Bollen & Brand Model - Empirical Example

In this document, we'll apply Bollen and Brand's model to a second empirical data set. The data come from a paper by Orth, Clark, Donnellan, and Robins (2020) called, "Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models." Find the article here and its correspdonding files here. Below, we'll follow the same sequence that we applied in the article: `dp_mod1`, followed by `dp_mod2`, and finally `dp_mod3`.

Study Background & Data

Orth et al. (2020) compiled several data sets, each assessing self-esteem (SE) and depression (DE) on multiple people over four waves. Specifically, their Open Science Framework (OSF) folder houses data from the following studies:

- Berkeley Longitudinal Study (BLS)
- California Families Project, children sample (CFP-C)
- California Families Project, mothers sample (CFP-M)
- My Work and I (MWI)
- Your Personality (YP)

We'll use the CFP-C sample in the analysis below. The CFP is an ongoing study of 674 Mexican-origin families from Northern California. Orth et al. (2020) report that their data come from waves 1, 3, 5, and 7, which were captured over two years among 674 adolescents. Self-esteem was measured with the Self-Description Questionnaire (SDQ), and depression was measured with the Early Adolescent Temperament Questionnaire-Revised (EATQ). The scores listed in their data are scale aggregates (i.e., mean scores across items on SDQ and EATQ, respectively). Citations for Orth et al.'s excellent manuscript and original data sources are shown below.

- Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2020). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology.*

- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 27*, 151-161.

- Robins, R. W., & Conger, K. J. (2017). *California Families Project [Sacramento and Woodland, California]: Item-level (producer) codebook.* Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

- Meier, L. L., & Spector, P. E. (2013). Reciprocal effects of work stressors and counterproductive work behavior: A five-wave longitudinal study. *Journal of Applied Psychology, 98*, 529-539.

- Orth, U., & Luciano, E. C. (2015). Self-esteem, narcissism, and stressful life events: Testing for selection and socialization. *Journal of Personality and Social Psychology, 109*, 707-721.

Data Cleaning

Load the data, and load necessary libraries.

```
library(tidyverse)
library(ggplot2)
library(lavaan)
```

```r
library(kableExtra)
library(tseries)
library(plm)

df <- read.table("CFP-C.dat")
```

Orth et al.'s item dictionary states the following:

- In the data files for the CFP-C, the order of variables is as follows: ID, SE1, SE2, SE3, SE4, DE1, DE2, DE3, DE4.

So, the first column lists participant ID's, the second self-esteem at time 1, the third self-esteem at time 2, and so on until depression at time 4. Notice that the data are already in wide form. Let's rename the columns to cohere with these labels.

```r
names(df) <- c("id",
                'se.1',
                'se.2',
                'se.3',
                'se.4',
                'de.1',
                'de.2',
                'de.3',
                'de.4')
```

In their description of the data, the authors state that missing values are coded as `-999`. Moreover, participant ID's were replaced with `-999`. You can see these aspects by peaking at the data.

```r
head(df) %>% kable() %>% kable_styling()
```

id	se.1	se.2	se.3	se.4	de.1	de.2	de.3	de.4
-999	2.88000	3.120000	3.00	3.20	1.666667	1.333333	1.333333	1.166667
-999	3.48000	2.920000	2.80	2.60	2.166667	1.666667	1.833333	2.166667
-999	3.12000	3.360000	2.76	-999.00	2.166667	1.500000	1.833333	-999.000000
-999	3.32000	3.320000	3.20	3.08	2.000000	1.833333	1.666667	1.833333
-999	3.47619	3.280000	3.20	3.32	3.000000	1.333333	1.166667	1.166667
-999	2.76000	3.333333	3.52	3.56	2.333333	1.666667	1.333333	1.333333

We need to replace the ID column with actual ID values (each row is a unique participant – it doesn't matter that these new IDs won't match original participant IDs because we are not merging any data). We also need to replace variables with values of `-999` with `NA`.

