Geo 873 – 001: Seminar in Human-Environment Geography

12:40 am – 3:30 pm; Geo 120

Zoom Link will be added to facilitate renewed needs, staring on Feb. 22, 2023

Introduction of Structural Equation Modeling (SEM)

- SEM basics(3/1)(Petri Nokelainen, University of Tampere, Finland; petri.nokelainen@uta.fi)

(https://www.researchgate.net/profile/Salam_Hmood/post/structural_equation_modeling/attachment/59d6551c79197b80779ac68a/AS%3A524275057479680%401502008322465/download/sem_en.ppt)

- SEM Demonstration in R;
  (1) Venkatesh Kolluru, University of South Dakota; Email: Venkatesh.Kolluru@coyotes.usd.edu
  (2) Gabriela Shirkey, LEES Lab, MSU; Email: shirkeyg@msu.edu
- Homework and term papers: assessment and scheduling
- Case studies in SES research

Reading

Venkatesh, K., John, R., Chen, J., Jarchow, M., Amirkhiz, R. G., Giannico, V., ... & Yuan, J. (2022). Untangling the impacts of socioeconomic and climatic changes on vegetation greenness and productivity in Kazakhstan. *Environmental Research Letters*, 17(9), 095007.

Meet with potential employers at the

GEOGRAPHY CAREER DAY AND JOB FAIR

Department of Geography, Environment and Spatial Sciences

Friday, March 31, 2023
2:30-4:30 PM
Geography Building

673 Auditorium Rd, East Lansing, MI 48824

REGISTER TODAY

ALL MAJORS WELCOME

Questions? Email geo@msu.edu

Department of Geography,
Environment, and Spatial Sciences
MICHIGAN STATE UNIVERSITY

#MSUGeography

Comments on Homework 1-2

- Please use written language
- Figures and tables need detailed, self-explanatory titles
- Use the spelling tools before submission
- Precision uses of words and statements
- “Do Not” copy from published work
- Professional use of citations and reference listing
- Submit future homework in Microsoft Word
- Q/A?

updated (both R and RStudio), please load the following packages:

- Lavaan
- semptools
- semPlot
- seminr
- dplyr

For those less familiar with R, it is OK if the package requires additional package downloads like auxiliary packages. If you have trouble downloading or want to refresh yourself on R before we meet, check out this link: https://researchwithfawad.com/index.php/lp-courses/seminr-lecture-series/an-introduction-to-r-and-r-studio/

For the PLS-SEM demo, please also access the following materials. Feel free to explore before class, but these are just for additional information:

1. YouTube series on using the SEMiR package for PLS-SEM: https://www.youtube.com/watch?v=PogcECaqxgM
2. Download the book and tutorial accompaniment: https://link.springer.com/book/10.1007/978-3-030-80519-7

Schedule for the remaining weeks

Mar 22-29:

- **Global climate change** and consequences (ecosystems and societies)
- **Homework 3**

Apr 5:

- **Globalization** & local practices: food, energy, water and **infectious diseases**
- **Renewable energy**: Bioenergy, solar, wind, earth heat, etc.

Apr 12

- Guest Lecture by Dr. Suraj Upadhaya, Iowa State University
- **Food** and agriculture
- **Homework 4**

Apr 19

- Guest lecture by Dr. Jack Liu, CHANS & Telecoupling
- **Water** (transboundary water)
- Q/A on student presentations

Apr 26

- Student presentations (10-15 min each)
- Teaching evaluation

Model Constructing

- In this presentation, I will use both the LISREL 8 –software and AMOS 5 for SEM analysis and PRELIS 2 –software (Jöreskog et al., 1985) for preliminary data analysis.
- All the previously mentioned approaches to SEM use the same pattern for constructing the model:
  1. model hypotheses,
  2. model specification,
  3. model identification and
  4. model estimation.

1. Model Hypotheses

Table 1. Variable Description

	Item	Summary variable	Sample statement
SUPPORTIVE MANAGEMENT	X1	Participative Leadership	It is easy to be touch with the leader of the training programme.
	X2	Elaborative Leadership	This organization improves it's members professional development.
	X3	Encouraging Leadership	My superior appreciates my work.
FUNCTIONAL GROUP	X4	Collaborative Activities	My teacher colleagues give me help when I need it.
	X5	Teacher – Student Connections	Athmosphere on my lectures is pleasant and spontaneous.
	X6	Group Spirit	The whole working community co-operates effectively.

1. Model Hypotheses

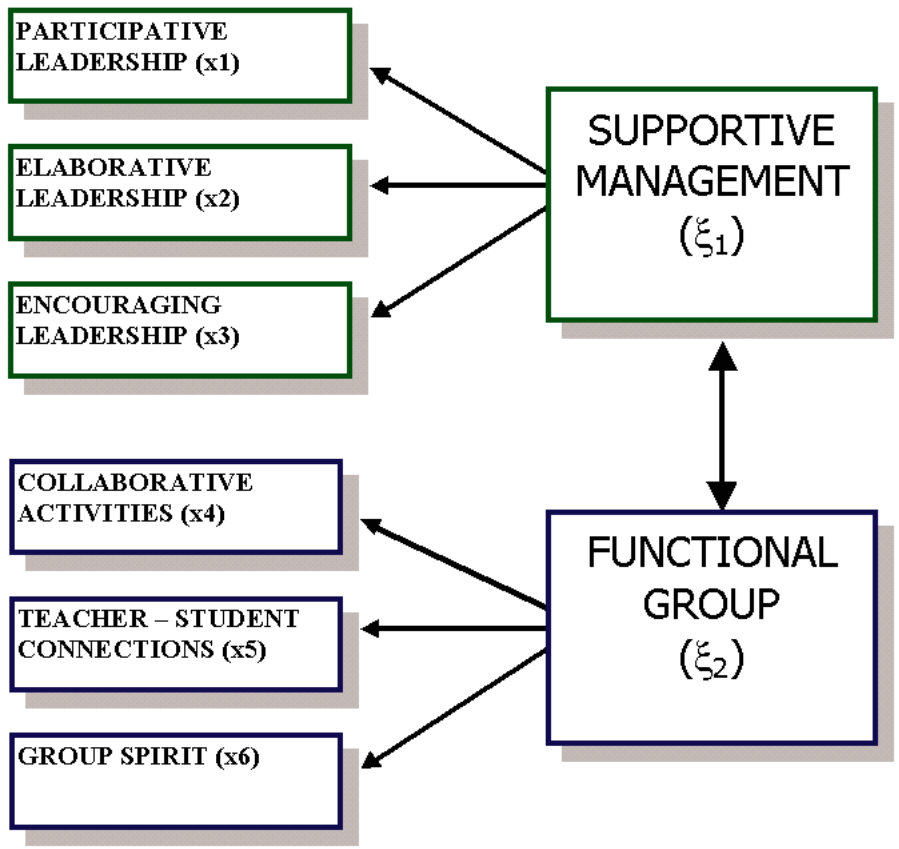
- A sample of the data is presented in Table 2.

