WebSEM: Structural Equation Modeling Online

Zhiyong Zhang and Ke-Hai Yuan

August 27, 2012

118 Haggar Hall, Department of Psychology, University of Notre Dame

Thanks

The development of the path diagram interface was supported by the Center for Creative Computing and the Institute for Scholarship in the Liberal Arts.

DISCLAIMER: WebSEM is provided by psychstat.org "as is" and "with all faults." psychstat.org makes no representations or warranties of any kind concerning the safety, suitability, lack of viruses, inaccuracies, typographical errors, or other harmful components of this WebSEM. There are inherent dangers in the use of any software, and you are solely responsible for determining whether WebSEM is compatible with your equipment and other software installed on your equipment. You are also solely responsible for the protection of your equipment and backup of your data, and THE PROVIDER will not be liable for any damages you may suffer in connection with using, modifying, or distributing WebSEM.

ISBN 978-162314090-8

Contents

Chapter 1

Basics of WebSEM

1.1 Registration

To use WebSEM, you have to register as a user. The benefits to register include:

- Save and upload data on our server
- Protect your data and others' data
- Access your data through WebDav on Windows, Mac, and mobile device
- Use our online SEM analysis
- Use our forum

We approve the registration of each user individually. Therefore, please use your real information including your email and address. If we cannot verify your information, your registration will not be approved.

Your information and data will be transferred safely between your computer and our server as we use SSL to encrypt your data. However, please do not store sensitive information such as SSN, bank account number, etc. on our server. We are not responsible for the lost of such information.

1.1.1 How to register?

To register, go to our website: `https://websem.psychstat.org`. You will be redirected to a page as shown in Figure 1.1.1. Since you do not have an account yet, click on the link `Register`. Then, you will see a registration page as shown in Figure 1.1.2. You need to fill in all the information on this page. Note that the image verification letter will be different on your screen and you have to type the one shown on your screen.

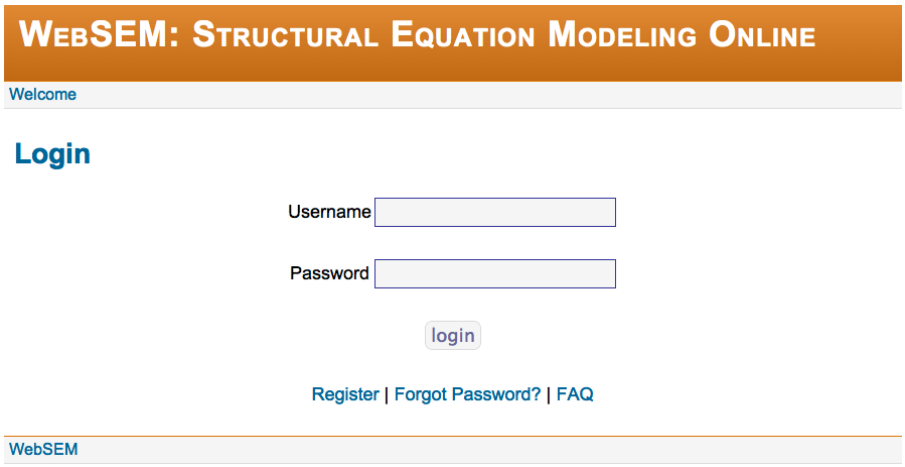


Figure 1.1.1: The login page

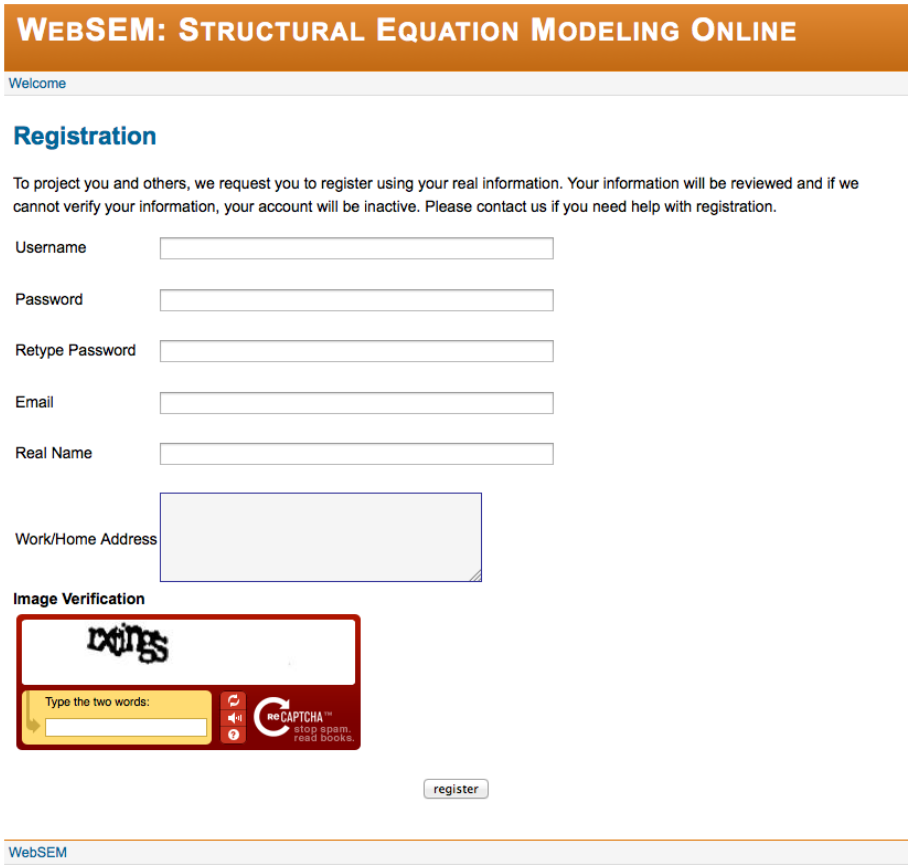


Figure 1.1.2: The registration page

After filling in all the information, click `register` at the bottom. If your registration is successful, you will see information like

`Your registration is completed. You will be notified after your registration is approved. Thanks.`

Shortly, you will receive an email from the address `websem@psychstat.org` with the following information

```
Dear XXX:
Thank you for registering with us. Your user name is
   xxxx.
Your account is *PENDING APPROVAL* and will be
   activated soon.
Thank You!
```

After we verify your information and approve your registration, you will receive the following information:

Your account at WebSEM has been approved. You can login through https://websem.psychstat.org/login.php now.

Now you can start to use WebSEM.

