social research UPDATE

- Soft Modelling creates optimal linear relationships among constructs specified by a conceptual model.
- Manifest (observed) variables in soft modelling may be measured at nominal, ordinal, or interval levels; and models may include variables measured at different levels.
- Spurious relationships, common in complex models with many variables, are easily identified with soft modelling procedures.
- Soft Modelling is useful in quasi-experimental designs and observational studies with either very small or very large sample sizes.

Soft Modelling the Predictors of Drug Treatment Use

R. Frank Falk and Peggy Tonkin

This paper explains soft modelling and illustrates the technique with an example from a study of the predictors of alcohol and drug treatment use. The four points above represent some of the strengths of soft modelling for data analysis in non-experimental research.

Soft modelling is a form of structural equation modelling that imposes few assumptions about the level of measurement, the sample size, or the scope of the theory. In relaxing these assumptions, one can argue for neither a confirmatory conclusion nor a probability index of overall model fit. Instead, the researcher is able to see how measurements fit together to form a construct with the use of principal components. Estimates of the multiple R squared are given for the predictive relationship between components. These results are based on a conceptual model that the researcher must specify in advance.

We were interested in the effects of enrolment in a government disability program and the availability of medical insurance on the number of times respondents reported involvement in some type of drug or alcohol treatment program over the course of a two year study period. The respondents were followed over a two-year period to evaluate the impact of a policy change that eliminated the disability classification for drug and alcohol addiction. The subjects in this sample were dropped from the supplemental income program and had the option of reap-

plying under some other disability or losing their cash and medical insurance benefits. Samples were drawn from nine city or county areas in five of the continental United States beginning in December, 1996. Our baseline measurements were taken immediately preceding the official cut-off date of January 1, 1997 and every six months after, for a total of five interview waves. Nine hundred and fifty seven subjects had complete data across all five of the data points.

Our conceptual model is presented in Figure 1. We were interested in predicting treatment use across the four interview waves after the baseline from data on the severity of problems experienced in several life domains: the respondent's perceived need for drug or alcohol treatment; the type (modality) of treatment, if any, in which the respondent was involved at baseline; high risk drug behaviours; and the respondent's medical insurance and disability income status after baseline.

Manifest variables that measure each construct are grouped together. A principal components analysis assigns a weight to each manifest variable and the weighted variables are added together to give a score for each construct. The constructs are then correlated, and a path analysis is computed. This is a conceptual presentation of the mathematics of soft modelling; the actual computation is more complex (Falk & Miller, 1992, Lomoeller, 1989, Wold 1980). The computer program we use for soft-model-

R. Frank Falk, Ph.D. is a professor of sociology at the University of Akron in Akron, Ohio, USA. In 2000 he was an International Visiting Fellow in Social Research Methods at the University of Surrey. His areas of interest include structural equation modeling, computer assisted analysis of qualitative data, and adult emotional development. He is currently developing a coding system for the analysis of biographical data. Dr. Falk has given numerous presentations, conducted workshops, and published a book and several articles on latent variable path analysis with partial least squares.

Peggy Tonkin, Ph. D. is a research associate at the University of Akron Institute for Health and Social Policy. Her areas of research include substance abuse prevention and treatment, and the sociology of education.

Figure 1

Path Model for Treatment Type

N = 957 (unweighted)