Table 2. A Sample of the Raw Data Set

	Variables					
	Supportive Management			Functional Group		
Teachers	Participative Leadership (x1)	Elaborative Leadership (x2)	Encouraging Leadership (x3)	Collaborative Activities (x4)	Teacher-student Connections (x5)	Group Spirit (x6)
1.	2.75	3.25	4.00	2.60	3.00	2.00
2.	3.25	3.75	5.00	3.40	4.00	3.00
3.	3.50	3.75	4.00	3.60	4.75	3.00
. . .						
319	5.00	1.00	3.00	3.00	3.00	5.00

1. Model Hypotheses

Figure 4. Hypothesized Model

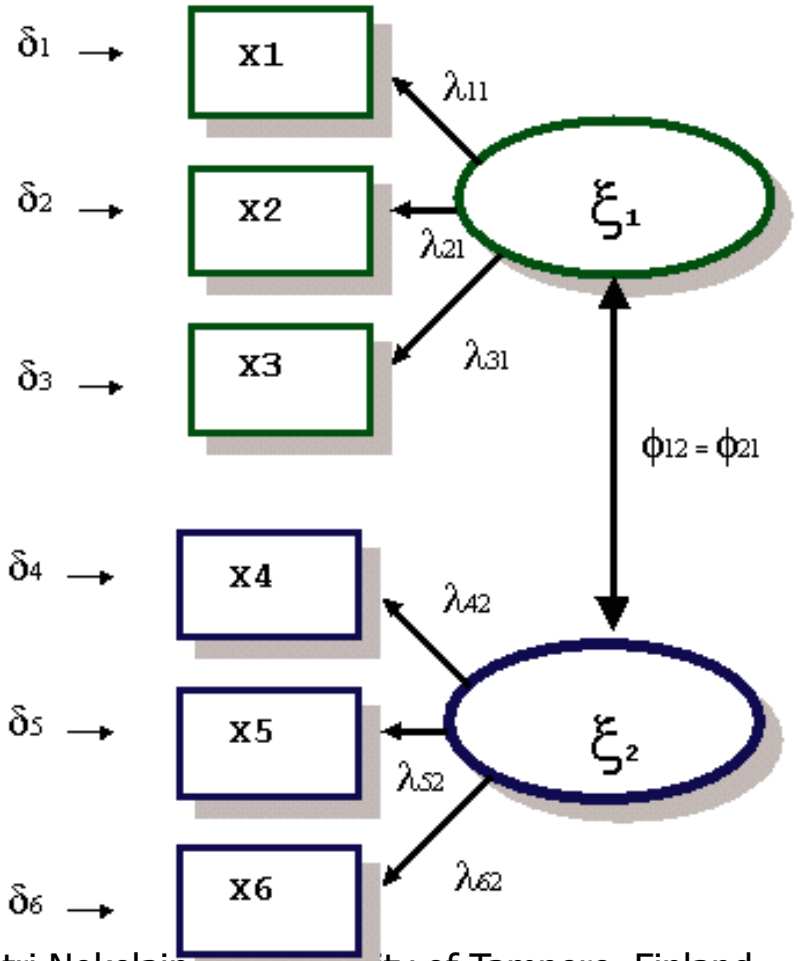


1. Model Hypotheses

- Two main hypotheses of interest are:
  - Does a two-factor model fit the data?
  - Is there a significant covariance between the supportive and functional factors?

2. Model Specification

Figure 5. Measurement Model



2. Model Specification

- Specification of the confirmatory factor model requires making formal and explicit statements about
  - the number of common factors,
  - the number of observed variables,
  - the variances and covariances among the common factors,
  - the relationships among observed variables and latent factors,
  - the relationships among residual variables and
  - the variances and covariances among the residual variables. (Jöreskog et al., 1989.)

3. Model Identification

- Identification is a theoretical property of a model, which depends neither on data or estimation.
  - When our model is identified we obtain unique estimates of the parameters.
- “Attempts to estimate models that are not identified result in arbitrary estimates of the parameters.” (Long, 1983, p. 35.)

3. Model Identification

- We gain constantly an identified model if
  - each observed variable in the model measures only one latent factor and
  - factor scale is fixed (Figure 6) or one observed variable per factor is fixed (Figure 7). (Jöreskog et al., 1979, pp. 196-197; 1984.)

4. Model Estimation

- When identification is approved, estimation can proceed.
- If the observed variables are normal and linear and there are more than 100 observations (319 in our example), Maximum Likelihood estimation is applicable.

4. Model Estimation

Figure 4. Hypothesized Model

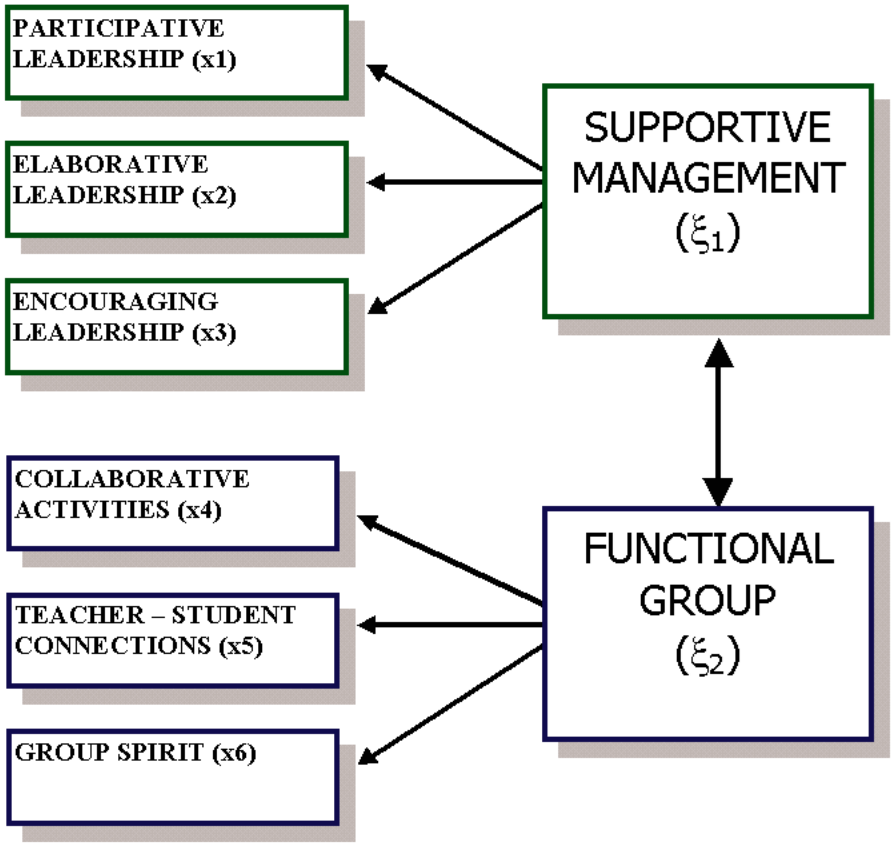
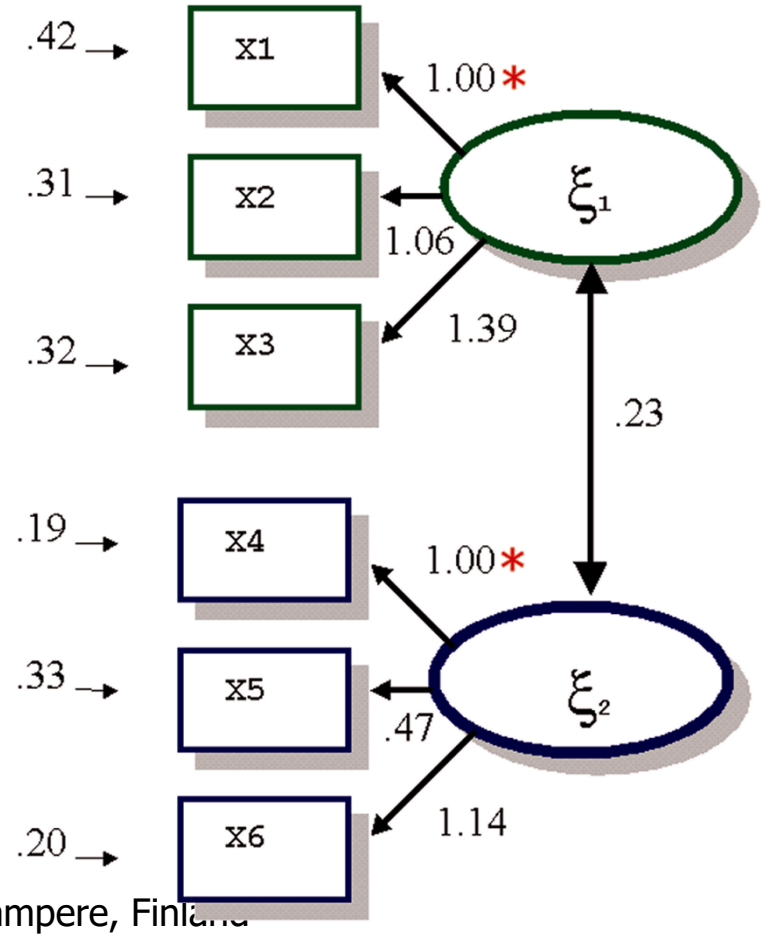


