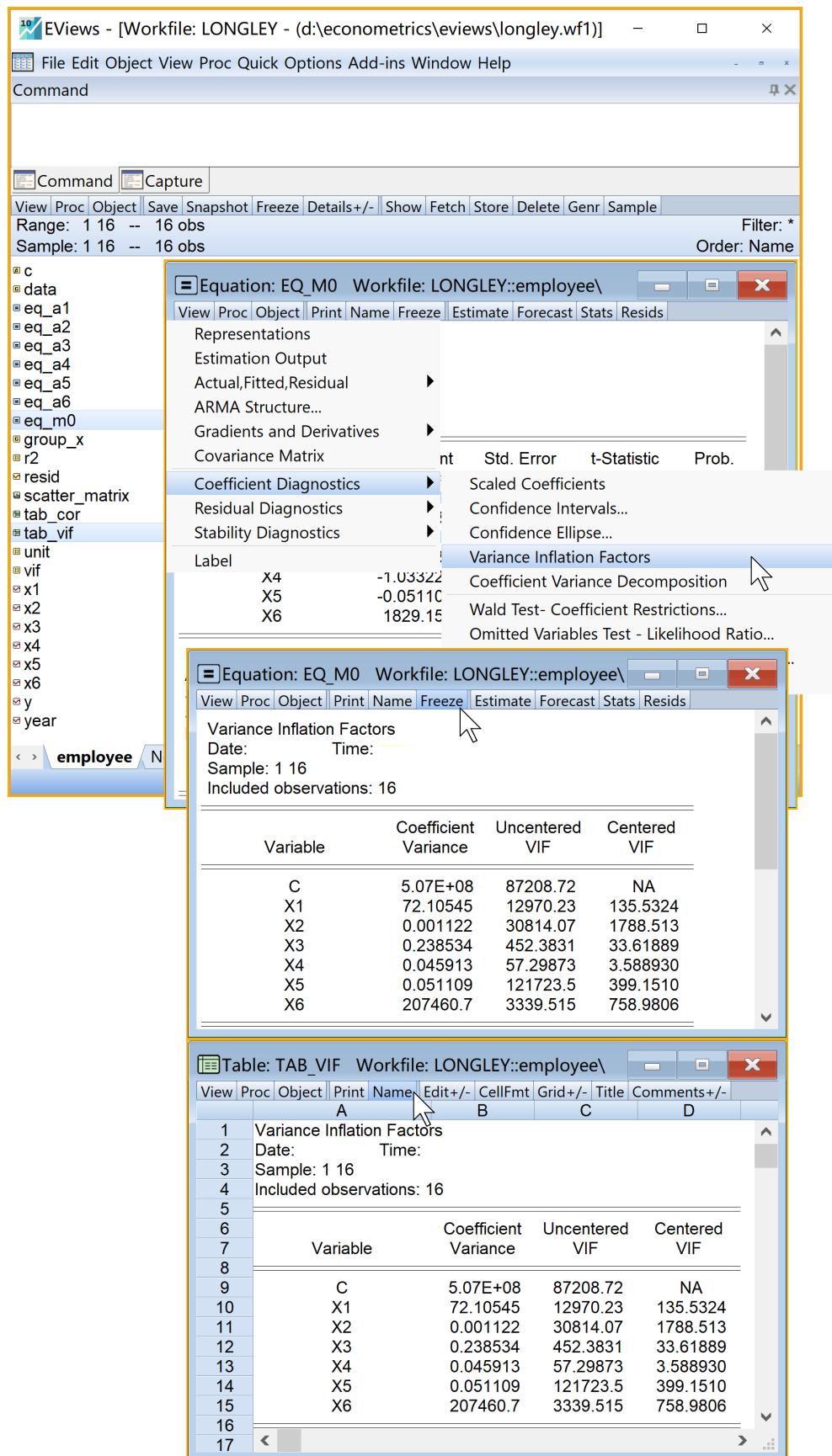


图 14: 辅助回归模型 A1A6 的 VIF 系统提取操作
25

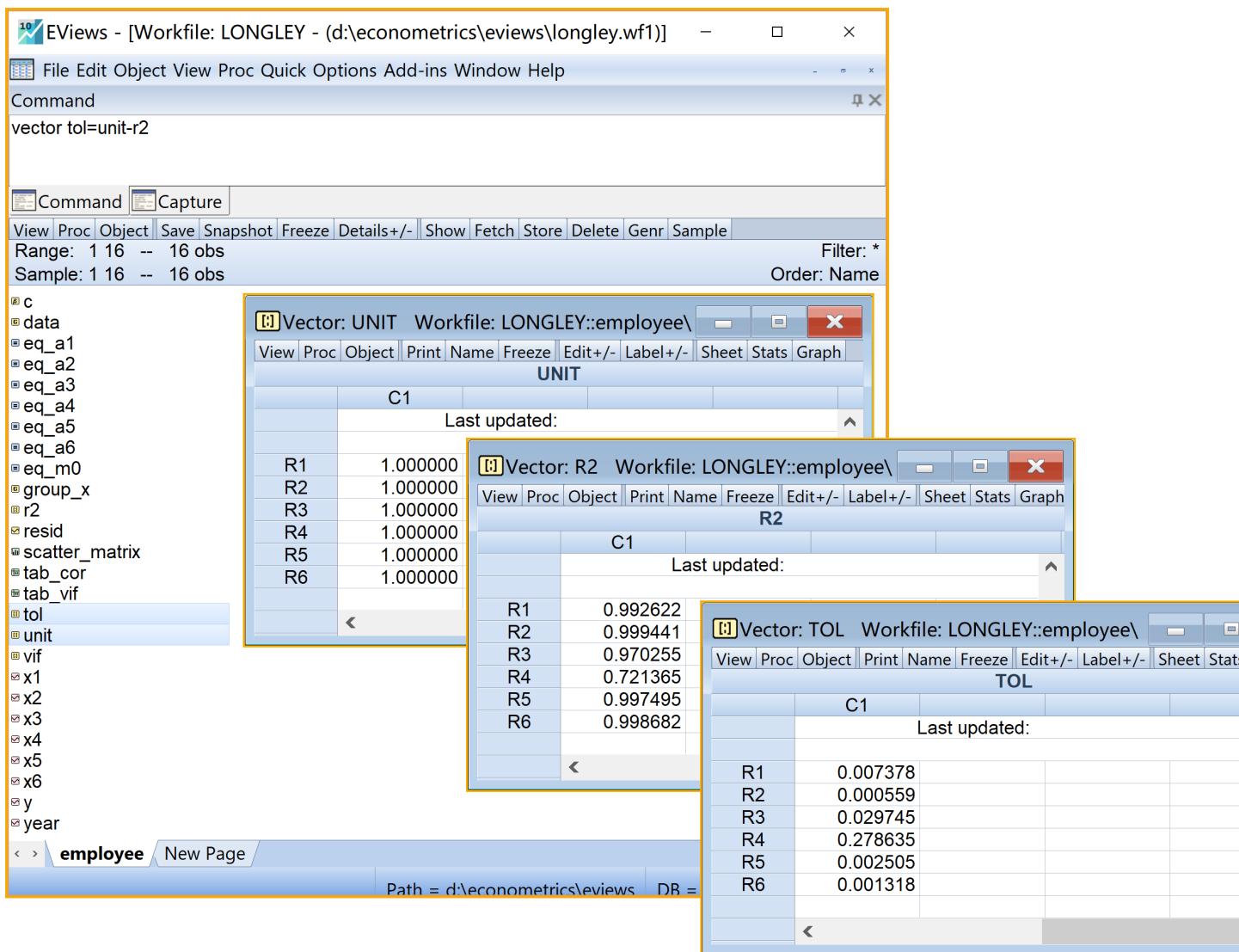


图 15: 辅助回归模型 A1 A6 的 TOL 手动提取操作

- 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
- 理论提示：该诊断方法由 Belsley, Kuh and Welsch (BKW) 2004 提出，具体细节可以参考Eviews 帮助文档³。注意，Eviews 分析的病态数是基于矩阵 $(\mathbf{X}'\mathbf{X})^{-1}$ ，而不是基于矩阵 \mathbf{X}

$$\widehat{\text{var}}_{\text{cov}}(\hat{\beta}) = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$$

