

## Online Appendix

Online appendix to accompany Pek, J. & Chalmers, R. P. (2015). Visualizing Confidence Bands for Semiparametrically Estimated Nonlinear Relations among Latent Variables. *Journal of Educational and Behavioral Statistics*. doi: 10.3102/1076998615589129

### **plotSEMM: A Web Application in R ‘shiny’**

We introduce a web application for users wishing to employ structural equation mixture models (SEMM) as a semiparametric model (SPM) in an exploratory fashion to describe bivariate latent functions without specification of their form. This utility expands on two existing tools, an online utility and an R (R Core Team, 2013) package ‘plotSEMM’, that generates two plots to aid in exploring potentially nonlinear latent relationships (Pek, Sterba, Kok, & Bauer, 2009). The first plot superimposes the aggregate latent function of Equation (9) on the model-implied bivariate contour plot of the two latent variables, and presents model-implied marginal distributions of the latent predictor and outcome at the margins of the plot. Additionally, the locally linear or within-class regression functions of Equation (7) and within-class marginal distributions for the latent variables could be plotted or suppressed. The second plot presents mixing probabilities across fixed or conditional values of the latent predictor  $\eta_1$  as defined in Equation (9) along with the marginal distribution of  $\eta_1$ . This second plot provides the added information on how the aggregate function shifts from one latent class to another across the range of the latent predictor  $\eta_1$ .

**New Features.** The new utility subsumes these plots and re-packages the existing R functions as an interactive web application with the ‘shiny’ package (RStudio, Incorporation, 2014). In addition to the two plots described above, plots of the aggregate function along with different types of approximate confidence bands (CBs) such as the delta method Wald-type or parametric bootstrap confidence intervals (CIs) or confidence envelopes (CEs) could be generated. Specifically, four different kinds of CBs are available for plotting: (a) delta method Wald-type CIs, (b) parametric bootstrap CIs, (c) delta method Wald-type CEs and (d) parametric bootstrap CEs. The utility can also generate point-wise CIs about latent predicted values  $E[\eta_2|\eta_1]$  with user-specified values on  $\eta_1$ . Importantly, because CEs could inform of the form of the unspecified latent function, the utility also includes an implementation of a line finding algorithm (Pek & Chalmers, 2015) which can diagnose potential bivariate nonlinearity.

**Launching the Web Application.** The web application is an updated version of the ‘plotSEMM’ package, which requires some familiarity with R, as the application is launched within the R environment. Entering the following command lines in R

```

1 install.packages("plotSEMM")
2 library(plotSEMM)
3 plotSEMM_GUI()

```

would install ‘plotSEMM’ and launch the web application as shown as a screenshot in Figure A. Note that the first line of code only needs to be run once. With user input, the web interface dynamically generates R code which is submitted to R on the user’s computer to generate plots within the web browser; R should be left open to run in the background while using the web application. Note that the web application has been optimized for Google Chrome and Mozilla Firefox.

