Using Structural Equation Modeling to Model Pedestrian’s Injury Severity Level

Zhi Chen, PhD Candidate

Xiao Qin, Ph.D., PE

Robert James Schneider, Ph.D.

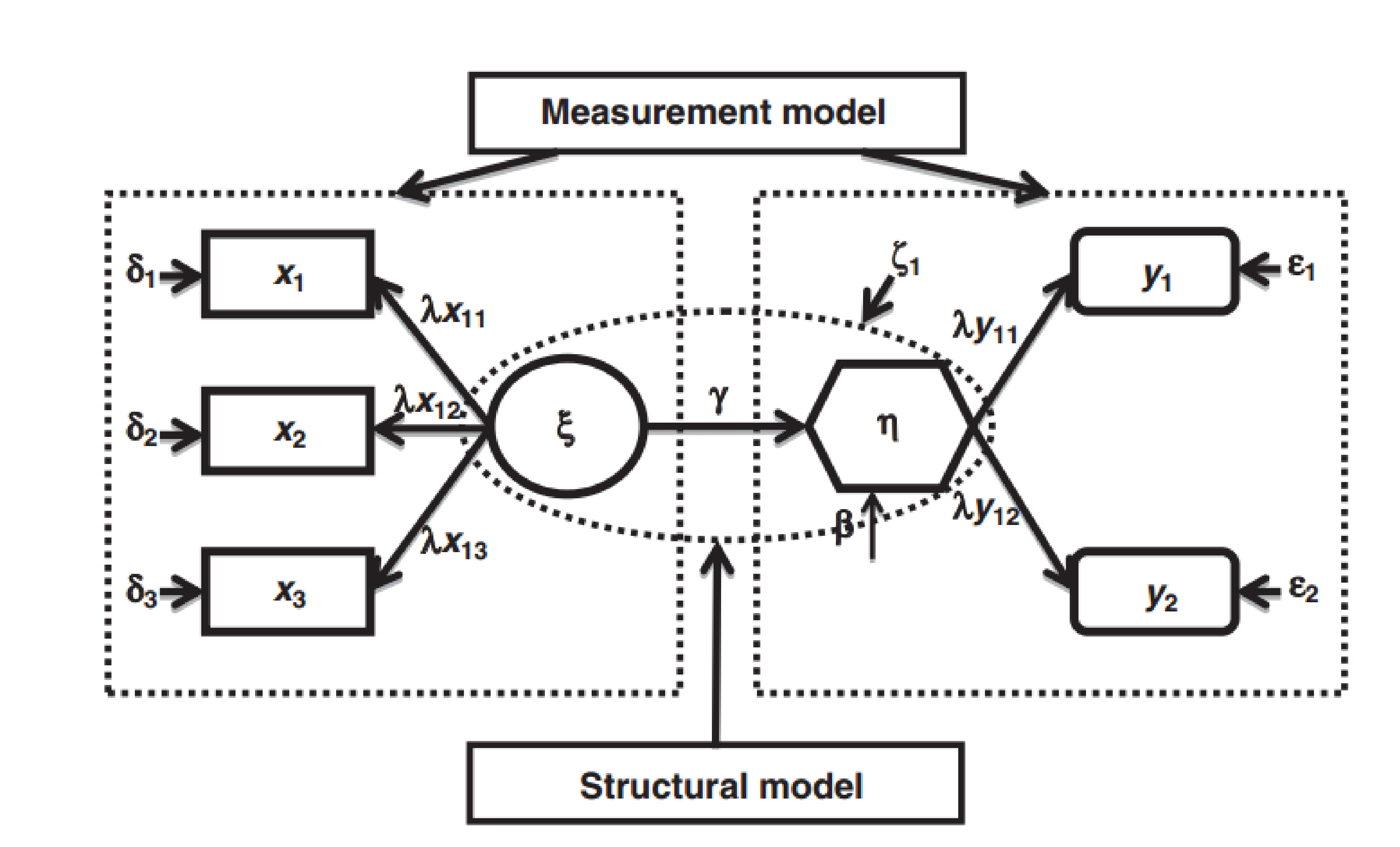
University of Wisconsin-Milwaukee

Research Questions

- How can we incorporate the knowledge of crash mechanism into safety modeling and perform value-added analysis?
- How can we investigate the intrinsic and often indirect relationships between contributing factors and crash outcomes?

Methods

Methodology: Structural Equation Modeling (SEM) with the assumption that each path represents a linear relationship.



