Modeling the Distillation Process in SEM

Laura Castro-Schilo, Ph.D.

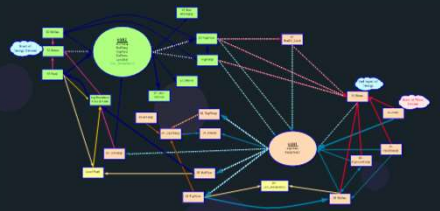
Sr. Research Statistician Developer

JMP Division, SAS Institute

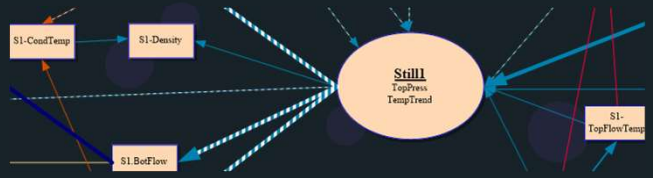
jmp STATISTICAL DISCOVERY FROM SAS

Key Features of Project

1. Clear theory of how processes affect each other



2. Theory stipulates variables have dual roles: inputs (X) and outputs (Y)



3. Each node in diagram is a process measured repeatedly –aka time series

Why Structural Equation Modeling?

General Framework

- Model variances, covariances, and means among variables
- Test theories of multivariate relations among variables
- Test direct and indirect effects
- Account for measurement error
- Specify and model latent (unobserved) variables
- Use cutting edge algorithms for missing data
- Intuitive path diagrams represent statistical models

Structural Equation Modeling

General Framework

- Path diagrams convey statistical models intuitively
  - E.g., Simple linear regression:

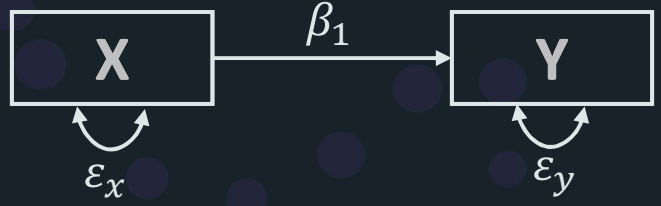
Structural Equation Modeling

General Framework

- Path diagrams convey statistical models intuitively
  - E.g., Simple linear regression:

$$Y_i = \beta_1 X_i + \varepsilon_{yi}$$
$$X_i = \varepsilon_{xi}$$

** means/intercepts omitted for simplicity*



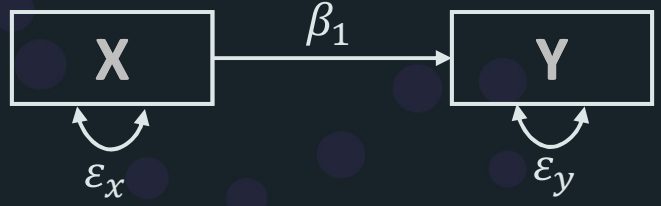
Structural Equation Modeling

General Framework

- Path diagrams convey statistical models intuitively
  - E.g., Simple linear regression:

$$Y_i = \beta_1 X_i + \varepsilon_{yi}$$
$$X_i = \varepsilon_{xi}$$

** means/intercepts omitted for simplicity*



- SEMs imply a covariance structure:

	X	***Y***
X	ε_x	
Y	$\beta_1\varepsilon_x$	$\beta_1^2\varepsilon_x + \varepsilon_y$

- Fit assessed comparing model-implied covariance vs. sample covariance

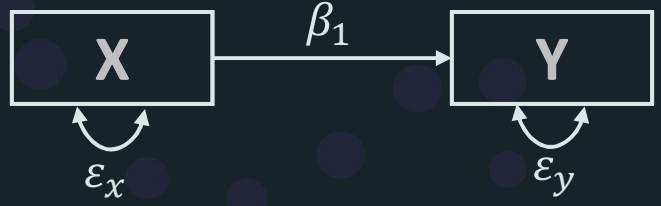
Structural Equation Modeling

General Framework

- Path diagrams convey statistical models intuitively
  - E.g., Simple linear regression:

$$Y_i = \beta_1 X_i + \varepsilon_{yi}$$
$$X_i = \varepsilon_{xi}$$

** means/intercepts omitted for simplicity*



- SEMs imply a covariance structure:

