TUTORIAL NOTES ON
STRUCTURAL CAUSAL MODELING USING DAGs

December 2018

T. Eapen

Structural Causal Modeling (SCM) using DAGs, or Graphical Causal Modeling (Elwert 2013), is based on the seminal work of Pearl (2009). It has origins in the path analysis of Wright (1921). Other early contributions include those of Haavelmo (1943), Simon (1954) and Strotz and Wold (1960) in economics, and Hubert Blalock (1962) in social sciences. Structural Causal Modeling can be viewed as a graph-theoretic approach to structural equation modeling, where the absence (or presence) of arrows between two variable represents a presumption of absence (or presence) of causal relationship between the two variables (Bollen and Pearl 2013). Developed as a non-parametric approach to causal inference, it's extensions to linear (Gaussian) structural equation modeling has been discussed in Pearl (2012), Bollen and Pearl (2013), and Chen, Pearl and Kline (2018). Relationship to structural models in economics is discussed by Pearl (2015), Heckman and Pinto (2015) and White and Chalak (2009).

Strutural Causal Modeling using DAGs has recently extensive application in epidemiology (Textor et al 2016), medical sciences (Shrier and Platt 2008; Wang and Bautista 2015), sociology (Elewert 2013), as well as some applications in psychology (van Kampen 2013) and economics (White and Chalak 2009; White and Lu 2011). A related application of DAGs is in causal discovery as part of the Tetrad Project (van Kampen et al 2014).

In Structural Causal Modeling, the model, which represents the causal assumption of the researcher, is represented using a Directed Acyclical Graph (Lauritzen 1996). The use of DAGs provide a principled approach to identifying minimally sufficient covariates (adjustment sets) that can be used as control variables to eliminate confounding. This is achieved through the application of graph-theoretic principles such as d-separation and back-door criterion (Pearl 2009)

The DAG represents the qualitative causal assumptions of the researchers and the substantive knowledge the researcher has about the world. The DAG encodes all the causal assumptions in the model through the presence/absence of directed arrows (Pearl 2009; Elwert 2013). These arrows can be unidirectional (X→Y), representing a causal effect or a bidirectional (X←→Y) representing a confounder[1].

The existence of a directed arrow between two nodes X→Y represents a *weak* assumption about the causal effect of X on Y. This is considered to be a weak (conservative) assumption, as it does not preclude the absence of a relationship between the two variable. In contrast, the absence of an arrow between two variable represents a *strong* assumption made by the modeler that the two variables are not causally related – hence independent (X_||_Y). Once the graphical model is formulated, it's implications can be tested again data.

There are three key advantages of Graphical Causal Modelling using DAGs. These are: (1) availability of local testable implications (2) principled approach to identifying conditioning variables (3) provision model implied IVs and conditional IVs to address confounding bias.

Firstly, the model provides certain testable implications on the data. These represents quantitative restrictions on the probability distribution on datasets that are compatible with the DAG model. These testable implications are in the form of conditional independence between variables that must hold true in the data, for the data to be consistent with the model. These implications emerge from the d-separation criterion that will be discussed later. These implications can be considered to be local tests of model fit. A consequent advantage of graphical SCM models compared to traditional SEM models (Bollen 1989) is that it allows for identification for part of the model even if the overall model is unidentified (Chen, Pearl and Kline 2018).

Second, it provides a principled approach to identifying conditioning variables to estimate coefficient. It is common in regression models to control for confounding bias by introducing control variables. However, the inappropriate selection of such conditioning variables can introduce new biases. The use of SCM enables a graph-theoretic approach to identifying minimal conditioning (or adjustment) sets. This can eliminate biases such as

[1] The birectional arrow can be replaced by an unobserved (latent variable) X←U→Y

Berkson's bias (also called M-bias or endogenous selection) and over-conditioning bias (Pearl 2009), an issue that is frequently overlooked in empirical studies (Elwert and Winship 2014).