Data Processing

- **Crash Data**: **1,480** intersection-related crashes between vehicles and pedestrians in Wisconsin from 2011 to 2013
- **Data Category**
  - Environment/Context: lighting, weather, road surface, intersection control
  - Driver: age, gender, impairment conditions, driving action (e.g. going straight, turning left/right)
  - Pedestrian: age, gender
  - Crash: pedestrian injury severity, vehicle damage
- **Data Conversion**
  - Categorical → dummy variables

    Daylight, dawn/dusk, lighting and dark, and with daylight condition as the base, three dummy variables were created to indicate three other conditions having difference influence on crash outcome
  - Continuous → categorical → dummy variables

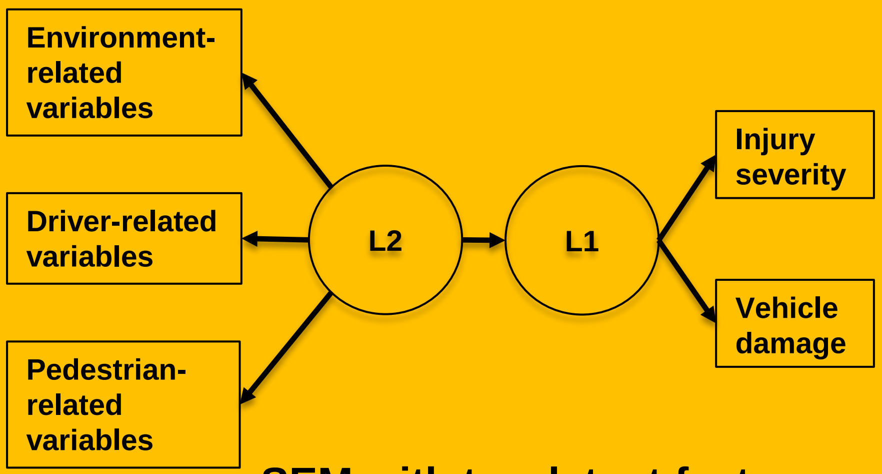
    pedestrian age was first converted into children, youth, middle-aged, and elderly, and then converted into dummy variables with middle-aged as the base, assuming that the trend is different between groups but consistent within the group.
  - Categorical → ordinal variable

    injury severity in KABCO scale is coded into ordinal rankings, 1,2,3,4,5 with 1 indicating PDO (K) and 5 indicating fatal (O) with the assumption that the ordinal variable represents the extent of crash-related variables

