COMP3224

Coursework III

Alberto Tamajo

Student ID: 30696844

March 2022
Electronics and Computer Science Department
University of Southampton

Contents

1 Introduction

This is a 6-pages report providing solutions to the exercises proposed in the Exercise sheet. It also contains a list of Appendices that delve deeper into the causal reasoning concepts encountered in the exercises. It is highly recommended to read the Appendices to appreciate the quality of this report and the rationale under which the tasks of this exercise have been carried out. Furthermore, the Appendices prove a deep understanding of the material covered in the lectures. Notice that some of the Appendices provide a whole more formal and clear re-elaboration of Section 3.8 of "Causal inference in Statistics: A Primer. By Judea Pearl, Madeleyn Glymour, Nicholas P. Jewell". Indeed, I have realised that some statements in the book lack generality or are badly presented.

2 Exercise 1

2.1 Introduction

Exercise 1 requires us to show, by means of a practical example, how the existence of a confounder introduces bias in regression coefficients. Specifically, firstly, we are demanded to set up a linear Structural Causal Model (U, V, F) whose corresponding Graph Causal Model is identical to the one shown in the Exercise sheet. Secondly, we are required to use such SCM to generate a dataset D. Finally, we need to demonstrate that while regressing Y on X and Z let us recover the total causal coefficient of X on Y, regressing only on X gives us a spurious causal effect.

2.2 Linear SCM setup

In order to show empirically that confounders introduce bias in regression coefficients, the linear SCM $M = (U, V, F)$ is employed, where:

$$\begin{aligned}
&U = \{U_X \sim \mathcal{N}(0,1)\ ,\ U_Y \sim \mathcal{N}(0,1)\ ,\ U_Z \sim \mathcal{N}(0,1)\}\\
&V = \{X, Y, Z\}\\
&F = \{f_Z = 1.0 + U_Z\ ,\ f_X = 1.5Z + 10 + U_X\ ,\ f_Y = -1.5X + 3.0Z + 4.0 + U_Y\}
\end{aligned}$$

2.3 Real total causal coefficient

Before demonstrating that confounders introduce bias in causal coefficients, we need to know the real magnitude of the total causal effect of X on Y in M. Sub-Appendix A.2 shows that the Graph Causal Model counterparts of linear SCMs can be annotated with path coefficients. Thus, we can produce a GCM G_M annotated with the corresponding path coefficients for M. G_M is illustrated in Figure 1.

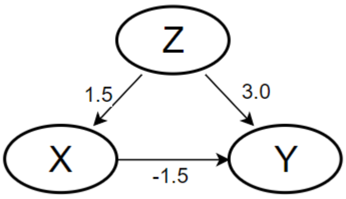

Figure 1: G_M, the GCM counterpart of M

Sub-Appendix A.2 also demonstrates that the total causal coefficient of X on Y is simply the sum of the products of the path coefficients of the edges on every non-backdoor path from X to Y. As a consequence, given that the only non-backdoor path between X and Y in G_M is the path $X \rightarrow Y$, then the total causal coefficient of X on Y is the path coefficient on the arrow from X to Y. Thus, the total causal effect is -1.5.

In the specific case of G_M, the total causal effect of X on Y coincides with the direct causal effect of X on Y.

2.4 Total causal coefficient identification and estimation with the backdoor criterion from observational data

Sub-Appendix A.3 states that we can identify and estimate the total causal coefficient of X on Y in linear SCMs with the help of the backdoor criterion. It also outlines a three-stage process to perform such identification and estimation. In what follows, this three-stage process is used to recover the total causal effect of X on Y from a dataset D generated through M.

The first stage consists of finding a set of variables in G_M that satisfies the backdoor criterion from X to Y. There exists only one backdoor path from X to Y, which is $X \leftarrow Z \rightarrow Y$. d-separation tell us that conditioning on Z blocks the path $X \leftarrow Z \rightarrow Y$. Therefore, the set $\{Z\}$ satisfies the backdoor criterion from X to Y.

With the help of the second stage's instructions, we formulate the regression equation

$$Y = \beta_X X + \beta_Z Z$$

Note that in G_M, the set P is just the empty set. Thus, we do not need to adjust for any other variable except for Z in the regression equation. For more information on the definition of the set P, the reader is referred to Sub-Sub-Appendix A.4.1.

Finally, we have to find the least-squares solution of Y with respect to the dataset D. The value of β_X after the regression process estimates the true total causal coefficient of X on Y.

Figure 3 proves empirically that as the size of D gets larger and larger, the process outlined so far better estimates the actual total causal coefficient of X on Y in G_M.

