Identification of Causal Effects in Linear SEMs using the Instrumental Variable Function

Hei Chan and Manabu Kuroki
The Institute of Statistical Mathematics
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
{hei,mkuroki}@ism.ac.jp

Abstract

In this paper, we discuss the problem of identification of causal effects in linear Structural Equation Models (SEMs) based on causal diagrams. The two main methods for solving this problem are algebraic methods, which attempts to solve the set of simultaneous equations based on Wright's method of path analysis, and graphical methods, which tests for certain graphical criteria in the causal diagram of the SEM. We will introduce the Instrumental Variable Function, which we will show to correspond to both algebraic and graphical methods and is useful in understanding the identification problem.

Introduction

Structural Equation Models (SEM) is a useful tool for causal analysis, and is widely used in areas of social science such as economics (Bollen 1989; Duncan 1975). Research by scientists, social scientists, and computer scientists in this area has allowed the problem to be applied in real-life models.

In a linear SEM, the relationships between observed variables are expressed in linear equations. The structure of the equations is such that they not only express the linear relationships between the variables, together with a stochastic error term for unobserved factors, but also the causal dependence among the observed variables, which can be graphically represented by a causal diagram, which is a directed acyclic graph (DAG). For each variable Y, its structural equation where it appears on the left-hand side, the presence (and absence) of a variable X on the right-hand side specifies that X is (or is not) a direct cause of Y.

A fundamental problem in linear SEMs is to estimate the strength of a causal effect (either the direct effect, the total effect, or the effect conditional on certain variables) from one variable to another variable, from the combination of observed data and the causal diagram of the SEM. This is called the *identification problem* (Fisher 1966). The identification problem can be solved by solving the set of simultaneous equations based on Wright's method of path analysis (Wright 1934) using algebraic methods such as Gröbner basis computations (Sullivant, Garcia-Puente, and Spielvogel 2010). A causal effect is identifiable if and only if it has a unique solution. While this algebraic method is sound and complete, the graphical structure of the SEM is not taken into account, and it has been shown to be computationally expensive given a large number of variables.

It is also possible to check if a causal effect is identifiable by testing for certain graphical criteria in the causal diagram of the linear SEM (Brito and Pearl 2002c; 2002a; Tian 2004; 2005; 2007b). In particular, these graphical tests include the back-door criterion (Pearl 2009), the instrumental variable (Bowden and Turkington 1984; Brito and Pearl 2002b), and the path-specific instrumental variable (Chan and Kuroki 2010). However, it has not been shown whether the satisfaction of these graphical test is a necessary condition for the identification of a causal effect.

In this paper, we will introduce the Instrumental Variable Function. Given certain variables, the function is computed from the covariances and causal effects between these variables. We then show that solving a set of simultaneous equations based on the Instrumental Variable Function is equivalent to solving the set of simultaneous equations based on Wright's method of path analysis, and thus provides a solution that is sound and complete. Next, we will show how the Instrumental Variable Function is related with graphical tests of the identification problem, by expressing the equations that are given in graphical methods for computing the causal effect in terms of the Instrumental Variable Function. Now we first provide the preliminary definitions for this paper, including linear SEMs, statistical terms such as covariances, and Wright's method of analysis.

Linear SEMs

Given a causal diagram $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ which is a DAG, we define a linear *structural equation model* (SEM), where the vertices represent observed variables, the directed edges represent direct causal relationships, and the bidirected edges represent correlations of error terms (regarding the graph theoretic terminology used in this paper, see (Pearl 2009)). Since $\mathcal{G}$ is a DAG, the SEM is *recursive*, and we assume $\mathbf{V} = (V_1, \ldots, V_n)$ is in topological order. Given $V_i \neq V_j$, we define the sets $V_j \in Desc(V_i)$ (and $V_i \in Anc(V_j)$) if and only if V_j is a *descendent* of V_i (and V_i is an ancestor of V_j).

To define the set of equations governing the linear SEM, a set of *parameters* is specified based on $\mathcal{G}$. For each variable $V \in \mathbf{V}$, we specify its variance as σ_V^2. For each edge $E \in \mathbf{E}$, we specify its edge parameter, denoted as θ_E, which has different meanings for directed and bi-directed edges:

- For any $V_i, V_j \in \mathbf{V}$ where $V_i \neq V_j$, if the directed edge $E = V_i \to V_j \in \mathbf{E}$, we define the parameter of the edge as $\theta_E = \lambda_{V_i \to V_j}$, which is the *direct causal effect* from V_i to V_j. Otherwise, we define $\lambda_{V_i \to V_j} = 0$.
- For any $V_i, V_j \in \mathbf{V}$ where $V_i \neq V_j$, if the bi-directed edge $E = V_i \leftrightarrow V_j \in \mathbf{E}$, we define the parameter of the edge as $\theta_E = \omega_{V_i \leftrightarrow V_j}$, which is the *correlation of error terms* between V_i and V_j. Otherwise, we define $\omega_{V_i \leftrightarrow V_j} = 0$.

The linear equation for each variable $V \in \mathbf{V}$ in the SEM is then given as:

$$V = \sum_{V_{pa} \to V \in \mathbf{E}} \lambda_{V_{pa} \to V} \cdot V_{pa} + \epsilon_V, \tag{1}$$

where V_{pa} is a parent of V, and the error term ϵ_V is assumed to be normally distributed with mean 0. The error terms of V_i and V_j are not independent of each other if and only if there is an bi-directed edge between V_i and V_j.

Next, for any $V_i, V_j \in \mathbf{V}$, we define the *covariance* between V_i and V_j as $\sigma_{V_i V_j}$, and the *conditional covariance* between V_i and V_j given $\mathbf{V_z} \subseteq \mathbf{V}$ as $\sigma_{V_i V_j | \mathbf{V_z}}$, with the following recursive condition for any $V_z \in \mathbf{V_z}$:

$$\sigma_{V_i V_j | \mathbf{V_z}} = \sigma_{V_i V_j | \mathbf{V_z} \setminus \{V_z\}} - \frac{\sigma_{V_i V_z | \mathbf{V_z} \setminus \{V_z\}} \sigma_{V_j V_z | \mathbf{V_z} \setminus \{V_z\}}}{\sigma_{V_z V_z | \mathbf{V_z} \setminus \{V_z\}}}, \tag{2}$$

with ending conditions $\sigma_{V_i V_j | \emptyset} = \sigma_{V_i V_j}$. Note that if V_i and V_j are conditional independent given $\mathbf{V_z}$, i.e., V_i and V_j are d-separated given $\mathbf{V_z}$ in $\mathcal{G}$, then $\sigma_{V_i V_j | \mathbf{V_z}} = 0$. By definition, the covariance matrix is symmetric, i.e., $\sigma_{V_i V_j} = \sigma_{V_j V_i}$ and the diagonal values are the variances of the variables, $\sigma_{V_i V_i} = \sigma_V^2$.

