Using R for Structural Equation Model:  
A transaction cost measurement

Pairach Piboonrugnroj and Stephen M. Disney

Logistics Systems Dynamics Group,  
Cardiff University

August 16th, 2011
Supply Chain Management Matters

- Businesses are trying to reduce their *transaction costs* to improve their business performance and relationships.

> Supplier $\rightarrow$ Manufacturer $\rightarrow$ Retailer $\rightarrow$ Customers

- However, a measurement of transaction cost is still limited.
  - Only in manufacturing context NOT in service sector.
  - Aspects of governance problem and opportunity costs are missing.
Motivation

Software Matters

- SEM can be fitted by various software but costly
  - Lisrel8.8 = USD 396
  - IBM SPSS Amos = USD 695
  - Mplus6.11 = USD 195-350 (student price)
  - StataSE12 = USD 895
- Can I use R to run SEM?
  - Identical output to those of other commercial software?
  - Any difficulties or problems from the non-technical user aspect?
Aims

1. To develop the measurement of *transaction costs*.
2. To empirically test such a measurement.
3. To compare the use of R packages for SEM with other software *via a non-technical aspect*, an outside R!
Analysis Method

- Structural Equation Model (Bollen, 1989) = Factor Analysis + Regression (Path Analysis)
- Proprietary software i.e.,

![Amos 18](https://www.spss.com)
Analysis Method

- Packages in R (R Development Core Team, 2011)
  - sem (John Fox, 2006)
  - OpenMx 1.0.7
  - lavaan 0.4-9 (Yves Rosseel, 2011)
Data

- Questionnaire survey
- Tourism industry of Thailand
- 53 usable responses
Figure: Using Lisrel to fit the model
AMOS: Model Specification
Figure: Using Amos to fit the model
Figure: Using Amos to fit the model
'sem' Package: Code

```r
## Introduction to Structural Equation Modelling ##
## R useR! conference August 2011 ##

install.packages("sem")
library(sem)

# 1. Load data
data.tc <- read.csv("/Users/pairachpiboongrungrjoj/Documents/PhD/Analysis/R/useR/cleandata.csv")

# Input covariance matrix
data.tc.1 <- cor(data.tc)

# hotel.csv <- csv(data.sem)
hotel.cor <- cor(hoteldata)

# path parameter start-value
model.TC.1 <- specify(model())

TC <- TC1, gamma1, NA # measurement item
TC <- TC2, gamma2, NA
TC <- TC3, gamma3, NA
TC <- TC4, gamma4, NA
TC <- TC5, gamma5, NA
TC <- TC6, gamma6, NA
TC <- TC7, gamma7, NA
TC <- TC8, gamma8, NA
TC <- TC9, gamma9, NA
TC <- TC10, gamma10, NA
TC <- TC11, gamma11, NA
TC <- TC12, gamma12, NA
TC <- TC13, gamma13, NA

TC1 <- TC1, e1, NA # measurement error
TC2 <- TC2, e2, NA
TC3 <- TC3, e3, NA
TC4 <- TC4, e4, NA
TC5 <- TC5, e5, NA
TC6 <- TC6, e6, NA
TC7 <- TC7, e7, NA
TC8 <- TC8, e8, NA
TC9 <- TC9, e9, NA
TC10 <- TC10, e10, NA
TC11 <- TC11, e11, NA
TC12 <- TC12, e12, NA
TC13 <- TC13, e13, NA
TC <- TC, NA, 1

model.TC.1

sem.TC.1 <- sem(model.TC.1, data.tc.1, S)

summary(sem.TC.1)

std.coef(sem.TC.1)
```

Pairach Piboonrungroj and Stephen M. Disney
Logistics Systems Dynamics Group, Cardiff University

