Rapid formation of regression tables for research purposes

Roy Wada

UCLA/RAND

October 2007

Reasons for Regression Tables

- Regression tables are apparently produced for journal publication
  - "formats it as in journal articles "

    John Gallup's –outreg–
  - "pretty looking publication-style regression tables"

    Ben Jann's –estout–
  - "publication quality tables"

    Ian Watson –tabout–

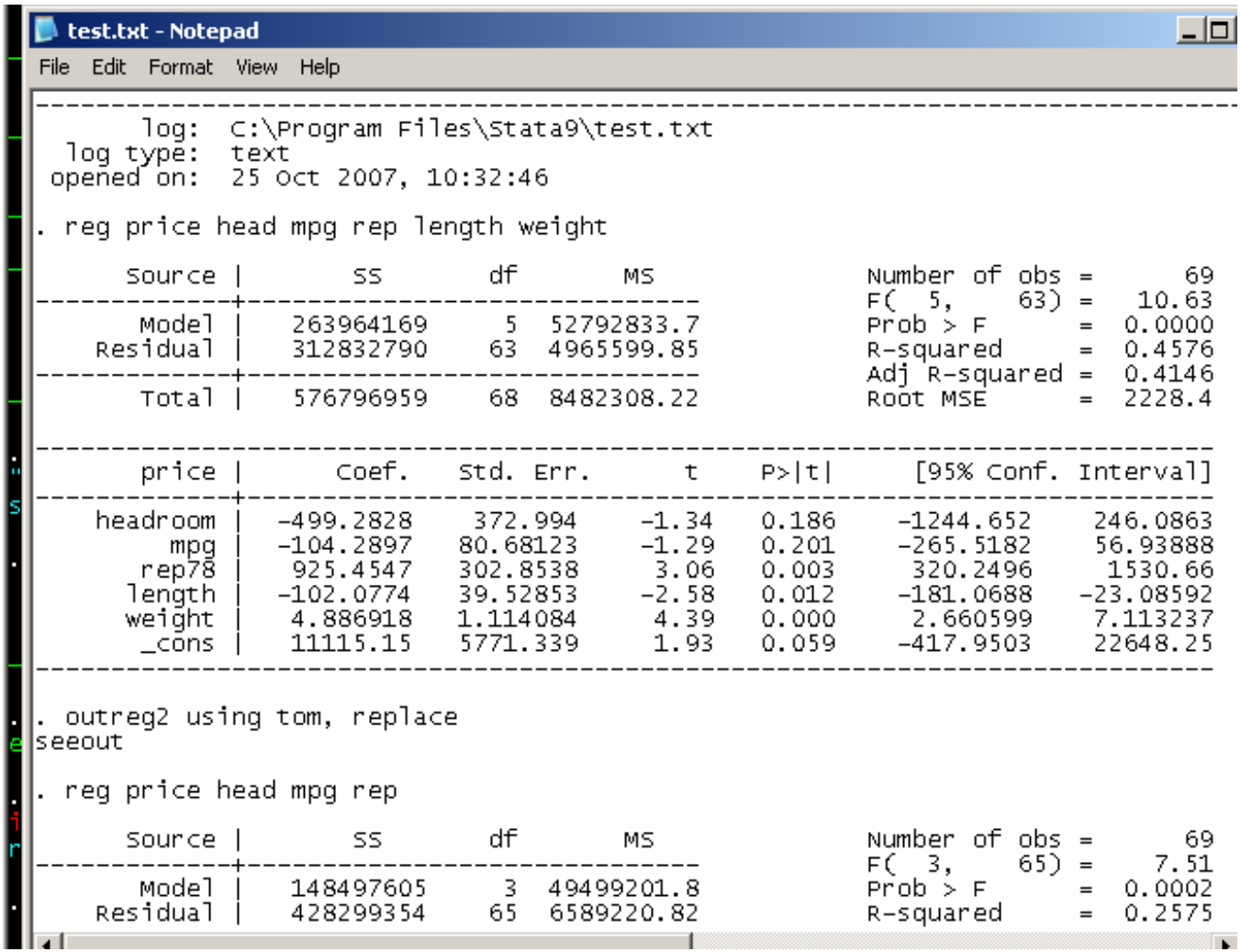
Reasons for Regression Tables

- How many regression tables do you publish?
- How often do you publish anyway?
- Why does it attract so much attention from Stata users?

Reasons for Regression Tables

- Log files are serially produced
- Researchers have traditionally printed log files and flipped them back-and-forth

Example of Log File



test.txt - Notepad

File Edit Format View Help

```
-------------------------------------------------------------------------------
      log:  C:\Program Files\Stata9\test.txt
  log type:  text
 opened on:  25 Oct 2007, 10:32:46

. reg price head mpg rep length weight

      Source |       SS       df       MS              Number of obs =      69
-------------+------------------------------           F(  5,    63) =   10.63
       Model |   263964169     5  52792833.7            Prob > F      =  0.0000
    Residual |   312832790    63  4965599.85            R-squared     =  0.4576
-------------+------------------------------           Adj R-squared =  0.4146
       Total |   576796959    68  8482308.22            Root MSE      =  2228.4

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    headroom |  -499.2828    372.994    -1.34   0.186    -1244.652    246.0863
         mpg |  -104.2897   80.68123    -1.29   0.201    -265.5182    56.93888
       rep78 |   925.4547   302.8538     3.06   0.003     320.2496     1530.66
      length |  -102.0774   39.52853    -2.58   0.012    -181.0688   -23.08592
      weight |   4.886918   1.114084     4.39   0.000     2.660599    7.113237
       _cons |   11115.15   5771.339     1.93   0.059    -417.9503    22648.25
------------------------------------------------------------------------------

. outreg2 using tom, replace
seeout

. reg price head mpg rep

      Source |       SS       df       MS              Number of obs =      69
-------------+------------------------------           F(  3,    65) =    7.51
       Model |   148497605     3  49499201.8            Prob > F      =  0.0002
    Residual |   428299354    65  6589220.82            R-squared     =  0.2575
```

