DISCUSSION PAPER SERIES

IZA DP No. 7628

Causal Analysis after Haavelmo

James Heckman
Rodrigo Pinto

September 2013

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Causal Analysis after Haavelmo

James Heckman
University of Chicago,
University College Dublin, American Bar Foundation and IZA

Rodrigo Pinto
University of Chicago

Discussion Paper No. 7628
September 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 7628
September 2013

ABSTRACT

Causal Analysis after Haavelmo[*]

Haavelmo's seminal 1943 paper is the first rigorous treatment of causality. In it, he distinguished the definition of causal parameters from their identification. He showed that causal parameters are defined using *hypothetical* models that assign variation to some of the inputs determining outcomes while holding all other inputs fixed. He thus formalized and made operational Marshall's (1890) *ceteris paribus* analysis. We embed Haavelmo's framework into the recursive framework of Directed Acyclic Graphs (DAG) used in one influential recent approach to causality (Pearl, 2000) and in the related literature on Bayesian nets (Lauritzen, 1996). We compare an approach based on Haavelmo's methodology with a standard approach in the causal literature of DAGs – the "*do-calculus*" of Pearl (2009). We discuss the limitations of DAGs and in particular of the *do-calculus* of Pearl in securing identification of economic models. We extend our framework to consider models for simultaneous causality, a central contribution of Haavelmo (1944). In general cases, DAGs cannot be used to analyze models for simultaneous causality, but Haavelmo's approach naturally generalizes to cover it.

JEL Classification: C10, C18

Keywords: causality, identification, do-calculus, directed acyclic graphs, simultaneous treatment effects

Corresponding author:

James Heckman
The University of Chicago
Department of Economics
1126 E. 59th St.
Chicago, IL 60637
USA
E-mail: jh@uchicago.edu

[*] We thank the guest editor of this issue Olav Bjerkholt for very helpful comments, participants at the Haavelmo symposium, Oslo, Norway, December, 2011, as well as Steven Durlauf, Maryclare Griffin and Cullen Roberts, three anonymous referees, participants in the seminar on Quantitative Research Methods in Education, Health and Social Sciences at the University of Chicago, March 2013, and participants at a seminar at UCL, September 4, 2013. Steve Stigler gave us helpful bibliographic references. This research was supported in part by the American Bar Foundation, a grant from the European Research Council DEVHEALTH-269874, and NICHD R37-HD065072. The views expressed in this paper are those of the authors and not necessarily those of the funders or commentators mentioned here.

1 Trygve Haavelmo's Causality

Trygve Haavelmo made fundamental contributions to understanding the formulation and identification of causal models. In two seminal papers (1943, 1944), he formalized the distinction between correlation and causation,[1] laid the foundation for counterfactual policy analysis and distinguished the concept of "fixing" from the statistical operation of conditioning—a central tenet of structural econometrics. He developed an empirically operational version of Marshall's notion of *ceteris paribus* (1890) which is a central notion of economic theory.

In Haavelmo's framework, the causal effects of inputs on outputs are determined by the impacts of *hypothetical* manipulations of inputs on outputs which he distinguishes from correlations between inputs and outputs in observational data. The causal effect of an input is defined using a hypothetical model that abstracts from the empirical data generating process by making hypothetical variation in inputs that are independent of all other determinants of outputs. As a consequence, Haavelmo's notion of causality relies on a thought experiment in which the model that governs the observed data is extended to allow for independent manipulation of inputs, irrespective of whether or not they vary independently in the data.

Haavelmo formalized Frisch's notion that "causality is in the mind."[2] Causal effects

[1] To our knowledge, the first recorded statement of the distinction correlation and causation is due to Fechner (1851), who distinguished "causal dependency" from what he called "functional relationship". See Heidelberger (2004, p. 102). In later work, Yule (1895, footnote 2, p. 605) discussed the distinction between correlation and causation in a discussion of the effect of relief payments on pauperism. We thank, respectively, Olav Bjerkholt and Steve Stigler for these references.

[2] This notion is central to structural econometrics. It was developed by Frisch and participants in his laboratory going back to at least 1930:

> "*. . . we think of a cause as something imperative which exists in the* exterior world. *In my opinion this is fundamentally wrong. If we strip the word cause of its animistic mystery, and leave only the part that science can accept, nothing is left except a* certain way of thinking, *an intellectual trick . . . which has proved itself to be a useful weapon . . . the scientific . . . problem of causality is essentially a problem regarding our way of thinking, not a problem regarding the nature of the exterior world.*" (Frisch 1930, p. 36, published 2011)

Writing in the heyday of the Frisch-Haavelmo-inspired Cowles Commission in the late 1940's, Koopmans and Reiersøl distinguished descriptive statistical inference form structural estimation in the following statement.

> "*In many fields the objective of the investigator's inquisitiveness is not just a "population" in the sense of a distribution of observable variables, but a physical structure projected behind this distribution, by which the latter is thought to be generated. The word "physical" is used merely to convey that the structure concept is based on the investigator's ideas as to the "explanation" or "formation" of the phenomena studied, briefly, on his theory of these phenomena, whether*

are not empirical statements or descriptions of actual worlds, but descriptions of hypothetical worlds obtained by varying—hypothetically—the inputs determining outcomes. Causal relationships are often suggested by observed phenomena, but they are abstractions from it. [3]

This paper revisits Haavelmo's notions of causality using the mathematical language of Directed Acyclic Graphs (DAGs). We start with a recursive framework less general than that of Haavelmo (1943). This allows us to represent causal models as Directed Acyclic Graphs which are intensively studied in the literature on Bayesian networks (Howard and Matheson, 1981; Lauritzen, 1996; Pearl, 2000). We then consider the general non-recursive framework of Haavelmo (1943, 1944) which cannot, in general, be framed within the context of DAGs.

Following Haavelmo, we distinguish hypothetical models that are used to define causal parameters as idealizations of empirical models that govern data generating processes. This enables us to discuss causal concepts such as "fixing" using an intuitive approach that draws on Haavelmo's notion of causality. Identification relies on linking the parameters defined in a hypothetical model using data generated by an empirical model.

This paper makes the following contributions to the literature on causality: (1) We build a framework for the study of causality inspired by Haavelmo's concept of hypothetical variation of inputs; (2) In doing so, we express Haavelmo's notion of causality in the mathematical language of DAGs; (3) For this class of models, we compare the simplicity of Haavelmo's

> *they are classified as physical in the literal sense, biological, psychological, sociological, economic or otherwise.*" (Koopmans and Reiersøl 1950, p. 165)

See Simon (1953), Heckman (2008) and Freedman et al. (2010), for later statements of this point of view.

[3] All models—empirical or hypothetical—are idealized thought experiments. There are no formalized rules for creating models, causal or empirical. Analysts may differ about the inputs and relationships in either type of model. A model is more plausible the more phenomena it predicts and the deeper are its foundations in established theory. Causal models are idealizations of empirical models which are in turn idealizations of phenomena. Some statisticians reject the validity of hypothetical models and seek to define causality using empirical methods (Sobel, 2005). As an example we can cite the "Rubin model" of Holland (1986), which equates establishing causality with the empirical feasibility of conducting experiments. This approach confuses definition of causal parameters with their identification from data. We refer to Heckman (2005, 2008) for a discussion of this approach.

framework with the well-known causal framework of the *do-calculus* proposed by Pearl (2000) which is beginning to be used in economics (see e.g. Margolis et al., 2012; White and Chalak, 2009); (4) We then discuss the limitations of the use of DAGs for econometric identification. We show that even in recursive models, the methods that rely solely on the information in DAGs do not exploit identification strategies based on functional restrictions and exclusion restrictions that are generated by economic theory. This limitation produces apparent non-identification in classically identified econometric models. We show how Haavelmo's approach naturally extends to notions of simultaneous causality while the DAG approach is fundamentally recursive.

Our paper is mainly on the methodology of causality. We do not create a new concept of causality, but rather propose a new framework within which to discuss it. We show that Haavelmo's approach is a complete framework for the study of causality which accommodates the main tools of identification used in the current literature in econometrics whereas other approaches do not.

We show that the causal operation of fixing described in Haavelmo (1943) and Heckman (2005, 2008) is equivalent to statistical conditioning when embedded in a hypothetical model that assigns independent variation to inputs with regard to all variables not caused by those inputs. Pearl (2009) uses the term *do* for the concept of fixing a variable. We show the relationship between statistical conditioning in a hypothetical model and the do-operator. Fixing, in our framework, differs from the operation of the do-operator because it targets specific causal links instead of variables that operate across multiple causal links. A benefit of targeting causal links is that it simplifies the analysis of the subsets of causal relationships associated with an input variable when compared to the do-operator.

Haavelmo's approach allows for a precise yet intuitive definition of causal effects. With it, analysts can identify causal effects by applying standard statistical tools. In contrast with the do-calculus, application of Haavelmo's concepts eliminates the need for additional extra-statistical graphical/statistical rules to achieve identification of causal parameters.

Haavelmo's approach also covers the case of simultaneous causality in its full generality whereas frameworks for causal analysis currently used in statistics cannot, except through introduction and application of *ad hoc* rules.

This paper is organized in the following way. Section 2 reviews Haavelmo's causal framework. Section 3 uses a modern framework of causality to assess Haavelmo's contributions to the literature. Section 4 examines how application of this framework differs from Pearl's do-calculus (2009) and enables analysts to apply the standard tools of probability and statistics without having to invent extra-statistical rules. It gives an example of the identification of causal effects that considers Pearl's "Front-Door" criteria. Section 5 discuss the limitations of DAGs in implementing the variety of sources of identification available to economists. We focus on the simplest cases of confounding models where instrumental variables are available. Section 6 extends the discussion to a simultaneous equations framework. Section 7 concludes.

2 Haavelmo's Causal Framework

We review the key concepts of causality developed by Haavelmo (1943, 1944)—starting with a recursive model. A causal model is based on a system of structural equations that define causal relationships among a set of variables. In the language of Frisch (1938), these structural equations are *autonomous* mechanisms represented by deterministic functions mapping inputs to outputs. By autonomy we mean, as did Frisch, that these relationships remain invariant under external manipulations of their arguments. They are functions in the ordinary usage of the term in mathematics. They produce the same values of the outcomes when inputs are assigned to a fixed set of values, however those values are determined. Even though the functional form of a structural equation may be unknown, the causal directions among the variables of a structural equation are assumed to be known. They are determined by thought experiments that may sometimes be validated in data. The variables chosen as arguments in a structural equation are assumed to account for all causes of the associated output variable.

Haavelmo developed his work on causality for aggregate economic models. He considered mean causal effects and, for the sake of simplicity, invoked linearity, assumed uniformity of responses to inputs across agents, and focused on continuous variables. More recent approaches generalize his framework.

Haavelmo formalized the distinction between correlation and causation using a simple model. In order to examine his ideas, consider three variables Y, X, U associated with error terms $\boldsymbol{\epsilon} = (\epsilon_U, \epsilon_X, \epsilon_Y)$ such that X, Y are observed by the analyst while variable U, ϵ are not.[4] He assumed that U is a confounding variable that causes Y and X. We represent this model through the following structural equations:

$$Y = f_Y(X, U, \epsilon_Y), \quad X = f_X(U, \epsilon_X), \quad \text{and } U = f_U(\epsilon_U),$$

where $\boldsymbol{\epsilon}$ is a vector of mutually independent error terms with cumulative distribution function Q_ϵ. Thus, if X, U, ϵ_Y take values of x, u, e_Y, then Y must take the value $y = f_Y(x, u, e_Y)$. By iterated substitution we can express all variables in terms of $\boldsymbol{\epsilon}$**.** Moreover, the mutual independence assumption of error terms implies that ϵ_Y is independent of (X, U) as $X = f_X(f_U(\epsilon_U), \epsilon_X)$ and $U = f_U(\epsilon_U)$. Notationally, we write $(X, U) \perp\!\!\!\perp \epsilon_Y$ where $\perp\!\!\!\perp$ denotes statistical independence. In the same fashion, we have that $\epsilon_X \perp\!\!\!\perp U$ but X is not independent of ϵ_U.

Haavelmo defines the causal effect of X on Y as being generated by a *hypothetical manipulation* of variable X that does not affect the values that U or $\boldsymbol{\epsilon}$ take. This is called *fixing* X by a hypothetical manipulation.[5] Notationally, outcome Y when X is fixed at x is denoted

[4] This framework allows for uncertainty on the part of agents if realizations of the uncertain variables are captured through variables X and U. In that sense the model can be characterized as a method for examining *ex-post* relationships between variables. For a discussion of causal analysis of *ex-post* versus *ex-ante* models, see, e.g., Hansen and Sargent (1980) and Heckman (2008).

[5] Haavelmo (1943) did not explicitly use the term "fixing." He set U (in our notation) to a specified value and manipulated X in his "hypothetical model." Specifically, Haavelmo set $U = 0$ but the point of evaluation is irrelevant in the linear case he analyzed.

by $Y(x) = f_Y(x, U, \epsilon_Y)$ and its expectation is given by $\mathbb{E}_{(U,\epsilon_Y)}(Y(x)) = \mathbb{E}(f(x, U, \epsilon_Y))$, where $\mathbb{E}_{(U,\epsilon_Y)}(\cdot)$ means expectation over the distribution of random variables U and ϵ_Y. The average causal effect of X on Y when X takes values x and x' is given by $\mathbb{E}_{(U,\epsilon_Y)}(Y(x)) - \mathbb{E}_{(U,\epsilon_Y)}(Y(x'))$. For notational simplicity, we henceforth suppress the subscript on $\mathbb{E}$ denoting the random variable with respect to which the expectation is computed.

Conditioning is a statistical operation that accounts for the dependence structure in the data. Fixing is an abstract operation that assigns independent variation to the variable being "fixed". The standard linear regression framework is convenient for illustrating these ideas and in fact is the one used by Haavelmo (1943).

Consider the standard linear model $Y = X\beta + U + \epsilon_Y$ where $\mathbb{E}(\epsilon_Y) = 0$ represent the data generating process for Y. The expectation of outcome Y when X is *fixed* at x is given by $\mathbb{E}(Y(x)) = x\beta + \mathbb{E}(U)$. This equation corresponds to Haavelmo's (1943) hypothetical model. The expectation of Y when X is *conditioned* on x is given by $\mathbb{E}(Y|X = x) = x\beta + \mathbb{E}(U|X = x)$, as $\mathbb{E}(\epsilon_Y|X = x) = 0$ because $\epsilon_Y \perp\!\!\!\perp X$. If $\mathbb{E}(U|X = x) = 0$ and elements of X are not collinear, then OLS identifies β and $\mathbb{E}(Y|X = x) = \mathbb{E}(Y(x)) = x\beta$ and β generates the average treatment effect of a change in X on Y. Specifically, $(x - x')\beta$ is the average difference between the expectation of Y when X is fixed at x and x'.

