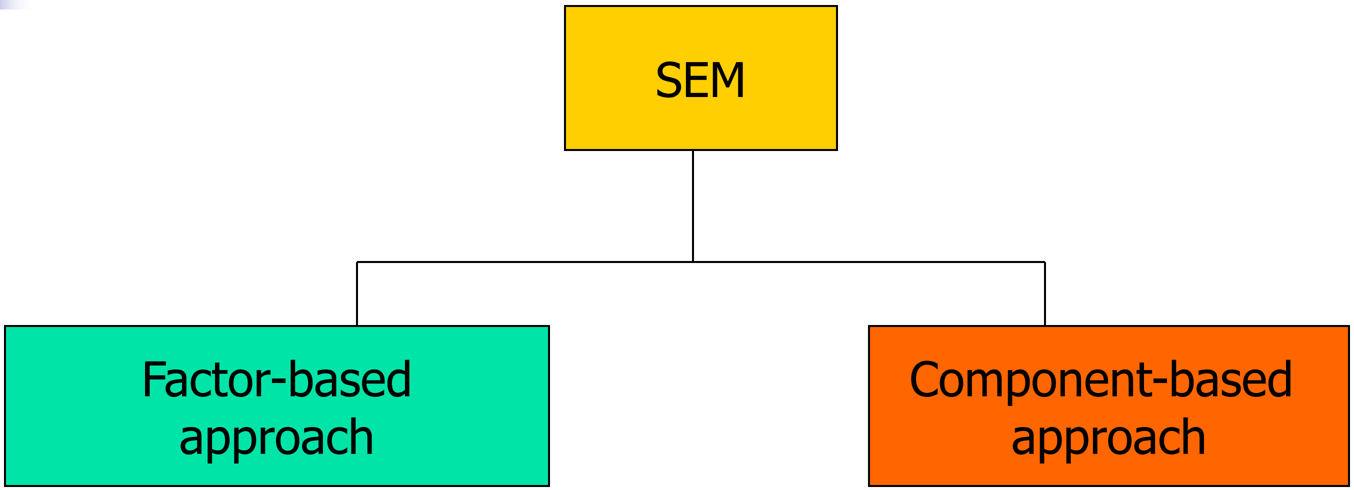
Generalized Structured Component Analysis: A Component-based Approach to Structural Equation Modeling

Heungsun Hwang
Department of Psychology
McGill University

Structural Equation Models

- Structural Equation Models (SEM) are used for the specification and analysis of interdependencies among manifest variables and hypothesized underlying theoretical constructs, often called latent variables.
  - aka path analysis with latent variables

Two Approaches to SEM



Latents ≈ Factors

- Covariance structure analysis

Latents ≈ Components

- PLS path modeling
- GSCA

Generalized Structured Component Analysis (Hwang & Takane, 2004)

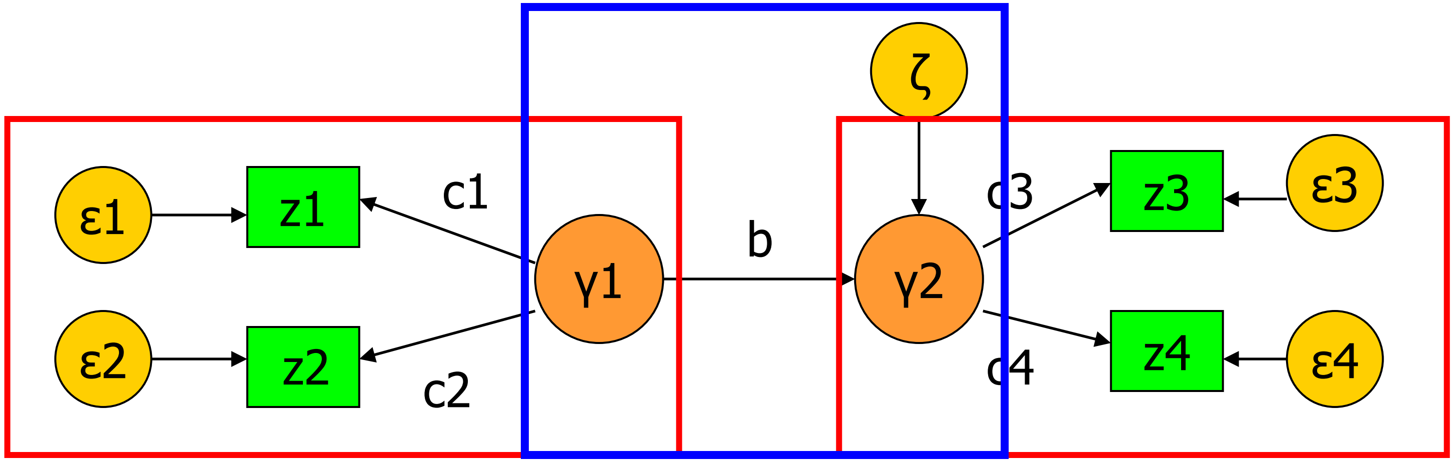
- Model specification
  - Involves three sub-models
  - Combines the sub-models into a single one
- Parameter estimation
  - A single least-squares optimization function
  - Alternating least-squares algorithm

The GSCA Model: Submodels

$$\boldsymbol{\gamma} = \mathbf{B}\boldsymbol{\gamma} + \boldsymbol{\zeta}$$

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ b & 0 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

Structural/inner model



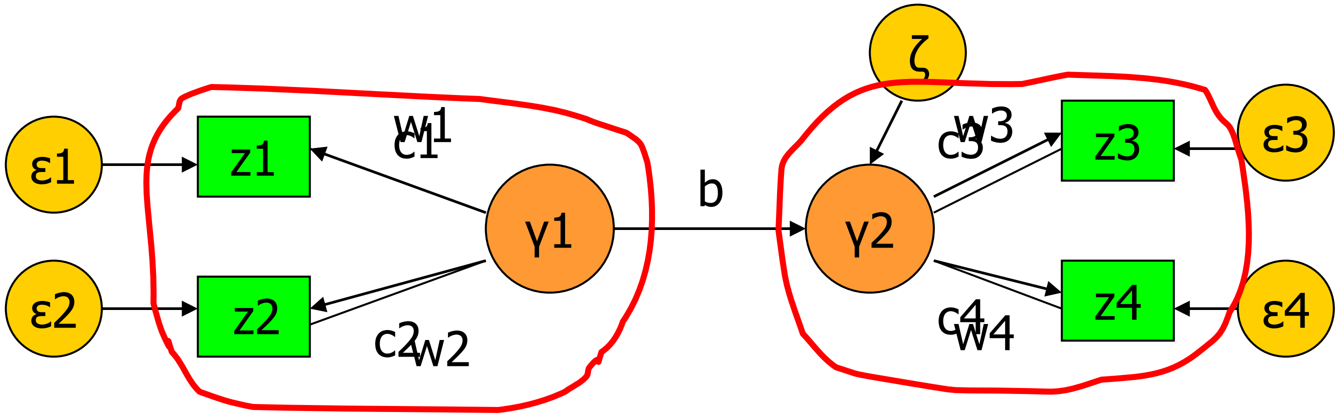
Measurement/outer model

$$\mathbf{z} = \mathbf{C}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} c_1 & 0 \\ c_2 & 0 \\ 0 & c_3 \\ 0 & c_4 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

