计量经济学Eviews实验指导书

Lab 8 虚拟变量回归模型

胡华平

2018/5/4

# 虚拟变量回归模型

## 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。

## 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。

Definition 1 (定量变量) 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。

Definition 2 (定性变量) 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。

Definition 3 (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。

### 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。

Definition 4 (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，假设定性变量具有个属性，对于任意属性，可以定义如下的虚拟变量：

Definition 5 (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，假设定性变量具有个属性，可以用如下一组虚拟变量完全表达该定性变量： $$

例如，定性变量肤色（）具有3个属性（），具体为，则可以构造出如下的虚拟变量体系[[1]](#footnote-27)：

### 如何理解虚拟变量回归模型？

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]](#footnote-28)。

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]](#footnote-29)
  + 完全混合模型
  + 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]](#footnote-30)：

* 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是否取对数，可以划分为[[5]](#footnote-31)：

* 经典线性模型：因变量为
* 半对数模型：因变量为

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1.5.5）
* 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1.5.6）

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1.5.3和节1.5.4）：

* 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**

## 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
   1. 残差序列观察法（描点图法）：绘制序列的描点图（dot plot）
   2. 残差序列观察法（描点图法）：确定滞后阶数并分别绘制序列与序列的散点图（scatter plot）
   3. 辅助回归法：构建残差序列对序列的辅助回归方程
   4. 自相关和偏相关分析法：Eviews菜单操作对残差序列进行自相关和偏相关分析（**注意滞后阶数的选择**）
   5. Durbin-Watson检验法：分析Eviews报告中的D-W统计量
   6. 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)
3. 自相关问题模型的矫正方法：
   1. 广义最小二乘法(GLS)：一阶差分法变换
   2. 广义最小二乘法(GLS)：基于残差辅助方程近似得到
   3. 广义最小二乘法(GLS)：基于D-W统计量近似计算得到
   4. 广义最小二乘法(GLS)：基迭代法近似计算得到
   5. 一致标准误校正法（HAC）：尼威-威斯特(Newey-West)校正法

## 实验案例——印度工人工资

**印度工人工资**：表1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1 印度工人工资(n=114)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| obs | wage | age | edu | dpt | sex |
| 1 | 117.00 | 26 | primary | permanent | female |
| 2 | 375.00 | 42 | primary | permanent | female |
| 3 | 175.00 | 33 | primary | permanent | female |
| 4 | 100.00 | 33 | primary | permanent | female |
| 5 | 162.50 | 30 | primary | permanent | female |
| 110 | 25.00 | 18 | illiteracy | temporary | male |
| 111 | 25.00 | 11 | illiteracy | temporary | male |
| 112 | 75.00 | 45 | illiteracy | temporary | male |
| 113 | 53.84 | 14 | illiteracy | temporary | male |
| 114 | 50.00 | 26 | illiteracy | temporary | male |

变量说明见表2：

表2 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 |

## 主要实验步骤

### 导入数据并进行预处理

* 目标：
* 思路：
* 新建Eviews工作文件（见图1）
  + 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：
    1. 依次操作：FileNewWorkfile
    2. 进行workfile create引导设置：
       - workfile structure type: unstructured/undatede
       - data range：114
       - workfile names(optional):
         * WF: worker（**建议命名**）
         * Page: indian（**建议命名**）
* Eviews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）：
    1. Excel找到数据。Excel表格中仅保留自己需要的数据（obs, wage, age, edu, dpt, sex）
    2. Eviews导入数据。FileImportImport From File：d:/econometrics/data/Lab8-indian-wage.xlsx

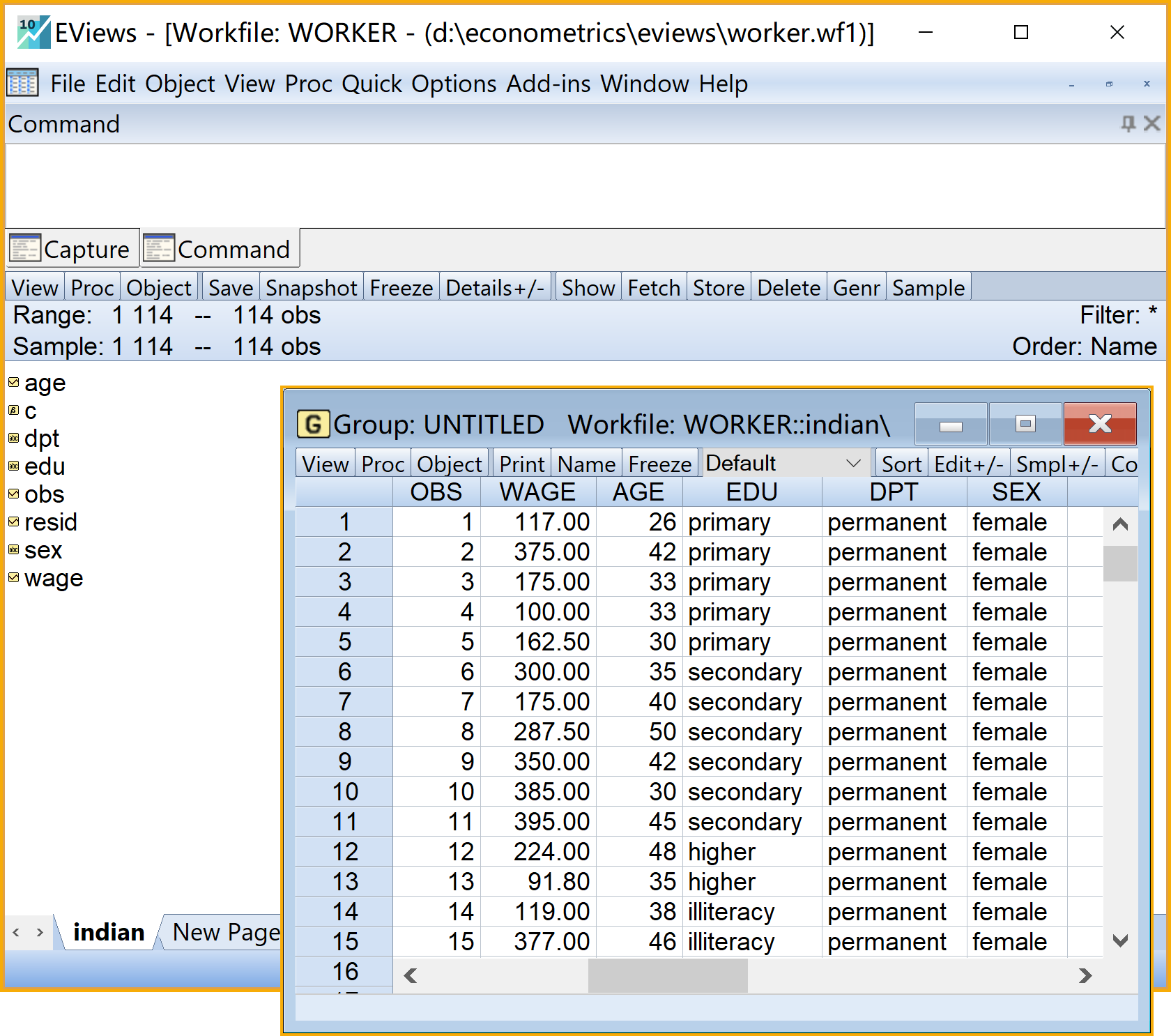


图1 导入数据的Eviews视窗

### 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2）
     1. 命令视窗（Command）依次输入命令（建议分别命名为edu\_d1、edu\_d2、edu\_d3和edu\_d4）
        + series edu\_d1=@recode(edu="illiteracy",1,0)'
        + series edu\_d2=@recode(edu="primary",1,0)'
        + series edu\_d3=@recode(edu="secondary",1,0)
        + series edu\_d4=@recode(edu="higher",1,0)'
     2. 运行命令：命令行中按Enter键
     3. 查看结果（以组group的形式查看）：
        + 按住键盘Ctrl+依次点击edu、edu\_d1、edu\_d2、edu\_d3、edu\_d4
        + 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3）
     1. 命令视窗（Command）依次输入命令（建议分别命名为dpt\_d1和dpt\_d2）
        + series dpt\_d1=@recode(dpt="temporary",1,0)'
        + series dpt\_d2=@recode(dpt="permanent",1,0)
     2. 运行命令：命令行中按Enter键
     3. 查看结果（以组group的形式查看）：
        + 按住键盘Ctrl+依次点击dpt、dpt\_d1、dpt\_d2
        + 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4）
     1. 命令视窗（Command）依次输入命令（建议分别命名为sex\_d1和sex\_d2）
        + series sex\_d1=@recode(sex="female",1,0)'
        + series sex\_d2=@recode(sex="male",1,0)
     2. 运行命令：命令行中按Enter键
     3. 查看结果（以组group的形式查看）：
        + 按住键盘Ctrl+依次点击sex、sex\_d1、sex\_d2
        + 点击鼠标右键 Open as Group
  4. 说明（[Eviews代码行的解读](http://www.eviews.com/Learning/dummies.html)^[具体细节请参看Eviews在线学习文档，网址http://www.eviews.com/Learning/dummies.html）：
     1. 代码series edu\_d1=@recode(edu="primary",1,0)表示给创建一个序列（Series）对象edu\_d1,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\_d1。
     2. 代码@recode(edu="primary",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primary，则相应赋值为1，或者就相应赋值为0。