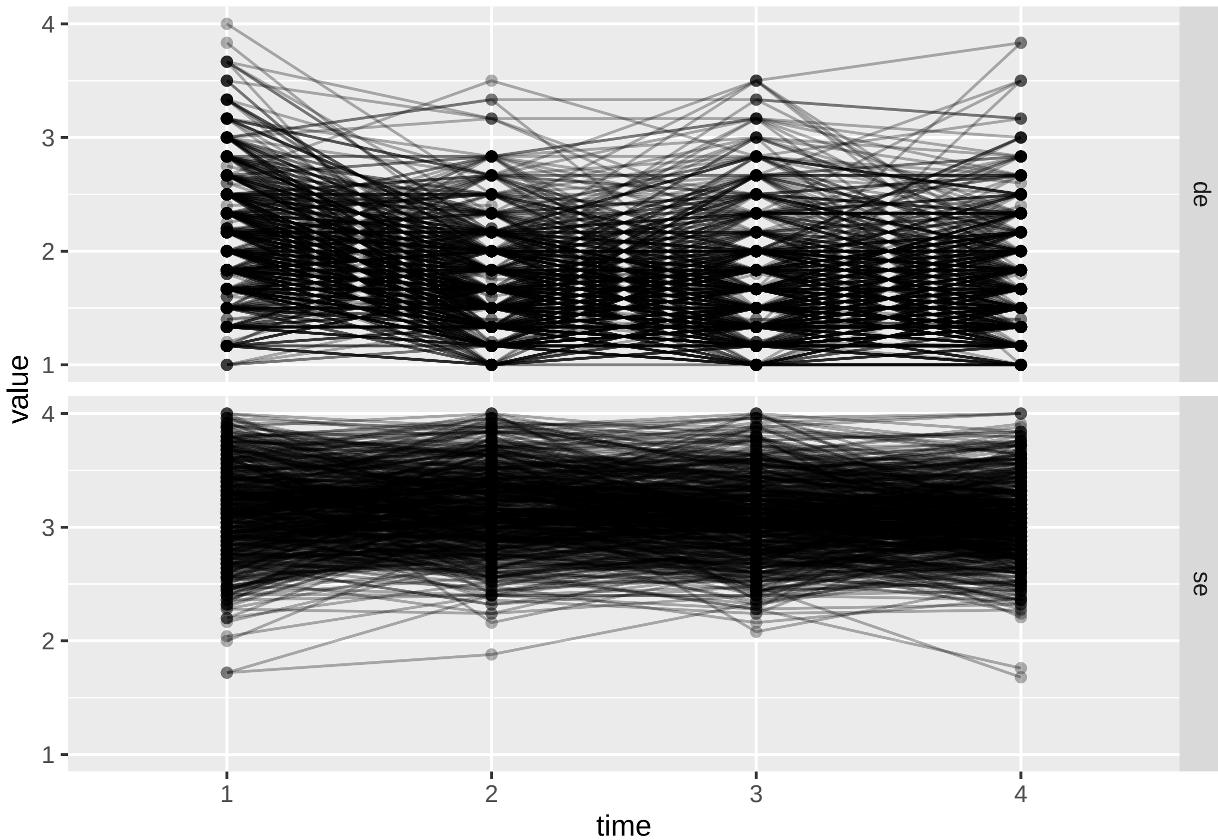
```r
rp_na <- function(x){ifelse(x == -999, NA, x)}
df <- df %>%
  # id values
  mutate(id = 1:nrow(df)) %>%
  # replace -999 with NA
  mutate_if(is.numeric, rp_na)
```

Great. Let's plot the data by changing it to long form.

```r
df_plot <- df %>%
  pivot_longer(cols = c(se.1,se.2,se.3,se.4,
                        de.1,de.2,de.3,de.4),
              names_to = "variable",
              values_to = "value") %>%
```

```r
  separate(variable, into = c("variable",
                              "time"),
           sep = "\\.")

ggplot(df_plot, aes(x = time, y = value, group = id)) +
  geom_point(alpha = 0.3) +
  geom_line(alpha = 0.3) +
  facet_grid(vars(variable))
```