The GSCA Model: Submodels

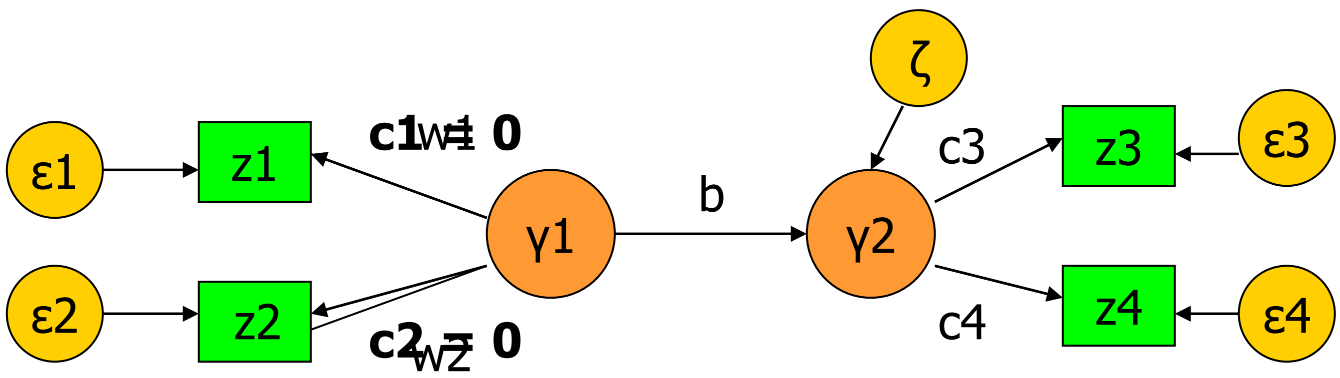
In GSCA, latent variables are defined as components or weighted sums of manifest variables.



$\gamma_1 = z_1 w_1 + z_2 w_2$ $\gamma_2 = z_3 w_3 + z_4 w_4$

$$\boldsymbol{\gamma} = \mathbf{Wz} \qquad \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & 0 & 0 \\ 0 & 0 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix}$$

What about formative indicators?



Measurement/outer model

$$\mathbf{z} = \mathbf{C}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & c_3 \\ 0 & c_4 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

The GSCA Model: Submodels

- Measurement model: $$\mathbf{z} = \mathbf{C}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$
- Structural model: $$\boldsymbol{\gamma} = \mathbf{B}\boldsymbol{\gamma} + \boldsymbol{\zeta}$$
- Weighted relation: $$\boldsymbol{\gamma} = \mathbf{W}\mathbf{z}$$

The GSCA Model

$$\boldsymbol{\gamma} = \mathbf{W}\mathbf{z}$$

$$\begin{bmatrix} \mathbf{z} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{C} \\ \mathbf{B} \end{bmatrix} \boldsymbol{\gamma} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\zeta} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{I} \\ \mathbf{W} \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{C} \\ \mathbf{B} \end{bmatrix} \mathbf{W}\mathbf{z} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\zeta} \end{bmatrix}$$

$$\mathbf{V}\mathbf{z} = \mathbf{A}\mathbf{W}\mathbf{z} + \mathbf{e}$$

GSCA: Parameter estimation

- The unknown parameters of GSCA (**W** & **A**) are estimated such that the sum of squares of the residuals ($\mathbf{e}_i$) is as small as possible.
- This is equivalent to minimizing the following single least-squares criterion:

$$\phi = \sum_{i=1}^{N} (\mathbf{V}\mathbf{z}_i - \mathbf{A}\mathbf{W}\mathbf{z}_i)'(\mathbf{V}\mathbf{z}_i - \mathbf{A}\mathbf{W}\mathbf{z}_i)$$

GSCA: Parameter estimation

- An alternating least squares (ALS) algorithm was developed to minimize this criterion.

- The ALS algorithm alternates two main steps:
  - Step 1: **A** is updated for fixed **W** in an LS sense
  - Step 2: **W** are updated for fixed **A** in an LS sense

- The bootstrap method is used for estimating standard errors or confidence intervals.

GSCA: Overall model fit measures

- GSCA provides overall model fit measures:

$$\text{FIT} = 1 - \left[\sum_{i=1}^{N} (\mathbf{Vz}_i - \mathbf{AWz}_i)'(\mathbf{Vz}_i - \mathbf{AWz}_i) / \sum_{i=1}^{N} (\mathbf{z}_i'\mathbf{V}'\mathbf{Vz}_i) \right]$$

$$\text{AFIT} = 1 - (1 - \text{FIT})(NJ/(NJ - K))$$

FIT and AFIT take into account the variance of the data explained by a model specification.

GSCA: Overall model fit measures

- Two additional measures are available to show the closeness between the sample covariances and the covariances reproduced by the model parameter estimates.

$$\text{GFI} = 1 - \frac{tr\left[(\mathbf{S} - \hat{\mathbf{\Sigma}})^2\right]}{tr(\mathbf{S}^2)}$$

$$\text{SRMR} = \sqrt{2\sum_{j=1}^{J}\sum_{q=1}^{j}\frac{\left[(s_{jq} - \hat{\sigma}_{jq})/(s_{jj}s_{qq})\right]^2}{J(J+1)}}$$

GSCA: Overall model fit measures

- From the generalized structured component analysis model, we have

$$\mathbf{Vz} = \mathbf{AWz} + \mathbf{e}$$

$$(\mathbf{V} - \mathbf{AW})\mathbf{z} = \mathbf{Qz} = \mathbf{e}$$

$$\mathbf{z} = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{e} = \mathbf{\Omega e}$$