Root Mean Square = .043
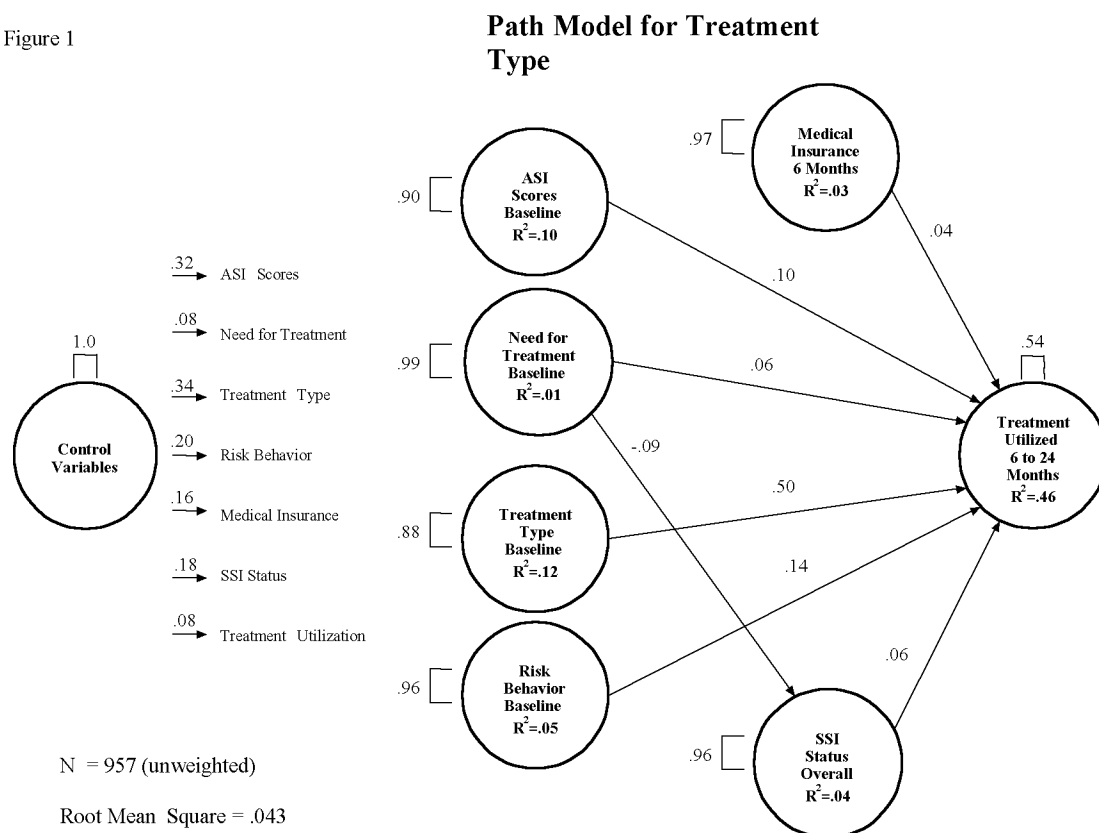
ling is *Latent Variable Path Analysis with Partial Least Squares Estimation* (Lomoeller, 1988) and is available from Dr. Jack McArdle at the University of Virginia (http://krypton.psyc.virginia.edu/jack.mcardle).

Figure 1 illustrates the proposed relationships (14 paths in all) between our eight constructs. The proposed relationships in the model come from the general literature on treatment entry and retention and the expected effects of the policy change. Preliminary analyses revealed significant differences between sample locations on treatment use; therefore, site was entered as a control variable and allowed to predict each of the other seven constructs. The paths from the demographic control construct are shortened to provide ease in reading. Our conceptual model follows the control construct in temporal order: addiction severity index scores, the need for treatment, treatment type at baseline, risk behaviours, medical insurance and disability income status are each allowed

to predict treatment use. The need for treatment is allowed to predict disability income status.

Before we consider each of the hypothesised relationships between these constructs, we should be satisfied with the reliability of our measurements. Sixteen variables were measured as indicants of the eight constructs. The indicants included nominal, interval, and ratio levels of measurement. They included measurements from well-established instruments such as the Addiction Severity Index Scores (see McLelland, Luborsky, Cacciola, and Griffith 1984). For a complete description of the study, see Tonkin, Swartz, and Kappagoda, forthcoming.

In our study, most of the measurements proved to be reliable (see Table 1). Principal component loadings above 0.55 are generally desirable. However, several observed variables with very small loadings were left in the model because our goal was to examine the comparative effects of each of the predictors. Low loadings indicated that the vari-

ables are not contributing substantially to the construct and thus are not good measures of the proposed construct. For example, the loading of 0.05 for involvement in a jail or prison treatment program at baseline indicates this variable is not a substantial contributor to the treatment type construct. Conversely, participation in methadone maintenance at baseline is the defining measure for the baseline treatment modality construct with a loading of 0.98.

