Barcelona crime longitudinal data quality

Alexandru Cernat

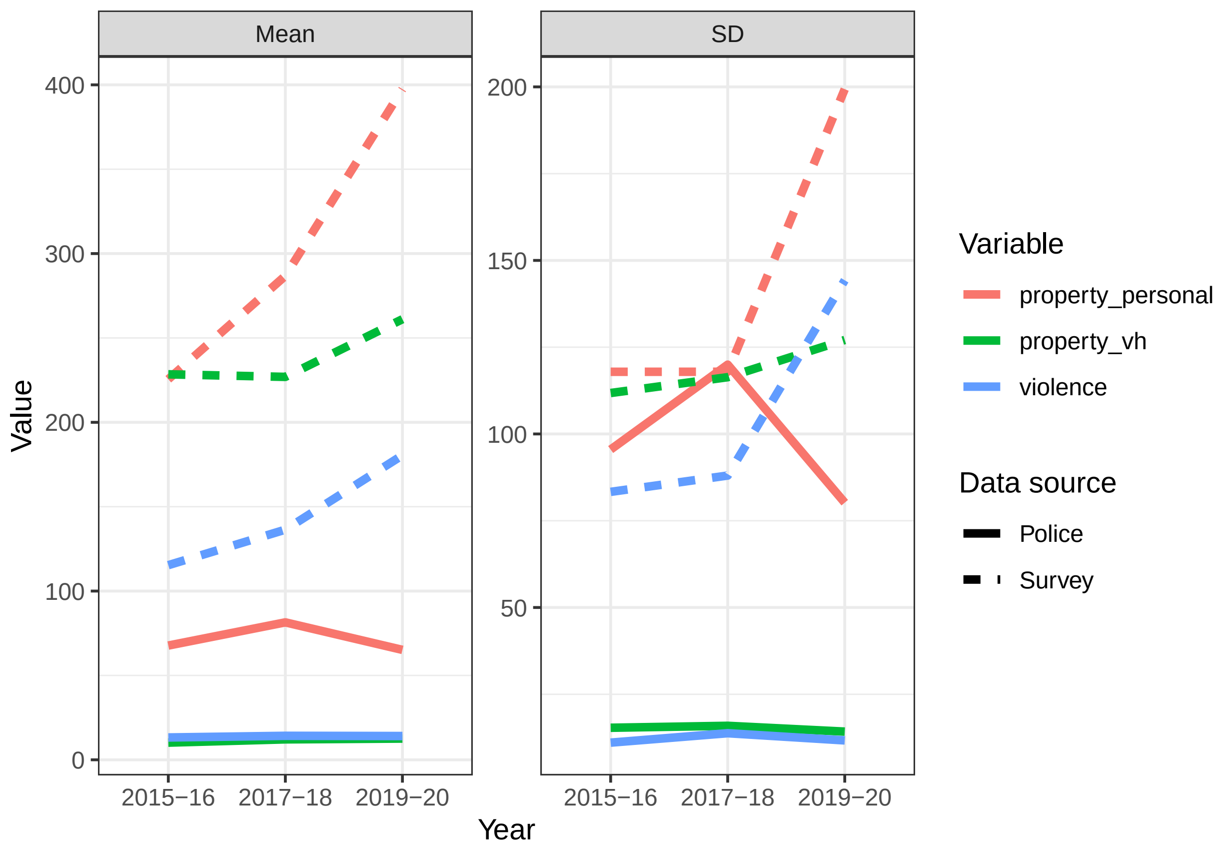
2021-03-09

Here I explore the Barcelona data that has three types of crimes: vehicle, property personal and violence collected in surveys and official data over regions over 6 years. We group years by two in order to avoid having regions with 0s.