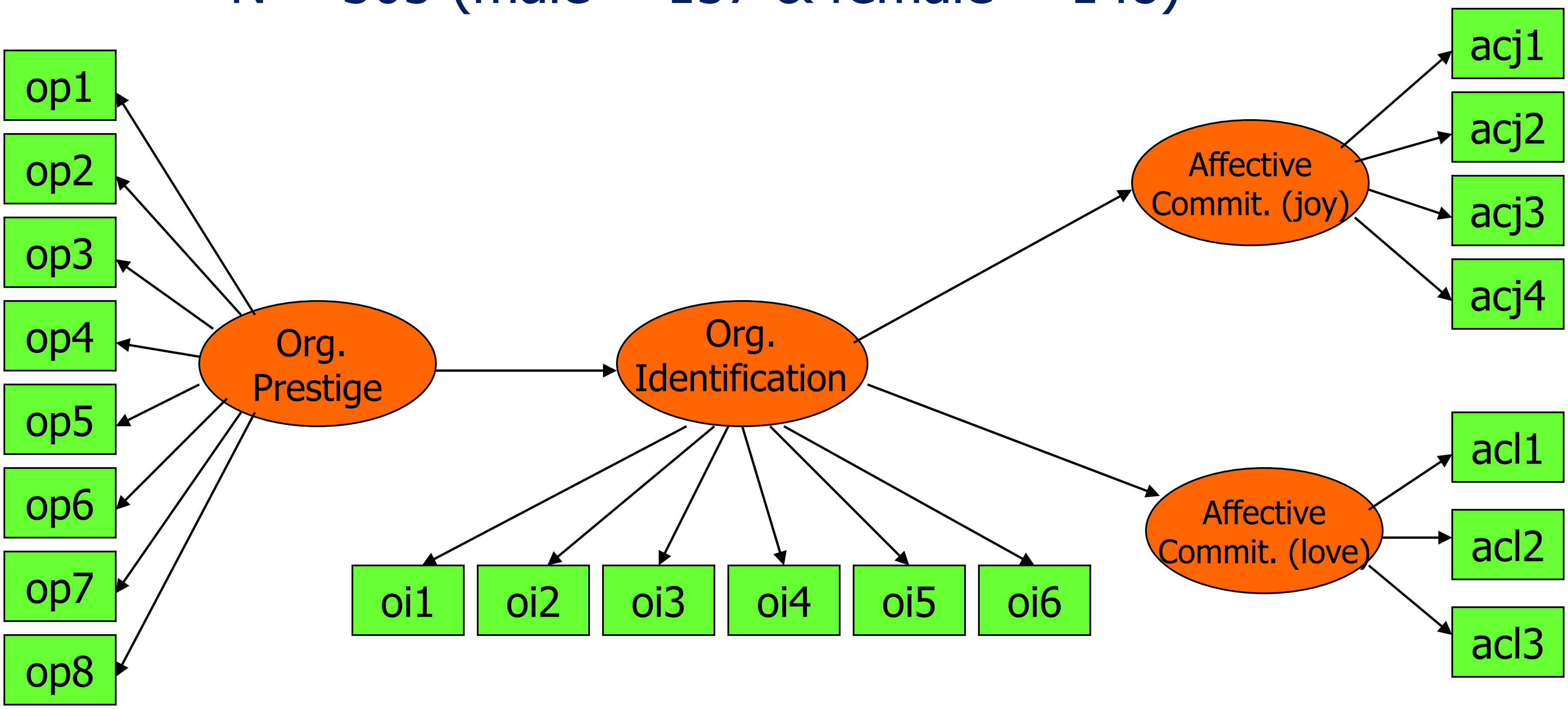
Then, $\mathbf{\Sigma} = \mathbf{\Omega}\, E(\mathbf{ee}')\, \mathbf{\Omega}'$.

Additional Features

- Constrained analysis
- Multiple group analysis with the imposition of cross-group constraints
- Higher-order latent variables
- Total and indirect effects

Example: Organizational Identification Data

N = 305 (male = 157 & female = 148)



Example: Organizational Identification Data

- For the unconstrained, 2-group model,
  - FIT = .53 (se = .01, 95% CI = .50 -.55)
  - AFIT = .53 (se = .01, 95% CI = .50-.55)
  - GFI = .99 (se = .00, 95% CI = .99 - .99)
  - SRMR = .06 (se = .00, 95% CI = .06 - .08)

Weight estimates (unconstrained model)

Latent	Indicator	Male				Female			
		Estimate	SE	|t|	CI	Estimate	SE	|t|	CI
OP	op1	.15	.01	11.31	.13-.18	.15	.02	8.98	.12-.18
	op2	.15	.01	11.57	.13-.18	.17	.01	11.42	.14-.20
	op3	.17	.02	8.89	.14-.21	.14	.01	10.37	.12-.17
	op4	.14	.02	9.18	.11-.17	.15	.01	14.96	.13-.18
	op5	.17	.02	10.60	.14-.20	.16	.01	12.25	.14-.18
	op6	.17	.01	11.75	.14-.20	.17	.01	14.23	.15-.20
	op7	.15	.02	8.94	.12-.18	.15	.01	12.48	.13-.18
	op8	.16	.01	11.53	.14-.19	.15	.01	13.38	.12-.17
OI	oi1	.20	.03	6.59	.15-.26	.24	.03	7.39	.17-.31
	oi2	.19	.02	8.65	.16-.24	.22	.03	8.16	.17-.27
	oi3	.17	.02	7.38	.12-.22	.22	.02	9.60	.18-.27
	oi4	.30	.04	8.44	.23-.37	.24	.03	8.69	.20-.30
	oi5	.20	.03	6.35	.15-.27	.26	.03	8.28	.22-.34
	oi6	.18	.03	6.53	.15-.27	.21	.03	7.18	.16-.27
AC_J	acj1	.31	.02	12.73	.26-.36	.29	.03	9.31	.23-.35
	acj2	.29	.03	11.52	.24-.34	.41	.03	14.01	.35-.47
	acj3	.36	.03	14.14	.30-.40	.36	.03	11.85	.30-.43
	acj4	.30	.03	8.85	.23-.38	.30	.03	9.21	.22-.35
AC_L	acl1	.42	.03	15.82	.37-.48	.48	.05	10.49	.39-.58
	acl2	.39	.03	11.73	.33-.45	.38	.07	5.73	.22-.50
	acl3	.44	.04	11.63	.38-.52	.52	.05	9.92	.41-.61

Loading estimates (unconstrained model)