Last, to make the tutorial as simple as possible, we'll select complete cases and rename the object to match the label we used in the article. Note that missing data is a serious issue. We are not suggesting that listwise deletion is an appropriate method. We are simply trying to make the models and code easy to understand.

```r
ids_no_missing <- df %>%
  filter(is.na(se.1) == F,
         is.na(se.2) == F,
         is.na(se.3) == F,
         is.na(se.4) == F) %>%
  filter(is.na(de.1) == F,
         is.na(de.2) == F,
         is.na(de.3) == F,
         is.na(de.4) == F
  ) %>%
  pull(id)

df <- df %>%
  filter(id %in% ids_no_missing) -> data_f
```

Modeling

Check stationarity for both variables. Our data is already in wide format, so we need to change it to long format to run the ADF test.

```
data_f_long <- data_f %>%
  pivot_longer(cols = c(se.1,se.2,se.3,se.4,
                        de.1,de.2,de.3,de.4),
               names_to = "variable",
               values_to = "value") %>%
  separate(variable,
           into = c("variable","time"),
           sep = "\\.") %>%
  pivot_wider(names_from = "variable",
              values_from = "value")

data_dickey_structure <- plm.data(data_f_long,
                                  index = c("id", "time"))
```

Run the tests.

```
adf.test(data_dickey_structure$se)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  data_dickey_structure$se
## Dickey-Fuller = -10.882, Lag order = 12, p-value = 0.01
## alternative hypothesis: stationary
```

Self-esteem is stationary,

```
adf.test(data_dickey_structure$de)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  data_dickey_structure$de
## Dickey-Fuller = -12.286, Lag order = 12, p-value = 0.01
## alternative hypothesis: stationary
```

as is depression.

Example 1

The first example models a concurrent effect from self-esteem to depression.

```
dp_mod1 <- "

eta_y =~ 1*de.2 + 1*de.3 + 1*de.4

de.2 ~ rho_y*de.1 + b1*se.2
de.3 ~ rho_y*de.2 + b1*se.3
de.4 ~ rho_y*de.3 + b1*se.4

se.2 ~~ se.3 + se.4
se.3 ~~ se.4
```

```r
de.1 ~~ se.2 + se.3 + se.4
eta_y ~~ de.1 + se.2 + se.3 + se.4

"

dp_mod1_fit <- sem(dp_mod1, data = data_f)
summary(dp_mod1_fit, fit.measures = T, standardized = T)
```

```
## lavaan 0.6-6 ended normally after 70 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         24
##   Number of equality constraints                     4
##
##   Number of observations                           518
##
## Model Test User Model:
##
##   Test statistic                                15.736
##   Degrees of freedom                                 8
##   P-value (Chi-square)                           0.046
##
## Model Test Baseline Model:
##
##   Test statistic                              1063.950
##   Degrees of freedom                                21
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.993
##   Tucker-Lewis Index (TLI)                       0.981
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -1636.262
