



TYPES of PROC REG Outputs

This application shows how PROC REG can produce SAS output files. It also presents how to use an alternative SAS method, ODS or Output Delivery System, to create SAS output files with statistics not shown in the ordinary procedures. This note uses a simplified example taken from Bali, Cakici, Yan and Zhang (JF 2005), Table III. It runs a regression that tests whether stock market excess return can be predicted with macroeconomic variables. The regression is the following:

$$\text{RETW} = a_0 + a_1 * \text{TERM} + a_2 * \text{DEF}$$

Where:

RETW = Market Value Weighted Return minus 3-month T-bill Rate

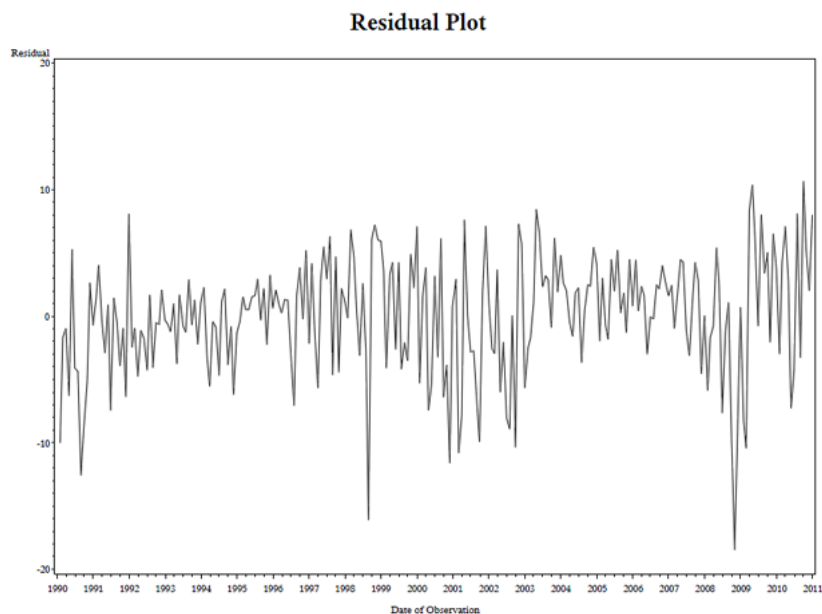
TERM = Lag of Term Premium calculated as the difference between 10-year Treasury bond yield and 3-month T-bill rate.

DEF = Lag of Default Yield calculated as the difference between Moody's Baa yield and 10-year T-bill yield.

Data comes from CRSP (value weighted returns) and Federal Reserve Bank Reports (interest rates). The three standard ways that PROC REG has to generate SAS data sets are: (1) OUTEST option, (2) OUTSSCP option, and (3) OUTPUT statement. These methods and ODS options are shown below :

```
/*Run the Regression by using Bali, Cakici, Yan and Zhang (JF 2005)*/  
proc reg data=x1 outest=outest1 outsscp=sum2;  
  *Use ODS to create output sets;  
  ods output ParameterEstimates= parameterestimates ANOVA = anova ;  
  ods output DWStatistic= dwstatistic FitStatistics= fitstatistics;  
  model retw = term def / dw ;  
  output out=output1 predicted=retw_hat residual=r_residual;  
run;  
ods output close;
```

As specified by parameters predicted=retw_hat residual=r_residual, the output data output1 contains predicted value of dependent variable (retw_hat) and residual (r_residual) for each time period of t. Given such information, one can achieve a simple residual plot by using output1 and SAS procedure, gplot.