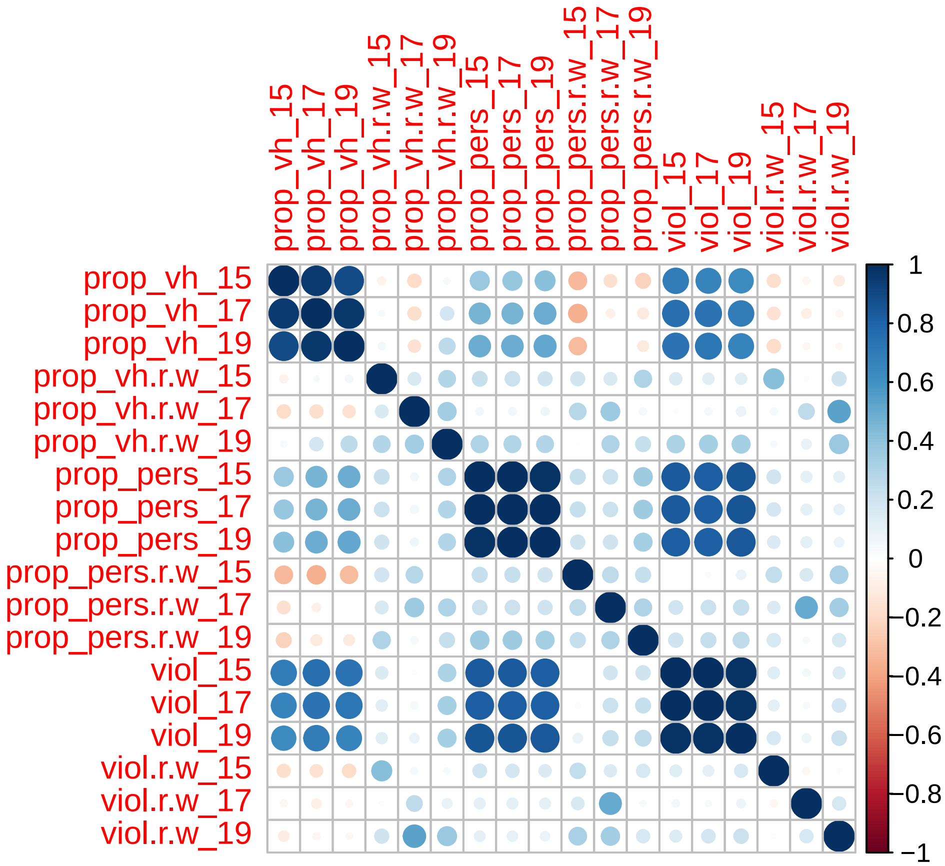
Here I concentrate on the weighted estimates from the survey (ending in ".r.w") and the official data.

Descriptives

First some descriptives. Bellow we observe quite big differences by data source.