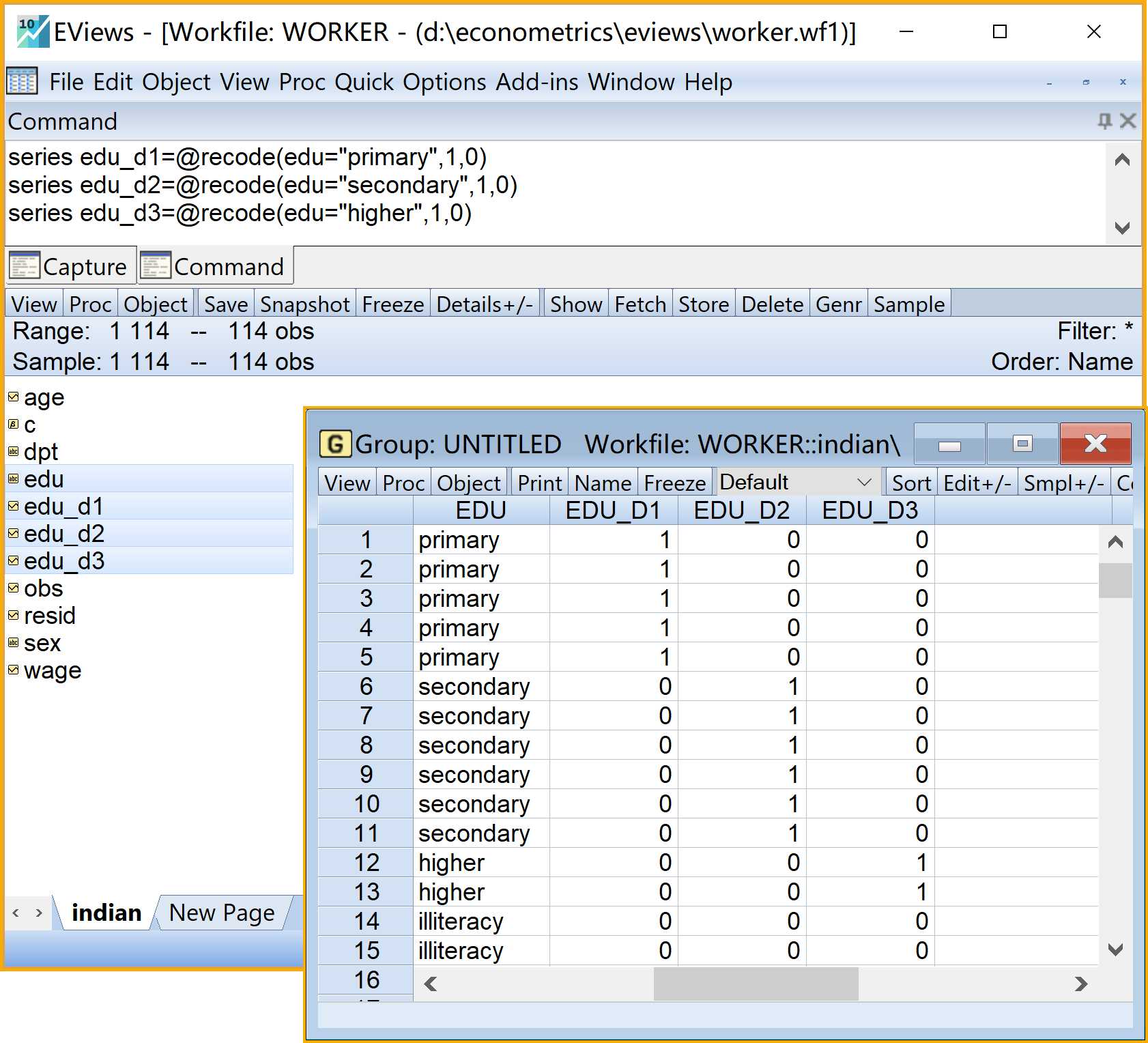


图2 定性变量edu用虚拟变量体系表达

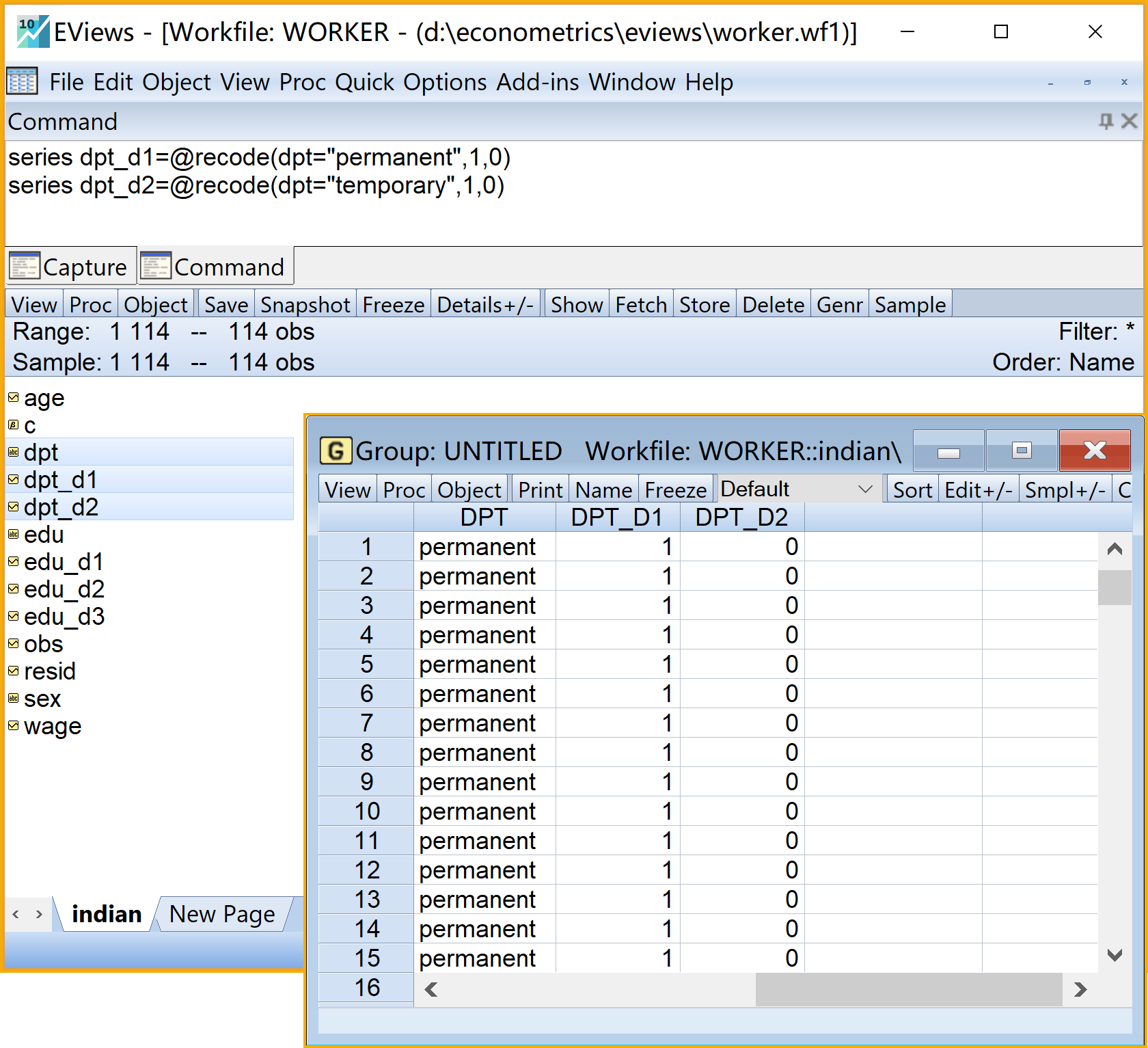


图3 定性变量dpt用虚拟变量体系表达

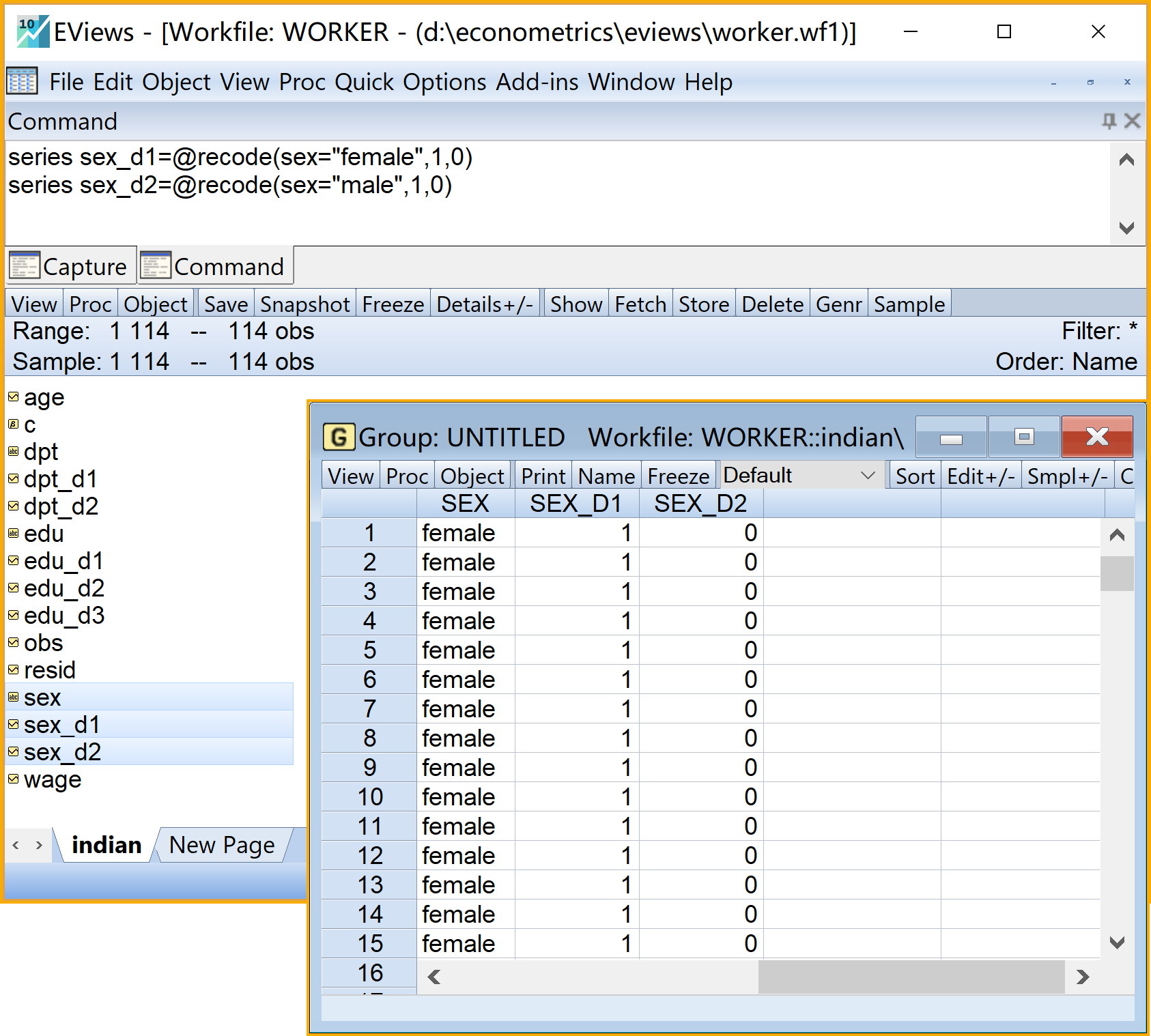


图4 定性变量sex用虚拟变量体系表达

#### 操作解读

## Loading required package: psych

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]](#footnote-38)。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]](#footnote-39)。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\_D1（见表3）、dpt\_D1和sex\_D1（见表4）。

表3 用虚拟变量系统完全表达定性变量edu

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | edu | edu\_D1 | edu\_D2 | edu\_D3 | edu\_D4 |
| 1 | primary | 0 | 1 | 0 | 0 |
| 2 | primary | 0 | 1 | 0 | 0 |
| 6 | secondary | 0 | 0 | 1 | 0 |
| 7 | secondary | 0 | 0 | 1 | 0 |
| 12 | higher | 0 | 0 | 0 | 1 |
| 13 | higher | 0 | 0 | 0 | 1 |
| 14 | illiteracy | 1 | 0 | 0 | 0 |
| 15 | illiteracy | 1 | 0 | 0 | 0 |

表4 用虚拟变量系统分别完全表达定性变量dpt和sex

|  |  |  |  |
| --- | --- | --- | --- |
|  | dpt | dpt\_D1 | dpt\_D2 |
| 1 | permanent | 0 | 1 |
| 2 | permanent | 0 | 1 |
| 3 | permanent | 0 | 1 |
| 112 | temporary | 1 | 0 |
| 113 | temporary | 1 | 0 |
| 114 | temporary | 1 | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | sex | sex\_D1 | sex\_D2 |
| 1 | female | 1 | 0 |
| 2 | female | 1 | 0 |
| 3 | female | 1 | 0 |
| 112 | male | 0 | 1 |
| 113 | male | 0 | 1 |
| 114 | male | 0 | 1 |

### 只含有虚拟变量的回归模型（考虑基础组的情形）

#### 加法模型

* 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
* 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
* 理论提示：
  + 模型1：只含有虚拟变量的、加法形式的**经典回归模型**见方程(1)
  + 模型2：只含有虚拟变量的、加法形式的**半对数回归模型**见方程(2)
* Eviews操作1（只含有虚拟变量的、加法形式的**经典回归模型**见方程(1)，菜单操作实现具体见图5）：
  1. 确定参照组为[**文盲&短期合同&女性**]，则如下虚拟变量将**不进入**回归模型
     1. edu\_d1
     2. dpt\_d1
     3. sex\_d1
  2. 设置回归模型。进入引导设置Equation Estimation specification
     1. Equation specification：输入命令wage c edu\_d2 edu\_d3 edu\_d4 dpt\_d2 sex\_d2
     2. Estimation settings：
        + Method: 下拉选择LS - Least Squares (NLS and ARMA)
        + Sample: （默认设置）
     3. 点击完成：OK
     4. 命名保存方程对象：（建议命名为eq\_only\_plus）
     5. 查看结果：双击eq\_only\_plus