ODS output table ParameterEstimates contains statistics for coefficient estimates, such as standard error, t-value, and p-value. ODS output table fitstatistics includes some statistics for the regression model, such as RMSE, R-square, and adjusted R-square. And ODS output table dwstatistic includes Durbin-Watson statistics and etc. With those values provided by different ODS output datasets, we can generate a standardized regression report (the table scheme is imposed by Microsoft Words).

| Variable | coef |
|-------------|----------------------|
| Intercept | -6.32*** [-7.109] |
| TERM | 0.255 [0.472] |
| DEF | 0.741* [1.895] |
| R-Square | 0.098 |
| Adj R-Sq | 0.091 |
| DW | 1.622 |
| No. of obs. | 252 |

The complete sample code to implement Bali, Cakici, Yan and Zhang (JF 2005) regression model, and generate residual plot and regression result report table is provided following:

```

/* ***** WRDS RESEARCH APPLICATIONS ***** */
/* ***** Summary : TYPES of PROC REG Outputs ***** */
/* Date : Revision of August 2011 */
/* Author : Luis Palacios and Rui Dai, WRDS */
/* Note : RETW = Market Value Weighted Return minus 3-month T-bill Rate */
/* Note : TERM = Lag of Term Premium calculated as the difference between */
/* 10-year Treasury bond yield and 3-month T-bill rate. */
/* Note : DEF = Lag of Default Yield calculated as the difference between */
/* Moody's Baa yield and 10-year T-bill yield. */
/* ***** */

proc sql;
  create table x1 as
  select distinct a.date, (a.vwretd)*100 - b.tcmnom_m3 as RETW,
    b.tcmnom_y10 - b.tcmnom_m3 as TERM, b.baa_na - b.tcmnom_m3 as DEF
  from crsp.msi as a, frb.rates_monthly as b
  where intck('month',b.date,a.date)= 1 and year(a.date)>=1990
  order by date;
  *the intck function allows to merge the frb.rates_monthly data as LAG variables;
quit;

proc reg data=x1 outest=outest1 outsscp=sum2;
  *Use ODS to create output sets;
  ods output ParameterEstimates= parameterestimates ANOVA = anova ;
  ods output DWStatistic= dwstatistic FitStatistics= fitstatistics;
  model retw = term def / dw ;
  output out=output1 predicted=retw_hat residual=r_residual;
run;
ods output close;

/* ***** Codes below are used to demonstrate how to use the Proc Reg and ODS output to ***** */
/* generate the customized outputs such as ***** */
/* 1)Residual graph in PDF ***** */
/* 2)Standard regression estimate report in CSV ***** */
/* ***** */

* 1) use ODS to generate a residual plot pdf file;
options orientation=landscape device=pdf nodate; /*PDF file layout configuration*/
ods listing close; /*closes the LISTING destination to conserve resources*/
ods pdf file='plot.pdf'; /*assign a name to plot pdf*/
goptions reset=global cback=white colors=(black); /*Graphic layout configuration 1*/
title "Residual Plot";
symbol1 color=black interpol=join; /*Graphic layout configuration 2*/
proc gplot data=output1; format date year.;
  plot r_residual*date;
run;
ods pdf close; /*closes the PDF destination*/

* 2) dance with Proc Reg output files;
data param1; set ParameterEstimates; /*parameterEstimates contains model estimates*/
if Prob<.01 then str="***"; else if Prob<.05 then str="**";
else if Prob<.1 then str="*";
coef=strip(roundz(Estimate,0.001)||str); idx=_N_;
proc sort; by Variable coef;
run;

proc transpose data=param1 out=param2; /*formatting estimate report 1*/
var Estimate tValue; by Variable coef idx;
run;

data param3; set param2; /*formatting estimate report 2*/
if _NAME_="tValue" then do coef=cats("[" ,roundz(col1,0.001)," ]"); Variable=""; end;
proc sort; by idx _NAME_;
run;

data fit1; set fitstatistics(obs=2 keep=Label2 nvalue2);/*fitstatistics: fit statistics*/
rename Label2=variable; coef=cats(roundz(nValue2, 0.001)," "); drop nValue2;
run;

data dwat1; set dwstatistic(obs=2 keep=Label1 nvalue1);
/*dwstatistic: Durbin-Watson statistic*/
rename Label1=variable; coef=cats(roundz(nValue1, 0.001)," "); drop nValue1;
run;

data param4; set param3(keep=variable coef) fit1 dwat1; run;

ods listing close; /*closes the LISTING destination to conserve resources*/
ods csv file='estimates.csv'; /*assign a name to estimate CSV*/
proc print data=param4; run;
ods csv close; /*closes the CSV destination*/
ods listing;

/* ***** Material Copyright Wharton Research Data Services ***** */
/* ***** All Rights Reserved ***** */