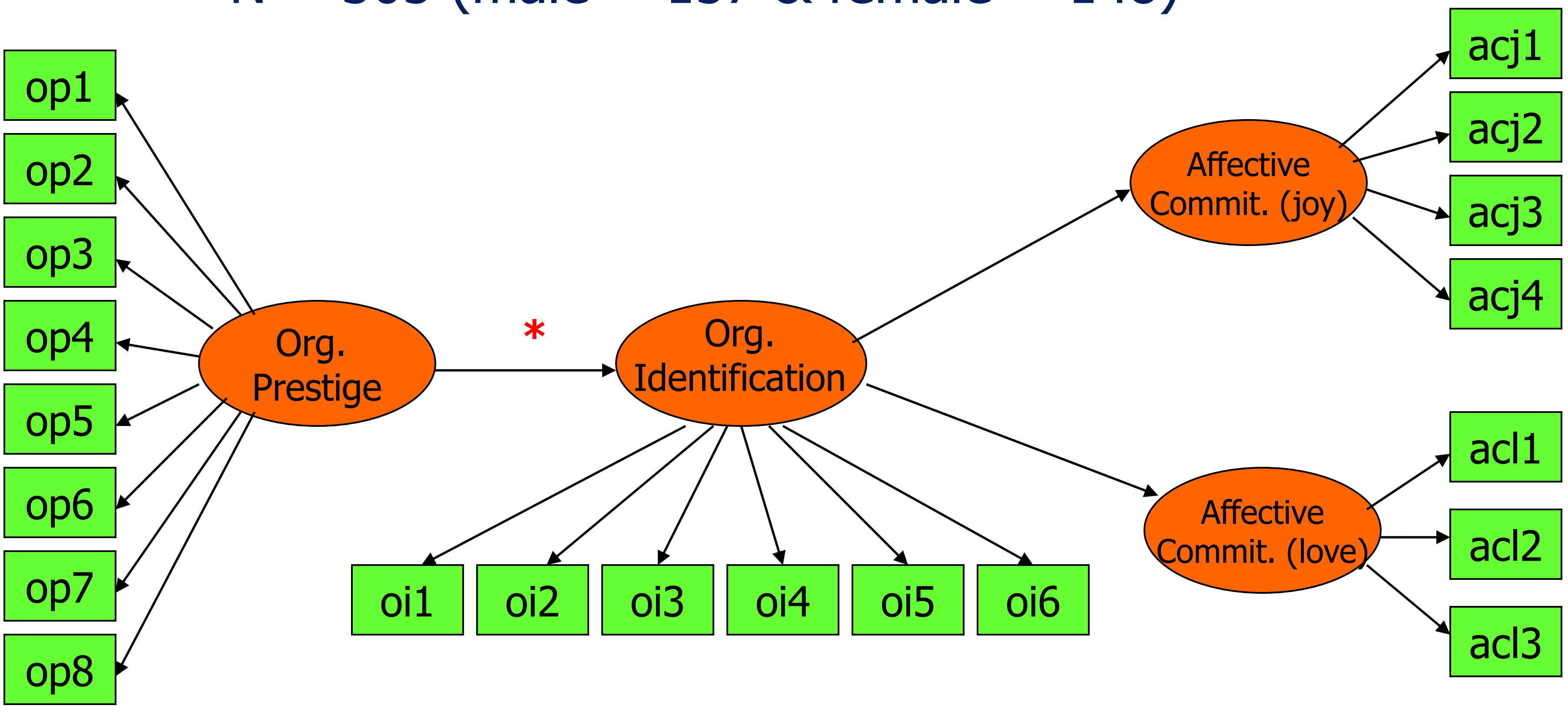
Latent	Indicator	**Male**				**Female**			
		Estimate	SE	|t|	CI	Estimate	SE	|t|	CI
OP	op1	.80	.05	15.05	.64-.87	.76	.04	20.60	.68-.82
	op2	.79	.04	20.37	.69-.84	.85	.02	39.07	.80-.89
	op3	.73	.05	13.42	.61-.82	.80	.03	22.97	.72-.85
	op4	.79	.05	14.53	.64-.86	.82	.04	22.49	.74-.87
	op5	.80	.05	17.72	.68-.86	.81	.04	19.20	.71-.87
	op6	.84	.03	29.34	.78-.89	.84	.04	20.88	.74-.90
	op7	.77	.04	19.40	.66-.83	.78	.04	21.73	.70-.83
	op8	.83	.04	19.94	.74-.90	.77	.05	14.65	.62-.85
OI	oi1	.83	.03	26.21	.75-.88	.73	.04	16.58	.62-.80
	oi2	.79	.03	24.11	.73-.85	.71	.05	13.64	.58-.80
	oi3	.64	.06	11.37	.53-.72	.65	.05	12.91	.55-.72
	oi4	.88	.02	43.48	.84-.92	.73	.06	12.48	.59-.82
	oi5	.85	.03	25.45	.77-.90	.75	.03	21.46	.69-.82
	oi6	.76	.05	16.32	.66-.83	.74	.06	12.63	.58-.81
AC_J	acj1	.79	.03	22.61	.71-.85	.69	.06	12.07	.58-.78
	acj2	.77	.05	16.29	.67-.84	.83	.02	34.10	.78-.88
	acj3	.85	.03	32.37	.79-.90	.76	.05	16.53	.65-.84
	acj4	.76	.03	22.18	.70-.82	.62	.07	8.45	.42-.74
AC_L	acl1	.84	.04	23.04	.76-.89	.75	.05	15.01	.63-.84
	acl2	.76	.05	16.11	.66-.83	.64	.11	5.87	.34-.80
	acl3	.81	.04	22.18	.72-.88	.77	.06	13.23	.65-.85

Path coefficient estimates (unconstrained model)

	Group	Estimate	SE	|t|	CI
OP → OI	Male	.37	.09	4.23	.22-.55
OI → AC_J		.71	.05	13.82	.60-.80
OI → AC_L		-.46	.09	5.25	-.63 - -.31
OP → OI	Female	.35	.07	5.27	.23 - .45
OI → AC_J		.47	.04	10.84	.40 - .56
OI → AC_L		-.34	.08	4.03	-.48 - -.14

Example: Organizational Identification Data

N = 305 (male = 157 & female = 148)



Example: Organizational Identification Data

- For the constrained, 2-group model,
  - FIT = .53 (se = .01, 95% CI = .50 -.56)
  - AFIT = .53 (se = .01, 95% CI = .50-.56)
  - GFI = .99 (se = .00, 95% CI = .99 - .99)
  - SRMR = .06 (se = .00, 95% CI = .06 - .09)

Example: Organizational Identification Data

- We also performed a paired t-test to compare the mean FIT values between the unconstrained and constrained models based on 100 bootstrap samples.
  - $t(99) = -.36$, $p = .72$, 95% CI = -.00 - .00.
  - There was no statistically significant mean difference in FIT between the constrained and unconstrained models.

Loading estimates (constrained model)

Latent	Indicator	Male				Female			
		Estimate	SE	|t|	CI	Estimate	SE	|t|	CI
OP	op1	.78	.03	23.49	.68-.82	.78	.03	23.49	.68-.82
	op2	.82	.02	35.87	.76-.86	.82	.02	35.87	.76-.86
	op3	.77	.03	25.45	.70-.82	.77	.03	25.45	.70-.82
	op4	.80	.03	23.16	.74-.87	.80	.03	23.16	.74-.87
	op5	.80	.03	27.44	.76-.86	.80	.03	27.44	.76-.86
	op6	.84	.02	33.88	.79-.89	.84	.02	33.88	.79-.89
	op7	.78	.03	29.47	.72-.82	.78	.03	29.47	.72-.82
	op8	.80	.03	24.71	.72-.86	.80	.03	24.71	.72-.86
OI	oi1	.78	.03	30.21	.72-.84	.78	.03	30.21	.72-.84
	oi2	.75	.03	23.96	.68-.80	.75	.03	23.96	.68-.80
	oi3	.64	.04	15.73	.54-.71	.64	.04	15.73	.54-.71
	oi4	.81	.03	27.80	.74-.86	.81	.03	27.80	.74-.86
	oi5	.80	.03	30.15	.74-.84	.80	.03	30.15	.74-.84
	oi6	.75	.03	21.49	.68-.81	.75	.03	21.49	.68-.81
AC_J	acj1	.74	.03	21.86	.67-.80	.74	.03	21.86	.67-.80
	acj2	.80	.03	30.45	.75-.85	.80	.03	30.45	.75-.85
	acj3	.81	.03	35.47	.75-.85	.81	.03	35.47	.75-.85
	acj4	.70	.04	19.67	.63-.78	.70	.04	19.67	.63-.78
AC_L	acl1	.79	.03	25.53	.71-.84	.79	.03	25.53	.71-.84
	acl2	.71	.05	13.77	.59-.78	.71	.05	13.77	.59-.78
	acl3	.79	.03	26.57	.73-.85	.79	.03	26.57	.73-.85

Path coefficient estimates (constrained model)