具体Eviews报告见6：

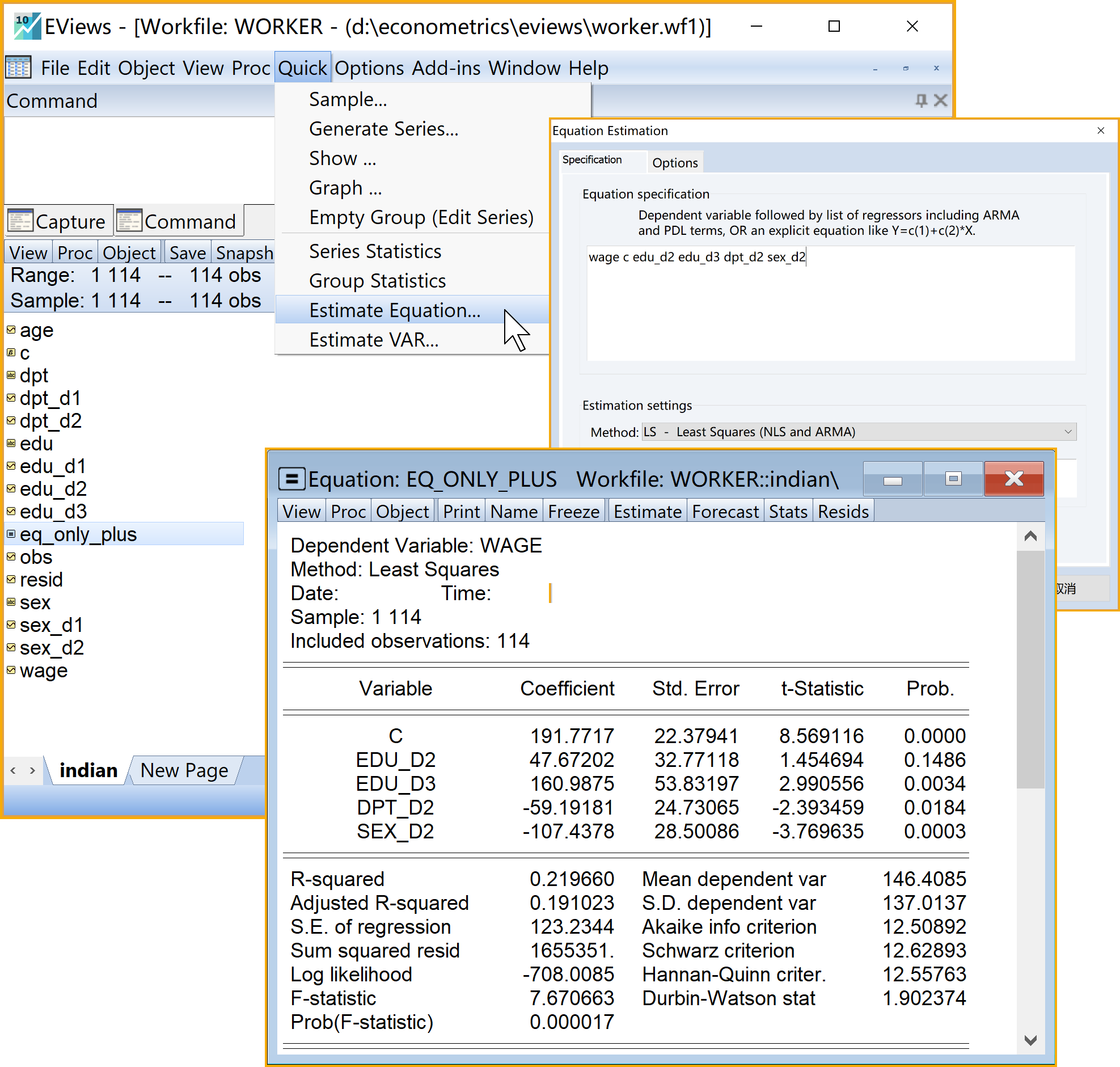


图5 只含虚拟变量的、加法形式的经典线性回归模型Eviews实现

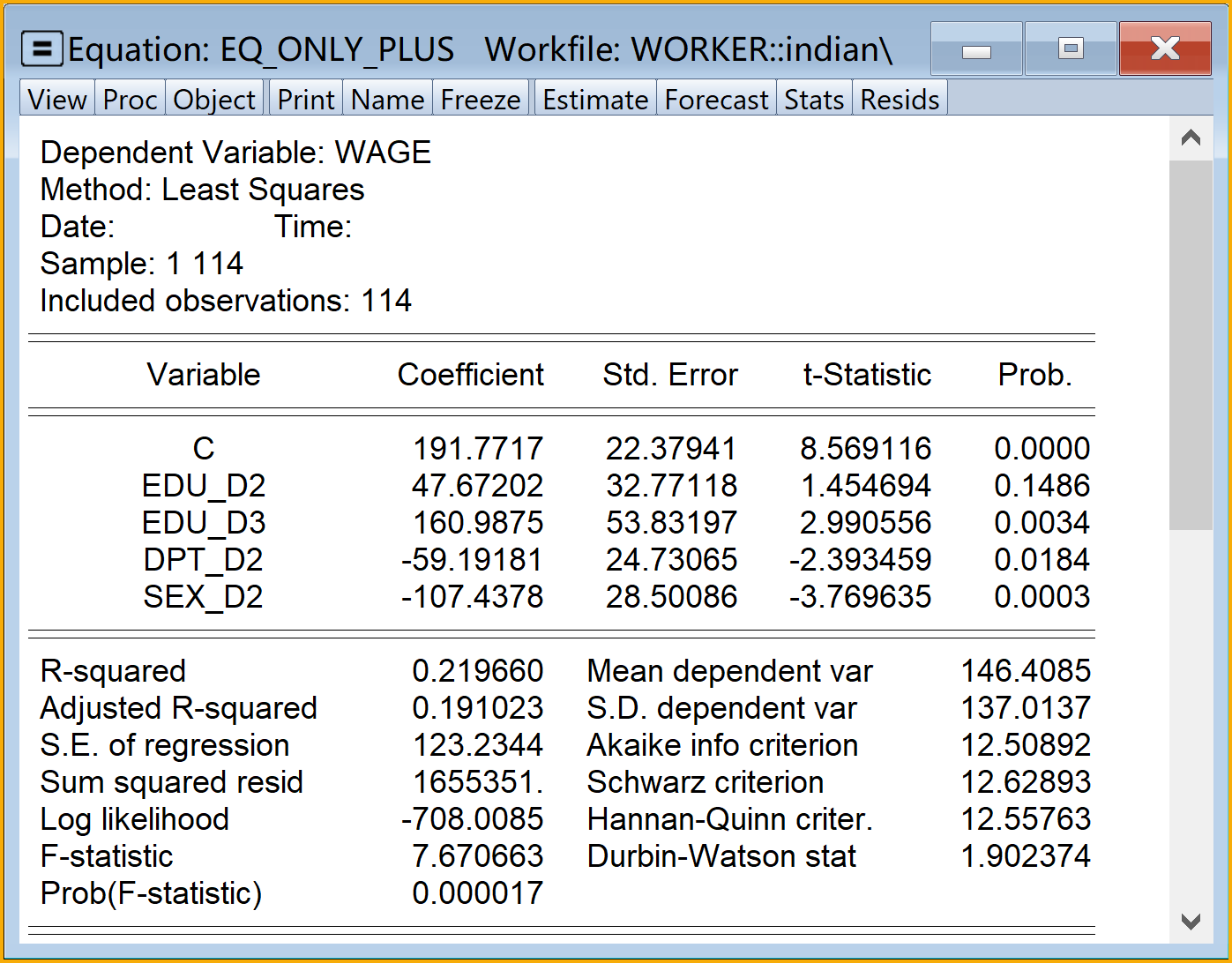


图6 只含虚拟变量的、加法形式的经典线性回归模型Eviews报告

* Eviews操作2（只含有虚拟变量的、加法形式的**半对数回归模型**见方程(2)，菜单操作实现具体见图7）：
  1. 确定参照组为[**文盲&短期合同&女性**]，则如下虚拟变量将**不进入**回归模型
     1. edu\_d1
     2. dpt\_d1
     3. sex\_d1
  2. 设置回归模型。进入引导设置Equation Estimation specification
     1. Equation specification：输入命令log(wage) c edu\_d2 edu\_d3 edu\_d4 dpt\_d2 sex\_d2
     2. Estimation settings：
        + Method: 下拉选择LS - Least Squares (NLS and ARMA)
        + Sample: （默认设置）
     3. 点击完成：OK
     4. 命名保存方程对象：（建议命名为eq\_only\_plus\_log）
     5. 查看结果：双击eq\_only\_plus\_log