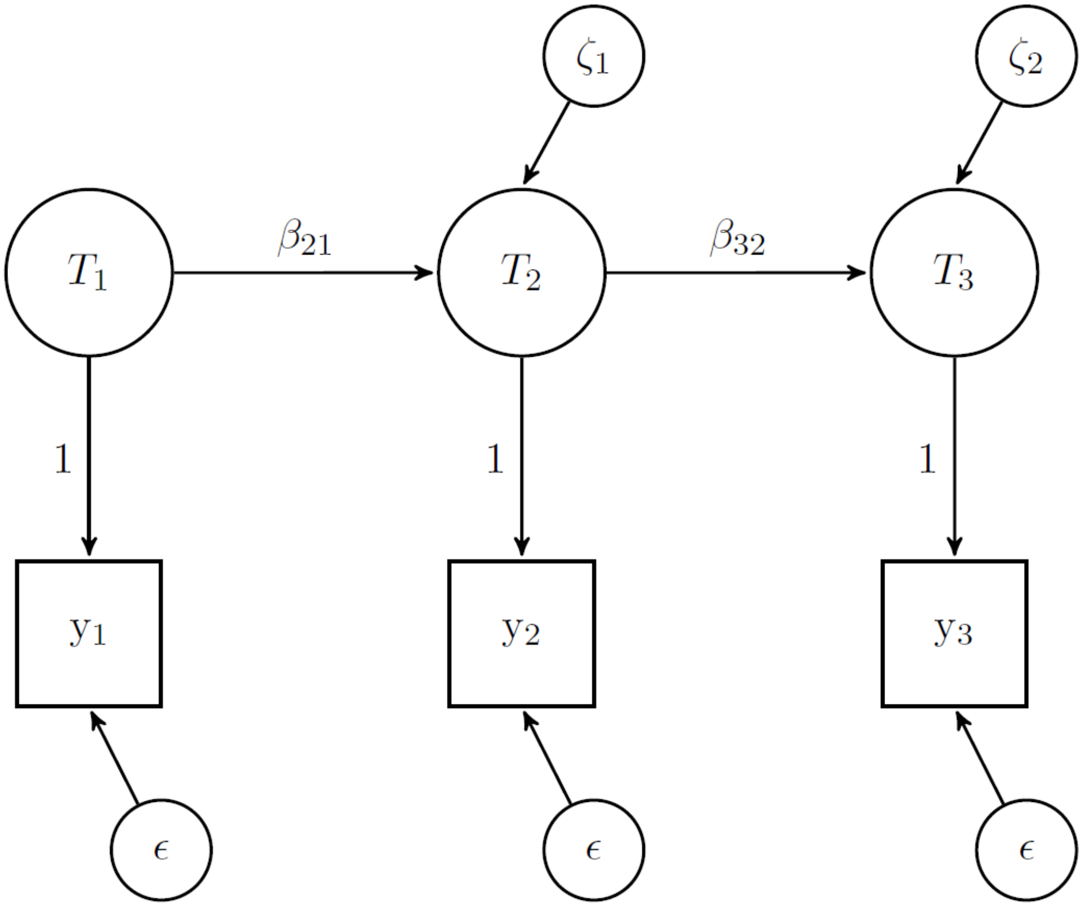


When we look at correlations also we see pretty striking patterns. First of all the consistency within measure is much higher for police data than survey data. Then, the relationship between of measures across data sources is very low. This could be problematic for any modeling.