```

APPENDIX:

The following is the data characteristics for some dataset generated by proc reg output and ODS Output:

Proc Reg Output Sets:

Output from OUTEST option

The **OUTEST** option produces a SAS dataset that contains estimated regression coefficients and optional statistics.

```
proc print data=outest1 noobs;  
  Title 'Data from OUTEST option';  
run;
```

Output from OUTEST option (OUTEST_OUT.cfm)

This file (outest1.sas7bdat) has the following variables:

| option | Description |
|-----------|---|
| _MODEL_ | a character variable containing the label of the corresponding MODEL statement |
| _TYPE_ | a character variable with the value 'PARAMS' for every observation |
| _DEPVAR_ | the name of the dependent variable |
| _RMSE_ | the root mean squared error or the estimate of the standard deviation of the error term |
| Intercept | the estimated intercept, unless the NOINT option is specified |

The dependent variable in each modelit is given a value of -1

If you specify the TABLEOUT option, the following statistics listed by _TYPE_ are added after the estimates:

| option | Description |
|--------|---|
| STDERR | the standard error of the estimate |
| T | the t statistic for testing if the estimate is zero |
| PVALUE | the associated p-value |
| LnB | the lower confidence for the estimate |
| UnB | the upper confidence for the estimate |

Specifying the option ADJRSQ, AIC, BIC, CP, EDF, GMSEP, JP, MSE, PC, RSQUARE, SBC, SP, or SSE in the PROC REG or MODEL statement automatically outputs these statistics and the model R2 for each model selected, regardless of the model selection method. Additional variables, in order of occurrence, are as follows.

| option | Description |
|----------|---|
| _IN_ | the number of regressors in the model not including the intercept |
| _P_ | the number of parameters in the model including the intercept, if any |
| _EDF_ | the error degrees of freedom |
| _SSE_ | the error sum of squares, if the SSE option is specified |
| _MSE_ | the mean squared error, if the MSE option is specified |
| _RSQ_ | the R2 statistic |
| _ADJRSQ_ | the adjusted R2, if the ADJRSQ option is specified |
| _CP_ | the Cp statistic, if the CP option is specified |
| _SP_ | the Sp statistic, if the SP option is specified |
| _JP_ | the Jp statistic, if the JP option is specified |
| _PC_ | the PC statistic, if the PC option is specified |
| _GMSEP_ | the GMSEP statistic, if the GMSEP option is specified |
| _AIC_ | the AIC statistic, if the AIC option is specified |
| _BIC_ | the BIC statistic, if the BIC option is specified |
| _SBC_ | the SBC statistic, if the SBC option is specified |

Output from OUTPUT statement

The **OUTPUT** statement creates a new SAS data set with a variety of statistics and diagnostic measures that are calculated for each observation in the data set.

```
proc print data=output1 (obs=10) noobs;  
  Title 'Data from OUTPUT Statement';  
run;
```

| DATE | RETW | TERMDEF | retw_hatr_residual |
|------------------|-------|--------------|--------------------|
| 19900131-14.8915 | -0.04 | 1.94-5.15400 | -9.7375 |
| 19900228-6.4099 | 0.31 | 2.04-5.01697 | -1.3929 |
| 19900330-5.5860 | 0.47 | 2.14-4.90767 | -0.6783 |
| 19900430-10.9986 | 0.42 | 2.04-5.00092 | -5.9977 |
| 19900531 0.8536 | 0.75 | 2.26-4.76366 | 5.6173 |
| 19900629-8.4296 | 0.75 | 2.40-4.64334 | -3.7863 |
| 19900731-8.9305 | 0.49 | 2.23-4.82740 | -4.1031 |
| 19900831-17.0596 | 0.60 | 2.33-4.72540 | -12.3342 |
| 19900928-13.0744 | 1.06 | 2.72-4.32306 | -8.7513 |
| 19901031-8.8503 | 1.29 | 3.04-4.01446 | -4.8359 |