'sem' Package: Output

```r
# print result (fit indices, parameters, hypothesis tests)
> summary(sem.TC.1)

Model Chi-square = 44.54  DF = 14 Pr(>ChiSq) = 4.8429e-05
Chi-square (null model) = 243.45  DF = 21
Goodness-of-fit index = 0.81983
Adjusted goodness-of-fit index = 0.63967
RMSEA index = 0.20482  90% CI: (NA, NA)
Bentler-Bonett NFI = 0.81205
Tucker-Lewis NNFI = 0.79407
Bentler CFI = 0.86271
SRMR = 0.087909
BIC = -11.164

Normalized Residuals
 Min. 1st Qu.  Median  3rd Qu.  Max.    
-0.8220  -0.2840   0.0000   0.0125   0.3350   0.8620

Parameter Estimates

| Parameter   | Estimate | Std Error | t value | Pr(>|z|) |
|-------------|----------|-----------|---------|----------|
| gamma1      | 0.74793  | 0.122606  | 6.1033  | 1.0587e-09 |
| gamma2      | 0.81548  | 0.116348  | 7.1802  | 6.9611e-13 |
| gamma3      | 0.818532 | 0.118532  | 6.7746  | 1.2472e-11 |
| gamma6      | 0.86297  | 0.33085e-14 | 7.5856 | 3.3085e-14 |
| gamma7      | 0.62515  | 0.129333  | 4.8337  | 1.3492e-06 |
| gamma11     | 0.71731  | 0.125624  | 5.7374  | 9.6159e-09 |
| gamma13     | 0.69637  | 0.126639  | 5.3249  | 3.2965e-08 |
| e1          | 0.44088  | 0.108493  | 4.0844  | 1.1633e-05 |
| e2          | 0.38211  | 0.109599  | 3.5703  | 1.4628e-04 |
| e3          | 0.35577  | 0.116086  | 3.0508  | 3.9470e-05 |
| e6          | 0.25528  | 0.071294  | 3.5806  | 3.4218e-04 |
| e7          | 0.60918  | 0.127416  | 4.7811  | 1.7438e-06 |
| e11         | 0.48547  | 0.108776  | 4.4630  | 8.0804e-06 |
| e13         | 0.51847  | 0.119840  | 4.4801  | 7.4592e-06 |

Iterations = 14
> # standardised coefficients (loadings)
> std.coef(sem.TC.1)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Std. Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma1</td>
<td>0.7473390</td>
</tr>
<tr>
<td>gamma2</td>
<td>0.8354804</td>
</tr>
</tbody>
</table>
```

Pairach Piboonrugroj and Stephen M. Disney
Logistics Systems Dynamics Group, Cardiff University

'OpenMx' package

Figure: OpenMx using Path Specification
'OpenMx' package

\[ R = ALA' + U \]

Figure: OpenMx using Matrix Specification
'OpenMx' package: Code

```r
require(OpenMx)
hoteldata <- read.csv("/Users/pairachpiboonrungroj/Documents/PhD/Analysis/R/useR/cleandataCFA.csv")
manifests <- names(hoteldata)
lats <- c("TC")
CFA.TC <- mxModel("One Factor",
type="RAM",
manifestVars = manifests,
latentVars = lats,
mxPath(from=lats, to=manifests),
mxPath(from=manifests, arrows=2),
mxPath(from=lats, arrows=2, 
free=FALSE, values=1.0),
mxData(cov(hoteldata), type="cov", 
numObs=53))
FactorFit.TC <- mxRun(CFA.TC)
summary(FactorFit.TC)
```

**Figure:** OpenMx using Path Specification
'OpenMx' package: Output1

```r
> require(OpenMx)
> hoteldata <- read.csv("/Users/pairachpiboonrungroj/Documents/PhD/Analysis/R/userR/cleandataCFA.csv")
> manifests <- names(hoteldata)
> latents <- c("TC")
> CFA.TC <- mxModel("One Factor",
+ type="RAM",
+ manifestVars = manifests,
+ latentVars = latents,
+ mxPath(from=latents, to=manifests),
+ mxPath(from=manifests, arrows=2),
+ mxPath(from=latents, arrows=2,
+ free=FALSE, values=1.0),
+ mData(cov(hoteldata), type="cov",
+ numObs=53))
> FactorFit.TC <- mxRun(CFA.TC)
> summary(FactorFit.TC)
Running One Factor

data:
$'One Factor.data'
$'One Factor.data'$cov

<table>
<thead>
<tr>
<th></th>
<th>TC1</th>
<th>TC2</th>
<th>TC3</th>
<th>TC6</th>
<th>TC7</th>
<th>TC11</th>
<th>TC13</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC1</td>
<td>2.945573</td>
<td>1.937228</td>
<td>1.345428</td>
<td>1.851597</td>
<td>1.448113</td>
<td>1.219158</td>
<td>1.284107</td>
</tr>
<tr>
<td>TC2</td>
<td>1.937228</td>
<td>2.311321</td>
<td>1.695573</td>
<td>1.754354</td>
<td>1.358491</td>
<td>1.186865</td>
<td>1.192671</td>
</tr>
<tr>
<td>TC3</td>
<td>1.345428</td>
<td>1.695573</td>
<td>2.354862</td>
<td>1.626996</td>
<td>1.276488</td>
<td>1.658563</td>
<td>1.379173</td>
</tr>
<tr>
<td>TC6</td>
<td>1.851597</td>
<td>1.754354</td>
<td>1.626996</td>
<td>2.566763</td>
<td>1.349149</td>
<td>1.416564</td>
<td>1.973512</td>
</tr>
<tr>
<td>TC7</td>
<td>1.448113</td>
<td>1.358491</td>
<td>1.276488</td>
<td>1.349419</td>
<td>2.980406</td>
<td>1.404572</td>
<td>1.485849</td>
</tr>
<tr>
<td>TC11</td>
<td>1.219158</td>
<td>1.186865</td>
<td>1.658563</td>
<td>1.416564</td>
<td>1.404572</td>
<td>2.383164</td>
<td>1.701379</td>
</tr>
<tr>
<td>TC13</td>
<td>1.284107</td>
<td>1.192671</td>
<td>1.379173</td>
<td>1.973512</td>
<td>1.485849</td>
<td>1.701379</td>
<td>3.05806</td>
</tr>
</tbody>
</table>
```
'OpenMx' package: Output2