In this paper, it is assumed that a causal diagram $\mathcal{G}$ and the corresponding linear SEM are faithful to each other; that is, the conditional independence relationships in the linear SEM are also reflected in $\mathcal{G}$, and vice versa (Spirtes, Glymour, and Scheines 2000).

Finally, for any $V_i, V_j \in \mathbf{V}$, we define the *total causal effect* from V_i to V_j as $\tau_{V_i V_j}$, which is the sum of products of direct causal effects along all directed paths from V_i to V_j, denoted by $dp(V_i, V_j)$:

$$\tau_{V_i V_j} = \sum_{p \in dp(V_i, V_j)} \prod_{V_k \to V_{k+1} \in p} \lambda_{V_k \to V_{k+1}}. \tag{3}$$

By definition, we have $\tau_{V_i V_i} = 1$, and $\tau_{V_i V_j} = 0$ if $V_i \neq V_j$ and $V_j \notin Desc(V_i)$. We also define the *conditional causal effect* from V_i to V_j given $\mathbf{V_z} \subseteq \mathbf{V}$ as $\tau_{V_i V_j | \mathbf{V_z}}$, which is the sum of products of direct causal effects along all directed paths from V_i to V_j which do not pass through any variable in $\mathbf{V_z}$, denoted by $dp(V_i, V_j \mid \mathbf{V_z})$:

$$\tau_{V_i V_j | \mathbf{V_z}} = \sum_{p \in dp(V_i, V_j | \mathbf{V_z})} \prod_{V_k \to V_{k+1} \in p} \lambda_{V_k \to V_{k+1}}. \tag{4}$$

The conditional causal effect $\tau_{V_i V_j | \mathbf{V_z}}$ can also be computed with the following recursive condition for any $V_z \in \mathbf{V_z}$:

$$\tau_{V_i V_j | \mathbf{V_z}} = \tau_{V_i V_j | \mathbf{V_z} \setminus \{V_z\}} - \tau_{V_i V_z | \mathbf{V_z} \setminus \{V_z\}} \tau_{V_z V_j | \mathbf{V_z} \setminus \{V_z\}}, \tag{5}$$

with ending conditions $\tau_{V_i V_j | \emptyset} = \tau_{V_i V_j}$. We can easily show that for all $i < j$, $\tau_{V_j V_i | \mathbf{V_z}} = 0$, and $\lambda_{V_i \to V_j} = \tau_{V_i V_j | \mathbf{V} \setminus \{V_i, V_j\}} = \tau_{V_i V_j | \mathbf{V}_{i+1}^{j-1}} = \tau_{V_i V_j | Desc(V_i) \cap Anc(V_j)}$.

Path Analysis

Given a linear SEM and its causal diagram $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, we can express the covariance between any two variables in terms of the parameters of the linear SEM using *Wright's method of path analysis* (Wright 1934), by accounting for all unblocked paths between the two variables.

First we make some definitions. For any unblocked path p, we define the term $t(p)$ as the product of the parameters of all edges on the path:

$$t(p) = \prod_{E \in p} \theta_E. \tag{6}$$

The parameter θ_E is either $\lambda_{V_k \to V_{k+1}}$, $\lambda_{V_{k+1} \to V_k}$, or $\omega_{V_k \leftrightarrow V_{k+1}}$, depending on the directionality of the edge E with respect to V_k and V_{k+1}. Note that, by our definition of the total causal effect (Equation 3) and the conditional causal effect (Equation 4), we have:

$$\tau_{V_i V_j} = \sum_{p \in dp(V_i, V_j)} t(p); \tag{7}$$

$$\tau_{V_i V_j | \mathbf{V_z}} = \sum_{p \in dp(V_i, V_j | \mathbf{V_z})} t(p). \tag{8}$$

We also define the term $s(p)$ as:

$$s(p) = r(p) t(p) = r(p) \prod_{E \in p} \theta_E, \tag{9}$$

where the value $r(p)$ is defined as follows:

- If p does not contain a bi-directed edge, i.e., it is a directed tree and thus has a root V_r, then $r(p) = \sigma_{V_r V_r} = \sigma_{V_r}^2$;
- If p contains a bi-directed edge, then $r(p) = 1$.

We now introduce the equations from Wright's Method of path analysis by expressing them in terms of $s(p)$.

Wright's Method of path analysis Given $V_i, V_j \in \mathbf{V}$, where the set of all unblocked paths between V_i and

V_j in $\mathcal{G}$ is denoted as $up(V_i, V_j)$, we have:

$$\sigma_{V_iV_j} = \sum_{p \in up(V_i,V_j)} s(p) = \sum_{p \in up(V_i,V_j)} \left(r(p) \prod_{E \in p} \theta_E \right). \tag{10}$$

As $\mathbf{V} = (V_1, \ldots, V_n)$ is in topological order, we can recursively compute the covariances and the total causal effects following the topological order. This is because any path between V_i and V_j that goes through $V_{k'}$ where $k' > i, j$ cannot be an unblocked path, as $V_{k'} \notin Anc(V_i) \cup Anc(V_j)$ and thus must be a collider.

Given $V_i, V_j \in \mathbf{V}$ where $i < j$, and $p \in up(V_i, V_j)$, let V_k be the vertex before V_j in p, and p_1 be the sub-path of p from V_i to V_k, and $p_2 \in \{V_k \to V_j, V_k \leftrightarrow V_j\}$ be the sub-path of p from V_k to V_j (note that we must have $k < j$). We have the following relations:

$$\begin{aligned} & p = p_1 \star p_2 \in up(V_i, V_j) \\ & \quad \Leftrightarrow \quad (p_1 \in up(V_i, V_k) \wedge p_2 = V_k \to V_j \in \mathbf{E}) \\ & \qquad\qquad \vee (p_1 \in dp(V_k, V_i) \wedge p_2 = V_k \leftrightarrow V_j \in \mathbf{E}), \\ & p = p_1 \star p_2 \in dp(V_i, V_j) \\ & \quad \Leftrightarrow \quad p_1 \in dp(V_i, V_k) \wedge p_2 = V_k \to V_j \in \mathbf{E}, \end{aligned}$$

for some $k < j$. Since $dp(V_k, V_i) = \emptyset$ if $k > i$, and $dp(V_i, V_k) = \emptyset$ if $i > k$, we have the following result.