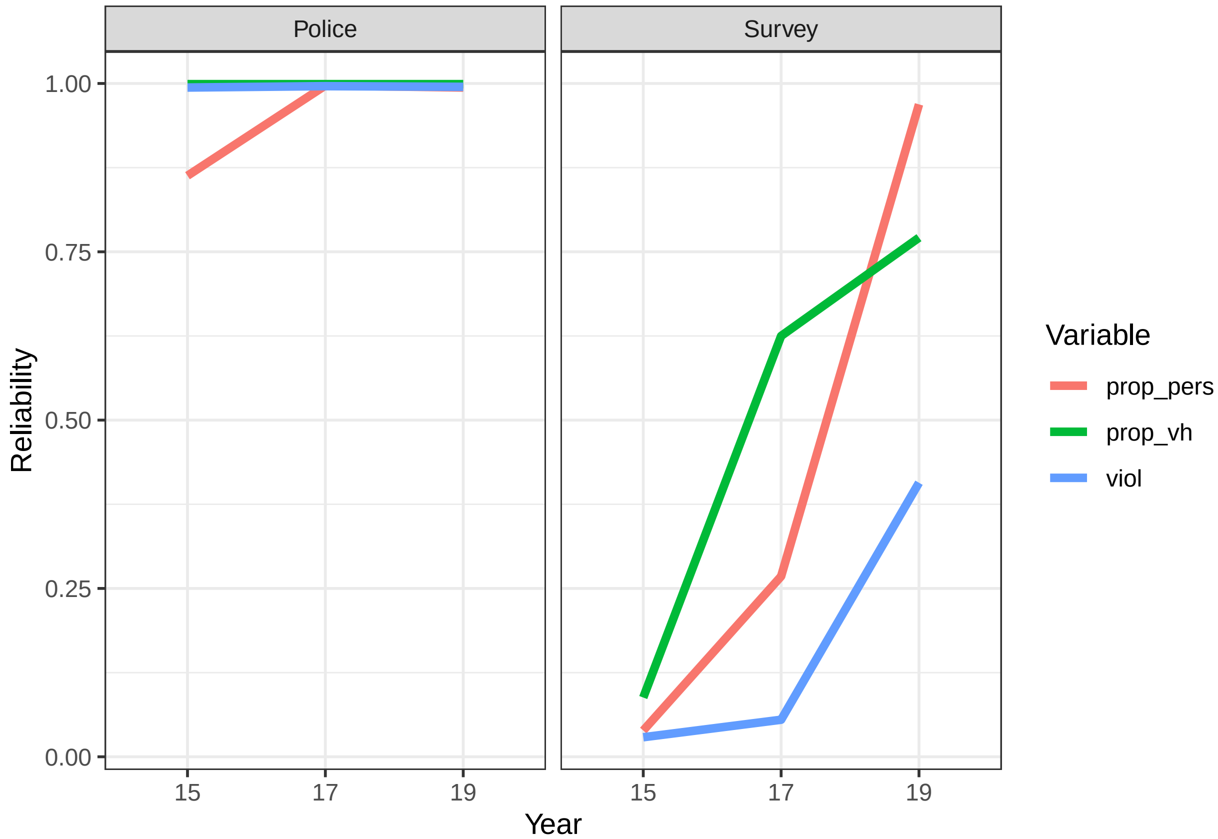
Quasi-simplex

A first way to look at the quality of the data is to use the quasi-simplex model. There assumes an auto-regressive model of true scores and estimates reliability by assuming equal variance of error over time (see book chapter on assumptions of Quasi-simplex for more info).



I estimate the models using `blavaan` which does in the background SEM using Stan for estimation. I tried to use ML but it leads to negative variances (relative common occurence for these models).

Here we plot the reliability by variable, data source and wave as estimated by quasi-simplex. Reliabilities are extremely for the survey data (as expected given the correlation matrices above).

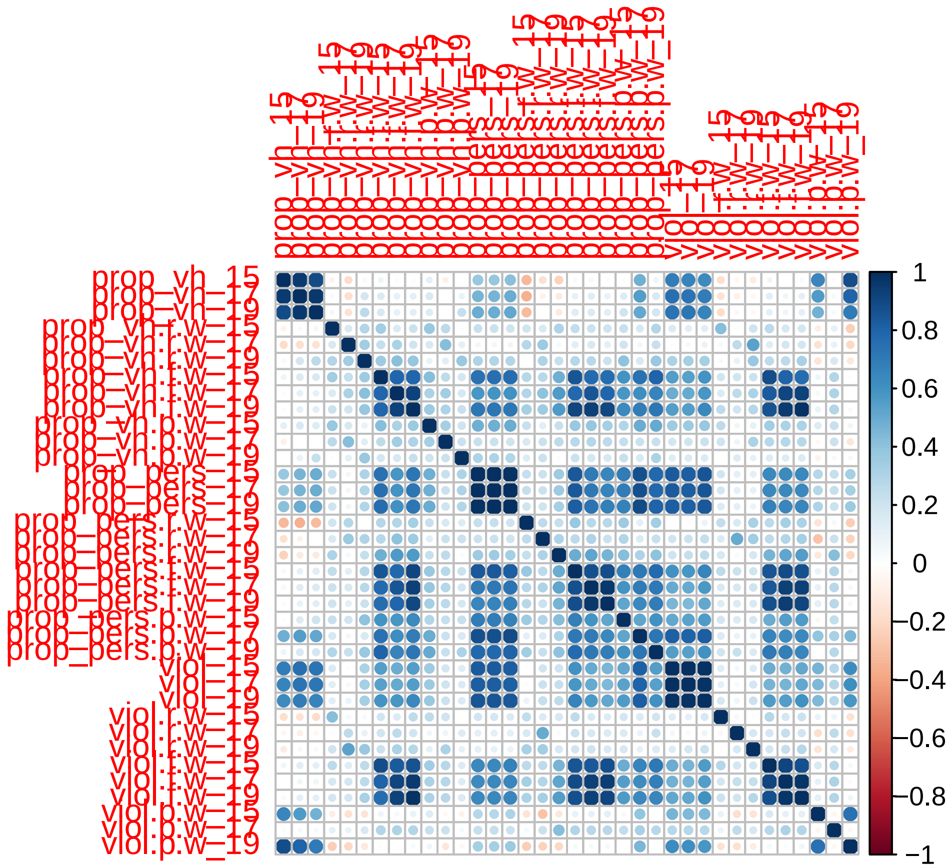


This is also obvious if we average reliability estimates.