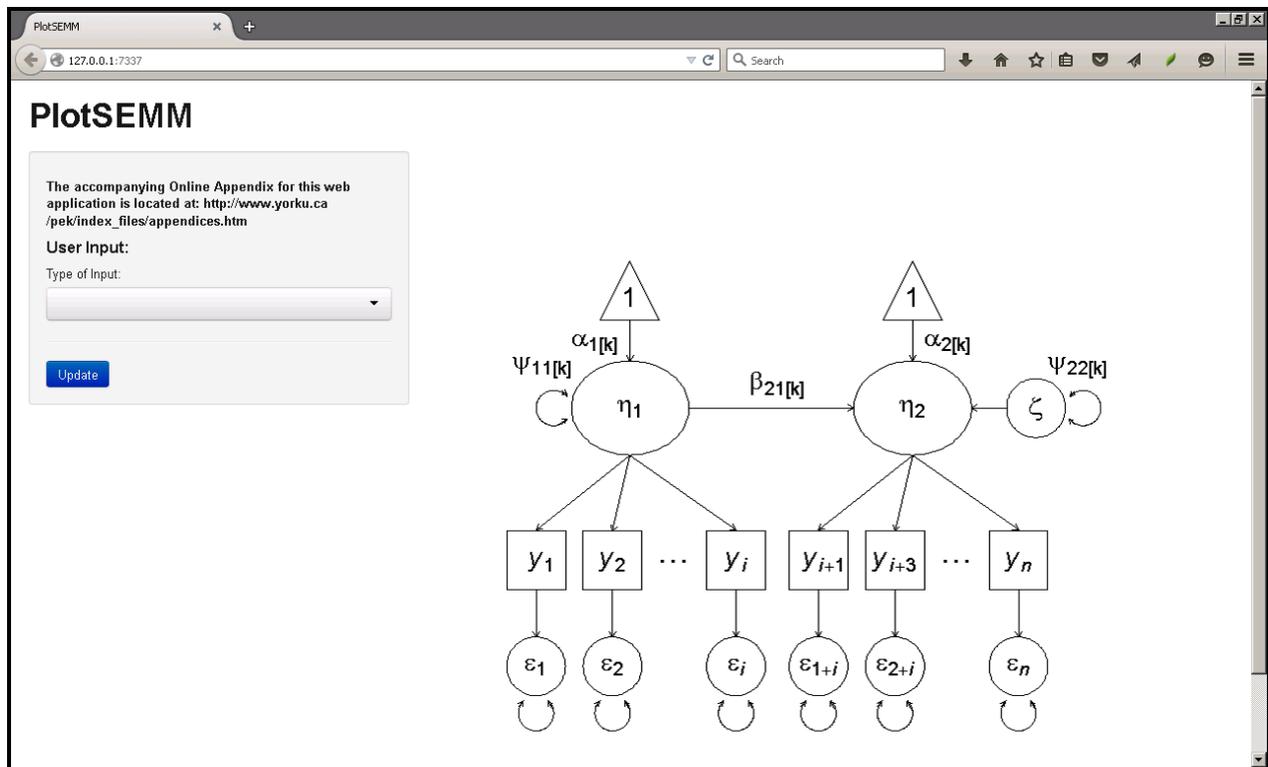


Figure A: Screenshot of the web application.

**User Input and Controlling Plots.** There are two approaches to inputting parameter estimates to generating plots, which is selected by clicking on the drop down menu under “Type of Input”, choosing either (a) Manual Input or (b) Mplus Files, and then clicking on the “Update” button. Clicking on “Update” will refresh the web page where more information for input is requested.

**Manual Input.** The manual approach mirrors the existing online utility, and can

be used with any statistical software that fits SEMMs to data such as Mplus (Muthén & Muthén, 2011) and OpenMx (Boker et al., 2011). When this option is selected, the number of classes for the model should be specified and updated by clicking on the “Update” button. Then, for each class in the model, users are required to input point estimates for the probability of class membership ( $\pi$ ), mean of the latent predictor ( $\alpha_1$ ), intercept ( $\alpha_2$ ) and slope ( $\beta_{21}$ ) of the linear regression of the latent outcome  $\eta_2$  on the latent predictor  $\eta_1$ , the variance of the latent predictor ( $\psi_{11}$ ) and the variance of the latent outcome ( $\psi_{22}$ ). Point estimates that are keyed in requires a number preceding decimals (i.e., ‘0.5’ and not ‘.5’).

After manually keying in these point estimates, users may choose to have either the contour plot or the probability plot generated by making a selection under “Type of Output” in the drop down menu and clicking on the “Update” button. The first plot includes the bivariate contour plot of the latent variables, the marginal distributions of the latent variables and the aggregate latent function. The probability plot includes the mixing probabilities at conditional values of the latent predictor  $\eta_1$  and the marginal distribution of  $\eta_1$ . By default, within-class information for the contour plot will be shown. Class information of the contour plot could be suppressed by first checking the option “Override various plotting defaults” and clicking on “Update”. After the web page refreshes, within-class information is suppressed by unchecking the option “Show class specific distributions, regression lines, and mixing probabilities for Contour plot” and clicking on “Update”.

Plots of CEs about the aggregate latent function require information from the  $p \times p$  estimated asymptotic covariance matrix of estimates  $\widehat{VAR}(\hat{\theta}_p)$ , making the manual input approach cumbersome and error prone. Hence, the second method and default of reading Mplus files directly allows for the automatic generation of CBs.