Corollary 1 *Given $\mathbf{V} = (V_1, \ldots, V_n)$ in topological order, for all $V_i, V_j \in \mathbf{V}$ where $i < j$, we have:*

$$\sigma_{V_iV_j} = \sum_{k=1}^{j-1} \sigma_{V_iV_k} \lambda_{V_k \to V_j} + \sum_{k=1}^{i} \tau_{V_kV_i} \omega_{V_k \leftrightarrow V_j}, \tag{11}$$

$$\tau_{V_iV_j} = \sum_{k=i}^{j-1} \tau_{V_iV_k} \lambda_{V_k \to V_j}. \tag{12}$$

Identification of Causal Effects

In the previous section, we compute the covariances and the total causal effects between the observed variables from the parameters of the linear SEM. In this section, we try to compute a causal effect from the covariances between the observed variables.

We first define the covariances and total causal effects as functions of the parameters of the linear SEM. Given a linear SEM and its causal diagram $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, with $\mathbf{V} = (V_1, \ldots, V_n)$ in topological order, we define its parameters as the set, $\Theta_{\mathcal{G}} = \{\lambda_{V_i \to V_j} : V_i \to V_j \in \mathbf{E}\} \cup \{\omega_{V_i \leftrightarrow V_j} : V_i \leftrightarrow V_j \in \mathbf{E}\} \cup \{\sigma^2_{V_i} : V_i \in \mathbf{V}\} \in \Re^{|\mathbf{E}|+|\mathbf{V}|}$. The covariances and the total causal effects are defined as functions of $\Theta_{\mathcal{G}}$, using the recursive methods according to Equation 11 and Equation 12, for all $V_i, V_j \in \mathbf{V}$ where $i < j$:

$$\begin{aligned} f_{\sigma_{V_iV_j}}(\Theta_{\mathcal{G}}) \;=\; & \sum_{k=1}^{j-1} f_{\sigma_{V_iV_k}}(\Theta_{\mathcal{G}}) f_{\lambda_{V_k \to V_j}}(\Theta_{\mathcal{G}}) \\ & + \sum_{k=1}^{i} f_{\tau_{V_kV_i}}(\Theta_{\mathcal{G}}) f_{\omega_{V_k \leftrightarrow V_j}}(\Theta_{\mathcal{G}}), \end{aligned} \tag{13}$$

$$f_{\tau_{V_iV_j}}(\Theta_{\mathcal{G}}) = \sum_{k=i}^{j-1} f_{\tau_{V_iV_k}}(\Theta_{\mathcal{G}}) f_{\lambda_{V_k \to V_j}}(\Theta_{\mathcal{G}}), \tag{14}$$

where the functions $f_{\lambda_{V_k \to V_j}}(\Theta_{\mathcal{G}})$ and $f_{\omega_{V_k \leftrightarrow V_j}}(\Theta_{\mathcal{G}})$ return the values of the corresponding parameters in $\Theta_{\mathcal{G}}$ (0 if the edge does not exist in $\mathcal{G}$). Moreover, we also have:

$$f_{\sigma_{V_iV_i}}(\Theta_{\mathcal{G}}) = f_{\sigma^2_{V_i}}(\Theta_{\mathcal{G}}), \tag{15}$$

for all $V_i \in \mathbf{V}$, where the functions $f_{\sigma^2_{V_i}}(\Theta_{\mathcal{G}})$ return the variance of V_i. We can also define the functions $f_\Sigma(\Theta_{\mathcal{G}})$ and $f_T(\Theta_{\mathcal{G}})$ to return the whole matrices of Σ and T respectively, using the equations as defined above, and other equations which ensure the necessary properties of the covariances (symmetry) and total causal effects (trivial values 1 and 0). Similar functions for conditional covariances and conditional causal effects can also be defined according to Equation 2 and Equation 5.

The identification problem can be defined as follows. A total causal effect $\tau_{V_iV_j}$ (or conditional causal effect $\tau_{V_iV_j|\mathbf{V_z}}$) is *universally identifiable* if there exists a unique solution given Σ, i.e., $f_\Sigma(\Theta_{\mathcal{G}}) = f_\Sigma(\Theta'_{\mathcal{G}}) \Rightarrow f_{\tau_{V_iV_j}}(\Theta_{\mathcal{G}}) = f_{\tau_{V_iV_j}}(\Theta'_{\mathcal{G}})$ for all $\Theta_{\mathcal{G}}, \Theta'_{\mathcal{G}} \in \Re^{|\mathbf{E}|+|\mathbf{V}|}$, or equivalently, there exists a map Φ where $\forall \Theta_{\mathcal{G}} \in \Re^{|\mathbf{E}|+|\mathbf{V}|} \left(f_{\tau_{V_iV_j}}(\Theta_{\mathcal{G}}) = \Phi \circ f_\Sigma(\Theta_{\mathcal{G}}) \right)$. Moreover, a causal effect is *generically identifiable*, if for a dense open subset $R \subset \Re^{|\mathbf{E}|+|\mathbf{V}|}$ there exists a map Φ where $\forall \Theta_{\mathcal{G}} \in R \left(f_{\tau_{V_iV_j}}(\Theta_{\mathcal{G}}) = \Phi \circ f_\Sigma(\Theta_{\mathcal{G}}) \right)$ (Sullivant, Garcia-Puente, and Spielvogel 2010). In our paper and in most papers on the problem of identification in linear SEMs, the notion of identifiability is understood to be generic identifiability. One method to solve the identification problem is to use algebraic methods such as Gröbner basis computations, by solving the set of simultaneous equations based on Wright's method of path analysis (Equation 10) to check if there is a unique solution for the causal effect (Sullivant, Garcia-Puente, and Spielvogel 2010).

Instrumental Variable Function

Given a linear SEM and its causal diagram $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, we define the *Instrumental Variable Function*. We will later describe its properties and explain why we propose this function.