2.5 Spurious total coefficient identification and estimation

The process outlined in the previous sub-section is the correct one. We need to regress Y on X and Z to estimate the total causal coefficient of X on Y. However, if we regressed Y solely on X, what total causal coefficient would we obtain?

Figure 3 clearly shows that regressing Y on X produces a spurious result. The reasoning for this is that regressing on Z is necessary to block the backdoor path from X to Y; otherwise, the coefficient β_X resulting from the regression process ends up capturing spurious information which does not have a causative nature.

3 Exercise 2

3.1 Introduction

This exercise involves altering the SCM M constructed in the previous task and showing how the backdoor criterion can recover the strength of the causal effect of X on Y in the new SCM. The modifications to be made consist of introducing an additional interaction term in Y's function and replacing the Gaussian random variable Z with one drawn from a binomial distribution. Appendix H provides some more explanations regarding the effects of adding the additional interaction term in Y's function.

3.2 SCM setup

By following the guidelines of this exercise, a new SCM $M = (U, V, F)$ is obtained, where:

$$\begin{aligned}
&U = \{U_X \sim \mathcal{N}(0,1)\ ,\ U_Y \sim \mathcal{N}(0,1)\ ,\ U_Z \sim \text{Bernoulli}(0.5)\} \\
&V = \{X, Y, Z\} \\
&F = \{f_Z = 1.0 + U_Z\ ,\ f_X = 1.5Z + 10 + U_X\ ,\ f_Y = -1.5X + 3.0Z + 3.0XZ + 4.0 + U_Y\}
\end{aligned}$$

3.3 Real total causal coefficient

We need to learn the real total causal coefficient of X on Y in M prior to empirically proving that the backdoor criterion enables its estimation. In the previous task, we have leveraged the fact that in linear SCMs the total causal coefficient of X on Y is simply the sum of the products of the path coefficients of the edges on every non-backdoor path from X to Y. However, this cannot be done here. Indeed, M is not a linear SCM due to the extra term in f_Y.

Sub-Appendix A.2 provides a formula, named DCC, to compute the direct causal coefficient of a variable on another regardless of the type of functions contained in a SCM. Given that in M there are no mediating variables between X and Y, then the total causal coefficient of X on Y coincides with the direct causal effect. Hence, we can use the DCC equation to learn how X affects Y:

$$
\begin{aligned}
DCC &= \mathbb{E}[Y \mid do(X = x+1), do(Z = z)] - \mathbb{E}[Y \mid do(X = x), do(Z = z)] \\
&= \mathbb{E}[-1.5(x+1) + 3.0z + 3.0(x+1)z + 4.0 + U_Y] - \mathbb{E}[-1.5x + 3.0z + 3.0xz + 4.0 + U_Y] \\
&= \mathbb{E}[-1.5(x+1) + 3.0z + 3.0(x+1)z + 4.0 + U_Y - (-1.5x + 3.0z + 3.0xz + 4.0 + U_Y)] \\
&= \mathbb{E}[-1.5(x+1) + 3.0z + 3.0(x+1)z + 4.0 + U_Y + 1.5x - 3.0z - 3.0xz - 4.0 - U_Y)] \\
&= \mathbb{E}[-1.5 + 3.0z] \\
&= \mathbb{E}[-1.5] + 3.0\,\mathbb{E}[z] \\
&= -1.5 + 3.0 \cdot 0.5 \\
&= 0
\end{aligned}
$$

The DCC formula demonstrates that the total causal coefficient of X on Y is 0.

3.4 Total causal coefficient identification and estimation with the backdoor criterion from observational data

Sub-Appendix A.3 claims that we can identify and estimate the total causal coefficient of a variable on another in linear SCMs with the help of the backdoor criterion. It also describes a three-stage process to perform such identification and estimation. Despite M being a non-linear SCM, the approach outlined in Sub-Appendix A.3 can be carried out anyway as it is far more general. In what follows, this three-stage process is used to recover the total causal effect of X on Y from a dataset D generated through M.

The first stage involves finding a set of variables that satisfies the backdoor criterion from X to Y. For the same reasons as in the previous exercise, $\{Z\}$ satisfies the backdoor criterion from X to Y.

With the help of the second stage's guidelines, the following regression equation is formulated:

$$Y = \beta_X X + \beta_Z Z$$

Note that the set P is just the empty set. Thus, we do not need to adjust for any other variable except for Z in the regression equation. For more information on the definition of the set P, the reader is referred to Sub-Sub-Appendix A.4.1.

