Mini Tutorial: Properly Perform M-fold CV

Datafile and Saved Model Definition

Using saved model results, this tutorial will illustrate how M-fold CV should *not* be performed.

Download this datafile: LMTr&Val.xls

Using this saved model definition: CVRsqRight.txt

Enabling advanced options

Once XLSTAT-Pro is activated, go to the menu **Options**, and in the tab **Advanced** enable the option named **Show the advanced buttons in the dialog boxes**.

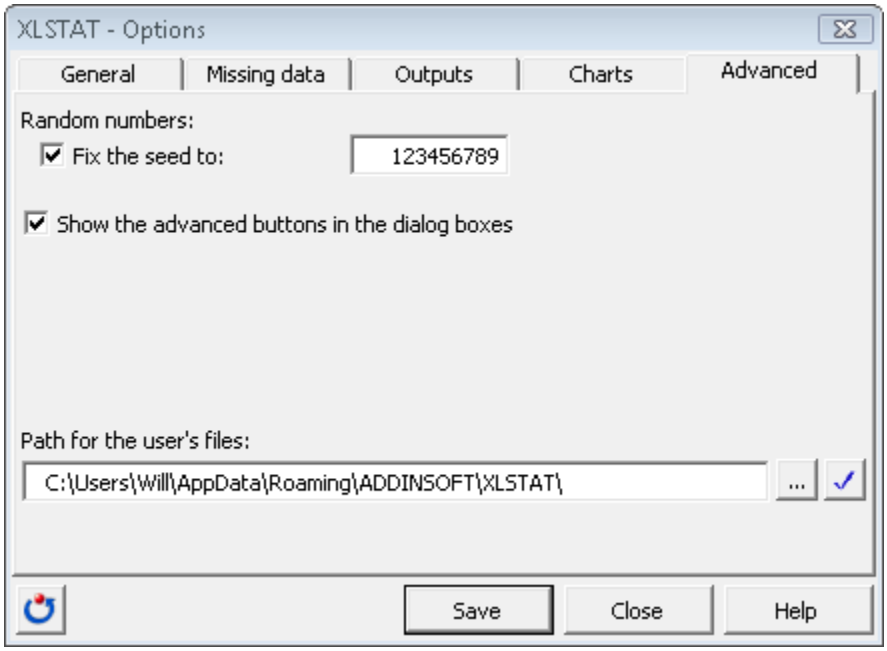
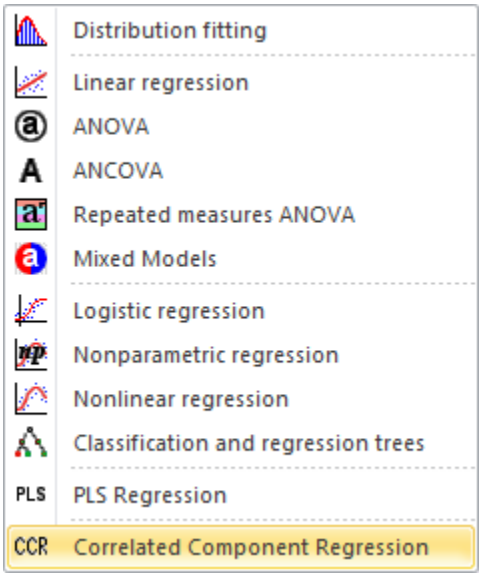


Figure 1. Advanced tab of XLSTAT Options Dialog Box

Opening a Previously Setup Correlated Component Regression

To activate the Correlated Component Regression dialog box, first start XLSTAT by clicking on the button in the Excel toolbar, then select the **XLSTAT / Modeling data / Correlated Component Regression** command in the Excel menu or click the corresponding button on the **Modeling data** toolbar.



Once you have clicked the button, the **Correlated Component Regression** dialog box is displayed with the Method=CCR.LM (linear regression model) selected by default.

When the dialog box is open click on the blue button to load the code.