All the variables in the original data set are included in the new data set, along with variables created in the OUTPUT statement. The OUTPUT statement has the option to include other variables as Residuals or Predicted Values, for example. To do that, a keyword has to specifies for each desired statistic (see the following list of keywords) followed by an equal sign, and the variable or variables to contain the statistic (In the example above, two keywords were used :PREDICTED and RESIDUAL). A complete list of keywords is listed below.

List of Keywords:

| option | Description |
|-----------|---|
| COOKD | Cook's D influence statistic |
| COVRATIO | standard influence of observation on covariance of betas |
| DFFITS | standard influence of observation on predicted value |
| H | leverage, $xi(X'X)^{-1}xi'$ |
| LCL | lower bound of a % confidence interval for an individual prediction |
| LCLM | lower bound of a % confidence interval for the expected value(mean) of the dependent variable |
| PREDICTED | predicted values |
| PRESS | ith residual divided by $(1-h)$, where h is the leverage, and where the model has been refit without the ith observation |
| RESIDUAL | residuals, calculated as ACTUAL minus PREDICTED |
| RSTUDENT | a studentized residual with the current observation deleted |
| STDI | standard error of the individual predicted value |
| STDP | standard error of the mean predicted value |
| STDR | standard error of the residual |
| STUDENT | studentized residuals, which are the residuals divided by their standard errors |
| UCL | upper bound of a % confidence interval for an individual prediction |
| UCLM | upper bound of a % confidence interval for the expected value (mean) of the dependent variable |

W

Output from OUTSSCP option

The **OUTSSCP** option produces a output SAS data set containing Sums of Squares and Crossproducts. Observations are identified by the character variable `_NAME_`. The data set contains all variables used in MODEL statements.

```
proc print data=sum2 noobs;
  title 'Data from OUTSSCP option';
run;
```

| _TYPE_ | _NAME_ | Intercept | TERM | DEF | RETW |
|--------|-----------|-----------|---------|----------|----------|
| SSCP | Intercept | 192.000 | 338.18 | 738.53 | -624.48 |
| SSCP | TERM | 338.180 | 851.30 | 1578.60 | -823.86 |
| SSCP | DEF | 738.530 | 1578.60 | 3194.87 | -2057.16 |
| SSCP | RETW | -624.477 | -823.86 | -2057.16 | 6063.37 |
| N | | 192.000 | 192.00 | 192.00 | 192.00 |

ODS Output:

Output from ODS DWStatistic

The ODS output **DWStatistic** produces a SAS dataset that contains Durbin-Watson statistic and p-value to test whether or not the errors have first-order autocorrelation (however, this is not relevant for our example since it has lagged variables). In order to get the DW statistics in a SAS data set, the option DW should be included in the MODEL statement.

```
proc print data= dwat1 noobs;
  title 'DWStatistic';
run;
```

DWStatistic

| Model | Dependent | Label1 | cValue1 | nValue1 |
|--------|-----------|---------------------------|---------|------------|
| MODEL1 | RETW | Durbin-Watson D | 1.787 | 1.786793 |
| MODEL1 | RETW | Number of Observations | 192 | 192.000000 |
| MODEL1 | RETW | 1st Order Autocorrelation | 0.094 | 0.093631 |

Output from ODS ParameterEstimates

The ODS output **ParameterEstimates** produces a SAS dataset that contains per each estimated parameter their standard error, T-statistics and p-value.

```
proc print data= param1 noobs;
  title 'ParameterEstimates';
run;
```