Finally, we have to find the least-squares solution of Y with respect to the dataset D. The value of β_X after the regression process estimates the true total causal coefficient of X on Y.

Figure 4 proves empirically that as the size of D gets larger and larger, the process outlined so far better estimates the actual total causal coefficient of X on Y.

4 Exercise 3

4.1 Introduction

Exercise 3 demands us to demonstrate through an example that it is possible to identify and estimate the total causal effect of a variable on another by means of the front-door criterion. Precisely, to start with, a linear SCM M needs to be constructed such that M's corresponding GCM is equivalent to the one illustrated in the Exercise sheet. Successively, we need to employ M to generate a dataset D. Although the confounder variable Z is contained in M, Z should not appear in D. The reasoning for this relies on this exercise's objective. Indeed, this exercise is designed to show us that even in the presence of a hidden variable we cannot adjust for, the causal effect's magnitude can still be recovered if a mediating variable exists. The last step consists of regressing Y and M separately and computing a product of coefficients so to estimate the total causal effect of T on Y.

4.2 Linear SCM setup

To carry out this exercise, the linear SCM $M = \{U, V, F\}$ is utilised, where:

$$
\begin{aligned}
&U = \{U_T \sim \mathcal{N}(0,1)\ ,\ U_Y \sim \mathcal{N}(0,1)\ ,\ U_Z \sim \mathcal{N}(0,1)\ ,\ U_M \sim \mathcal{N}(0,1)\} \\
&V = \{T, Y, Z, M\} \\
&F = \{f_Z = -2.0 + U_Z\ ,\ f_T = -3.0Z + 1.0 + U_T\ ,\ f_M = 15.0T - 3.0 + U_M\ ,\ f_Y = 2.0M + -4.0Z + 4.0 + U_Y\}
\end{aligned}
$$

4.3 Real total causal coefficient

The actual total causal coefficient of T on Y in M needs to be learned before empirically proving that the front-door criterion allows us to estimate it. Sub-Appendix A.2 shows that the Graph Causal Model counterparts of linear SCMs can be annotated with path coefficients. Thus, we can produce a GCM G_M annotated with the corresponding path coefficients for M. G_M is illustrated in Figure 2.

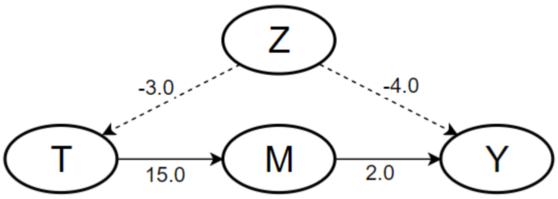

Figure 2: G_M, the GCM counterpart of M

Sub-Appendix A.2 also demonstrates that the total causal coefficient of T on Y is simply the sum of the products of the path coefficients of the edges on every non-backdoor path from T to Y. As a consequence, given that the only non-backdoor path between T and Y in G_M is the path $T \to M \to Y$, then the total causal coefficient of T on Y is the product of the path coefficients on the arrows from T to M and from M to Y. Thus, the total causal effect is 30.0. Unlike Exercise 1, the total causal effect of T on Y in G_M does not coincide with the direct causal effect of T on Y. In fact, the latter is equivalent to 0 since there is no direct edge from T to Y.

4.4 Total causal coefficient identification and estimation with the front-door criterion from observational data

The identification and estimation procedure applied in Exercise 1 cannot be adopted here as we are pretending that the confounder variable Z is unobserved. Fortuitously for us, Sub-Appendix A.3 claims that we can identify and estimate the total causal coefficient of a variable on another in linear SCMs with the help of the front-door criterion. It also outlines a four-stage process to perform such identification and estimation. In what follows, this four-stage process is used to recover the total causal effect of X on Y from a dataset D generated through M.

The first stage consists of finding a set of variables in G_M that satisfies the front-door criterion from T to Y. It turns out that the set $\{M\}$ intercepts all directed paths from T to Y, there exists no unblocked path from T to $\{M\}$ and all backdoor paths from $\{M\}$ to Y are blocked by T. Consequently, $\{M\}$ satisfies the front-door criterion.

Using the instructions of the second step, we formulate the regression equation

$$Y = \beta_M M + \beta_{T_1} T$$

Note that in G_M, the set P_1 is just the empty set. Thus, we do not need to adjust for any other variable except for T in the regression equation. For more information on the definition of the set P_1, the reader is referred to Sub-Sub-Appendix A.4.3.

The third step instructs us on how to formulate the second regression equation

$$M = \beta_{T_2} T$$

Note that in G_M, the set P_2 is just the empty set. Thus, we do not need to include any other variable except for T in the regression equation. For more information on the set P_2, the reader should read Sub-Sub-Appendix A.4.3.