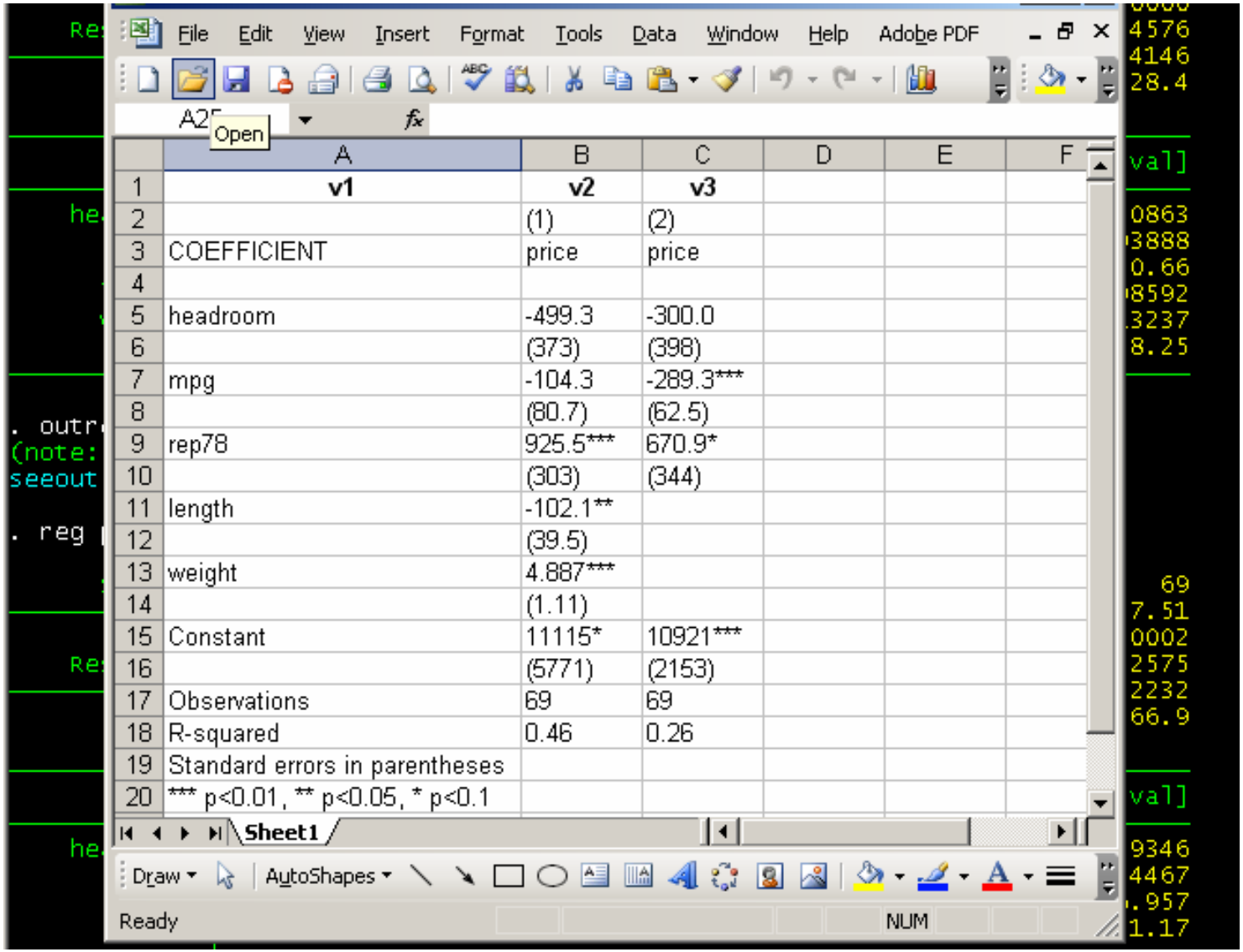
Reasons for Regression Tables

- Regression tables facilitates analysis by placing outputs next to each other
  - Robustness of results
  - (Mis)specification check
  - Testing for collinearity

Regression Output: –reg–

```
. reg price head mpg rep length weight

      Source |       SS       df       MS              Number of obs =      69
-------------+------------------------------           F(  5,    63) =   10.63
       Model |  263964169      5  52792833.7            Prob > F      =  0.0000
    Residual |  312832790     63  4965599.85            R-squared     =  0.4576
-------------+------------------------------           Adj R-squared =  0.4146
       Total |  576796959     68  8482308.22            Root MSE      =  2228.4

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    headroom |  -499.2828    372.994    -1.34   0.186    -1244.652    246.0863
         mpg |  -104.2897   80.68123    -1.29   0.201    -265.5182    56.93888
       rep78 |   925.4547   302.8538     3.06   0.003     320.2496     1530.66
      length |  -102.0774   39.52853    -2.58   0.012    -181.0688   -23.08592
      weight |   4.886918   1.114084     4.39   0.000     2.660599    7.113237
       _cons |   11115.15   5771.339     1.93   0.059    -417.9503    22648.25
------------------------------------------------------------------------------

. reg price head mpg rep

      Source |       SS       df       MS              Number of obs =      69
-------------+------------------------------           F(  3,    65) =    7.51
       Model |  148497605      3  49499201.8            Prob > F      =  0.0002
    Residual |  428299354     65  6589220.82            R-squared     =  0.2575
-------------+------------------------------           Adj R-squared =  0.2232
       Total |  576796959     68  8482308.22            Root MSE      =  2566.9

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    headroom |  -300.0293   398.0516    -0.75   0.454    -1094.993    494.9346
         mpg |  -289.3462   62.53921    -4.63   0.000    -414.2456   -164.4467
       rep78 |   670.8971   343.5213     1.95   0.055    -15.16242    1356.957
       _cons |   10921.33   2153.003     5.07   0.000     6621.487    15221.17
```