Thirdly, the SCM approach provides a graph theoretic approach to handling confounding bias (endogeneity) by identifying implied instrument variables (IVs) as well as conditional IVs. (Angrist, Imbens, and Rubin 1996; Brito and Pearl 2002; Van Der Zander, Liśkiewicz, and Textor 2015)

Building Blocks of Structural Causal Models

The use of graphs for graphical analysis can be traced to the work of Sewell Wright and his pioneering path tracing rules (Wright 1921). Method of vanishing partial correlations to assess model fit (Blalock 1962; Simon 1977). We begin with the definition of key components of every SCM model: nodes, paths, and, routes (Chen and Pearl 2014; Chen Pearl and Kline 2018; Thoemmes et al 2018). In a graphical model, each node represents a variable. A path consists of a directed arrow that connected two nodes. The path represents a possible influence of one variable on the other. A sequence of paths is termed as a route.

A graphical model can be decomposed into four building blocks involving two paths between three variables X, Y, Z. These are chains (X→Z→Y), inverse chains (X←Z←Y), forks (X←Z→Y) and colliders (X→Z←Y). These structures imply certain independencies in the data. The collider represents blocking of information in the path at the middle variable Z. This implies an unconditional independence (X_|_Y). Conditioning on Z opens up this closed path, introducing a spurious relation between X and Y (M-bias or Berkson bias). The other three structure chains, inverse chains, and forks do not imply any unconditional independencies. Instead, they imply the conditional independencies given by (X_|_Y|Z). That is conditioning on the middle variable closes the path between X and Y and renders the two variable independent. These basic rules for conditioning can be extended to a set of variables. An important implication is that sensitivity to different control sets is not necessarily a test of robustness as conditioning on a collider or it's descendants can introduce a spurious relationship between two variables (Chen and Pearl 2014).

Partial correlation coefficient between X and Y is the correlation between the two variables after controlling for Z. The partial regression coefficients can be calculated from the pairwise correlation coefficients using Cramer (1946) (Chen and Pearl 2014). If the conditioning set Z involves a large number of variables, the above relation can be used recursively. However, this is unwieldy. If the objective is to obtain partial correlation, the d-separation criteria that employs a graph theoretic approach is more convenient.

D-Separation Criterion

The d-separation criterion is provided by Pearl (1988). It enables us to derived the model implied conditional independences and zero partial correlation by analyzing the path diagram. This is used to obtain testable implications which can be used as goodness of fit test, as well as aid us in in the identification analysis

Three rules of d-separation are the following (Chen and Pearl 2014; Chen, Pearl and Kline 2018):

(1) Variables X and Y are d-separated given a conditioning set Z if there are no active routes between X and Y given Z.
(2) A path between two variables is d-separated if (1) the path is a chain or fork and one of the variables in the middle of the path is controlled or (2) the path contains a collider.
(3) If X and Y are d-separated given a set Z in the path diagram, G, this implies that X and Y are independent given Z. In a linear model, this implies that the partial correlation coefficient =0.

The d-separation criteria impose restrictions on the causal model in the form of conditional or unconditional independencies. These conditional independencies must exist in the data for it to be compatible with a given model. If one or more of the implied independencies does not hold in the data, this means that the causal model has been incorrectly specified. Thus independencies implied in the model, identified by the d-separation criterion provides very strong tests of the credibility of the model.

A conditional independence implied by the model (such as X_|_Y|Z) can be tested by regressing both X and Y on Z and then testing for non-zero correlation between the residuals. If the data is normally distributed, then linear regression can be used, else non-parametric regression techniques should be used to compute the residual. The R package dagitty provides this feature (Textor et al 2016)

Identification and Single Door Criterion Criterion

The d-separation criterion can be utilized to identify causal relationship between two variables. This is achieved by blocking all the non-causal (or back-door paths) that introduce spurious relationship between X and Y. This criterion can be used to identify the minimal conditioning set.

Single Door Criterion. (Pearl 2009, Chen, Pearl and Kline 2018): Let G be a DAG and α be path coefficient in the path X→Y. If there exists a set of variable W such that

(1) W contains no descendent of Y
(2) W d-separates X from Y in subgraph G_α obtained by removing the edge X→Y from G. Then α is given by the regression coefficient β (Y,X|W).

A related criterion called back door criterion provides the conditions for estimating a total effect using OLS even when the overall model is not identified (Chen and Pearl 2014; Pearl 2009).