Definition 1 *Given $V_w, V_y \in \mathbf{V}$ with $V_w \neq V_y$, and $\mathbf{V_x} \subseteq \mathbf{V}$, we define the **Instrumental Variable Function**, $\upsilon_{V_yV_w\|\mathbf{V_x}}$, as:*

$$\begin{aligned} & \upsilon_{V_yV_w\|\mathbf{V_x}} \\ = \; & \rho_{V_yV_w\|\mathbf{V_x}\setminus\{V_w\}} \\ & - \sum_{V_x \in \mathbf{V_x}\setminus\{V_y\}} \rho_{V_xV_w\|\mathbf{V_x}\setminus\{V_w\}} \tau_{V_xV_y|\mathbf{V_x}\setminus\{V_x,V_y\}}, \end{aligned} \tag{16}$$

where given $V_i, V_j \in \mathbf{V}$ and $\mathbf{V_k} \subseteq \mathbf{V}$, we define:

$$\rho_{V_iV_j\|\mathbf{V_k}} = \sigma_{V_iV_j} - \sum_{V_k \in \mathbf{V_k}} \sigma_{V_iV_k} \tau_{V_kV_j|\mathbf{V_k}\setminus\{V_k\}}. \tag{17}$$

The Instrumental Variable Function obeys the two following basic properties (see Appendix for proof). First it is symmetric, i.e.:

$$\upsilon_{V_y V_w \| \mathbf{V_x}} = \upsilon_{V_w V_y \| \mathbf{V_x}}. \quad (18)$$

Moreover, given $V_a \notin \{\{V_w, V_y\} \cup Anc(V_w) \cup Anc(V_y)\}$, we have:

$$\upsilon_{V_y V_w \| \mathbf{V_x} \cup \{V_a\}} = \upsilon_{V_y V_w \| \mathbf{V_x}}, \quad (19)$$

i.e., any variable V_a which is neither V_w, V_y, nor an ancestor of V_w or V_y can be removed from $\mathbf{V_x}$ without affecting its value.

We now state the following important theorem.

Theorem 1 *Given $\mathbf{V} = (V_1, \ldots, V_n)$ in topological order, with $V_w, V_y \in \mathbf{V}$ where $w < y$, we have:*

$$\upsilon_{V_w V_y \| \mathbf{V}} = \omega_{V_w \leftrightarrow V_y}. \quad (20)$$

The proof is as follows. Rearranging from Equation 11, we have:

$$\sigma_{V_x V_y} - \sum_{k=1}^{y-1} \sigma_{V_x V_k} \lambda_{V_k \to V_y} = \sum_{k=1}^{x} \tau_{V_k V_x} \omega_{V_k \leftrightarrow V_y},$$

for all $x = 1, \ldots, w$. We denote the left-hand-side as $A(x)$ and the right-hand-side as $B(x)$. From Equation 19, we have $\upsilon_{V_w V_y \| \mathbf{V}} = \upsilon_{V_w V_y \| \mathbf{V}_1^y}$, where we define $\mathbf{V}_i^y = \{V_i, \ldots, V_y\}$. From Equation 16, we have:

$$\begin{aligned} & \upsilon_{V_w V_y \| \mathbf{V}_1^y} \\ = \ & \rho_{V_w V_y \| \mathbf{V}_1^{y-1}} \\ & - \sum_{V_x \in \mathbf{V}_1^y \setminus \{V_w\}} \rho_{V_x V_y \| \mathbf{V}_1^{y-1}} \tau_{V_x V_w | \mathbf{V}_1^y \setminus \{V_x, V_w\}} \\ = \ & \rho_{V_w V_y \| \mathbf{V}_1^{y-1}} - \sum_{x=1}^{w-1} \rho_{V_x V_y \| \mathbf{V}_1^{y-1}} \lambda_{V_x \to V_w}, \end{aligned}$$

since $\tau_{V_x V_w | \mathbf{V}_1^y \setminus \{V_x, V_w\}} = \lambda_{V_x \to V_w}$ for all $V_x \in \mathbf{V}_1^{w-1}$, and $\tau_{V_x V_w | \mathbf{V}_1^y \setminus \{V_x, V_w\}} = 0$ for all $V_x \in \mathbf{V}_{w+1}^y$. From Equation 17, for all $V_i \in \mathbf{V}_1^w$, we have:

$$\begin{aligned} \rho_{V_i V_y \| \mathbf{V}_1^{y-1}} &= \sigma_{V_i V_y} - \sum_{V_k \in \mathbf{V}_1^{y-1}} \sigma_{V_i V_k} \tau_{V_k V_y | \mathbf{V}_1^{y-1} \setminus \{V_k\}} \\ &= \sigma_{V_i V_y} - \sum_{k=1}^{y-1} \sigma_{V_i V_k} \lambda_{V_k \to V_y}, \end{aligned}$$

since $\tau_{V_k V_y | \mathbf{V}_1^{y-1} \setminus \{V_k\}} = \lambda_{V_k \to V_y}$ for all $V_k \in \mathbf{V}_1^{y-1}$. Therefore, we have:

$$\begin{aligned} & \upsilon_{V_w V_y \| \mathbf{V}} \\ = \ & \sigma_{V_w V_y} - \sum_{k=1}^{y-1} \sigma_{V_w V_k} \lambda_{V_k \to V_y} \\ & - \sum_{x=1}^{w-1} \left(\sigma_{V_x V_y} - \sum_{k=1}^{y-1} \sigma_{V_x V_k} \lambda_{V_k \to V_y} \right) \lambda_{V_x \to V_w}, \end{aligned}$$

which is equal to $A(w) - \sum_{x=1}^{w-1} A(x) \lambda_{V_x \to V_w} = B(w) - \sum_{x=1}^{w-1} B(x) \lambda_{V_x \to V_w}$, which can be expressed as:

$$\begin{aligned} & \sum_{k=1}^{w} \tau_{V_k V_w} \omega_{V_k \leftrightarrow V_y} \\ & - \sum_{x=1}^{w-1} \left(\sum_{k=1}^{x} \tau_{V_k V_x} \omega_{V_k \leftrightarrow V_y} \right) \lambda_{V_x \to V_w} \\ = \ & \tau_{V_w V_w} \omega_{V_w \leftrightarrow V_y} \\ & + \sum_{k=1}^{w-1} \left(\tau_{V_k V_w} - \sum_{x=k}^{w-1} \tau_{V_k V_x} \lambda_{V_x \to V_w} \right) \omega_{V_k \leftrightarrow V_y} \\ = \ & \tau_{V_w V_w} \omega_{V_w \leftrightarrow V_y} \\ = \ & \omega_{V_w \leftrightarrow V_y}, \end{aligned}$$

because from Equation 12, we have:

$$\tau_{V_k V_w} = \sum_{x=k}^{w-1} \tau_{V_k V_x} \lambda_{V_x \to V_w},$$

for all $k = 1, \ldots, w-1$. This means $\upsilon_{V_w V_y \| \mathbf{V}} = \omega_{V_w \leftrightarrow V_y}$, and completes the proof.

Solving the Identification Problem using the Instrumental Variable Function

We now show that solving a set of equations based on the Instrumental Variable Function is equivalent to solving the set of equations based on Wright's method of path analysis, meaning that it provides a sound and complete method to solve the identification problem.