	X	Y
X	ε_x	
Y	$\beta_1\varepsilon_x$	$\beta_1^2\varepsilon_x + \varepsilon_y$

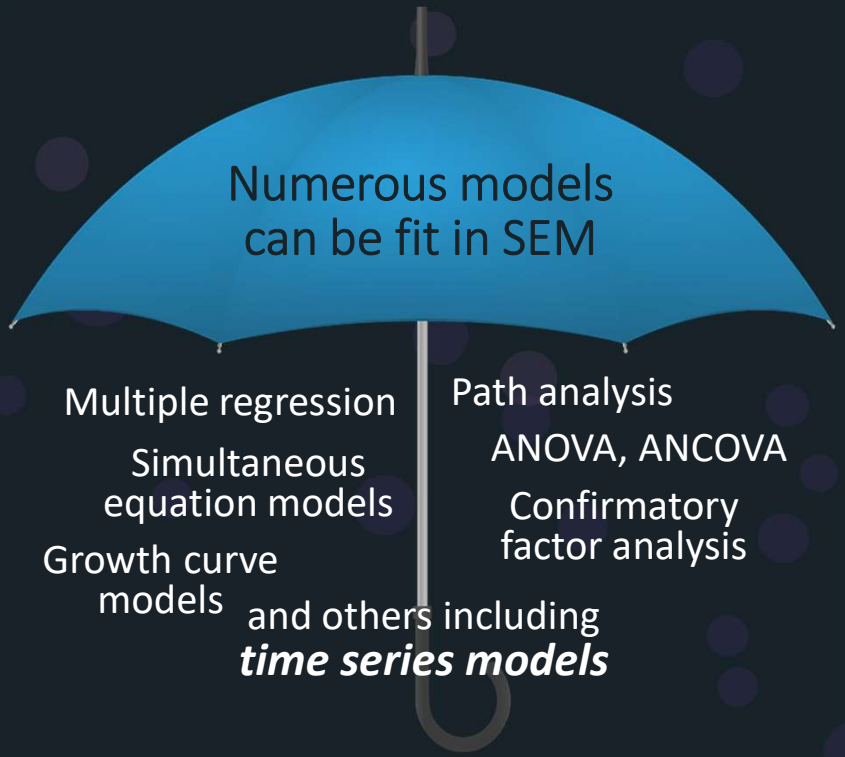
Compare with sample

	X	Y
X	1.40	
Y	0.76	1.14

- Fit assessed comparing model-implied covariance vs. sample covariance

Structural Equation Modeling

General Framework



Time Series Analysis

ARIMA Models (Box & Jenkins, 1970)

- Time Series:
  - Collection of data where there is dependence on previous data points
  - Equally spaced time intervals
  - Account for dependence by regressing on the past
- Autoregressive, AR(p), processes:

$$Y_t = \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \varepsilon_{yt}$$

- AR(1) process:

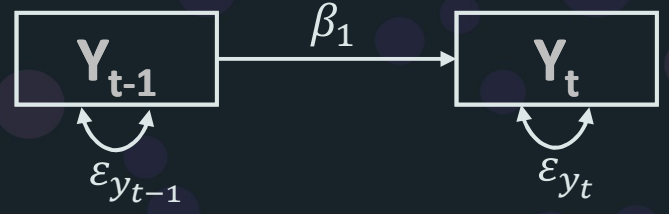
$$Y_t = \beta_1 Y_{t-1} + \varepsilon_{yt}$$