	Group	Estimate	SE	|t|	CI
OP → OI		.37	.06	6.50	.21-.46
OI → AC_J	Male	.71	.05	15.55	.61-.80
OI → AC_L		-.46	.08	5.89	-.60 - -.31
OP → OI		.37	.06	6.50	.21 - .46
OI → AC_J	Female	.47	.05	9.90	.40 - .56
OI → AC_L		-.33	.07	4.48	-.46 - -.14

Extensions of GSCA

- GSCA has been extended to improve data-analytic flexibility and generality.
  - Fuzzy clusterwise GSCA (Hwang et al., 2007)
  - Multilevel GSCA (Hwang et al., 2007)
  - Regularized GSCA (Hwang, 2008)
  - Nonlinear GSCA (Hwang & Takane, 2008)
  - GSCA with latent interactions (Hwang et al., 2010)
  - Kernel GSCA (Suk & Hwang, 2012)
  - Dynamic GSCA (Jung et al., 2012)
  - Functional GSCA (Suk & Hwang, 2013)

Functional Data

- Functional data represent data collected in the form of curves, surfaces, images, or anything else varying over a continuum.
  - The continuum can be time, space, wavelength, probability, and so forth.

- Functional data can be characterized by high-frequency repeated measurements that involve smooth but often intricate trajectories of change over a continuum.

Functional Data

- Due to ever-increasing technological advances in measurement, functional data become ubiquitous in the natural, social, and health sciences.
- A few examples in psychology include:
  - motor control data (Mattar & Ostry, 2010)
  - musical perception data (Vines et al., 2005)
  - eye-tracking data (Jackson & Sirois, 2009)
  - neuroimaging data (Tian, 2010; Hwang et al., 2012).

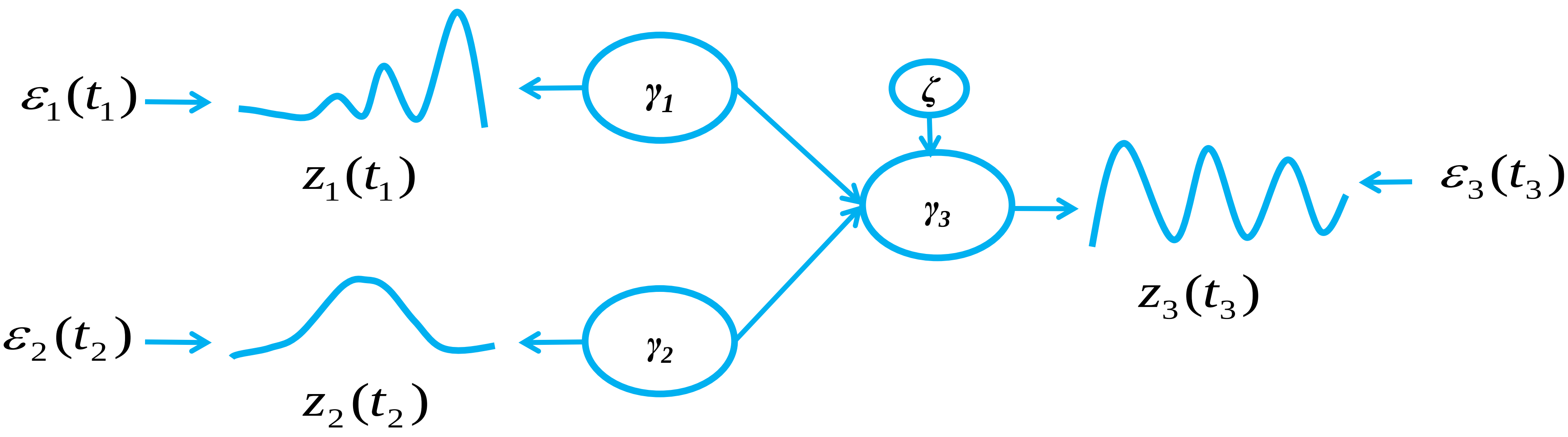
Functional GSCA

- Functional GSCA was developed for (component-based) structural equation modeling of functional data.

- In functional GSCA, manifest variables are replaced by functions, while latent variables are components of functions.

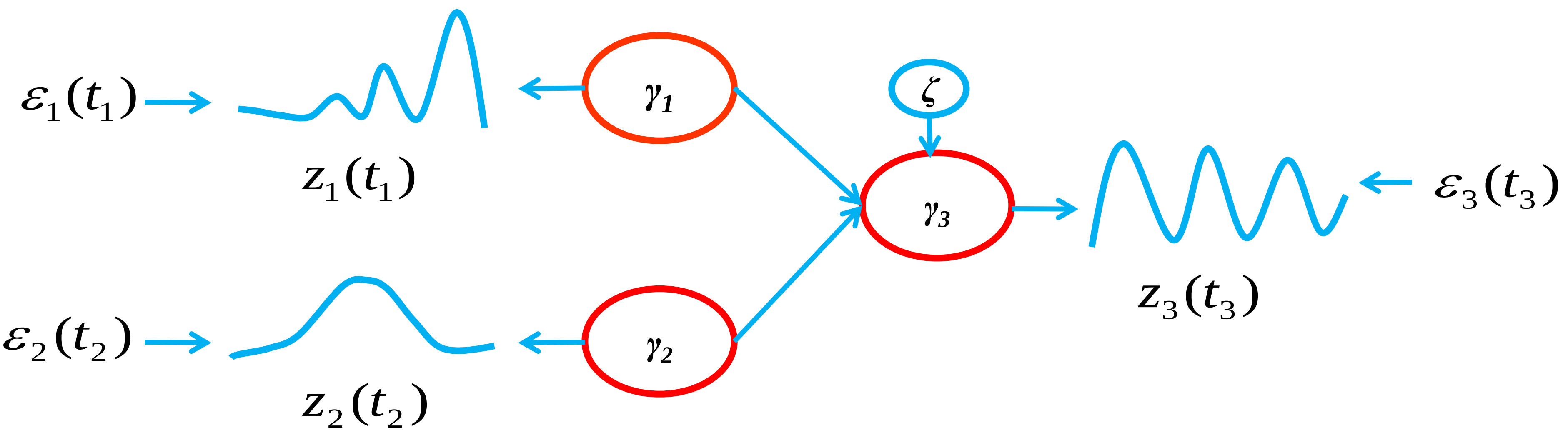
Functional GSCA

- A prototype model that has three blocks of functional data.