group	var	reliability	reliability_source
Police	prop_pers	0.95	0.98
Police	prop_vh	1.00	0.98
Police	viol	1.00	0.98
Survey	prop_pers	0.43	0.36
Survey	prop_vh	0.49	0.36
Survey	viol	0.16	0.36

Given this and the point David made in a recent email let's have a look at survey data about places where crimes happen (".f.w") and looking only at crimes reported to the police (".p.w").

The quasi-simplex does not work for the ".f.w" (determinant of covariance is 0 for some reason). So here I compere the roiginal survye measures with the ".p.w" ones.



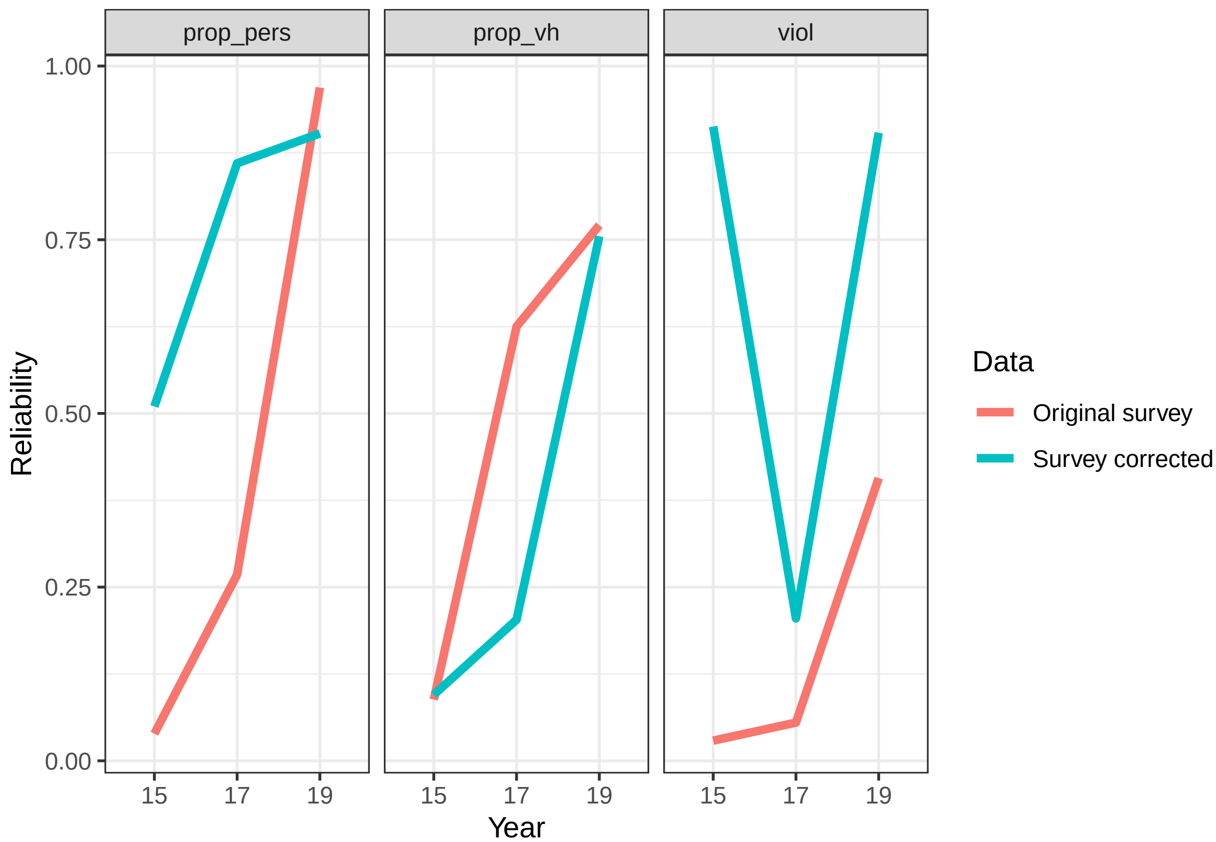

```
## blavaan (0.3-15) results of 1000 samples after 500 adapt/burnin iterations
##
##   Number of observations                            73
##
##   Number of missing patterns                         1
##
##   Statistic                                 MargLogLik         PPP
##   Value                                      -1447.154       0.134
##
## Latent Variables:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##   t1 =~
##     prop_vh.p.w_15    1.000                                2.581    0.095
##   t2 =~
##     prop_vh.p.w_17    1.000                                5.640    0.203
##   t3 =~
##     prop_vh.p.w_19    1.000                               31.275    0.755
##      Rhat    Prior
##
##        NA
##
##        NA
##
##        NA
##
## Regressions:
```

```
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##   t2 ~
##     t1               1.868    6.974   -14.77   14.498    0.855    0.855
##   t3 ~
##     t2               5.521    3.499    1.824   14.835    0.996    0.996
##      Rhat    Prior
##
##     1.002    normal(0,10)
##
##     1.001    normal(0,10)
##
## Intercepts:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##    .prop_vh.p.w_15   38.329    3.171     32.2   44.568   38.329    1.404
##    .prop_vh.p.w_17   39.811    3.347    33.13   46.502   39.811    1.434
##    .prop_vh.p.w_19   45.034    4.774    35.86   54.047   45.034    1.087
##     t1               0.000                               0.000    0.000
##    .t2               0.000                               0.000    0.000
##    .t3               0.000                               0.000    0.000
##      Rhat    Prior
##     1.001    normal(0,32)
##     1.001    normal(0,32)
##     1.000    normal(0,32)
##        NA
##        NA
##        NA
##
## Variances:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##    .prp_v.._15 (a)  739.037   85.542  594.252  923.993  739.037    0.991
##    .prp_v.._17 (a)  739.037   85.542  594.252  923.993  739.037    0.959
##    .prp_v.._19 (a)  739.037   85.542  594.252  923.993  739.037    0.430
##     t1               6.660   12.837    0.009   50.882    1.000    1.000
##    .t2               8.576   15.031    0.004   53.815    