Time Series Analysis

Implemented in SEM

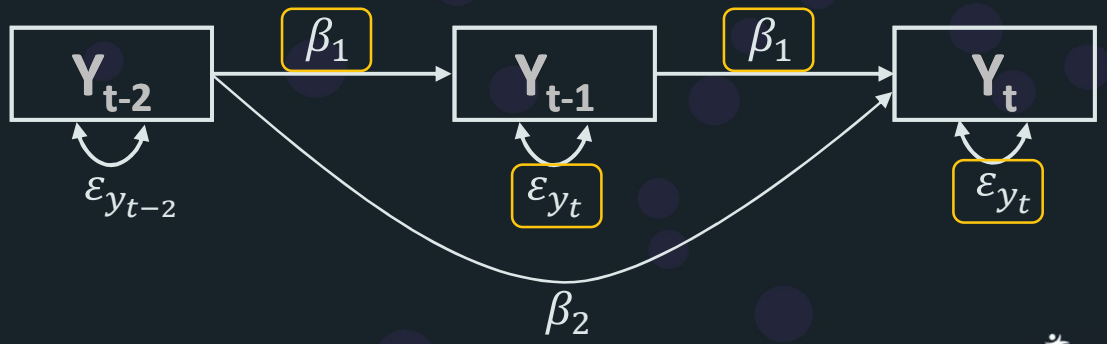
- AR(1) process:

$$Y_t = \beta_1 Y_{t-1} + \varepsilon_{y_t}$$



- AR(2) process:

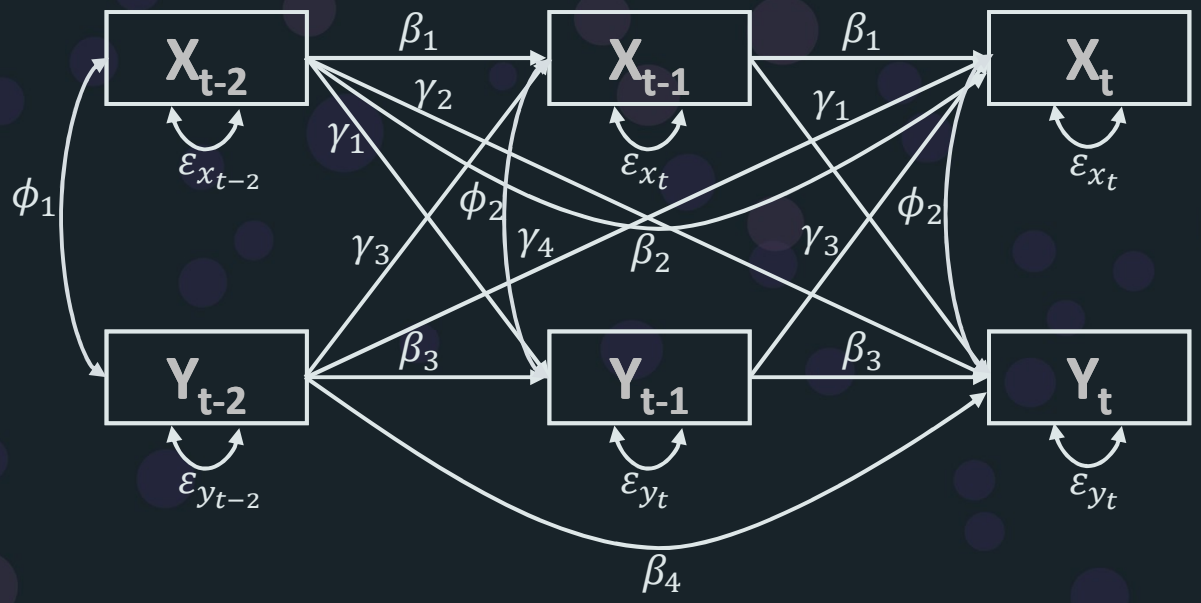
$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \varepsilon_{y_t}$$