1.1.2 Update your information

At any time, you can update your information after login. Click the Profile link at the bottom, you will see a page like

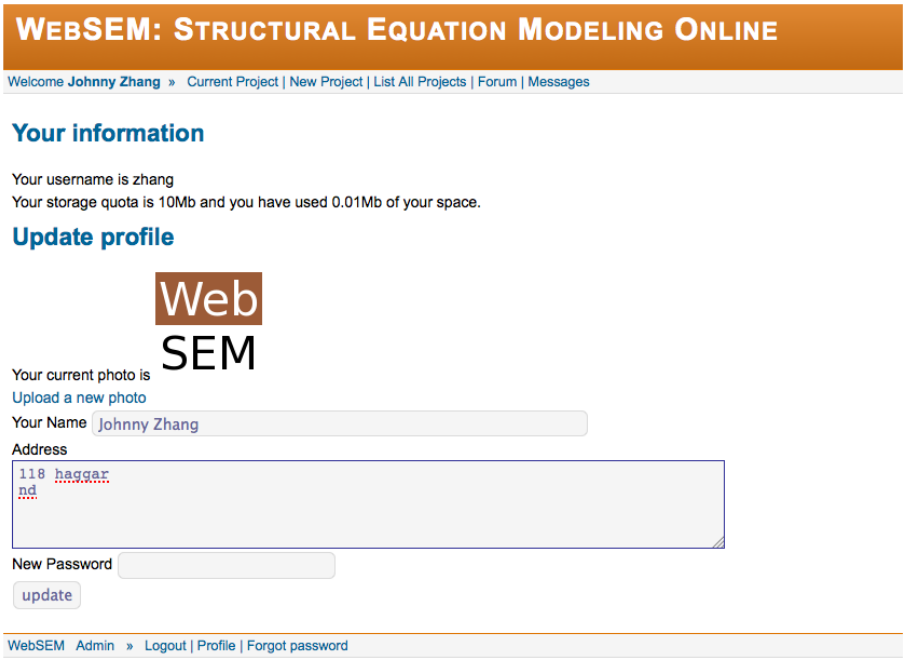


Figure 1.1.3: The project index page

This page shows:

- Your username
- Your storage quota. Initially, you will have 10Mb of space to use. WebSEM is meant to conduct SEM analysis and should not be used as a storage media.
- Your used space. If you use more than 10Mb space, you can still access your analysis but you cannot conduct new analysis. You may free up your space by deleting some old projects or files.

On this page, you can

- Change your photo by uploading a new one.

- Change your name and address.
- Change your password.

1.2 Use WebSEM

After registration and approval, you can login WebSEM on the login page shown in Figure 1.1.1. After login, you will see a page like Figure 1.2.1.

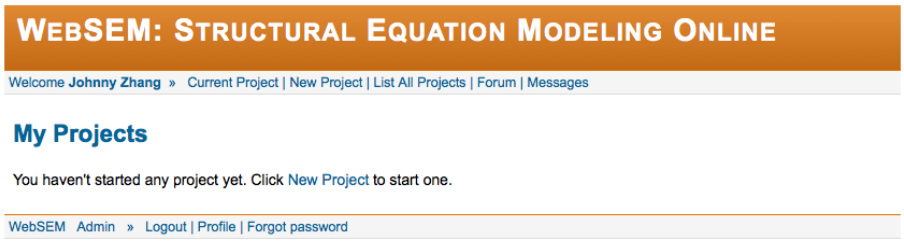


Figure 1.2.1: The project index page

Now you can create a new project by clicking on the link `New Project`. You will be directed to the webpage as shown in Figure 1.2.2. On this page, you can type in the name of a project you want to create. In the example, we used `Mediation Analysis`.

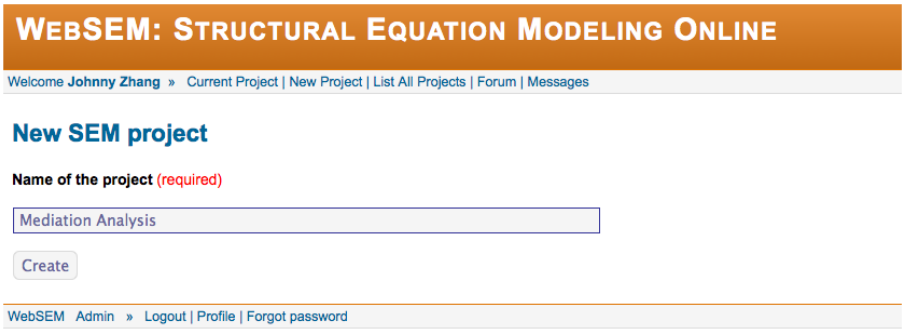


Figure 1.2.2: The project index page

By clicking `Create`, you will see a page as in Figure 1.2.3.

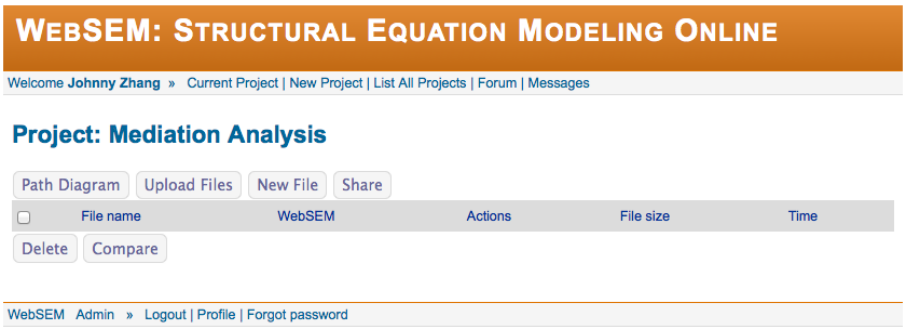


Figure 1.2.3: The project index page

On this page, there are the following buttons:

- `Path Diagram`: A new path diagram can be created.
- `Upload Files`: One can upload one or more files such as data file to be analyzed.
- `New File`: One can create a new file using our online editor.
- `Share`: One can share the project with collaborators.
- `Delete`: One can delete one or more files.
- `Compare`: One can compare the contents of two text files.

WebSEM is an integrated data analysis environment. Suppose we want to create a new file called `simmed.R`. In this file, we will write some R contents to generate data for mediation model.

1.2.1 R editor and run R

WebSEM can be used as an R editor to run R online. For example, to edit `simmed.R`, we do it in a webpage as in Figure 1.2.4.