Regression Table: -shellout-



	A	B	C	D	E	F
1	**v1**	**v2**	**v3**			
2		(1)	(2)			
3	COEFFICIENT	price	price			
4						
5	headroom	-499.3	-300.0			
6		(373)	(398)			
7	mpg	-104.3	-289.3***			
8		(80.7)	(62.5)			
9	rep78	925.5***	670.9*			
10		(303)	(344)			
11	length	-102.1**				
12		(39.5)				
13	weight	4.887***				
14		(1.11)				
15	Constant	11115*	10921***			
16		(5771)	(2153)			
17	Observations	69	69			
18	R-squared	0.46	0.26			
19	Standard errors in parentheses					
20	*** p<0.01, ** p<0.05, * p<0.1					

Regression Output: –reg3–

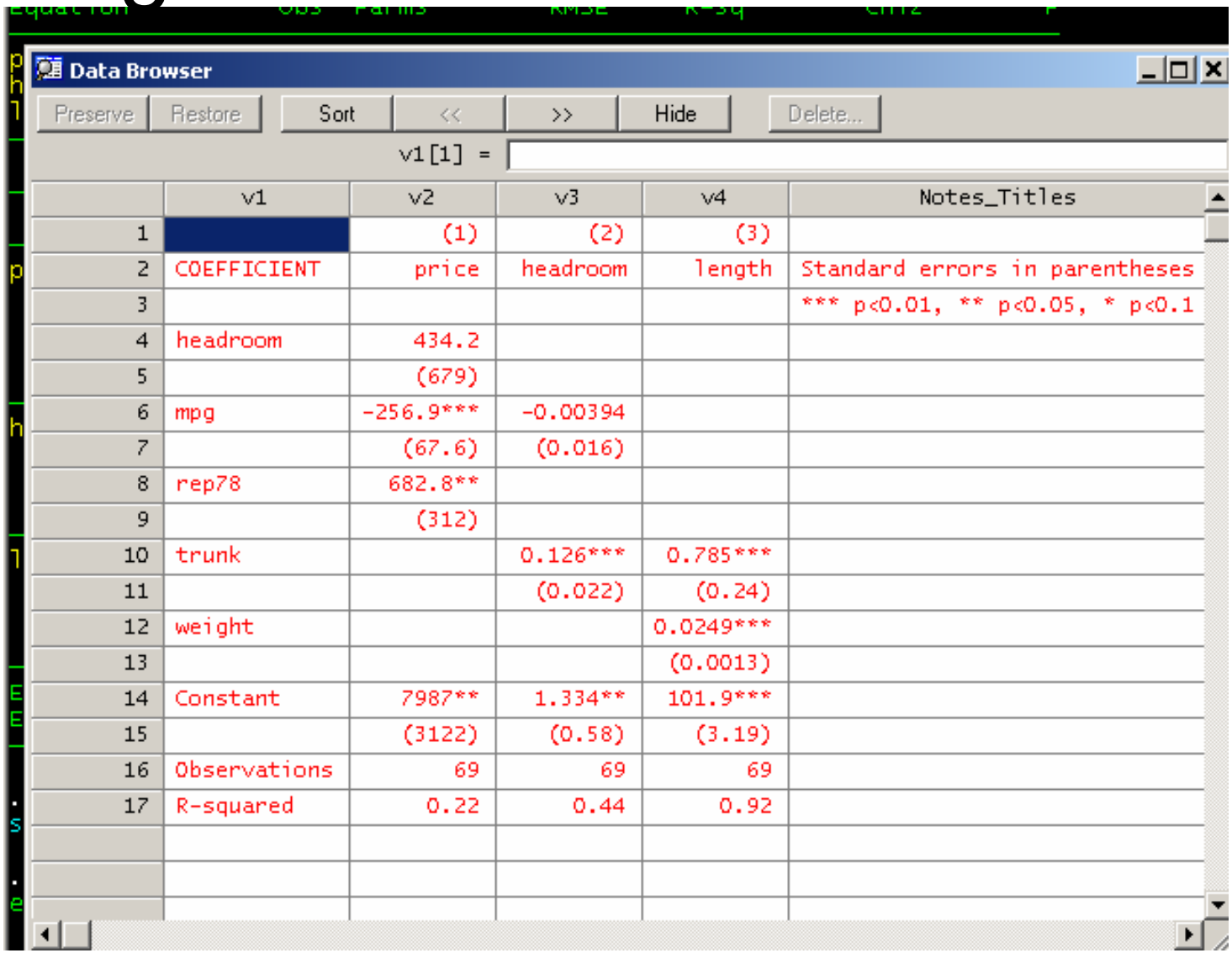
```
. do "C:\DOCUME~1\Me\LOCALS~1\Temp\STD04000000.tmp"

. reg3 (price=head mpg rep) (head=trunk mpg) (length=trunk weight)

Three-stage least-squares regression
```