##   Loglikelihood unrestricted model (H1)      -1628.394
##
##   Akaike (AIC)                                3312.524
##   Bayesian (BIC)                              3397.523
##   Sample-size adjusted Bayesian (BIC)         3334.039
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.043
##   90 Percent confidence interval - lower         0.005
##   90 Percent confidence interval - upper         0.075
##   P-value RMSEA <= 0.05                          0.594
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.029
##
```

```
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   eta_y =~
##     de.2              1.000                                0.219    0.460
##     de.3              1.000                                0.219    0.417
##     de.4              1.000                                0.219    0.448
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   de.2 ~
##     de.1    (rh_y)    0.199    0.033    6.093    0.000    0.199    0.235
##     se.2      (b1)   -0.467    0.046  -10.254    0.000   -0.467   -0.363
##   de.3 ~
##     de.2    (rh_y)    0.199    0.033    6.093    0.000    0.199    0.180
##     se.3      (b1)   -0.467    0.046  -10.254    0.000   -0.467   -0.320
##   de.4 ~
##     de.3    (rh_y)    0.199    0.033    6.093    0.000    0.199    0.213
##     se.4      (b1)   -0.467    0.046  -10.254    0.000   -0.467   -0.335
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   se.2 ~~
##     se.3              0.076    0.007   11.312    0.000    0.076    0.573
##     se.4              0.060    0.006    9.600    0.000    0.060    0.465
##   se.3 ~~
##     se.4              0.071    0.006   11.220    0.000    0.071    0.567
##   de.1 ~~
##     se.2             -0.025    0.009   -2.717    0.007   -0.025   -0.120
##     se.3             -0.026    0.009   -2.956    0.003   -0.026   -0.131
##     se.4             -0.010    0.009   -1.137    0.255   -0.010   -0.050
##   eta_y ~~
##     de.1              0.032    0.009    3.450    0.001    0.148    0.262
##     se.2              0.006    0.007    0.866    0.387    0.026    0.071
##     se.3             -0.006    0.006   -0.874    0.382   -0.026   -0.072
##     se.4             -0.005    0.006   -0.742    0.458   -0.021   -0.060
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .de.2              0.124    0.010   12.085    0.000    0.124    0.549
##    .de.3              0.156    0.012   12.889    0.000    0.156    0.566
##    .de.4              0.117    0.009   12.558    0.000    0.117    0.488
##     de.1              0.316    0.020   16.093    0.000    0.316    1.000
##     se.2              0.137    0.008   16.093    0.000    0.137    1.000
##     se.3              0.129    0.008   16.093    0.000    0.129    1.000
##     se.4              0.123    0.008   16.093    0.000    0.123    1.000
##     eta_y             0.048    0.009    5.551    0.000    1.000    1.000
```

