Shiny Interface to the RobStatTM Package

Gregory Brownson

March 2, 2020

1 Introduction

This document describes a user interface (Shiny UI) to the **RobStatTM** R package companion to the *Robust Statistics: Theory and Methods* book (Maronna et al. 2019 second edition). The Shiny UI supports use of functionality in **RobStatTM** without needing to set many R function arguments or understanding the complexities of the underlying functions. The simplicity of the Shiny UI will allow students and practitioners who are new to robust statistical methods to easily use them and compare robust results with classical methods. With the Shiny UI, you will be able to load a data set either from an existing R package or a local .csv file and then run the data through different methods of analysis. Currently, the Shiny UI provides an interface to robust location and scale, robust linear regression, robust covariance estimation, and robust principal component analysis.

Package Dependencies

The Shiny UI depends on the following packages:

- DT
- fit.models
- ggplot2
- grid
- gridExtra
- gtools
- shiny
- shinyjs
- xts

With the above packages all installed, start the Shiny UI interface with the single line of code:

ShinyUI()

This opens the Shiny user interface, enabling you to import data, and use functions in RobStatTM to fit models and analyze results.

Figure 1 displays the initial form of the Shiny user interface window. The dialog in this window is where you would select or upload a data set, and then navigate to one of the model tabs for further analysis. The subsequent sections contain detailed instructions for the functionality of each component of the Shiny UI.

RobStatTM	Data	Location-Scale	Models 🗸	About	Help
Data Source					
R Package	e				
 Upload 					
Library Name	e				
RobStatTM		-			
Select Datas	et				
algae		-			
Load Data					

Figure 1: Initial screen after starting the Shiny UI

2 Loading Data

You may load the data into the Shiny UI from an existing R package, or upload a csv file, by selecting from an "R Package", or "Upload" as shown in Figure 2. Once you select a package from the list under **Library Name**, the **Select Dataset** input will be populated by the names of existing of datasets in that package. After selecting a dataset, you should click on the **Load Data** button to load the data into the application for further analysis, and the data will be displayed in a table on the right. Select the data set **mineral** in the dropdown **Select Dataset** and click on **Load Data**. Then you will get the data display result shown in Figure 2.

Data Source	Show 10 - entries	Search:										
 R Package Upload 		copper 🔶	zinc 🔶									
Library Name	1	102	4									
RobStatTM -	2	96	56									
Select Dataset	3	265	2									
mineral 🗸	4	185	8									
Load Data	5	229	26									
()	6	20	1									
	7	49	9									
	8	28	9									
	9	128	28									
	10	83	16									
	Showing 1 to 10 of 53 entries	Previous 1 2 3 4 5	6 Next									

Figure 2: Interface to load a data set from an existing R package

Figure 3 shows the options for uploading a csv file for analysis with RobStatTM. Since some csv files do not have the

default format (comma-delimited values, double quote escape character), advanced options are available to specify the separator or escape character. If the data is a time series, then you must check the corresponding box as shown in the example. The first column of data must contain the date or time indices to be read as a time series. For this tutorial, we use the data set **hfunds.ts.csv** which contains returns for five hedge funds, and upload it as a time series (Instructions to find this dataset are provided below). The window in the right-hand panel displays the data in a table like format, where you may view the returns for each date. Each column corresponds to one of the five different types of hedge funds: emerging markets (EM), Private Equity (PE), U.S. High Yield (USHY), Alternative Investments (AI), and Bond (BND). To find the **hfunds** dataset, first you need to type find.package('RobStatTM') in the command line to get the path to the **RobStatTM** package. From that location, the file is located in *etc* subdirectory. For data sets with a large number of variables, you can view the data in full screen mode.

Source	Show 10 • ent	ries		Search:	ı:				
t Package Ipload		EM 🔶	PE	USHY 🔶	AI ≑	в			
oose CSV File	1989-11-30	0.026	0.013	-0.003	-0.019	-0.			
rowse hfunds4.ts.csv	1989-12-31	0.058	0.008	-0.014	-0.045	0			
Upload complete	1990-01-31	-0.022	-0.079	-0.034	-0.005	-0.			
leader	1990-02-28	-0.026	0.012	-0.021	0.02	-0.			
	1990-03-31	0.051	0.024	0.017	0.019	0.00			
comma	1990-04-30	-0.023	-0.039	-0.007	-0.003	-0.			
emicolon	1990-05-31	0.028	0.087	0.011	0.008	0.			
te	1990-06-30	-0.003	-0.008	0.022	-0.004	0.			
O None	1990-07-31	1990-07-31 -0.008 -0.017		0.017	-0.023	0.			
puble Quote ngle Quote	1990-08-31	-0.144	-0.102	-0.054	-0.005	-0.			
Time Series? Dad Data	Showing 1 to 10 of	Previous	1 2	3 4 5	15	N			

Figure 3: Interface to upload a csv file to the application

3 Robust Location and Scale Estimators

Click on the **Location-Scale** tab to open the dialog which allows you to compute classical estimates for mean and standard deviation and/or robust estimates for location and scale for a single variable, and compare results when you choose to compute both classical and robust estimates. The Shiny UI interfaces with the locScaleM() function in **RobStatTM** to compute robust estimates of location and scale.