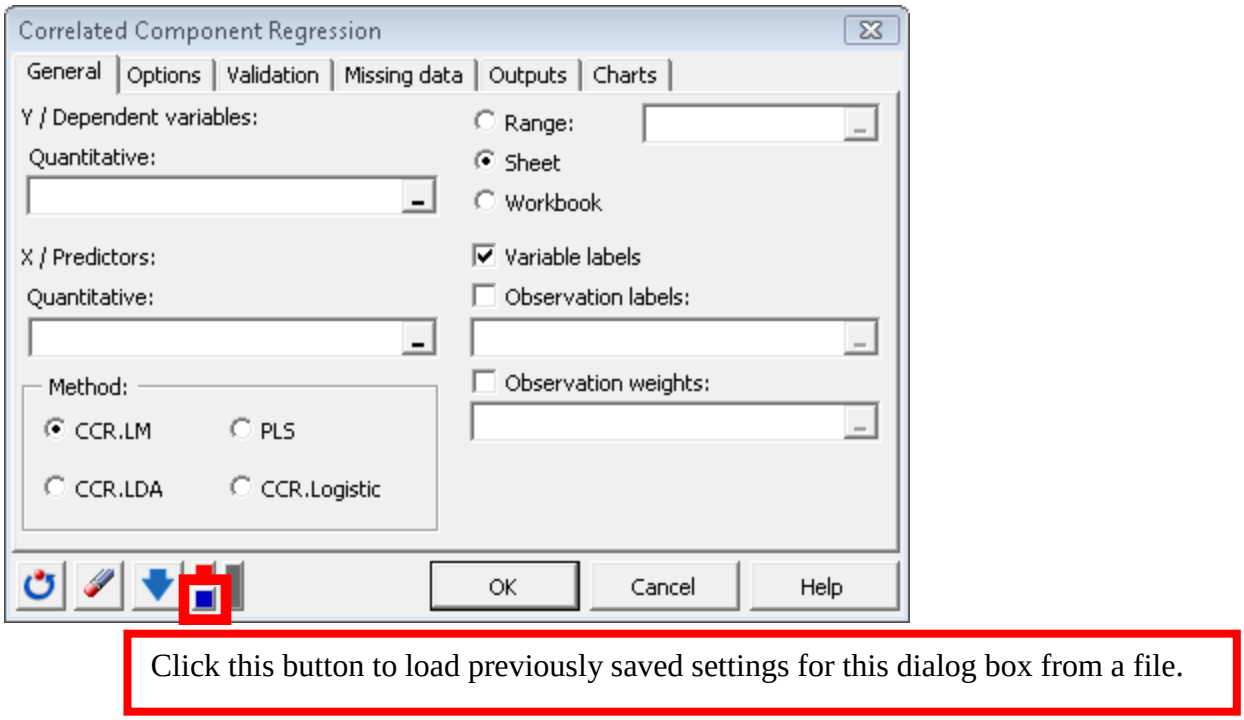


Figure 2. Loading previously saved model settings for CCR dialog box.