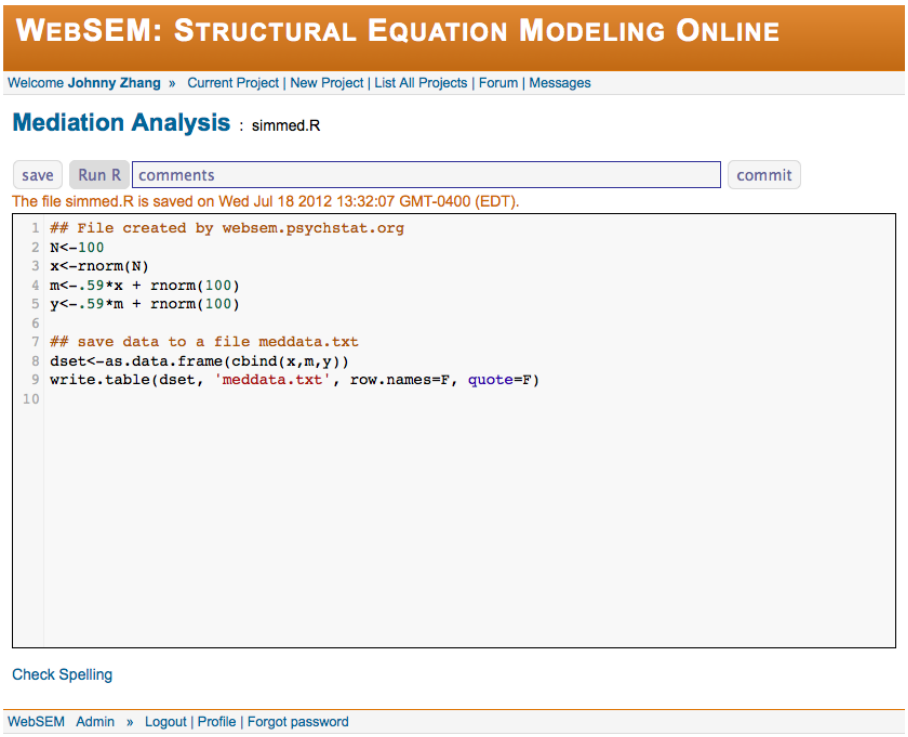


Figure 1.2.4: The R editor

On this page, there are the following buttons:

- `save`: save the current content to a file
- `Run R`: submit the content in the current editor to R to run the analysis
- `commit`: save a copy. One can save the history of editing by using commit. The comments can be changed to indicate edits.

After clicking on the `Run R` button, a link `Click to see the R output` will be shown. One can follow the link to see the output of the R analysis.

We can go back to the project page by clicking on `Current Project`. Now there are three files as shown in Figure 1.2.5. Note that one can `Edit`, `View`, `Delete`, `Download`, and `Rename` each file. If one has saved the history of the files, one can also view its editing history by clicking on `History`. By selecting two files and clicking on `Compare`, one can also compare the contents of the two files.

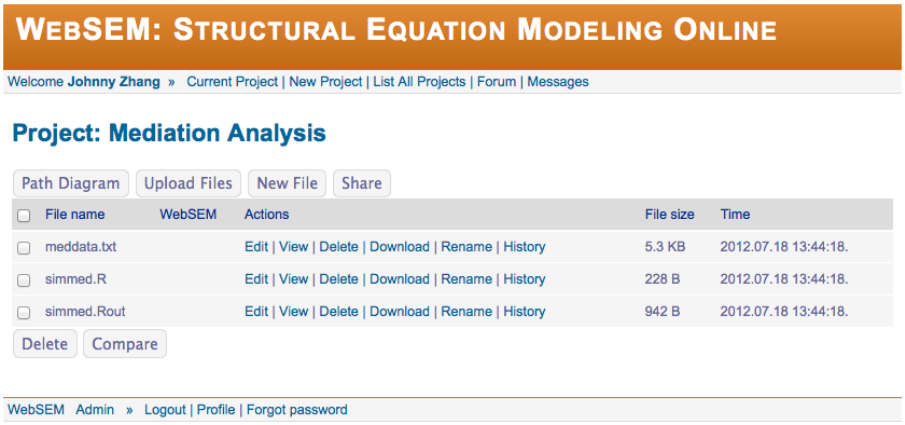


Figure 1.2.5: The project index page

1.2.2 Path diagram

By clicking on `Path Diagram`, a webpage will open for a user to draw path diagram for SEM analysis. A path diagram webpage looks like Figure .

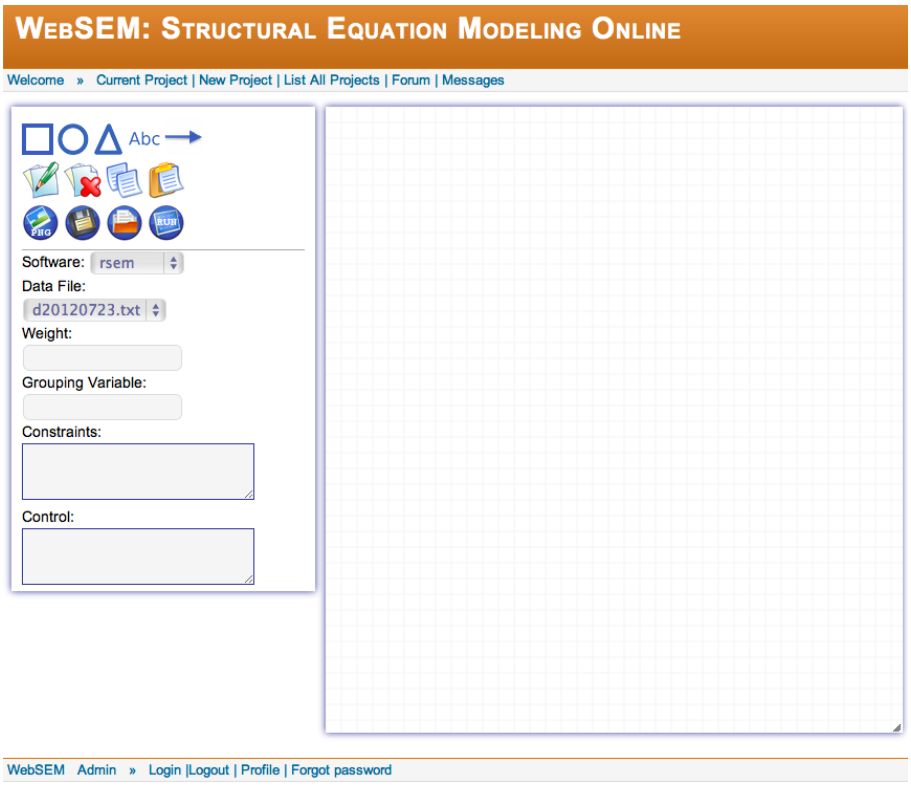


Figure 1.2.6: The path diagram drawing page

On the left panel, there are many different buttons.