具体Eviews报告见8：

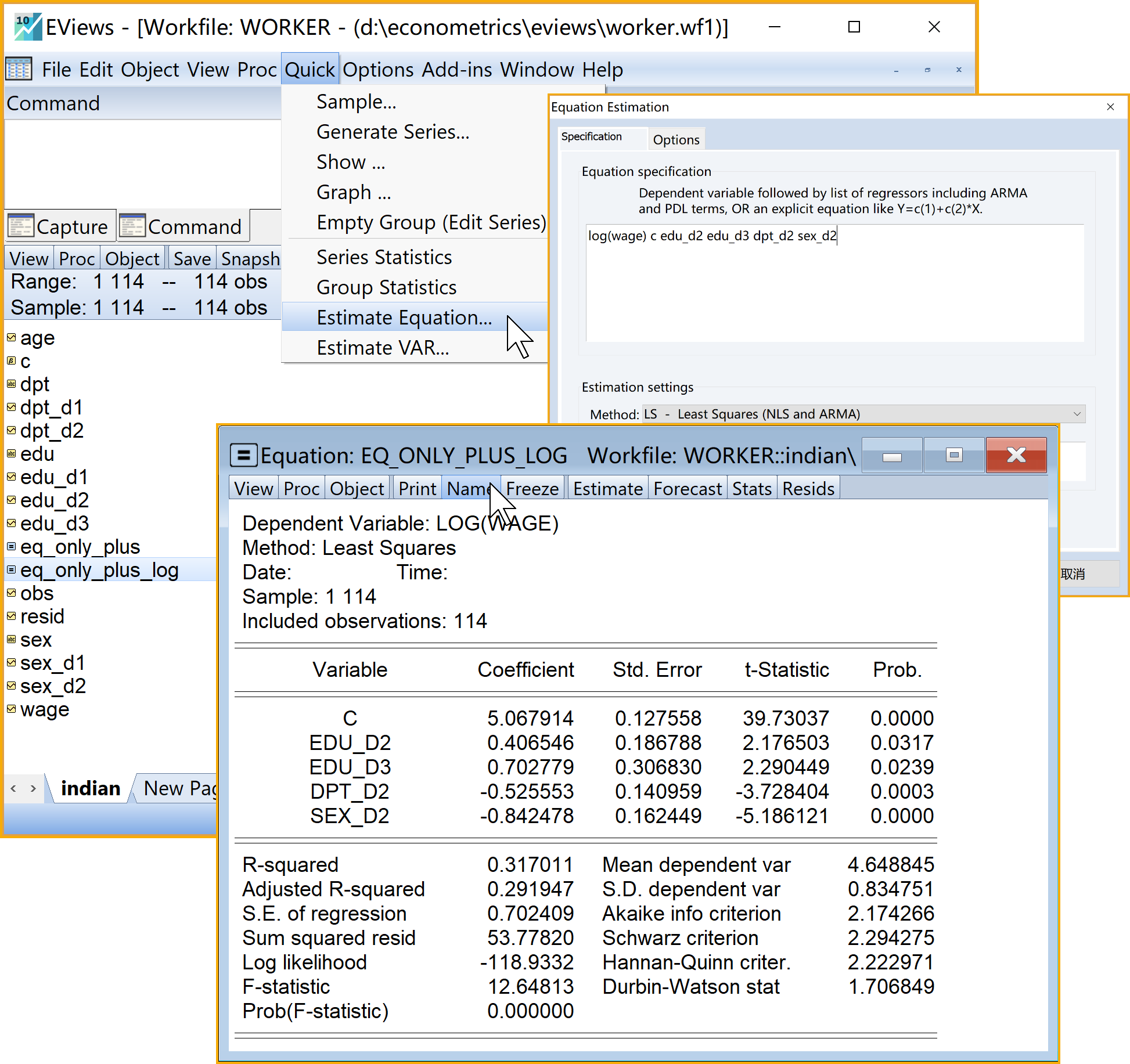


图7 只含虚拟变量的、加法形式的半对数线性回归模型Eviews实现

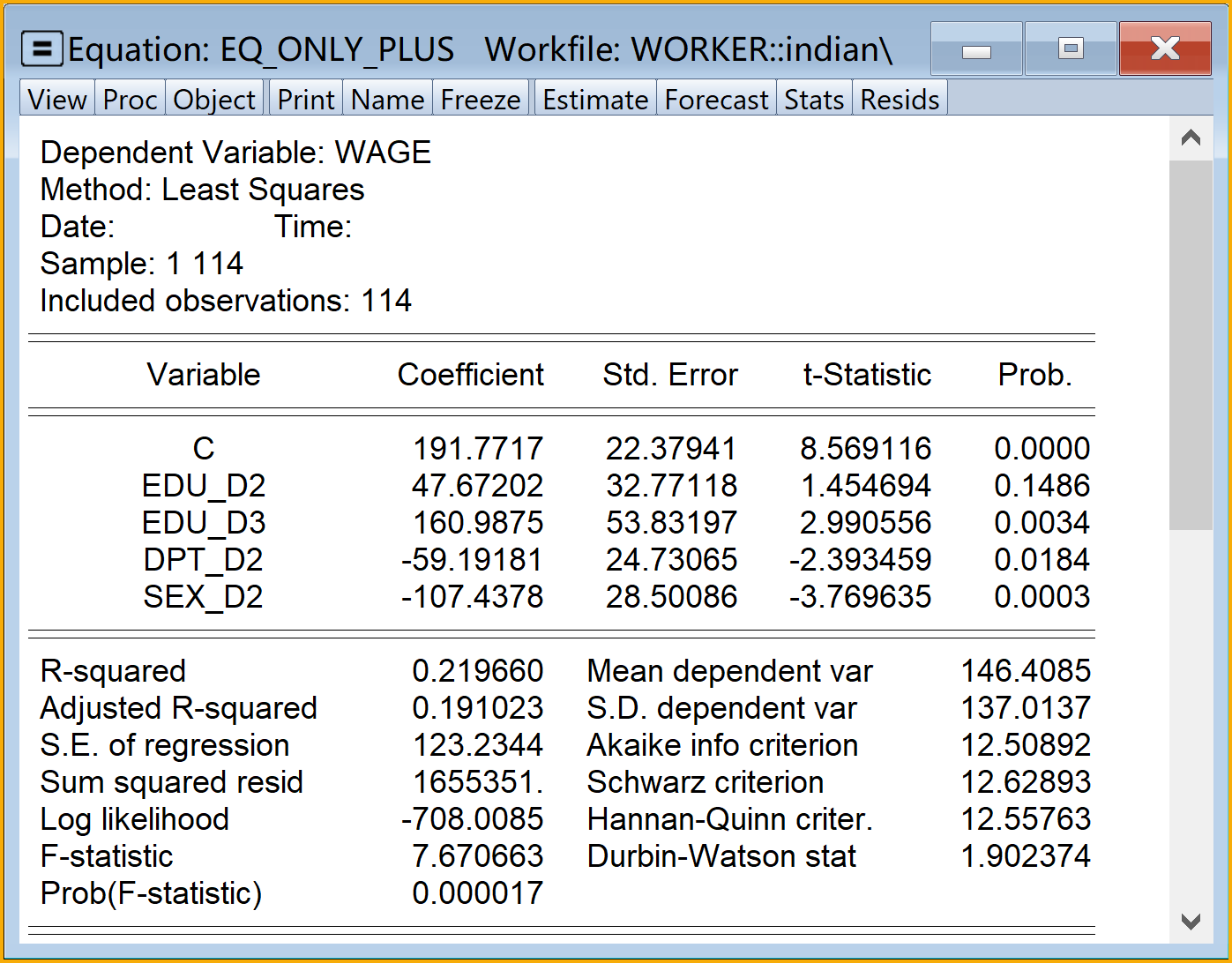


图8 只含虚拟变量的、加法形式的半对数线性回归模型Eviews报告

##### 报告解读

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{alignedat}{5} & =+&130.36+&10.47edu\_D2+&49.71edu\_D3+&162.59edu\_D4+&59.37dpt\_D2 \ & &(7.3397)&(0.3126)&(1.4818)&(2.9944)&(2.3899) \ & &(17.7609)&(33.4850)&(33.5470)&(54.2989)&(24.8399) \ & -&106.49sex\_D2 \ & &(-3.7005) \ & &(28.7784) \ & & & R^2=0.2204& {R^2}=0.1843& F^{}=6.1053 \end{alignedat}

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{alignedat}{5} & =+&4.53+&0.06edu\_D2+&0.42edu\_D3+&0.71edu\_D4+&0.53dpt\_D2 \ & &(44.7474)&(0.3045)&(2.1853)&(2.2995)&(3.7187) \ & &(0.1012)&(0.1909)&(0.1912)&(0.3095)&(0.1416) \ & -&0.84sex\_D2 \ & &(-5.1040) \ & &(0.1640) \ & & & R^2=0.3176& {R^2}=0.2860& F^{}=10.0528 \end{alignedat}

#### 乘法模型

* 目标：把定性变量的虚拟变量以交叉项的形式引入模型方程，解释回归报告
* 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
* 理论提示：
  + 模型1：只含有虚拟变量的、乘法形式的**经典回归模型**见方程(3)和方程(4)；
  + 模型2：只含有虚拟变量的、乘法形式的**半对数回归模型**见方程(5)
* Eviews操作1（只含有虚拟变量的、乘法形式的**经典回归模型**见方程(3)，菜单操作实现具体见图9）：
  1. 确定参照组为[**文盲&短期合同&女性**]，则如下虚拟变量将**不进入**回归模型
     1. edu\_d1
     2. dpt\_d1
     3. sex\_d1
  2. 设置回归模型。进入引导设置Equation Estimation specification
     1. 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
     2. Estimation settings：
        + Method: 下拉选择LS - Least Squares (NLS and ARMA)
        + Sample: （默认设置）
     3. 点击完成：OK
     4. 命名保存方程对象：（建议命名为eq\_only\_prod）
     5. 查看结果：双击eq\_only\_prod

具体Eviews报告见10：

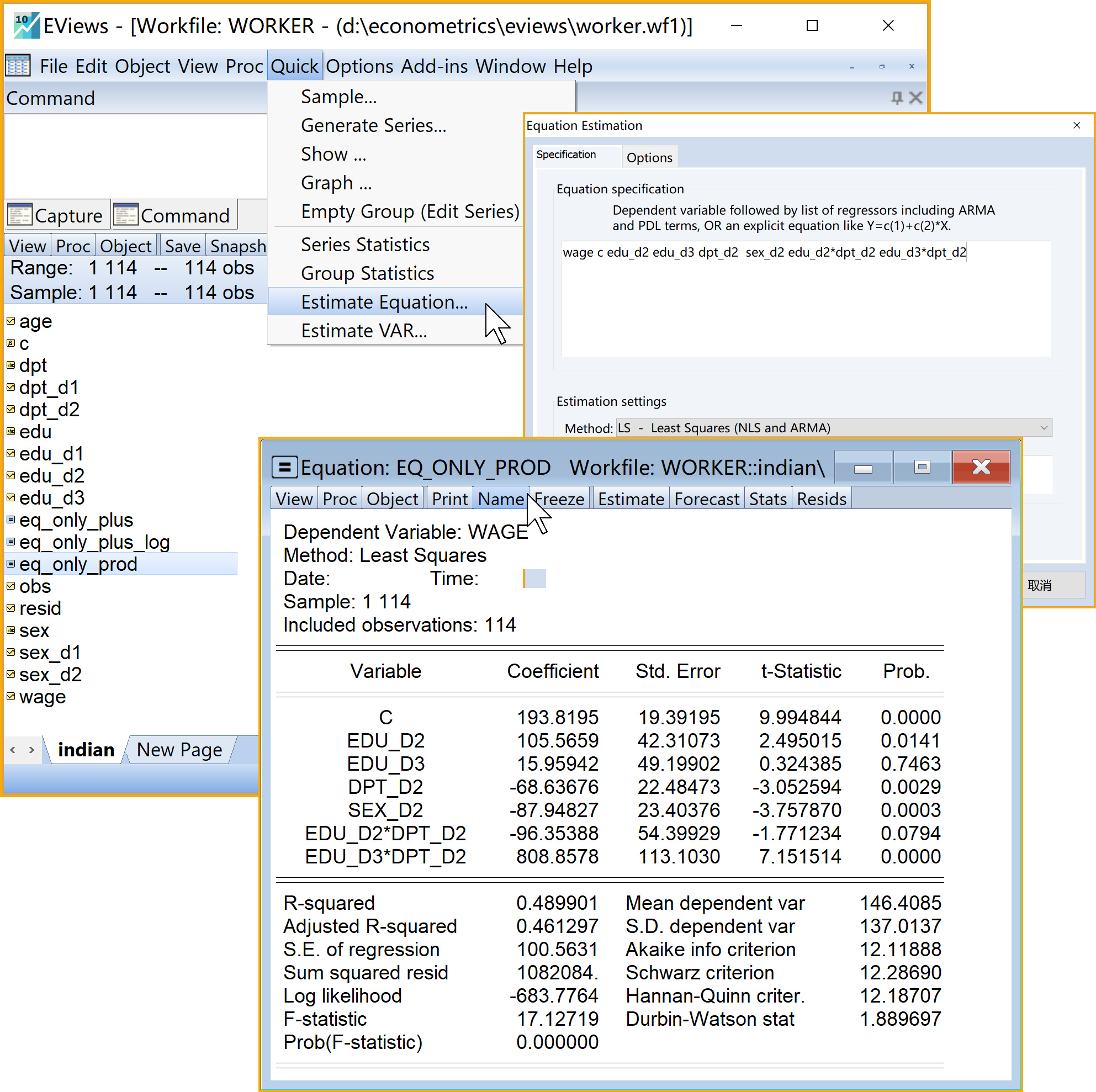


图9 只含虚拟变量的、乘法形式的经典线性回归模型Eviews实现

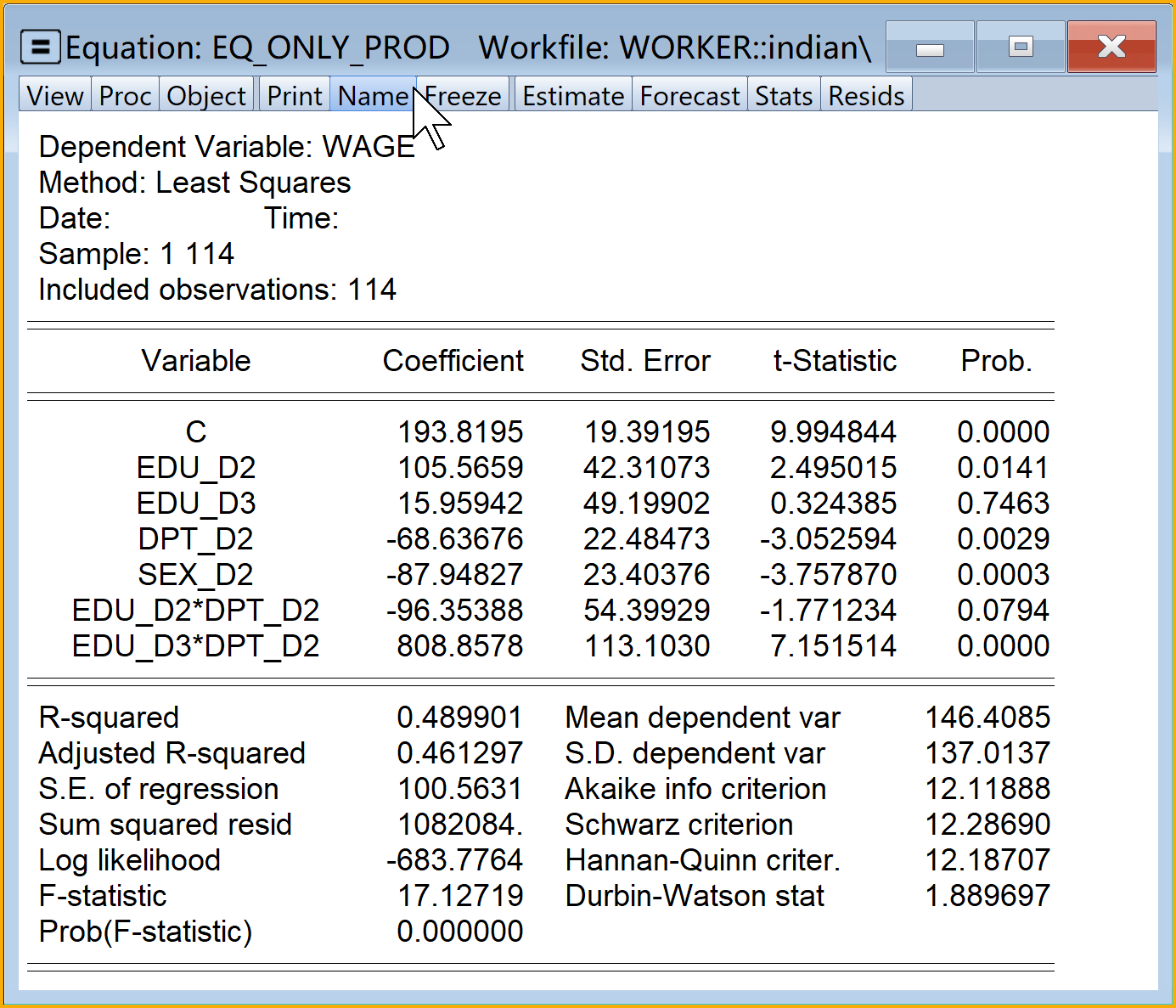


图10 只含虚拟变量的、乘法形式的经典线性回归模型Eviews报告

* Eviews操作2（只含有虚拟变量的、乘法形式的**半对数回归模型**见方程(5)，菜单操作实现具体见图11）：
  1. 确定参照组为[**文盲&短期合同&女性**]，则如下虚拟变量将**不进入**回归模型
     1. edu\_d1
     2. dpt\_d1
     3. sex\_d1
  2. 设置回归模型。进入引导设置Equation Estimation specification
     1. 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
     2. Estimation settings：
        + Method: 下拉选择LS - Least Squares (NLS and ARMA)
        + Sample: （默认设置）
     3. 点击完成：OK
     4. 命名保存方程对象：（建议命名为eq\_only\_prod\_log\_part）
     5. 查看结果：双击eq\_only\_prod\_log\_part