In the final stage, the least-squares solutions of Y and M need to be computed separately with respect to the dataset D. The values of β_M and β_{T_2} after the regression process estimate the true total causal coefficients of M on Y and T on M, respectively. $\beta_M \times \beta_{T_2}$ gives us the actual total causal coefficient of T on Y.

This whole process makes perfect sense if we refer to the method used in the previous sub-section to compute the true total causal coefficient of X on Y. Here, we have only decomposed that process into two sub-processes. First, we have found the total causal coefficient of T on M. Then, the total causal coefficient of M on Y. By multiplying the previous two causal coefficients, we perform a computation that is equivalent to summing all the products of the path coefficients of the edges on every non-backdoor path from T to Y.

Figure 5 proves empirically that as the size of D approaches infinity, the process outlined here outputs an estimation that better approximates the total causal coefficient of T on Y in G_M.

5 Exercise 4

5.1 Introduction

This exercise involves reusing the SCM M instantiated in the previous task to generate a dataset D and recover the total causal coefficient of T on Y with the help of the backdoor criterion. Unlike in Exercise 3, here, we do not pretend that Z is a hidden variable; this is the reason why the identification and estimation process can be carried out through the backdoor criterion. Moreover, we are tasked to compare and evaluate under what noise conditions in M which method between the one adopted here and the one applied in the previous exercise provides more reliable estimates.

5.2 Linear SCM setup

The linear SCM M used to generate the dataset D is identical to the one constructed in the previous exercise. Its GCM counterpart G_M is equal to the GCM illustrated in the Exercise sheet.

5.3 Real total causal coefficient

Thanks to the previous exercise, we already know that the total causal coefficient of T on Y in G_M is 30.0.

5.4 Total causal coefficient identification and estimation with the backdoor criterion from observational data

In what follows, the three-stage process illustrated in Sub-Appendix A.3 is employed to recover the total causal effect of T on Y from a dataset D generated through M.

The first stage consists of finding a set of variables in G_M that satisfies the backdoor criterion from T to Y. There exists only one backdoor path from T to Y, which is $T \leftarrow Z \rightarrow Y$. d-separation tell us that conditioning on Z blocks the path $T \leftarrow Z \rightarrow Y$. Therefore, the set $\{Z\}$ satisfies the backdoor criterion from T to Y.

In the second stage, we formulate the regression equation

$$Y = \beta_T T + \beta_Z Z$$

In G_M, the set P is just the empty set. Thus, we do not need to adjust for any other variable except for Z in the regression equation. For more information on the definition of the set P, the reader is referred to Sub-Sub-Appendix A.4.1.

The last step involves computing the least-squares solution of Y with respect to the dataset D. The value of β_T after the regression process estimates the true total causal coefficient of T on Y.

Figure 6 proves empirically that as the size of D gets increasingly larger, the process outlined so far better estimates the total causal coefficient of T on Y in G_M.

5.5 Comparison between the backdoor criterion and front-door criterion approach

In order to compare and evaluate under what noise conditions in M which method between backdoor and front-door adjustment gives better estimates, several experiments have been conducted. Specifically, four experiments have been carried out taking inspiration from the causal reasoning concept of interventions. In each experiment, we have intervened on the standard deviation of a single exogenous variable, making it larger, while keeping all the other exogenous variables' standard deviations constant. Subsequently, a dataset D has been generated, and the causal coefficient estimation of T on Y has been computed using both the backdoor and front-door adjustment. The results of these four experiments are illustrated in what follows:

- when the standard deviation of the exogenous variable U_Y is increased, both backdoor and front-door adjustment output worse estimates. Though, front-door adjustment estimates are poorer than backdoor ones. Figure 7 illustrates this claim.
- when the standard deviation of the exogenous variable U_M is increased, both backdoor and front-door adjustment give worse estimates. However, backdoor adjustment estimates are poorer than front-door ones. Figure 8 illustrates this claim.
- when the standard deviation of the exogenous variable U_Z is increased, both backdoor and front-door adjustment estimates change only slightly.

- when the standard deviation of the exogenous variable U_T is increased, both backdoor and front-door estimates vary only a little.

Note that when all exogenous variables are standard normal distributions, the backdoor adjustment estimates are slightly better than front-door ones. In Appendix K, some intuitive explanations are given in order to let the reader gain an understanding of why the experiments have led to the above results.