The model fit statistics are $\chi^2(8) = 15.74$; $p < 0.05$; RMSEA $= 0.042$; CFI $= 0.99$; SRMR $= 0.029$, and the standardized coefficient estimates are as follows. Depression had a positive autoregressive effect ($B = 0.20$, $SE = 0.033$, $p < 0.05$). The concurrent relationship between self-esteem and satisfaction was negative ($B =$ -0.47, $SE = 0.046$, $p < 0.05$), such that low self-esteem was association with higher depression at t.

Example 2

The second example demonstrates a lagged, two-variable dynamic panel. Self-esteem will now be evaluated as a lag-one predictor of depression.

```
dp_mod2 <- "

eta_y =~ 1*de.2 + 1*de.3 + 1*de.4

de.2 ~ rho_y*de.1 + b1*se.1
de.3 ~ rho_y*de.2 + b1*se.2
de.4 ~ rho_y*de.3 + b1*se.3

se.1 ~~ se.2 + se.3
se.2 ~~ se.3

de.1 ~~ se.1 + se.2 + se.3
eta_y ~~ de.1 + se.1 + se.2 + se.3

"

dp_mod2_fit <- sem(dp_mod2, data = data_f)
summary(dp_mod2_fit, fit.measures = T, standardized = T)
```

```
## lavaan 0.6-6 ended normally after 61 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         24
##   Number of equality constraints                     4
##
##   Number of observations                           518
##
## Model Test User Model:
##
##   Test statistic                                53.512
##   Degrees of freedom                                 8
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                               899.606
##   Degrees of freedom                                21
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.948
##   Tucker-Lewis Index (TLI)                       0.864
##
## Loglikelihood and Information Criteria:
```

```
##
##   Loglikelihood user model (H0)              -1818.849
##   Loglikelihood unrestricted model (H1)      -1792.092
##
##   Akaike (AIC)                                3677.697
##   Bayesian (BIC)                              3762.697
##   Sample-size adjusted Bayesian (BIC)         3699.213
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.105
##   90 Percent confidence interval - lower         0.079
##   90 Percent confidence interval - upper         0.132
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.039
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   eta_y =~
##     de.2              1.000                               0.255    0.535
##     de.3              1.000                               0.255    0.492
##     de.4              1.000                               0.255    0.518
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   de.2 ~
##     de.1    (rh_y)    0.230    0.036    6.360    0.000    0.230    0.272
##     se.1      (b1)    0.107    0.043    2.517    0.012    0.107    0.092
##   de.3 ~
##     de.2    (rh_y)    0.230    0.036    6.360    0.000    0.230    0.212
##     se.2      (b1)    0.107    0.043    2.517    0.012    0.107    0.077
##   de.4 ~
##     de.3    (rh_y)    0.230    0.036    6.360    0.000    0.230    0.243
##     se.3      (b1)    0.107    0.043    2.517    0.012    0.107    0.078
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   se.1 ~~
##     se.2              0.068    0.007    9.278    0.000    0.068    0.446
##     se.3              0.052    0.007    7.619    0.000    0.052    0.355
##   se.2 ~~
##     se.3              0.076    0.007   11.312    0.000    0.076    0.573
##   de.1 ~~
##     se.1             -0.046    0.010   -4.441    0.000   -0.046   -0.199
##     se.2             -0.025    0.009   -2.717    0.007   -0.025   -0.120
```

```
##     se.3              -0.026    0.009   -2.956    0.003   -0.026   -0.131
##   eta_y ~~
##     de.1               0.040    0.010    3.821    0.000    0.156    0.278
##     se.1              -0.033    0.008   -4.331    0.000   -0.128   -0.312
##     se.2              -0.046    0.007   -6.535    0.000   -0.179   -0.485
##     se.3              -0.055    0.007   -7.941    0.000   -0.216   -0.603
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .de.2              0.134    0.011   11.973    0.000    0.134    0.591
##    .de.3              0.169    0.013   12.954    0.000    0.169    0.631
##    .de.4              0.141    0.011   13.010    0.000    0.141    0.583
##     de.1              0.316    0.020   16.093    0.000    0.316    1.000
##     se.1              0.168    0.010   16.093    0.000    0.168    1.000
##     se.2              0.137    0.008   16.093    0.000    0.137    1.000
##     se.3              0.129    0.008   16.093    0.000    0.129    1.000
##     eta_y             0.065    0.011    5.779    0.000    1.000    1.000
```

Model fit statistics are $\chi^2(8) = 53.51$, $p < 0.05$; RMSEA = 0.11; CFI = 0.95; SRMR = 0.039, and the standardized parameter estimates are 0.23 ($SE = 0.036$, $p < 0.05$) for the autoregressive effect of depression and 0.11 ($SE = 0.042$, $p < 0.05$) for the lag-one effect of self-esteem.