具体Eviews报告见12：

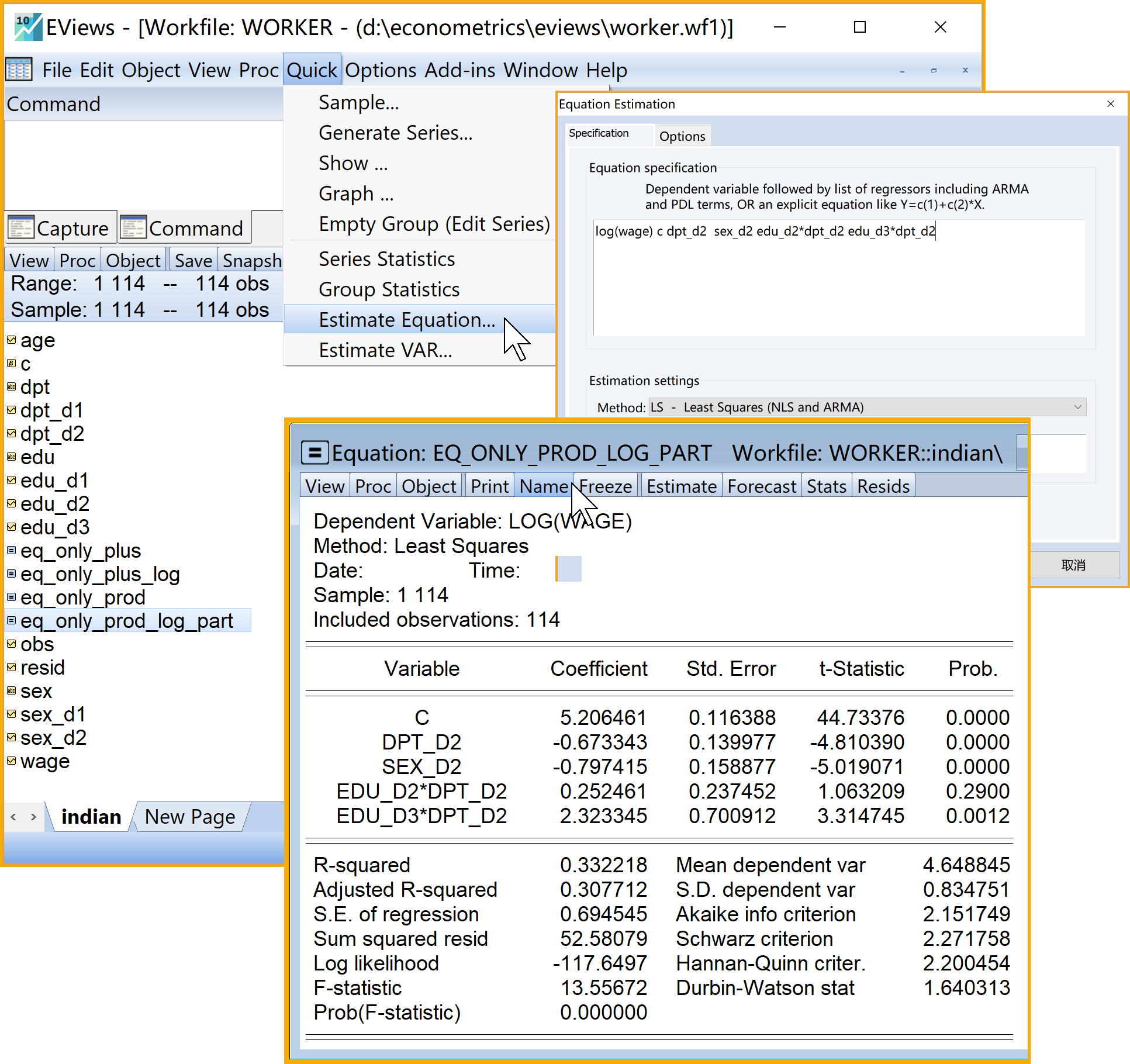


图11 只含虚拟变量的、乘法形式的半对数线性回归模型Eviews实现

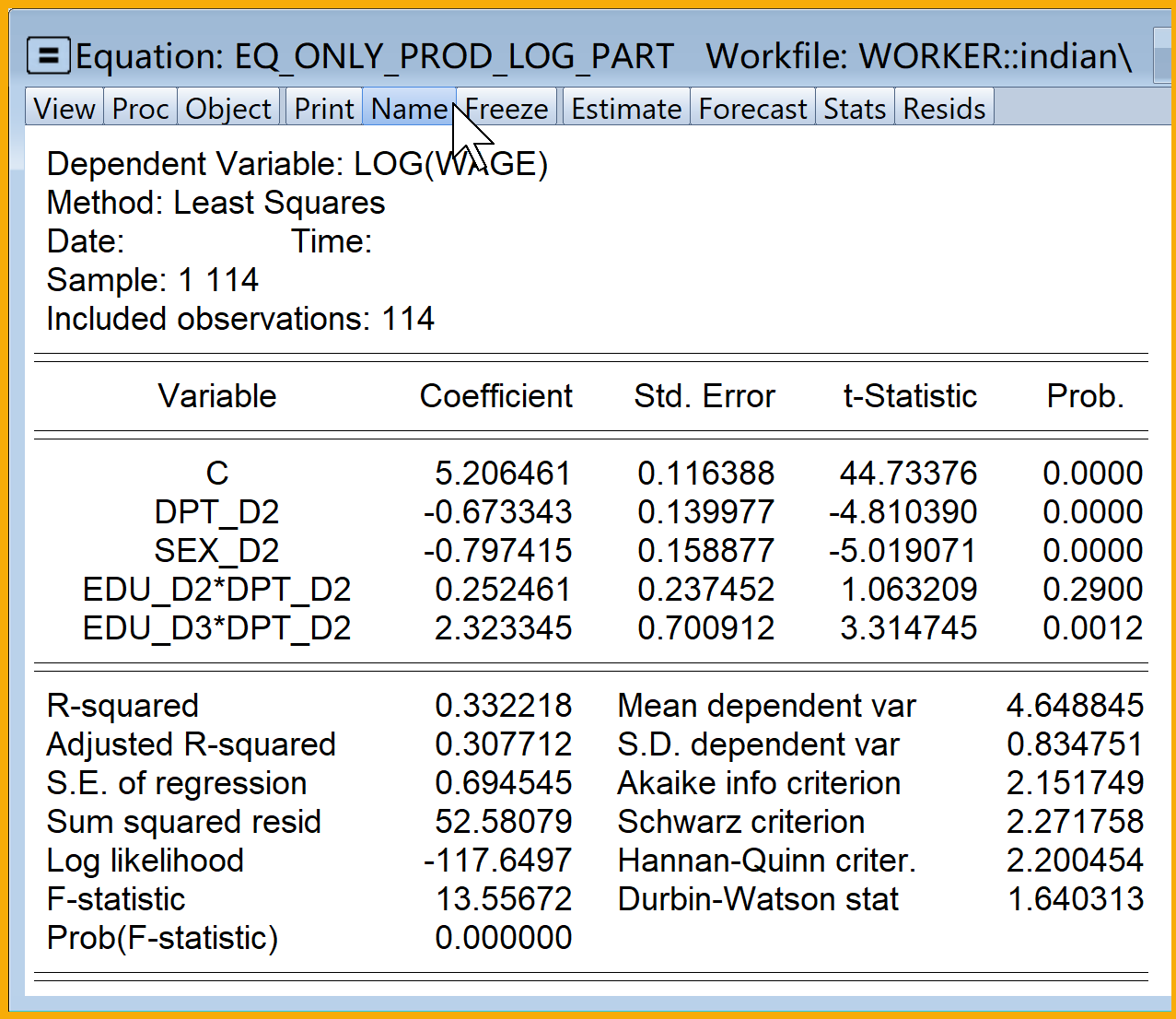


图12 只含虚拟变量的、乘法形式的半对数线性回归模型Eviews报告

##### 报告解读

只含虚拟变量的、乘法形式的经典回归模型（全部变量进入）Eviews结果简要报告如下：

\begin{alignedat}{5} & =+&120.70+&23.03edu\_D2+&13.72edu\_D3+&829.30edu\_D4+&74.78dpt\_D2 \ & &(7.9082)&(0.7060)&(0.3895)&(8.0985)&(2.9864) \ & &(15.2625)&(32.6264)&(35.2241)&(102.4020)&(25.0404) \ & -&88.19sex\_D2-&32.62edu\_D2:dpt\_D2+&90.22edu\_D3:dpt\_D2-&814.86edu\_D4:dpt\_D2 \ & &(-3.6975)&(-0.5459)&(1.6171)&(-7.1313) \ & &(23.8515)&(59.7485)&(55.7896)&(114.2653) \ & & & R^2=0.4925& {R^2}=0.4538& F^{}=12.7366 \end{alignedat}

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{alignedat}{5} & =+&120.70+&74.78dpt\_D2-&88.19sex\_D2+&23.03edu\_D2+&13.72edu\_D3 \ & &(7.9082)&(2.9864)&(-3.6975)&(0.7060)&(0.3895) \ & &(15.2625)&(25.0404)&(23.8515)&(32.6264)&(35.2241) \ & +&829.30edu\_D4-&32.62dpt\_D2:edu\_D2+&90.22dpt\_D2:edu\_D3-&814.86dpt\_D2:edu\_D4 \ & &(8.0985)&(-0.5459)&(1.6171)&(-7.1313) \ & &(102.4020)&(59.7485)&(55.7896)&(114.2653) \ & & & R^2=0.4925& {R^2}=0.4538& F^{}=12.7366 \end{alignedat}

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{alignedat}{5} & =+&4.52+&0.52dpt\_D2-&0.79sex\_D2+&0.06edu\_D2+&0.27edu\_D3 \ & &(43.4667)&(3.0729)&(-4.8492)&(0.2838)&(1.1078) \ & &(0.1040)&(0.1706)&(0.1625)&(0.2222)&(0.2399) \ & +&2.34edu\_D4+&0.00dpt\_D2:edu\_D2+&0.38dpt\_D2:edu\_D3-&1.96dpt\_D2:edu\_D4 \ & &(3.3514)&(0.0097)&(1.0126)&(-2.5246) \ & &(0.6975)&(0.4070)&(0.3800)&(0.7783) \ & & & R^2=0.3656& {R^2}=0.3173& F^{}=7.5648 \end{alignedat}

### 同时含有虚拟变量和定量变量的回归模型（考虑基础组的情形）

#### 加法模型

* 理论提示：
  + 模型1：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(6)
  + 模型2：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(7)
* Eviews操作1（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(6)，菜单操作实现具体见图13）：
  1. 确定参照组为[**文盲&短期合同&女性**]，则如下虚拟变量将**不进入**回归模型
     1. edu\_d1
     2. dpt\_d1
     3. sex\_d1
  2. 设置回归模型。进入引导设置Equation Estimation specification
     1. Equation specification：输入命令wage c edu\_d2 edu\_d3 edu\_d4 dpt\_d2 sex\_d2 age
     2. Estimation settings：
        + Method: 下拉选择LS - Least Squares (NLS and ARMA)
        + Sample: （默认设置）
     3. 点击完成：OK
     4. 命名保存方程对象：（建议命名为eq\_both\_plus）
     5. 查看结果：双击eq\_both\_plus