6 Exercise 5

6.1 Introduction

This exercise asks us to define two linear functions $X \to Y$ so that to generate two datasets D_1, D_2 and fit two regression lines each. The first regression line regresses in the forward causal direction while the second one regresses in the backward effect direction. The only difference between the two linear functions lies in their noise variables ϵ_Y. Indeed, while the first linear function employs a gaussian noise, the second one uses a uniform noise. Subsequently, we need to comment on any differences we notice between the correlation of the residuals and the independent variable for both regression lines of each linear function. This exercise introduces us to some Structure identifiability methods (Appendix F) and lets us understand that, under some causal model assumptions, the joint distribution $P(X, Y)$ allows us to identify the underlying SCM.

6.2 Linear functions setup

In order to generate the datasets D_1 and D_2, the following linear functions $X \to Y$ are employed, respectively:

$$Y_1 = 1.5X_1 + U_{Y_1}$$
$$Y_2 = 1.5X_2 + U_{Y_2}$$

where $X_1 \sim \mathcal{N}(0,1)$, $X_2 \sim \mathcal{U}[0,1]$, $U_{Y_1} \sim \mathcal{N}(0,1)$ and $U_{Y_2} \sim \mathcal{U}[0,1]$

6.3 Structure identifiability for the first linear function

We know that the causal direction for dataset D_1 is $X_1 \to Y_1$. This holds because D_1's generative function defines Y_1 in terms of X_1. As such, X_1 is the causal variable and Y_1 is the effect variable. However, let us suppose for a moment that we do not know the underlying generative mechanism of D_1. If we assume a linear generative function, can we recover the causal direction from the joint distribution $P(X_1, Y_1)$ described by D_1? It turns out that we cannot identify the underlying structure of the causal mechanism because of the claim made by the *Identifiability of linear non-Gaussian models theorem* (Sub-Appendix F.3). Indeed, both X_1 and N_{Y_1} are gaussian distributions. Consequently, D_1 admits the backward effect model as well as the forward causal model. The fact that the causal mechanism underlying D_1 cannot be identified from the observed joint distribution is also demonstrated in Sub-Appendix M.1 via application of the *Independence of residuals test* (Appendix F.4). The dependence relations between the residuals and the independent variables are not proven through high-order statistical tools but by drawing correlation scatter plots and residuals' density plots.

6.4 Structure identifiability for the second linear function

We know that the causal direction for dataset D_2 is $X_2 \to Y_2$. However, let us suppose for a moment that we do not know the underlying generative mechanism of D_2. If we assume a linear generative function, can we recover the causal direction from the joint distribution $P(X_2, Y_2)$ described by D_2? It turns out that we can identify the underlying structure of the causal mechanism because of the claim made by the Identifiability of linear non-Gaussian models theorem (Appendix F.3). Indeed, none of X_1 and N_{Y_1} are gaussian distributions. Consequently, D_2 does not admit the backward effect model but only the forward causal model. The fact that the causal mechanism underlying D_2 can be identified from the observed joint distribution is demonstrated in Sub-Appendix M.2 via application of the *Independence of residuals test* (Appendix F.4). The dependence relations between the residuals and the independent variables are not proven through high-order statistical tools but by drawing correlation scatter plots and residuals' density plots.

Appendices

A Causal inference in linear Structural Causal Models

So far, we have looked at the backdoor criterion, front-door criterion and d-separation. A nice property of these causal reasoning tools is that they are valid regardless of the type of functions contained in a given Structural Causal Model(SCM). Indeed, their derivation is based on Graph Causal Models(GCM), which make no assumptions about the form of relationship between two variables, only that the relationship exists.

For a summary description of SCMs and GCMs, the reader is referred to Appendix D and E, respectively.

In this coursework, we leverage the backdoor criterion, front-door criterion and d-separation to apply causal inference in the context of linear SCMs. Specifically, we will see what causal assumptions and implications look like when dealing with SCMs containing only linear functions.

A.1 Structural Equation Modelling methods

Even though regression equations make no causal assumptions by default, we can apply the causal inference methods learnt so far, such as the backdoor criterion, front-door criterion and d-separation, to express the causal effect of a variable on another in terms of regression coefficients. Thus, it is possible to transform a causal inference problem into a regression problem, and this mapping from the causal inference domain to the regression domain comes with several advantages. The main benefit is that many software packages exist nowadays that allow us to compute partial regression coefficients very efficiently. As such, we can compute the strength of causal effects very efficiently as well.

All the methods that combine regression equations and causal reasoning are labelled as Structural Equation Modelling(SEM) methods. In the next sub-appendices, we will see in detail how to map a causal inference problem into a regression problem.