Multivariate Time Series Analysis

Implemented in SEM

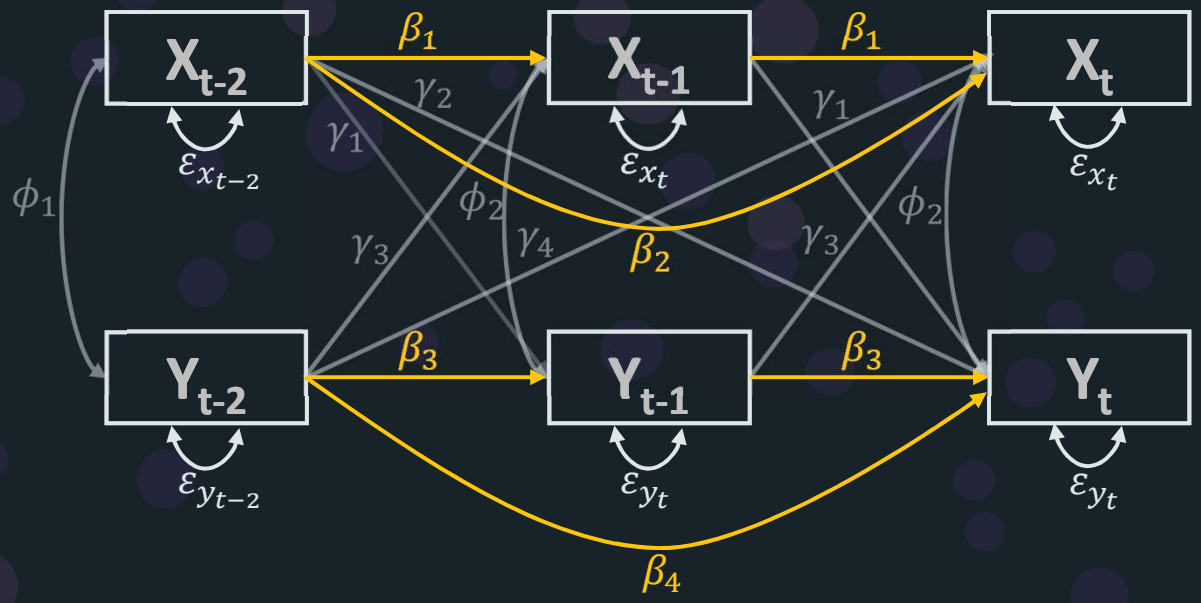
- Vector Autoregressive Process: VAR(2)



Multivariate Time Series Analysis

Implemented in SEM

- Vector Autoregressive Process: VAR(2)

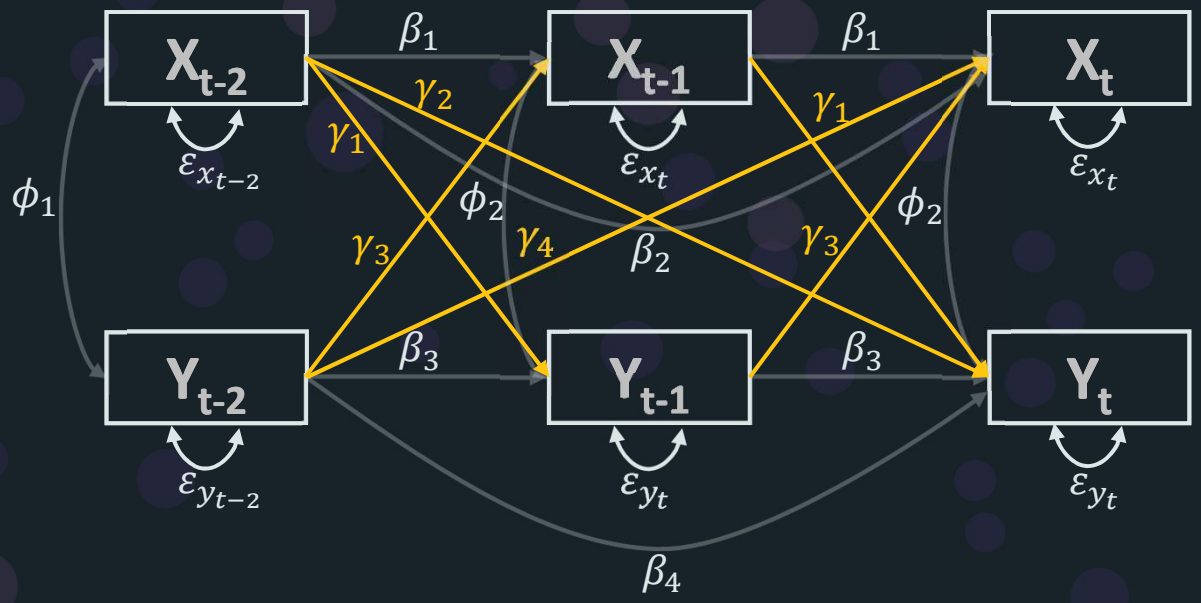


Autoregressive effects

Multivariate Time Series Analysis

Implemented in SEM

- Vector Autoregressive Process: VAR(2)