The difficulty of identifying the average causal effect of X on Y when $\mathbb{E}(U|X) \neq 0$ (and thereby $\mathbb{E}(Y|X = x) \neq \mathbb{E}(Y(x))$) stems from the potential confounding effects of unobserved variable U on X. In this case, the standard Least Squares estimator does not generate an autonomous causal or structural parameter because $plim(\hat{\beta}) = \beta + \text{cov}(X, U)/\,\text{var}(X)$ depends on the covariance between X and U. While the concept of a causal effect does not rely on the properties of the data generating process, the identification of causal effects does.

Without linearity, one needs an assumption stronger than $\mathbb{E}(U|X = x) = 0$ to obtain $\mathbb{E}(Y|X = x) = \mathbb{E}(Y(x))$. Indeed if one assumes no confounding effects of U, that is to say that X and U are independent ($X \perp\!\!\!\perp U$), then one can show that fixing is equivalent to statistical conditioning:

$$\begin{aligned}
\mathbb{E}(Y|X=x) &= \int f_Y(x,u,\epsilon_Y)dQ_{(U,\epsilon_Y)|X=x}(u,\epsilon_Y) \\
&= \int f_Y(x,u,\epsilon_Y)dQ_U(u)dQ_{\epsilon_Y}(\epsilon_Y) \\
&= \mathbb{E}(f_Y(x,U,\epsilon_Y)) \\
&= \mathbb{E}(Y(x)),
\end{aligned}$$

where $dQ_{(U,\epsilon_Y)|X=x}(u,\epsilon_Y)$ denotes the cumulative joint distribution function of U, ϵ_Y conditional on $X = x$ and the second equality comes from as the fact that U, X and ϵ_Y are mutually independent. If $X \perp\!\!\!\perp (U, \epsilon_Y)$ holds, we can use observational data to identify the mean value of Y fixing $X = x$ by evaluating the expected value of Y conditional on $X = x$. Note that in general, the value obtained depends on the functional form of $f_Y(x, u, \epsilon_Y)$.

Haavelmo's notation has led to some confusion in the statistical literature. His argument was aimed at economists of the 1940s and does not use modern notation. Haavelmo's key definitions and ideas are given by examples rather than by formal definitions. We restate and clarify his argument in this paper.

To simplify the exposition, assume that all variables are discrete and let $\mathbf{P}$ denote their probability measure. The factorization of the joint distribution of Y, U conditional on X is given by $\mathbf{P}(Y, U|X = x) = \mathbf{P}(Y|U, X = x)\,\mathbf{P}(U|X = x)$. In contrast, in the abstract operation of fixing X is assumed not to affect the marginal distribution of U. That is to say that $U(x) = U$. Therefore the joint distribution of Y, U when X is fixed at x is given by $\mathbf{P}(Y(x), U(x)) = \mathbf{P}(Y(x), U) = \mathbf{P}(Y|U, X = x)\,\mathbf{P}(U)$.

Fixing lies outside the scope of standard statistical theory and is often a source of confusion. Indeed, even though the probabilities $\mathbf{P}(Y|U, X = x)$ and $\mathbf{P}(U)$ are well defined, neither the causal operation of fixing nor the resulting joint distribution follow from standard statistical arguments.[6] Conditioning *is* equivalent to fixing under independence of X and

[6] See Pearl (2009) and Spirtes et al. (2000) for discussions.

U. In this case the conditional joint distribution of Y and U becomes $\mathbf{P}(Y, U|X = x) = \mathbf{P}(Y|U, X = x)\,\mathbf{P}(U|X = x) = \mathbf{P}(Y|U, X = x)\,\mathbf{P}(U)$.

To gain more intuition on the difference between fixing and standard statistical theory express the conditional expectation $\mathbb{E}(Y|X = x)$ as the integral across ϵ over a restricted set $\mathcal{A}^C$. By iterated substitution, we can write Y as $Y = f_Y(f_X(f_U(\epsilon_U), \epsilon_X), f_U(\epsilon_U), \epsilon_Y)$. Thus

$$\mathbb{E}(Y|X = x) = \frac{\int_{\mathcal{A}^C} f_Y(f_X(f_U(\epsilon_U), \epsilon_X), f_U(\epsilon_U), \epsilon_Y) dQ_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon})}{\int_{\mathcal{A}^C} dQ_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon})} \tag{1}$$

$$\text{where } \mathcal{A}^C = \{\boldsymbol{\epsilon} = (\epsilon_U, \epsilon_X, \epsilon_Y) \in \text{supp}(\boldsymbol{\epsilon})\ ;\ f_X(f_U(\epsilon_U), \epsilon_X) = x\}. \tag{2}$$

Fixing, on the other hand, is written as the integral across ϵ over its full support:

$$\mathbb{E}(Y(x)) = \frac{\int_{\mathcal{A}^F} f_Y(x, f_U(\epsilon_U), \epsilon_Y) dQ_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon})}{\int_{\mathcal{A}^F} dQ_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon})} \tag{3}$$

$$\text{where } \mathcal{A}^F = \{\boldsymbol{\epsilon} = (\epsilon_U, \epsilon_X, \epsilon_Y) \in \text{supp}(\boldsymbol{\epsilon})\} \quad \text{and} \quad \int_{\mathcal{A}^F} dQ_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) = 1. \tag{4}$$

Fixing differs from conditioning in terms of the difference in the integration sets $\mathcal{A}^F$ and $\mathcal{A}^C$. While conditional expectation (1) is a standard operation in statistics, the operation used to define fixing is not. Equation (1) is an expectation conditional on the event $f_X(f_U(\epsilon_U), \epsilon_X) = x$, which affects the integration set $\mathcal{A}^C$ as given in (2). Fixing (3), on the other hand, integrates the function $f_Y(x, f_U(\epsilon_U), \epsilon_Y)$ across the whole support of ϵ as given in (4). The inconsistency between fixing and conditioning in the general case comes from the fact that fixing X is equivalent to setting the expression $f_X(f_U(\epsilon_U), \epsilon_X)$ to x without changing the probability measures of ϵ_U, ϵ_X associated with the operation of conditioning on the event $X = x$.

This paper interprets Haavelmo's approach by introducing a hypothetical model that enables analysts to examine fixing using standard tools of probability. The *hypothetical model* departs from the data generating process by exploiting autonomy and creating a *hypothetical* variable that has the desired property of independent variation with regard to

U. The hypothetical model is an idealization of the empirical model. Standard statistical tools apply to both the data generating process and the hypothetical model.

To formalize Haavelmo's notions of causality, let a hypothetical model with error terms $\boldsymbol{\epsilon}$ and four variables including Y, X, U but also a new variable $\tilde{X}$ with the property that $\tilde{X} \perp\!\!\!\perp (X, U, \boldsymbol{\epsilon})$.[7] Invoking autonomy, the hypothetical model shares the same structural equation as the empirical one but departs from it by replacing X with an $\tilde{X}$-input, namely $Y = f_Y(\tilde{X}, U, \epsilon_Y)$. The hypothetical model is not a wildly speculative departure from the empirical data generating process but an expanded version of it. Thus $(Y|X = x, U = u) = f_Y(x, u, \epsilon_Y)$ in the empirical model and $(Y|\tilde{X} = x, U = u) = f_Y(x, u, \epsilon_Y)$ in the hypothetical model. The hypothetical model has the same marginal distribution of U as the empirical model. The joint distributions of variables in the empirical model $\mathbf{P}_{\mathrm{E}}$ and the hypothetical model $\mathbf{P}_{\mathrm{H}}$ may differ.

The hypothetical model clarifies the notion of fixing in the empirical model. Fixing in the empirical model is based on non-standard statistical operations. However, the distribution of the outcome Y when X is fixed at x in the empirical model can be interpreted as standard statistical conditioning in the hypothetical model, namely, $\mathbf{P}_{\mathrm{E}}(Y(x)) = \mathbf{P}_{\mathrm{H}}(Y|\tilde{X} = x)$. The next section formalizes this notion using one modern language of causality.[8]

3 Haavelmo's Framework Recast in a Modern Framework of Causality

We recast Haavelmo's model in the framework of Directed Acyclic Graphs (DAGs). DAGs are studied in Bayesian Networks (Howard and Matheson, 1981; Lauritzen, 1996) and are often used to define and estimate causal relationships (Lauritzen, 2001). The literature on

[7] We could express $\tilde{X} = f_{\tilde{X}}(\epsilon_{\tilde{X}})$ to be notationally consistent.

[8] Frisch's (1938) notion of invariance used by Haavelmo is called SUTVA in one model of causality popular in statistics. See Holland (1986) and Rubin (1986).

causality based on DAGs was advanced by Judea Pearl (2000, 2009).[9]

In this fundamentally recursive framework, a causal model consists of a set of variables $\mathcal{T} = \{V_1, \ldots, V_n\}$ associated with a set of mutually independent error terms $\boldsymbol{\epsilon} = \{\epsilon_1, \ldots, \epsilon_n\}$ and a system of autonomous structural equations $\{f_1, \ldots, f_n\}$. Variable set $\mathcal{T}$ includes both observed and unobserved variables. Variable set $\mathcal{T}$ also include both external and internal variables. We clarify these concepts in the following way.

Causal relationships between a dependent variable $V_i \in \mathcal{T}$ and its arguments are defined by $V_i = f_i(Pa(V_i), \epsilon_i)$, where $Pa(V_i) \subset \mathcal{T}$ and $\epsilon_i \in \boldsymbol{\epsilon}$ are called parents of V_i and are said to directly cause V_i. If $Pa(V) = \varnothing$ then variable V is not caused by any variable in $\mathcal{T}$. In this case, V is an *external variable* determined outside the system, otherwise the variable is called an *internal or endogenous variable.* The error terms in $\boldsymbol{\epsilon}$ are not caused by any variable and are introduced to avoid degenerate conditioning statements among variables in $\mathcal{T}$. For simplicity of notation, we keep the error terms $\boldsymbol{\epsilon}$ implicit, except when it clarifies matters to do so. We assume that all random variables in this section and the next are discrete valued although this requirement is easily relaxed.

Causal relationships are represented by a graph G where each node corresponds to a variable $V \in \mathcal{T}$. Nodes are connected by arrows from $Pa(V)$ to V and represent causal influences between variables. Descendants of a variable V, i.e. $D(V) \subset \mathcal{T}$, consist of all variables connected to V by arrows of the same direction arising from V. Graph G is called a DAG if no variable is a descendant of itself, i.e., $V \notin D(V), \ \forall \ V \in \mathcal{T}$. Observe that this assumption rules out simultaneity—a central feature of Haavelmo's approach. Children of a variable V are the set of variables that have V as a parent, namely, $Ch(V) = \{V' \in \mathcal{T}; V \in Pa(V')\}$.

Causal relationships are translated into statistical relationships in a DAG through a property termed the Local Markov Condition (LMC) (Kiiveri et al., 1984; Lauritzen, 1996). LMC states that a variable is independent of its non-descendants conditional on its parents.

[9] Chalak and White (2012) present generalizations of this approach.

LMC (5) also holds among variables in $\mathcal{T}$ under the assumption that error terms $\{\epsilon_1, \dots, \epsilon_n\}$ are mutually independent (Pearl, 1988; Pearl and Verma, 1994), namely:

$$\textbf{LMC}\text{: for all } V \in \mathcal{T}, \quad V \perp\!\!\!\perp (\mathcal{T} \setminus D(V)) \mid Pa(V). \tag{5}$$

We use Dawid's (1979) notation to denote conditional independence. If W, K, Z are subsets of $\mathcal{T}$, the expression $W \perp\!\!\!\perp K|Z$ means that each variable in W is statistically independent of each variable in K conditional on all variables in Z.

The conditional independence relationships generated by LMC (5) can be further manipulated using the Graphoid relations.[10] A benefit of LMC (5) is that we can factorize the joint distribution of variables $\mathbf{P}(V_1, \dots, V_n)$. Under a recursive model, we can assume without loss of generality that variables $(V_1, \dots, V_n, \dots, V_N)$ are ordered so that $(V_1, \dots, V_{n-1})$ are non-descendants of V_n and thereby $Pa(V_n) \subset (V_1, \dots, V_{n-1})$. Thus:

$$\mathbf{P}(V_1, \dots, V_n) = \prod_{V_n \in \mathcal{T}} \mathbf{P}(V_n | V_1, \dots, V_{n-1}) = \prod_{V_n \in \mathcal{T}} \mathbf{P}(V_n | Pa(V_n)), \tag{6}$$

where the last equality comes from applying LMC (5).

Table 1 uses the Haavelmo model described in Section 2 to illustrate the concepts discussed here. Table 1 presents two models and seven panels separated by a series of horizontal lines. The first panel names the models. The second panel presents the structural equations generating the models. Columns 1 and 2 are based on structural equations that have the

[10] The Graphoid relationships are a set of elementary conditional independence relationships presented by Dawid (1979):

$$\begin{aligned}
\text{Symmetry: } & X \perp\!\!\!\perp Y|Z \Rightarrow Y \perp\!\!\!\perp X|Z. \\
\text{Decomposition: } & X \perp\!\!\!\perp (W, Y)|Z \Rightarrow X \perp\!\!\!\perp Y|Z. \\
\text{Weak Union: } & X \perp\!\!\!\perp (W, Y)|Z \Rightarrow X \perp\!\!\!\perp W|(Y, Z). \\
\text{Contraction: } & X \perp\!\!\!\perp Y|Z \text{ and } X \perp\!\!\!\perp W|(Y, Z) \Rightarrow X \perp\!\!\!\perp (W, Y)|Z. \\
\text{Intersection: } & X \perp\!\!\!\perp W|(Y, Z) \text{ and } X \perp\!\!\!\perp Y|(W, Z) \Rightarrow X \perp\!\!\!\perp (W, Y)|Z. \\
\text{Redundancy: } & X \perp\!\!\!\perp Y|X.
\end{aligned}$$

The intersection relation is only valid for variables with strictly positive probability distributions. See also Dawid (2001).

same functional form, but different inputs. The third panel represents the associated model as a DAG. Squares represent observed variables, circles represent unobserved variables. (Except in the first panel, the components of $\boldsymbol{\epsilon}$ are kept implicit in the table.) The fourth panel displays the parents in $\mathcal{T}$ for each variable. The fifth panel shows the conditional independence relationships generated by the application of LMC (5) and the sixth panel presents the factorization of the joint distribution. The seventh and final panel provides the joint distribution of variables when X is fixed at x and the corresponding joint distribution for the hypothetical models. The content of the last panel is discussed further in this section.