A.2 Direct and total causal coefficients in linear SCMs

A.2.1 Linear SCMs

A linear SCM is a Structural Causal Model containing only linear functions. Formally, a linear SCM is a Structural Causal Model (U, V, F) such that for every $f \in F$, f is a linear function. Thus, every variable $X \in V$ is affected in a linear fashion by its parents $par(X)$.

A.2.2 Direct causal coefficients

A nice property of linear SCMs is that we can fully specify the functions in the model by annotating their graph causal model counterparts with path coefficients. As such, given that f_Y is the linear function of variable Y, then

$$f_Y = \sum_{i}^{N} \beta_i X_i + U_Y$$

where X_i is the i^th parent of Y and β_i is the path coefficient on the arrow from X_i to Y.

β_i, the path coefficient on the arrow from X_i to Y, is the direct causal coefficient of X_i on Y. To see why this is so, we can refer to Appendix C. The latter claims that in order to find the direct causal effect of X_i on Y, we should keep any mediating variable Z steady. However, Appendix C considers the causal effect of X_i on Y from a probability perspective. Since our aim here is to find the direct causal coefficient of X_i on Y, we should keep any mediating variable Z steady and, additionally, any parent of variable Y except for X_i itself. This amounts to computing the partial derivative of f_Y with respect to X_i:

$$\frac{\partial f_Y}{\partial X_i} = \frac{\partial}{\partial X_i} \sum_{i}^{N} \beta_i X_i + U_Y = \beta_i$$

As we can see, the result is β_i. Intuitionally, the partial derivative of f_Y with respect to X_i tells us how Y changes with respect to X_i given that all of its other parent variables are kept constant. Therefore, we are isolating the direct causal effect of X_i on Y. This provides us with the direct causal effect coefficient.

Equivalently, we can also prove that β_i is the direct cause coefficient of X_i on Y by using the CDE formula in Appendix C, exchanging probabilities for expectations and intervening on any parent of Y as follows:

$$
\begin{aligned}
DCC = &\ \mathbb{E}[Y \mid do(X_i = x_i + 1), do(X_1 = x_1), ..., do(X_{i-1} = x_{i-1}), ..., do(X_{i+1} = x_{i+1}), ..., do(X_N = x_N)] - \\
&\ \mathbb{E}[Y \mid do(X_i = x_i), do(X_1 = x_1), ..., do(X_{i-1} = x_{i-1}), ..., do(X_{i+1} = x_{i+1}), ..., do(X_N = x_N)] \\
= &\ (\beta_i(x_i + 1) + \beta_1 x_1 + ... + \beta_{i-1}x_{i-1} + \beta_{i+1}x_{i+1} + \beta_N x_N + U_Y) - (\beta_i x_i + \beta_1 x_1 + ... + \beta_{i-1}x_{i-1} + \beta_{i+1}x_{i+1} + \beta_N x_N + U_Y) \\
= &\ \beta_i(x_i + 1) - \beta_i(x_i) \\
= &\ \beta_i
\end{aligned}
$$

DCC stands for Direct Causal Coefficient. It computes the difference in Y's value when increasing X_i by one unit and keeping all other variables constant. Thus, it gives us the direct causal coefficient of X_i on Y.

A.2.3 Total causal coefficients

We have seen that the path coefficient β_i is the direct causal coefficient of X_i on Y. However, frequently, we wish to find the total causal coefficient of X_i on Y. The total causal coefficient accounts for both the direct causal effect of X_i on Y and the indirect causal effect through a set of mediating variables. Computing the total causal coefficient of X_i on Y involves three steps:

- Find every non-backdoor path from X_i to Y.
- For each path, multiply all coefficients along the edges.
- Sum all the products together.

We do not prove the above claim as it can be easily derived with a little algebra. The reader is referred to ”Causal inference in Statistics: A Primer. By Judea Pearl, Madeleyn Glymour, Nicholas P. Jewell” for more details about the proof.

A.3 Causal coefficients identification and estimation from observational data

So far, we have seen how to express direct and total causal coefficients in terms of path coefficients, assuming that the latter are known a priori. Here, we tackle two more complicated problems, known as *causal coefficients identification* and *causal coefficients estimation.* Solving a causal coefficient identification problem amounts to identifying total or direct causal coefficients from observational data by writing down a regression equation. Once a regression equation is laid down, computing its least-squares solution gives us an estimation of the causal coefficient of interest. The latter step is used to solve a causal coefficient estimation problem.