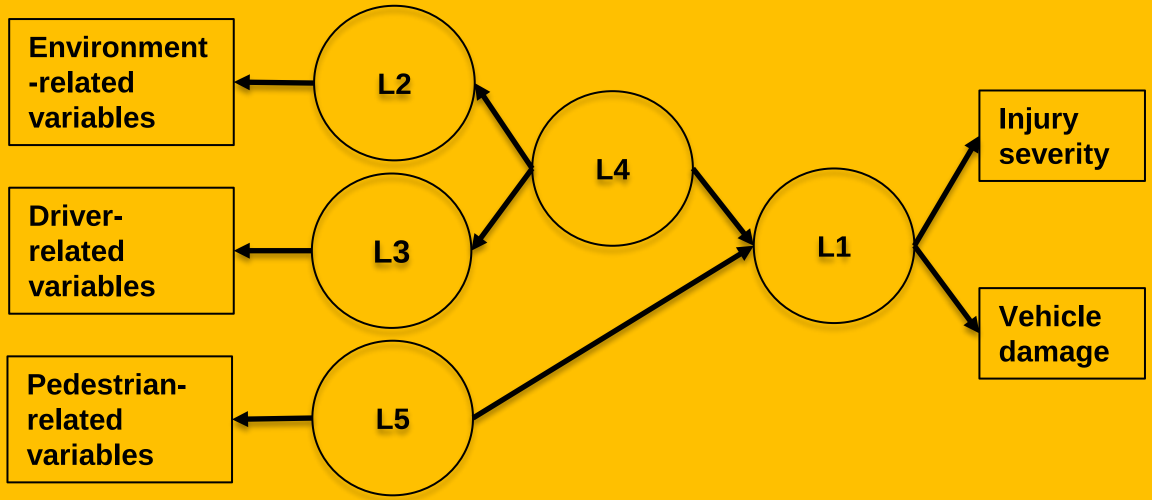
Model Structure

Crash Mechanism

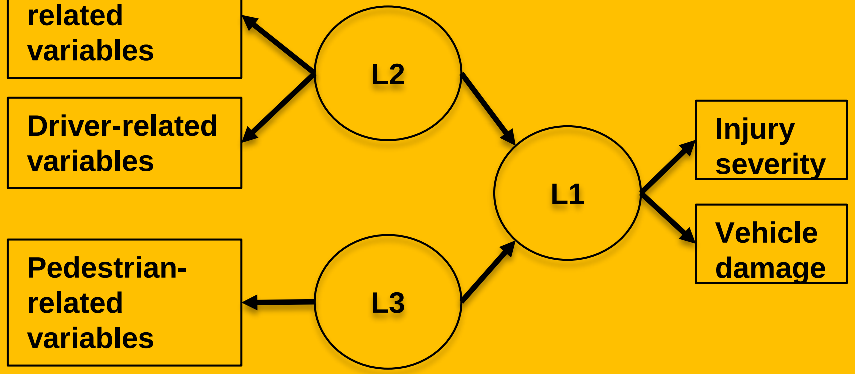
The collision impact between the vehicle and pedestrian and the pedestrian’s conditions lead to the crash outcome such as the injury severity and vehicle damage. The impact force is largely determined by vehicle speed and mass.