ParameterEstimates

| Model | Dependent Variable | DF | Estimate | StdErr | tValue | Prob | |
|--------|--------------------|-----------|----------|----------|---------|-------|--------|
| MODEL1 | RETW | Intercept | 1 | -6.81555 | 1.31775 | -5.17 | <.0001 |
| MODEL1 | RETW | TERM | 1 | 0.14596 | 0.71988 | 0.20 | 0.8395 |
| MODEL1 | RETW | DEF | 1 | 0.85948 | 0.61166 | 1.41 | 0.1616 |

Output from ODS ANOVA

The ODS output **ANOVA** produces a SAS dataset that contains the variables from the ANOVA table including the F-statistics.

```
proc print data= anova1 noobs;
  title 'ANOVA';
run;
```

ANOVA

| Model | Dependent | Source | DF | SS | MS | FValue | ProbF |
|-------|-----------|--------|----|----|----|--------|-------|
|-------|-----------|--------|----|----|----|--------|-------|

| | | | | | | | |
|--------|------|-----------------|---------------|-----------|-----------|------|--------|
| MODEL1 | RETW | Model | 2 | 336.72066 | 168.36033 | 8.61 | 0.0003 |
| MODEL1 | RETW | Error | 1893695.54627 | 19.55315 | — | — | — |
| MODEL1 | RETW | Corrected Total | 1914032.26693 | — | — | — | — |

More options for ODS output files are available in for PROC REG. ODS table names of the SAS reg procedure (also available at support.sas.com (<http://support.sas.com/>)).

ODS Table Names

Table 76.10 ODS Tables Produced by PROC REG

V

| ODS Table Name | Description | Statement | Option |
|---------------------|---|-----------|---|
| ACovEst | Consistent covariance of estimates matrix | MODEL | ALL, ACOV |
| ACovTestANOVA | Test ANOVA using ACOV estimates | TEST | ACOV (MODEL statement) |
| ANOVA | Model ANOVA table | MODEL | Default |
| CanCorr | Canonical correlations for hypothesis combinations | MTEST | CANPRINT |
| CollinDiag | Collinearity Diagnostics table | MODEL | COLLIN |
| CollinDiagNoInt | Collinearity Diagnostics for no intercept model | MODEL | COLLINOINT |
| ConditionBounds | Bounds on condition number | MODEL | (SELECTION=BACKWARD FORWARD STEPWISE MAXR MINR) and DETAILS |
| Corr | Correlation matrix for analysis variables | PROC | ALL, CORR |
| CorrB | Correlation of estimates | MODEL | CORRB |
| CovB | Covariance of estimates | MODEL | COVB |
| CrossProducts | Bordered model $\mathbf{X}'\mathbf{X}$ matrix | MODEL | ALL, XPX |
| DWStatistic | Durbin-Watson statistic | MODEL | ALL, DW |
| DependenceEquations | Linear dependence equations | MODEL | Default if needed |
| Eigenvalues | MTest eigenvalues | MTEST | CANPRINT |
| Eigenvectors | MTest eigenvectors | MTEST | CANPRINT |
| EntryStatistics | Entry statistics for selection methods | MODEL | (SELECTION=BACKWARD FORWARD STEPWISE MAXR MINR) and DETAILS |
| ErrorPlusHypothesis | MTest error plus hypothesis matrix $\mathbf{H}+\mathbf{E}$ | MTEST | PRINT |
| ErrorSSCP | MTest error matrix \mathbf{E} | MTEST | PRINT |
| FitStatistics | Model fit statistics | MODEL | Default |
| HypothesisSSCP | MTest hypothesis matrix | MTEST | PRINT |
| InvMTestCov | $\text{Inv}(\mathbf{L} \text{ Ginv}(\mathbf{X}'\mathbf{X}) \mathbf{L}')$ and $\text{Inv}(\mathbf{Lb}-\mathbf{c})$ | MTEST | DETAILS |
| InvTestCov | $\text{Inv}(\mathbf{L} \text{ Ginv}(\mathbf{X}'\mathbf{X}) \mathbf{L}')$ and $\text{Inv}(\mathbf{Lb}-\mathbf{c})$ | TEST | PRINT |
| InvXPX | Bordered $\mathbf{X}'\mathbf{X}$ inverse matrix | MODEL | I |
| MTestCov | $\mathbf{L} \text{ Ginv}(\mathbf{X}'\mathbf{X}) \mathbf{L}'$ and $\mathbf{Lb}-\mathbf{c}$ | MTEST | DETAILS |
| MTransform | MTest matrix \mathbf{M} , across dependents | MTEST | DETAILS |
| MultStat | Multivariate test statistics | MTEST | Default |
| NObs | Number of observations | | Default |
| OutputStatistics | Output statistics table | MODEL | ALL, CLI, CLM, INFLUENCE, P, R |
| PartialData | Partial regression leverage data | MODEL | PARTIALDATA |