- Eviews 操作（菜单操作实现，具体见图16）：

- 1) 打开主方程：双击方程（equation）对象 eq_mo
- 2) 进入功能菜单：
 - 选择分析菜单： \Rightarrow View \Rightarrow Coefficient Diagnostics \Rightarrow 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 （具体 Eviews 报告见17）

4.4 对存在多重共线性的模型进行修正

4.4.1 简单剔除变量法

- 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：

- 酉情删除：经济学和实践观察法
- 巧妙删除：变量变换法
- 规则删除：逐步回归法

4.4.1.1 经济学和实践观察法

- 提示：
 - 理论提示：利用先验信息（成为研究领域的专家！）酉情删除特定变量，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)

³ 具体细节可以参考 Eviews 在线帮助文档，网址 http://www.eviews.com/help/helpintro.html#page/content/testing-Coefficient_Diagnostics.html

EViews - [Workfile: LONGLEY - (d:\econometrics\reviews\longley.wf1)]

File Edit Object View Proc Quick Options Add-ins Window Help

Command

Command Capture

View | Proc | Object | Save Snapshot | Freeze | Details+/- | Show | Fetch | Store | Delete | Genr | Sample

Range: 1 16 -- 16 obs Filter: *

Sample: 1 16 -- 16 obs Order: Name

c
data
eq_a1
eq_a2
eq_a3
eq_a4
eq_a5
eq_a6
eq_m0
group_X
r2
resid
scatter_matrix
tab_cor
tab_cvd
tab_vif
tol
unit
vif
x1
x2
x3
x4
x5
x6
y
year

Equation: EQ_M0 Workfile: LONGLEY::employee

Representations
Estimation Output
Actual,Fitted,Residual
ARMA Structure...
Gradients and Derivatives
Covariance Matrix
Coefficient Diagnostics
Residual Diagnostics
Stability Diagnostics
Label

	nt	Std. Error	t-Statistic	Prob.
X4	-1.03322			
X5	-0.05110			
X6	1829.15			

R-squared: 0.99547
Adjusted R-squared: 0.99246
S.E. of regression: 304.854
Sum squared resid: 836424.1
Log likelihood: -109.6174
F-statistic: 330.2853
Prob(F-statistic): 0.000000

Coefficient Variance Decomposition

Path = d:\econometrics\reviews DR = none WF = longley

Equation: EQ_M0 Workfile: LONGLEY::employee

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

Date: 04/08/18 Time: 11:55
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

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

Table: TAB_CVD Workfile: LONGLEY::employee

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

A	B	C	D	E	F	G	H
1	Coefficient Variance Decomposition						
2	Date:	Time:					
3	Sample: 1 16						
4	Included observations: 16						
5							
6	Eigenvalues	5.07E+08	201598.2	28.96906	0.037097	0.008008	1.32E-05
7	Condition	6.63E-17	1.67E-13	1.16E-09	9.05E-07	4.19E-06	0.002542
8							
9							
10	Variance Decomposition Proportions						
11							
12							
13	Variable	1	2	3	4	5	6
14							
15	C	1.000000	4.61E-09	4.72E-15	5.83E-22	1.50E-21	3.04E-24
16	X1	0.595056	0.003296	0.401648	1.07E-07	2.73E-09	6.97E-12
17	X2	0.512335	0.477718	0.008255	0.000714	3.46E-05	0.000915
18	X3	0.383038	0.533542	0.003177	0.059528	0.020715	5.86E-09
19	X4	0.012803	0.289451	0.132728	0.498229	0.066790	5.18E-08
20	X5	0.923977	0.052811	0.022710	0.000263	2.17E-08	0.000238
21	X6	0.028271	0.971729	1.67E-10	8.26E-14	8.55E-15	9.26E-19
22							
23							