SEM with two latent factors



SEM with five latent factors

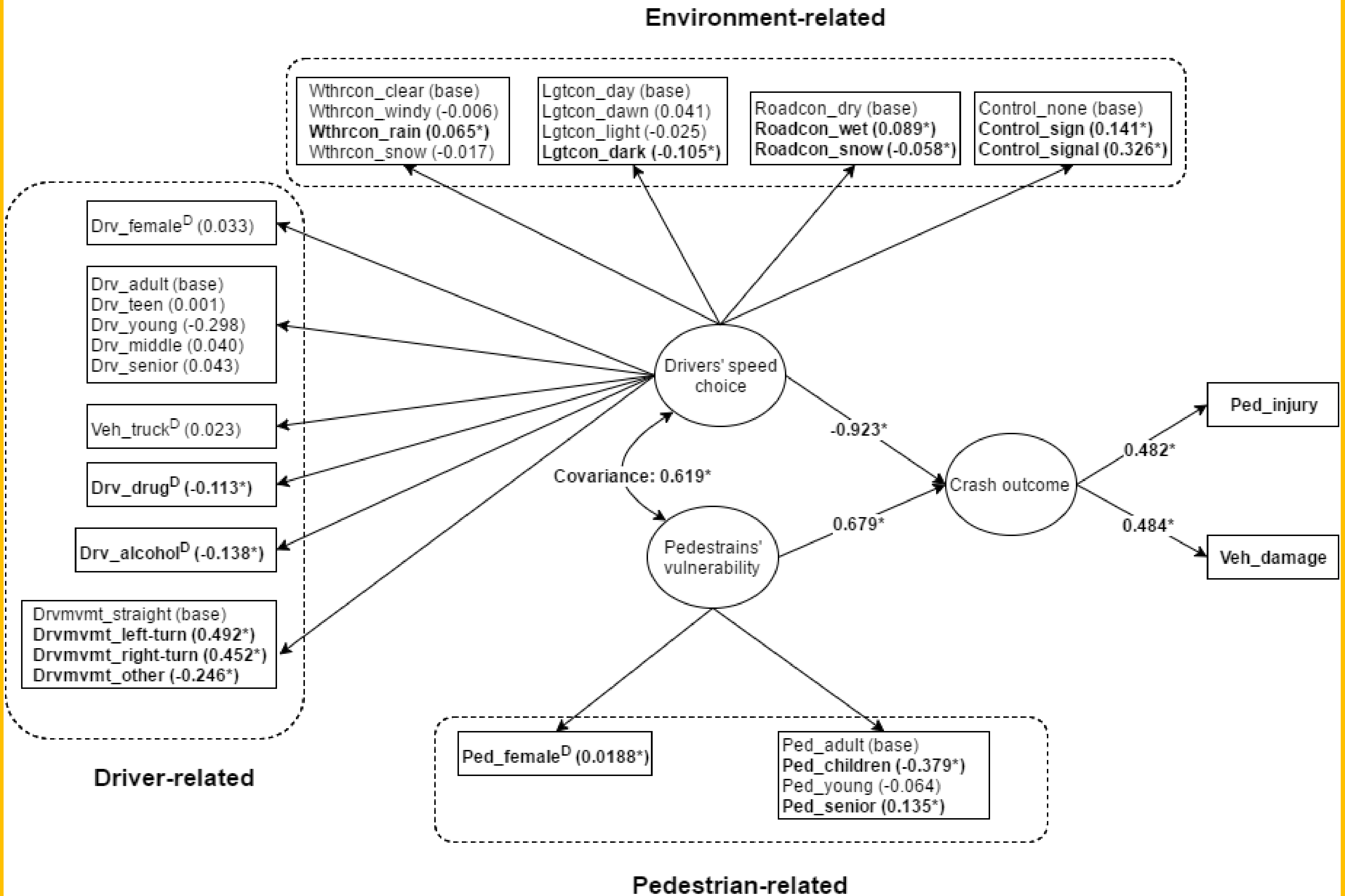


SEM with three latent factors

Findings

- Both SEM models with two and three latent factors converged, while the model with five latent factors failed to converge after many attempts;
- The SEM model with three latent factors has better goodness-of-fit than the one with two latent factors, suggesting the three latent factors structure may be better at explaining the underlying crash mechanism.
- The injury severity of pedestrian and vehicle damage can be used to measure the actual crash outcome which is a latent variable modeled by other latent variables.
- Driver’s speed choice can be measured by environmental factors, driver characteristics, driver’s conditions (drunk/impaired); pedestrian’s vulnerability is measured by the pedestrian characteristics and conditions;
- Driver’s poor speed choice and increased pedestrian’s vulnerability will cause more severe pedestrian injuries.

Findings (cont.)



Note:
D: indicates a dummy flag variable (e.g. Drv_female is 1 if the driver is female, and is 0 otherwise);
* indicates the estimate is significant at 5% significance level.

Future Research Needs

- More investigation is needed for the underlying crash mechanism of vehicle-pedestrian crashes and appropriate model structures that are close to the real crash mechanism;
- In linear SEMs, observed variables are assumed to be continuous, but they are discrete in nature, so generalized SEMs should be developed to capture their discreteness.
- The recommendation on the sample size is 10 to 20 times of variables, so for complicated models with many variables, a large sample size is desirable.
- Design studies to validate the latent variables and model structures.

Implications for Practice

- The SEM empowers the design of safety studies by imposing a structure between observations.
  - Convert an input-output model to a variable structure (sequential, parallel, etc.)
  - Incorporate the engineering knowledge about crash mechanisms into the modeling process.
- Guide the safety data collection and analysis.
- Better data, better structure, and better information for better decision and policy-making.
  - Improve intersection design and enforcement policies to assist drivers in making better decisions when approaching intersections.

Reference

[1] Wang, K., and X. Qin. Use of structural equation modeling to measure severity of single-vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2432, 2014, pp. 17-25.

[2] Kim, K., P. Pant, and E. Yamashita. Measuring influence of accessibility on accident severity with structural equation modeling. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2236, 2011, pp. 1-10.

[3] Lee, C., and M. Abdel-Aty. Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. *Accident Analysis & Prevention,* Vol. 37, No. 4, 2005, pp. 775-786.

[4] *Structural Equation Modeling.* Stata Press. http://www.stata.com/manuals14/sem.pdf.

[5] Acock, A. C. Discovering structural equation modeling using Stata. *Stata Press books*, 2013.

[6] Kline, R. B. *Principles and practice of structural equation modeling*. Guilford publications, 2015.

[7] Bollen, K. A. *Structural equations with latent variables*. John Wiley & Sons, 2014.

Summary Statistics

Variable	Description	Frequency	Percentage
Environment Factors			
Wthrcond	Weather condition		
Clear	Clear	860	58.1
Windy	Windy or cloudy	382	25.8
Rain	Rain	183	12.4
Snow	Snow or sleet	55	3.7
Lgtcond	Lighting condition		
Day	Daylight	951	64.3
Dawn	Dawn or dusk	47	3.2
Light	Night with street light	449	30.3
Dark	Night without street light	33	2.2
Roadcond	Road surface condition		
Dry	Dry	1116	75.4
Wet	Wet or muddy	292	19.7
Snow	Snow or sleet	72	4.9
Control	Intersection control		
None	None	436	29.5
Sign	Stop/Yield sign	341	23.0
Singal	Singal	703	47.5