free parameters:

<table>
<thead>
<tr>
<th>name</th>
<th>matrix</th>
<th>row</th>
<th>col</th>
<th>Estimate</th>
<th>Std.Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>TC1</td>
<td>TC1</td>
<td>1.2836649</td>
<td>0.2103818</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>TC2</td>
<td>TC1</td>
<td>1.2708069</td>
<td>0.1765347</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>TC3</td>
<td>TC1</td>
<td>1.2322648</td>
<td>0.1813500</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>TC6</td>
<td>TC1</td>
<td>1.3825805</td>
<td>0.1822115</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>TC7</td>
<td>TC1</td>
<td>1.0792584</td>
<td>0.2232472</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>TC11</td>
<td>TC1</td>
<td>1.1073405</td>
<td>0.1929606</td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>TC13</td>
<td>TC1</td>
<td>1.2130293</td>
<td>0.2195215</td>
</tr>
<tr>
<td>8</td>
<td>S</td>
<td>TC1</td>
<td>TC1</td>
<td>1.2978175</td>
<td>0.2958840</td>
</tr>
<tr>
<td>9</td>
<td>S</td>
<td>TC2</td>
<td>TC2</td>
<td>0.6082638</td>
<td>0.1837838</td>
</tr>
<tr>
<td>10</td>
<td>S</td>
<td>TC3</td>
<td>TC3</td>
<td>0.8363844</td>
<td>0.2033757</td>
</tr>
<tr>
<td>11</td>
<td>S</td>
<td>TC6</td>
<td>TC6</td>
<td>0.6552330</td>
<td>0.1828873</td>
</tr>
<tr>
<td>12</td>
<td>S</td>
<td>TC7</td>
<td>TC7</td>
<td>1.8156069</td>
<td>0.3796282</td>
</tr>
<tr>
<td>13</td>
<td>S</td>
<td>TC11</td>
<td>TC11</td>
<td>1.1569600</td>
<td>0.2591265</td>
</tr>
<tr>
<td>14</td>
<td>S</td>
<td>TC13</td>
<td>TC13</td>
<td>1.5343629</td>
<td>0.3423551</td>
</tr>
</tbody>
</table>

observed statistics: 28
estimated parameters: 14
degrees of freedom: 14
-2 log likelihood: 517.56
saturated -2 log likelihood: 473.0199
number of observations: 53
chi-square: 44.54013
p: 4.842873e-05
AIC (Mx): 16.54013
BIC (Mx): -5.521976
adjusted BIC: 
RMSEA: 0.2028773
timestamp: 2011-08-16 11:19:23
frontend time: 1.606380 secs
backend time: 0.01884508 secs
independent submodels time: 0.0001170635 secs
wall clock time: 1.715351 secs
cpu time: 1.715351 secs
openmx version number: 1.0.7-1706
lavaan Package