To recap, assume that we are now given two $n \times n$ matrices: the covariances $\Sigma = \{\sigma_{V_i V_j}\}$, which is symmetric, i.e., $\sigma_{V_i V_j} = \sigma_{V_j V_i}$ for all $i = 1, \ldots, n$; and the total causal effects $T = \{\tau_{V_i V_j}\}$, whose entries on the diagonal are all 1 and below the diagonal are all 0, i.e., $\tau_{V_i V_i} = 1$ and $\tau_{V_j V_i} = 0$, for all $i, j = 1, \ldots, n$ and $i < j$. The two matrices are induced by the parameters $\Theta_{\mathcal{G}}$ if $\Sigma = f_\Sigma(\Theta_{\mathcal{G}})$ and $T = f_T(\Theta_{\mathcal{G}})$, i.e., the following equations are satisfied for all $i, j = 1, \ldots, n$ and $i < j$:

$$\begin{aligned} \sigma_{V_i V_j} &= f_{\sigma_{V_i V_j}}(\Theta_{\mathcal{G}}), \\ \tau_{V_i V_j} &= f_{\tau_{V_i V_j}}(\Theta_{\mathcal{G}}), \\ \sigma_{V_i V_i} &= f_{\sigma^2_{V_i}}(\Theta_{\mathcal{G}}). \end{aligned}$$

In particular, the first set of equations are based on Wright's method of path analysis.

Given the Instrumental Variable Function defined in Definition 1 as a function of Σ and T, we show, in the proof of Theorem 1, that if $\Sigma = f_\Sigma(\Theta_{\mathcal{G}})$ and $T = f_T(\Theta_{\mathcal{G}})$, for all $i, j = 1, \ldots, n$ and $i < j$, we have:

$$\upsilon_{V_i V_j \| \mathbf{V}} = f_{\omega_{V_i \leftrightarrow V_j}}(\Theta_{\mathcal{G}}).$$

We now show that $\Sigma = f_\Sigma(\Theta_{\mathcal{G}})$ and $T = f_T(\Theta_{\mathcal{G}})$ if the following equations are satisfied for all $i, j = 1, \ldots, n$ and $i < j$:

$$\begin{aligned} \upsilon_{V_i V_j \| \mathbf{V}} &= f_{\omega_{V_i \leftrightarrow V_j}}(\Theta_{\mathcal{G}}), \\ \tau_{V_i V_j} &= f_{\tau_{V_i V_j}}(\Theta_{\mathcal{G}}), \\ \sigma_{V_i V_i} &= f_{\sigma^2_{V_i}}(\Theta_{\mathcal{G}}). \end{aligned}$$

This is equivalent to showing that $\sigma_{V_iV_j} = f_{\sigma_{V_iV_j}}(\Theta_\mathcal{G})$ for all $i, j = 1, \ldots, n$ and $i < j$ if the above equations are all satisfied.

Since we have $\upsilon_{V_kV_j\|\mathbf{V}} = \omega_{V_k \leftrightarrow V_j}$ for all $k = 1, \ldots, i$, we have:

$$\sum_{k=1}^{i} \tau_{V_kV_i}\upsilon_{V_kV_j\|\mathbf{V}} = \sum_{k=1}^{i} \tau_{V_kV_i}\omega_{V_k\leftrightarrow V_j}.$$

From previous proof, we have:

$$\upsilon_{V_kV_j\|\mathbf{V}} = A(k) - \sum_{x=1}^{k-1} A(x)\lambda_{V_x\to V_k},$$

where:

$$A(x) = \sigma_{V_xV_j} - \sum_{y=1}^{j-1} \sigma_{V_xV_y}\lambda_{V_y\to V_j}.$$

Therefore, the left-hand-side can be expressed as:

$$\begin{aligned}
& \sum_{k=1}^{i} \tau_{V_kV_i}\upsilon_{V_kV_j\|\mathbf{V}} \\
= \; & \sum_{k=1}^{i} \tau_{V_kV_i}\sigma_{V_kV_j} - \sum_{k=1}^{i}\sum_{y=1}^{j-1} \tau_{V_kV_i}\sigma_{V_kV_y}\lambda_{V_y\to V_j} \\
& - \sum_{k=1}^{i}\sum_{x=1}^{k-1} \tau_{V_kV_i}\sigma_{V_xV_j}\lambda_{V_x\to V_k} \\
& + \sum_{k=1}^{i}\sum_{x=1}^{k-1}\sum_{y=1}^{j-1} \tau_{V_kV_i}\sigma_{V_xV_y}\lambda_{V_y\to V_j}\lambda_{V_x\to V_k} \\
= \; & \tau_{V_iV_i}\sigma_{V_iV_j} - \sum_{y=1}^{j-1} \tau_{V_iV_i}\sigma_{V_iV_y}\lambda_{V_y\to V_j} \\
& + \sum_{x=1}^{i-1} \tau_{V_xV_i}\sigma_{V_xV_j} - \sum_{x=1}^{i-1}\sum_{y=1}^{j-1} \tau_{V_xV_i}\sigma_{V_xV_y}\lambda_{V_y\to V_j} \\
& - \sum_{x=1}^{i-1}\sum_{k=x+1}^{i} \tau_{V_kV_i}\sigma_{V_xV_j}\lambda_{V_x\to V_k} \\
& + \sum_{x=1}^{i-1}\sum_{y=1}^{j-1}\sum_{k=x+1}^{i} \tau_{V_kV_i}\sigma_{V_xV_y}\lambda_{V_y\to V_j}\lambda_{V_x\to V_k} \\
= \; & \sigma_{V_iV_j} - \sum_{y=1}^{j-1} \sigma_{V_iV_y}\lambda_{V_y\to V_j} \\
& + \sum_{x=1}^{i-1}\left(\tau_{V_xV_i} - \sum_{k=x+1}^{i} \lambda_{V_x\to V_k}\tau_{V_kV_i}\right)\sigma_{V_xV_j} \\
& - \sum_{x=1}^{i-1}\sum_{y=1}^{j-1}\left(\tau_{V_xV_i} - \sum_{k=x+1}^{i} \lambda_{V_x\to V_k}\tau_{V_kV_i}\right)\sigma_{V_xV_j}\lambda_{V_y\to V_j} \\
= \; & \sigma_{V_iV_j} - \sum_{y=1}^{j-1} \sigma_{V_iV_y}\lambda_{V_y\to V_j}.
\end{aligned}$$

This is because:

$$\tau_{V_xV_i} = \sum_{k=x+1}^{i} \lambda_{V_x\to V_k}\tau_{V_kV_i},$$

for all $x = 1, \ldots, i-1$, since we have $\tau_{V_xV_i} = f_{\tau_{V_xV_i}}(\Theta_\mathcal{G})$, for all $x, i = 1, \ldots, n$ and $x < i$ (we can prove this by recursion, similar to how we prove Equation 12). Therefore, we have:

$$\sigma_{V_iV_j} - \sum_{y=1}^{j-1} \sigma_{V_iV_y}\lambda_{V_y\to V_j} = \sum_{k=1}^{i} \tau_{V_kV_i}\omega_{V_k\leftrightarrow V_j},$$

which is equivalent to Equation 11. Therefore, we have $\sigma_{V_iV_j} = f_{\sigma_{V_iV_j}}(\Theta_\mathcal{G})$ for all $i, j = 1, \ldots, n$ and $i < j$, and we have the following theorem.