Therefore, if we need to estimate a total or direct causal coefficient from observational data, we need to carry out a two-step process. Firstly, a regression equation needs to be formulated. Secondly, the least-squares solution of the regression equation needs to be computed to find out the estimation of the causal coefficient. This process is more formally outlined in what follows.

Let D be an observed dataset that we assume is generated by a causal mechanism M, such that M is a linear SCM (U, V, F). We are interested in estimating either the direct or total causal coefficient β_X of X on Y, where $X, Y \in V$. We need to carry out a two-step process to estimate β_X. The first step, known as causal coefficient identification, involves writing down a regression equation

$$Y = \beta_X X + \beta_{Z_1} Z_1 + ... + \beta_{Z_N} Z_N$$

where $Z_1, ..., Z_N \in V$ are to be chosen according to some criteria so that to get non-spurious responses. The criteria for selecting $Z_1, ..., Z_N$ are discussed in the next section.

The second step, named causal coefficient estimation, consists of finding the least-squares solution of

$$Y = \beta_X X + \beta_{Z_1} Z_1 + ... + \beta_{Z_N} Z_N$$

with respect to the observed dataset D. The value of β_X give us the estimation of either the direct or total causal coefficient of X on Y.

A.4 Criteria for causal coefficients identification

Given an observed dataset D and its underlying causal mechanism M, such that M is a linear SCM (U, V, F), the first step in the estimation of the direct or total causal coefficient of $X \in V$ on $Y \in V$ is called causal coefficient identification. The latter involves writing down a regression equation

$$Y = \beta_X X + \beta_{Z_1} Z_1 + ... + \beta_{Z_N} Z_N$$

where $Z_1, ..., Z_N \in V$. The selection of variables $Z_1, ..., Z_N$ depends on whether we are interested in the direct or total causal coefficient of X on Y and the topology of the Graph Causal Model corresponding to M.

A.4.1 Total causal coefficients identification with the backdoor criterion

The backdoor criterion gives us the set of variables Z we need to adjust for to determine the total causal effect of X on Y. It turns out that we can exploit this criterion to determine the total causal coefficient of X on Y in linear SCMs. In principle, once we obtain the set Z, we can estimate the conditional expectation of Y given X and Z. Then, we can average over Z and obtain a measure of the total causal effect of X on Y. All of this needs to be translated into the language of regression. Luckily, the translation is relatively simple. First, we need to find a set Z that satisfies the backdoor criterion from X to Y. Then, we formulate the regression equation

$$Y = \beta_X X + \beta_{Z_1} Z_1 + ... + \beta_{Z_N} Z_N + \beta_{P_1} P_1 + ... + \beta_{P_T} P_T$$

where $\{Z_1, ..., Z_N\} = Z$ and $\{P_1, ..., P_T\} = P$. The set P is defined as follows

$$P = \{P_i \mid P_i \text{ is an ancestor of } Y \text{ along the path containing the } i^{th} \text{ incoming edge of } Y \text{ such that this path is not a backdoor path from } X \text{ to } Y \text{ or a causal path from } X \text{ to } Y, i = 1, ..., T\}$$

If we regress Y on X, Z and P as in the regression equation above, the coefficient β_X give us the estimation of the true total causal coefficient of X on Y. The reasoning for this process is similar to the one behind the justification of the backdoor criterion. Regressing on Z blocks all backdoor paths from X and Y, thus preventing the coefficient β_X from capturing the spurious information that those paths contain. Similarly, regressing on P prevents β_X from absorbing causal information stemming from other ascendants of Y.

A.4.2 Direct causal coefficients identification with the backdoor criterion

In order to determine the direct causal effect of X on Y in linear SCMs, we need to apply a procedure similar to the one in the previous section, except that now, we also need to block all the indirect paths from X to Y. Specifically, first, we need to discover a set of variables Z_B that satisfies the backdoor criterion from X to Y. Afterwards, we need to find a set of variables Z_D that d-separates X and Y in the Graph Causal Model without the direct edge from X to Y. In the end, we formulate the regression equation

$$Y = \beta_X X + \beta_{Z_1} Z_1 + ... + \beta_{Z_N} Z_N + \beta_{P_1} P_1 + ... + \beta_{P_T} P_T$$

where $\{Z_1, ..., Z_N\} = Z_B \cup Z_D$ and $\{P_1, ..., P_T\} = P$. The set P is defined as follows

$$P = \{P_i \mid P_i \text{ is an ancestor of } Y \text{ along the path containing the } i^{th} \text{ incoming edge of } Y \text{ such that this path is not a backdoor path from } X \text{ to } Y \text{ or a causal path from } X \text{ to } Y, i = 1, ..., T\}$$

If we regress Y on X, $Z_B \cup Z_D$ and P as in the regression equation above, the coefficient β_X give us the estimation of the true direct causal coefficient of X on Y.