0.270    0.270
##    .t3               8.590   20.263    0.002    58.72    0.009    0.009
##      Rhat    Prior
##     1.001 gamma(1,.5)[sd]
##     1.001
##     1.001
##     1.005 gamma(1,.5)[sd]
##     1.000 gamma(1,.5)[sd]
##     1.001 gamma(1,.5)[sd]
##
## blavaan (0.3-15) results of 1000 samples after 500 adapt/burnin iterations
##
##   Number of observations                            73
##
##   Number of missing patterns                         1
##
##   Statistic                                 MargLogLik         PPP
##   Value                                      -1438.666       0.320
##
## Latent Variables:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
```

```
##   t1 =~
##     prp_prs.p.w_15    1.000                              13.600    0.510
##   t2 =~
##     prp_prs.p.w_17    1.000                              38.648    0.860
##   t3 =~
##     prp_prs.p.w_19    1.000                              48.069    0.903
##      Rhat    Prior
##
##        NA
##
##        NA
##
##        NA
##
## Regressions:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##   t2 ~
##     t1                2.835    0.738    1.826    4.657    0.997    0.997
##   t3 ~
##     t2                1.242    0.131    1.009    1.511    0.998    0.998
##      Rhat    Prior
##
##     1.001    normal(0,10)
##
##     1.000    normal(0,10)
##
## Intercepts:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##    .prp_prs.p.w_15   22.112    3.240   15.717   28.397   22.112    0.829
##    .prp_prs.p.w_17   28.890    4.864   19.327   38.243   28.890    0.643
##    .prp_prs.p.w_19   38.148    5.749   26.912   49.095   38.148    0.716
##     t1                0.000                               0.000    0.000
##    .t2                0.000                               0.000    0.000
##    .t3                0.000                               0.000    0.000
##      Rhat    Prior
##     0.999    normal(0,32)
##     1.000    normal(0,32)
##     1.000    normal(0,32)
##        NA
##        NA
##        NA
##
## Variances:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##    .prp_p.._15 (a)  525.707   62.432  418.526  662.474  525.707    0.740
##    .prp_p.._17 (a)  525.707   62.432  418.526  662.474  525.707    0.260
##    .prp_p.._19 (a)  525.707   62.432  418.526  662.474  525.707    0.185
##     t1              184.962   79.163   61.237  362.193    1.000    1.000
##    .t2                7.493   15.628    0.002   53.688    0.005    0.005
##    .t3                8.391   18.078    0.003   55.088    0.004    0.004
##      Rhat    Prior
##     1.001 gamma(1,.5)[sd]
##     1.001
##     1.001
```

```