具体Eviews报告见14：

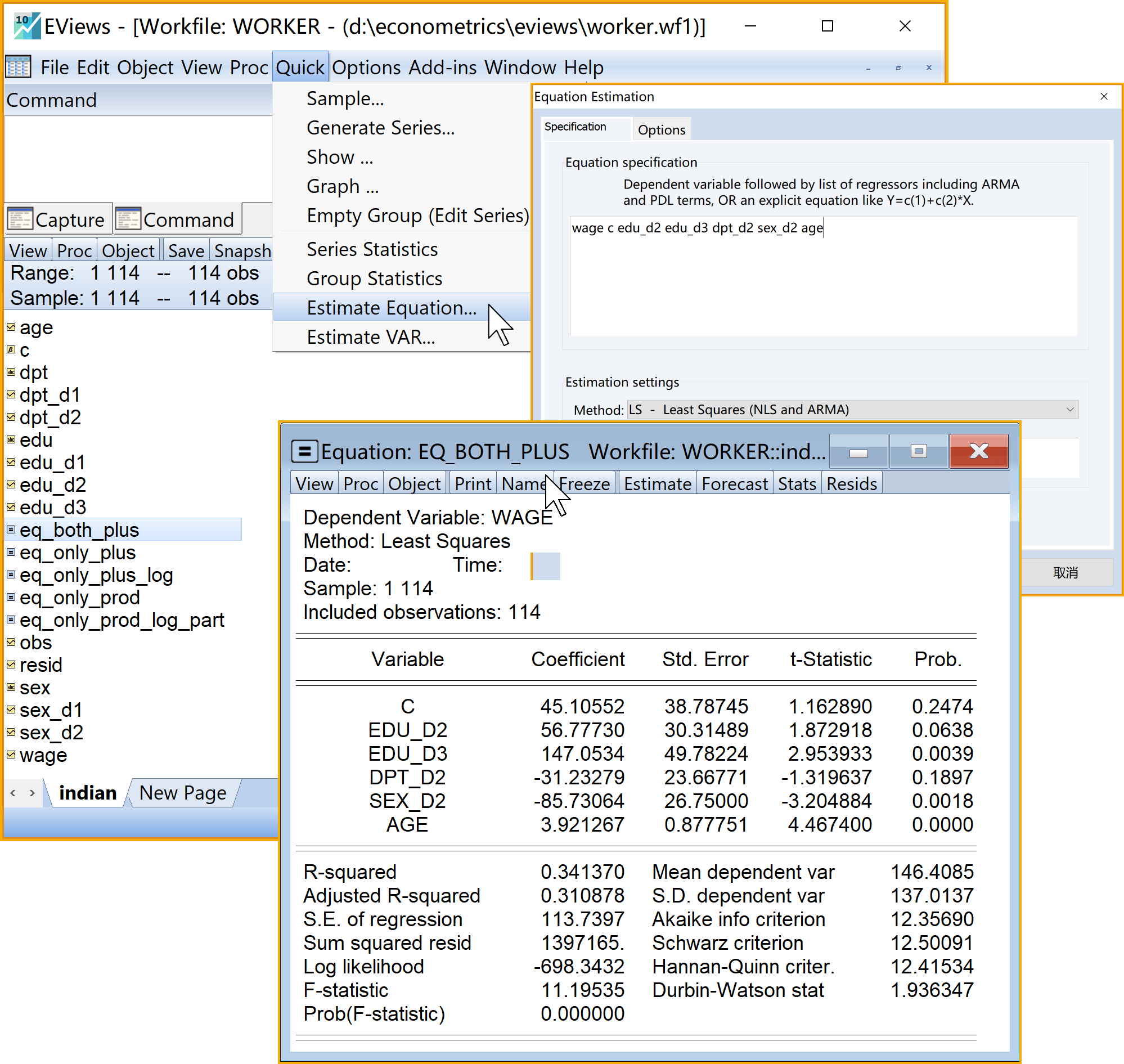


图13 同时含虚拟变量和定量变量的、加法形式的经典线性回归模型Eviews实现

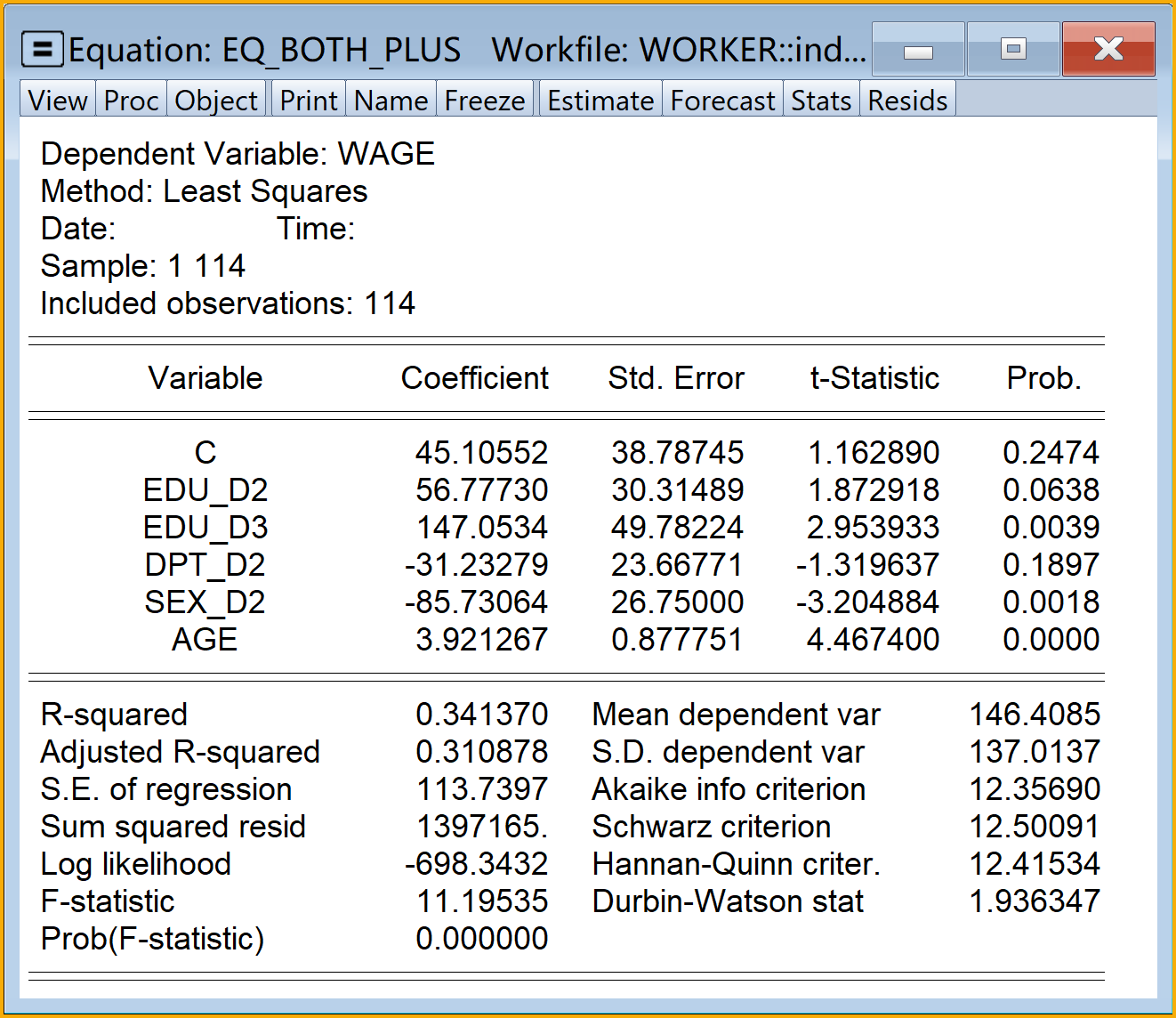


图14 同时含虚拟变量和定量变量的、加法形式的经典线性回归模型Eviews报告

* Eviews操作2（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(7)，菜单操作实现具体见图15）：
  1. 确定参照组为[**文盲&短期合同&女性**]，则如下虚拟变量将**不进入**回归模型
     1. edu\_d1
     2. dpt\_d1
     3. sex\_d1
  2. 设置回归模型。进入引导设置Equation Estimation specification
     1. Equation specification：输入命令log(wage) c edu\_d2 edu\_d3 edu\_d4 dpt\_d2 sex\_d2 age
     2. Estimation settings：
        + Method: 下拉选择LS - Least Squares (NLS and ARMA)
        + Sample: （默认设置）
     3. 点击完成：OK
     4. 命名保存方程对象：（建议命名为eq\_both\_plus\_log）
     5. 查看结果：双击eq\_both\_plus\_log