Figure 9. Parameter Estimates



Conclusions

- SEM has proven to be a very versatile statistical toolbox for educational researchers when used to confirm theoretical structures.
- Perhaps the greatest strength of SEM is the requirement of a prior knowledge of the phenomena under examination.
  - In practice, this means that the researcher is testing a theory which is based on an exact and explicit plan or design.
  - One may also notice that relationships among factors examined are free of measurement error because it has been estimated and removed, leaving only common variance.
  - Very complex and multidimensional structures can be measured with SEM; in that case SEM is the only *linear* analysis method that allows complete and simultaneous tests of all relationships.

Conclusions

- Disadvantages of SEM are also simple to point out.
  - Researcher must be very careful with the study design when using SEM for *exploratory* work.
  - As mentioned earlier, the use of the term ‘causal modeling’ referring to SEM is misleading because there is nothing causal, in the sense of inferring causality, about the use of SEM.
  - SEM's ability to analyze more complex relationships produces more complex models: Statistical language has turned into jargon due to vast supply of analytic software (LISREL, EQS, AMOS).
  - When analyzing scientific reports methodologically based on SEM, usually a LISREL model, one notices that they lack far too often decent identification inspection which is a prerequisite to parameter estimation.

Conclusions

- Overgeneralization is always a problem – but specifically with SEM one must pay extra attention when interpreting causal relationships since *multivariate normality* of the data is assumed.
  - This is a severe limitation of linear analysis in general because the reality is seldom linear.
- We must also point out that SEM is based on covariances that are not stable when estimated from small (<200 observation) samples.
- On the other hand, too large (>200 observations) sample size is also a reported problem (e.g., Bentler et al., 1983) of the significance of χ^2.

Conclusions

- SEM programs allow calculation of modification indices which help researcher to fit the model to the data.
  - Added or removed dependencies must be based on theory!
  - Overfitting model to the data reduces generalizability!
- Following slides demonstrate the effect of sample size and model modification (according to modification indices).
  - Example 2 in the course exercise booklet.

SEM applications in SES research

Review | Open Access | Published: 22 November 2016

Applications of structural equation modeling (SEM) in ecological studies: an updated review

Yi Fan ✉, Jiquan Chen, Gabriela Shirkey, Ranjeet John, Susie R. Wu, Hogeun Park & Changliang Shao

Ecological Processes **5**, Article number: 19 (2016) | Cite this article

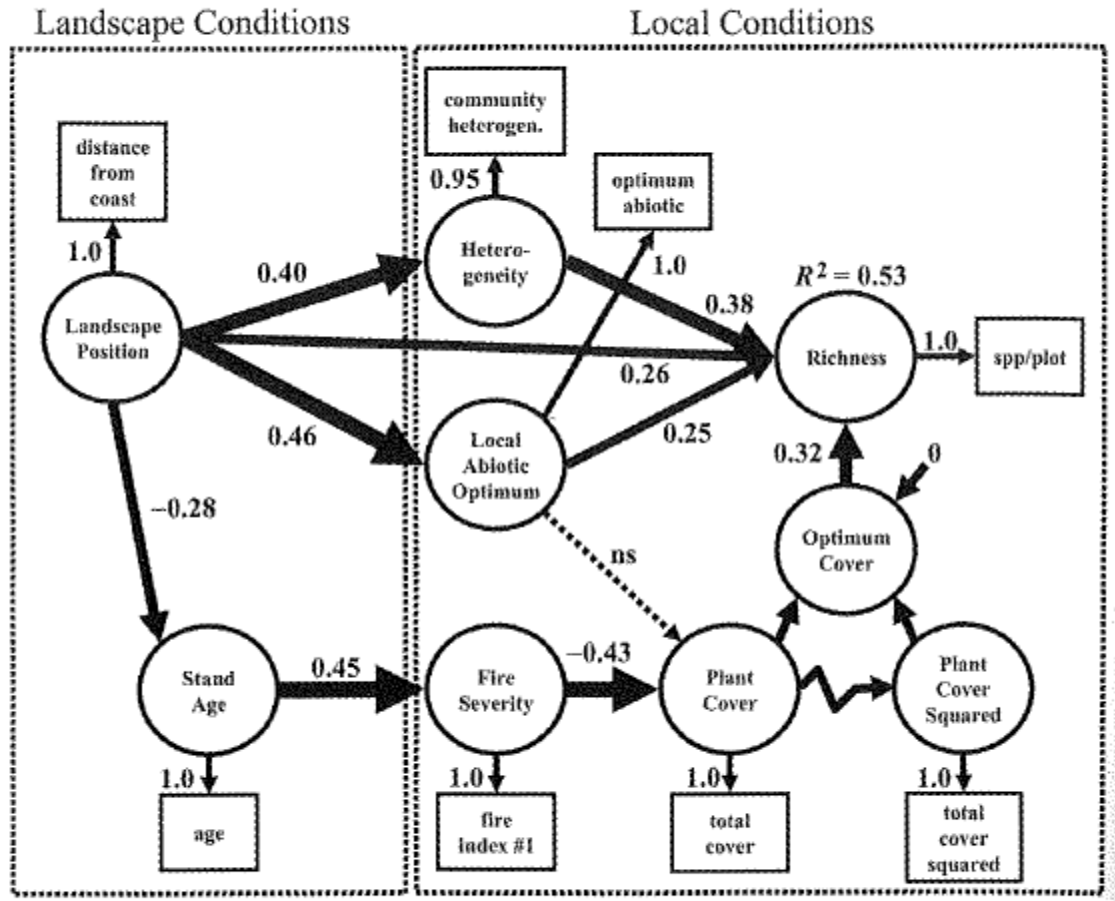
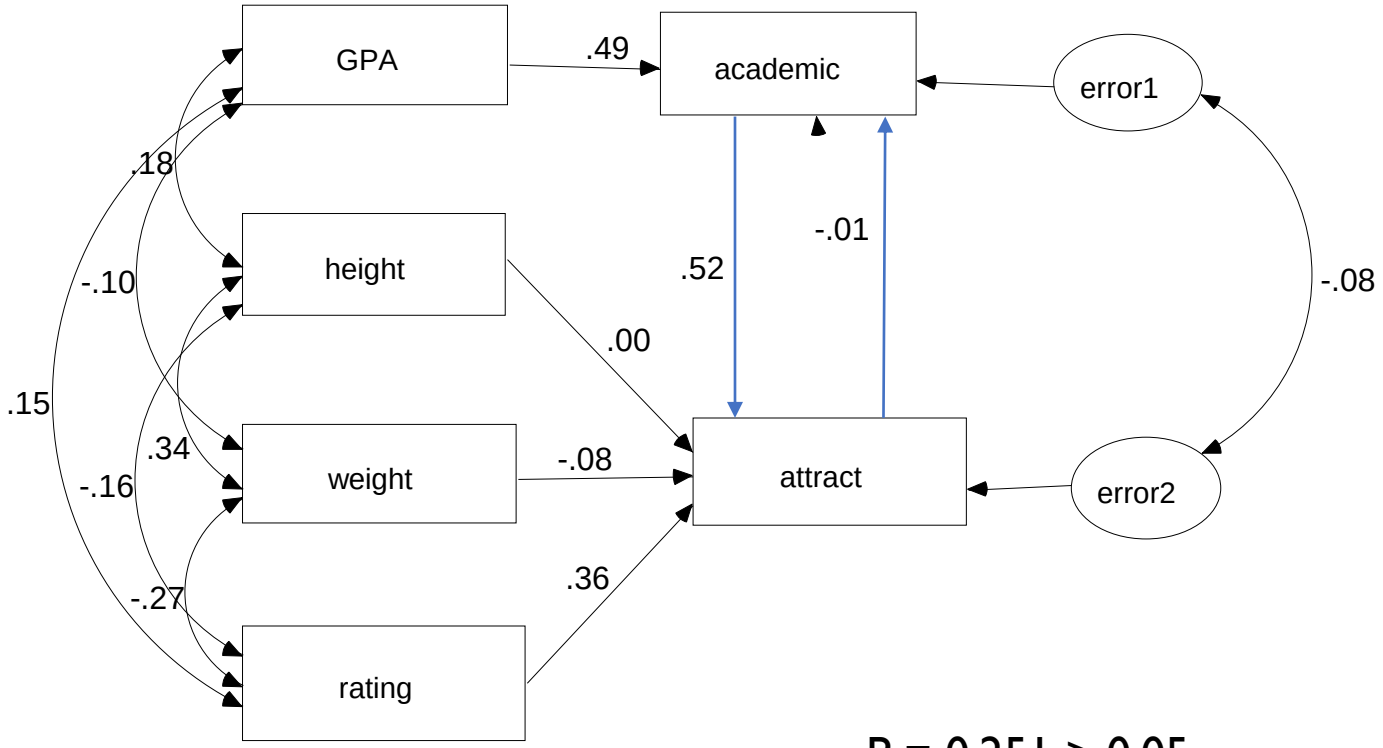