**Mplus Files.** The approach of reading in Mplus files requires users to point to the location of the folder containing Mplus output files under “Directory containing Mplus Files”. These are the .out and two files which contain information about the  $p$  point estimates  $\hat{\theta}_p$  and the asymptotic covariance matrix  $\widehat{VAR}(\hat{\theta}_p)$ . To obtain the proper files, the following three lines of code should be present in the Mplus input file. Note that the directory should contain only files from a *single* model. If the contour and probability plots are desired, only the Mplus .out file is required; plots of the CBs require all three files.

```

1 | output: tech1 tech3;
2 | savedata: tech3=acov.dat;
3 | savedata: results=result.dat;

```

By specifying `tech1`, the .out file would contain information on the parameter specifi-

cation of the SEMM which is essential for the application to link parameters to their estimates. With the `tech3` command, estimates of the  $p \times p$  asymptotic covariance matrix of estimates  $\widehat{VAR}(\hat{\theta}_p)$  are made available and output into the `acov.dat` file with the `savedata` command. Finally, the third line of code outputs the  $p$  point estimates into the `result.dat` file which contains more precise point estimates than present in the `.out` file.

The directory containing the three Mplus files can be input into the web application, or with the `setwd()` statement before running `plotSEMM_GUI()`. Then, either the contour plot, the probability plot, or a plot of CBs is generated by making the intended selection in the drop down menu under “Type of Output”. Clicking on the “Update” button will refresh the web page with the specified plot. Output options for the contour and probability plots are the same as those for the manual input approach.

By default, 95% delta method Wald-type CIs and CEs are plotted along with the bold aggregate function; the red colored open circles depict point-wise Wald-type CIs and the red colored dashed lines represent the simultaneous Wald-type CEs. Alternatively, 90% CBs could be generated by selecting “90%” instead of “95%” in the drop down menu under “Confidence Level”. Additionally, parametric bootstrap CIs and CEs may be produced by checking the “Parametric Bootstrap Confidence Intervals” and “Parametric Bootstrap Confidence Envelope” options, respectively. These are depicted as blue colored crosses and blue colored dot-dashed lines, respectively. Note that parametric bootstrap CIs and CEs are empirically generated and tend to be computationally intensive, especially for the latter. The time taken to estimate parametric bootstrap CEs for models with a large number of  $p$  parameters is highly dependent on the specifications of the computer used, and could take up to 30 minutes or more to estimate.

Confidence envelopes provide users with the ability to diagnose nonlinearity of the recovered latent regression by conducting a test about the null hypothesis  $H_0 : E[\eta_2|\eta_1 = \alpha_0 + \beta_0\eta_1]$ , where  $\alpha_0$  and  $\beta_0$  can take on any real value. This null hypothesis is rejected when the CE does not contain any linear function, suggesting nonlinearity. Conversely, when a line is enclosed within the CE, linearity of the latent function is suspect. A graphical implementation of this test is available to users in the form of a line finding algorithm, which is engaged by default for the delta method CE. To suppress the algorithm, users could uncheck the “Run Line Finding Algorithm” option and click on the “Update” button. When linearity is suspect, a line that is located within the CE will be plotted. This line will be solid bold and pink in color for the delta method CE. In contrast, this line will be a solid, bold and light blue in color for the parametric bootstrap CE. When no line is

plotted, nonlinearity of the latent function is suggested, and the message “No Line was Found Within the Confidence Envelope(s)” will be displayed on top of the graphic.

**Additional Options.** The web application allows for some flexibility in modifying axis labels, changing the location of the legend as well as the number of points used to generate the plots. These options are available when the option “Override various plotting defaults” is checked and the “Update” button is clicked. For all plots, font size in the plots could be modified by changing the value under “Font size (cex).” However, not all options are available for all three plots. Specifically, both axes may be relabeled for the contour and CB plots, and only the latent predictor may be renamed for the probability plot. Also, only class information in the contour plot can be suppressed. The number of points used to generate the graphics could be changed for all three plots. Increasing the number of points to plot in “Number of points” reduces the granularity of the plots, but increases the computational time needed to estimate parametric bootstrap CEs and complete running the line finding algorithm. Additionally, targeted simulations (Pek & Chalmers, 2015) confirm that as the number of points increase, results of the line finding algorithm become more precise. When CEs are graphed, users could also uncheck the option “Allow plot title to indicate ‘No Line was Found within the Confidence Envelope(s)’” to suppress this information.