- : square for observed variables
- : circle for latent variables
- : for intercepts or means
- Abc: for text
- : for drawing paths
- : for editing a shape or path
- : for deleting a shape or path

-  : copy
-  : paste
-  : generate a png figure for the path diagram
-  : save the path diagram
-  : open/load a path diagram
-  : run an analysis

There are two ways to draw a path diagram. On non-touch screen device, one can draw a path diagram using the context menu. On the drawing canvas, right click the mouse, a menu in Figure 1.2.7 will pop out. Clicking a shape will draw it at the point of the curse.

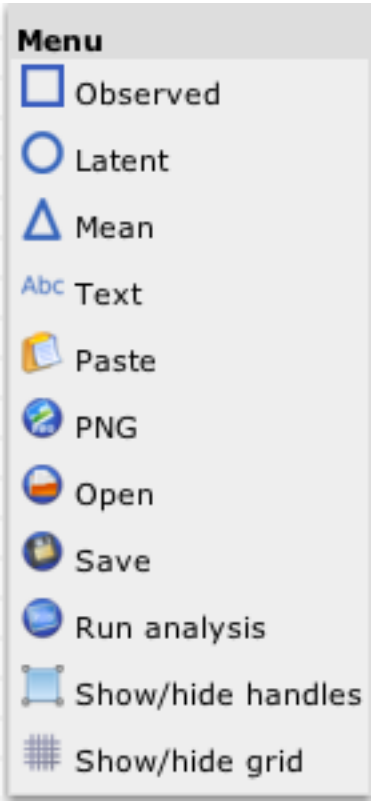


Figure 1.2.7: The drawing menu

On touch-screen device such as iPad, one can click on the button on the left panel directly and a shape will be drawn on the left corner of the canvas.

After drawing a shape, one can conduct the following operations: move, resize, copy, edit (rename). Those buttons can be dragged to the right panel to form path diagram. To drag a shape, click it to select the shape. Then drag, and drop it to the right position. To change the name of a shape, use the right click of a mouse and a menu will pop up as shown in Figure 1.2.8. One can `Edit text` for the name, `Delete` the shape, `Copy` the shape, or `Draw path` starting with this shape.

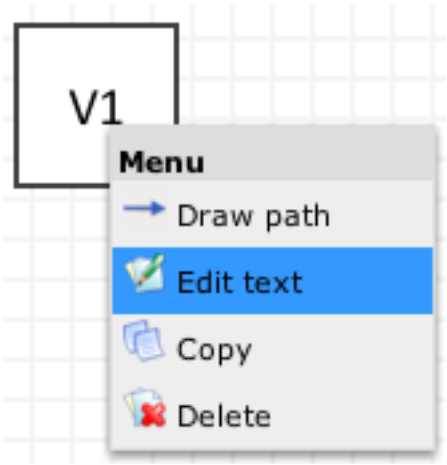


Figure 1.2.8: The menu for a shape

To draw the paths, first click on a square/circle/triangle, then call the menu in Figure 1.2.8 by right click. One then select `Draw path` then a link will show up and drag to another shape to form a path. If the shape goes to is itself, a double-headed arrow will be drawn on the shape to indicate its variance. After drawing, one can change the path by clicking on the control handle (the small square in the middle). For example, one can change the direction of arrow by clicking on it (including left, right, double-headed). One can also use right click to call the menu in Figure 1.2.9 for the path.

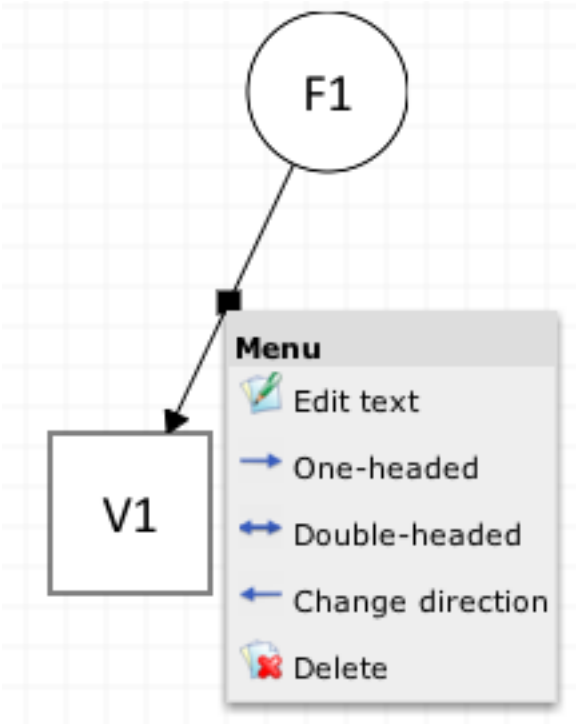


Figure 1.2.9: Menu for a path

Overall, a path diagram for the mediation model looks like Figure 1.2.10.

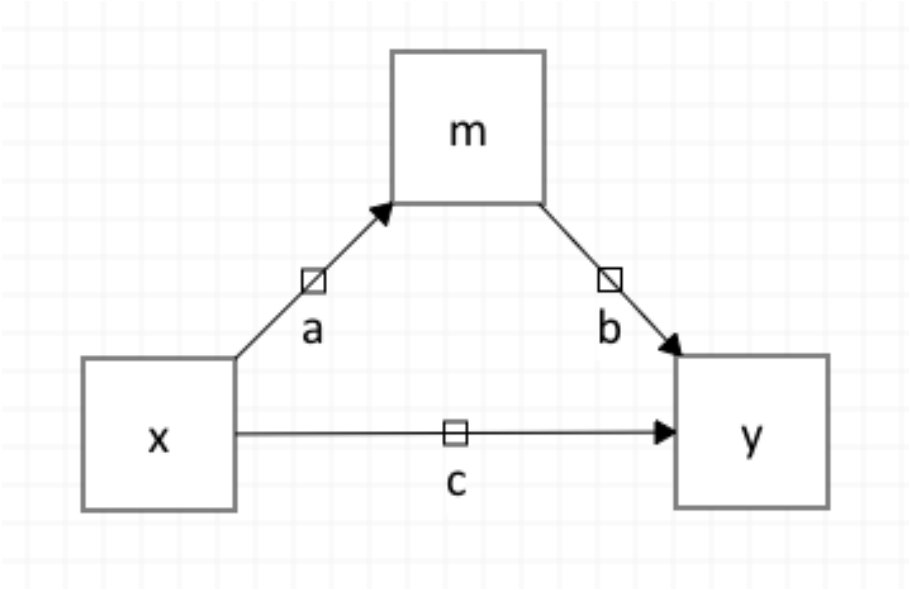


Figure 1.2.10: Path diagram for a mediation model

1.2.3 SEM analysis

With the path diagram, one can run an SEM analysis. WebSEM currently support the R package `rsem` and `lavaan` for SEM analysis. One can select the software to use by clicking on the drop selection box for `Software`. Here, we select `lavaan` for mediation analysis. One also need to specify the dat a file to use. WebSEM automatically list all file with extension name `.txt`. Therefore, your data file has to have an extension name `.txt` and only text file can be used. In this case, the simulated data `meddata.txt` are used. Since this is a mediation analysis, we need to calculate the mediation effect `ab`. Therefore, in the Constraints field, we type `ab==a*b`. Note that `a` and `b` are path names in the path diagram. The complete specification for mediation analysis is shown in Figure 1.2.11.