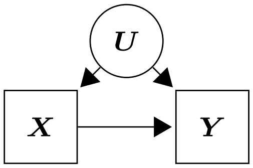
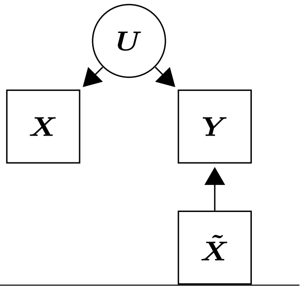
Using this framework, we can discuss the concept of fixing introduced in Section 2 in greater generality. Following Section 2, we define the causal operation of fixing a variable in a model represented by a graph G by the intervention that sets a value to this variable in $\mathcal{T}$ in a fashion that does not affect the distribution of its non-descendants. In other words, fixing a random variable (or a set of random variables) $X \in \mathcal{T}$ to x translates to setting $X = x$ for *all* X-inputs in the structural equations associated with variables in $Ch(X)$. Pearl (2009) uses the term *doing* for what we call *fixing*. We use his notation in writing Equation (7). The post-intervention distribution of variables in $\mathcal{T}$ when X is fixed at x is given by

$$\mathbf{P}(\mathcal{T} \setminus \{X\} | do(X) = x) = \prod_{V \in \mathcal{T} \setminus \{\{X\} \cup Ch(X)\}} \mathbf{P}(V|Pa(V)) \prod_{V \in Ch(X)} \mathbf{P}(V|Pa(V) \setminus \{X\}, X = x). \tag{7}$$

Versions of Equation (7) can be found in Pearl (2001); Robins (1986); Spirtes et al. (2000). In this instance, $\mathrm{do}(X) = x$ is equivalent to conditioning $\tilde{X}$ at $\tilde{X} = x$.

As noted in Section 2, standard arguments of statistical conditioning are unable to describe the probability laws governing the fixing operation used in Equation (7). Our solution to this problem draws on Haavelmo's insight that causality is a property of hypothetical models in which causal effects on output variables are generated through hypothetical independent variations of inputs. Specifically, we show that the fixing operation is easily translated into statistical conditioning under the Hypothetical model described in Section 3.1.

Table 1: **Haavelmo Empirical and Hypothetical Models**

1. Haavelmo Empirical Model	2. Haavelmo Hypothetical Model
$\mathcal{T} = \{U, X, Y\}$ $\boldsymbol{\epsilon} = \{\epsilon_U, \epsilon_X, \epsilon_Y\}$ $Y = f_Y(X, U, \epsilon_Y)$ $X = f_X(U, \epsilon_X)$ $U = f_U(\epsilon_U)$	$\mathcal{T} = \{U, X, Y, \tilde{X}\}$ $\boldsymbol{\epsilon} = \{\epsilon_U, \epsilon_X, \epsilon_Y\}$ $Y = f_Y(\tilde{X}, U, \epsilon_Y)$ $X = f_X(U, \epsilon_X)$ $U = f_U(\epsilon_U)$
	
$Pa(U) = \varnothing,$ $Pa(X) = \{U\}$ $Pa(Y) = \{X, U\}$	$Pa(U) = Pa(\tilde{X}) = \varnothing,$ $Pa(X) = \{U\}$ $Pa(Y) = \{\tilde{X}, U\}$
	$Y \perp\!\!\!\perp X \mid (\tilde{X}, U)$ $X \perp\!\!\!\perp (\tilde{X}, Y) \mid U$ $\tilde{X} \perp\!\!\!\perp U$
$\mathbf{P}_{\mathrm{E}}(Y, X, U) =$ $\mathbf{P}_{\mathrm{E}}(Y \mid X, U)\, \mathbf{P}_{\mathrm{E}}(X \mid U)\, \mathbf{P}_{\mathrm{E}}(U)$	$\mathbf{P}_{\mathrm{H}}(Y, X, U, \tilde{X}) =$ $\mathbf{P}_{\mathrm{H}}(Y \mid \tilde{X}, U)\, \mathbf{P}(X \mid U)\, \mathbf{P}_{\mathrm{H}}(U)\, \mathbf{P}_{\mathrm{H}}(\tilde{X})$
$\mathbf{P}_{\mathrm{E}}(Y, U \mid X \textit{ fixed} \text{ at } x) =$ $\mathbf{P}_{\mathrm{E}}(Y \mid X = x, U)\, \mathbf{P}_{\mathrm{E}}(U)$	$\mathbf{P}_{\mathrm{H}}(Y, U, X \mid \tilde{X} = x) =$ $\mathbf{P}_{\mathrm{H}}(Y \mid \tilde{X} = x, U)\, \mathbf{P}(X \mid U)\, \mathbf{P}_{\mathrm{H}}(U)$

This table has two columns and seven panels separated by horizontal lines. Each column presents a causal model. The first panel names the models. The second panel presents the structural equations generating the models. In this row alone we make $\boldsymbol{\epsilon}$ explicit. In the other rows it is kept implicit to avoid clutter. Columns 1 and 2 are based on structural equations that have the same functional form, but have different inputs. The third panel represents the model as a DAG. Squares represent observed variables, circles represent unobserved variables. The fourth panel presents the parents in $\mathcal{T}$ of each variable. The fifth panel shows the conditional independence relationships generated by the application of the Local Markov Condition. The sixth panel presents the factorization of the joint distribution of variables in the Bayesian Network. The last panel of column 1 presents the joint distribution of variables when X is *fixed* at x. This entails a thought experiment implicit in Haavelmo (1943). The last panel of column 2 gives the joint distribution of variables generated by the hypothetical model when $\tilde{X}$ is conditioned at value $\tilde{X} = x$.

3.1 The Hypothetical Model

Our approach is based on a hypothetical model that is used to study causal effects. To recall, we use the term *empirical model* to designate the data generating process and the term *hypothetical model* to designate the model used to characterize causal effects.

The hypothetical model is based on the empirical model. It shares the same structural equations and same distribution of error terms as the empirical model. The hypothetical model differs from the empirical model in two ways. First, it appends to the empirical model an external variable (or a set of external variables) termed a hypothetical variable(s). Second, it replaces the action of existing inputs. If $X \in \mathcal{T}$ is the target variable to be fixed in the empirical model, then the newly created hypothetical variable $\tilde{X}$ replaces the X-input of one, some or all variables in $Ch(X)$. In other words, children of X in the empirical model will have their X-input replaced by a $\tilde{X}$-input in the hypothetical model. We assume that X and $\tilde{X}$ have common supports.

Table 1 illustrates the concept of a hypothetical model using the Haavelmo model introduced in Section 2. Column 2 presents the hypothetical model associated with the Haavelmo empirical model presented in the first column.

For the sake of clarity, we use G_{E} for the DAG representing the empirical model and $\mathcal{T}_{\mathrm{E}}$ for its associated set of variables. We use $Pa_{\mathrm{E}}, D_{\mathrm{E}}, Ch_{\mathrm{E}}$ for the parents, descendants, and children with DAG G_{E}. We use $\mathbf{P}_{\mathrm{E}}$ for the probability measure of variables in $\mathcal{T}_{\mathrm{E}}$. For the corresponding counterparts in the hypothetical model we use $G_{\mathrm{H}}, \mathcal{T}_{\mathrm{H}}, Pa_{\mathrm{H}}, D_{\mathrm{H}}, Ch_{\mathrm{H}}$ and $\mathbf{P}_{\mathrm{H}}$.

We now list some salient features of the hypothetical model. Let $\tilde{X}$ denote the hypothetical variable (or variables) associated with $X \in \mathcal{T}_{\mathrm{E}}$. We expand the list of variables in the hypothetical model so that $\mathcal{T}_{\mathrm{H}} = \mathcal{T}_{\mathrm{E}} \cup \{\tilde{X}\}$. The hypothetical variable can replace some or all of the input X for variables in $Ch_{\mathrm{E}}(X)$, i.e., $Ch_{\mathrm{H}}(\tilde{X}) \subseteq Ch_{\mathrm{E}}(X)$. Children of X in the empirical model can be partitioned among X and $\tilde{X}$ in the hypothetical model: $Ch_{\mathrm{E}}(X) = Ch_{\mathrm{H}}(X) \cup Ch_{\mathrm{H}}(\tilde{X})$. [11] As a consequence we also have that $D_{\mathrm{E}}(X) = D_{\mathrm{H}}(X) \cup D_{\mathrm{H}}(\tilde{X})$, that is, X-descendants of the empirical model constitute the X and $\tilde{X}$ descendants in the hypothetical model. Parental sets of the hypothetical model are defined by $Pa_{\mathrm{H}}(V) = Pa_{\mathrm{E}}(V)\ \forall\ V \in \mathcal{T}_{\mathrm{E}} \backslash Ch_{\mathrm{H}}(\tilde{X})$ and $Pa_{\mathrm{H}}(V) = \{Pa_{\mathrm{E}}(V) \backslash \{X\}\} \cup \{\tilde{X}\}\ \forall\ V \in Ch_{\mathrm{H}}(\tilde{X})$. Moreover, $\tilde{X}$ is an external variable, that is, $Pa_{\mathrm{H}}(\tilde{X}) = \varnothing$. The hypothetical model is also a DAG. Thus LMC (5) holds and the joint distribution of the variables in $\mathcal{T}_{\mathrm{H}}$ can be factorized using equation (6). By sharing the same structural equations and distribution of error terms, the conditional probabilities of the hypothetical model can be written as:

$$\mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V)) = \mathbf{P}_{\mathrm{E}}(V|Pa_{\mathrm{E}}(V))\ \forall\ V \in \mathcal{T}_{\mathrm{E}} \setminus Ch_{\mathrm{H}}(\tilde{X}) \tag{8}$$

and

$$\mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V) \setminus \{\tilde{X}\}, \tilde{X} = x) = \mathbf{P}_{\mathrm{E}}(V|Pa_{\mathrm{E}}(V) \setminus \{X\}, X = x)\ \forall\ V \in Ch_{\mathrm{H}}(\tilde{X}). \tag{9}$$

Equations (8)–(9) arise because the distribution of a variable $V \in \mathcal{T}_{\mathrm{E}}$ conditional on its parents is determined by the distribution of its error terms, which is the same for hypothetical and empirical models.

We now link the probability measures of the empirical and hypothetical models. Theorem **T-1** uses LMC (5) and Equation (8) to show that the distribution of non-descendants of $\tilde{X}$ are the same in both hypothetical and empirical models:

Theorem T-1. Let $\tilde{X}$ be the hypothetical variable in the hypothetical model represented by G_{H} associated with variable X in empirical model G_{E}. Let W, Z be any disjoint set of

[11] As an example, let a simple empirical model for mediation analysis consist of three variables: an input variable X, a mediation variable M caused by X and an outcome of interest Y caused by X and M. This model is represented as a DAG in Model 1 of Table 2 and $Ch_{\mathrm{E}}(X) = \{M, Y\}$. Suppose we are interested in the indirect effect, that is the effect of X on Y that operates exclusively by changes in M while holding the distribution of X unaltered. The hypothetical model for the evaluation of the indirect causal effect assigns the causal link of X on M to the hypothetical variable $\tilde{X}$. Namely, X still causes Y, but $\tilde{X}$ causes M. This hypothetical model is represented by Model 3 of Table 2. In this model $Ch_{\mathrm{H}}(X) = \{Y\}$, $Ch_{\mathrm{H}}(\tilde{X}) = \{M\}$ and $Ch_{\mathrm{E}}(X) = Ch_{\mathrm{H}}(X) \cup Ch_{\mathrm{H}}(\tilde{X})$.

variables in $\mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X})$ then:

$$\mathbf{P}_{\mathrm{H}}(W|Z) = \mathbf{P}_{\mathrm{H}}(W|Z, \tilde{X}) = \mathbf{P}_{\mathrm{E}}(W|Z) \, \forall \, \{W, Z\} \subset \mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X}).$$

Proof. See Appendix □

Theorem **T-1** also holds for the set of variables that are non-descendants of X according to the empirical model, which are a subset of $\mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X})$. Thus $\mathbf{P}_{\mathrm{H}}(W|Z) = \mathbf{P}_{\mathrm{H}}(W|Z, \tilde{X}) = \mathbf{P}_{\mathrm{E}}(W|Z)$ for all $\{W, Z\} \subset \mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{E}}(X)$.

The following theorem uses Theorem **T-1** and Equations (8)–(9) to show that the distribution of variables conditional on X and $\tilde{X}$ taking the same value x in the hypothetical model is equal to the distribution of the variables conditional on $X = x$ in the empirical model:

Theorem T-2. Let $\tilde{X}$ be the hypothetical variable in the hypothetical model represented by G_{H} associated with variable X in empirical model G_{E} and let W, Z be any disjoint[12] set of variables in $\mathcal{T}_{\mathrm{E}}$ then:

$$\mathbf{P}_{\mathrm{H}}(W|Z, X = x, \tilde{X} = x) = \mathbf{P}_{\mathrm{E}}(W|Z, X = x) \, \forall \, \{W, Z\} \subset \mathcal{T}_{\mathrm{E}}.$$

Proof. See Appendix.[13] □

The hypothetical variable is created to have the desired independent variation to generate causal effects. As a consequence, the operation of fixing a variable in the empirical model is translated into statistical conditioning in the hypothetical model. In particular, if we replace

[12] Disjoint (i.e., distinct) from X

[13] We also note the following result:

Corollary C-1. Let $\tilde{X}$ be uniformly distributed in the support of X and let W, Z be any disjoint set of variables in $\mathcal{T}_{\mathrm{E}}$ then:

$$\mathbf{P}_{\mathrm{H}}(W|Z, X = \tilde{X}) = \mathbf{P}_{\mathrm{E}}(W|Z) \, \forall \, \{W, Z\} \subset \mathcal{T}_{\mathrm{E}}.$$

Proof. See Appendix. We thank an anonymous referee for suggesting this result and its proof. □

the X-input by a $\tilde{X}$-input for all children of X, as suggested by the operation of fixing or "doing," we have that the distribution of an outcome $Y \in \mathcal{T}_{\mathrm{E}}$ of the empirical model when variable X is fixed at x (for all its children) is equivalent to the distribution of Y conditioned on the hypothetical variable $\tilde{X}$ being assigned to value x. This is captured by the following theorem:

Theorem T-3. Let $\tilde{X}$ be the hypothetical variable in G_{H} associated with variable X in the empirical model G_{E}, such that $Ch_{\mathrm{H}}(\tilde{X}) = Ch_{\mathrm{E}}(X)$, then:

$$\mathbf{P}_{\mathrm{H}}(\mathcal{T}_{\mathrm{E}} \setminus \{X\} | \tilde{X} = x) = \mathbf{P}_{\mathrm{E}}(\mathcal{T}_{\mathrm{E}} \setminus \{X\} | do(X) = x).$$

Proof. See Appendix □

One benefit of the hypothetical model is its greater flexibility for the study of causal effects. While the do-operator targets all causal relationships involving a variable X, the hypothetical variable allows analysts to target causal relationships of X separately. Indeed we can choose which variable in $Ch(X)$ will be caused by a hypothetical variable $\tilde{X}$, which in turn replaces some of the X inputs. This flexibility facilitates the investigation of causal effects in models that examine different causal paths associated with a single input, such as mediation analysis.[14]