图 16: 主回归模型系数方差分解的操作过程

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

Variable	Associated Eigenvalue						
	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

Variable	Associated Eigenvalue						
	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 17: 主回归模型系数方差分解的 Eviews 报告

- * 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关；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18）：
 - 1) 变量变换，生成实际 GNP (= 名义 GNP/CPI)（见图18）
 - a. 命令视窗 (Command) 输入命令: series x_rgnp=X2/X1
 - b. 运行命令: 命令行中按 Enter 键
 - c. 查看计算结果
 - 2) 引导设置 Equation Estimation \Rightarrow specification
 - a. Equation specification: 输入命令 Y c x_rgnp X3 X4 X5
 - 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_adj_man
主回归分析结果如图：

4.4.1.2 变量变换法

4.4.1.2.1 一阶差分法 (first difference form):

- 理论提示：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：

$$Y_t = \beta_1 + \beta_2 X_{2,t} + \beta_3 X_{3,t} + u_t \quad \text{原模型} \quad (11)$$

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + u_{t-1} \quad \text{滞后 1 阶变量模型} \quad (12)$$

$$Y_t - Y_{t-1} = \beta_2(X_{2,t} - X_{2,t-1}) + \beta_3(X_{3,t} - X_{3,t-1}) + (u_t - u_{t-1}) \quad \text{一阶差分模型} \quad (13)$$

$$Y_t^* = \beta_2 X_{2,t}^* + \beta_3 X_{3,t}^* + v_t \quad \text{精简化模型} \quad (14)$$

- 提醒：“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。

4.4.1.2.2 比率变换法 (ratio transformation)