| | | | | |
|--------------------|---|-------|---|---|
| ParameterEstimates | Model parameter estimates | MODEL | Default if SELECTION= is not specified | |
| RemovalStatistics | Removal statistics for selection methods | MODEL | (SELECTION=BACKWARD STEPWISE MAXR MINR) and DETAILS | |
| ResidualStatistics | Residual statistics and PRESS statistic | MODEL | ALL, CLI, CLM, INFLUENCE, P, R | |
| SelParmEst | Parameter estimates for selection methods | MODEL | SELECTION=BACKWARD FORWARD STEPWISE MAXR MINR | |
| SelectionSummary | Selection summary for FORWARD, BACKWARD, and STEPWISE methods | MODEL | SELECTION=BACKWARD FORWARD STEPWISE | V |
| SeqParmEst | Sequential parameter estimates | MODEL | SEQB | |
| SimpleStatistics | Simple statistics for analysis variables | PROC | ALL, SIMPLE | |
| SpecTest | White's heteroscedasticity test | MODEL | ALL, SPEC | |
| SubsetSelSummary | Selection summary for R-square, Adj-RSq, and Cp methods | MODEL | SELECTION=RSQUARE ADJRSQ CP | |
| TestANOVA | Test ANOVA table | TEST | Default | |
| TestCov | $L \text{ Ginv}(\mathbf{X'X}) L'$ and $Lb-c$ | TEST | PRINT | |
| USSCP | Uncorrected SSCP matrix for analysis variables | PROC | ALL, USSCP | |



(<http://www.wharton.upenn.edu>)

About WRDS (<https://wrds-www.wharton.upenn.edu/pages/about>)

WRDS FAQs (<https://wrds-www.wharton.upenn.edu/pages/wrds-faqs>)

WRDS News (<https://wrds-web.wharton.upenn.edu/wrds/news/index.cfm>)

3 Ways to use WRDS (<https://wrds-www.wharton.upenn.edu/pages/3-ways-use-wrds>)

Account Types on WRDS (<https://wrds-www.wharton.upenn.edu/pages/wrds-account-types>)

Terms of Use (<https://wrds-web.wharton.upenn.edu/wrds/about/terms.cfm>)

Account Preferences (<https://wrds-web.wharton.upenn.edu/wrds/mywrds/preferences.cfm>)

Info / Support Request (https://wrds-web.wharton.upenn.edu/wrds/about/external_support_request.cfm)

Privacy Policy (<https://wrds-www.wharton.upenn.edu/pages/wrds-privacy-policy>)

WRDS Demo (<https://wrds-www.wharton.upenn.edu/demo/>)

Conference Calendar (<http://www.whartonwrds.com/about/conferences/>)

Best Paper Awards (<http://www.whartonwrds.com/best-paper-award-winners/>)

Wharton Research Data Services

Unless otherwise noted, all material is © 1993 - 2017, The Wharton School, University of Pennsylvania. All rights reserved.