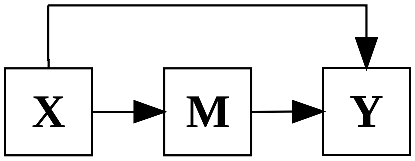
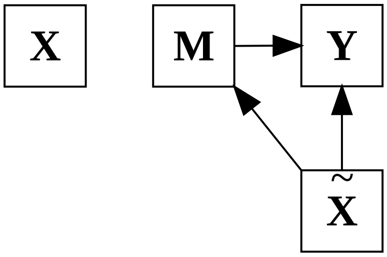
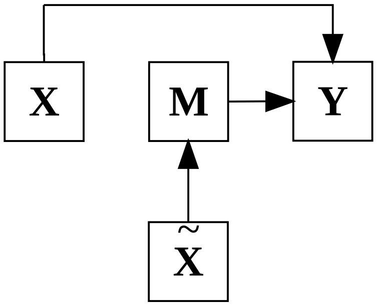
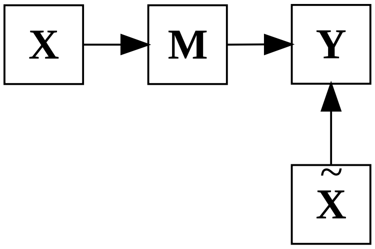
To show this, consider a simple empirical model for mediation consisting of three variables, an input variable X, a mediation variable M caused by X and an outcome of interest Y caused by X and M. Its structural equations are given by $Y = f_Y(X, M, \epsilon_Y)$, $M = f_M(X, \epsilon_M), X = f_X(\epsilon_X)$ and its DAG is represented as Model 1 of Table 2 (with $\boldsymbol{\epsilon}$ kept implicit). The total causal effect of X on Y when X is fixed at x compared to when it is fixed at x' is given by $TE(x, x') = \mathbb{E}_{\mathrm{E}}(Y(x) - Y(x'))$ where $Y(x) = f_Y(x, M(x), \epsilon_Y)$, $M(x) = f_M(x, \epsilon_M)$ and $\mathbb{E}_{\mathrm{E}}$ denotes the expectation over the probability measure of the empirical model. The hypothetical model for the evaluation of the total causal effect considers all

[14] Robins and Richardson (2011) analyze the mediation framework of Model 1 of Table 2 to examine four broad classes of graphical causal models.

causal links of X in the hypothetical variable $\tilde{X}$. It is represented by Model 2 of Table 2. Using this model, the total causal effect is given by $TE(x, x') = \mathbb{E}_{\mathrm{H}}(Y|\tilde{X} = x) - \mathbb{E}_{\mathrm{H}}(Y|\tilde{X} = x')$, where $\mathbb{E}_{\mathrm{H}}$ denotes the expectation over the probability measure of the hypothetical model in Model 2 of Table 2. Suppose that we are interested in the indirect effect, that is the effect of X on Y that operates exclusively by changing M while keeping the distribution of X unaltered from what it is in the empirical model, i.e., $\mathbb{E}_{\mathrm{E}}(f_Y(X, M(x), \epsilon_Y) - f_Y(X, M(x'), \epsilon_Y))$. This effect is $\mathbb{E}_{\mathrm{H}}(Y|\tilde{X} = x) - \mathbb{E}_{\mathrm{H}}(Y|\tilde{X} = x')$ derived from the hypothetical model presented in model 3 of Table 2. Hypothetical model 4 of Table 2 gives the causal graph for the direct effect of X on Y that operates exclusively through changes in X conditioning on M in the empirical model.[15]

[15] See Heckman and Pinto (2013) for further discussion of mediation models.

Table 2: **Models for Mediation Analysis**

1. Empirical Mediation Model	2. Hypothetical Model for Total Effect of X on Y
	
3. Hypothetical Model for Indirect Effect of X on Y	4. Hypothetical Model for Direct Effect of X on Y
	

This table shows four models represents by DAGs. To simplify the displays we keep the unobservables in $\boldsymbol{\epsilon}$ implicit. Model 1 represents the empirical model for mediation analysis. The remaining three models are hypothetical models that target different causal effects of X on Y. Model 2 represents the hypothetical model for the analysis of total effect of X on Y. Model 3 examines the indirect effect X on Y. Model 4 examines the direct effect of X on Y.

The hypothetical model does not suppress the variable we seek to fix, but rather creates a new hypothetical variable that allows us to examine a variety of causal effects. This approach provides a natural framework within which to examine counterfactual outcomes that involve both fixing and conditioning. Specifically, the expected value of an outcome Y when an input X is fixed at x conditional on $X = x'$ is readily defined by $\mathbb{E}_{\mathrm{H}}(Y|\tilde{X} = x, X = x')$ in the hypothetical model. By characterizing causality through a hypothetical model we avoid the necessity of defining new mathematical tools outside standard statistical analysis. The next section illustrates this point by identifying the causal effects of the "Front-Door" model of Pearl (2009) using his "*do-calculus*" and the standard statistical tools that can be used to analyze the hypothetical model.

The hypothetical model allows analysts to clearly distinguish the definition of causal effects from their identification in data. Causal effects are translated into statistical conditioning in the hypothetical model. Identification of causal effects requires analysts to relate the hypothetical and empirical distributions in a fashion that allows the evaluation of causal effects examined in the hypothetical model using data generated by the empirical model. For example, a standard technique for doing so is matching:

Lemma L-1. Matching: Let Z, W be any disjoint set of variables in $\mathcal{T}_{\text{E}}$ and let $\tilde{X}$ be a hypothetical variable in model G_{H} associated with $X \in \mathcal{T}_{\text{E}}$ in model G_{E} such that, in the hypothetical model, $X \perp\!\!\!\perp W|(Z, \tilde{X})$, then

$$\mathbf{P}_{\text{H}}(W|Z, \tilde{X} = x) = \mathbf{P}_{\text{E}}(W|Z, X = x).$$

Proof. See Mathematical Appendix. □

Variables Z of Lemma **L-1** are called matching variables. In statistical jargon, it is said that matching variables solve the problem of confounding effects between a treatment indicator X and outcome W. Matching is commonly used to identify treatment effects in propensity score matching models.[16] In these models, the conditional independence relation of Lemma **L-1** is assumed to be true. Pearl (1993) describes a graphical test called the "Back-Door" criterion that can be applied to a DAG in order to check if a set of variables satisfy the matching assumptions of Lemma **L-1**. The next section illustrates the use of Lemma **L-1**.

[16] See, e.g., Rosenbaum and Rubin (1983).

4 The Do Calculus and Haavelmo's Notation of Causality

To illustrate the points made in the previous section, we give the rules of the "*do-calculus*" and compare the identification strategies associated with the *do-calculus* with an approach using the hypothetical model. The *do-calculus*, developed by Pearl (1995), consists of three graphical and statistical rules that operate on the empirical model $\mathbf{P}_{\mathrm{E}}$ and that supplement standard statistics. In special cases they solve the problem of identifying causal effects in Bayesian Networks. The concept of a hypothetical model—central to the Frisch-Haavelmo approach—is not used in the literature on DAGs. It is commonly used to process the information of a causal model that can be represented by a DAG. Examples of this literature are Huang and Valtorta (2006, 2008) and Tian and Pearl (2002, 2003).

To review these methods, we introduce the graphical and statistical notation used to define the do-calculus. Let X, Y, Z be arbitrary disjoint sets of variables (nodes) in a causal graph G. $G_{\overline{X}}$ denotes a modification of DAG G obtained by deleting the arrows pointing to X, $G_{\underline{X}}$ denotes the modified DAG obtained by deleting the arrows emerging from X and $G_{\overline{X},\underline{Z}}$ denotes the DAG obtained by deleting arrows pointing to X and emerging from Z. Table 4 presents an example of the application of these rules for the Front-Door model, which is described in Table 3. The Front-Door model is described in greater detail in the next Section.

Let G be a DAG and let X, Y, Z, W be any disjoint sets of variables. The do-calculus rules are:

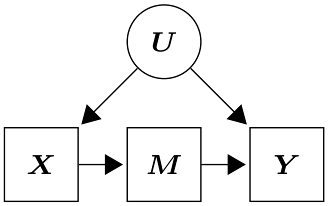
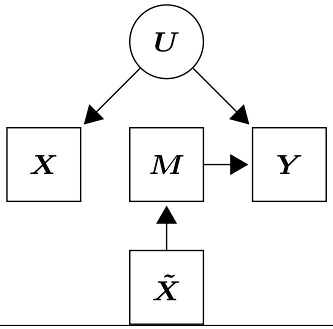
- **Rule 1:** Insertion/deletion of observations:

  $Y \perp\!\!\!\perp Z|(X, W)$ under $G_{\overline{X}} \Rightarrow \mathbf{P}(Y|do(X), Z, W) = \mathbf{P}(Y|do(X), W)$.

- **Rule 2:** Action/observation exchange:

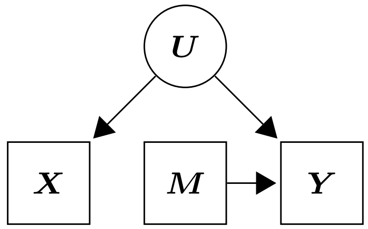
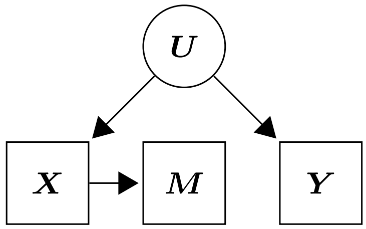
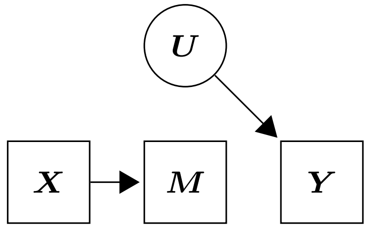
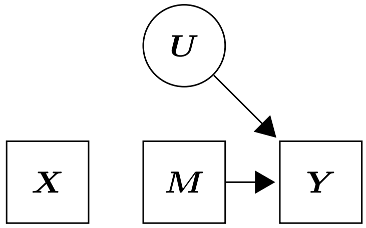
  $Y \perp\!\!\!\perp Z|(X, W)$ under $G_{\overline{X},\underline{Z}} \Rightarrow \mathbf{P}(Y|do(X), do(Z), W) = \mathbf{P}(Y|do(X), Z, W)$.

Table 3: **"Front-Door" Empirical and Hypothetical Models**

1. Pearl's "Front-Door" Empirical Model	2. Our Version of the "Front-Door" Hypothetical Model
$\mathcal{T} = \{U, X, M, Y\}$ $\boldsymbol{\epsilon} = \{\epsilon_U, \epsilon_X, \epsilon_M, \epsilon_Y\}$ $Y = f_Y(M, U, \epsilon_Y)$ $X = f_X(U, \epsilon_X)$ $M = f_M(X, \epsilon_M)$ $U = f_U(\epsilon_U)$	$\mathcal{T} = \{U, X, M, Y, \tilde{X}\}$ $\boldsymbol{\epsilon} = \{\epsilon_U, \epsilon_X, \epsilon_M, \epsilon_Y\}$ $Y = f_Y(M, U, \epsilon_Y)$ $X = f_X(U, \epsilon_X)$ $M = f_M(\tilde{X}, \epsilon_M)$ $U = f_U(\epsilon_U)$
	
$Pa(U) = \varnothing,$ $Pa(X) = \{U\}$ $Pa(M) = \{X\}$ $Pa(Y) = \{M, U\}$	$Pa(U) = Pa(\tilde{X}) = \varnothing,$ $Pa(X) = \{U\}$ $Pa(M) = \{\tilde{X}\}$ $Pa(Y) = \{M, U\}$
$Y \perp\!\!\!\perp X \| (M, U)$ $M \perp\!\!\!\perp U \| X$	$Y \perp\!\!\!\perp (\tilde{X}, X) \| (M, U)$ $M \perp\!\!\!\perp (U, X) \| \tilde{X}$ $X \perp\!\!\!\perp (M, \tilde{X}, Y) \| U$ $U \perp\!\!\!\perp (M, \tilde{X})$ $\tilde{X} \perp\!\!\!\perp (X, U)$
$\mathbf{P}_{\mathrm{E}}(Y, M, X, U) =$ $\mathbf{P}_{\mathrm{E}}(Y \| M, U)\,\mathbf{P}_{\mathrm{E}}(X \| U)\,\mathbf{P}_{\mathrm{E}}(M \| X)\,\mathbf{P}_{\mathrm{E}}(U)$	$\mathbf{P}_{\mathrm{H}}(Y, M, X, U, \tilde{X}) =$ $\mathbf{P}_{\mathrm{H}}(Y \| M, U)\,\mathbf{P}(X \| U)\,\mathbf{P}_{\mathrm{H}}(M \| \tilde{X})\,\mathbf{P}_{\mathrm{H}}(U)\,\mathbf{P}_{\mathrm{H}}(\tilde{X})$
$\mathbf{P}_{\mathrm{E}}(Y, M, U \| do(X) = x) =$ $\mathbf{P}_{\mathrm{E}}(Y \| M, U)\,\mathbf{P}_{\mathrm{E}}(M \| X = x)\,\mathbf{P}_{\mathrm{E}}(U)$	$\mathbf{P}_{\mathrm{H}}(Y, M, U, X \| \tilde{X} = x) =$ $\mathbf{P}_{\mathrm{H}}(Y \| M, U)\,\mathbf{P}(X \| U)\,\mathbf{P}_{\mathrm{H}}(M \| \tilde{X} = x)\,\mathbf{P}_{\mathrm{H}}(U)$

This table has two columns and seven panels separated by horizontal lines. Each column presents a causal model. The first panel names the models. The second panel presents the structural equations generating the model. In this row alone we make $\boldsymbol{\epsilon}$ explicit. In the other it is kept implicit to avoid clutter. Columns 1 and 2 are based on structural equations that have the same functional form, but have different inputs. The third panel represents the model as a DAG. Squares represent observed variables, circles represent unobserved variables. The fourth panel presents the parents in $\mathcal{T}$ of each variable. The fifth panel shows the conditional independence relationships generated by the application of the Local Markov Condition. The sixth panel presents the factorization of the joint distribution of variables in the Bayesian Network. The last panel of column 1 presents the joint distribution of variables when X is fixed at x using the "do operator." The last panel of column 2 gives the joint distribution of variables generated by the hypothetical models associated with empirical model 1 when $\tilde{X}$ is conditioned at $\tilde{X} = x$.