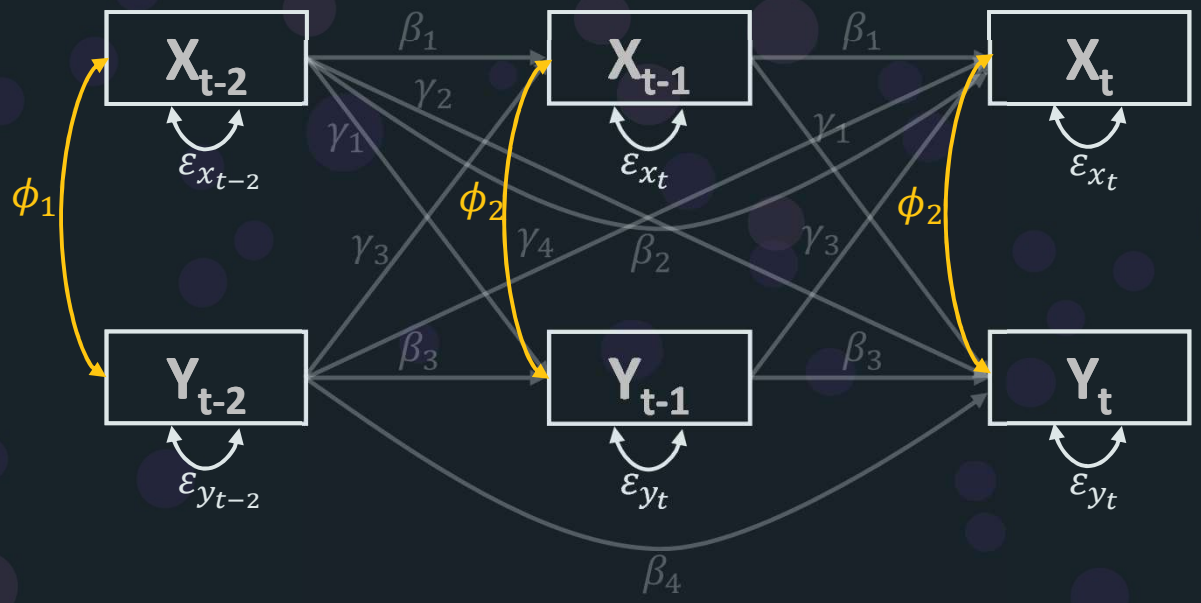
Cross-lagged effects

jmp STATISTICAL DISCOVERY FROM SAS

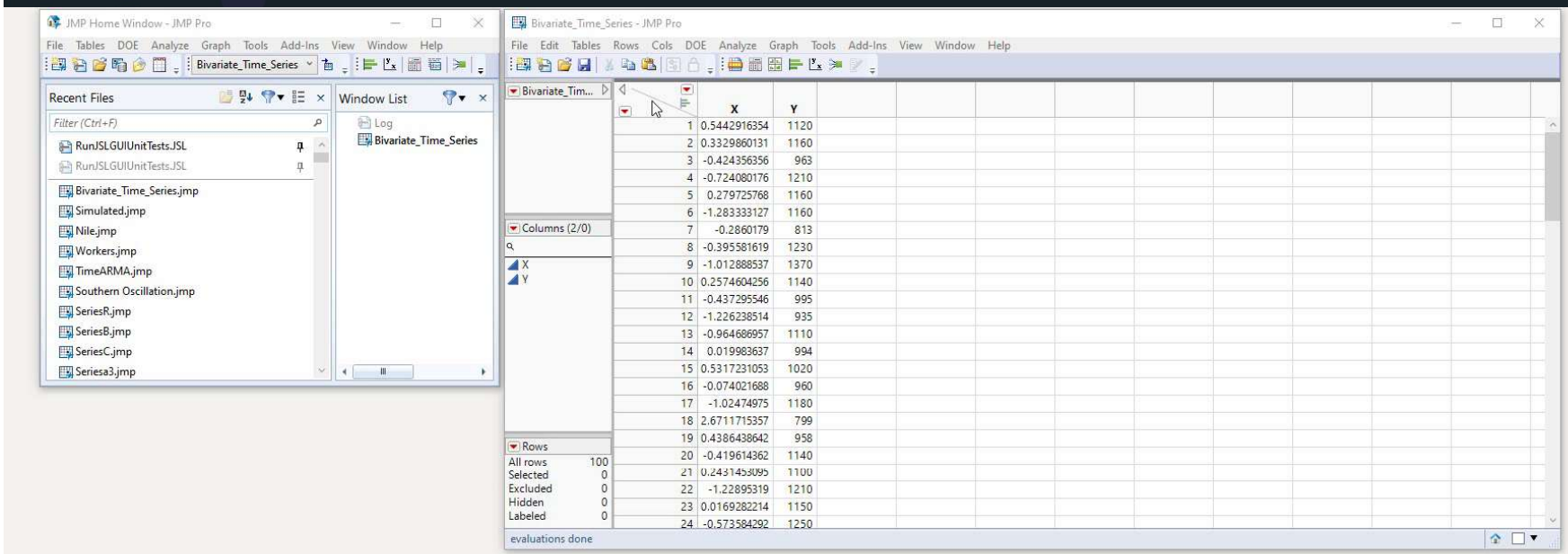
Multivariate Time Series Analysis

Implemented in SEM

- Vector Autoregressive Process: VAR(2)



Within-time covariances and residual covariances


JMP Home Window - JMP Pro
File Tables DOE Analyze Graph Tools Add-Ins View Window Help
Bivariate_Time_Series
Recent Files
Filter (Ctrl+F)
RunJSLGUIUnitTests.JSL
RunJSLGUIUnitTests.JSL
Bivariate_Time_Series.jmp
Simulated.jmp
Nile.jmp
Workers.jmp
TimeARMA.jmp
Southern Oscillation.jmp
SeriesR.jmp
SeriesB.jmp
SeriesC.jmp
Seriesa3.jmp
Window List
Log
Bivariate_Time_Series
Bivariate_Time_Series - JMP Pro
File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View Window Help
Bivariate_Tim...
Columns (2/0)
X
Y
Rows
All rows 100
Selected 0
Excluded 0
Hidden 0
Labeled 0
X Y
1 0.5442916354 1120
2 0.3329860131 1160
3 -0.424356356 963
4 -0.724080176 1210
5 0.279725768 1160
6 -1.283333127 1160
7 -0.2860179 813
8 -0.395581619 1230
9 -1.012888537 1370
10 0.2574604256 1140
11 -0.437295546 995
12 -1.226238514 935
13 -0.964686957 1110
14 0.019983637 994
15 0.5317231053 1020
16 -0.074021688 960
17 -1.02474975 1180
18 2.6711715357 799
19 0.4386438642 958
20 -0.419614362 1140
22 -1.22895319 1210
23 0.0169282214 1150
24 -0.573584292 1250
evaluations done

Multivariate Time Series Analysis

Distillation Process Data

- 26 processes, ~45k measurements (10-minute intervals)
- Explore univariate time series models in JMP:
  - AR(1) or AR(2) fit best for most processes
  - VAR(2) for SEM
- Data pre-processing:
  - Missing data
    - SEM estimation can handle it BUT computationally intensive
    - Selected subset of complete data: ~13k
  - Large scale differences
    - Standardized all processes
  - Create lagged variables
- Model specification
  - Equality constraints are a challenge (tedious and error prone)
  - JSL script to generate JSL for SEM

Multivariate Time Series Analysis

Distillation Process Data