Theorem 2 *The following set of equations for all $i, j = 1, \ldots, n$ and $i < j$:*

$$\begin{aligned}
\sigma_{V_iV_j} &= f_{\sigma_{V_iV_j}}(\Theta_\mathcal{G}), \\
\tau_{V_iV_j} &= f_{\tau_{V_iV_j}}(\Theta_\mathcal{G}), \\
\sigma_{V_iV_i} &= f_{\sigma^2_{V_i}}(\Theta_\mathcal{G}),
\end{aligned}$$

are satisfied if and only if for all $i, j = 1, \ldots, n$ and $i < j$:

$$\begin{aligned}
\upsilon_{V_iV_j\|\mathbf{V}} &= f_{\omega_{V_i\leftrightarrow V_j}}(\Theta_\mathcal{G}), \\
\tau_{V_iV_j} &= f_{\tau_{V_iV_j}}(\Theta_\mathcal{G}), \\
\sigma_{V_iV_i} &= f_{\sigma^2_{V_i}}(\Theta_\mathcal{G}).
\end{aligned}$$

For the equations based on the Instrumental Variable Function, the right-hand-sides only contain the function $f_{\omega_{V_i\leftrightarrow V_j}}(\Theta_\mathcal{G})$, which is trivially 0 if $V_i \leftrightarrow V_j \notin \mathbf{E}$. The advantage of using the set of equations based on the Instrumental Variable Function is that for the problem of finding the solution space of any subset of causal effects, and hence for the problem of identification of causal effects, we can consider only the subset of equations where the right-hand-side is trivially 0. Given Σ and T, assume that parameters $\Theta_\mathcal{G}$ satisfy the following set of equations for all $i, j = 1, \ldots, n$ and $i < j$:

$$\begin{aligned}
\upsilon_{V_iV_j\|\mathbf{V}} &= 0, \text{ if } V_i \leftrightarrow V_j \notin \mathbf{E}, \\
\tau_{V_iV_j} &= f_{\tau_{V_iV_j}}(\Theta_\mathcal{G}), \\
\sigma_{V_iV_i} &= f_{\sigma^2_{V_i}}(\Theta_\mathcal{G}).
\end{aligned}$$

Since the above functions of $\Theta_\mathcal{G}$ do not involve any of the parameters $\omega_{V_i\leftrightarrow V_j}$ for $V_i \leftrightarrow V_j \in \mathbf{E}$, there must exist $\Theta'_\mathcal{G}$, where for all $i, j = 1, \ldots, n$ and $i < j$:

$$\begin{aligned}
\upsilon_{V_iV_j\|\mathbf{V}} &= f_{\omega_{V_i\leftrightarrow V_j}}(\Theta'_\mathcal{G}), \\
f_{\lambda_{V_i\to V_j}}(\Theta_\mathcal{G}) &= f_{\lambda_{V_i\to V_j}}(\Theta'_\mathcal{G}), \\
f_{\sigma^2_{V_i}}(\Theta_\mathcal{G}) &= f_{\sigma^2_{V_i}}(\Theta'_\mathcal{G}),
\end{aligned}$$

and the following equations are satisfied:

$$\begin{aligned}
\upsilon_{V_iV_j\|\mathbf{V}} &= f_{\omega_{V_i\leftrightarrow V_j}}(\Theta'_\mathcal{G}), \\
\tau_{V_iV_j} &= f_{\tau_{V_iV_j}}(\Theta'_\mathcal{G}), \\
\sigma_{V_iV_i} &= f_{\sigma^2_{V_i}}(\Theta'_\mathcal{G}).
\end{aligned}$$

Therefore, we have $\Sigma = f_\Sigma(\Theta'_\mathcal{G})$ and $T = f_T(\Theta_\mathcal{G}) = f_T(\Theta'_\mathcal{G})$. Moreover, $\Theta'_\mathcal{G}$ also satisfies the first set of equations, since they are a subset of the last set of equations. Therefore, we have the following theorem.

Theorem 3 *The following set of equations for all $i, j = 1, \ldots, n$ and $i < j$:*

$$\begin{aligned} \sigma_{V_iV_j} &= f_{\sigma_{V_iV_j}}(\Theta_\mathcal{G}), \\ \tau_{V_iV_j} &= f_{\tau_{V_iV_j}}(\Theta_\mathcal{G}), \\ \sigma_{V_iV_i} &= f_{\sigma^2_{V_i}}(\Theta_\mathcal{G}), \end{aligned}$$

are satisfied if and only if for all $i, j = 1, \ldots, n$ and $i < j$:

$$\begin{aligned} \upsilon_{V_iV_j\|\mathbf{V}} &= 0, \text{ if } V_i \leftrightarrow V_j \notin \mathbf{E}, \\ \tau_{V_iV_j} &= f_{\tau_{V_iV_j}}(\Theta_\mathcal{G}), \\ \sigma_{V_iV_i} &= f_{\sigma^2_{V_i}}(\Theta_\mathcal{G}). \end{aligned}$$

Relations with Graphical Solutions of the Identification Problem

Besides algebraic methods, it is also possible to check whether the causal effect $\tau_{V_xV_y}$, for $V_y \in \mathbf{V}$ and $V_x \in \mathbf{V} \setminus \{\{V_y\} \cup Desc(V_y)\}$, is identifiable by testing for certain graphical criteria in the causal diagram $\mathcal{G}$ of the linear SEM. These tests include the back-door criterion (Pearl 2009), the instrumental variable (Bowden and Turkington 1984; Brito and Pearl 2002b), and the path-specific instrumental variable (Chan and Kuroki 2010).