##     1.000 gamma(1,.5)[sd]
##     1.000 gamma(1,.5)[sd]
##     1.000 gamma(1,.5)[sd]
##
## ** WARNING ** blavaan (0.3-15) did NOT converge after 500 adapt+burnin iterations
## ** WARNING ** Proceed with caution
##
##   Number of observations                            73
##
##   Number of missing patterns                         1
##
##   Statistic                                 MargLogLik         PPP
##   Value                                      -2108.091       0.114
##
## Latent Variables:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##   t1 =~
##     viol.p.w_15       1.000                              31.439    0.913
##   t2 =~
##     viol.p.w_17       1.000                               2.936    0.205
##   t3 =~
##     viol.p.w_19       1.000                              29.720    0.904
##      Rhat    Prior
##
##        NA
##
##        NA
##
##        NA
##
## Regressions:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##   t2 ~
##     t1                0.034    0.104   -0.149    0.155    0.362    0.362
##   t3 ~
##     t2               10.075   27.352  -35.542   38.951    0.995    0.995
##      Rhat    Prior
##
##     4.670    normal(0,10)
##
##     5.857    normal(0,10)
##
## Intercepts:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##    .viol.p.w_15      14.103    3.848     6.64    21.66   14.103    0.410
##    .viol.p.w_17       9.799    1.711    6.564   13.143    9.799    0.684
##    .viol.p.w_19      27.637   12.711    2.704   52.702   27.637    0.841
##     t1                0.000                               0.000    0.000
##    .t2                0.000                               0.000    0.000
##    .t3                0.000                               0.000    0.000
##      Rhat    Prior
##     1.000    normal(0,32)
##     1.010    normal(0,32)
##     1.000    normal(0,32)
```

```
##         NA
##         NA
##         NA
##
## Variances:
##                    Estimate  Post.SD pi.lower pi.upper   Std.lv  Std.all
##    .vil.p.w_15 (a)  196.510   36.077  140.673   281.73  196.510    0.166
##    .vil.p.w_17 (a)  196.510   36.077  140.673   281.73  196.510    0.958
##    .vil.p.w_19 (a)  196.510   36.077  140.673   281.73  196.510    0.182
##     t1              988.384  174.605  683.453 1360.856    1.000    1.000
##    .t2                7.491    3.281    2.943   15.485    0.869    0.869
##    .t3                8.303   19.913    0.006   51.333    0.009    0.009
##      Rhat     Prior
##     1.009 gamma(1,.5)[sd]
##     1.009
##     1.009
##     1.001 gamma(1,.5)[sd]
##     1.011 gamma(1,.5)[sd]
##     1.000 gamma(1,.5)[sd]
```



var	group	reliability	reliability_source
prop_pers	Original survey	0.43	0.36
prop_pers	Survey corrected	0.76	0.59
prop_vh	Original survey	0.49	0.36
prop_vh	Survey corrected	0.35	0.59
viol	Original survey	0.16	0.36

var	group	reliability	reliability_source
viol	Survey corrected	0.67	0.59

Longitudinal variance decomposition