具体Eviews报告见16：

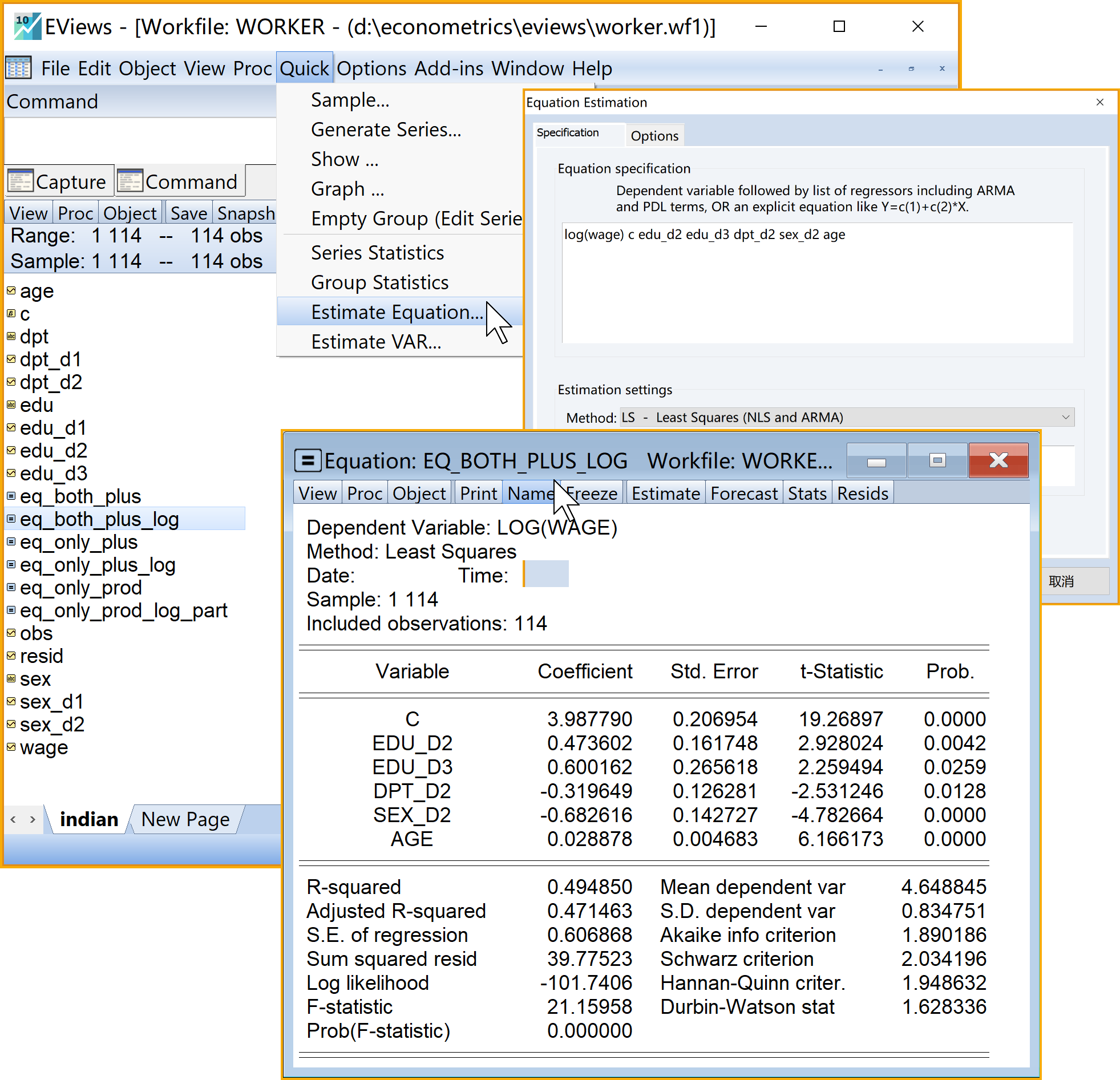


图15 同时含虚拟变量和定量变量的、加法形式的半对数线性回归模型Eviews实现

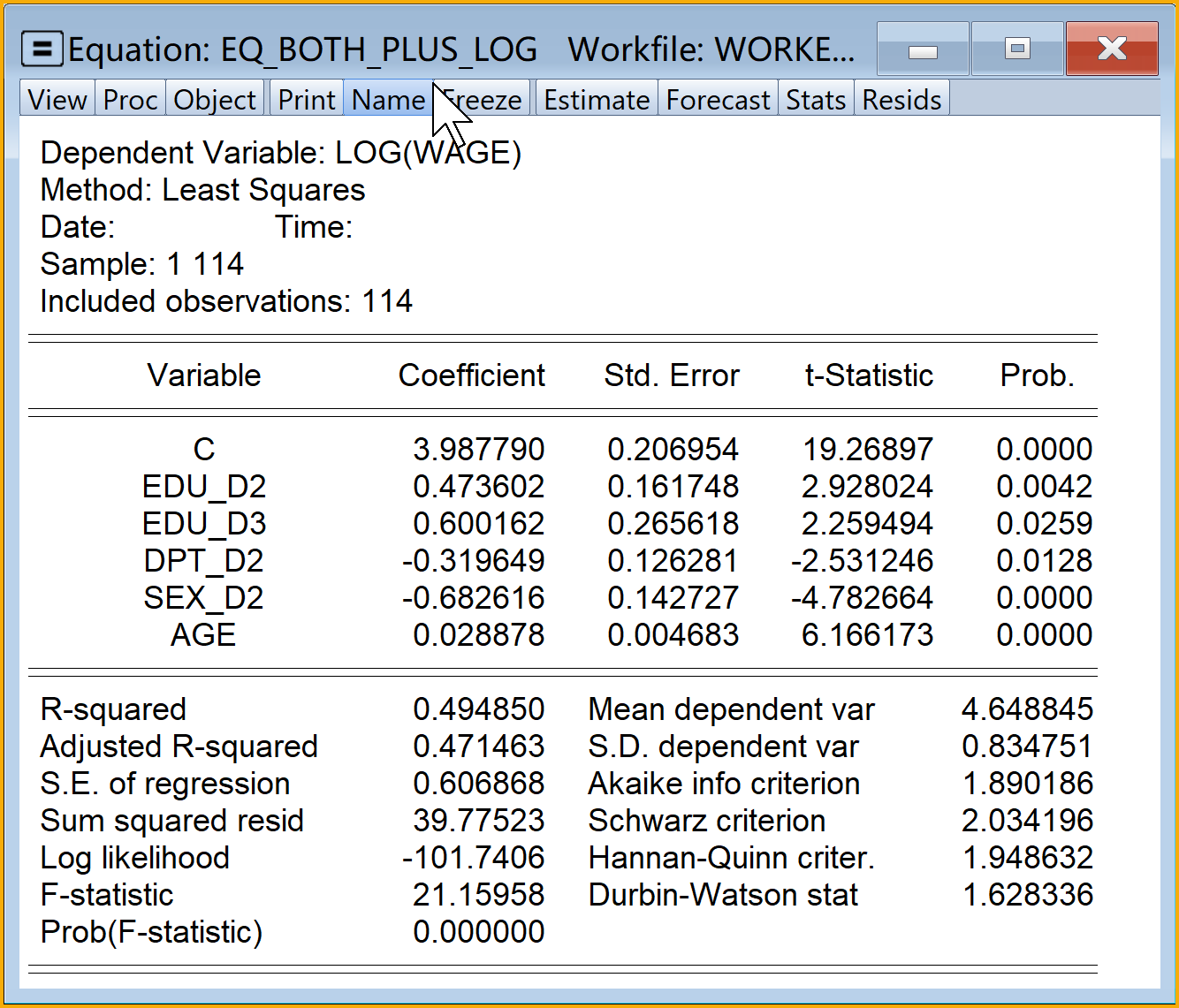


图16 同时含虚拟变量和定量变量的、加法形式的半对数线性回归模型Eviews报告

##### 报告解读

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{alignedat}{5} & =+&6.79+&23.96edu\_D2+&61.59edu\_D3+&150.49edu\_D4+&31.16dpt\_D2 \ & &(0.2130)&(0.7734)&(1.9867)&(3.0054)&(1.3141) \ & &(31.8931)&(30.9789)&(31.0035)&(50.0725)&(23.7120) \ & -&83.20sex\_D2+&3.99age \ & &(-3.0819)&(4.5129) \ & &(26.9981)&(0.8835) \ & & & R^2=0.3450& {R^2}=0.3083& F^{}=9.3945 \end{alignedat}

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{alignedat}{5} & =+&3.62+&0.16edu\_D2+&0.51edu\_D3+&0.62edu\_D4+&0.32dpt\_D2 \ & &(21.3135)&(0.9530)&(3.0586)&(2.3342)&(2.5264) \ & &(0.1699)&(0.1651)&(0.1652)&(0.2668)&(0.1263) \ & -&0.67sex\_D2+&0.03age \ & &(-4.6302)&(6.2268) \ & &(0.1438)&(0.0047) \ & & & R^2=0.4991& {R^2}=0.4710& F^{}=17.7694 \end{alignedat}

#### 乘法模型

* 理论提示：
  + 模型1：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(8)
  + 模型2：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(10)
* Eviews操作1（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(8)，菜单操作实现具体见图17）：
  1. 确定参照组为[**文盲&短期合同&女性**]，则如下虚拟变量将**不进入**回归模型
     1. edu\_d1
     2. dpt\_d1
     3. sex\_d1
  2. 设置回归模型。进入引导设置Equation Estimation specification
     1. 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
     2. Estimation settings：
        + Method: 下拉选择LS - Least Squares (NLS and ARMA)
        + Sample: （默认设置）
     3. 点击完成：OK
     4. 命名保存方程对象：（建议命名为eq\_both\_prod）
     5. 查看结果：双击eq\_both\_prod

具体Eviews报告见18：

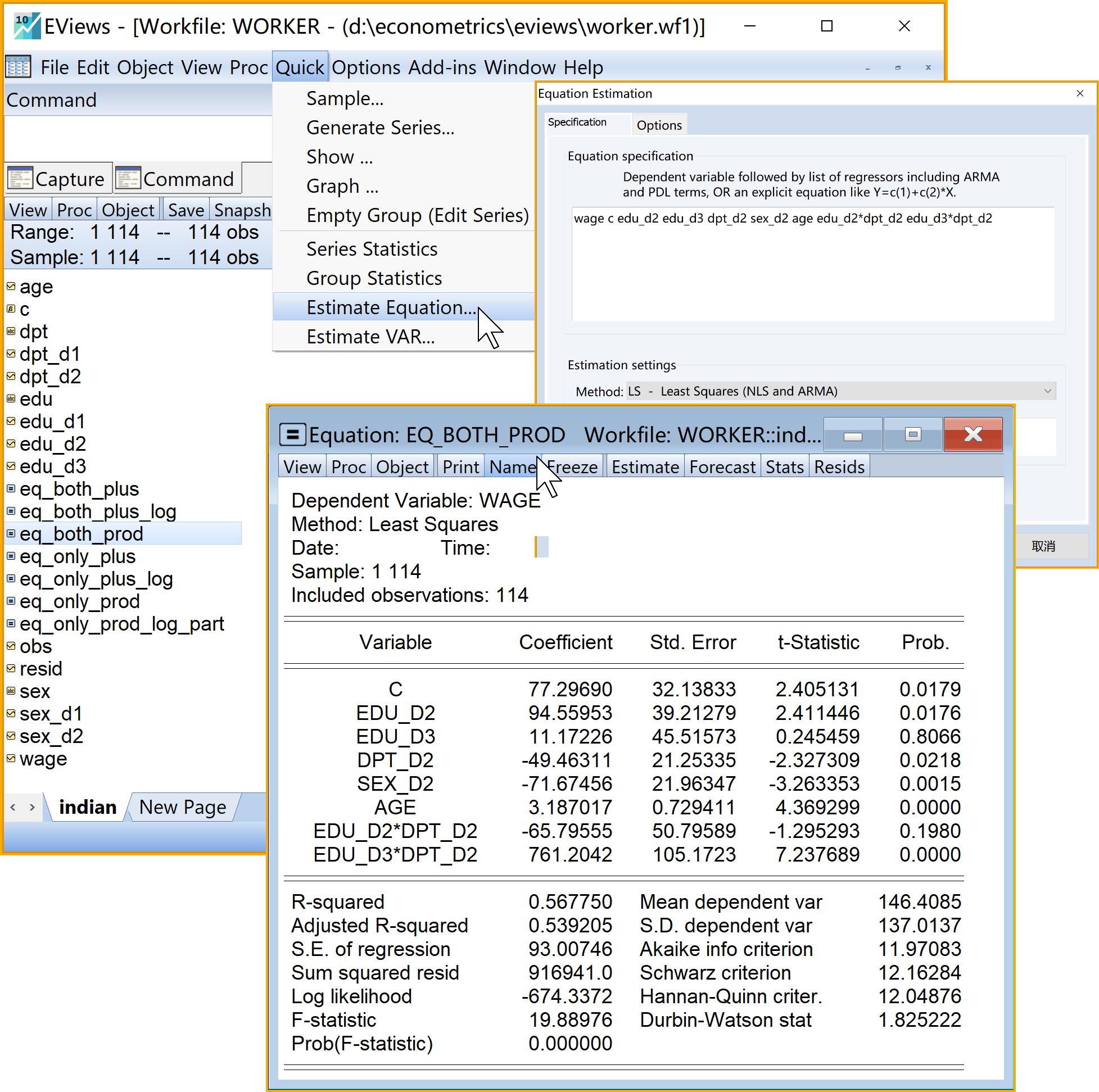


图17 同时含虚拟变量和定量变量的、乘法形式的经典线性回归模型Eviews实现

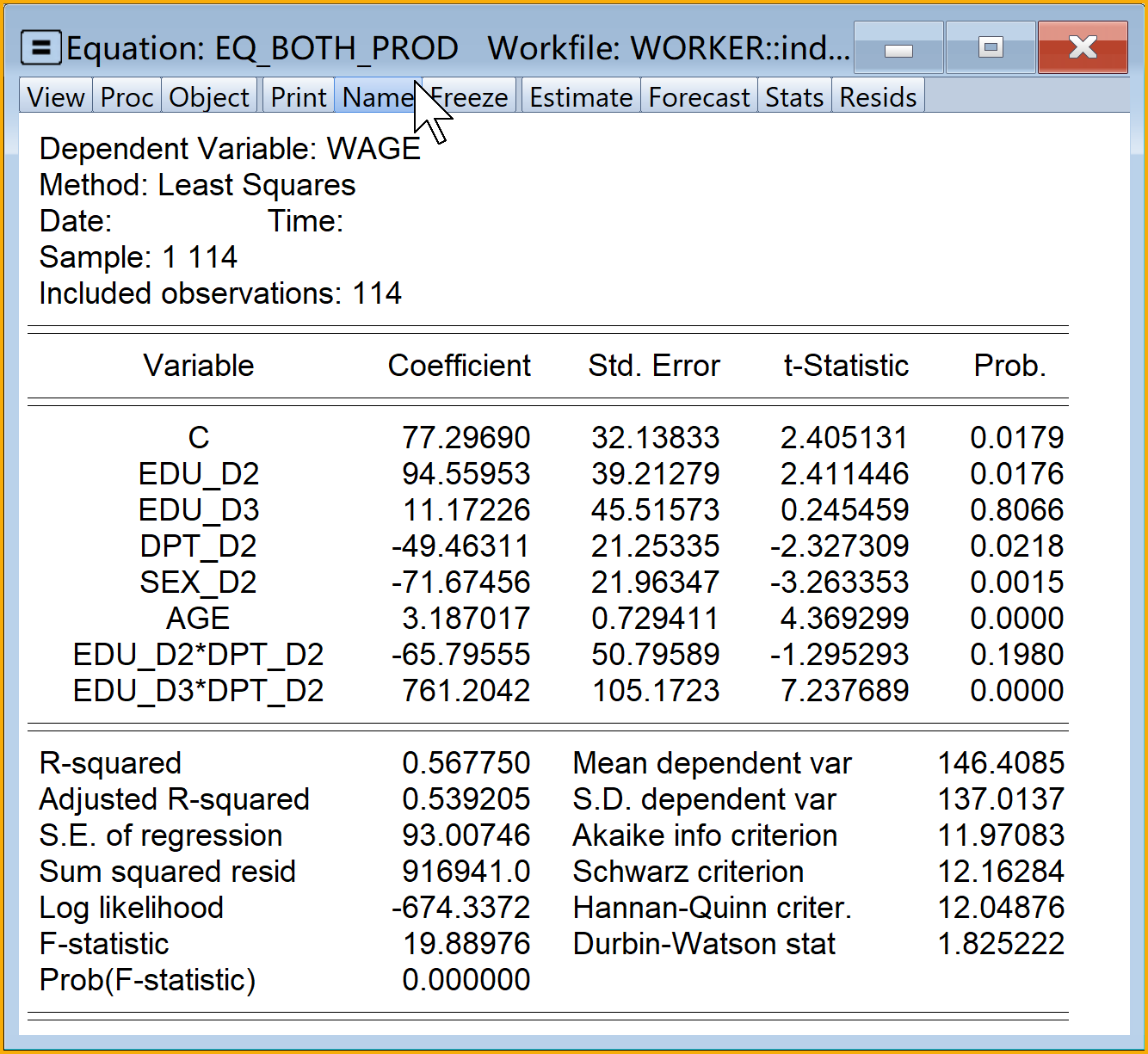


图18 同时含虚拟变量和定量变量的、乘法形式的经典线性回归模型Eviews报告

* Eviews操作2（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(10)，菜单操作实现具体见图19）：
  1. 确定参照组为[**文盲&短期合同&女性**]，则如下虚拟变量将**不进入**回归模型
     1. edu\_d1
     2. dpt\_d1
     3. sex\_d1
  2. 设置回归模型。进入引导设置Equation Estimation specification
     1. 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
     2. Estimation settings：
        + Method: 下拉选择LS - Least Squares (NLS and ARMA)
        + Sample: （默认设置）
     3. 点击完成：OK
     4. 命名保存方程对象：（建议命名为eq\_both\_prod\_log\_part）
     5. 查看结果：双击eq\_both\_prod\_log\_part