Point-wise CIs are useful for evaluating hypotheses about the effect of a single latent predictor value  $\eta_1$  on the latent outcome  $E[\eta_2|\eta_1]$ ;  $H_0 : E[\eta_2|\eta_1 = \eta_{10}] = \eta_{20}$ , where  $\eta_{10}$  is a conditional value of the latent predictor under the null hypothesis and  $\eta_{20}$  is the predicted latent outcome value under the null hypothesis. The utility can compute 95% and 90% delta method Wald-type or parametric bootstrap CIs based on user-specified values on  $\eta_1$ . When “Type of Output” is selected as “Confidence Bands (Mplus Input Only)”, a value of  $\eta_1$  could be entered under “Conditional Value of Latent Predictor ( $\eta_1$ ) for Confidence Interval(s) about  $E[\eta_2|\eta_1]$ ”. The estimated user-specified CI(s), which are consistent with the type of CI(s) selected to be plotted, would also be reported under the plot of CB(s). These estimated user-specified CI(s) would have the same confidence level as the CB(s) graphed, which is determined under “Confidence Level”.

Plots generated in web browsers could be output by right clicking on the plot and selecting the typical available options. For instance, a plot could be saved by selecting “Save Image As...”, and typing in the desired file extension such as `.png`, `.jpg`, and `.tiff`. Such images could be cropped or modified in an image editor. Higher quality images (for publication) could be constructed from data generated by the application. An `.Rdata` file containing all the information required to produce the three types of plots may be output into

the working directory in R by checking the option “Save the data used to plot the graphics to working directory in R.” The name of the output `.Rdata` file may be modified under “Saved data file name”; `.Rdata` files may be converted to other file types for alternative graphing software. Variable descriptions in the data file are provided in Table A.

**Table A**

Variables exported into the `.Rdata` file.

Variable	Description
<code>Eta1</code>	Conditional values on the latent predictor $\eta_1$
<code>Eta2</code>	Values on the latent outcome $\eta_2$
<code>agg_denEta1</code>	Marginal distribution of the latent predictor $\eta_1$
<code>agg_denEta2</code>	Marginal distribution of the latent outcome $\eta_2$
<code>agg_pred</code>	Predicted values of the aggregate function $\hat{E}[\eta_2 \eta_1]$
<code>class_pred</code>	Predicted values of the locally linear function by class
<code>contour</code>	Values of the bivariate density between $\eta_1$ and $\eta_2$
<code>class_prob</code>	Mixing probabilities at conditional values of $\eta_1$ by class
<code>class_denEta1</code>	Within-class distribution of the latent predictor $\eta_1$
<code>class_denEta2</code>	Within-class distribution of the latent outcome $\eta_2$
<code>bs_CIlo</code>	Lower bound values of the parametric bootstrap CI
<code>bs_CIhi</code>	Upper bound values of the parametric bootstrap CI
<code>delta_CIlo</code>	Lower bound values of the delta method Wald-type CI
<code>delta_CIhi</code>	Upper bound values of the delta method Wald-type CI
<code>delta_CElo</code>	Lower bound values of the delta method Wald-type CE
<code>delta_CEhi</code>	Upper bound values of the delta method Wald-type CE
<code>alpha</code>	Type I error rate
<code>bs_CElo</code>	Lower bound values of the parametric bootstrap CE
<code>bs_CEhi</code>	Upper bound values of the parametric bootstrap CE

## Examples

Two data sets of 500 cases each were generated to show how the SEMM can recover linear and nonlinear (quadratic) latent functions without their explicit specification, and how CEs used in conjunction with the line finding algorithm can diagnose latent nonlinearity. Details on data generation and model fitting are provided in Pek, et al. (2009). All models were estimated using Mplus 7.1 (Muthén & Muthén, 2011), and the Mplus files containing the information needed to generate these plots are provided online at [http://www.yorku.ca/pek/index\\_files/appendices.htm](http://www.yorku.ca/pek/index_files/appendices.htm). Mplus files for the empirical example in the manuscript are also provided.