Figure 7.5. Relationship between herbaceous richness and various factors following fire in southern California chaparral (from Grace and Keeley 2006). This model includes an explicit nonlinearity between Plant Cover and Richness, along with a composite variable summarizing the relationship called Optimum Cover.

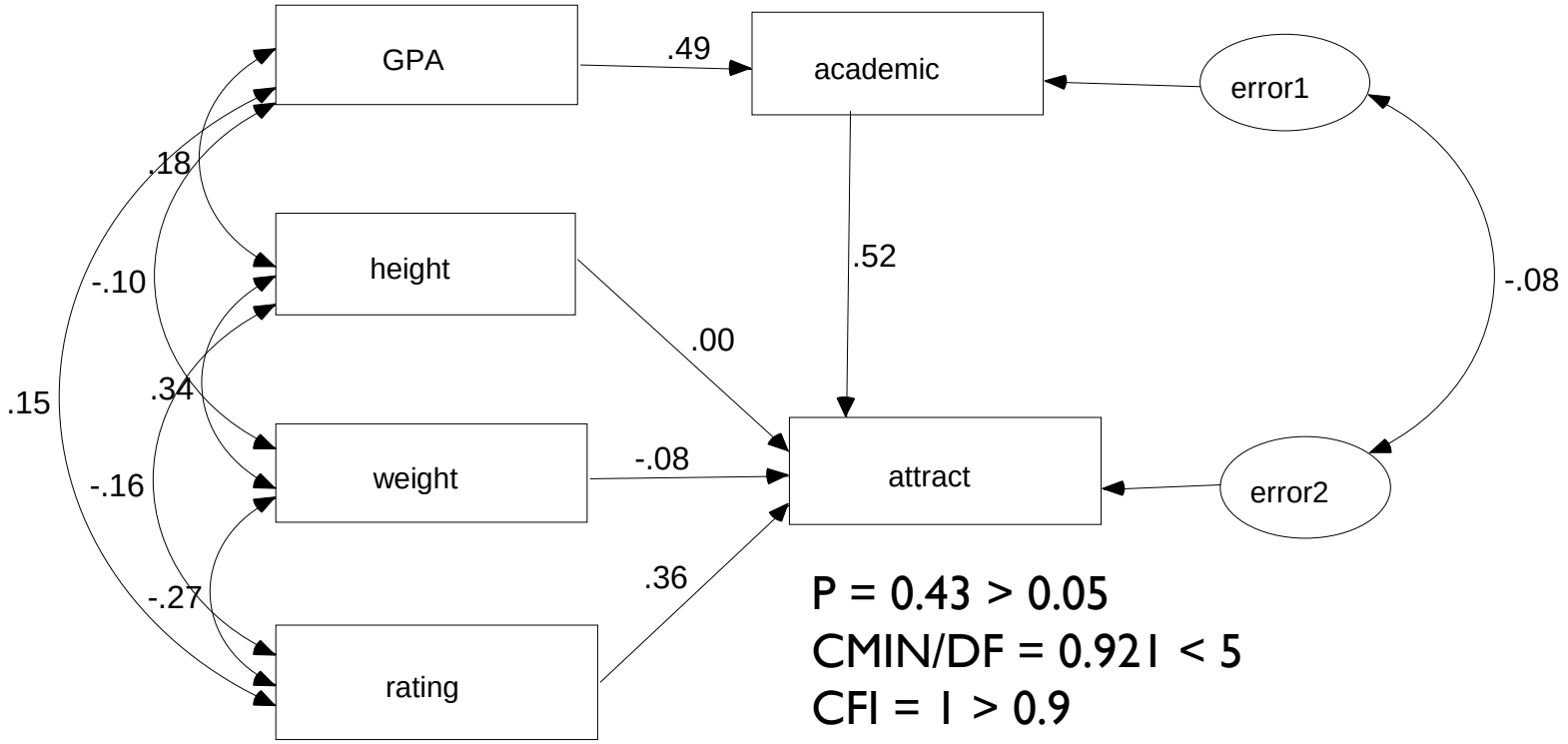
Grace. 2006

Model A



P = 0.251 > 0.05
CMIN/DF = 1.381 < 5
CFI = 0.996 > 0.9
NFI = 0.988 > 0.8
RMSEA = 0.043 < 0.06
Conclusion: Model is accepted