Figure: SEM description in lavaan website
lavaan package: model syntax

```r
# An analysis for useR conference 2011

# 1. Load data
hoteldata <- read.csv("/Users/pairachpiboonrungrong/Docs/Documents/PhD/Analysis/R/useR/cleandata.csv")

# 2. Install Package
install.packages("lavaan")

# 3. Load Package
library(lavaan)

# 4. Structural Model

TC.Model <- '                  
# latent variable definitions
  cost =~ TC1 + TC2 + TC3 + TC6 + TC7 + TC11 + TC13

fitTC <- sem(TC.Model, data = hoteldata)
summary(fitTC, standardized = TRUE, fit.measures=True)
```

**Figure: lavaan model syntax**
lavaan package: output1

Lavaan (0.4-9) converged normally after 24 iterations

Number of observations: 53
Estimator: ML
Minimum Function Chi-square: 45.307
Degrees of freedom: 14
P-value: 0.000

Chi-square test baseline model:
Minimum Function Chi-square: 248.133
Degrees of freedom: 21
P-value: 0.000

Full model versus baseline model:
Comparative Fit Index (CFI): 0.862
Tucker-Lewis Index (TLI): 0.793

Loglikelihood and Information Criteria:
Loglikelihood user model (H0): -601.149
Loglikelihood unrestricted model (H1): -576.451
Number of free parameters: 14
Akaike (AIC): 1230.299
Bayesian (BIC): 1257.883
Sample-size adjusted Bayesian (BIC): 1213.909

Root Mean Square Error of Approximation:
RMSEA: 0.06
90 Percent Confidence Interval: 0.141, 0.274
P-value RMSEA <= 0.05: 0.000

Standardized Root Mean Square Residual:
SRMR: 0.068

Figure: lavaan model syntax
Parameter estimates:

<table>
<thead>
<tr>
<th>Latent variables:</th>
<th>Information Standard Errors</th>
<th>Expected Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std.err</td>
</tr>
<tr>
<td>cost = -</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>TC1</td>
<td>0.983</td>
<td>0.160</td>
</tr>
<tr>
<td>TC2</td>
<td>0.960</td>
<td>0.162</td>
</tr>
<tr>
<td>TC3</td>
<td>1.077</td>
<td>0.168</td>
</tr>
<tr>
<td>TC6</td>
<td>0.841</td>
<td>0.186</td>
</tr>
<tr>
<td>TC7</td>
<td>0.863</td>
<td>0.165</td>
</tr>
<tr>
<td>TC11</td>
<td>0.945</td>
<td>0.185</td>
</tr>
</tbody>
</table>

Variance:

| TC1    | 1.273 | 0.280 | 4.543 | 0.000 | 1.273 | 0.441 |
| TC2    | 0.685 | 0.170 | 4.037 | 0.000 | 0.685 | 0.382 |
| TC3    | 0.821 | 0.192 | 4.278 | 0.000 | 0.821 | 0.355 |
| TC6    | 0.843 | 0.172 | 3.745 | 0.000 | 0.843 | 0.255 |
| TC7    | 1.781 | 0.308 | 4.945 | 0.000 | 1.781 | 0.609 |
| TC11   | 1.135 | 0.244 | 4.644 | 0.000 | 1.135 | 0.485 |
| TC13   | 1.505 | 0.321 | 4.693 | 0.000 | 1.505 | 0.510 |
| cost   | 1.617 | 0.521 | 3.104 | 0.002 | 1.000 | 1.000 |

Figure: lavaan model syntax
Result Comparison

One latent variable: TC
Seven manifest variables: TC1 - TC7

Table: SEM outputs from two proprietary software and three R packages

<table>
<thead>
<tr>
<th>Fit Indices</th>
<th>LISREL</th>
<th>AMOS</th>
<th>sem</th>
<th>OpenMx</th>
<th>lavaan</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>44.54</td>
<td>44.540</td>
<td>44.54</td>
<td>44.54013</td>
<td>45.397</td>
</tr>
<tr>
<td>(df)</td>
<td>(14)</td>
<td>(14)</td>
<td>(14)</td>
<td>(14)</td>
<td>(14)</td>
</tr>
<tr>
<td>CFI</td>
<td>0.86</td>
<td>0.863</td>
<td>0.86271</td>
<td>NA</td>
<td>0.862</td>
</tr>
<tr>
<td>GFI</td>
<td>0.82</td>
<td>0.82</td>
<td>0.81983</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>NFI</td>
<td>0.82</td>
<td>0.817</td>
<td>0.81705</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>NNFI</td>
<td>0.79</td>
<td>NA</td>
<td>0.79407</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>TLI</td>
<td>NA</td>
<td>0.794</td>
<td>NA</td>
<td>NA</td>
<td>0.793</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.20</td>
<td>0.205</td>
<td>0.20482</td>
<td>0.2028773</td>
<td>0.206</td>
</tr>
<tr>
<td>SRMR</td>
<td>0.07</td>
<td>NA</td>
<td>0.067909</td>
<td>NA</td>
<td>0.068</td>
</tr>
</tbody>
</table>
Conclusion 1

- Considering opportunity cost and governance problem, this study proposed and tested a new measurement for transaction cost but yet well fitted one (very small sample size).