- 理论提示：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：

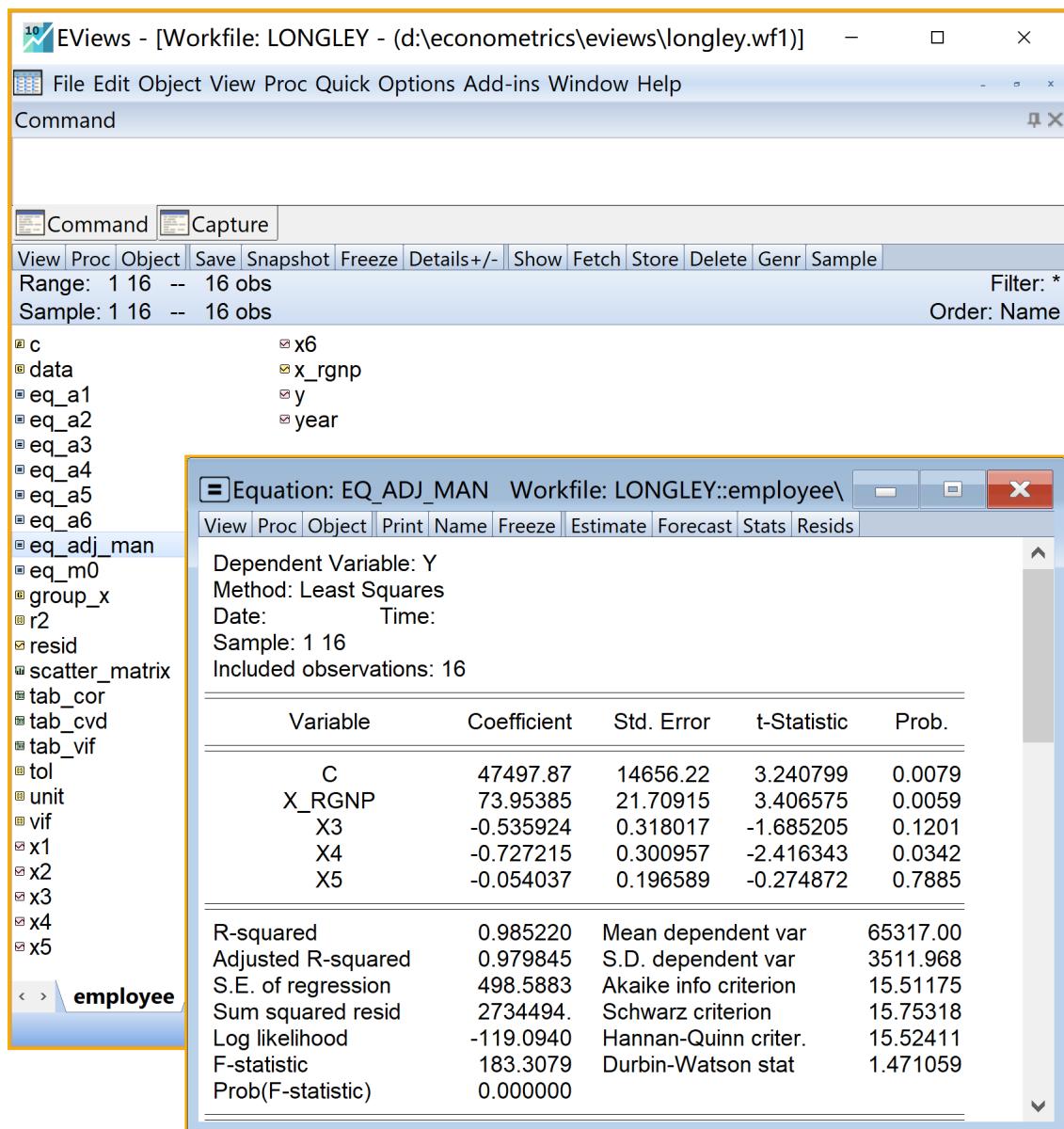


图 18: 多重共线性问题模型的矫正：酌情删除法

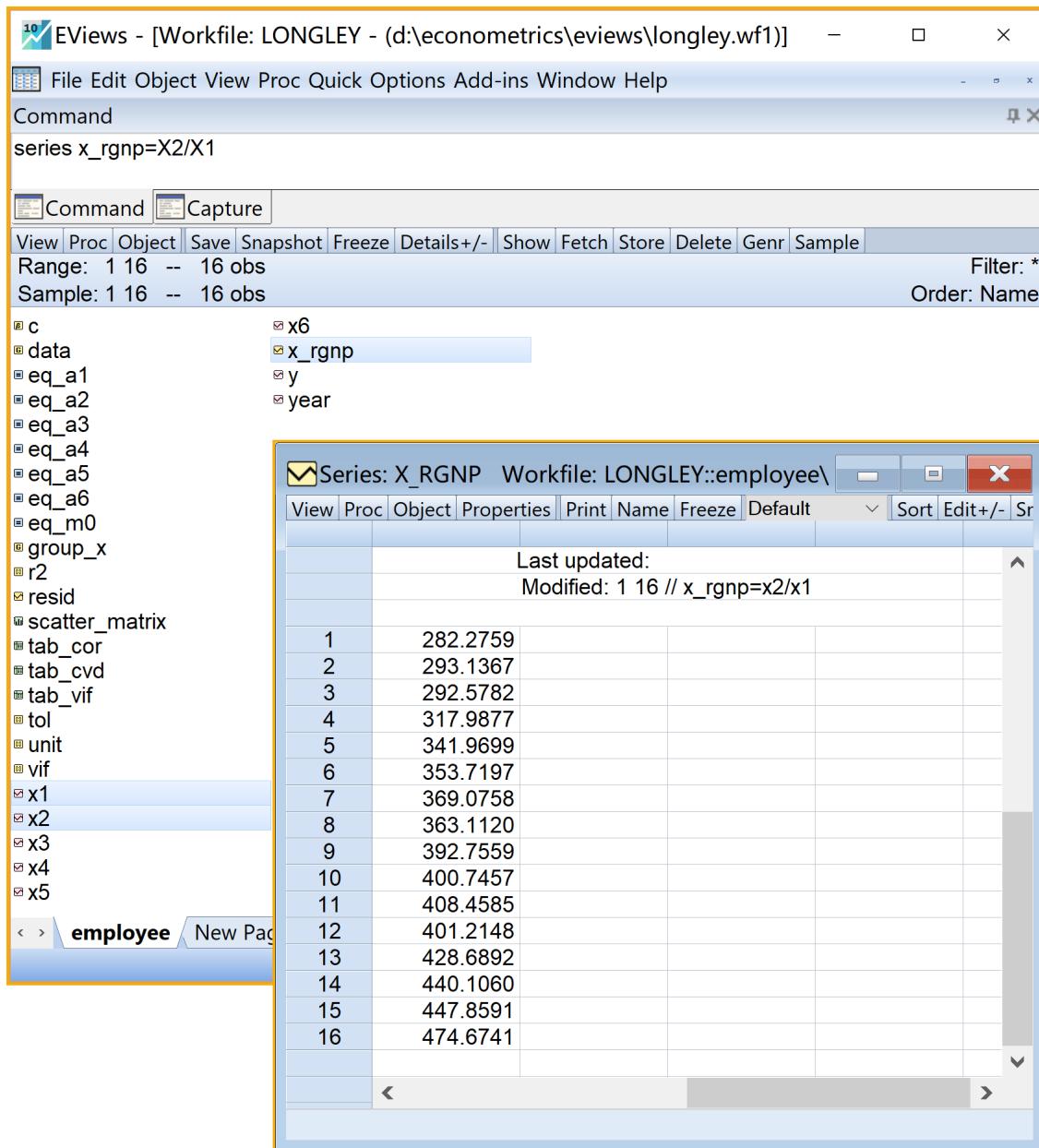


图 19: 多重共线性模型的变量变换操作

- 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。

$$Y_t = \beta_1 + \beta_2 X_{2,t} + \beta_3 X_{3,t} + u_t \quad (15)$$

$$\frac{Y_t}{X_{3,t}} = \frac{\beta_1}{X_{3,t}} + \beta_2 \frac{X_{2,t}}{X_{3,t}} + \frac{u_t}{X_{3,t}} \quad (16)$$

$$Y_t^* = \beta_1^* + \beta_2^* X_{2,t}^* + v_t \quad (17)$$

- 提醒: “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。

4.4.1.3 逐步最小二乘回归法 (Stepwise Least Squares Regression)

- 提示:

- 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。

- 操作提示⁴:

- * 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20):

- 1) 依次选择 \Rightarrow Quick \Rightarrow Estimation Equation
- 2) 引导设置 Equation Estimation \Rightarrow Specification
 - a. 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)
 - b. 输入自变量 (List of search regressors): c X1 c X2 c X3 c X4 c X5 c X6
 - c. 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 \Rightarrow Options
 - a. 方法设置 (Selection Method):
 - 下拉选择 Stepwise
 - 点击选择 Backwards
 - b. 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

⁴具体细节可参看 Eviews 说明书, 网址 http://www.eviews.com/help/helpintro.html#page/content/Regress2-Stepwise_Less_Squares_Regression.html

- 其他: 默认设置
 - c. 权重设置 (Weights): 默认设置
 - d. 步数设置 (Maximum steps): 默认设置
 - e. 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21)

4.4.2 主成分法 (Principal Components) (自学)