Computing Location and Scale

To use the Location-Scale dialog, you first select a **Variable** for whatever dataset you loaded, and then choose a **Method** from the options of *Classical*, *Robust*, or *Both*. The classical and robust methods are both computed and then compared when *Both* is chosen. Furthermore, if either *Robust* or *Both* are selected, then drop-down lists for ρ and ψ functions (rho and psi), and asymptotic efficiency choices appear, with defaults "mopt" (modified optimal) and 0.95, respectively. To compute the estimates and display the summary, you will click on the **Results** button. Figure 4 shows

an example where you first load the data set **flour** using the method explained in Section 2, select the variable V1, and click on Results, thereby using the default values of rho and psi, and asymptotic efficiency.

Location-Scale

Calculate the robust location and scale for a single variable in a data set.

Dataset flour Choose Data	Comparison of Location (SE): Classical 4.28 (1.08) Robust 3.12 (0.113) Comparison of Scale: Classical 5.3 Robust 0.695
Variable	
V1 •	
Method	
 Both 	
 Classical 	
⊖ Robust	
Туре	
 Both 	
⊖ Scale	
Rho and Psi Functions	
mopt 💌	
Asymptotic Efficiency	
0.85	
0.9	
0.95	
○ 0.99	
Results	

Figure 4: Comparison of Classical and robust estimates for location and scale for the flour dataset

Note that the results are displayed in the main panel on the right. This is the layout for each model summary in the Shiny UI. For location and scale, the summary displays estimates for location, the standard error of the location estimate, and the estimate for scale.

4 Robust Linear Regression

Click on the **Robust Linear Regression** tab, located under the **Models** tab in the menu as shown in Figure 5 (all subsequent models are located under the **Models** tab). We now show how to do a computation and comparison of classical and robust linear regression model fits, using the **mineral** data set in the RobStatTM package that we showed you how to load in Section 2.

RobStatTM Data Location-Scale	Models - About Help
Robust Linear Regression Calculate the robust coefficients of several fac	Robust Linear Regression Robust Covariance Robust PCA
Model Plotting	

Figure 5: Navigation to Robust Linear Regression menu

Fitting Linear Models

There are four linear model fitting methods to choose from in the UI: least-squares (LS), M, MM, and distance constrained maximum-likelihood (DCML). The Shiny UI provides an interface to the implementation of these methods through the lm() function in the **stats** package and the lmrobM(), lmrobdetMM(), and lmrobdetDCML() functions in **RobStatTM**. You have the option to compare any two models by checking the **Add Second Method** box and selecting inputs for a second model. In Figure 6, the classical least-squares (LS) and robust MM regression methods are being compared. You must select dependent and independent variables for the model, and then the regression formula will automatically populate the text box. When comparing regressions, the formulas do not need to be identical. You fit the models and view the summary by pressing the **Results** button at the bottom of the model panel, and then the results will be displayed in the panel that appears at the right.

Robust Linear Regression

alculate li obust regr	inear regres ression.	sion coefficients using	a robust regression, and compare them to the coefficients of a least squares regression, or another
Model	Plots		
Dataset			Residual Statistics:
mineral			Method Min 1Q Median 3Q Max
0	Dete		MM -16.54 -5.421 1.049 7.919 116.8
Choose	Data		Coefficients:
🗹 Add	Second Met	thod	Method Estimate Std. Error t value Pr(>ltl)
Method	d	Method	LS 7.961 2.702 2.946 0.004838 **
LS	-	MM -	(Intercept) MM 15.19 2.258 6.728 1.464e-08 ***
			LS 0.1346 0.01983 6.787 1.181e-08 ***
Depend	dent	Dependent	MM 0.01261 0.02232 0.5648 0.5747
zinc	-	zinc 👻	Signif. codes: 0 **** 1e-03 *** 0.01 ** 0.05 ** 0.1 ** 1
Indepe	ndent	Independent	Residual Scale Estimates:
coppe	er	copper	MM: 9.994 on 51 degrees of freedom
Formul	a	Formula	Multiple R-squared:
zinc ~	coppei	zinc ~ copp	LS: 0.4746 MM: 0.008241
		Robust Regression Choices Family mopt	
		Efficiency	
		0.95	
Resul	Its		