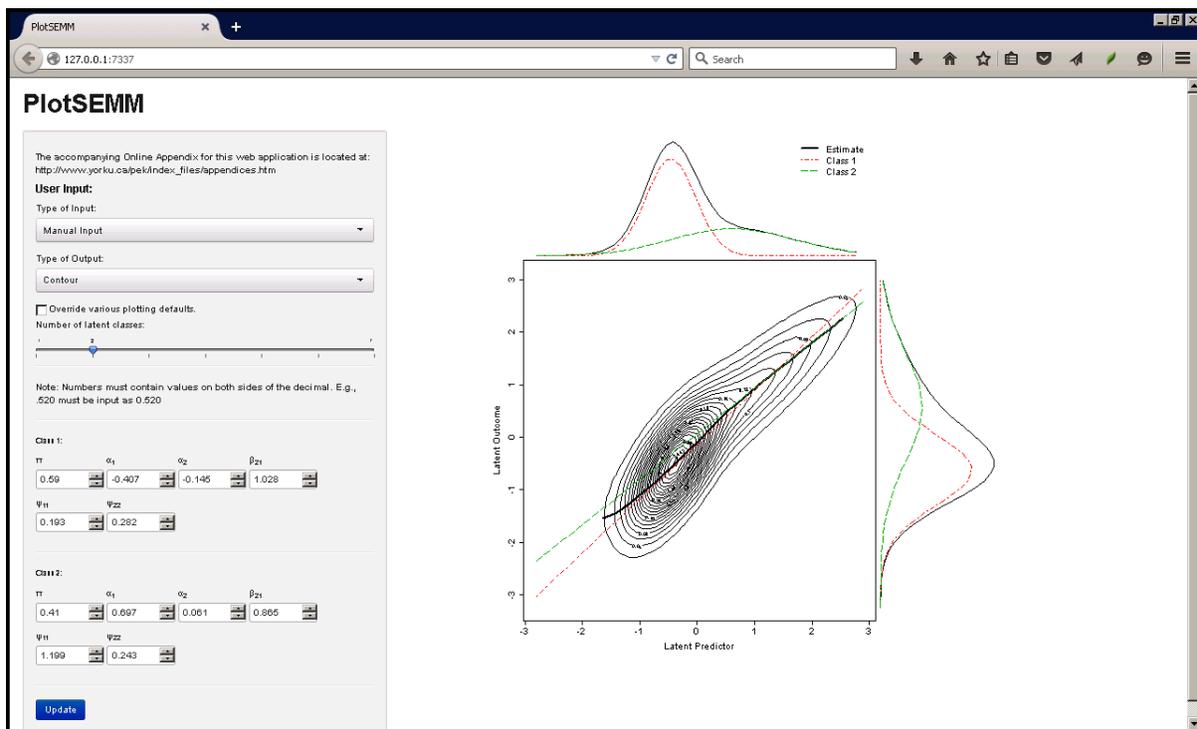


Figure B: Screenshot of the generated linear function, bivariate contour plot, and marginal distributions of the latent variables.

**Linear Function.** We first illustrate the approach of manually keying in estimates into the web application. To the left of Figure B are values of estimates manually input to generate the plot. Here, two classes were found to be optimal for the linear function. For each class, 6 estimates are input, and a total of 12 estimates are keyed in. Clicking on the “Update” button returns the contour plot as shown. To suppress class information, rename the axes, move the legend and alter the number of points plotted, the “Override various plotting defaults” should be checked, and the web page refreshed by clicking on “Update”. Figure C is a screenshot of the probability plot, with the option to override defaults selected.