Functional GSCA: Submodels

- Weighted relation:



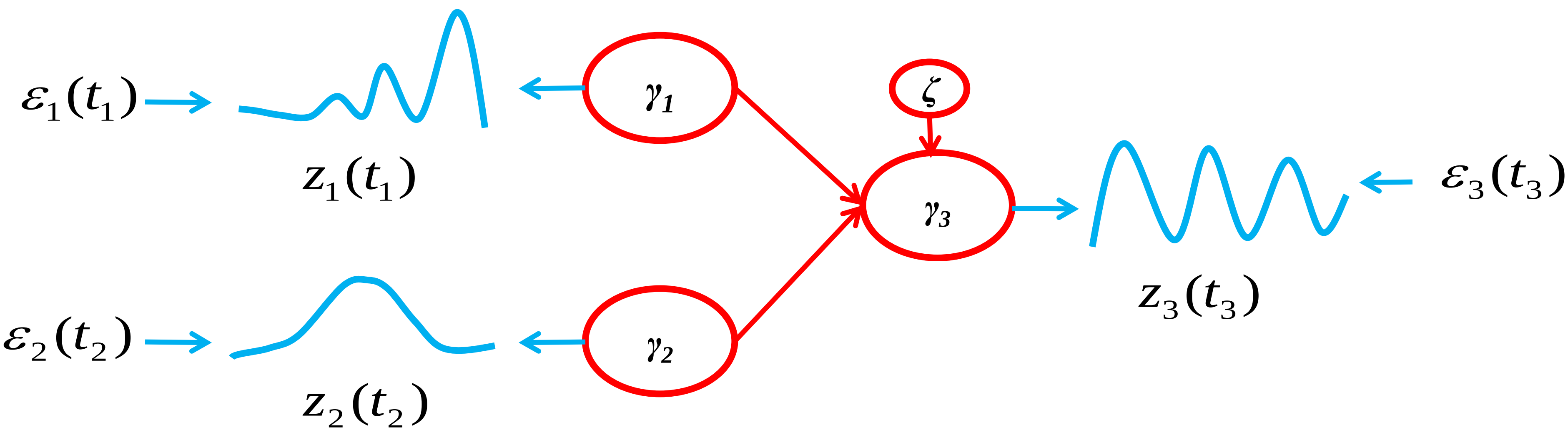
$$\gamma_1 = \int z_1(t_1) w_1(t_1) dt_1$$

$$\gamma_2 = \int z_2(t_2) w_2(t_2) dt_2$$

$$\gamma_3 = \int z_3(t_3) w_3(t_3) dt_3$$

Functional GSCA: Submodels

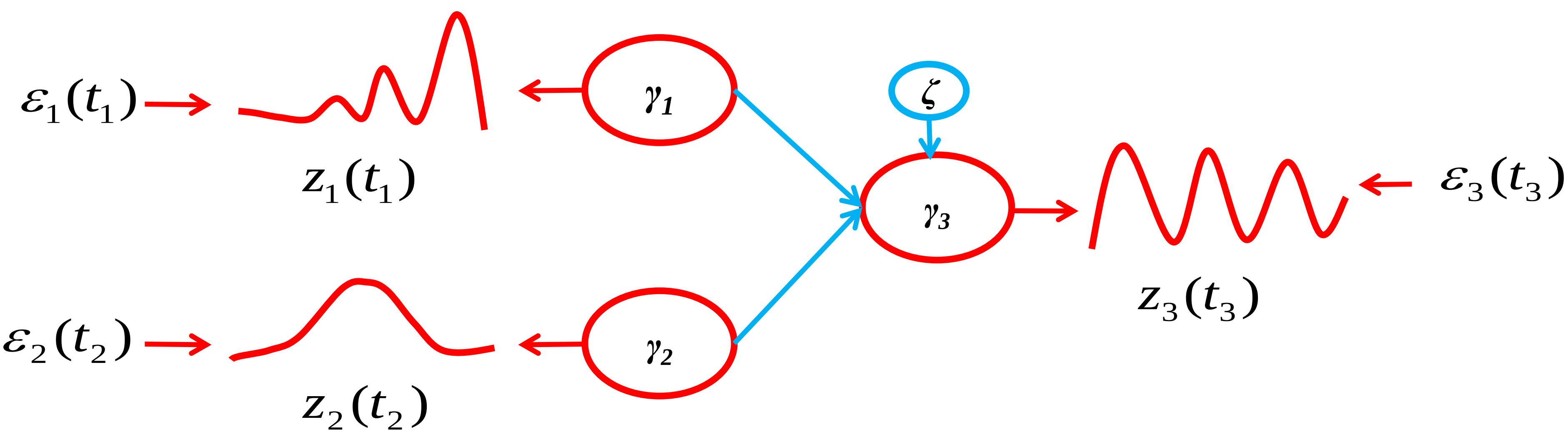
- Structural model:



$$\gamma_3 = \gamma_1 b_1 + \gamma_2 b_3 + \zeta$$

Functional GSCA: Submodels

- Measurement model:



$$z_1(t_1) = \gamma_1 c_1(t_1) + \varepsilon_1(t_1)$$

$$z_2(t_2) = \gamma_2 c_2(t_2) + \varepsilon_2(t_2)$$

$$z_3(t_3) = \gamma_3 c_3(t_3) + \varepsilon_3(t_3)$$

Functional GSCA: Parameter Estimation

- We minimize the following penalized least-squares function to estimate parameters.

$$\phi = \sum_{i=1}^{N}\sum_{p=1}^{P}\int \varepsilon_{ip}^{2}(t_p)dt_p + \sum_{i=1}^{N}\zeta_i^2$$

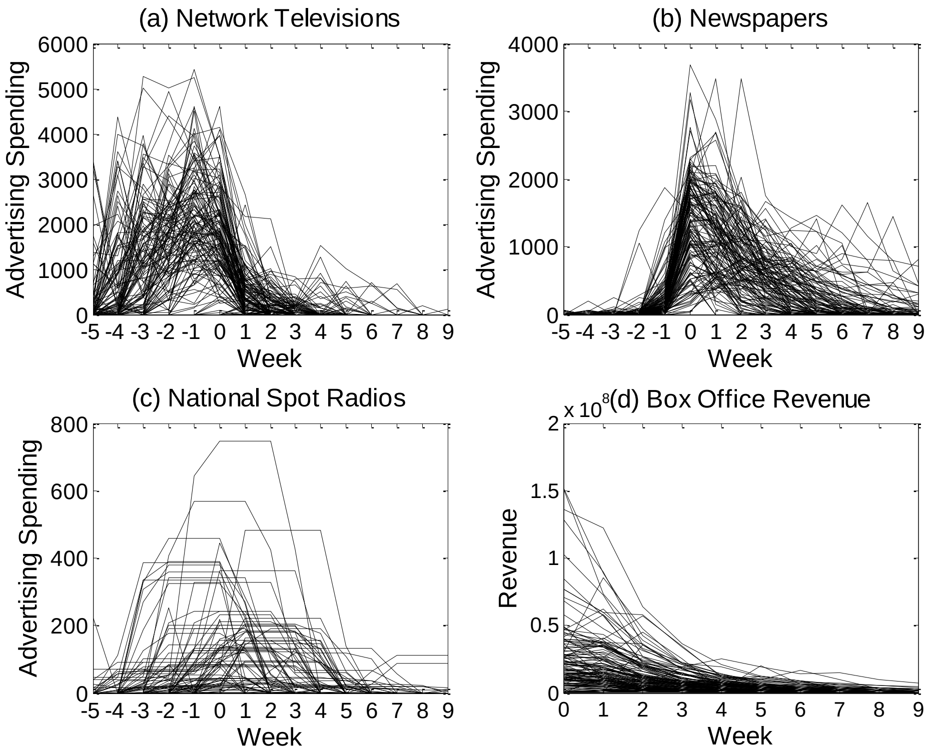
$$+ \lambda_w \sum_{p=1}^{P}\int (D^2 w_p(t_p))^2 dt_p + \lambda_c \sum_{p=1}^{P}\int (D^2 c_p(t_p))^2 dt_p$$