Figure 6: Least-squares and robust MM regression comparison using mineral data from RobStatTM

The standard errors, t-statistics, and p-values for the robust coefficients for robust fits are computed using a robust covariance matrix for the independent variables, as an important step to ensure that those quantities are robust themselves (see the RobStatTM book Section 5.6 for details). Furthermore, the *proportion of variance explained by the model*, or R^2 , for robust fits is a robust version of classical least-squares R^2 . To run this example, you have to make zinc is the dependent variable and copper is the independent variable for the LS and MM estimators. The default "mopt" choice for rho and psi, with the default asymptotic efficiency of 95% is used for the MM estimate. The equivalent command-line code for obtaining the results in the right hand side of Figure 6 is provided below.

summary(fm)

Plots

Model	Plots							
Plots	luals v Fi	t						
Response v. Fit								
Resid	luals Norn	nal QQ Plot						
Std. I	Residuals	v. Robust Distances						
Estim	ated Resi	dual Density						
Std. I	Residuals	v. Index (Time)						
🗹 Scatt	er with Ov	erlaid Fits						
Options								
🗹 QQ F	lot Envelo	ope						
View F	Plots							

Figure 7: Linear Regression plotting options

After fitting a linear regression model, or a pair of such models, you may graphically analyze the results using the **Plots** tab. In the dialog above, you may choose from several different plots. Note that the *Scatter with Overlaid Fits*

plot is only available in the case of two univariate regressions. The following figures show each of the plots selected in the dialog of Figure 7.



Residuals versus Fitted Values

Figure 8: Scatterplot of LS and robust MM residuals versus fitted values using mineral data from RobStatTM



Response versus Fitted Values

Figure 9: Scatterplot of response values versus LS and robust MM fitted values using mineral data from RobStatTM

Normal QQ Plots of Residuals



Figure 10: LS and robust MM normal QQ plots of residuals using mineral data from RobStatTM

Standardized Residuals versus Robust Distances



Figure 11: LS and robust MM Plots of standardized residuals versus square-root Mahalanobis distances using **mineral** data from **RobStatTM**

Probability Density Estimates of Residuals



Figure 12: Probability density estimates of LS and robust MM residuals using mineral data from RobStatTM



Standardized Residuals versus Index (Time)

Figure 13: LS and robust MM standardized residuals using mineral data from RobStatTM

Scatterplots with Overlaid Fits



Figure 14: Scatterplot of data with overlaid LS and robust MM fits using mineral data from RobStatTM

5 Robust Covariance

The Shiny UI allows you to use the **RobStatTM** functions **covClassic()** and **covRob()**. Since these two methods only work numeric data, the options for **Variables** are limited to those with numeric values only. Note that the options in the **Method** group allows you to choose either the classical method alone, the robust method alone, or use the default of both methods, which is highly recommended. If *Robust*, or *Both* is chosen, then you use the Robust Covariance Estimator drop-down list to make an estimator choice. You may select *MM*, *Rocke*, or *Auto* as the estimator. If you use the default choice *Auto*, then a Rocke-estimator is used if the number of variables is greater than or equal to 10, and an MM-estimator is used otherwise. For details concerning these estimators, see Sections 6.4.4 and 6.5, respectively in the book Robust Statistics: Theory and Methods, 2nd editio Figure 15 displays the comparison of classical and robust covariance matrix and location estimates along with classical and robust eigenvalues associated with those covariance matrix estimate for variables in the **wine** dataset.