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
price	69	3	2556.461	0.2182	25.71	0.0000
headroom	69	2	.6355204	0.4370	52.67	0.0000
length	69	2	6.580633	0.9151	797.13	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
price						
headroom	434.165	679.224	0.64	0.523	-897.0896	1765.42
mpg	-256.8845	67.55487	-3.80	0.000	-389.2896	-124.4794
rep78	682.7653	311.8087	2.19	0.029	71.63147	1293.899
_cons	7987.222	3122.007	2.56	0.011	1868.202	14106.24
headroom						
trunk	.1256041	.0217683	5.77	0.000	.082939	.1682693
mpg	-.0039376	.0161147	-0.24	0.807	-.0355218	.0276465
_cons	1.334476	.5796718	2.30	0.021	.1983401	2.470612
length						
trunk	.7846249	.2415988	3.25	0.001	.3111	1.25815
weight	.0248736	.0013024	19.10	0.000	.0223208	.0274263
_cons	101.9446	3.187924	31.98	0.000	95.69635	108.1928

```
Endogenous variables:  price headroom length
Exogenous variables:   mpg rep78 trunk weight
```

Regression Table: –seeout–



Data Browser

Preserve | Restore | Sort | << | >> | Hide | Delete...

v1[1] =

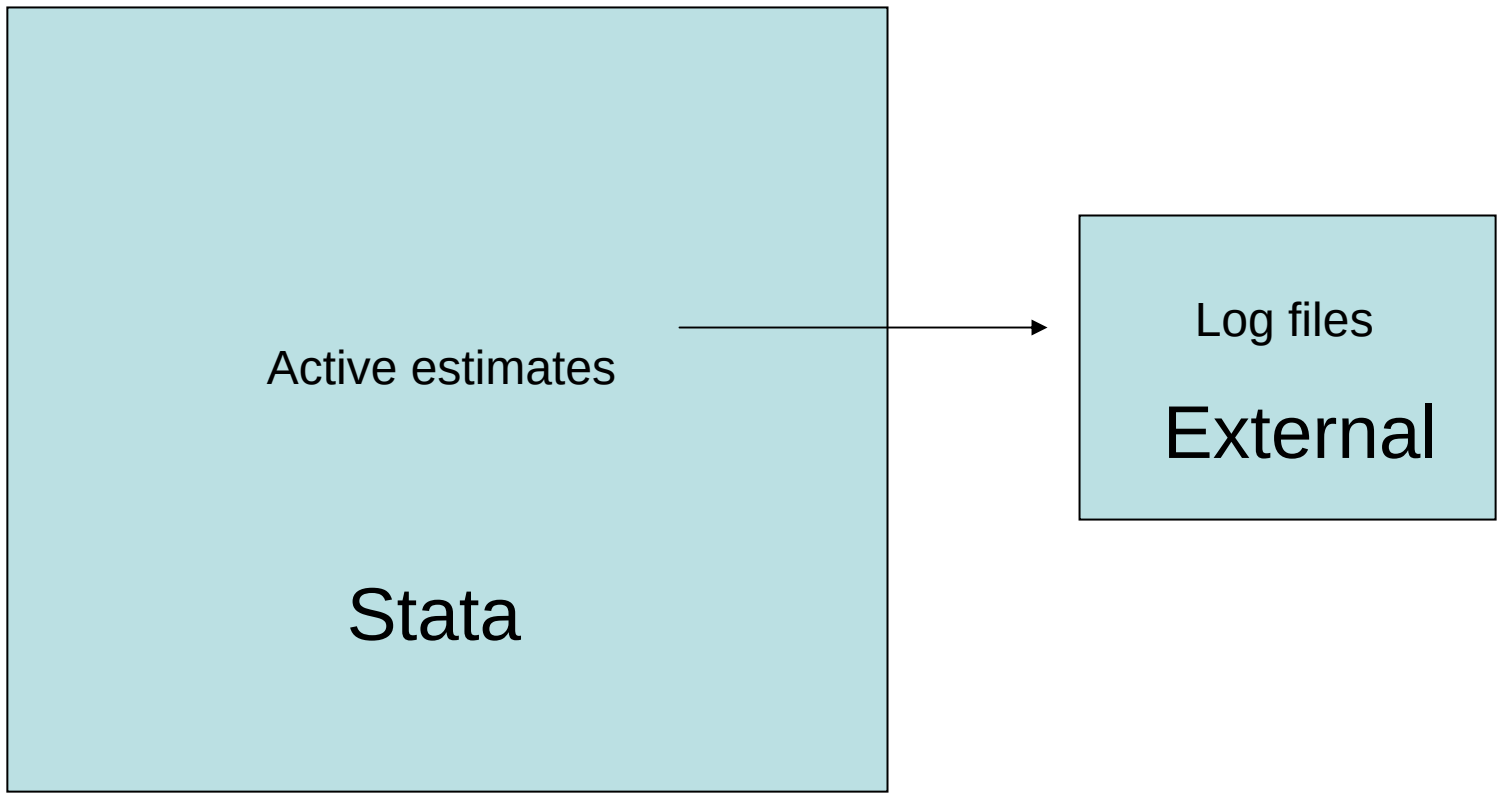
	v1	v2	v3	v4	Notes_Titles
1		(1)	(2)	(3)	
2	COEFFICIENT	price	headroom	length	Standard errors in parentheses
3					*** p<0.01, ** p<0.05, * p<0.1
4	headroom	434.2			
5		(679)			
6	mpg	-256.9***	-0.00394		
7		(67.6)	(0.016)		
8	rep78	682.8**			
9		(312)			
10	trunk		0.126***	0.785***	
11			(0.022)	(0.24)	
12	weight			0.0249***	
13				(0.0013)	
14	Constant	7987**	1.334**	101.9***	
15		(3122)	(0.58)	(3.19)	
16	Observations	69	69	69	
17	R-squared	0.22	0.44	0.92	

Regression Tables as Information Management tool

- Improperly formatted Information is less useful than properly formatted one
- Management aspects of information with respect to regression tables
  - Storage
  - Interface (delivery to end-user)
    - Access
    - Display