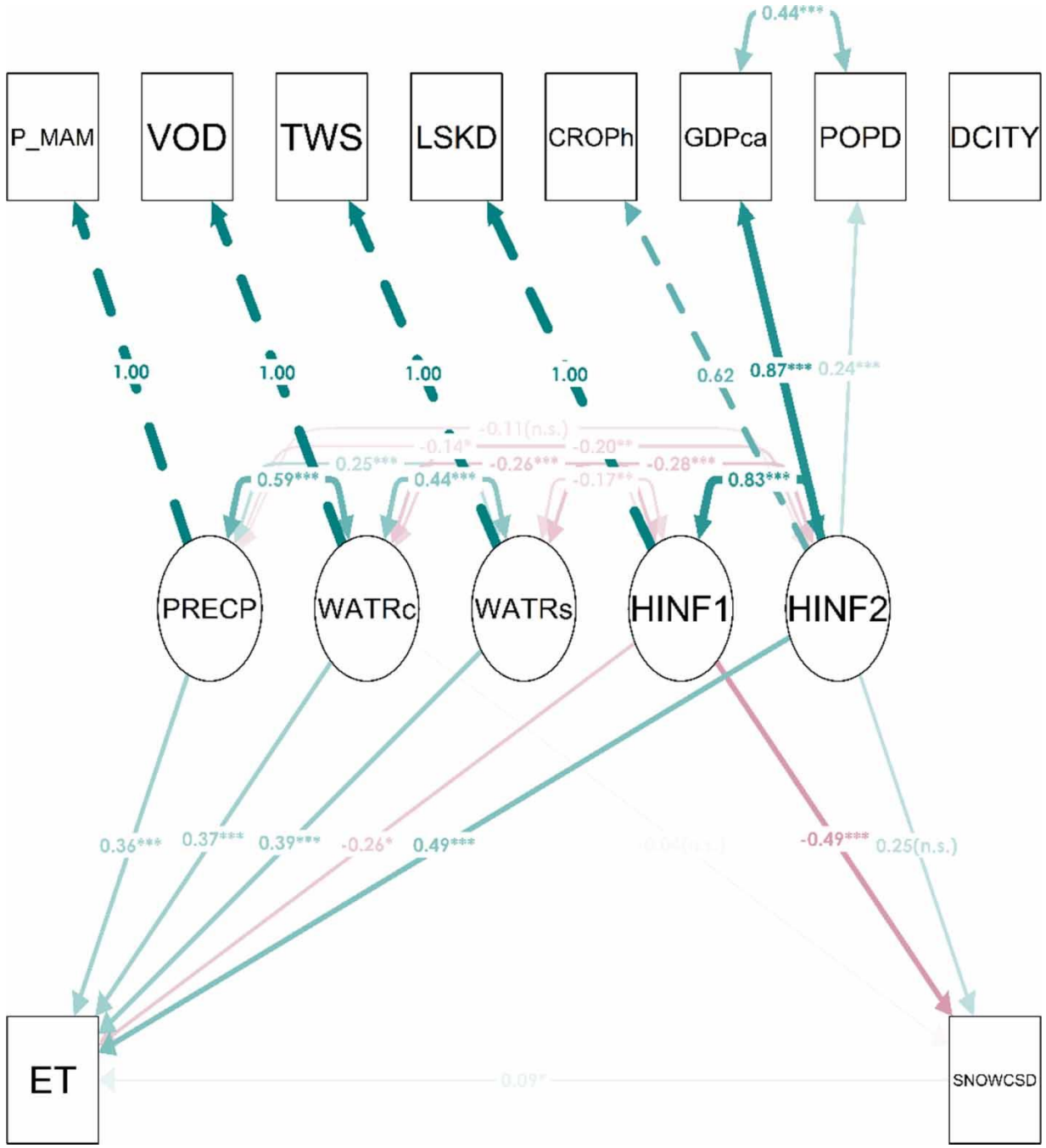
Model B: Respecification



P = 0.43 > 0.05
CMIN/DF = 0.921 < 5
CFI = 1 > 0.9
NFI = 0.988 > 0.8
RMSEA = 0.000 < 0.06
Conclusion: Model is accepted

Model B is better than Model A

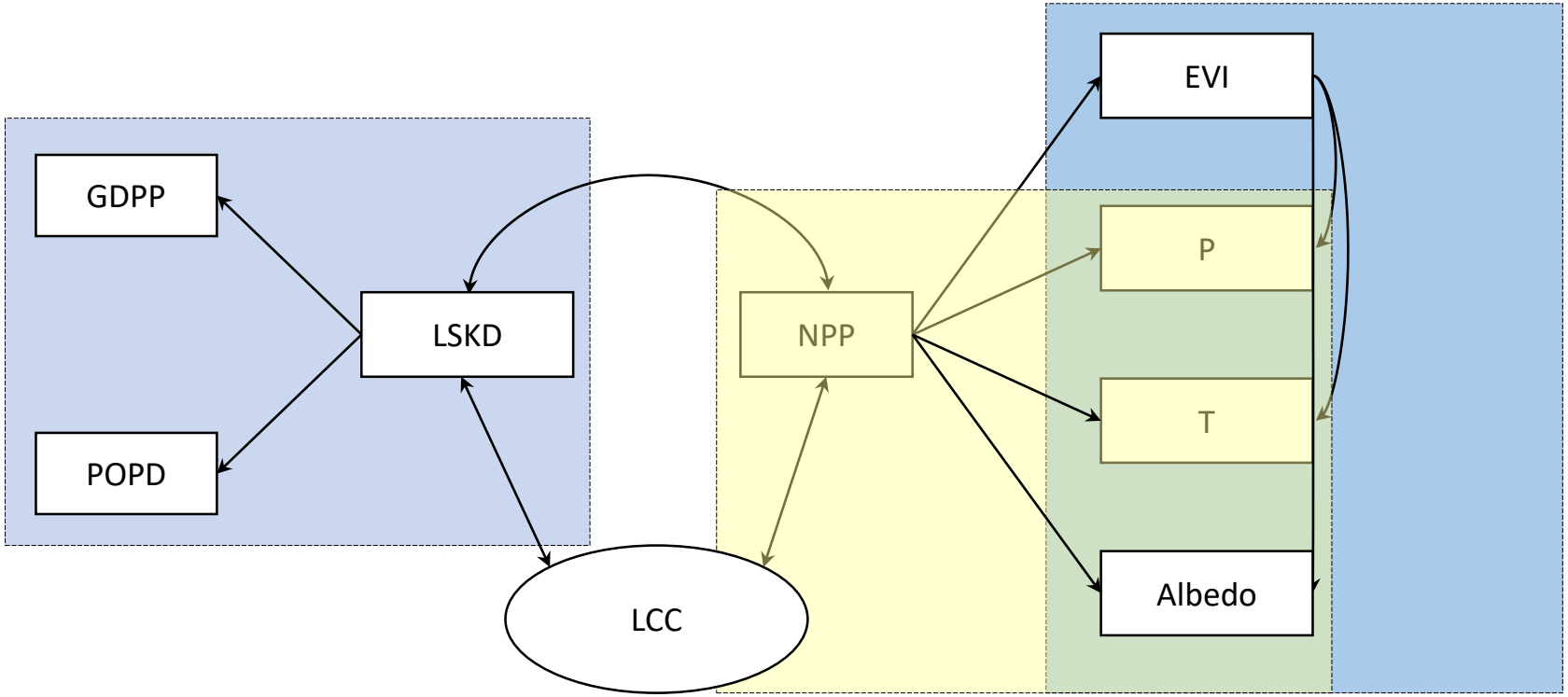
Figure 7. Structural equation modeling examining percent snow cover variability ($SNOWc_{SD}$) as a moderator between the evapotranspiration (ET) and precipitation (PRECP), water content (WATRc), water storage (WATRs), human influence—1 (HINF1) and human influence—2 (HINF2) as constructs (i.e. latent variables). Model fit—chi-square (χ^2; degrees of freedom = 19) = 55.26, comparative fit index = 0.96; Tucker–Lewis index = 0.90; standardized root mean square residual = 0.07. All parameter estimates are standardized (full forms in appendix)



Venkatesh et al. 2022

Hypothesis:

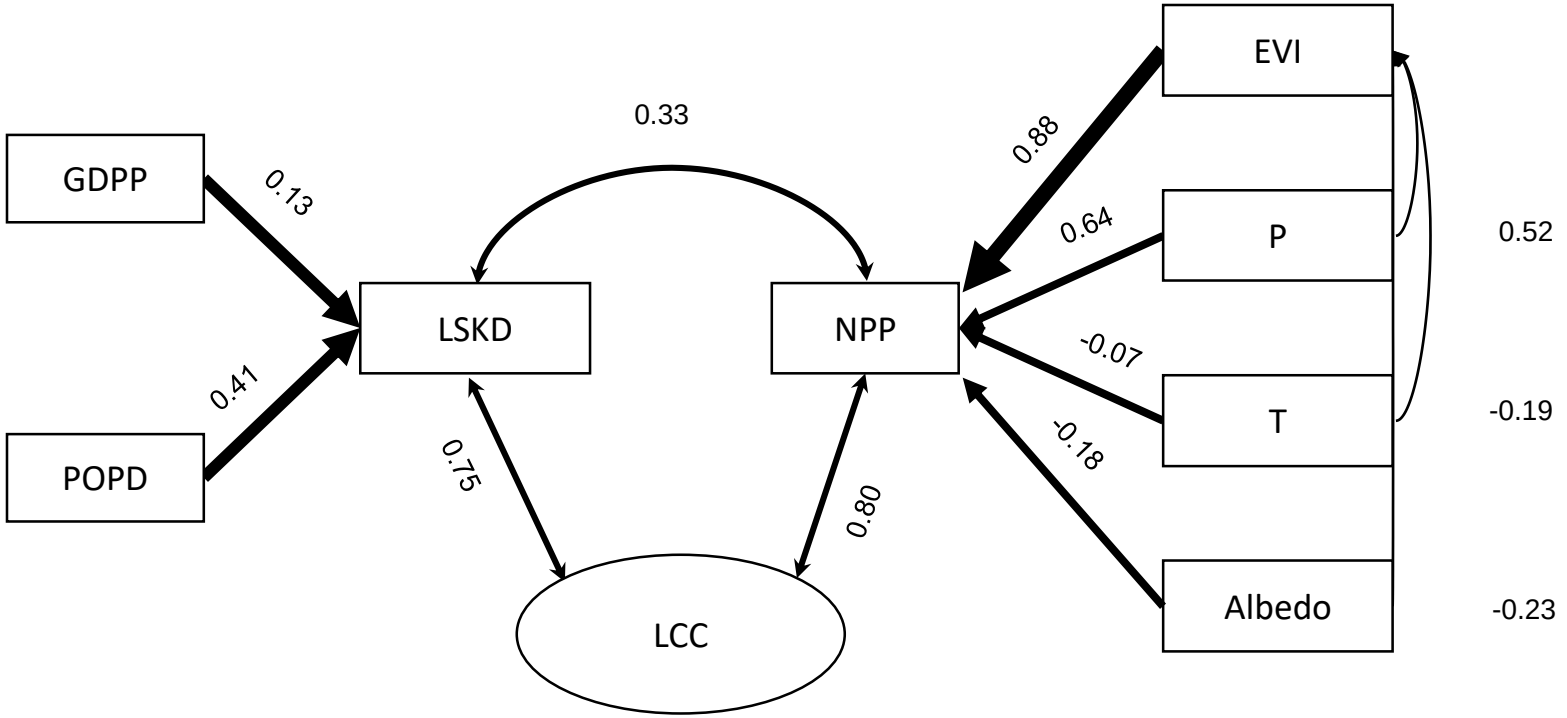
Structural Equation Modeling (SEM)



Chen et al. ERL, 2015

Mongolia Plateau

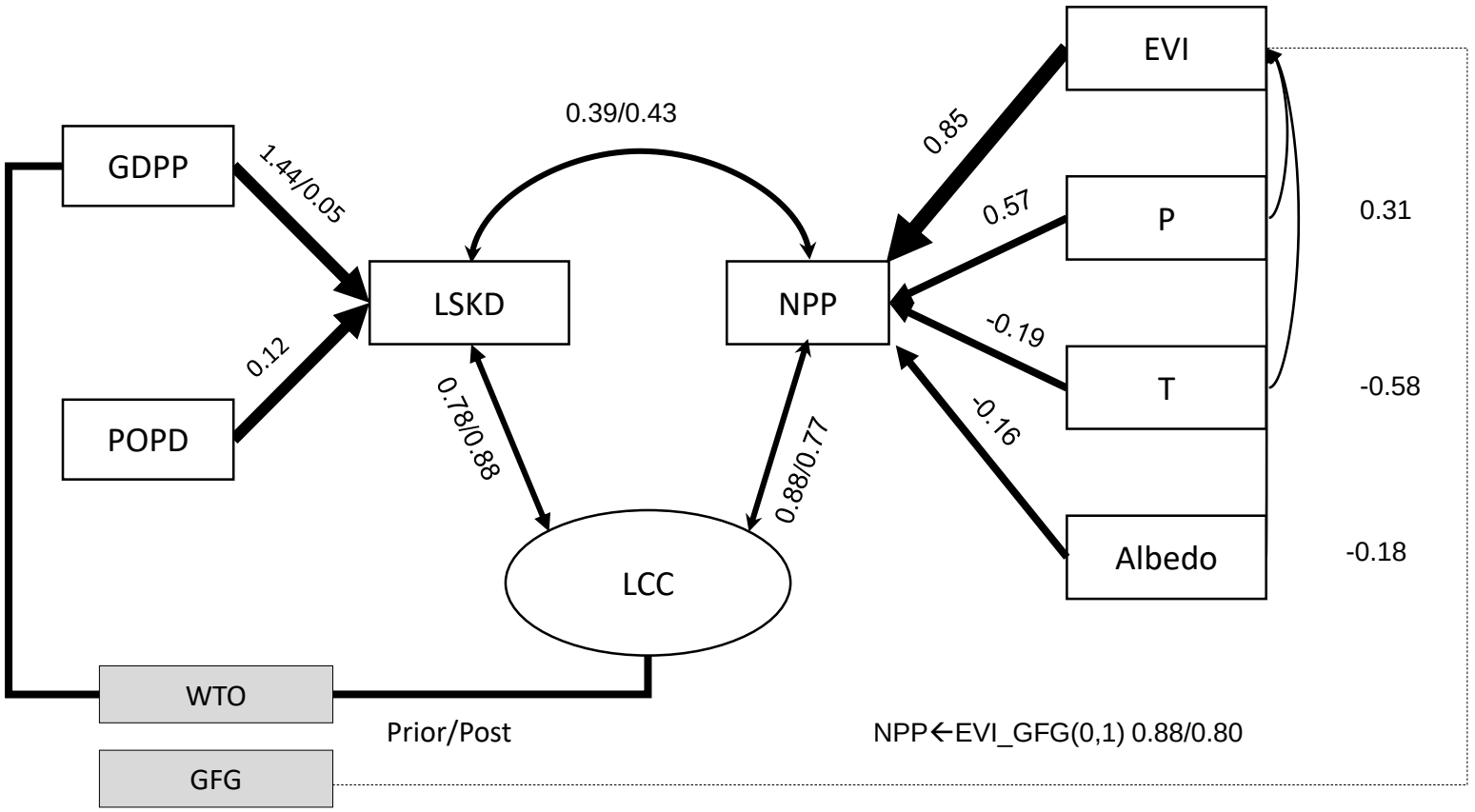
The Structural Equation Modeling of the CNH system