具体Eviews报告见20：

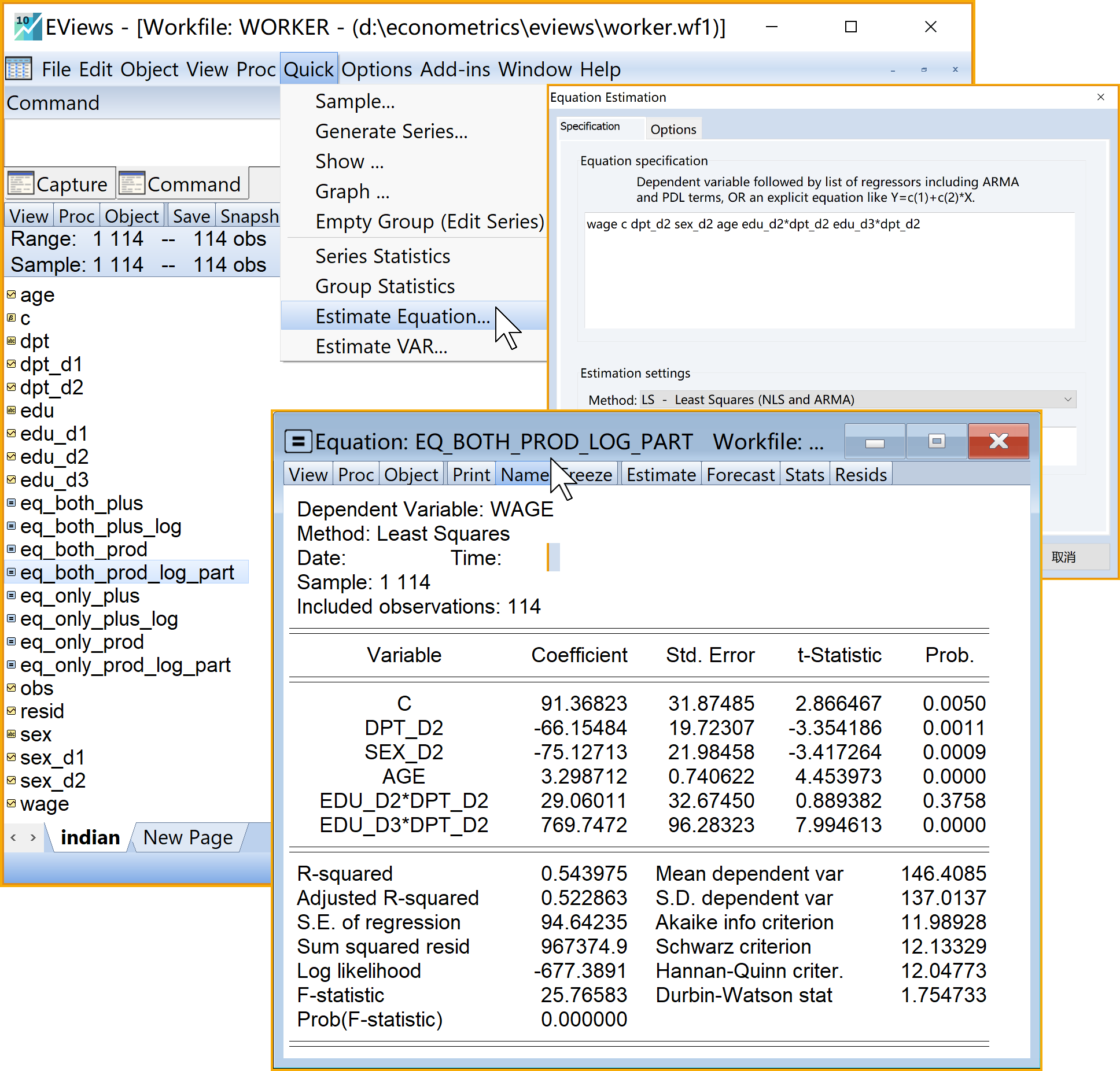


图19 同时含虚拟变量和定量变量的、乘法形式的半对数线性回归模型Eviews实现

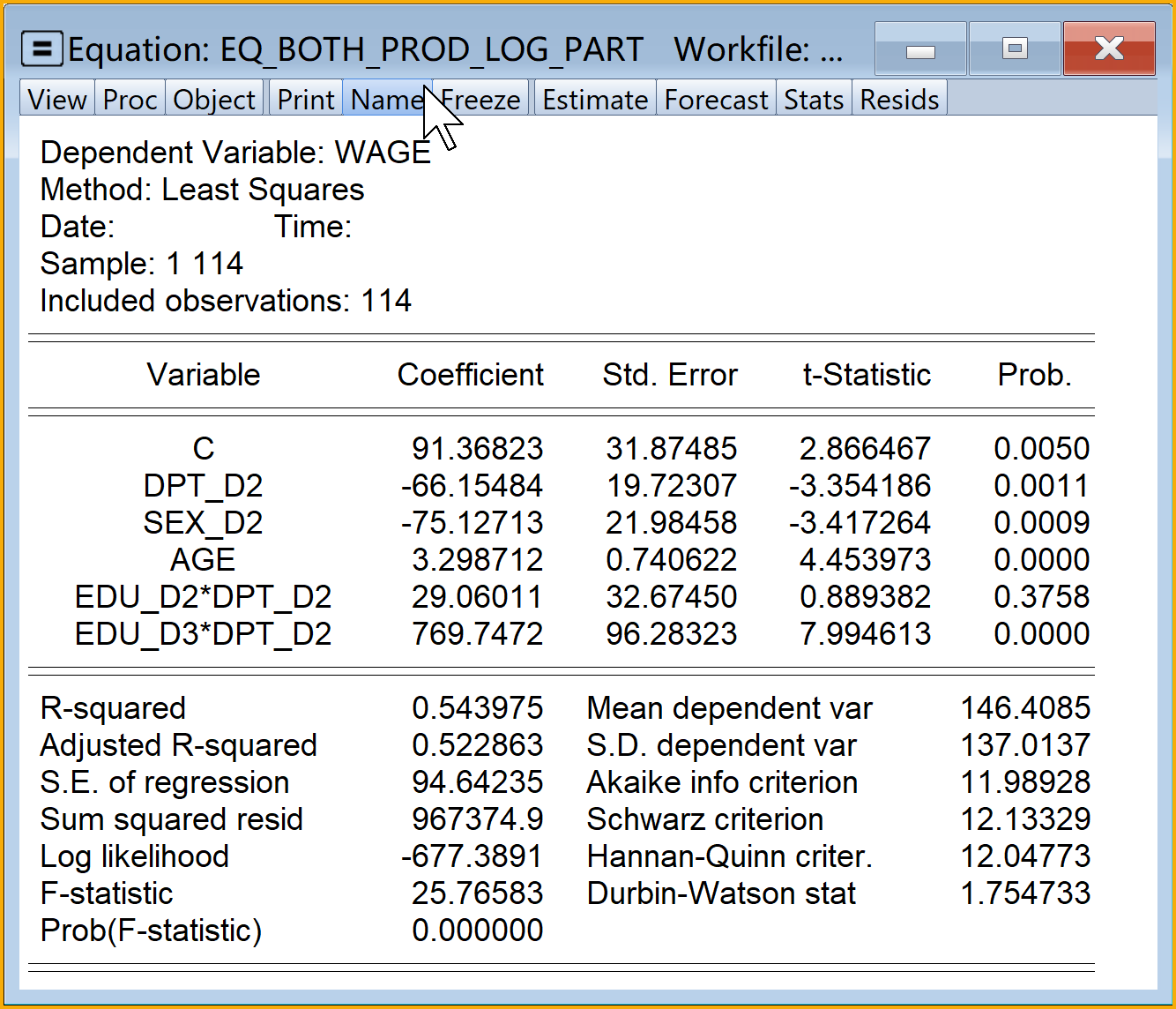


图20 同时含虚拟变量和定量变量的、乘法形式的半对数线性回归模型Eviews报告

##### 报告解读

同时含虚拟变量和定量变量的、乘法形式的经典回归模型（全部变量进入）Eviews结果简要报告如下：

\begin{alignedat}{5} & =+&19.43+&32.29edu\_D2+&35.59edu\_D3+&777.93edu\_D4+&54.46dpt\_D2 \ & &(0.7215)&(1.0706)&(1.0830)&(8.1760)&(2.3131) \ & &(26.9248)&(30.1607)&(32.8591)&(95.1488)&(23.5466) \ & -&70.34sex\_D2+&3.25age-&26.81edu\_D2:dpt\_D2+&59.80edu\_D3:dpt\_D2-&766.28edu\_D4:dpt\_D2 \ & &(-3.1451)&(4.4121)&(-0.4864)&(1.1521)&(-7.2326) \ & &(22.3647)&(0.7361)&(55.1151)&(51.9083)&(105.9477) \ & & & R^2=0.5725& {R^2}=0.5355& F^{}=15.4756 \end{alignedat}

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结果简要报告如下：

\begin{alignedat}{5} & =+&19.43+&54.46dpt\_D2-&70.34sex\_D2+&3.25age+&32.29edu\_D2 \ & &(0.7215)&(2.3131)&(-3.1451)&(4.4121)&(1.0706) \ & &(26.9248)&(23.5466)&(22.3647)&(0.7361)&(30.1607) \ & +&35.59edu\_D3+&777.93edu\_D4-&26.81dpt\_D2:edu\_D2+&59.80dpt\_D2:edu\_D3-&766.28dpt\_D2:edu\_D4 \ & &(1.0830)&(8.1760)&(-0.4864)&(1.1521)&(-7.2326) \ & &(32.8591)&(95.1488)&(55.1151)&(51.9083)&(105.9477) \ & & & R^2=0.5725& {R^2}=0.5355& F^{}=15.4756 \end{alignedat}

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结果简要报告如下： \begin{alignedat}{5} & =+&3.65+&0.35dpt\_D2-&0.63sex\_D2+&0.03age+&0.14edu\_D2 \ & &(21.0939)&(2.3114)&(-4.4157)&(5.8925)&(0.7352) \ & &(0.1730)&(0.1513)&(0.1437)&(0.0047)&(0.1938) \ & +&0.45edu\_D3+&1.90edu\_D4+&0.05dpt\_D2:edu\_D2+&0.12dpt\_D2:edu\_D3-&1.55dpt\_D2:edu\_D4 \ & &(2.1475)&(3.1022)&(0.1519)&(0.3711)&(-2.2738) \ & &(0.2112)&(0.6114)&(0.3542)&(0.3336)&(0.6808) \ & & & R^2=0.5244& {R^2}=0.4833& F^{}=12.7417 \end{alignedat}

### 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：
  + 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。
* 理论提示：
  + 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(11)和无基础组的时间序列季节虚拟变量模型(12)。其中，为定量变量。
* Eviews操作（略）

### 分段线性回归模型（piecewise linear regression）

* 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：
  + **分段**现象：在经济关系中，当解释变量的值达到某一水平/阀值之前，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/阀值的解释变量对被解释变量的影响的一类线性回归模型。
* 理论提示：
  + 一个阀值的分段线性回归模型(13)
  + 两个阀值的分段线性回归模型(14)
* Eviews操作（略）

## 作业题

**印度工人工资**：表5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5 印度工人工资(n=114)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| obs | wage | age | edu | dpt | sex |
| 1 | 117.00 | 26 | primary | permanent | female |
| 2 | 375.00 | 42 | primary | permanent | female |
| 3 | 175.00 | 33 | primary | permanent | female |
| 4 | 100.00 | 33 | primary | permanent | female |
| 5 | 162.50 | 30 | primary | permanent | female |
| 110 | 25.00 | 18 | illiteracy | temporary | male |
| 111 | 25.00 | 11 | illiteracy | temporary | male |
| 112 | 75.00 | 45 | illiteracy | temporary | male |
| 113 | 53.84 | 14 | illiteracy | temporary | male |
| 114 | 50.00 | 26 | illiteracy | temporary | male |

变量说明见表6：

表6 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 |

请回答如下问题:

1. 请对数据表5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6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。
   1. 请写出加法形式的有截距虚拟变量回归模型（PRM）
   2. 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）
   3. 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
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。
   1. 请写出该乘法形式的有截距虚拟变量回归模型（PRM）
   2. 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）
   3. 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
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。
   1. 请写出该乘法形式的有截距虚拟变量回归模型（PRM）
   2. 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?
      * 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。
   3. 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。
   1. 请写出该乘法形式的有截距虚拟变量回归模型（PRM）
   2. 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?
      * 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。
   3. 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

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，也可以用个虚拟变量**充分表达**该定性变量。 [↑](#footnote-ref-27)
2. **方差分析模型**（Analysis of variance, ANOVA）常用来分析定量化的因变量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 [↑](#footnote-ref-28)
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）。 [↑](#footnote-ref-29)
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。 [↑](#footnote-ref-30)
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。 [↑](#footnote-ref-31)
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**） [↑](#footnote-ref-38)
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”。 [↑](#footnote-ref-39)