Table 4: **Do-calculus and the Front-Door Model**

1. Modified Front-Door Model $G_{\underline{X}} = G_{\overline{M}}$	2. Modified Front-Door Model $G_{\underline{M}}$
	
$(Y, M) \perp\!\!\!\perp X\|U$ $(X, U) \perp\!\!\!\perp M$	$(X, M) \perp\!\!\!\perp Y\|U$ $(Y, U) \perp\!\!\!\perp M\|X$
3. Modified Front-Door Model $G_{\overline{X},\underline{M}}$	4. Modified Front-Door Model $G_{\overline{X},\overline{M}}$
	
$(X, M) \perp\!\!\!\perp (Y, U)$	$(Y, M, U) \perp\!\!\!\perp X$ $U \perp\!\!\!\perp M$

This table shows four models represented by DAGs ($\boldsymbol{\epsilon}$ are kept implicit to avoid notational clutter). Squares represent observed variables, circles represent unobserved variables. Each DAG is generated by the deletion of arrows of the original Front-Door model (first column of Table 3) according to the rules of the do-calculus. Below each model, we show conditional independent relations generated by the application of the Local Markov Condition (5) to variables of the models.

- **Rule 3:** Insertion/deletion of actions:

  $Y \perp\!\!\!\perp Z|(X, W)$ under $G_{\overline{X},\overline{Z(W)}} \Rightarrow \mathbf{P}(Y|do(X), do(Z), W) = \mathbf{P}(Y|do(X), W)$,

  where $Z(W)$ is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

These rules are intended to supplement standard statistical tools with a new set of "do" operations. We illustrate the use of the do-calculus in the next section.

4.1 Comparing Analyses Based on the Do-calculus with those from the Hypothetical Model

We compare the do-calculus and an analysis based on our hypothetical model by identifying the causal effects of Pearl's "Front-Door model". That model consists of four variables: (1) an external unobserved variable U; (2) an observed variable X caused by U; (3) an observed variable M caused by X; and (4) an outcome Y caused by U and M. The Front-Door model is presented in the first column of Table 3.

We are interested in identifying the distribution of the outcome Y when X is fixed at x. By identification we mean expressing the quantity $\mathbf{P}(Y|do(X))$ in terms of the distribution of observed variables.

The do-calculus identifies $\mathbf{P}(Y|do(X))$ through four steps which we now perform. Steps 1, 2 and 3 identify $\mathbf{P}(M|do(X))$, $\mathbf{P}(Y|do(M))$ and $\mathbf{P}(Y|M, do(X))$ respectively. Step 4 uses the first three steps to identify $\mathbf{P}(Y|do(X))$.

1. Invoking LMC (5) for variable M of DAG $G_{\underline{X}}$, (DAG 1 of Table 4) generates $X \perp\!\!\!\perp M$. Thus, by Rule 2 of the do-calculus, we obtain $\mathbf{P}(M|do(X)) = \mathbf{P}(M|X)$.

2. Invoking LMC (5) for variable M of DAG $G_{\overline{M}}$, (DAG 1 of Table 4) generates $X \perp\!\!\!\perp M$. Thus, by Rule 3 of the do-calculus, $\mathbf{P}(X|do(M)) = \mathbf{P}(X)$. In addition, applying LMC (5) for variable M of DAG $G_{\underline{M}}$, (DAG 2 of Table 4) generates $M \perp\!\!\!\perp Y|X$. Thus, by Rule 2 of do-calculus, $\mathbf{P}(Y|X, do(M)) = \mathbf{P}(Y|X, M)$.

$$\text{Therefore } \mathbf{P}(Y|do(M)) = \sum_{x' \in \mathrm{supp}(X)} \mathbf{P}(Y|X = x', do(M))\, \mathbf{P}(X = x'|do(M))$$
$$= \sum_{x' \in \mathrm{supp}(X)} \mathbf{P}(Y|X = x', M)\, \mathbf{P}(X = x'),$$

   where "supp" means support.

3. Invoking LMC (5) for variable M of DAG $G_{\overline{X},\underline{M}}$, (DAG 3 of Table 4) generates

$Y \perp\!\!\!\perp M|X$. Thus, by Rule 2 of the do-calculus, $\mathbf{P}(Y|M, do(X)) = \mathbf{P}(Y|do(M), do(X))$. In addition, applyingLMC (5) for variable X of DAG $G_{\overline{X},\overline{M}}$, (DAG 4 of Table 4) generates $(Y, M, U) \perp\!\!\!\perp X$. By weak union and decomposition, we obtain $Y \perp\!\!\!\perp X|M$. Thus by Rule 3 of the do-calculus, we obtain that $\mathbf{P}(Y|do(X), do(M)) = \mathbf{P}(Y|do(M))$. Thus $\mathbf{P}(Y|M, do(X)) = \mathbf{P}(Y|do(M), do(X)) = \mathbf{P}(Y|do(M))$.

4. We collect the results from the three previous steps to identify $\mathbf{P}(Y|do(X))$ from observed data:

$$
\begin{aligned}
&\mathbf{P}(Y|do(X) = x) \\
&\quad = \sum_{m \in \text{supp}(M)} \mathbf{P}(Y|M, do(X) = x)\, \mathbf{P}(M|do(X) = x) \\
&\quad = \sum_{m \in \text{supp}(M)} \underbrace{\mathbf{P}(Y|do(M) = m, do(X) = x)}_{\textbf{Step 3}} \mathbf{P}(M = m|do(X) = x) \\
&\quad = \sum_{m \in \text{supp}(M)} \underbrace{\mathbf{P}(Y|do(M) = m)}_{\textbf{Step 3}} \mathbf{P}(M = m|do(X) = x) \\
&\quad = \sum_{m \in \text{supp}(M)} \underbrace{\left(\sum_{x' \in \text{supp}(X)} \mathbf{P}(Y|X = x', M)\, \mathbf{P}(X = x') \right)}_{\textbf{Step 2}} \underbrace{\mathbf{P}(M = m|X = x)}_{\textbf{Step 1}}.
\end{aligned}
$$

In this fashion, we use the do-calculus to identify the desired causal parameter. It is instructive to compare this proof of identification with one based on the approach of Haavelmo. We identify the causal effects of X on Y for the Front-Door model using a hypothetical model. We replace the relationship of X on M by a hypothetical variable $\tilde{X}$ that causes M. We use $\mathbf{P}_{\text{E}}$ to denote the probability of the Front-Door model that generates the data (Column 1 of Table 3) and $\mathbf{P}_{\text{H}}$ for the hypothetical model (Column 2 of Table 3). As before, we seek to identify $\mathbf{P}_{\text{H}}(Y|\tilde{X})$ (the equivalent of $\mathbf{P}(Y|do(X))$) from observed distributions in the empirical model.

We first present a lemma that states three useful conditional independence relations of the hypothetical model. The lemma is based on the application of LMC (5) and the Graphoid relationships:

Lemma L-2. In the Front-Door hypothetical model, (1) $Y \perp\!\!\!\perp \tilde{X}|M$, (2) $X \perp\!\!\!\perp M$, and (3) $Y \perp\!\!\!\perp \tilde{X}|(M, X)$

Proof. By LMC (5) for X, we obtain $(Y, M, \tilde{X}) \perp\!\!\!\perp X|U$. By LMC (5) for Y we obtain $Y \perp\!\!\!\perp (X, \tilde{X})|(M, U)$. By Contraction applied to $(Y, M, \tilde{X}) \perp\!\!\!\perp X|U$ and $Y \perp\!\!\!\perp (X, \tilde{X})|(M, U)$ we obtain $(Y, X) \perp\!\!\!\perp \tilde{X}|(M, U)$. By LMC (5) for U we obtain $(M, \tilde{X}) \perp\!\!\!\perp U$. By Contraction applied to $(M, \tilde{X}) \perp\!\!\!\perp U$ and$(Y, M, \tilde{X}) \perp\!\!\!\perp X|U$ we obtain$(X, U) \perp\!\!\!\perp (M, \tilde{X})$. The second relationship in the Lemma is obtained by Decomposition. In addition, by Contraction on $(Y, X) \perp\!\!\!\perp \tilde{X}|(M, U)$ and $(M, \tilde{X}) \perp\!\!\!\perp U$ we obtain $(Y, X, U) \perp\!\!\!\perp \tilde{X}|M$. The two remaining conditional independence relationships of the Lemma are obtained by Weak Union and Decomposition.[17] □

[17] One can also prove Lemma **L-2** using Pearl's *d-Separation* criteria. According to Pearl (2009), a path p connecting X and Y is said to be *d-Separated* (or blocked) by a set of nodes Z if and only if

1. a path p contains a chain $i \to m \to j$ or a fork $i \leftarrow m \to j$ such that the middle node m is in Z, or
2. a path p contains an inverted fork (or collider) $i \to m \leftarrow j$ such that the middle node m is not in Z and such that no descendant of m is in Z.

A set Z is said to *d-separate* X from Y if and only if Z blocks every path from a node in X to a node in Y. If X and Y are d-Separated by Z according to a graph G, then $Y \perp\!\!\!\perp Y|Z$ in G. We are examining the Hypothetical Model described by second column of Table 2. Variables Y and $\tilde{X}$ are connected by a single path $\tilde{X} \to M \to Y$. Thus we have that $Y \perp\!\!\!\perp \tilde{X}|M$, according to part 1 of the d-Separation criteria. Moreover, we can also state that $Y \perp\!\!\!\perp \tilde{X}|(M, X)$ as X is not a collider nor a decendant of a collider (part 2 of the d-Separation criteria). Finally, there is no path that connects X and M of the form $X \to \ldots \to M$ nor $X \leftarrow \ldots \leftarrow M$. Thus we can state that $X \perp\!\!\!\perp M$ according to part 1 of the d-Separation criteria.

Applying these results,

$$\begin{aligned}
&\mathbf{P}_{\mathrm{H}}(Y|\tilde{X}=x)\\
&\quad= \sum_{m\in\mathrm{supp}(M)} \mathbf{P}_{\mathrm{H}}(Y|M=m,\tilde{X}=x)\,\mathbf{P}_{\mathrm{H}}(M=m|\tilde{X}=x)\\
&\quad= \sum_{m\in\mathrm{supp}(M)} \mathbf{P}_{\mathrm{H}}(Y|M=m)\,\mathbf{P}_{\mathrm{H}}(M=m|\tilde{X}=x)\\
&\quad= \sum_{m\in\mathrm{supp}(M)} \left(\sum_{x'\in\mathrm{supp}(X)} \mathbf{P}_{\mathrm{H}}(Y|X=x',M=m)\,\mathbf{P}_{\mathrm{H}}(X=x'|M=m)\right)\mathbf{P}_{\mathrm{H}}(M=m|\tilde{X}=x)\\
&\quad= \sum_{m\in\mathrm{supp}(M)} \left(\sum_{x'\in\mathrm{supp}(X)} \mathbf{P}_{\mathrm{H}}(Y|X=x',M=m)\,\mathbf{P}_{\mathrm{H}}(X=x')\right)\mathbf{P}_{\mathrm{H}}(M=m|\tilde{X}=x)\\
&\quad= \sum_{m\in\mathrm{supp}(M)} \left(\sum_{x'\in\mathrm{supp}(X)} \mathbf{P}_{\mathrm{H}}(Y|X=x',\tilde{X}=x',M=m)\,\mathbf{P}_{\mathrm{H}}(X=x')\right)\mathbf{P}_{\mathrm{H}}(M=m|\tilde{X}=x)\\
&\quad= \sum_{m\in\mathrm{supp}(M)} \left(\sum_{x'\in\mathrm{supp}(X)} \underbrace{\mathbf{P}_{\mathrm{E}}(Y|M,X=x')}_{\text{by Theorem T-2}}\,\underbrace{\mathbf{P}_{\mathrm{E}}(X=x')}_{\text{by Theorem T-1}}\right)\underbrace{\mathbf{P}_{\mathrm{E}}(M=m|X=x)}_{\text{by Matching L-1}}.
\end{aligned}$$

The second equality comes from relationship (1) $Y \perp\!\!\!\perp \tilde{X}|M$ of Lemma **L-2**. The fourth equality comes from relationship (2) $X \perp\!\!\!\perp M$ of Lemma **L-2**. The fifth equality comes from relationship (3) $Y \perp\!\!\!\perp \tilde{X}|(M,X)$ of Lemma **L-2**. The last equality links the distributions of the hypothetical model with the ones of the empirical model. The first term uses Theorem **T-2** to equate $\mathbf{P}_{\mathrm{H}}(Y|X=x',\tilde{X}=x',M=m)=\mathbf{P}_{\mathrm{E}}(Y|M,X=x')$. The second term uses the fact that X is not a child of $\tilde{X}$, thus by Theorem **T-1**, $\mathbf{P}_{\mathrm{H}}(X=x')=\mathbf{P}_{\mathrm{E}}(X=x')$. Finally, the last term uses Matching applied to M. Namely, LMC (5) for M generates $M \perp\!\!\!\perp X|\tilde{X}$ in the hypothetical model. Then, by Matching **L-1**, $\mathbf{P}_{\mathrm{H}}(M|\tilde{X}=x)=\mathbf{P}_{\mathrm{E}}(M|X=x)$.

It is clear from this example that, even though both frameworks produce the same final identification formula, the methods underlying them differ greatly. A key concept in the framework inspired by Haavelmo is the notion of a hypothetical model. Hypothetical models are the essential ingredients of science. Using this specification, identification is secured using the standard statistical tools involving the rules of conditional probability distributions. LMC and the graphoid relations generate conditional independence relationships that arise from the hypothetical model. Identification using the hypothetical model is transparent, and does not require additional causal rules.[18]

5 The Benefits and Limitations of DAGs

A major benefit of a DAG is its intuitive description of models as causal chains. DAG assumptions list the variables in a model and their causal relationships. A DAG does not generate or characterize any restrictions on functional forms or parametric specifications of the structural equations. In this sense, if an identification result is achieved, it is obtained under very weak conditions.

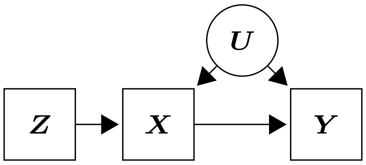
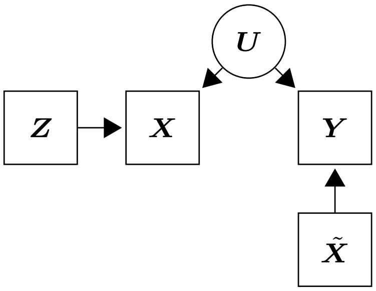
The generality of a DAG is also the source of its limitation. Methods that focus on identification of models solely described by DAGs lack the tools for invoking additional assumptions that would generate the identification of an *a priori* non-identified model. We clarify this point by considering the instrumental variable model.

The simplest instrumental variable model consists of four variables: (1) a confounding variable U that is external and unobserved; (2) an external instrumental variable Z; (3) an observed variable X caused by U and Z; and (4) an outcome Y caused by U and X. The empirical instrumental variable model is described in the first column of Table 5. Its hypothetical counterpart is presented in the second column of Table 5.