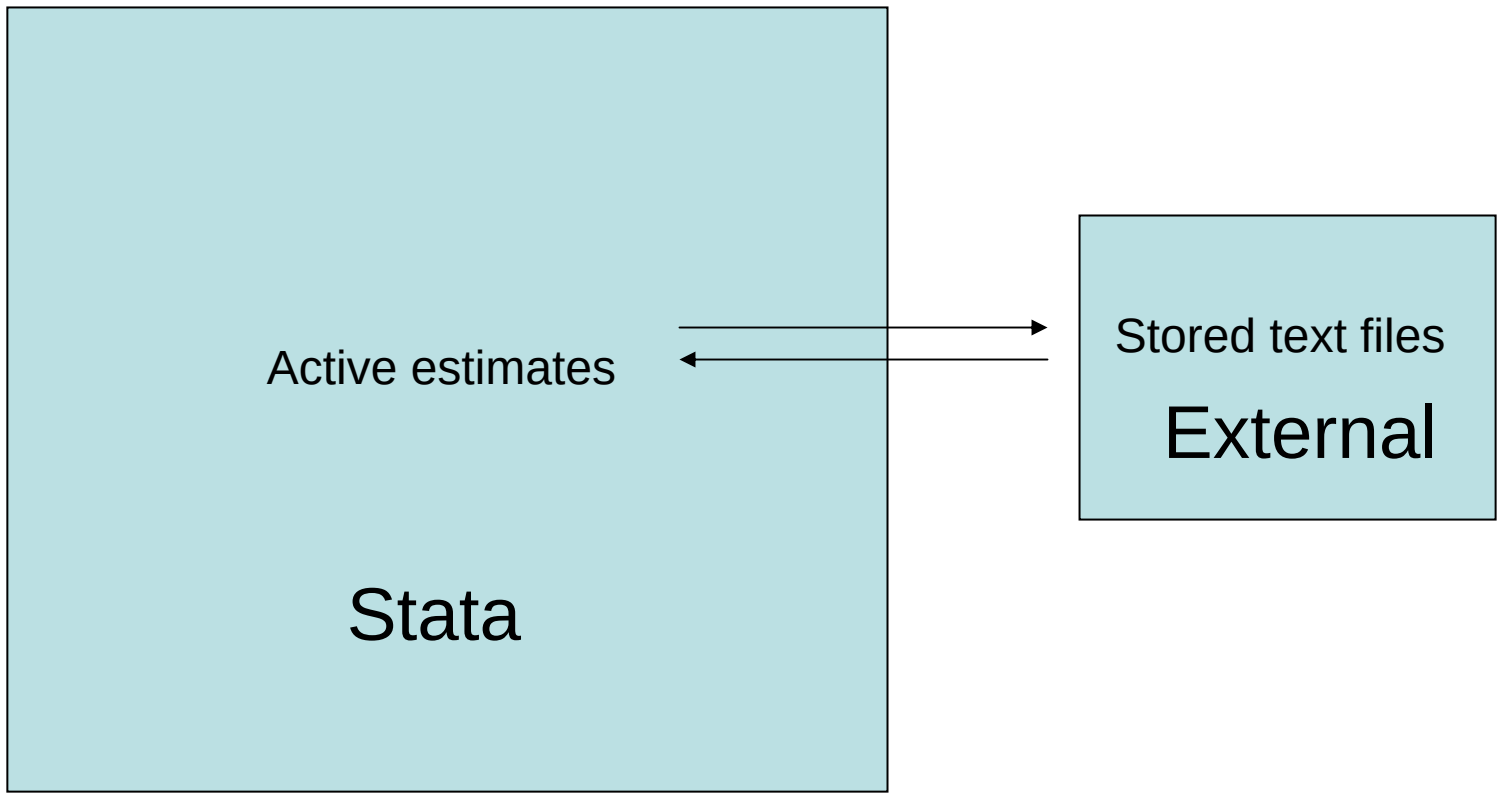
Regression Tables: Storage Aspect

- Traditional log files



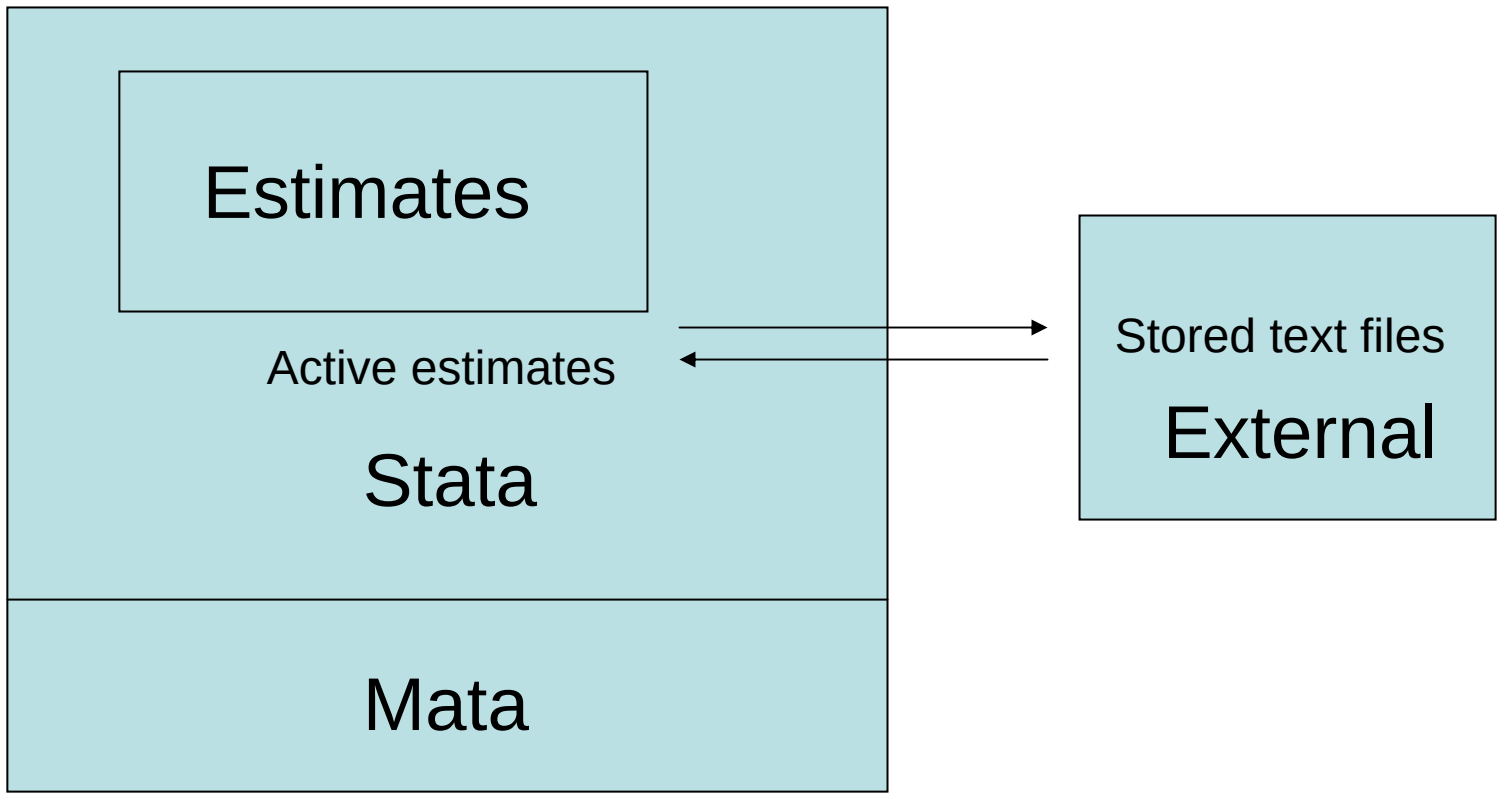
Regression Tables: Storage Aspect