Chen et al. ERL, 2015

Mongolia Plateau

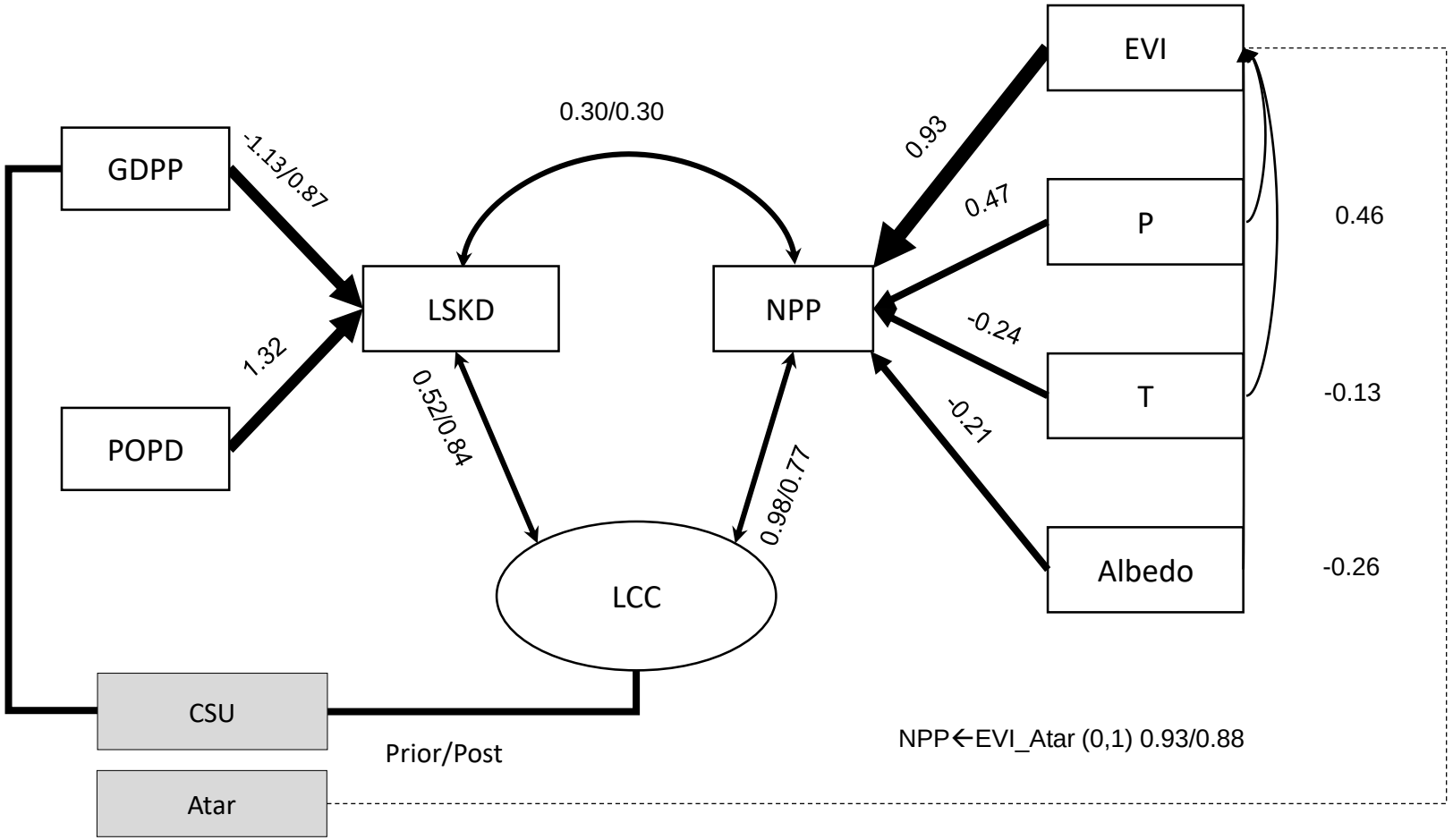
The Structural Equation Modeling of the CNH system



Chen et al. ERL, 2015

Mongolia Plateau

The Structural Equation Modeling of the CNH system



Chen et al. ERL, 2015

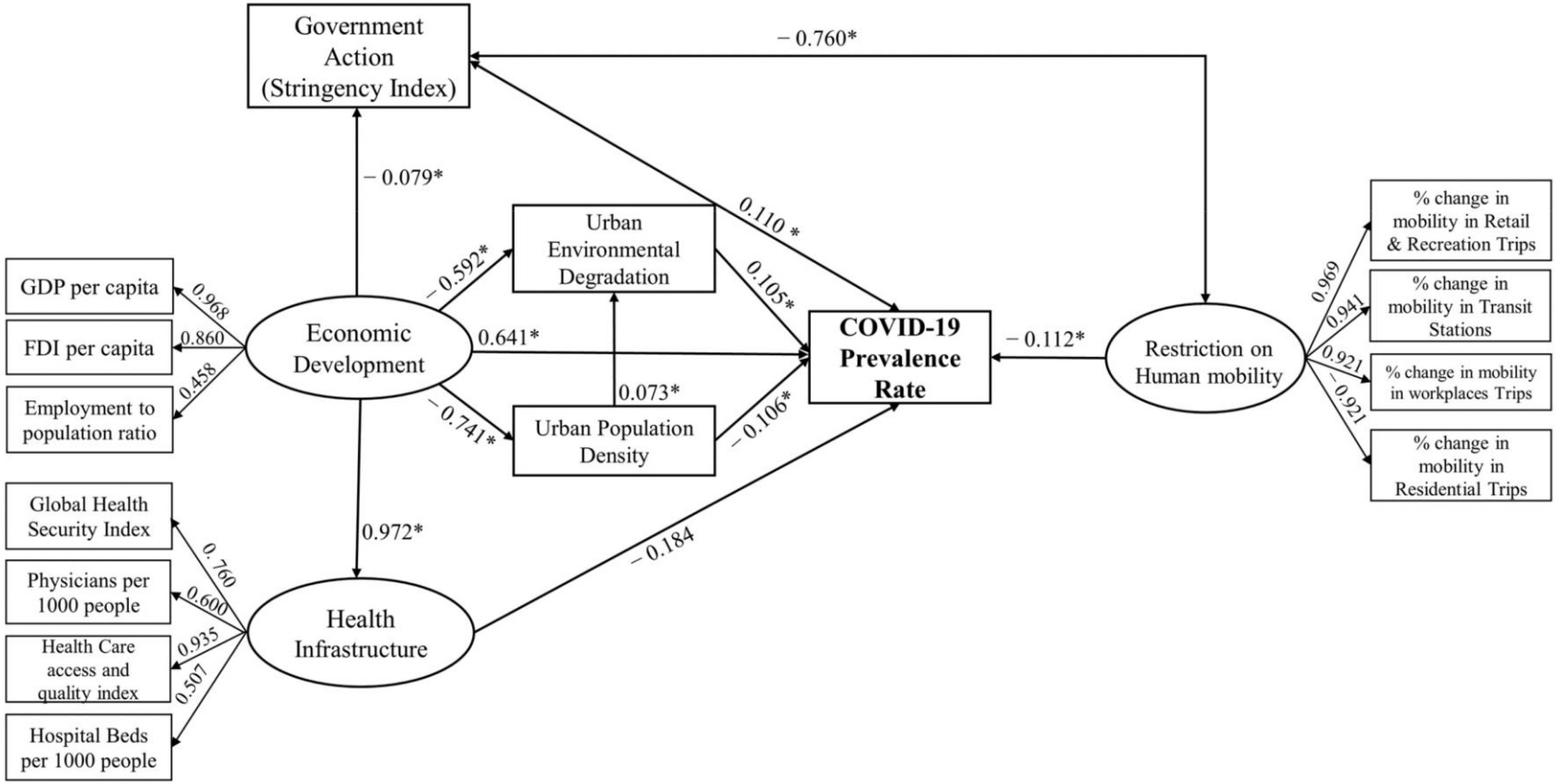


Figure 3. Empirical influences of major human system factors (economic, social, policy, health infrastructure, and urban environment status) on the prevalence rate (PR) of COVID-19 for the 151 countries from the 20-week study period. Observed variables are presented by boxes, latent variables by ovals and the standardized path coefficients and factor loads are listed next to arrows in the PLS-SEM model. The PR of COVID-19 was particularly related to economic development level, health infrastructure, and policies regarding restrictions on human mobility, but less associated with urban environment and urban population density. Fan et al. 2022.

Fig. 5. Partial Least Squares Structural Equation Modeling (PLS-SEM) of socioeconomic and biophysical drivers on urbanization in both Inner Mongolia (IM) and Mongolia (MG). Latent variables are circular shapes, and measured variables are squares. The path coefficients describe the relationship between variables. The IM model illustrates that the economy is a major driver of urbanization (R^2 = 0.422) whereas the MG model demonstrates that both economy and social goods drive urbanization (R^2 = 0.342).