The instrumental variable method is a fundamental ingredient of a huge literature on econometric identification (see, e.g., Matzkin, 2013). It is the basis for more sophisticated models such as the Generalized Roy model, which is widely used in econometrics in the analysis of selection bias and in evaluating social programs (Heckman, 1976, 1979, Heckman and Robb, 1985, Powell, 1994, Heckman and Vytlacil, 2007a,b). Examples of this literature are nonparametric control functions (see, e.g., Blundell and Powell, 2003) and identification through instrumental variables (Reiersöl, 1945).

[18] Pearl (2009) also considers a "Back Door model" and applies do-calculus to identify a model that can readily be defined by the Haavelmo approach and identified using conventional matching methods.

Table 5: **Instrumental Variable Empirical and Hypothetical Models**

1. Instrumental Variable Empirical Model	2. Instrumental Variable Hypothetical Model
$\mathcal{T} = \{U, X, Z, Y\}$ $\boldsymbol{\epsilon} = \{\epsilon_U, \epsilon_X, \epsilon_Z, \epsilon_Y\}$ $Y = g_Y(X, U, \epsilon_Y)$ $X = g_X(U, Z, \epsilon_X)$ $Z = g_Z(\epsilon_Z)$ $U = g_U(\epsilon_U)$	$\mathcal{T} = \{U, X, Z, Y, \tilde{X}\}$ $\boldsymbol{\epsilon} = \{\epsilon_U, \epsilon_X, \epsilon_Z, \epsilon_Y\}$ $Y = g_Y(\tilde{X}, U, \epsilon_Y)$ $X = g_X(U, Z, \epsilon_X)$ $Z = g_Z(\epsilon_Z)$ $U = g_U(\epsilon_U)$
	
$Pa(U) = Pa(Z) = \varnothing,$ $Pa(X) = \{U, Z\}$ $Pa(Y) = \{U, X\}$	$Pa(U) = Pa(U) = \varnothing,$ $Pa(X) = \{U, Z\}$ $Pa(Y) = \{U, \tilde{X}\}$
$Y \perp\!\!\!\perp Z \vert (X, U)$ $Z \perp\!\!\!\perp U$	$Y \perp\!\!\!\perp (X, Z) \vert (\tilde{X}, U)$ $Z \perp\!\!\!\perp (U, Y, \tilde{X}) \vert (\tilde{X}, U)$ $X \perp\!\!\!\perp (Y, \tilde{X}) \vert (Z, U)$ $U \perp\!\!\!\perp (Z, \tilde{X})$ $\tilde{X} \perp\!\!\!\perp (U, X, Z)$
$\mathbf{P}_{\mathrm{E}}(Y, Z, X, U) =$ $\mathbf{P}_{\mathrm{E}}(Y \vert X, U)\, \mathbf{P}_{\mathrm{E}}(X \vert U, Z)\, \mathbf{P}_{\mathrm{E}}(Z)\, \mathbf{P}_{\mathrm{E}}(U)$	$\mathbf{P}_{\mathrm{H}}(Y, Z, X, U, \tilde{X}) =$ $\mathbf{P}_{\mathrm{H}}(Y \vert \tilde{X}, U)\, \mathbf{P}_{\mathrm{H}}(X \vert U, Z)\, \mathbf{P}_{\mathrm{H}}(Z)\, \mathbf{P}_{\mathrm{H}}(U)\, \mathbf{P}_{\mathrm{H}}(\tilde{X})$
$\mathbf{P}_{\mathrm{E}}(Y, Z, U \vert do(X) = x) =$ $\mathbf{P}_{\mathrm{E}}(Y \vert X = x, U)\, \mathbf{P}_{\mathrm{E}}(Z)\, \mathbf{P}_{\mathrm{E}}(U)$	$\mathbf{P}_{\mathrm{H}}(Y, Z, X, U \vert \tilde{X} = x) =$ $\mathbf{P}_{\mathrm{H}}(Y \vert \tilde{X} = x, U)\, \mathbf{P}_{\mathrm{H}}(X \vert U, Z)\, \mathbf{P}_{\mathrm{H}}(Z)\, \mathbf{P}_{\mathrm{H}}(U)$

This table has two columns and seven panels separated by horizontal lines. Each column presents a causal model. The first panel names the model. The second panel presents the structural equations generating the model. In this row alone we make the $\boldsymbol{\epsilon}$ explicit. In the other rows it is kept implicit to avoid notational clutter. Columns 1 and 2 are based on structural equations that have the same functional form, but have different inputs. The third panel represents the model as a DAG. Squares represent observed variables and circles represent unobserved variables. The fourth panel presents the parents in $\mathcal{T}$ of each variable. The fifth panel shows the conditional independence relationships generated by the application of the Local Markov Condition. The sixth panel presents the factorization of the joint distribution of variables in the Bayesian Network. The last panel of column 1 presents the joint distribution of variables when X is fixed at x. (do$(X) = x$). The last panel of column 2 gives the joint distribution of variables generated by hypothetical models associated with empirical model 1 when $\tilde{X}$ is conditioned on $\tilde{X} = x$.

Chapters 3 and 5 of Pearl (2009) show that the instrumental variable model is not identified applying the rules of the do-calculus. Indeed, it is impossible to identify the causal effect of X on Y without additional information.

The non-identification of the instrumental variable model poses a major limitation for the identification literature that relies exclusively on DAGs. Identification of the instrumental variable model relies on assumptions outside the scope of the DAG literature. For example, we can use LMC (5) to obtain the following conditional independence relationships: $Y \perp\!\!\!\perp Z|(U, X)$ and $U \perp\!\!\!\perp Z$. These relationships in addition to $X \not\perp\!\!\!\perp Z$ satisfy the necessary criteria to apply the method of Two Stage Least Squares (TSLS). TSLS identifies the instrumental variable model under a linearity assumption. As a consequence, if we assume that the causal relationship of X and U on outcome Y are represented by a linear equation, i.e., $Y = X\beta + U$, then it is well-known that parameters β can be identified using $\operatorname{cov}(Z, Y)/\operatorname{cov}(Z, X)$ under standard rank conditions.

Linearity and homogeneity of the effects of X on Y across agents (i.e., β is the same across the values X, U take) are strong assumptions about the causal links that govern the relationship between Y and X. This assessment fostered a huge literature in economics devoted to methods that relax linearity and homogeneity and that allow coefficients to be correlated with regressors. Examples of this literature are Imbens and Angrist (1994), Vytlacil (2002), and Heckman and Vytlacil (2005, 2007a,b), who identify the instrumental variable model under more general conditions by making assumptions on the causal relationship of Z with X. Imbens and Angrist (1994) show that the instrumental variable model can be identified under a "monotonicity" assumption (increasing the values of an instrument has the same qualitative effect on all agents). Vytlacil (2002) shows that this assumption is equivalent to assuming an instrumental variable model in which the treatment assignment decision rule is separable in terms of unobserved characteristics of the agents and the instrumental variable. Heckman and Vytlacil (1999, 2005, 2007a,b) develop and apply this result.

Table 6 summarizes the common and distinct features of Pearl's do-calculus and the approach based on Haavelmo's hypothetical model. Both approaches use structural equation models in the sense of Koopmans and Reiersøl (1950). Both invoke autonomy and assume mutually independent errors $\boldsymbol{\epsilon}$. In recursive models, both use the Local Markov Condition and the Graphoid axioms. Both use "fixing" or the "do operator" to define counterfactuals.

Table 6: Summarizing the Do-calculus of Pearl (2009) and Haavelmo's Framework

Common Features of Haavelmo and Do-Calculus:		
Autonomy (Frisch, 1938) **Errors Terms:** $\boldsymbol{\epsilon}$ mutually independent **Statistical Tools:** LMC and Graphoid Axioms apply **Counterfactuals:** Fixing or Do-operator is a Causal, not statistical, operation		
Distinctive Features of Haavelmo and Do-Calculus:		
	Haavelmo	**Do-calculus**
Approach:	Thinks Outside the Box of the Empirical Model	Applies Complex Tools
Introduces:	Constructs a Hypothetical Model	Graphical Rules
Identification:	Connects P_H and P_E	Iteration of Do-Calculus Rules
Versatility:	Basic Statistical Principles Apply	Creates New Rules of Statistics

The approaches diverge in their analyses of identification. The approach based on Haavelmo creates a hypothetical variable $\tilde{X}$ and an associated hypothetical model that is "outside the box" of the empirical model. It applies standard probability calculus to the hypothetical model to connect the hypothetical model to the empirical model. Pearl's do-calculus creates a new set of extra-statistical tools to identify the causal parameters created by fixing or the "do operator." Our analysis shows that in the hypothetical model of Haavelmo, the special extra-statistical tools of the do-calculus are not required to identify causal parameters. The econometric approach identifies a broader range of models that cannot be identified using the rules of the "do-calculus."

6 Hypothetical Models and Simultaneous Equations

The literature on causality provides a framework for modeling causal processes that are based on DAGs. Less is known about Directed Cyclic Graphs (DCGs) that are used to represent Simultaneous Equations. Indeed, the fundamental Local Markov Condition no longer holds for DCGs (Spirtes, 1995). Nevertheless, the notion of fixing readily extends to a system of simultaneous equations.

Consider a system of two equations:

$$Y_1 = g_{Y_1}(Y_2, X_1, U_1), \tag{7a}$$

$$Y_2 = g_{Y_2}(Y_1, X_2, U_2). \tag{7b}$$

$\mathcal{T}_{\mathrm{E}} = \{Y_1, Y_2, X_1, X_2, U_1, U_2\}$. Our analysis can be readily generalized to systems with more than two equations, but for the sake of brevity, we focus on the two-equation case. To simplify notation, we keep $\boldsymbol{\epsilon}$ implicit.

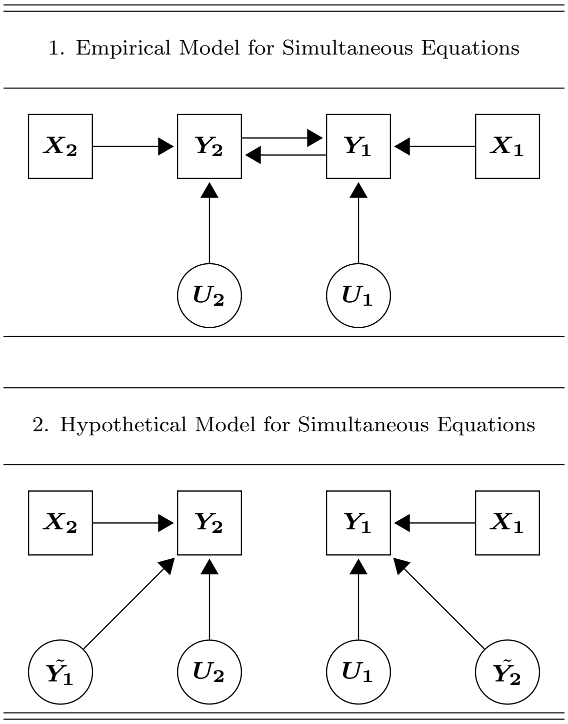
The empirical Simultaneous Equations Model of (7a) and (7b) is represented as Model 1 of Table 7. Many different versions of this model appear in the literature. For simplicity, we assume $U_1 \perp\!\!\!\perp U_2$ and $(U_1, U_2) \perp\!\!\!\perp (X_1, X_2)$.[19]

The hypothetical model associated with the causal operation of fixing both Y_2 and Y_1 is represented in Model 2 of Table 7. Under autonomy, the causal effect of Y_2 on Y_1 when Y_2 is fixed at y_2 is given by $Y_1(y_2) = g_{Y_1}(y_2, X, U_1)$. Symmetrically, $Y_2(y_1) = g_{y_2}(y_1, X, U_2)$. We define hypothetical random variables $\tilde{Y}_1, \tilde{Y}_2$. They replace the Y_1, Y_2 inputs in Equations (7a) and (7b) in the same fashion as discussed in previous sections. $(\tilde{Y}_1, \tilde{Y}_2) \perp\!\!\!\perp (X_1, X_2, U_1, U_2)$; and $\tilde{Y}_1 \perp\!\!\!\perp \tilde{Y}_2$. $\mathcal{T}_{\mathrm{H}} = \{\tilde{Y}_1, \tilde{Y}_2, Y_1, Y_2, X_1, X_2, U_1, U_2\}$. We assume a common support for (Y_1, Y_2) and $(\tilde{Y}_1, \tilde{Y}_2)$.

[19] These assumptions are made to simplify the analysis. A large literature relaxes these assumptions and develops identification criteria for cases where $U_1 \not\perp\!\!\!\perp U_2$ and $(U_1, U_2) \not\perp\!\!\!\perp (X_1, X_2)$. The literature considers a variety of specifications (see Matzkin, 2008). We maintain the assumptions that $U_1 \perp\!\!\!\perp U_2$ and $(U_1, U_2) \perp\!\!\!\perp (X_1, X_2)$ for simplicity.

In the same fashion as in the model previously discussed, the distribution of Y_1 when Y_2 is fixed at y_2 is given by $\mathbf{P}_{\mathrm{H}}(Y_1|\tilde{Y}_2 = y_2)$. The average causal effect of Y_2 on Y_1 when Y_2 is fixed at the two values of y_2 and y_2' is given by $\mathbb{E}_{\mathrm{H}}(Y_1|\tilde{Y}_2 = y_2) - \mathbb{E}_{\mathrm{H}}(Y_1|\tilde{Y}_2 = y_2')$, where $\mathbb{E}_{\mathrm{H}}$ denotes expectation over the probability measure $\mathbf{P}_{\mathrm{H}}$ of the hypothetical model. The hypothetical variation of $\tilde{Y}_2$ corresponds to the standard Marshallian and Walrasian thought experiments in which quantities or prices are fixed to trace out demand and supply curves (see, e.g., Mas-Colell et al., 1995). A symmetric analysis produces the causal effect of Y_1 on Y_2. Thus we obtain the counterpart to the counterfactuals defined for the recursive models earlier in this paper.

Table 7: **Models for Simultaneous Equations**


1. Empirical Model for Simultaneous Equations

2. Hypothetical Model for Simultaneous Equations

This table shows two models. (The ϵ are kept implicit.) Model 1 represents the empirical model for Simultaneous Equations where Y_1 and Y_2 cause each other. Model 1 is cyclic, and hence it is not a DAG. Model 2 represents one possible hypothetical model associated with the empirical model for Simultaneous Equations. In Model 2, the hypothetical variable $\tilde{Y}_2$ is associated with the causal link of Y_2 on Y_1 of Model 1 and the hypothetical variable $\tilde{Y}_1$ is associated with the causal link of Y_1 on Y_2 of Model 1.