- John Gallup's -outreg-



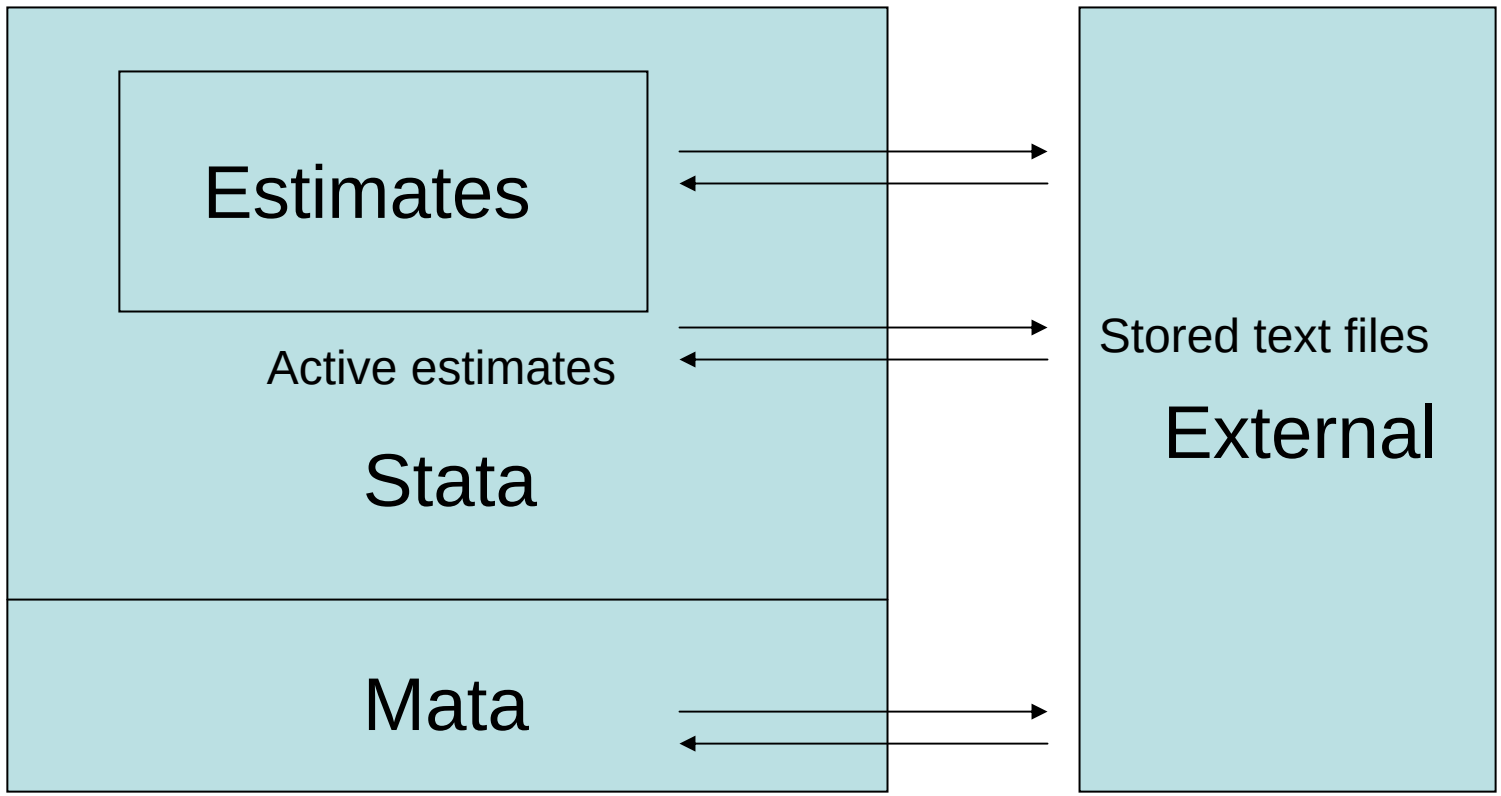
Regression Tables: Storage Aspect

- Estimates and Mata (post-Stata 8/Stata 9)