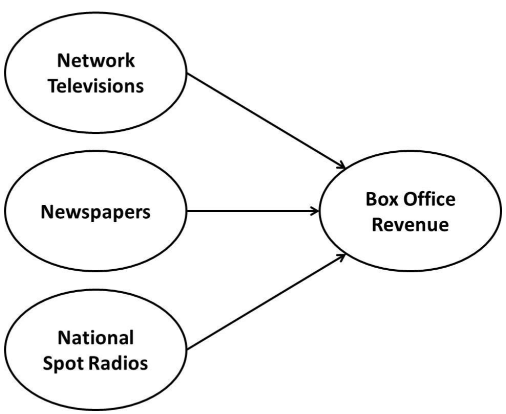
Functional GSCA: Parameter Estimation

- To minimize this function, we use an alternating penalized least-squares algorithm (e.g., Hwang, 2009), in combination with the basis-function expansion approach (Ramsay & Silverman, 2005, Chapter 3).

Example: Movie Revenue Data

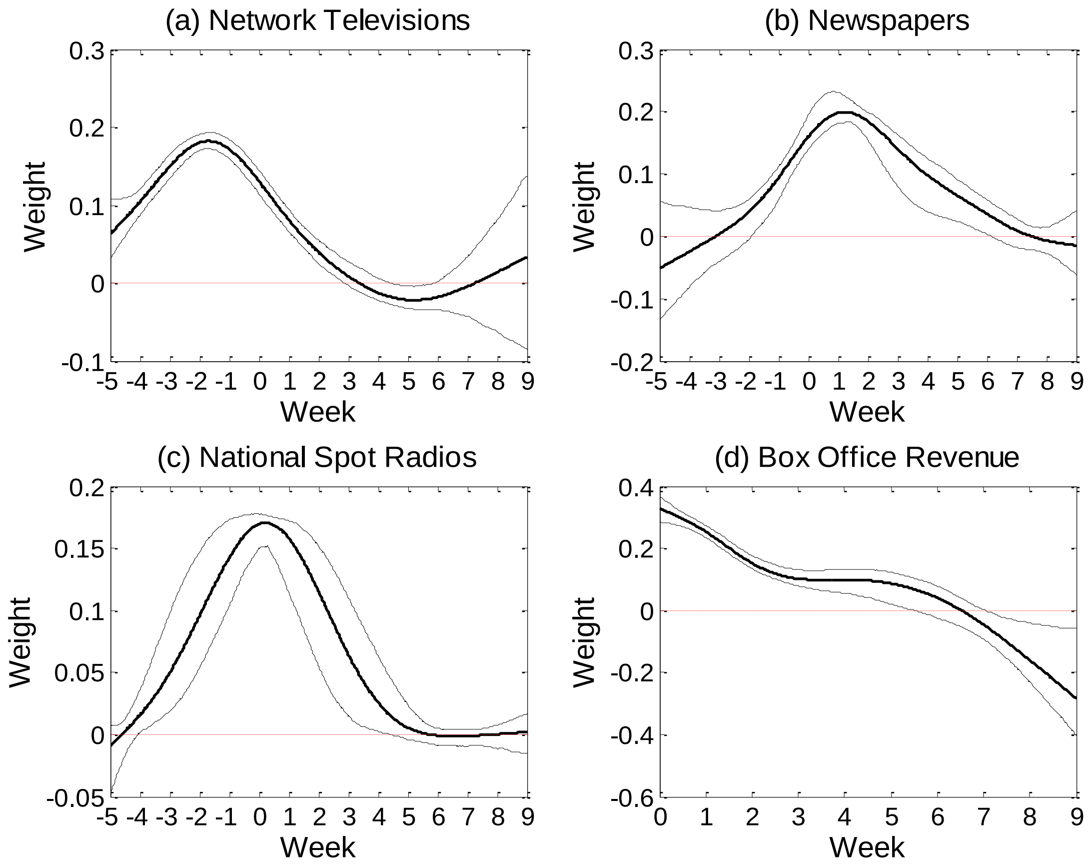
- We examine how advertising spending on different media affects the movie box office revenue.