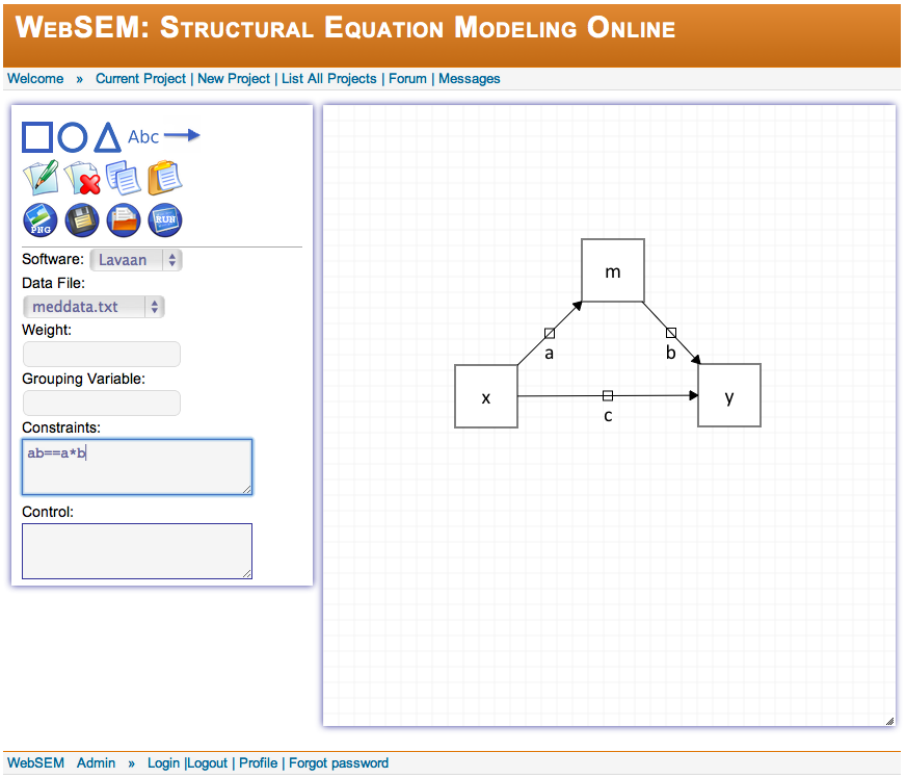


Figure 1.2.11: A complete path diagram ready for analysis

To run the analysis, click the button 

. Then a button  will show up. Click the button will take the user to the output of the analysis. The output looks like that in Figures 1.2.12 and 1.2.13.

WebSEM started at at 14:22:05 on Jul 18, 2012 . You may need to refresh this page for complete output for complex data analysis.

The current analysis was conducted by the WebSEM user johnny. To contact us, make sure to include the ticket no for this analysis **9eb26498863d1fb0f38517b4892c8c0e**

Table 1. Descriptive statistics (N=100)

Variables	Mean	SD	Min	Max	Skewness	Kurtosis
x	0.047789	1.0176	-2.1218	2.8185	0.10087	2.6664
m	0.0059712	1.1621	-2.6216	4.2457	0.39614	3.9173
y	-0.028349	1.1971	-2.2806	2.6547	0.11735	2.3899

Model information
Observed variables: x m y .
The weight is: 0 .
The software to be used is: lavaan .

The following output is from Lavaan.

```
lavaan (0.4-14) converged normally after 13 iterations

  Number of observations                           100

  Estimator                                         ML
  Minimum Function Chi-square                    0.000
  Degrees of freedom                                 0
  P-value                                        0.000

Chi-square test baseline model:

  Minimum Function Chi-square                   58.198
  Degrees of freedom                                 3
  P-value                                        0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    1.000
  Tucker-Lewis Index (TLI)                       1.000

Loglikelihood and Information Criteria:
```

Figure 1.2.12: Mediation analysis output part 1

```
  Loglikelihood user model (H0)              -429.832
  Loglikelihood unrestricted model (H1)      -429.832

  Number of free parameters                         5
  Akaike (AIC)                                869.665
  Bayesian (BIC)                              882.691
  Sample-size adjusted Bayesian (BIC)         866.899

Root Mean Square Error of Approximation:

  RMSEA                                         0.000
  90 Percent Confidence Interval          0.000  0.000
  P-value RMSEA <= 0.05                         1.000

Standardized Root Mean Square Residual:

  SRMR                                          0.000

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  m ~
    x         (a)     0.516    0.102    5.064    0.000
  y ~
    m         (b)     0.492    0.097    5.083    0.000
    x         (c)     0.148    0.110    1.338    0.181

Variances:
    m                 1.064    0.150
    y                 0.996    0.141

Defined parameters:
    ab                0.254    0.071    3.587    0.000
```

Figure 1.2.13: Mediation analysis output part 2

1.2.4 After analysis

Figure 1.2.14 is the project page after running the mediation analysis.

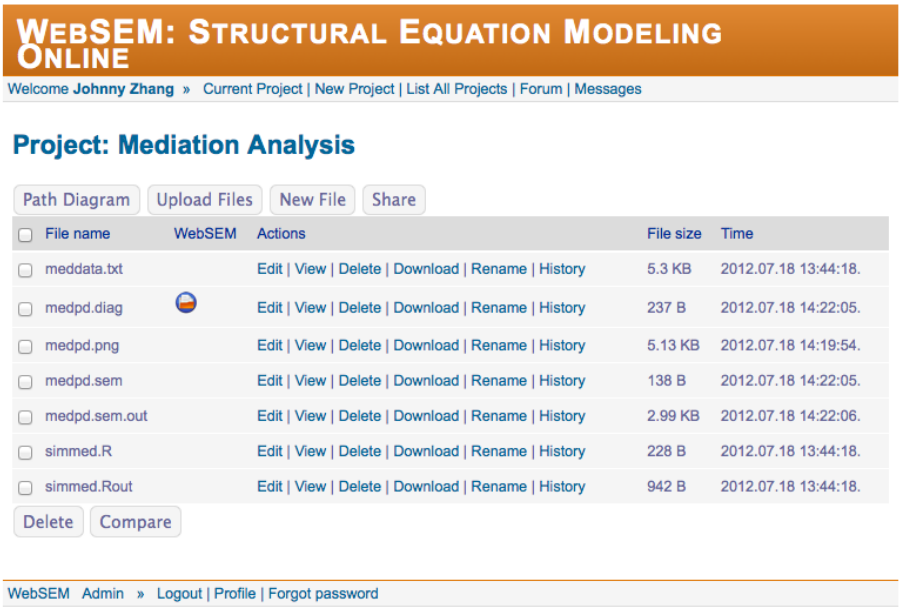


Figure 1.2.14: The menu for path diagram

It includes the following files:

- `meddata.txt`: simulated data file
- `medpd.diag`: path diagram for mediation analysis. One can open the path diagram directly by clicking on the icon.
- `medpd.png`: path diagram as a figure file
- `medpd.sem`: SEM input file generated from the path diagram. One can edit the file directly to run SEM analysis without drawing a path diagram.
- `medpd.sem.out`: SEM output for mediation analysis
- `simmed.R`: R codes to simulate mediation data
- `simmed.Rout`: R output to run the simulation codes.

One can also share the analysis with another user by clicking on `Share` button. The user you share the analysis with can run and edit your analysis but cannot delete any files.

1.2.5 Access your analysis directly on your computer

1.2.5.1 Through a web browser

You can always access your analysis using a web browser on any device including mobile devices such as Iphone and Ipad.

1.2.5.2 On a Mac through WebDav

In Finder, click `Go` - `Connect to server`. Then in the `Sever Address` of the pop out window, type `https://websem.psychstat.org/analysis/johnny`. Then in the login window, type in your user name and password.

1.2.5.3 On iPad or iPhone

1.2.5.4 On Windows PC

Chapter 2

Using scripts and more on model building

WebSEM also supports the use of scripts other than path diagram to run SEM analysis. For example, for the mediation analysis, the script looks like

```
MODEL:                                  1
x, 1, a, 1, m,1                         2
m, 1, b, 1, y,1                         3
x, 1, c, 1, y,1                         4
DATA:                                   5
meddata.txt                             6
WEIGHT:                                 7
                                        8
GROUP:                                  9
                                        10
CONSTRAINTS:                            11
ab==a*b                                 12
                                        13
CONTROL:                                14
                                        15
SOFTWARE:                               16
lavaan                                  17
```

Note that everything followed by colon represents keywords.

2.1 MODEL

The model can be generated by drawing path diagrams. A model is specified by paths with the following meaning

from	from type	label	arrow type	to	to type
variable name	1: observed (square)		1: one-headed		1: observed (square)
	2: latent (circle)		2: two-headed		2: latent (circle)
	3: triangle (intercept)				3: triangle (intercept)

- Each line represents one path with the elements in the table above. Each element is separated by comma.
- From is the starting variable, can be observed (1), latent (2) or intercept (3).
- To is the end variable.
- Label can have different specification:
  - No label: a free parameter to estimate
  - With a label: a free parameter to estimate and named by the label
  - `@`: a fixed parameter automatically at 1.
  - `@0.5`: a fixed parameter at 0.5
  - `par1 @ 0.5`: a fixed parameter at 0.5 with name `par1`
  - `?`: a free parameter
  - `?0.5`: a free parameter with starting value 0.5
  - `par2 ? 0.5`: a free parameter with starting value 0.5 and name par2
  - Same label represents the same parameters.
  - For multiple group analysis, multiple label can be given and separated by semicolon ";"
    - `a1;a2@.5;.3`: for the first group, a1 fixed at 0.5 and a2 fixed at 0.3
    - `d1;d1`: for both group, the parameter is the same at d1
    - `b1;b2?0;0`: starting values at 0 for both group.
- Arrow type: 1 single-headed and 2 double-headed.

2.1.1 Specifying a model using equations

One can also specify a model using equations to simply the scripts as shown below. Note that MODEL(EQ) is used to distinguish the current model specification from the path methods. To specify a regression, use "`=`" and to specify a factor model, use "`==`". For example, `y=x` means `y` regresses on `x`. `f==x1+x2+x3` means factor `f` is indicated by three `x`s.

```
MODEL(EQ):
y = x +m
m = x
DATA:
meddata.txt
WEIGHT:
```

```
GROUP:

CONSTRAINTS:
ab == a*b
CONTROL:

SOFTWARE:
lavaan
```

2.2 DATA

Data should be in text file and free format. Any line starting with "#" will be considered as comments and not treated as data. The first line should be variables names. Data should be separated by space. Here, a data file name is provided.

2.3 WEIGHT

The weight to be used in the robust analysis. The default weight is 0.1 is not specified.

2.4 GROUP

The grouping variable in a multiple group analysis. A parenthesis can be used to specify the base group. For example, `gender (1, 2)` means gender is the grouping variable and `1` is the base group.

2.5 CONSTRAINTS

The following constraints can be used: `>, <, =`. For example, `par1 > par2`. `par3 >0`. `par4 = par1*par1`.

New parameters can also be defined here using `==`. For example `ab == a*b`.

2.6 CONTROL

Not available yet.

2.7 SOFTWARE

The software to use for analysis. For example, `rsem` will use the robust method. `lavaan` can be used to use R package lavaan for sem analysis.

Chapter 3

Some examples

3.1 Mediation analysis

Note the following: (1) `ab=a*b` is the mediation effect and `abc=a*b+c` is the total effect. (2) `bmem` is selected for analysis. (3) `bootstrap=100` is used to bootstrap the analysis for 100 times.