Back-door criterion $\mathbf{V_z} \subseteq \mathbf{V} \setminus \{\{V_x\} \cup Desc(V_x) \cup \{V_y\} \cup Desc(V_y)\}$ satisfies the *back-door criterion* relative to (V_x, V_y), if after removing all outgoing directed edges from X from $\mathcal{G}$, i.e., in $\mathcal{G}' = (\mathbf{V}, \mathbf{E} \setminus \{V_x \to V : V \in \mathbf{V}\})$, V_x and V_y are d-separated given $\mathbf{V_z}$. Then the total causal effect $\tau_{V_xV_y}$ is identifiable, and we have the following equation:

$$\tau_{V_xV_y} = \frac{\sigma_{V_yV_x|\mathbf{V_z}}}{\sigma_{V_xV_x|\mathbf{V_z}}}. \tag{21}$$

Instrumental variable Given $\mathbf{V_z} \subseteq \mathbf{V} \setminus \{\{V_x\} \cup Desc(V_x) \cup \{V_y\} \cup Desc(V_y)\}$, $V_w \in \mathbf{V} \setminus \{Desc(V_x) \cup Desc(V_y)\}$ is an instrumental variable relative to (V_x, V_y), if after removing all outgoing directed edges from V_x from $\mathcal{G}$, i.e., in $\mathcal{G}' = (\mathbf{V}, \mathbf{E} \setminus \{V_x \to V : V \in \mathbf{V}\})$, V_w and V_y are d-separated given $\mathbf{V_z}$, and V_w and V_x are d-connected given $\mathbf{V_z}$. Then the total causal effect $\tau_{V_xV_y}$ is identifiable, and we have the following equation:

$$\tau_{V_xV_y} = \frac{\sigma_{V_yV_w|\mathbf{V_z}}}{\sigma_{V_xV_w|\mathbf{V_z}}}. \tag{22}$$

Path-specific instrumental variable Given $\mathbf{V_z} \subseteq \mathbf{V} \setminus \{\{V_x\} \cup Desc(V_x) \cup \{V_y\} \cup Desc(V_y)\}$, $V_w \in \mathbf{V}$ is a path-specific instrumental variable relative to (V_x, V_y), if after removing all outgoing directed edges from V_x and V_y from $\mathcal{G}$, i.e., in $\mathcal{G}' = (\mathbf{V}, \mathbf{E} \setminus \{\{V_x \to V : V \in \mathbf{V}\} \cup \{V_y \to V : V \in \mathbf{V}\}\})$, V_w and V_y are d-separated given $\mathbf{V_z}$, and V_w and V_x are d-connected given $\mathbf{V_z}$. Then the total causal effect $\tau_{V_xV_y}$ is identifiable if $\tau_{V_yV_w}$ and $\tau_{V_xV_w|V_y}$ are also identifiable, and we have the following equation:

$$\tau_{V_xV_y} = \frac{\sigma_{V_yV_w|\mathbf{V_z}} - \sigma_{V_yV_y|\mathbf{V_z}}\tau_{V_yV_w} - \sigma_{V_yV_x|\mathbf{V_z}}\tau_{V_xV_w|V_y}}{\sigma_{V_xV_w|\mathbf{V_z}} - \sigma_{V_xV_y|\mathbf{V_z}}\tau_{V_yV_w} - \sigma_{V_xV_x|\mathbf{V_z}}\tau_{V_xV_w|V_y}}. \tag{23}$$

To summarize, the above graphical identification methods are based on the following procedure. If some graphical test is satisfied, an equation relating this causal effect with certain covariances and other causal effects is satisfied, and this equation can be used (together with other equations if other causal effects are involved) to solve for this causal effect.

We can express the equations above in terms of the Instrumental Variable Function (proof omitted for length, to be included in longer version). The equation for the back-door criterion (Equation 21) can be expressed as:

$$\upsilon_{V_yV_x\|\{V_x\}\cup\mathbf{V_z}} = 0. \tag{24}$$

The equation for the instrumental variable (Equation 22) can be expressed as:

$$\upsilon_{V_yV_w\|\{V_x\}\cup\mathbf{V_z}} = 0. \tag{25}$$

The equation for the path-specific instrumental variable (Equation 23) can be expressed as:

$$\upsilon_{V_yV_w\|\{V_x\}\cup\{V_y\}\cup\mathbf{V_z}} = 0. \tag{26}$$

Note that the above graphical tests can be adapted for the identification of the direct causal effect $\lambda_{V_x \to V_y}$ and the conditional causal effect $\tau_{V_xV_y|\mathbf{V_z}}$. Moreover, they can all be extended for the identification of multiple causal effects $\tau_{V_{x_1}V_y} \ldots, \tau_{V_{x_k}V_y}$, as long as certain restrictions are satisfied (see *G-criterion* (Brito and Pearl 2006) and *accessory sets* (Tian 2007a)), which guarantees that the system of equations we use to solve are linearly independent. While we do not give the explicit tests and equations here, our results above can also be applied on these tests.

Conclusion

In this paper, we introduced the Instrumental Variable Function, which we showed to be useful in understanding the identification problem of causal effects in linear SEMs. We first show that this function gives us a set of simultaneous equations which we can use to solve for the causal effect, and we showed to be sound and complete. We then relate the equations from the graphical tests of identifiability to the Instrumental Variable Function. Our next step is thus to show that testing when the Instrumental Variable Function for certain variables is zero is related to testing for d-separation after removing certain edges, and providing an algorithm to generate a set of equations based on Instrumental Variable Function to solve the identification problem.

References

Bollen, K. A. 1989. *Structural Equations with Latent Variables*. John Wiley.

Bowden, R. J., and Turkington, D. A. 1984. *Instrumental Variables*. Cambridge University Press.

Brito, C., and Pearl, J. 2002a. Generalized instrumental variables. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, 85–93. Morgan Kaufmann Publishers.

Brito, C., and Pearl, J. 2002b. A graphical criterion for the identification of causal effects in linear models. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI)*, 533–538. AAAI Press.

Brito, C., and Pearl, J. 2002c. A new identification condition for recursive models with correlated errors. *Structural Equation Modeling* 9(4):459–474.

Brito, C., and Pearl, J. 2006. Graphical condition for identification in recursive SEM. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI)*, 47–54. AUAI Press.

Chan, H., and Kuroki, M. 2010. Using descendants as instrumental variables for the identification of direct causal effects in linear SEMs. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 73–80. JMLR.

Duncan, O. D. 1975. *Introduction to Structural Equation Models*. Academic Press.

Fisher, F. M. 1966. *The Identification Problem in Econometrics*. McGraw-Hill.

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition. 1st edition, 2000.

Spirtes, P.; Glymour, C.; and Scheines, R. 2000. *Causation, Prediction, and Search*. MIT Press.

Sullivant, S.; Garcia-Puente, L. D.; and Spielvogel, S. 2010. Identifying causal effects with computer algebra. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI)*. AUAI Press.

Tian, J. 2004. Identifying linear causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI)*, 104–111. AAAI Press.