Researchers, in general, should expect principal component loadings to be above 0.55 before examining the relationships between the theoretical constructs. Generally, three or more manifest variables are preferred as multiple indicants of a construct. When there is only one indicant for each construct, the programme becomes a standard path analysis procedure.

As we stated in the beginning, soft modelling creates optimal linear relationships among constructs specified by a conceptual model. It does this by an iterative process

Table 1.

Component Loadings and Weights for Measurement Model				Path Coefficients for Inner Model		
Components and Manifest Variables	**Loading**	**Residual**	**Weight**	**Predictor Constructs**	**Predicted Constructs**	**Path**
Demographic Controls				Demographic Controls	ASI Scores Base	0.32
Site	0.98	0.05			Need for Treatment Base	0.08
ASI Scores Basline					Treatment Type Base	0.34
Alcohol	0.19	0.96			Risk Behavior Base	0.20
Drug	0.70	0.51			Medical Insurance	0.16
Social	0.62	0.61			Disability Status	0.18
Psychological	0.64	0.59			Treatment Utilization	0.08
Legal	0.63	0.60		ASI Scores Baseline	Treatment Utilization	0.10
Need for Treatment Baseline				Need for Treatment Base	Disability Status	-0.09
Need for Treatment		0.00	1.00		Treatment Utilization	0.06
Treatment Type Baseline				Treatment Type Base	Treatment Utilization	0.50
Jail or Prison	0.05	1.00		Risk Behavior Base	Treatment Utilization	0.14
Methadone Maintenance	0.98	0.05		Medical Insurance	Treatment Utilization	0.04
Residential	0.14	0.98		SSI Status Overall	Treatment Utilization	0.06
Outpatient	0.25	0.94		**R-square**	ASI Scores Base =	0.10
Risk Behavior Baseline					Need for Treatment Base =	0.01
Inject Drugs	0.91	0.17			Treatment Type Base =	0.12
Share Needle	0.59	0.65			Risk Behavior Base =	0.05
Medical Insurance 6 months					Medical Insurance =	0.04
Insurance 6 months		0.00	1.00		Disability Status =	0.04
Treatment Utilization					Treatment Utilization =	0.46
Treatment Ratio		0.00	1.00		RMS Cov (E, U) =	0.04
Disability Status						
Benefits Ratio		0.00	1.00			

in which each predictor construct is made to be as good a predictor as it can be, while the predicted construct is made to be the best predicted construct it can be. To achieve this optimisation, the principle component loadings on the manifest variables are adjusted. This is accomplished in two stages by using the least squares minimisation procedure within each group of manifest variables and their construct and then between the constructs.

This optimisation procedure provides the researcher with three important pieces of information: the matrix of correlations between the constructs, the path coefficients between the constructs as specified in the model, and the multiple R square value for each predicted construct. With this information, we are able to evaluate our original conceptual model.

The correlation matrix allows us to assess if the relationships are going in the predicted direction. In our model, we assumed that all the proposed relationships would be in a positive direction. The path coefficients give the direct effects and direction of relationship between the predictor construct and the predicted construct. These are the equivalent of standardised regression coefficients or beta weights, and so may be directly compare with one another. They differ from the correlation coefficients because they take into account the correlations between the predictor constructs.

The multiple R square is the measure of the total amount of variance in the predicted construct that is accounted for by the predictor constructs. These provide an estimate of the effect of all the predictors taken together on the predicted construct. Traditional statistical tests of significance can be applied to the multiple R square.

With these three pieces of information, we can evaluate how well our model predicts the relationships in our original conceptualisations. The results noted in Table 1 provide the information necessary to evaluate our conceptual model. The first piece of

social research **UPDATE**

information to evaluate is whether the multiple R square is high enough to believe that our predicted constructs are in fact being predicted by their predictor constructs. In this case, only three of the seven constructs had statistically significant multiple R squares and our model predicted 10, 12, and 46 percent of the variance in the predicted constructs.