Under simultaneity, the graph for Model 1 is cyclic and the relationships that hold for

DAGs, such as the LMC (5), break down (Lauritzen and Richardson, 2002; Spirtes, 1995). Equations (7a) and (7b) cannot be represented as Directed Bayesian networks. The tools developed for DAGs do not directly apply and require modification. Equations (7a) and (7b) are fundamentally non-recursive and observed variables emerge from a feedback process.

A traditional assumption in the simultaneous equations literature is "completeness"—the existence of at least a local solution for Y_1 and Y_2 in terms of (X_1, X_2, U_1, U_2):

$$Y_1 = \phi_1(X_1, X_2, U_1, U_2), \tag{8a}$$

$$Y_2 = \phi_2(X_1, X_2, U_1, U_2).^{20} \tag{8b}$$

These are called "reduced form" equations (see, e.g., Matzkin, 2008, 2013). They inherit the autonomy properties of the structural equations.

The assumption of the existence of a reduced form is not innocuous even in the linear cases for continuous Y_1 and Y_2 analyzed by Haavelmo (1943, 1944) and the Cowles Foundation pioneers (see Koopmans et al., 1950). Heckman (1978), Tamer (2003), and Chesher and Rosen (2012) analyze the case in which Y_1 and Y_2 are discrete valued. Solutions (8a) and (8b) may not exist except under conditions given in those papers.[21] Alternatively, there may be multiple solutions giving rise to reduced form correspondences. In the case where no solutions exist, the model is incoherent as an equilibrium model unless additional assumptions are invoked. However, one can construct hypothetical models using Haavelmo's insights even in incoherent cases.[22]

[20] We use the term "completeness" in the sense of Koopmans et al. (1950); i.e., the existence of a local solution of equations (7a) and (7b). This concept is to be distinguished from the notion of completeness in the nonparametric IV literature (Newey and Powell, 2003) or in hypothesis testing (Lehmann and Romano, 2005).

[21] Linear probability model approximations to Equations (7a) and (7b), as advocated by Angrist and Pischke (2008), although widely used, are in general not autonomous. They can, however, be estimated and identified for incoherent models, creating the illusion of coherency through approximation error. See Heckman and MaCurdy (1985).

[22] This might be a conceptually unsatisfactory enterprise unless the data intended to be described by the model display disequilibrium cycling phenomena and a time sequence for the evolution of the system, e.g., $Y_1^{(t)}, Y_2^{(t+1)}, \ldots$, is postulated as functions of inputs where superscripts denote time-dated variables.

In addition, some frameworks for multivariate discrete data may not be sufficiently rich to distinguish correlation from causation. Heckman (1978) shows that log-linear models for discrete data used in statistics (see, e.g., Bishop et al., 1975) have too few parameters to make causal distinctions. He introduces a class of latent variable models in which such distinctions are possible.

Note further that even in models in which the reduced form equations are well defined, it is not possible, in general, to *simultaneously* vary $\tilde{Y}_1$ and $\tilde{Y}_2$ so that they (i) solve Equations (7a) and (7b) and (ii) also satisfy the requirement that $(\tilde{Y}_1, \tilde{Y}_2) \perp\!\!\!\perp (X_1, X_2, U_1, U_2)$. This is apparent from the reduced form equations (8a) and (8b) that, under completeness, the proposed variations must also satisfy. Nonetheless, $\tilde{Y}_2$ and $\tilde{Y}_1$ can be separately constructed to create hypothetical models corresponding to Equations (7a) and (7b) respectively. These equations exist as theoretical constructs independent of any particular equilibrium construct.[23]

Matzkin (2007, 2008, 2012, 2013) presents comprehensive and definitive treatments of alternative approaches for identifying simultaneous equations. Our analysis readily extends to systems with more than two equations, but for the sake of brevity we do not make the extension here.

[23] Under completeness, we can use a version of indirect least squares (ILS) to define causal parameters and identify them where the induced variation in $\tilde{Y}_1$ and $\tilde{Y}_2$ satisfy equilibrium conditions. Thus if X_1 and X_2 are disjoint, one can use ILS to identify from reduced form equations (8a) and (8b), assumed to be differentiable:

$$\underset{\text{(From 8a)}}{\frac{\partial Y_1}{\partial X_2}} = \frac{\partial g_{Y_1}(Y_2, X_1, U_1)}{\partial Y_2} \quad \underset{\text{(From 8b)}}{\frac{\partial Y_2}{\partial X_2}}$$

$$\frac{\frac{\partial Y_1}{\partial X_2}}{\frac{\partial Y_2}{\partial X_2}} = \frac{\frac{\partial \phi_1(\cdot)}{\partial X_2}}{\frac{\partial \phi_2(\cdot)}{\partial X_2}} = \frac{\partial g_{Y_1}(\cdot)}{\partial Y_2}$$

If X_1 and X_2 contain common elements, the method can be modified to use only the distinct elements in X_1 and X_2 in this analysis.

7 Summary and Conclusions

This paper examines Haavelmo's fundamental contributions to the study of causal inference. He produced the first formal analysis of the distinction between causation and correlation. He carefully distinguished the process of defining causality—a mental act that assigns hypothetical variation to inputs—from the act of identifying causal models from data. Haavelmo was remarkably clear about concepts that are still muddled in some quarters of statistics.[24]

Haavelmo shows us that causal effects of inputs on outputs are defined in abstract models that assign independent variation to inputs. He formalized Frisch's notion that causality is in the mind. We formalize his insight extending his analysis for linear models to more general models. This enables us to discuss causal concepts such as "fixing" using an intuitive approach that applies Haavelmo's ideas.

Following Haavelmo, we distinguish the definition of causal parameters from their identification. Our approach to defining causality relies on the assumption of autonomy joined with Haavelmo's notion of hypothetical random variables. Together they enable us to express the distribution of counterfactual outcomes using structural equations and the distributions of the data by replacing the variables whose causal effects we seek to establish with their hypothetical counterparts. Causal models thus defined apply standard statistical tools and do not require new procedures like the do-calculus that lie outside the scope of the standard tools of probability and statistics.

Identification in Haavelmo's model is achieved in recursive models by applying standard statistical tools to Bayesian Networks. We link the distributions of empirical and hypothetical models by expressing the quantities of interest in the hypothetical model into observed quantities in the empirical one.

We illustrate the benefits of Haavelmo's approach by comparing identification of the causal effects of Pearl's flagship Front-Door model (Pearl, 2009) using a method based on

[24] See, e.g., Holland (1986) and Sobel (2005) for examples of the confusion between models and identification strategies exemplified by the claim that no causal statements are possible unless persons are randomly assigned to treatment.

the Haavelmo approach and a method based on the do-calculus of Pearl (2009). While both methods generate the same estimator, the identification methods differ on both conceptual and methodological grounds. We discuss the limitations of methods of identification that rely on the fundamentally recursive approach of Directed Acyclic Graphs.

That framework cannot accommodate the fundamentally non-recursive framework of the simultaneous equations model without violating autonomy. We consider causality in the simultaneous equations model developed in the seminal research of Haavelmo (1943, 1944). The framework of simultaneous equations is fundamentally non-recursive and falls outside of the framework of Bayesian causal nets and DAGs. The analysis of causality in simultaneous equations models and the identification of causal parameters are central and enduring contributions of Haavelmo (1944).

References

Angrist, J. D. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Bishop, Y. M., S. E. Fienberg, and P. W. Holland (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Massachusetts: The MIT Press.

Blundell, R. and J. Powell (2003). Endogeneity in nonparametric and semiparametric regression models. In L. P. H. M. Dewatripont and S. J. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Volume 2. Cambridge, UK: Cambridge University Press.

Chalak, K. and H. White (2012). Causality, conditional independence, and graphical separation in settable systems. *Neural Computation 24*(7), 1611–1668.

Chesher, A. and A. Rosen (2012). Simultaneous equations for discrete outcomes: Coherence, completeness, and identification. Working Papers CWP21/12, cemmap.

Dawid, A. (2001). Separoids: A mathematical framework for conditional independence and irrelevance. *Annals of Mathematics and Artificial Intelligence 32*(1–4), 335–372.

Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society. Series B (Statistical Methodological) 41*(1), 1–31.

Fechner, G. T. (1851). Outline of a new principle of mathematical psychology. *Psychological Research 49*, 203–207.

Freedman, D., D. Collier, J. Sekhon, and P. Stark (2010). *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. Cambridge, UK: Cambridge University Press.

Frisch, R. (1930). *A Dynamic Approach to Economic Theory: The Yale Lectures of Ragnar Frisch, 1930.* New York, New York: Routledge. Olav Bjerkholt and Duo Qin (eds)., Published 2010.

Frisch, R. (1938). Autonomy of economic relations: Statistical versus theoretical relations in economic macrodynamics. Paper given at League of Nations. Reprinted in D.F. Hendry and M.S. Morgan (1995), *The Foundations of Econometric Analysis*, Cambridge University Press.

Haavelmo, T. (1943, January). The statistical implications of a system of simultaneous equations. *Econometrica 11*(1), 1–12.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica 12*(Supplement), iii–vi and 1–115.

Hansen, L. P. and T. J. Sargent (1980, February). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control 2*(1), 7–46.

Heckman, J. and R. Pinto (2013). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. Forthcoming, *Econometric Reviews.*

Heckman, J. J. (1976, December). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement 5*(4), 475–492.

Heckman, J. J. (1978, July). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*(4), 931–959.

Heckman, J. J. (1979, January). Sample selection bias as a specification error. *Econometrica 47*(1), 153–162.

Heckman, J. J. (2005, August). The scientific model of causality. *Sociological Methodology 35*(1), 1–97.

Heckman, J. J. (2008, April). Econometric causality. *International Statistical Review 76*(1), 1–27.

Heckman, J. J. and T. E. MaCurdy (1985, February). A simultaneous equations linear probability model. *Canadian Journal of Economics 18*(1), 28–37.

Heckman, J. J. and R. Robb (1985, October-November). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics 30*(1–2), 239–267.

Heckman, J. J. and E. J. Vytlacil (1999, April). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences 96*(8), 4730–4734.

Heckman, J. J. and E. J. Vytlacil (2005, May). Structural equations, treatment effects and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 4779–4874. Amsterdam: Elsevier.

Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effects in new environments. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 71, pp. 4875–5143. Amsterdam: Elsevier.

Heidelberger, M. (2004). *Nature from within: Gustav Theodor Fechner and his psychophysical worldview.* Pittsburgh, PA: University of Pittsburgh Press.

Holland, P. W. (1986, December). Statistics and causal inference. *Journal of the American Statistical Association 81*(396), 945–960.

Howard, R. A. and J. E. Matheson (1981). Principles and applications of decision analysis. In *Influence diagrams* (1 ed.)., pp. 720–762. Menlo Park, CA: Stanford Research Institute.

Huang, Y. and M. Valtorta (2006). A study of identifiability in causal Bayesian network. Technical report, University of South Carolina Department of Computer Science.

Huang, Y. and M. Valtorta (2008). On the completeness of an identifiability algorithm for semi-markovian models. *Annals of Mathematics and Artificial Intelligence 54*(4), 363–408.

Imbens, G. W. and J. D. Angrist (1994, March). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Kiiveri, H., T. P. Speed, and J. B. Carlin (1984). Recursive causal models. *Journal of the Australian Mathematical Society (Series A) 36*(1), 30–52.

Koopmans, T. C. and O. Reiersøl (1950, June). The identification of structural characteristics. *The Annals of Mathematical Statistics XXI*(2), 165–181.

Koopmans, T. C., H. Rubin, and R. B. Leipnik (1950). Measuring the equation systems of dynamic economics. In T. C. Koopmans (Ed.), *Statistical Inference in Dynamic Economic Models*, Number 10 in Cowles Commission Monograph, Chapter 2, pp. 53–237. New York: John Wiley & Sons.

Lauritzen, S. L. (1996). *Graphical Models.* Oxford, UK: Clarendon Press.

Lauritzen, S. L. (2001). Causal inference from graphical models. In O. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg (Eds.), *Complex Stochastic Systems*, pp. 63–107. London: Chapman and Hall.

Lauritzen, S. L. and T. S. Richardson (2002). Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(3), 321–348.

Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses* (Third ed.). New York: Springer Science and Business Media.

Margolis, M., J. List, and D. Osgood (2012, April). Endangered options and endangered species: what we can learn from a dubious design. Unpublished manuscript, Gettysburg College, Department of Economics.

Marshall, A. (1890). *Principles of Economics.* New York: Macmillan and Company.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995). *Microeconomic Theory.* New York: Oxford University Press.

Matzkin, R. L. (2007). Nonparametric identification. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B. Amsterdam: Elsevier.

Matzkin, R. L. (2008). Identification in nonparametric simultaneous equations models. *Econometrica 76*(5), 945–978.

Matzkin, R. L. (2012). Identification in nonparametric limited dependent variable models with simultaneity and unobserved heterogeneity. *Journal of Econometrics 166*(1), 106–115.

Matzkin, R. L. (2013). Nonparametric identification of structural economic models. *Annual Review of Economics 5.* Forthcoming.

Newey, W. K. and J. L. Powell (2003, September). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, CA: Morgan Kaufmann Publishers Inc.

Pearl, J. (1993). [Bayesian analysis in expert systems]: Comment: Graphical models, causality and intervention. *Statistical Science 8*(3), 266–269.

Pearl, J. (1995, December). Causal diagrams for empirical research. *Biometrika 82*(4), 669–688.

Pearl, J. (2000). *Causality.* Cambridge, England: Cambridge University Press.

Pearl, J. (2001). *Causality: Models, reasoning, and inference* (Reprinted with corrections ed.). New York: Cambridge University Press.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). New York: Cambridge University Press.

Pearl, J. and T. S. Verma (1994). A theory of inferred causation. In D. Prawitz, B. Skyrms, and D. Westerståhl (Eds.), *Logic, Methodology, and Philosophy of Science*, Volume IX, pp. 789–812. Amsterdam: Elsevier Science. Proceedings of the Ninth International Congress of Logic, Methodology, and Philosophy of Science, Uppsala, Sweden, August 7–14, 1991.

Powell, J. L. (1994). Estimation of semiparametric models. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, pp. 2443–2521. Amsterdam: Elsevier.

Reiersöl, O. (1945). Confluence analysis by means of instrumental sets of variables. *Arkiv för Matematik, Astronomi och Fysik 32A*(4), 1–119.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period: Application to control of the healthy worker survivor effect. *Mathematical Modelling 7*(9–12), 1393–1512.

Robins, J. M. and T. S. Richardson (2011). Alternative graphical causal models and the identification of direct effects. In P. E. Shrout, K. M. Keyes, and K. Ornstein (Eds.), *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures*, Chapter 6, pp. 103–158. New York, NY: Oxford University Press.

Rosenbaum, P. R. and D. B. Rubin (1983, April). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Rubin, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association 81*(396), 961–962.