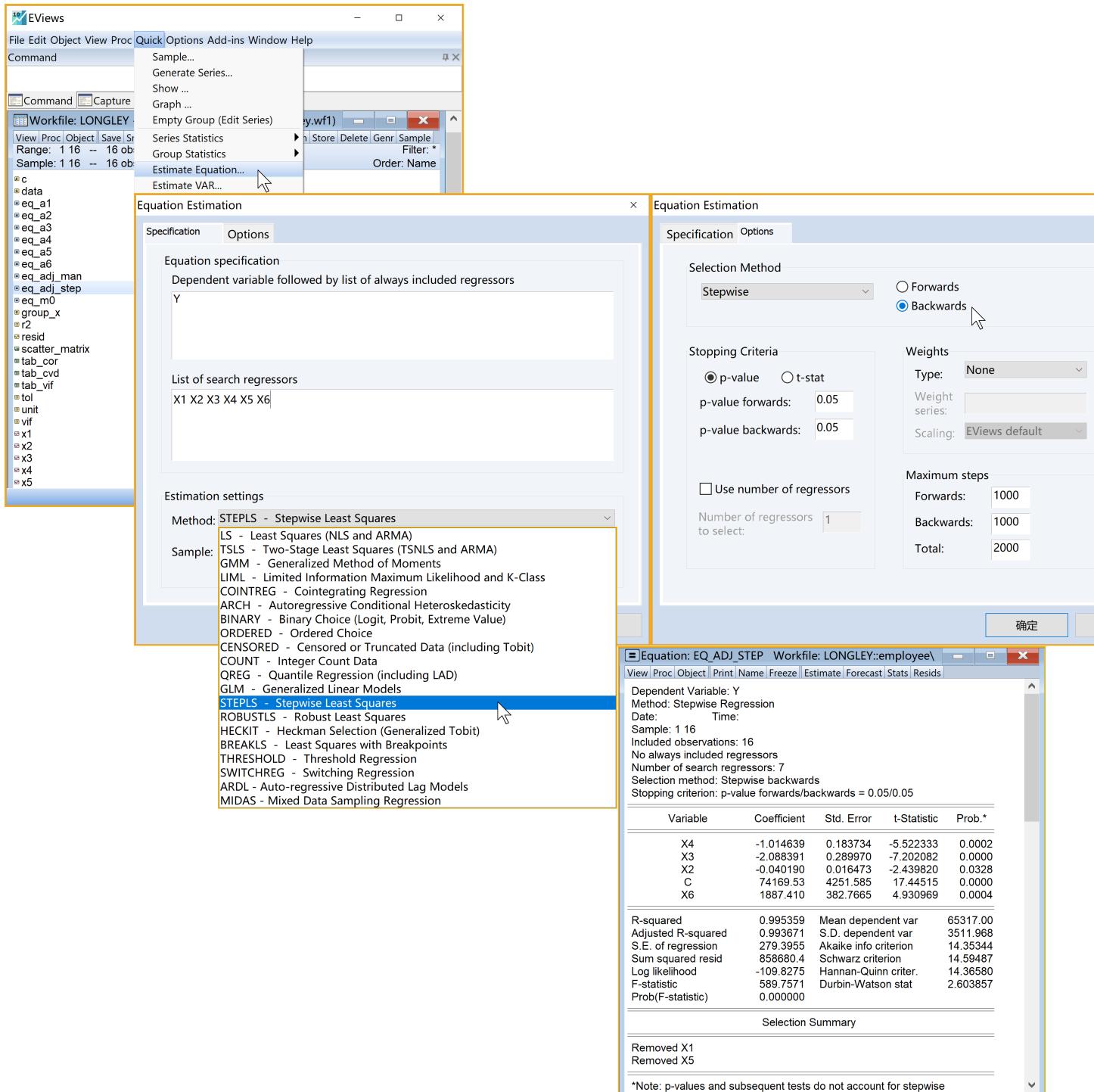


图 20: 多重共线性模型的矫正：后向逐步回归法

Dependent Variable: Y
 Method: Stepwise Regression
 Date: 11/16/2019 Time: 14:36
 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 21: 多重共线性模型的后向逐步回归法矫正报告