Regression Tables: Storage Aspect

- Estimates and Mata (post-Stata 8/Stata 9)



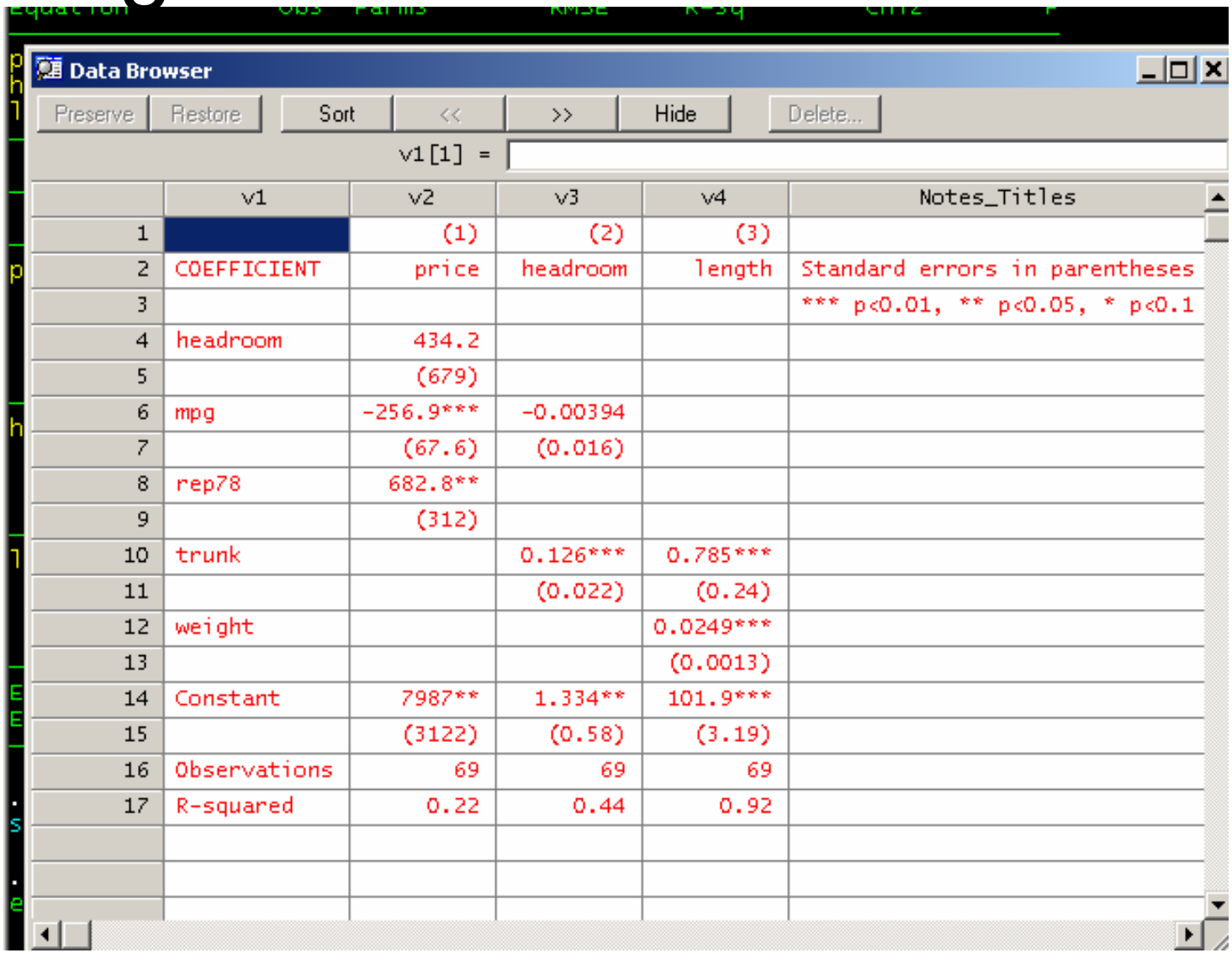
Regression Tables: Interface Aspect

- Display
  - Wrapping text problem
  - Publication-quality tables are not always the best for analysis (finger pointing)
- Access
  - Automated (least involvement by users)
  - External compatibility

Regression Tables: -outreg2- based on -outreg-

- Display
  - Automatically formatted display (digits, stars)
  - Spreadsheet format in -seeout- and -shellout-
- Access
  - Single command driven
  - One click hypertext
  - Excel, Word, LaTex, Text
  - Prefix command and stored options