Simon, H. A. (1953). Causal ordering and identifiability. In W. C. Hood and T. C. Koopmans (Eds.), *Studies in Econometric Method*, Chapter 3, pp. 49–74. New York, NY: John Wiley & Sons, Inc.

Sobel, M. E. (2005). Discussion: 'the scientific model of causality'. *Sociological Methodology 35*(1), 99–133.

Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. In *Proceedings of the Eleventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-95)*, pp. 491–498. San Francisco, CA: Morgan Kaufmann.

Spirtes, P., C. N. Glymour, and R. Scheines (2000). *Causation, Prediction and Search* (2 ed.). Cambridge, MA: MIT Press.

Tamer, E. (2003, January). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies 70*(1), 147–165.

Tian, J. and J. Pearl (2002). A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pp. 567–573. Cambridge, MA: AAAI Press.

Tian, J. and J. Pearl (2003). On the identification of causal effects. Technical report, Cognitive Systems Laboratory, University of California at Los Angeles.

Vytlacil, E. J. (2002, January). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica 70*(1), 331–341.

White, H. and K. Chalak (2009). Settable systems: An extension of Pearl's causal model with optimization, equilibrium, and learning. *Journal of Machine Learning Research 10*, 1759–1799.

Yule, G. U. (1895). On the correlation of total pauperism with proportion of out-relief. *The Economic Journal 5*(20), 603–611.

A Mathematical Appendix

Theorem **T-1**:

Proof. If V is non-descendant of $\tilde{X}$ in the hypothetical model, i.e. $V \in \mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X})$ then $V \in \mathcal{T}_{\mathrm{E}} \setminus Ch_{\mathrm{H}}(\tilde{X})$ as $Ch_{\mathrm{H}}(\tilde{X}) \subset D_{\mathrm{H}}(\tilde{X})$. Thus $\mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V)) = \mathbf{P}_{\mathrm{E}}(V|Pa_{\mathrm{E}}(V))$ from Equation (8). Moreover, it must be the case that parents of V are also non-descendants of $\tilde{X}$, i.e., $Pa_{\mathrm{H}}(V) \subset \mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X}) \subset \mathcal{T}_{\mathrm{E}} \setminus Ch_{\mathrm{H}}(\tilde{X}) \therefore Pa_{\mathrm{H}}(V) = Pa_{\mathrm{E}}(V)$ by Equation (8). Another way of saying this is that the parents of V are not children of $\tilde{X}$. Thus we can use factorization (6) to write:

$$\mathbf{P}_{\mathrm{H}}(\mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X})) = \prod_{V \in \mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X})} \mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V)) = \prod_{V \in \mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X})} \mathbf{P}_{\mathrm{E}}(V|Pa_{\mathrm{E}}(V)) = \mathbf{P}_{\mathrm{E}}(\mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X})).$$

As a consequence, $\mathbf{P}_{\mathrm{H}}(W) = \mathbf{P}_{\mathrm{E}}(W)$ for all $W \subset \mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X})$ and thereby

$$\mathbf{P}_{\mathrm{H}}(W = w|Z = z) = \frac{\mathbf{P}_{\mathrm{H}}(W = w, Z = z)}{\mathbf{P}_{\mathrm{H}}(Z = z)} = \frac{\mathbf{P}_{\mathrm{E}}(W = w, Z = z)}{\mathbf{P}_{\mathrm{E}}(Z = z)} = \mathbf{P}_{\mathrm{E}}(W = w|Z = z).$$

Conditioning on $\tilde{X}$ comes from that fact that $\tilde{X} \perp\!\!\!\perp (\mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X}))$, which is obtained by applying LMC (5) to $\tilde{X}$ in G_{H}. □

Theorem **T-2**:

Proof. In order to prove the theorem, we first partition the set of variables $\mathcal{T}_{\mathrm{E}}$ into four sets:

$$\mathcal{T}_{\mathrm{E}} = \{\underbrace{\mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{E}}(X)}_{\text{Set 1}}\} \cup \{\underbrace{D_{\mathrm{E}}(X) \setminus Ch_{\mathrm{E}}(X)}_{\text{Set 2}}\} \cup \{\underbrace{Ch_{\mathrm{H}}(X)}_{\text{Set 3}}\} \cup \{\underbrace{Ch_{\mathrm{H}}(\tilde{X})}_{\text{Set 4}}\}.$$

Set 1 consists of all variables in $\mathcal{T}_{\mathrm{E}}$ that are non-descendants of X in the empirical model and thereby nondescendants of $\tilde{X}$ in the hypothetical one. Set 2 consists of descendants of X but not directly caused by X, i.e., except its Children. Sets 3 and 4 are the Children of X and $\tilde{X}$ in the hypothetical model. Note that Sets 3 and 4 consist of all Children of X

in the empirical model as $Ch_{\mathrm{E}}(X) = Ch_{\mathrm{H}}(X) \cup Ch_{\mathrm{H}}(\tilde{X})$. We now examine the variables of each set separately:

1. For all $V \in \mathcal{T}_{\mathrm{H}} \setminus D_{\mathrm{E}}(X) \Rightarrow \{V, Pa_{\mathrm{H}}(V)\} \subset \mathcal{T}_{\mathrm{H}} \setminus D_{\mathrm{E}}(X) \subset \mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X})$, as $D_{\mathrm{H}}(\tilde{X}) \subset D_{\mathrm{E}}(X)$. Also $X \in \mathcal{T}_{\mathrm{E}} \setminus D_{\mathrm{H}}(\tilde{X})$. Thus by Theorem **T-1**: $\mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}, \tilde{X} = x, X = x) = \mathbf{P}_{\mathrm{E}}(V|Pa_{\mathrm{E}}(V), X = x)$.

2. $V \in D_{\mathrm{E}}(X) \setminus Ch_{\mathrm{E}}(X) \Rightarrow \tilde{X} \notin Pa_{\mathrm{H}}(V), X \notin Pa_{\mathrm{H}}(V)$, and $Pa_{\mathrm{H}}(V) = Pa_{\mathrm{E}}(V)$. Moreover, $X, \tilde{X}$ must be non-descendants of V due to the acyclic property of the empirical model on X. Thus, by LMC (5), $(\tilde{X}, X) \perp\!\!\!\perp V|Pa_{\mathrm{H}}(V)$. By Weak Union, $\tilde{X} \perp\!\!\!\perp V|(Pa_{\mathrm{H}}(V), X)$. Therefore $\mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V), \tilde{X} = x, X = x) = \mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V), X = x) = \mathbf{P}_{\mathrm{E}}(V|Pa_{\mathrm{E}}(V), X = x)$ by Equation (8).

3. $V \in Ch_{\mathrm{H}}(X) \Rightarrow \tilde{X} \notin Pa_{\mathrm{H}}(V)$ and $X \in Pa_{\mathrm{H}}(V) = Pa_{\mathrm{E}}(V)$. Also, $\tilde{X}$ is external, thus $\tilde{X} \perp\!\!\!\perp V|Pa_{\mathrm{H}}(V)$ by LMC (5) applied to V. Therefore $\mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V) \setminus X, \tilde{X} = x, X = x) = \mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V) \setminus X, X = x) = \mathbf{P}_{\mathrm{E}}(V|Pa_{\mathrm{E}}(V) \setminus X, X = x)$ by Equation (8) as $V \in Ch_{\mathrm{H}}(X) \subset \mathcal{T}_{\mathrm{E}} \setminus Ch_{\mathrm{H}}(\tilde{X})$.

4. $V \in Ch_{\mathrm{H}}(\tilde{X}) \Rightarrow \tilde{X} \in Pa_{\mathrm{H}}(V)$. Moreover, X must be a non-descendant of V due to the acyclic property of the empirical model on X. Thus, by LMC (5), $X \perp\!\!\!\perp V|Pa_{\mathrm{H}}(V)$. Therefore $\mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V) \setminus \tilde{X}, \tilde{X} = x, X = x) = \mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V) \setminus \tilde{X}, \tilde{X} = x) = \mathbf{P}_{\mathrm{E}}(V|Pa_{\mathrm{E}}(V) \setminus X, X = x)$ by Equation (9).

Grouping items 1–4, we have that for all $V \in \mathcal{T}_{\mathrm{H}}$, $\mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V), \tilde{X} = x, X = x) = \mathbf{P}_{\mathrm{E}}(V|Pa_{\mathrm{E}}(V), X = x)$. Thus we can use the factorization (6) to obtain:

$$\begin{aligned}
\mathbf{P}_{\mathrm{H}}(\mathcal{T}_{\mathrm{E}}|X = x, \tilde{X} = x) &= \prod_{V \in \mathcal{T}_{\mathrm{E}}} \mathbf{P}_{\mathrm{H}}(V|Pa_{\mathrm{H}}(V), \tilde{X} = x, X = x) \\
&= \prod_{V \in \mathcal{T}_{\mathrm{E}}} \mathbf{P}_{\mathrm{E}}(V|Pa_{\mathrm{E}}(V), X = x) \\
&= \mathbf{P}_{\mathrm{E}}(\mathcal{T}_{\mathrm{E}}|X = x). \qquad (12)
\end{aligned}$$

The claim of the theorem is a direct consequence of Equation (12). □

Corollary **C-1**:

Proof.

$$\begin{aligned}
\mathbf{P}_{\mathrm{H}}(\mathcal{T}_{\mathrm{E}}|X=\tilde{X}) &= \sum_{x\in\mathrm{supp}(X)} \mathbf{P}_{\mathrm{H}}(\mathcal{T}_{\mathrm{E}}|X=x,\tilde{X}=x)\frac{\mathbf{P}_{\mathrm{H}}(X=x,\tilde{X}=x)}{\sum_{x\in\mathrm{supp}(X)}\mathbf{P}_{\mathrm{H}}(X=x,\tilde{X}=x)}\\
&= \sum_{x\in\mathrm{supp}(X)} \mathbf{P}_{\mathrm{H}}(\mathcal{T}_{\mathrm{E}}|X=x,\tilde{X}=x)\frac{\mathbf{P}_{\mathrm{H}}(X=x)\mathbf{P}_{\mathrm{H}}(\tilde{X}=x)}{\sum_{x\in\mathrm{supp}(X)}\mathbf{P}_{\mathrm{H}}(X=x)\mathbf{P}_{\mathrm{H}}(\tilde{X}=x)}\\
&= \sum_{x\in\mathrm{supp}(X)} \mathbf{P}_{\mathrm{H}}(\mathcal{T}_{\mathrm{E}}|X=x,\tilde{X}=x)\frac{\mathbf{P}_{\mathrm{H}}(X=x)}{\sum_{x\in\mathrm{supp}(X)}\mathbf{P}_{\mathrm{H}}(X=x)}\\
&= \sum_{x\in\mathrm{supp}(X)} \mathbf{P}_{\mathrm{H}}(\mathcal{T}_{\mathrm{E}}|X=x,\tilde{X}=x)\mathbf{P}_{\mathrm{H}}(X=x)\\
&= \sum_{x\in\mathrm{supp}(X)} \mathbf{P}_{\mathrm{E}}(\mathcal{T}_{\mathrm{E}}|X=x)\mathbf{P}_{\mathrm{E}}(X=x)\\
&= \mathbf{P}_{\mathrm{E}}(\mathcal{T}_{\mathrm{E}}).
\end{aligned}$$

The second equality stems from $Pa_{\mathrm{H}}(\tilde{X})=\varnothing$ and X is not descendant of $\tilde{X}$, thus by LMC (5), $X \perp\!\!\!\perp \tilde{X}$. Therefore $\mathbf{P}_{\mathrm{H}}(X=x,\tilde{X}=x)=\mathbf{P}_{\mathrm{H}}(X=x)\mathbf{P}_{\mathrm{H}}(\tilde{X}=x)$. The third equality comes from the assumption that $\mathbf{P}_{\mathrm{H}}(\tilde{X}=x)$ is constant due to uniformity. The fourth equality comes from the fact that $\sum_{x\in\mathrm{supp}(X)}\mathbf{P}_{\mathrm{H}}(X=x)=1$. The first term of the fifth equality comes from an application of Theorem **T-2**. The second term of the fifth equality comes from Theorem **T-1** and the fact that $X\in\mathcal{T}_{\mathrm{E}}\setminus D_{\mathrm{H}}(\tilde{X})$. □

Theorem **T-3**:

Proof.

$$\begin{aligned}
\mathbf{P}_{\mathrm{H}}(\mathcal{T}_{\mathrm{E}} \setminus X|\tilde{X} = x) &= \prod_{V \in \mathcal{T}_{\mathrm{E}} \setminus \{X \cup Ch_{\mathrm{H}}(X)\}} \mathbf{P}_{\mathrm{H}}(V|Pa(V)) \prod_{V \in Ch_{\mathrm{H}}(X)} \mathbf{P}_{\mathrm{H}}(V|Pa(V) \setminus \tilde{X}, \tilde{X} = x) \\
&= \prod_{V \in \mathcal{T}_{\mathrm{E}} \setminus \{X \cup Ch_{\mathrm{E}}(X)\}} \mathbf{P}_{\mathrm{H}}(V|Pa(V)) \prod_{V \in Ch_{\mathrm{E}}(X)} \mathbf{P}_{\mathrm{H}}(V|Pa(V) \setminus \tilde{X}, \tilde{X} = x) \\
&= \prod_{V \in \mathcal{T}_{\mathrm{E}} \setminus \{X \cup Ch_{\mathrm{E}}(X)\}} \mathbf{P}_{\mathrm{E}}(V|Pa(V)) \prod_{V \in Ch_{\mathrm{E}}(X)} \mathbf{P}_{\mathrm{E}}(V|Pa(V) \setminus X, X = x) \\
&= \mathbf{P}_{\mathrm{E}}(\mathcal{T}_{\mathrm{E}} \setminus X|do(X) = x).
\end{aligned}$$

The first equality comes from the fact that the hypothetical model is a DAG, therefore we apply factorization (6). The second equality comes from the characteristic of the do-operator, which targets all causal links of a fixed variable X. Thus the hypothetical variable $\tilde{X}$ must replace all X inputs which is equivalent to $Ch_{\mathrm{H}}(\tilde{X}) = Ch_{\mathrm{E}}(X)$. The first and second terms of the third equality come as a consequence of Equations (8) and (9) respectively. The last equality comes from the definition of the do-operator. □

Lemma **L-1**:

Proof.

$$\begin{aligned}
\mathbf{P}_{\mathrm{H}}(W|Z, \tilde{X} = x) &= \mathbf{P}_{\mathrm{H}}(W|Z, \tilde{X} = x, X = x) && \text{by assumption } X \perp\!\!\!\perp W|(Z, \tilde{X}) \text{ in } G_{\mathrm{H}} \\
&= \mathbf{P}_{\mathrm{E}}(W|Z, X = x) && \text{by Theorem } \textbf{T-2}.
\end{aligned}$$

□