Dataset Comparison of Covariance/C	Correlation E	Estimates										
wine Method V1 V2	2 V3	V	4	V5	V6	V7	V8	V9	V10	V11	V12	V13
Classic 0.2136 -0	0.01289 -	-0.0156 -	-0.3746	0.7732	0.0659	0.07621	5.069e-04	0.05858	0.2337	0.004307	0.01152	36.92
Choose Data 0.2122 -0.	.008808 -0	0.01748 -	-0.2905	0.8164	0.05805	0.06654	-0.002612	0.04833	0.278	0.001568	0.02014	41.44
Variables V2	0.4741 0.0	004101	0.1053	0.5734	-0.01949	-0.05235	-0.00431	-0.02293	-0.2197	-0.03368	0.04259	-56.84
Robust -0.008808 (0.05706 0	0.02243	0.1564	0.4165	6.752e-04	-0.005884	0.003985	0.01916	-0.05268	-0.001188	0.009691	-3.756
V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 Classic -0.0156 0.	.004101	0.0516	0.3178	0.9124	3.697e-04	-0.006362	0.007414	-0.01362	-0.03495	0.00633	-0.006618	-1.486
Robust -0.01748 (0.02243 0	0.06255	0.434	1.027	0.001667	-0.00805	0.01045	-0.01908	-0.06034	0.006089	-0.009947	-3.877
Both Classic -0.3746 V4	0.1053	0.3178	6.484	6.372	-0.1925	-0.2906	0.05391	-0.1822	-0.6653	0.02758	-0.107	-69.06
Classical Robust -0.2905	0.1564	0.434	7.265	9.769	-0.1577	-0.2365	0.08199	-0.07052	-0.8828	0.05487	-0.1576	-113.7
O Robust V5	0.5734 0	0.9124	6.372	110.2	1.093	0.5147	0.1745	-0.2555	2.401	-0.1362	0.4523	-344
Robust 0.8164	0.4165	1.027	9.769	115.8	1.184	0.5914	0.3407	-0.243	2.312	-0.03503	0.2395	-323.1
Type V6 Detwork 0.0009 -0	0.01949 3.65	978-04 -	-0.1925	1.093	0.1149	0.1063	-4.0348-04	0.05219	0.2729	-0.008857	0.006435	22.10
Covariances Robust 0.05805 0.7	0.05225-04 0.0	001007 -	-0.1577	1.104	0.1311	0.1219	-0.002169	0.05249	0.3259	-0.01396	0.007929	24.42
O Correlations V7 Classic 0.07621 -	0.00200 -0.0	000302 -	0.2900	0.5147	0.1003	0.100	-0.002493	0.00907	0.3031	0.005942	-0.01237	26.4
Robust Covariance Estimator Classic 5.069e.04 -0.	0.003004 -0.	007414 0	0.2303	0.1745	-4.0346-04	-0.002403	0.000132	-0.004172	-0.01323	0.00336	-0.000079	-0 2370
V8 Robust - 0.002612 0	003085 0	01045 0	0.08100	0.3407	-0.002160	-0.006132	0.005423	-0.008835	-0.02210	0.003225	-0.006028	-2 477
Classic 0.05858 -	0.02293 -0	01362	-0 1822	-0 2555	0.05219	0.08987	-0.004172	0 1698	0.2168	0.004989	4 571e-04	12.98
Robust 0.04833 (0.01916 -0.	0.01908 -0	0.07052	-0.243	0.05249	0.09277	-0.008835	0.189	0.2685	-0.001597	0.003595	15.88
Classic 0.2337	-0.2197 -0	.03495 -	-0.6653	2.401	0.2729	0.3651	-0.01323	0.2168	1.534	0.004072	-0.08266	161.5
V10 Robust 0.278 -0	0.05268 -0.	.06034 -	-0.8828	2.312	0.3259	0.4428	-0.02219	0.2685	1.879	-0.007908	-0.09283	179.6
Classic 0.004307 -0	0.03368 0.	0.00633 0	0.02758	-0.1362	-0.008857	3.641e-04	0.00336	0.004989	0.004072	0.01357	-0.01292	9.118
V11 Robust 0.001568 -0.	.001188 0.0	006089 0	0.05487	-0.03503	-0.01398	-0.005842	0.003225	-0.001597	-0.007908	0.01303	-0.01293	5.826
Classic 0.01152 0	0.04259 -0.0	006618	-0.107	0.4523	0.006435	-0.01257	-0.008091	4.571e-04	-0.08266	-0.01292	0.1275	-27.54
V12 Robust 0.02014 0.	.009691 -0.0	009947 -	-0.1576	0.2395	0.007929	-0.008679	-0.006928	0.003595	-0.09283	-0.01293	0.1441	-25.89
Classic 36.92	-56.84	-1.486	-69.06	-344	22.15	33.5	-0.2379	12.98	161.5	9.118	-27.54	49070
Robust 41.44	-3.756	-3.877	-113.7	-323.1	24.42	36.4	-2.477	15.88	179.6	5.826	-25.89	49950
Robust Estimate of Location	1:											
Method V1 V2 V3	V4 V5	V6	٧7 V	va va	9 V10 V	/11 V12	V13					
Classic 13.74 2.011 2.456	6 17.04 106	5.3 2.84	2.982	0.29 1.4	899 5.528	1.062 3.158	3 1116					
Robust 13.73 1.779 2.466	6 17.15 106	6.8 2.854	3.002 0	0.2922 1.9	912 5.648	1.077 3.15	5 1145					
Eigenvalues:												
Method Eval. 1 Eval. 2 Ev	val. 3 Eval.	4 Eval. 5	Eval. 6	Eval. 7	Eval. 8 Ev	al. 9 Eval	. 10 Eval. 1	1 Eval. 12	Eval. 13			
Classic 49070 108.3 6	6.199 0.93	0.412	0.1817	0.1118	0.09973 0.0	7915 0.03	052 0.0203	4 0.006755	0.002164			
Robust 49960 114.6 6	6.424 1.17	6 0.1916	0.134	0.1106	0.08745 0.0	4865 0.02	748 0.0211	9 0.007717	0.002016			