- R packages can be used to fit SEM
  - Identical or almost to those of commercial software.
  - lavaan is probably the most useR-friendly package in R.
  - OpenMx offers alternative approach (Matrix specification) and powerful.

- Challenges of R packages for SEM
Conclusion 2

- Challenges of R packages for SEM
  - More user-friendly?
    - by just *Drawing* like AMOS? Proposed in OpenMx
    - by just *Clicking* as a plugin in Rcmdr (John Fox)
  - Publishing SEM research using R package(s)?
  - SEM is available in Stata12 (either drawing or coding)

- Comparing with more advance SEM model e.g., multiple group, multilevel or growth curve model
Acknowledgements

"The author is grateful to the Royal Thai Government through the Commission on Higher Education for financial support of this study."

Pairach Piboonrugnroj and Stephen M. Disney
Logistics Systems Dynamics Group, Cardiff University
Thank you very much

Any suggestion?
Result

- Cronbach's alpha was greater than 0.7
- Chi-square = 40.244, (d.f. = 37, \( p = 0.329 \))
  CFI = 0.989, TLI = 0.983 and RMSEA = 0.041.
- Coefficients: uncertainty 0.458 (\( p = 0.031 \)) and asset specificity 0.622 (\( p < 0.001 \))
## Construct measures with reliability and factor loadings

<table>
<thead>
<tr>
<th>Measurement Items</th>
<th>Factor Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asset specificity</strong> ((\alpha = .718))</td>
<td></td>
</tr>
<tr>
<td>In building the relationship with my firm, this supplier . . .</td>
<td></td>
</tr>
<tr>
<td>. . . has an operating process that has been tailored.</td>
<td>1.000</td>
</tr>
<tr>
<td>. . . has made specific investments in resources.</td>
<td>0.862</td>
</tr>
<tr>
<td><strong>Uncertainty</strong> ((\alpha = .702))</td>
<td></td>
</tr>
<tr>
<td>My firm can accurately predict the performance</td>
<td></td>
</tr>
<tr>
<td>of this supplier in our next transaction.</td>
<td>1.000</td>
</tr>
<tr>
<td>My firm knows that this supplier will adapt quickly,</td>
<td></td>
</tr>
<tr>
<td>should we have change our specifications at short notice.</td>
<td>0.693</td>
</tr>
<tr>
<td><strong>Transaction cost</strong> ((\alpha = .880))</td>
<td></td>
</tr>
<tr>
<td>It is very complicated and difficult to write a contract.</td>
<td>1.000</td>
</tr>
<tr>
<td>It took a significant effort to gather the critical information.</td>
<td>0.916</td>
</tr>
<tr>
<td>It is very difficult to monitor the performance of this supplier.</td>
<td>0.926</td>
</tr>
<tr>
<td>It takes a lot of effort to solve problems in our relationship.</td>
<td>0.890</td>
</tr>
<tr>
<td>This supplier tends to take advantage from my hotel with guile.</td>
<td>0.705</td>
</tr>
<tr>
<td>It is very difficult to assess the performance of this supplier.</td>
<td>0.742</td>
</tr>
<tr>
<td>We should better select other suppliers.</td>
<td>0.688</td>
</tr>
</tbody>
</table>
'sem' package

- sem

Figure: http://socserv.socsci.mcmaster.ca/jfox/Courses/Brazil-2008/index.html
Supply Chain Collaborations

Definition

"At least two firms in the same supply chain work together to achieve their mutual goals" (Mentzer et al., 2001; Simatupang and Sridharan, 2005).
## Costs and Benefits

- **Costs**: Finding the best suppliers (e.g., price, quality), Monitoring (QC) (Barratt, 2004; Holweg et al., 2005)
- **Benefits**: Better level of responsiveness and service level (Speckman, 1998; Holweg et al., 2005)