Park et al. 2017

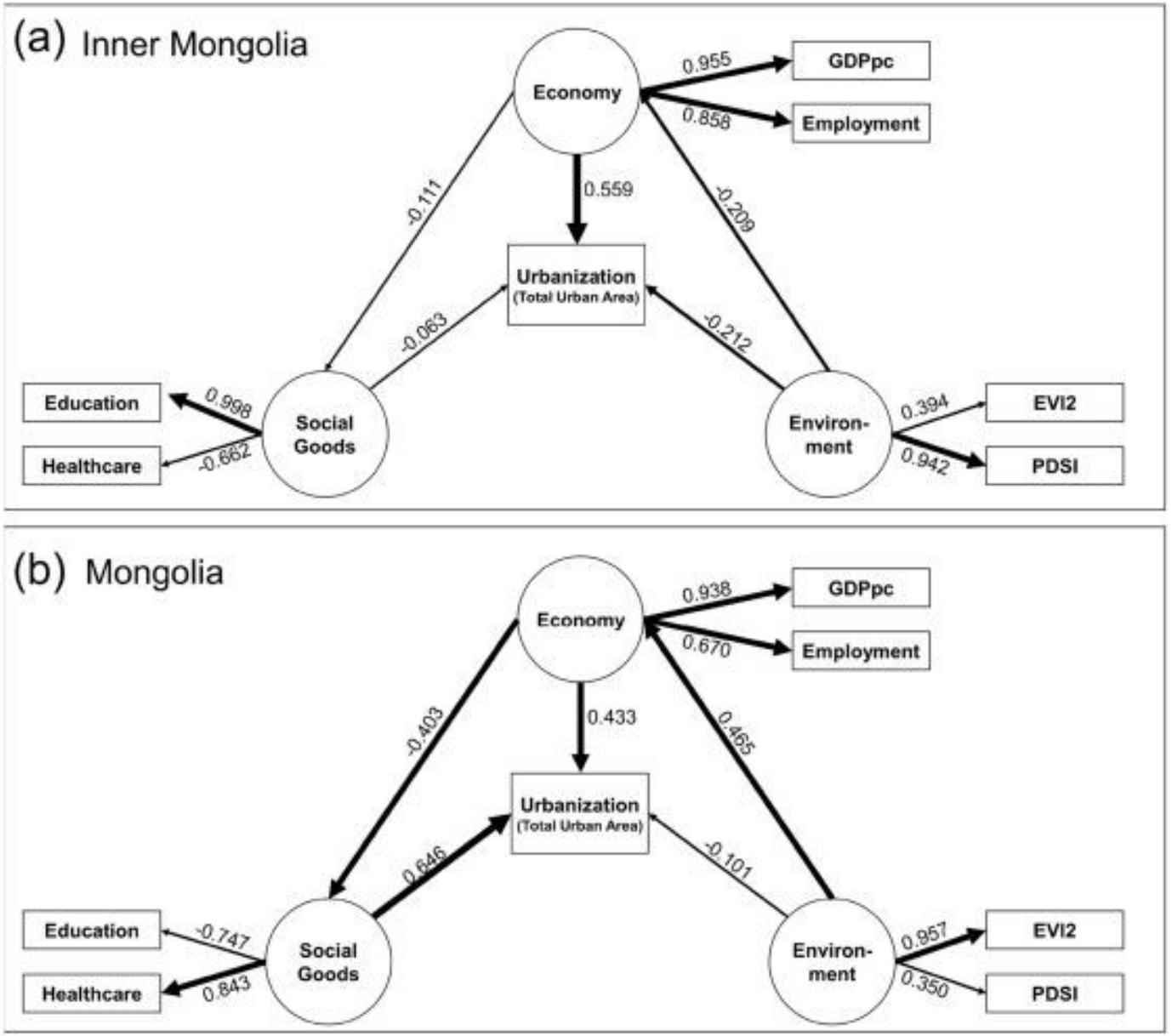
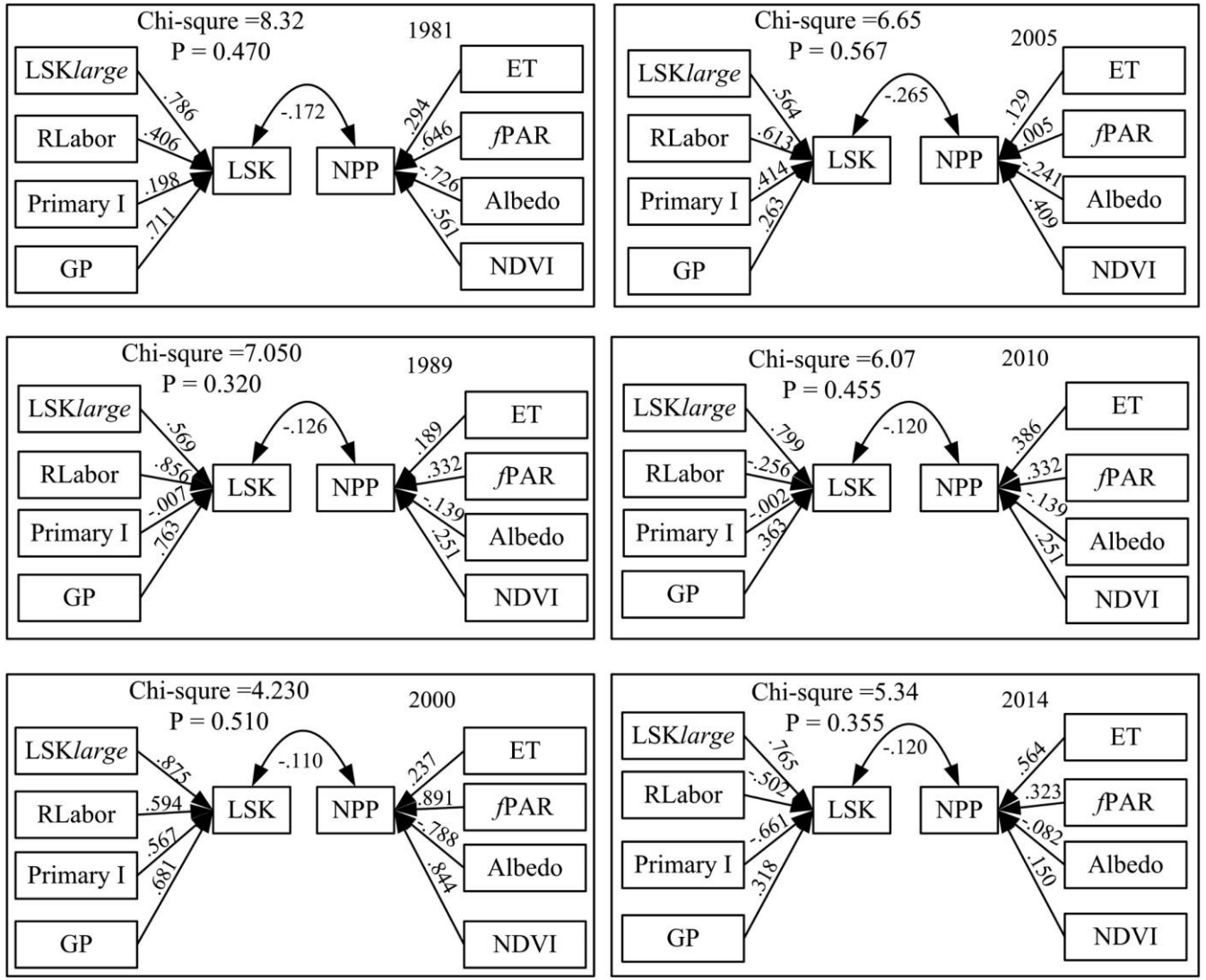


Figure 7. Dynamics of structural relationships based on structural equation modeling (SEM) for coupled changes of socioeconomic and environmental variables for the six time periods (1981, 1989, 2000, 2005, 2010, 2014) on the Tibetan Plateau. A one-way arrow indicates a hypothesized causal relationship between the two variables, while a two-way arrow indicates a feedback relationship. Absence of a line between any two variables implies that no hypothesis was proposed in this study. Livestock (LSK) is hypothesized to be influenced by LSK_{large}, R_{Labor}, Primary I, and GP, while ecosystem net primary productivity (NPP) is related to NDVI, *f*PAR, ET and Albedo (see figure 3). The residuals of LSK and NPP were assessed by the model. The partial regression coefficients indicated the strength and direction of these relationships.



Tian et al., 2018. Coupled dynamics of socioeconomic and environmental systems in Tibet. ERL.