Variable	Description	Frequency	Percentage
Driver Characteristics			
Drv_female	Female driver flag		
0	Male	818	55.3
1	Female	662	44.7
Age	Driver age		
Teen	<20	107	7.2
Young	20-24	189	12.8
Adult	25-44	526	35.5
Middle	45-64	457	30.9
Senior	>64	201	13.6
Veh_truck	Truck flag		
0	Passenger car	1250	84.5
1	Truck or bus	230	15.5
Drv_drug	Drug flag		
0	No	1459	98.2
1	Yes	21	1.8
Drv_alcohol	Alcohol flag		
0	No	1382	93.4
1	Yes	98	6.6
Drvmvmt	Driver movement		
Straight	Going straight	562	37.9
Left-turn	Turning left	534	36.1
Right-turn	turning right	321	21.7
Other	Other movements	63	4.3

Variable	Description	Frequency	Percentage
Pedestrian Characteristics			
Ped_female	Female pedestrian flag		
0	Male	762	51.5
1	Female	718	48.5
Age	Pedestrian age		
Children	<14	240	16.2
Young	15-24	364	24.6
Adult	25-64	701	47.4
Senior	>64	175	11.8
Crash Outcome			
Ped_injury	Injury severity		
1	O	52	3.5
2	C	562	37.9
3	B	629	42.5
4	A	194	13.1
5	K	43	2.9
Veh_damage	Vehicle damage		
1	None or minor	1357	91.7
2	Moderate	114	7.7
3	Severe	9	0.6

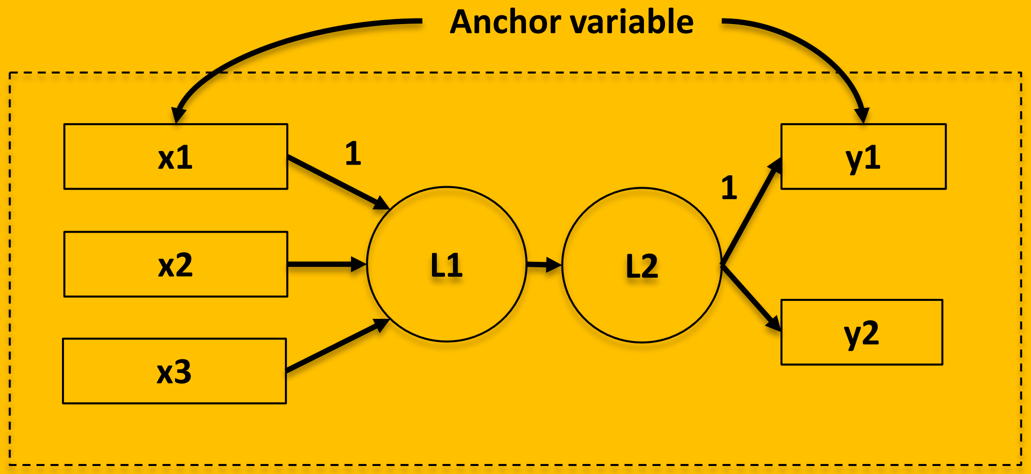
Findings (cont.)

- Compared to daylight, dark condition leads to significantly worse speed choice;
- Compared to dry road surface, wet surface leads to significantly better speed choice, and snowy/sleety surface has significantly negative influence;
- Drivers’ age and gender don’t affect their speed choice;
- Compared to intersections without signs or signals, drivers take better speed choice at intersections with signs or signals, and their speed choice tend to be better at signalized intersections than at those with stop/yield signs;
- Compared to going straight, drivers tend to take better speed choice when turning left or right;
- Drug and alcohol will worsen drivers’ speed choice;
- Compared to middle-aged pedestrians, elderly ones have higher vulnerability while children have lower vulnerability;
- Compared to male pedestrians, females tend to be more vulnerable.

Non-convergence Issue

- Why
  - The optimization procedure is stuck in the local optimum rather than the global optimum.
- Causes
  - Poor choice of anchor variables* or potential correlations between observed variables ;
  - Improper model structure.

* The latent variables are unobserved and hence their scales are unknown and it makes the SEM model unidentified. One common solution is to force the factor loading of one variable, the anchor variable, to be 1.



- Solutions
  - Conduct the factor analysis for observed variables linked to the same latent variable, and select the one with the maximum factor loading as the anchor variable;
  - Add covariance constraints on pairs of variables that show strong correlations;
  - Build simple and converged SEM models with proper structures first, and add more variables one by one to test the convergence of the extended model. If the extended model still converges, its structure should be still proper.