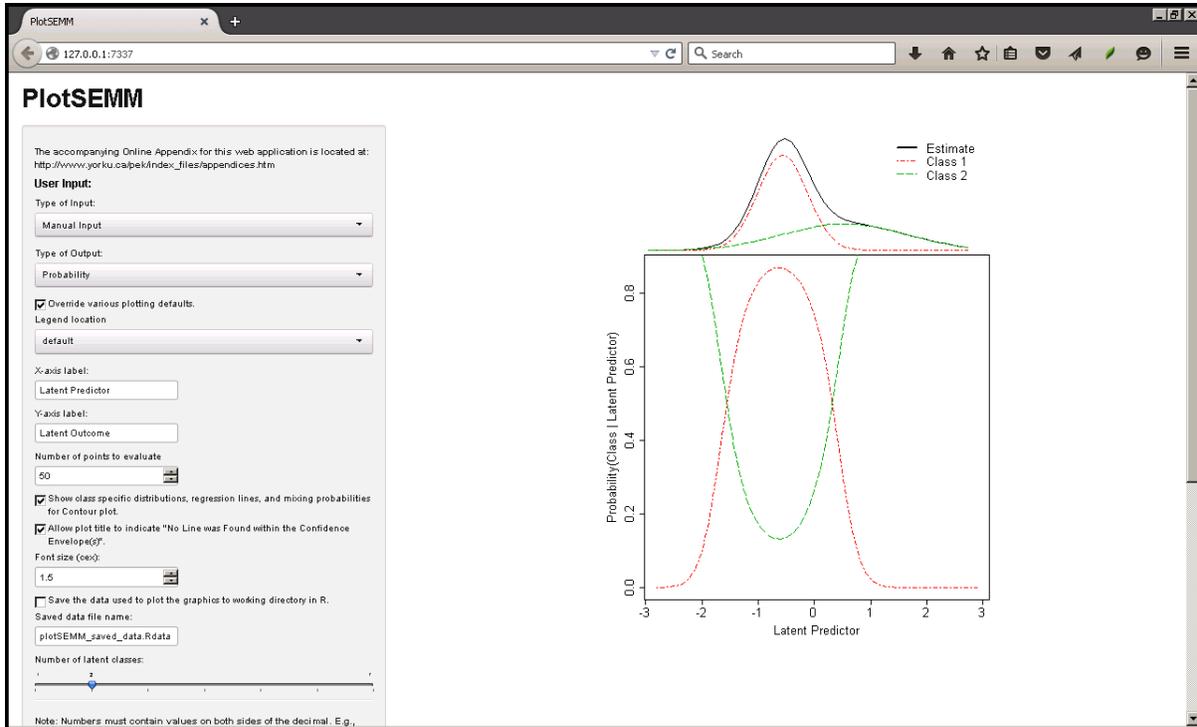


Figure C: Screenshot of other plotting options, with the probability plot.

Recall that CBs can only be generated when three Mplus files are used as input for the utility. In Figure D, a screenshot of the web application is provided for this type of input, and 95% delta method Wald-type as well as parametric bootstrap CIs are presented in the plot. Additionally, as  $\eta_1 = 1.6$  was specified under “Conditional Value of Latent Predictor ( $\eta_1$ ) for Confidence Interval(s) about  $E[\eta_2|\eta_1]$ ”, the point estimate for the predicted latent outcome  $\hat{E}[\eta_2|\eta_1 = 1.6] = 1.45$  and the 95% delta method Wald-type and parametric bootstrap CIs for  $E[\eta_2|\eta_1 = 1.6]$ , which are both (1.26, 1.64), are also reported under the graph. These user-specified estimates will not be reported when no value of  $\eta_1$  is provided (not shown).

In Figure E, 90% delta method Wald-type and parametric bootstrap CEs are presented. Recall that the population generating function for this data is linear. Here, the line finding algorithm was engaged, and both types of CEs were found to contain at least one linear function as expected; the pink bold line is contained within the delta method Wald-type CE and the light blue bold line is contained within the parametric bootstrap CE. Results from the line finding algorithm suggest linearity of the unspecified latent function as both types of CEs were found to contain at least one linear function.