Back-door Criterion for Total Effect

It may be possible to identify some of the model coefficients, even if the overall model is not identified. The back-door criterion (Pearl 2000; Chen and Pearl 2014) may be used in identification of a total effect. For any two variables X and Y in a causal diagram G, the total effect of X on Y is identifiable if there exists a set of measurements Z that

(1) No member of Z is a descendent of X; and
(2) Z d-separates X from Y in the subgraph Gx formed by deleting from G all arrows emanating from X

If the above two conditions are satisfied, then the total effect of X on Y is given by $\beta_{YX.Z}$

Instrumental and Conditional Instrumental Variables

The graph theoretic definition of Instrumental Variables in Provided by Brito and Pearl (2002) and Chen and Pearl (2014). A variable Z qualifies as an instrumental variable for coefficient α from X to Y if

(1) Z is d-separated from Y in the subgraph G_α obtained by removing the edge X→Y from G and
(2) Z is not d-separated from X in G_α

Conditional instruments are variable that can be utilized as IVs once they are conditioned on a set of set of other variables. In many models, an IV that meets exclusion restriction may not be available, instead a conditional IV may exist that satisfies the exclusion restriction (Van Der Zander, Liśkiewicz, and Textor 2015)

A variable Z is a conditional instrumental variable given a set W for coefficient α from X to Y if

(1) W contains only non-descendants of Y
(2) W d-separates Z from Y in the subgraph $G\alpha$ obtained by removing edge X→Y from G
(3) W does not d-separate Z from X in $G\alpha$

When Z is a conditional instrument for a α given a set W, then $\alpha = \beta_{ZY.W/}\beta_{ZX.W}$

The R package Dagitty provides a convenient way to identifying conditioning variables (Textor and Liskiewicz 2012) and conditional IVs (Van Der Zander et al 2015).

Use of Local Fit Measures in DAGs

In traditional path models/ SEMs fit measures are reported such as $\chi 2$ and various fit indices such as RMSEA and CFI. The use of local fit measures have been largely absent, with the exception of the use of modification indices (Thoemmes, Rosseel and Textor 2018). In graphical modeling using DAGs, the use of local fit measures provided by the d-separation criterion provides several advantages not provided by global fit measures. Local tests help us identify the parts of the graphical model that are supported by the data. Local tests can also be performed immediately after data collection and can sometime provide information about which part of the overall model does not fit the observed data.

Equivalent Models

Equivalent models are alternative models with the same data implications in relation to d-separation. Models are equivalent if they have the same skeleton and same v-structure (Pearl 2009). The presence of equivalent models pose a threat to the uniqueness of the model.

REFERENCES

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin (1996), “Identification of causal effects using instrumental variables,” Journal of the American Statistical Association 91(432), 444-455.

Blalock, H. (1962), “Four-variable causal models and partial correlations,” American Journal of Sociology, 68, 182–194.

Bollen, Kenneth A. (2012), “Instrumental variables in sociology and the social sciences." Annual Review of Sociology, 38, 37-72.

Bollen, Kenneth A., and Judea Pearl. "Eight myths about causality and structural equation models." In Handbook of causal analysis for social research, pp. 301-328. Springer, Dordrecht, 2013.

Bowden, Roger J., and Darrel A. Turkington (1984), “Instrumental Variables,” Econometric Society Monograph in Quantitative Economics.

Brito, Carlos, and Judea Pearl (2002), “Generalized instrumental variables,” Proceedings of the Eighteenth conference on Uncertainty in Artificial Intelligence, 85-93.

Chen, Bryant, Judea Pearl, and Rex Kline, (2019), “Graphical tools for linear path models,” Psychometrika (Forthcoming).

Chen, Bryant, and Judea Pearl (2014), “Graphical tools for linear structural equation modeling,” Tech. Rep. R-432, Department of Computer Science, University of California, Los Angeles, CA.

Elwert, Felix (2013), “Graphical Causal Models,” Handbook of causal analysis for social research, Springer, Dordrecht, 245-273.

Elwert, Felix, and Christopher Winship. "Endogenous selection bias: The problem of conditioning on a collider variable." Annual Review of Sociology, 40, 31-53.

Forster, Eve A., and Kevin N. Dunbar (2009), “Creativity evaluation through Latent Semantic Analysis,” Proceedings of the Annual Conference of the Cognitive Science Society, 602-607.

Gelman, Andrew, and Guido Imbens (2013), “Why ask why? Forward causal inference and reverse causal questions,” No. w19614, National Bureau of Economic Research.

Greenland, Sander (2000), “An introduction to instrumental variables for epidemiologists,” International Journal of Epidemiology, 29(4), 722-729.