Tian, J. 2005. Identifying direct causal effects in linear models. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI)*, 346–353. AAAI Press.

Tian, J. 2007a. A criterion for parameter identification in structural equation models. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence (UAI)*, 392–399. AUAI Press.

Tian, J. 2007b. On the identification of a class of linear models. In *Proceedings of the Twenty-Second National Conference on Artificial Intelligence (AAAI)*, 1284–1289. AAAI Press.

Wright, S. 1934. The method of path coefficients. *Annals of Mathematical Science* 5(3):161–215.

Appendix

For ease of notation, we define $\mathbf{V}_{\mathbf{k}\cup a} = \mathbf{V}_\mathbf{k} \cup \{V_a\}$ and $\mathbf{V}_{\mathbf{k}\setminus k} = \mathbf{V}_\mathbf{k} \setminus \{V_k\}$ (similar notations for multiple elements and operations).

The Instrumental Variable Function is symmetric, since we have:

$$\begin{aligned}
& \upsilon_{V_y V_w \| \mathbf{V_x}} \\
= \; & \rho_{V_y V_w \| \mathbf{V}_{\mathbf{x}\setminus w}} - \sum_{V_x \in \mathbf{V}_{\mathbf{x}\setminus y}} \rho_{V_x V_w \| \mathbf{V}_{\mathbf{x}\setminus w}} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\setminus x,y}} \\
= \; & \sigma_{V_y V_w} - \sum_{V_k \in \mathbf{V}_{\mathbf{x}\setminus w}} \sigma_{V_y V_k} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\setminus k,w}} \\
& - \sum_{V_x \in \mathbf{V}_{\mathbf{x}\setminus y}} \sigma_{V_x V_w} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\setminus x,y}} \\
& + \sum_{V_x \in \mathbf{V}_{\mathbf{x}\setminus y}} \sum_{V_k \in \mathbf{V}_{\mathbf{x}\setminus w}} \sigma_{V_x V_k} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\setminus k,w}} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\setminus x,y}} \\
= \; & \sigma_{V_w V_y} - \sum_{V_x \in \mathbf{V}_{\mathbf{x}\setminus y}} \sigma_{V_w V_x} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\setminus x,y}} \\
& - \sum_{V_k \in \mathbf{V}_{\mathbf{x}\setminus w}} \sigma_{V_k V_y} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\setminus k,w}} \\
& + \sum_{V_k \in \mathbf{V}_{\mathbf{x}\setminus w}} \sum_{V_x \in \mathbf{V}_{\mathbf{x}\setminus y}} \sigma_{V_k V_x} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\setminus x,y}} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\setminus k,w}} \\
= \; & \rho_{V_w V_y \| \mathbf{V}_{\mathbf{x}\setminus y}} - \sum_{V_k \in \mathbf{V}_{\mathbf{x}\setminus w}} \rho_{V_k V_y \| \mathbf{V}_{\mathbf{x}\setminus y}} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\setminus k,w}} \\
= \; & \upsilon_{V_w V_y \| \mathbf{V_x}}.
\end{aligned}$$

Moreover, given $V_a \notin \{\{V_w, V_y\} \cup Anc(V_w) \cup Anc(V_y)\}$, we have:

$$\begin{aligned}
& \upsilon_{V_y V_w \| \mathbf{V}_{\mathbf{x}\cup a}} \\
= \; & \sigma_{V_y V_w} - \sum_{V_k \in \mathbf{V}_{\mathbf{x}\cup a\setminus w}} \sigma_{V_y V_k} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\cup a\setminus k,w}} \\
& - \sum_{V_x \in \mathbf{V}_{\mathbf{x}\cup a\setminus y}} \sigma_{V_x V_w} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\cup a\setminus x,y}} \\
& + \sum_{V_x \in \mathbf{V}_{\mathbf{x}\cup a\setminus y}} \sum_{V_k \in \mathbf{V}_{\mathbf{x}\cup a\setminus w}} \sigma_{V_x V_k} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\cup a\setminus k,w}} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\cup a\setminus x,y}} \\
= \; & \sigma_{V_y V_w} - \sum_{V_k \in \mathbf{V}_{\mathbf{x}\setminus w}} \sigma_{V_y V_k} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\cup a\setminus k,w}} \\
& - \sum_{V_x \in \mathbf{V}_{\mathbf{x}\setminus y}} \sigma_{V_x V_w} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\cup a\setminus x,y}} \\
& + \sum_{V_x \in \mathbf{V}_{\mathbf{x}\setminus y}} \sum_{V_k \in \mathbf{V}_{\mathbf{x}\cup a\setminus w}} \sigma_{V_x V_k} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\cup a\setminus k,w}} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\cup a\setminus x,y}} \\
= \; & \sigma_{V_y V_w} - \sum_{V_k \in \mathbf{V}_{\mathbf{x}\setminus w}} \sigma_{V_y V_k} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\setminus k,w}} \\
& - \sum_{V_x \in \mathbf{V}_{\mathbf{x}\setminus y}} \sigma_{V_x V_w} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\setminus x,y}} \\
& + \sum_{V_x \in \mathbf{V}_{\mathbf{x}\setminus y}} \sum_{V_k \in \mathbf{V}_{\mathbf{x}\setminus w}} \sigma_{V_x V_k} \tau_{V_k V_w | \mathbf{V}_{\mathbf{x}\setminus k,w}} \tau_{V_x V_y | \mathbf{V}_{\mathbf{x}\setminus x,y}}
\end{aligned}$$

$$= \upsilon_{V_y V_w \| \mathbf{V_x}}.$$

The second step is due to $\tau_{V_a V_w | \mathbf{V}_{\mathbf{x} \cup a \setminus a, w}} = 0$ and $\tau_{V_a V_y | \mathbf{V}_{\mathbf{x} \cup a \setminus a, y}} = 0$ (since $V_a \notin \{\{V_w, V_y\} \cup Anc(V_w) \cup Anc(V_y)\}$). The third step is due to $\tau_{V_k V_w | \mathbf{V}_{\mathbf{x} \cup a \setminus k, w}} = \tau_{V_k V_w | \mathbf{V}_{\mathbf{x} \setminus k, w}}$ for all $V_k \in \mathbf{V_x}$, and $\tau_{V_x V_y | \mathbf{V}_{\mathbf{x} \cup a \setminus x, y}} = \tau_{V_x V_y | \mathbf{V}_{\mathbf{x} \setminus x, y}}$ for all $V_x \in \mathbf{V_x}$ (also since $V_a \notin \{\{V_w, V_y\} \cup Anc(V_w) \cup Anc(V_y)\}$ as V_a cannot block any potential causal effects from any variable to either V_w or V_y).