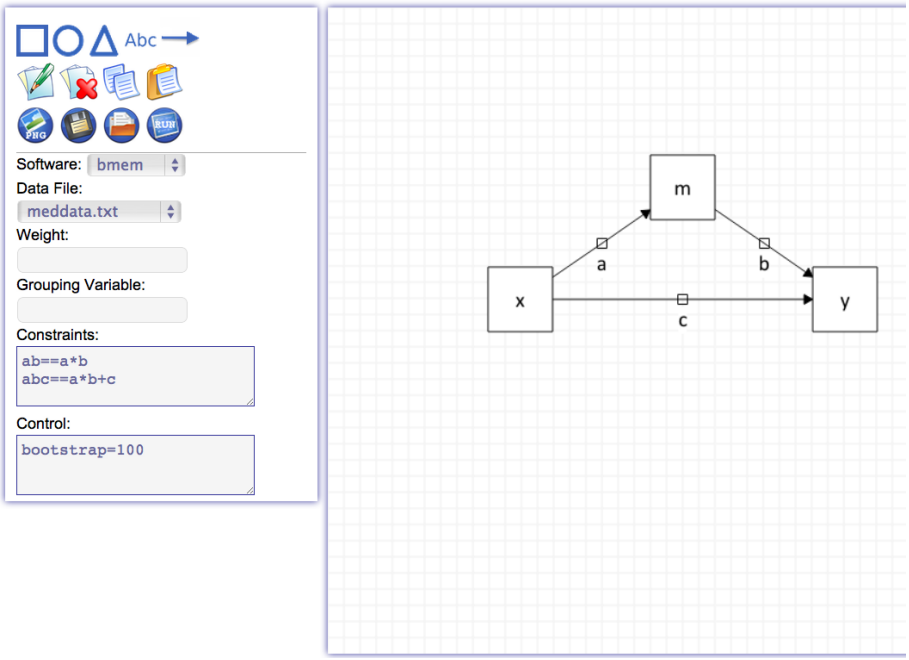


Figure 3.1.1: The interface for mediation analysis

The main output of the analysis is bias-corrected confidence intervals as shown below.

```
The 95% bias-corrected confidence intervals
     estimate  se.boot      2.5%   97.5%
a     0.86509 0.089942  0.675937 1.04051
b     0.29572 0.106734  0.111905 0.52782
c     0.42997 0.146703 -0.033745 0.65071
m~~m  0.85899 0.114341  0.602069 1.02165
y~~y  1.04643 0.115673  0.864759 1.30257
ab    0.25583 0.099242  0.097578 0.44818
abc   0.68580 0.105222  0.470111 0.87102
```

3.2 Factor analysis

1. Fixed loadings can be specified by changing the name of a path.
2. When factor means are estimated, the means/intercepts for observed variables are automatically fixed at 0.

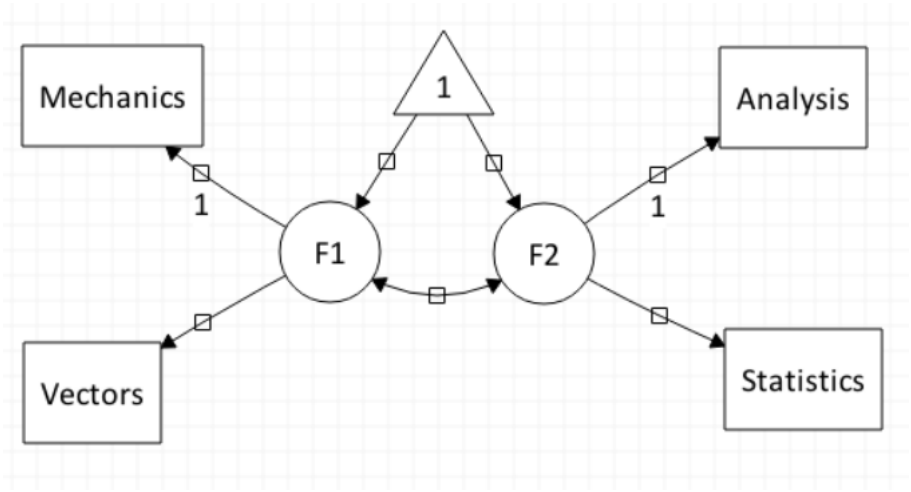


Figure 3.2.1: A factor model

3.2.1 The sample output

```
Descriptive statistics (N=88)
             Mean     sd Min Max Skewness Kurtosis
Mechanics  38.955 17.486   0  77 -0.32417   2.5008
Vectors    54.860 12.869   9  82 -0.73071   4.6615
Algebra    50.602 10.625  15  80 -0.31788   4.1573
Analysis   46.682 14.845   9  70 -0.62888   2.4997
Statistics 49.175 15.889  24  81  0.49404   2.1350
```

```
Model information
Observed variables: Analysis Statistics Mechanics Vectors .
Latent variables: F2 F1 .
The weight is: 0.1 .
The software to be used is: websem .
Results

  Statistic                                          ML
  Value                                           3.549
  Degrees of freedom                                  3
  P-value                                         0.314

  Statistic                                         RML
  Value                                           1.365
  Degrees of freedom                                  3
  P-value                                         0.714

  Statistic                                         AML
  Value                                           1.207
  Degrees of freedom                              0.340
  P-value                                         0.690

  Statistic                                        CRADF
  Value                                           1.417
  Degrees of freedom                                  3
  P-value                                         0.702

  Statistic                                          RF
  Value                                           0.469
  Degrees of freedom 1                            3.000
  Degrees of freedom 2                           85.000
  P-value                                         0.705

                   Estimate        SE  Z-value  P-value
Latent variables:
  F2 =~
    Analysis          1.000
    Statistics        0.876    0.054   16.263    0.000
  F1 =~
    Mechanics         1.000
    Vectors           1.289    0.058   22.358    0.000

Covariances:
  F2 ~~
    F1               78.812   23.622    3.336    0.001
```

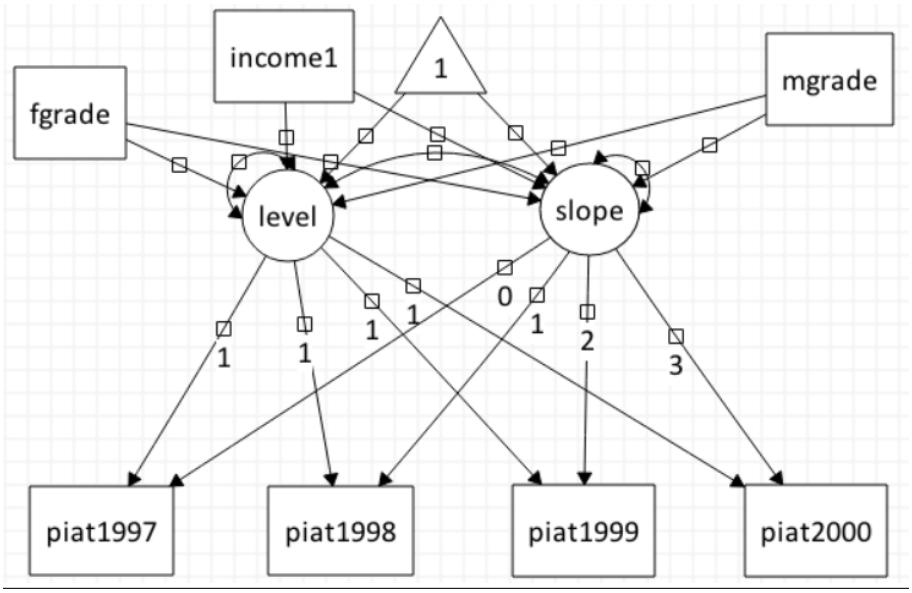
```
Intercepts:
    F1                39.447     1.832    21.528     0.000
    F2                47.192     1.660    28.430     0.000
    Analysis           0.000
    Statistics         0.000
    Mechanics          0.000
    Vectors            0.000

Variances:
    Analysis          10.546    55.674     0.189     0.850
    Statistics       203.807    41.614     4.898     0.000
    Mechanics        180.696    30.890     5.850     0.000
    Vectors           41.574    27.272     1.524     0.127
    F2               192.694    54.434     3.540     0.000
    F1                87.660    34.304     2.555     0.011
```

3.3 Robust growth curve model

1. The predictors are corrected by default.

Algorithm 3.1 A growth curve model with predictors



3.4 Multiple group analysis

1. Grouping variable needs to be specified. `gender(1,2)` the categories are put in the parentheses and the first number is the base group.