Greenland, Sander, Judea Pearl, and James M. Robins (1999), “Causal diagrams for epidemiologic research,” Epidemiology, 37-48.

Heckman, James, and Rodrigo Pinto (2015), “Causal analysis after Haavelmo,” Econometric Theory, 31(1), 115-151.

Pearl, Judea (1995), “Causal diagrams for empirical research,” Biometrika, 82(4), 669-688.

Pearl, Judea (1995), “On the testability of causal models with latent and instrumental variables,” Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, pp. 435-443. Morgan Kaufmann Publishers Inc., 1995.

Pearl, Judea (2009), Causality, Cambridge University Press.

Pearl, Judea (2012), “The causal foundation of structural equation modeling,” In R. Hoyle (Ed.), Handbook of structural equation modeling (pp. 68–91), Newbury Park: Sage.

Pearl, Judea (2015), “Trygve Haavelmo and the emergence of causal calculus,” Econometric Theory, 31(1), 152-179.

Pearl, Judea. "Causal diagrams for empirical research." Biometrika 82, no. 4 (1995): 669-688.

Perkovic, J. Textor, M. Kalisch and M. H. Maathuis (2015), “A Complete Generalized Adjustment Criterion,” Proceedings of UAI.

Richardson, Thomas, Spirtes, Peter (2002), “Ancestral graph Markov models,” Annals of Statistics, 30(4), 962-1030.

Shpitser, T. VanderWeele and J. M. Robins (2010), “On the validity of covariate adjustment for estimating causal effects, Proceedings of UAI.

Shrier, Ian, and Robert W. Platt (2008), “Reducing bias through directed acyclic graphs,” BMC Medical Research Methodology, 8(1).

Simon, Herbert A. (1954), “Spurious correlation: A causal interpretation,” Journal of the American Statistical Association, 49(267), 467-479.

Simon, Herbert A. (1977), “Causal ordering and identifiability,” Models of Discovery, Springer, Dordrecht.

Textor, Johannes, Benito van der Zander, Mark S. Gilthorpe, Maciej Liśkiewicz, and George TH Ellison (2016), “Robust causal inference using directed acyclic graphs: the R package ‘dagitty’,” International journal of epidemiology, 45(6), 1887-1894.

Textor, Johannes, and Maciej Liskiewicz (2012), “Adjustment criteria in causal diagrams: An algorithmic perspective,” arXiv preprint, arXiv:1202.3764.

Thoemmes, Felix, Yves Rosseel, and Johannes Textor (2018), “Local fit evaluation of structural equation models using graphical criteria,” Psychological Methods, 23(1).

Textor, Johannes, Juliane Hardt, and Sven Knüppel (2011), “DAGitty: a graphical tool for analyzing causal diagrams,” Epidemiology, 22(5).

Tikka, Santtu, and Juha Karvanen (2018), “Identifying causal effects with the R package causaleffect,” arXiv preprint, arXiv:1806.07161.

Van Der Zander, Benito, Maciej Liśkiewicz, and Johannes Textor (2015), “Efficiently finding conditional instruments for causal inference,” 3243-3249.

Van der Zander, Benito, and Maciej Liskiewicz (2016), “On Searching for Generalized Instrumental Variables,” AISTATS, 1214-1222.

Van der Zander, M. Liskiewicz and J. Textor (2014), “Constructing separators and adjustment sets in ancestral graphs,” Proceedings of UAI.

Van der Zander and M. Liskiewicz (2016), “Separators and Adjustment Sets in Markov Equivalent DAGs,” Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI’16), Phoenix, Arizona, USA.

van Kampen, Dirk. (2014), “The SSQ model of schizophrenic prodromal unfolding revised: An analysis of its causal chains based on the language of directed graphs,” European Psychiatry, 29(7), 437-448.

Wang, Lina, and Leonelo E. Bautista (2014), “Serum bilirubin and the risk of hypertension,” International Journal of Epidemiology, 44(1), 142-152.

White, Halbert, and Xun Lu (2011), “Causal diagrams for treatment effect estimation with application to efficient covariate selection, “Review of Economics and Statistics 93(4), 1453-1459.

Wright, Sewall (1921), “Correlation and causation,” Journal of Agricultural Research, 20(7), 557-585.

Zhang (2008), “Causal Reasoning with Ancestral Graphs,” Journal of Machine Learning Research, 1437-1474.