Figure 15: Classical and robust covariance estimates for selected variables in wine data

A command-line equivalent of the Shiny UI interface in Figure 15 is shown below

summary(cov.fm)

Plots

Estimates	Plots					
Plots	lies					
Mahalanobis Distances						
 ✓ Ellipses Matrix ✓ Chi-Square QQ Plot 						
 ☑ Image Display ☑ Distance-Distance Plot 						
View Plots	•					

Figure 16: Covariance plotting options

The Shiny UI provides the plot types displayed in Figure 16 for classical and robust covariance matrix estimates. For the wine data set, the plot types selected in Figure 16 are shown below.



Mahalanobis Distances

Figure 17: Square-root Mahalanobis distances for classical and robust covariance estimates for the wine data.

Distances Chi-Squared QQ Plot



Figure 18: Classical and robust square-root Mahalanobis distances chi-squared QQ plots for the wine data.

Distance-Distance scatterplot



Figure 19: Scatterplot of square-root classical and robust Mahalanobis distances for the wine data.

Figures 15, 17, 18, and 19 above correspond to Example 6.2 and Figure 6.3 in Maronna et al. You may find it of interest run the above example with the choice "Covariances" changed to "Correlations". Figures 20 and 21 below use the variables V2, V4, V8, V11, and V12 in the **wine** dataset, which were chosen due to their low correlation with each other and to reduce the dimensionality of the dataset.

Eigenvalues



Figure 20: Scree plot for classical and robust covariance estimates for the wine data.





6 Robust PCA

This section allows a simple computation and comparison of classical and robust principal component analysis.

Computing Robust Principal Components

The options for computing principal components are similar to the options for computing covariance. Classical principal components are computed using **prcomp()** and robust principal components are computed using the **pcaRobS()**

Ellipses Matrix

function in **RobStatTM**. You must select at least three **Variables** for analysis. If the option "Both" is selected for **Method**, then the classical and robust principal components will be computed and compared. Figure 22 contains an example comparing classical and robust principal components of the **bus** data.

Robust F	Principle Component Analysis																
Principal compo	onent analysis using classical and/or robust methods.																
Estimates	Plots																
Dataset		Importance of Components:															
wine			Method	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	
		Standard deviation	Classic	3.666	2.052	1.711	1.308	1.153	0.9792	0.8656	0.7259	0.6602	0.5778	0.5038	0.4206	0.3277	
Choose Data	ta		Robust	1.972	1.782	2.584	1.68	2.28	1.334	0.8705	0.8293	0.8378	0.7331	0.6051	0.629	0.4069	
Variables		Proportion of Variance	Classic	0.4947	0.1551	0.1078	0.06302	0.04895	0.0353	0.02759	0.0194	0.01605	0.01229	0.00934	0.00651	0.00395	
V1 V2 V	V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13		Classic	0.1432	0.1169	0.2459	0.104	0.1915	0.0655	0.0279	0.02533	0.02585	0.01979	0.01346	0.01457	0.0061	
		Cumulative Proportion	Robust	0.4547	0.2601	0.506	0.0200	0.8015	0.867	0.8949	0.9310	0.9461	0.9658	0.9093	0.990	1	
Method			11000001	0.1402	0.2001	0.000	0.01	0.0010	0.001	0.0010	0.0202	0.0401	0.0000	0.0700	0.0000		
Both																	
O Classical	d.																
 Robust 																	
Standard	dize																
Results																	

Figure 22: Classical and robust estimates for principal components for first ten variables in the bus data set

The command-line equivalent to the example in Figure 22 is:

Plots

Plots for principal component analysis are not available in the current version, and we hope to add them in a future update.

7 What's Next?

Improvements to the User Interface

We hope to evolve the Shiny UI as we receive feedback on the interface and consider new ideas for it.