Next we examine the contribution of each predictor construct to the multiple R squared, the contribution of each is referred to as the percent of variance contributed. It is approximated by multiplying the path coefficient by the correlation coefficient. When this value is less than 0.015 or less than 1.5 %, the predictor construct is not making an important contribution to the variance in the predicted construct and the path should be eliminated (Falk & Miller, 1992). In this example, the paths from medical insurance and disability income status should be dropped from the model because both the correlation values and the path values are small. The path from need for treatment to overall disability income status should also be dropped. However, they are left in the model because they are the constructs of most interest in this study. The path from need for treatment to treatment use is small, but the correlation coefficient between the two constructs is large. The product of the two is greater than 0.015; therefore, the path is left in the model.

Once all unimportant paths have been eliminated, the remaining path coefficients should be compared to their respective correlation coefficients for any suppressor effects. If the signs of the two coefficients do not agree, a suppressor effect is present. Models with many correlated variables are likely to have suppressors. However, these effects may not be true suppressor effects, but mathematical artefacts. The easiest way to determine which type of effect is present is to compare the multiple R squares before and after eliminating the suspect path. If the multiple R square does not change significantly after removing the path, the suppressor is an artefact and the path should be removed.

Worth noting is that much of the variance in our model is accounted for by differences among the nine sample sites. This is apparent from the relatively large path coefficients emanating from the demographic control construct. The construct for the Addiction Severity Index scores had a positive impact on treatment use as was hypothesised. The size of the impact was small but its relative size makes it one of the best predictors of treatment use in the model. By far the best predictor of treatment use is the type of treatment in which the subject participated at baseline. Since methadone maintenance defines this construct, the large path coefficient indicates that respondents who were in methadone maintenance at baseline were very likely to be in treatment at later points in the study. The need for treatment at baseline, medical insurance, and disability income status are the weakest predictors of treatment use. Only the path from need for treatment to disability income status produced a sign contrary to our predicted direction. Since the sign of the path coefficient matches the direction of the correlation coefficient, the observed relationship is in fact a negative one and not the result of a suppressor effect. The negative path indicates subjects who felt they needed treatment at baseline were less likely to be receiving disability income benefits after baseline.

A snapshot of our findings reveals that a statistically significant amount of variation was accounted for in three of our predicted constructs. Forty-six percent of the variation in treatment use after baseline is accounted for by the seven predictor constructs, with methadone maintenance accounting for more than half of the 46 percent. Less than five percent of the variance in each of the two main predictors of treatment use (medical insurance and disability income status) is accounted for by the four preceding predictors. The demographic control construct (sample site) is the best predictor of both these constructs. What the findings suggest is that controlling for site differences, treatment use at an earlier point in time is the best predictor of treatment at a later point in time. Specifically, involvement in a methadone maintenance program at baseline is the best predictor of treatment use over the course of the study.

References

Falk R. F. & Miller, N. B. (1992) *A Primer for Soft Modeling.* Akron, Ohio: The University of Akron Press.

Lomoeller, J. B. (1988) 'The PLS program system: Latent variables path analysis with partial least squares estimation.' *Multivariate Behavioral Research,* 23:125-127.

Lomoeller, J. B. (1989) *Latent Variable Path Modeling with Partial Least Square.* Heidelberg: Physica-Verlag.

McLellan, A.T., Luborsky, L., Cacciola, J., and Griffith, J. 1984. "New Data from the Addiction Severity Index: Reliability and Validity in three Centers." *Journal of Nervous and Mental Disorders* 173:412-423.

Tonkin, P., Swartz, J., and Kappagoda, S. (2001). Methodology of the Supplemental Security Income Study. *Contemporary Drug Problems* (forthcoming).

Wold, H.(1980) 'Soft modelling: Intermediate between traditional model building and data analysis.' *Mathematical Statistics vol.* 6:333-46.