Example 3

The last example demonstrates a reciprocal dynamic panel. Both self-esteem and depression will have autoregressive effects, and they will both act as lag-one inputs to the other state.

```
dp_mod3 <- "

eta_y =~ 1*de.2 + 1*de.3 + 1*de.4
eta_x =~ 1*se.2 + 1*se.3 + 1*se.4

de.2 ~ rho_y*de.1 + b1*se.1
de.3 ~ rho_y*de.2 + b1*se.2
de.4 ~ rho_y*de.3 + b1*se.3

se.2 ~ rho_x*se.1 + b2*de.1
se.3 ~ rho_x*se.2 + b2*de.2
se.4 ~ rho_x*se.3 + b2*de.3

de.1 ~~ de.1
se.1 ~~ se.1
de.1 ~~ se.1

eta_x ~~ eta_x
eta_y ~~ eta_y
eta_x ~~ eta_y

de.1 ~~ eta_x
de.1 ~~ eta_y
se.1 ~~ eta_x
se.1 ~~ eta_y
"

dp_mod3_fit <- sem(dp_mod3, data = data_f)
```

```r
summary(dp_mod3_fit, fit.measures = T, standardized = T)
```

```
## lavaan 0.6-6 ended normally after 56 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         28
##   Number of equality constraints                     8
##
##   Number of observations                           518
##
## Model Test User Model:
##
##   Test statistic                               103.591
##   Degrees of freedom                                16
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                              1210.972
##   Degrees of freedom                                28
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.926
##   Tucker-Lewis Index (TLI)                       0.870
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -1879.802
##   Loglikelihood unrestricted model (H1)      -1828.006
##
##   Akaike (AIC)                                3799.604
##   Bayesian (BIC)                              3884.603
##   Sample-size adjusted Bayesian (BIC)         3821.119
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.103
##   90 Percent confidence interval - lower         0.084
##   90 Percent confidence interval - upper         0.122
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.053
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
```

```
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   eta_y =~
##     de.2              1.000                               0.250    0.517
##     de.3              1.000                               0.250    0.479
##     de.4              1.000                               0.250    0.516
##   eta_x =~
##     se.2              1.000                               0.199    0.522
##     se.3              1.000                               0.199    0.565
##     se.4              1.000                               0.199    0.567
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   de.2 ~
##     de.1    (rh_y)    0.263    0.037    7.117    0.000    0.263    0.305
##     se.1      (b1)    0.155    0.043    3.560    0.000    0.155    0.131
##   de.3 ~
##     de.2    (rh_y)    0.263    0.037    7.117    0.000    0.263    0.244
##     se.2      (b1)    0.155    0.043    3.560    0.000    0.155    0.113
##   de.4 ~
##     de.3    (rh_y)    0.263    0.037    7.117    0.000    0.263    0.283
##     se.3      (b1)    0.155    0.043    3.560    0.000    0.155    0.112
##   se.2 ~
##     se.1    (rh_x)    0.229    0.042    5.389    0.000    0.229    0.246
##     de.1      (b2)    0.071    0.020    3.520    0.000    0.071    0.104
##   se.3 ~
##     se.2    (rh_x)    0.229    0.042    5.389    0.000    0.229    0.248
##     de.2      (b2)    0.071    0.020    3.520    0.000    0.071    0.097
##   se.4 ~
##     se.3    (rh_x)    0.229    0.042    5.389    0.000    0.229    0.229
##     de.3      (b2)    0.071    0.020    3.520    0.000    0.071    0.106
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   de.1 ~~
##     se.1             -0.046    0.010   -4.441    0.000   -0.046   -0.199
##   eta_y ~~
##     eta_x            -0.049    0.006   -8.383    0.000   -0.985   -0.985
##   eta_x ~~
##     de.1             -0.023    0.007   -3.393    0.001   -0.118   -0.210
##   eta_y ~~
##     de.1              0.036    0.010    3.479    0.001    0.144    0.255
##   eta_x ~~
##     se.1              0.037    0.006    6.074    0.000    0.186    0.454
##   eta_y ~~
##     se.1             -0.036    0.008   -4.767    0.000   -0.143   -0.348
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##     de.1              0.316    0.020   16.093    0.000    0.316    1.000
##     se.1              0.168    0.010   16.093    0.000    0.168    1.000
##     eta_x             0.039    0.007    5.546    0.000    1.000    1.000
##     eta_y             0.063    0.011    5.611    0.000    1.000    1.000
##    .de.2              0.142    0.011   12.458    0.000    0.142    0.605
```

```
##     .de.3              0.177    0.013    13.319    0.000    0.177    0.646
##     .de.4              0.136    0.011    12.894    0.000    0.136    0.575
##     .se.2              0.083    0.007    12.467    0.000    0.083    0.573
##     .se.3              0.063    0.005    11.585    0.000    0.063    0.508
##     .se.4              0.064    0.005    12.616    0.000    0.064    0.519
```

The fit indices for this model are $\chi^2(16) = 103.60$, $p < 0.05$; RMSEA $= 0.10$; CFI $= 0.93$; SRMR $= 0.05$, and the estimated coefficients are, respectively, 0.26 ($SE = 0.037$, $p < 0.05$) for the depression autoregressive effect, 0.23 ($SE = 0.042$, $p < 0.05$) for the self-esteem autoregressive effect, 0.16 ($SE = 0.043$, $p < 0.05$) for the lag-one effect of self-esteem on depression, and 0.071 ($SE = 0.02$, $p < 0.05$) for the lag-one effect of depression on self-esteem.