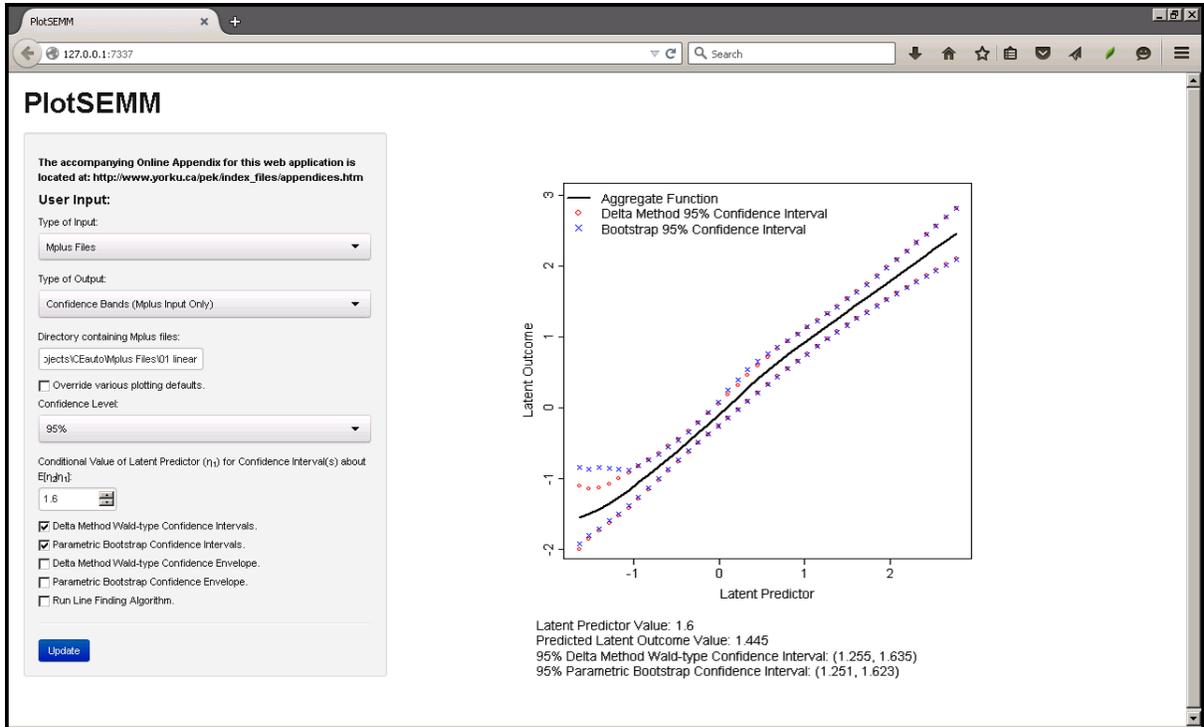


Figure D: Screenshot of linear function with 95% delta method Wald-type and parametric bootstrap confidence intervals, and 95% confidence intervals about  $E[\eta_2|\eta_1 = 1.6]$ .

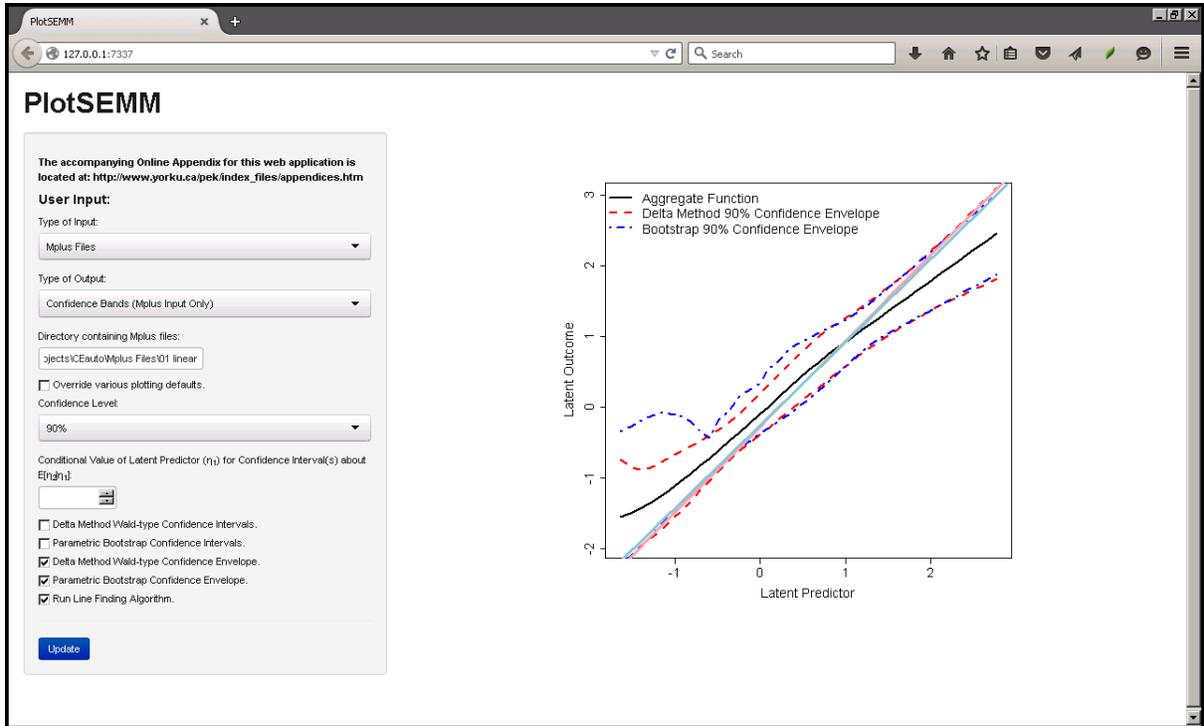


Figure E: Screenshot of linear function with 90% delta method Wald-type and parametric bootstrap confidence envelopes. The pink bold line and light blue bold line are contained within the Wald-type and parametric bootstrap confidence envelopes, respectively.