2. To equal a parameter for multiple groups, use the same as like `b;b`. Note labels are separated by `;`.

3. Indirect effects can also be calculated for multiple groups.

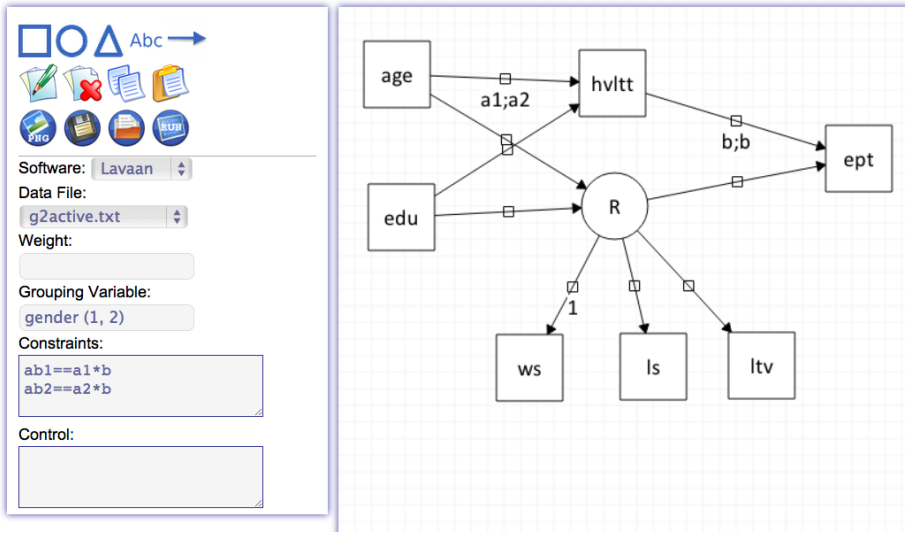


Figure 3.4.1: A multiple group analysis example

3.4.1 Sample output

```
Descriptive statistics (N=1114)
              Mean       sd Min Max Skewness Kurtosis
training   0.43268  0.49567   0   1  0.27177   1.0739
gender     1.76930  0.42147   1   2 -1.27848   2.6345
age       72.44704  5.14772  65  91  0.68823   2.9710
edu       13.73698  2.60001   6  20  0.41854   2.9679
hvltt     27.44794  4.76392   9  36 -0.66130   3.3883
ws        10.98923  4.96003   1  30  0.47037   3.1397
ls        11.65619  5.63229   0  30  0.33530   2.6471
ltv        6.47846  2.73574   0  15  0.26518   2.7027
ept       20.29623  5.12698   3  28 -0.72079   2.9922

Model information
Observed variables: age hvltt edu ept ws ls ltv .
Latent variables: R .
Grouping variable: gender .
The weight is: 0 .
The software to be used is: lavaan .
```

```
The following output is from Lavaan.
lavaan (0.4-14) converged normally after 101 iterations

  Number of observations per group
  1                                                257
  2                                                857

  Estimator                                         ML
  Minimum Function Chi-square                  827.975
  Degrees of freedom                                33
  P-value                                        0.000

Chi-square for each group:

  1                                            231.010
  2                                            596.966

Chi-square test baseline model:

  Minimum Function Chi-square                 3367.648
  Degrees of freedom                                40
  P-value                                        0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.761
  Tucker-Lewis Index (TLI)                       0.710

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)             -20993.998
  Loglikelihood unrestricted model (H1)     -20580.010

  Number of free parameters                         27
  Akaike (AIC)                               42041.996
  Bayesian (BIC)                             42177.420
  Sample-size adjusted Bayesian (BIC)        42091.661

Root Mean Square Error of Approximation:

  RMSEA                                          0.208
  90 Percent Confidence Interval          0.196  0.220
  P-value RMSEA <= 0.05                          0.000

Standardized Root Mean Square Residual:
```

```
  SRMR                                            0.173

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

Group 1 [1]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  R =~
    ept               0.818    0.032   25.819    0.000
    ws                1.000
    ls                1.083    0.021   51.922    0.000
    ltv               0.587    0.012   48.153    0.000

Regressions:
  hvltt ~
    age       (a1)    0.190    0.022    8.608    0.000
  ept ~
    hvltt      (b)    0.439    0.011   40.058    0.000
  R ~
    edu               0.888    0.081   10.929    0.000
    age              -0.020    0.017   -1.183    0.237
  hvltt ~
    edu               0.831    0.106    7.852    0.000

Intercepts:
    ept               0.000
    ws                0.000
    ls                0.000
    ltv               0.000
    hvltt             0.000
    R                 0.000

Variances:
    ept              14.806    1.437
    ws                7.849    0.936
    ls                7.789    0.993
    ltv               3.124    0.355
    hvltt            24.795    2.187
    R                12.057    1.308
```

```
Group 2 [2]:

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  R =~
    ept               0.704    0.027   25.675    0.000
    ws                1.000
    ls                1.068    0.011  101.577    0.000
    ltv               0.565    0.007   79.679    0.000

Regressions:
  hvltt ~
    age       (a2)    0.221    0.011   19.274    0.000
  ept ~
    hvltt      (b)    0.439    0.011   40.058    0.000
  R ~
    edu               0.949    0.051   18.682    0.000
    age              -0.027    0.009   -2.877    0.004
  hvltt ~
    edu               0.870    0.061   14.211    0.000

Intercepts:
    ept               0.000
    ws                0.000
    ls                0.000
    ltv               0.000
    hvltt             0.000
    R                 0.000

Variances:
    ept              12.038    0.633
    ws                4.891    0.389
    ls                6.700    0.483
    ltv               4.111    0.232
    hvltt            23.616    1.141
    R                14.066    0.799


Defined parameters:
    ab1               0.083    0.010    8.416    0.000
    ab2               0.097    0.006   17.368    0.000
```