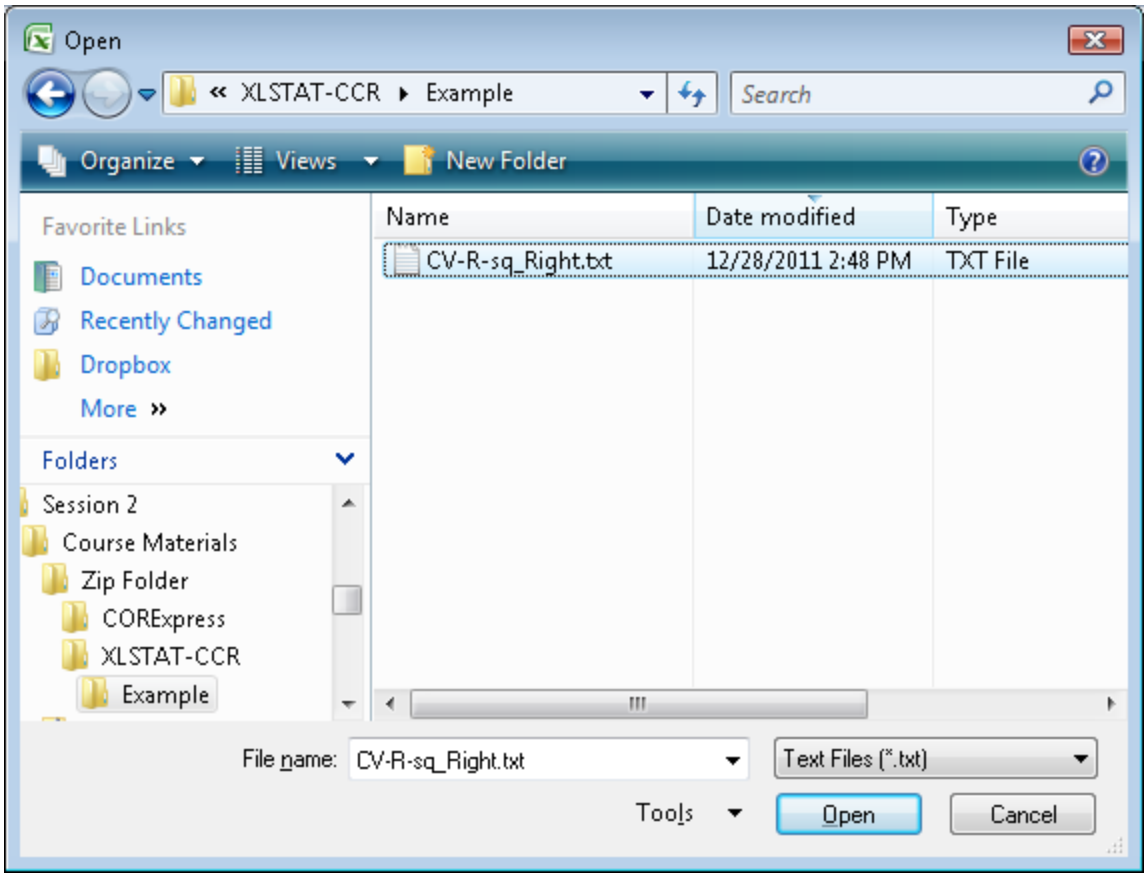


Figure 3. Loading previously saved model settings for CCR dialog box.

Open the text file “CVRsqRight.txt” to load previously saved settings for this dialog box from a file.

The control window will now show the saved model specifications and the corresponding model output, illustrating how to properly perform M-fold CV.

Now simply click on **OK.**

From the drop down menu, select ‘Cross-validation step-down table’. Scroll down to the ‘Cross-validation step-down plot’. As expected, we see higher CV R^2 values as the number of predictors is reduced from 9 to 8 to 7, etc.

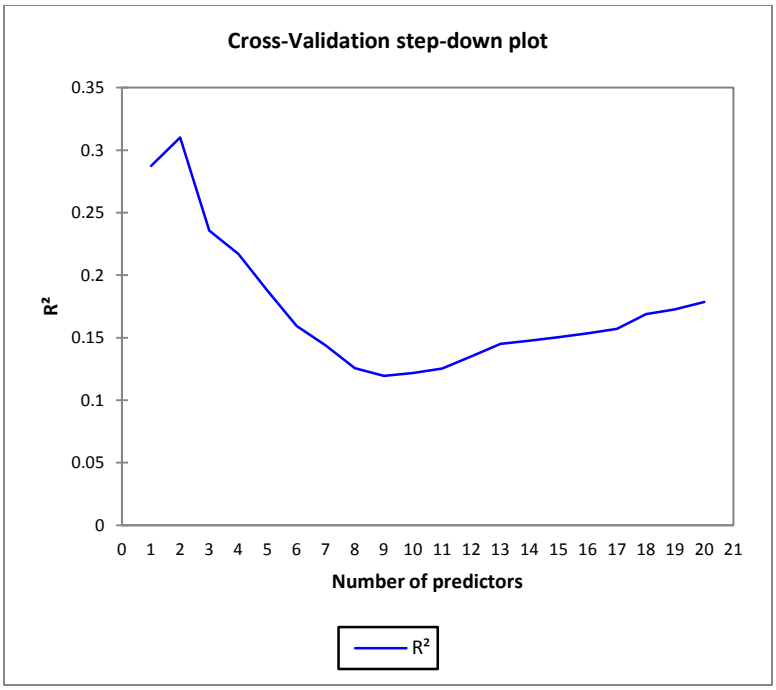


Figure 4. CV-R^2 Results obtained using Cross-validation the wrong way mistakenly suggests that the model with 9 predictors is best. (Results for simulated sample #1 with N=50)