**Nonlinear Function.** Use of the SEMM as an SPM to explore and recover bivariate latent regressions without their specification is enhanced when nonlinearity can be detected. Figure F below is a screenshot of the web application taking in Mplus files as input and returning a plot of the aggregate function along with 95% delta method Wald-type and parametric bootstrap CEs. As the population generating function is quadratic, the line finding algorithm could not detect a linear function within either CEs as indicated by “No Line was Found within the Confidence Envelope(s)” above the plot. Taken together, these results suggest that the recovered latent function is nonlinear.

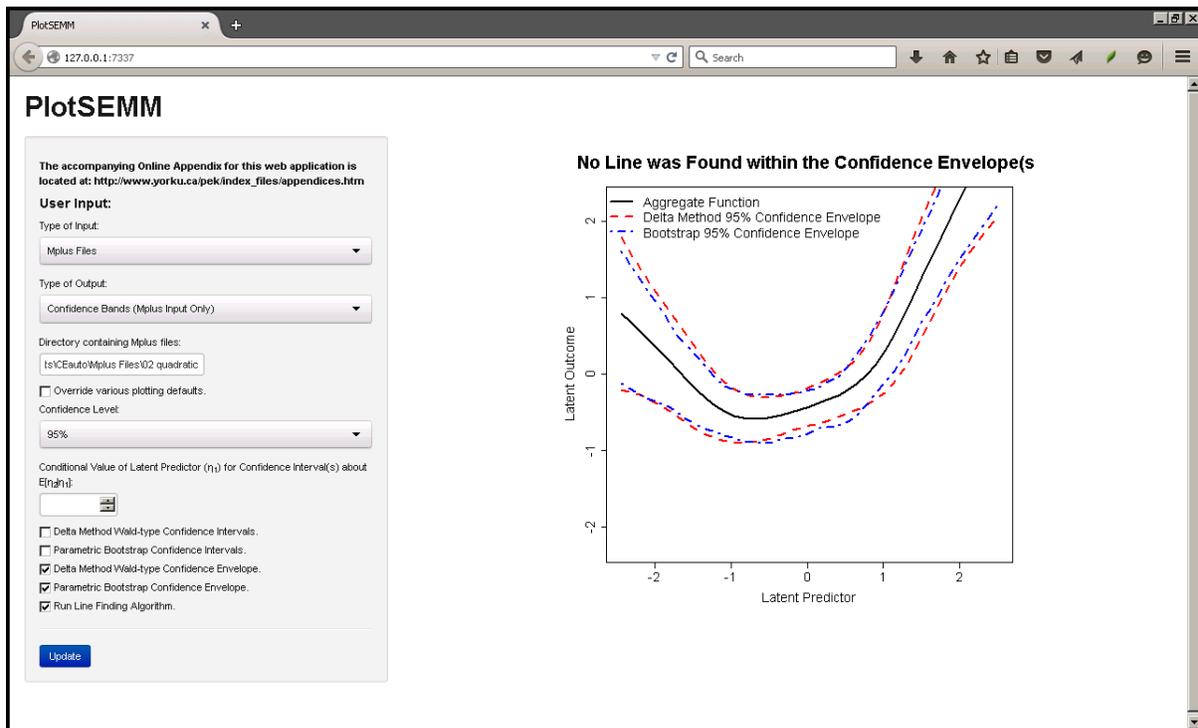


Figure F: Screenshot of nonlinear function with 90% delta method Wald-type and parametric bootstrap confidence envelopes.

## References

- Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., ... others (2011).  
Openmx: an open source extended structural equation modeling framework. *Psychometrika*, *76*, 306–317. doi: 10.1007/s11336-010-9200-6
- Muthén, L. K., & Muthén, B. O. (2011). Mplus user’s guide (6th ed.) [Computer software manual]. Los Angeles, CA: Muthén & Muthén.
- Pek, J., & Chalmers, R. P. (2015). Diagnosing nonlinearity with confidence envelopes for a semiparametric approach to modeling bivariate nonlinear relations among latent variables. *Structural Equation Modeling*. doi: 10.1080/10705511.2014.937790
- Pek, J., Sterba, S. K., Kok, B. E., & Bauer, D. J. (2009). Estimating and visualizing nonlinear relations among latent variables: a semiparametric approach. *Multivariate Behavioral Research*, *44*, 407–436. doi: 10.1080/00273170903103290
- R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- RStudio, Incorporation. (2014). shiny: Web application framework for R [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=shiny>