Regression Table: –seeout–



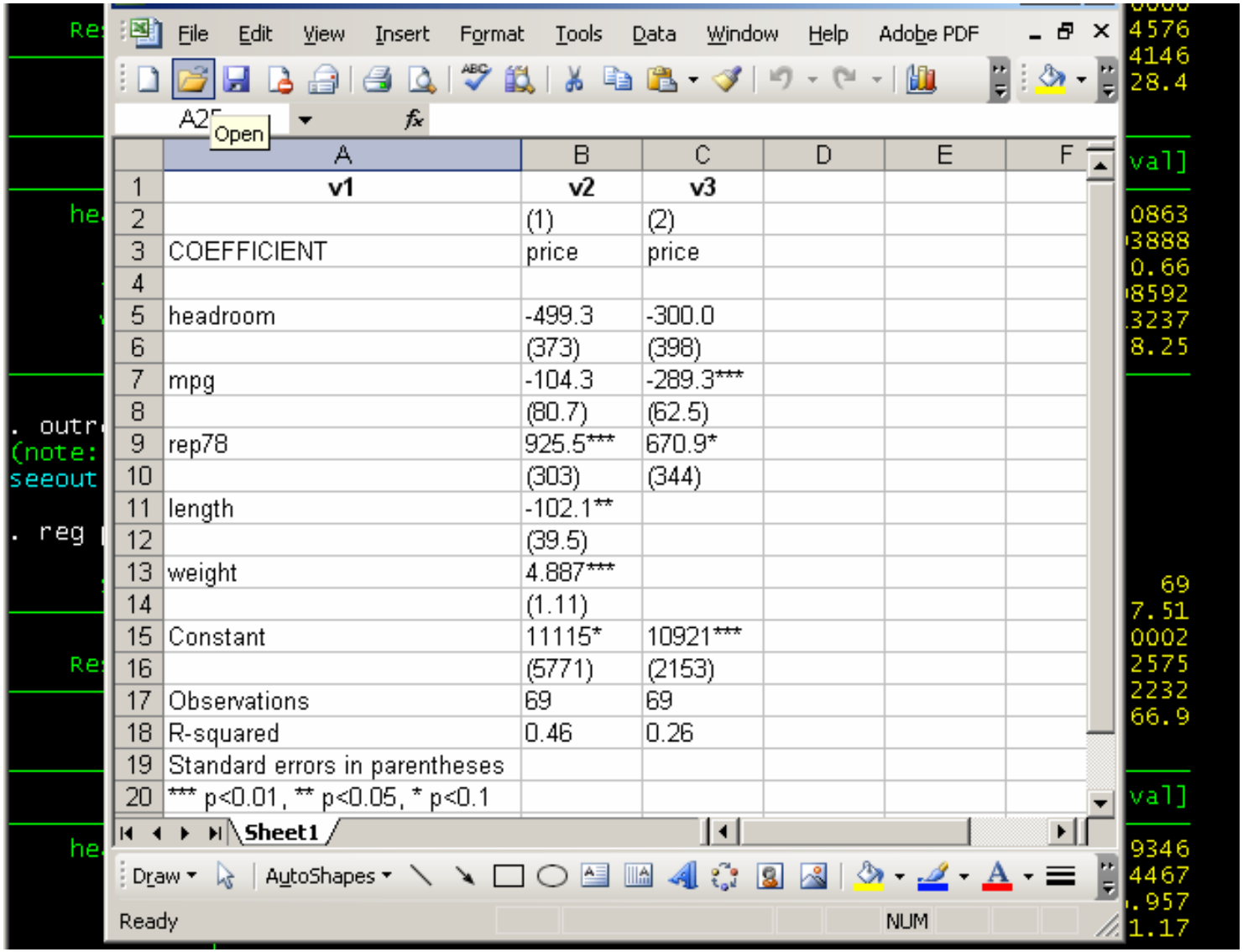
Data Browser

Preserve | Restore | Sort | << | >> | Hide | Delete...

v1[1] =

	v1	v2	v3	v4	Notes_Titles
1		(1)	(2)	(3)	
2	COEFFICIENT	price	headroom	length	Standard errors in parentheses
3					*** p<0.01, ** p<0.05, * p<0.1
4	headroom	434.2			
5		(679)			
6	mpg	-256.9***	-0.00394		
7		(67.6)	(0.016)		
8	rep78	682.8**			
9		(312)			
10	trunk		0.126***	0.785***	
11			(0.022)	(0.24)	
12	weight			0.0249***	
13				(0.0013)	
14	Constant	7987**	1.334**	101.9***	
15		(3122)	(0.58)	(3.19)	
16	Observations	69	69	69	
17	R-squared	0.22	0.44	0.92	

Regression Table: -shellout-



	A	B	C	D	E	F
1	**v1**	**v2**	**v3**			
2		(1)	(2)			
3	COEFFICIENT	price	price			
4						
5	headroom	-499.3	-300.0			
6		(373)	(398)			
7	mpg	-104.3	-289.3***			
8		(80.7)	(62.5)			
9	rep78	925.5***	670.9*			
10		(303)	(344)			
11	length	-102.1**				
12		(39.5)				
13	weight	4.887***				
14		(1.11)				
15	Constant	11115*	10921***			
16		(5771)	(2153)			
17	Observations	69	69			
18	R-squared	0.46	0.26			
19	Standard errors in parentheses					
20	*** p<0.01, ** p<0.05, * p<0.1					

Regression Tables: -outreg2-

```
. shellout using tom.xml

.
end of do-file

. do "C:\DOCUME~1\Me\LOCALS~1\Temp\STD04000000.tmp"

. reg price head mpg rep
```

Source	SS	df	MS
Model	148497605	3	49499201.8
Residual	428299354	65	6589220.82
Total	576796959	68	8482308.22

Number of obs	69
F(3, 65)	7.51
Prob > F	0.0002
R-squared	0.2575
Adj R-squared	0.2232
Root MSE	2566.9

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
headroom	-300.0293	398.0516	-0.75	0.454	-1094.993	494.9346
mpg	-289.3462	62.53921	-4.63	0.000	-414.2456	-164.4467
rep78	670.8971	343.5213	1.95	0.055	-15.16242	1356.957
_cons	10921.33	2153.003	5.07	0.000	6621.487	15221.17

```
. outreg2 using tom, excel
"tom.xml"
seeout

. seeout using tom
  seeout

. shellout using tom.xml
```

Regression Tables: Future Development

- Rethinking about traditional OLS display
  - Sum of Squares, test of F-statistics (Prob > F), 95% Confidence Intervals
- Automatic storage and production of regression tables
- Complicated command is no automation at all (easier to do it by hand!!)
- Spreadsheet-style interface dedicated to regression outputs/tables
- Mata-based programming