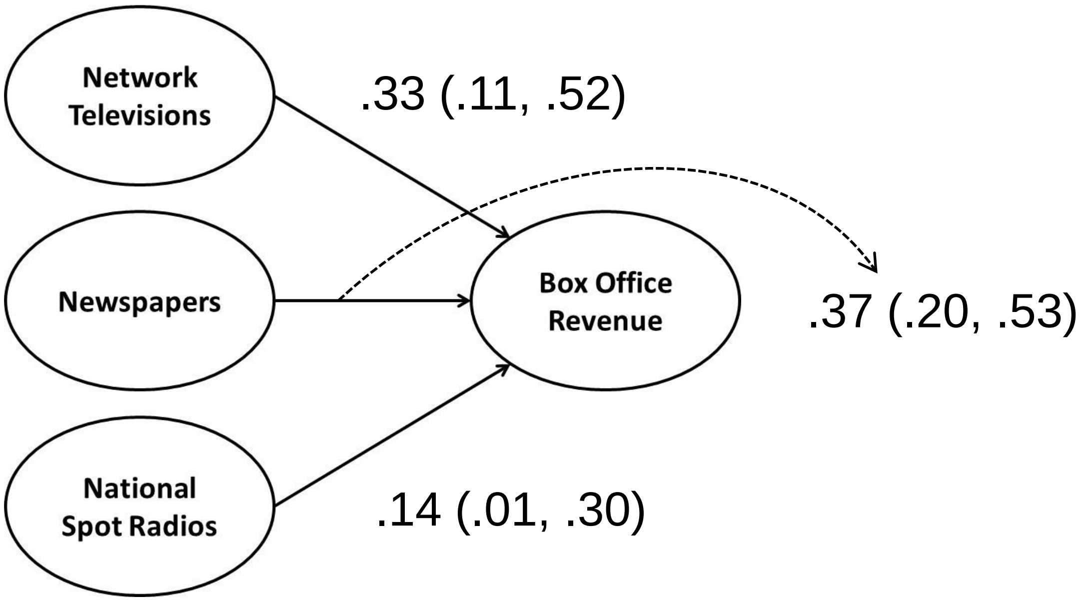
Example: Movie Revenue Data

- Weight functions estimated



Example: Movie Revenue Data

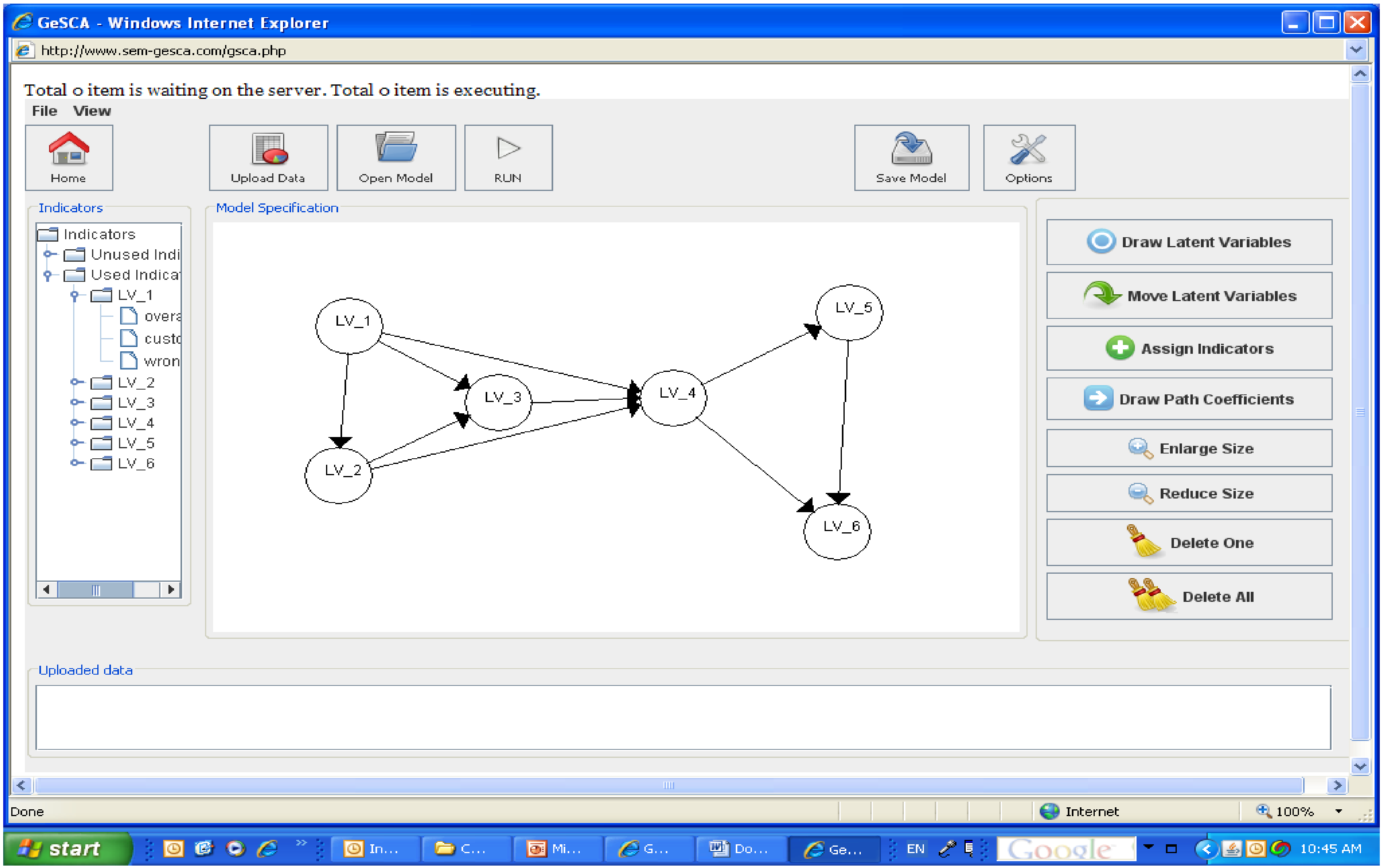
- Path coefficient estimates



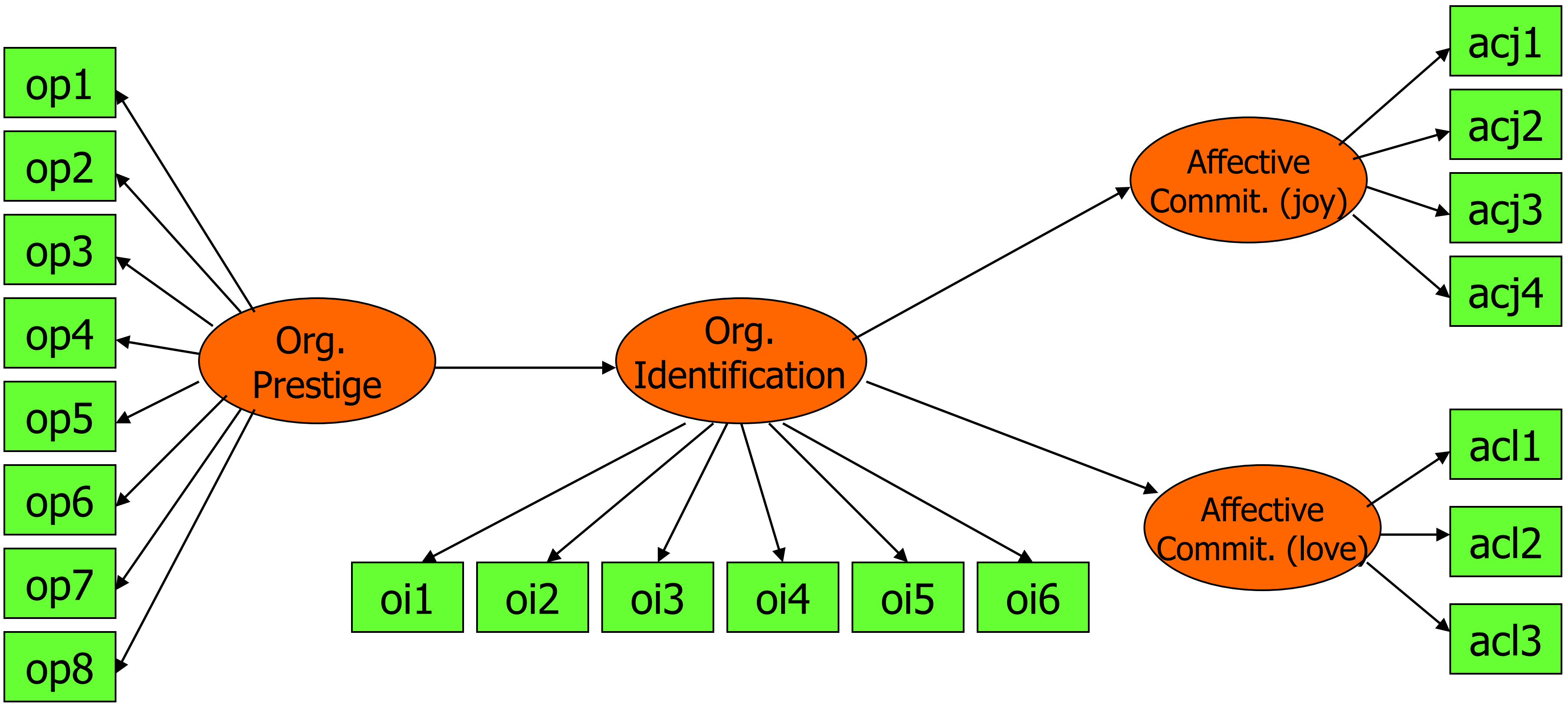
GeSCA

Generalized Structured Component Analysis

(www.sem-gesca.org)



Example: Organizational Identification Data



Thank you!