A.4.3 Total causal coefficients identification with the front-door criterion

While the backdoor criterion allows us to easily identify the total causal coefficients of X on Y, if the set of variables Z we need to adjust for is unobservable, we cannot block the backdoor paths from X to Y. In these scenarios, the front-door criterion allows us to identify a set of intermediate variables Z so to determine the total causal effect of X on Y through two consecutive applications of the backdoor criterion.

As we already know how to identify the total causal coefficient of X on Y in linear SCMs with the backdoor criterion, then we can also leverage the front-door criterion for the identification task. Intuitionally, we need to determine the total causal coefficient of Z on Y and then multiply it with the total causal coefficient of X on Z. This makes perfect sense as in Sub-Appendix A.2.3 we have seen that the total causal coefficient of X on Y is simply the sum of the products of the path coefficients of the edges on every non-backdoor path from X to Y. Unlike the previous identification criteria, this identification process involves three stages to determine the total causal coefficient of X on Y. For simplicity and without loss of generality, we are going to assume that the set of intermediary variables Z contains only one variable, Z_1. First, we need to find a set of variables Z that satisfies the front-door criterion between X and Y. Second, we need to formulate the regression equation

$$Y = \beta_{Z_1}Z_1 + \beta_{X_1}X + \beta_{P_{1,1}}P_{1,1} + ... + \beta_{P_{1,T}}P_{1,T}$$

where $\{P_{1,1}, ..., P_{1,T}\} = P_1$. The set P_1 is defined as follows

$$P_1 = \{P_{1,i} \mid P_{1,i} \text{ is an ancestor of } Y \text{ along the path containing the } i^{th} \text{ incoming edge of } Y \text{ such that this path is not a backdoor path from } Z_1 \text{ to } Y \text{ or a causal path from } Z_1 \text{ to } Y, i = 1, ..., T\}$$

Third, we need to formulate the regression equation

$$Z_1 = \beta_{X_2}X + \beta_{P_{2,1}}P_{2,1} + ... + \beta_{P_{2,T'}}P_{1,T'}$$

where $\{P_{2,1}, ..., P_{2,T}\} = P_2$. The set P_2 is defined as follows

$$P_2 = \{P_{2,i} \mid P_{2,i} \text{ is an ancestor of } Z_1 \text{ along the path containing the } i^{th} \text{ incoming edge of } Z_1 \text{ such that this path is not a backdoor path from } X \text{ to } Z_1 \text{ or a causal path from } X \text{ to } Z_1, i = 1, ..., T\}$$

If we regress Y on X, Z and P_1 as in the first regression equation above, the coefficient β_{Z_1} give us the estimation of the true total causal coefficient of Z on Y. Similarly, if we regress Z on X and P_2 as in the second regression equation, the coefficient β_{X_2} give us the estimation of the true total causal coefficient of X on Z. $\beta_{Z_1} \times \beta X_2$ is the estimation of the true total causal coefficient of X on Y.

B Regression Equations

In statistics, it is often necessary to find the prediction of a variable Y based on the value of other variables $X_1, X_2, ..., X_n$. The best possible prediction is given by the conditional expectation $E[Y|X_1 = x_1, ..., X_n = x_n]$. However, as it is often difficult to come up with the joint distribution $P(Y, X_1, ..., X_n)$ or the conditional distribution $P(Y|X_1, ..., X_n)$, we would like to make predictions directly from the data.

Regression equations allow us to do exactly this. Given a dataset $D = \{(y_t, x_{1,t}, ..., x_{n,t}) \mid t = 1, ..., T\}$, we typically try to find the least-squares regression line. That is, the line that takes observed values of $X_1, ..., X_n$ as input and gives values of Y as output, such that the square error between the predicted and actual values of Y is minimized, on average. Formally, the least-squares regression line $y = r_1x_1 + ... + r_nx_n + c$ is the line that minimises the following value:

$$\sum_t (y_t - y_t')^2 = \sum_t (y_t - r_1x_1 - ... - r_nx_n - c)^2$$

Let $1 \leq p \geq n$, r_p is also denoted as $R_{YX_p \cdot \{X_1,...,X_{p-1},X_{p+1},...,X_n\}}$. $R_{YX_p \cdot \{X_1,...,X_{p-1},X_{p+1},...,X_n\}}$ is called the partial regression coefficient of Y on X_p and describes the slope of Y on X_p given that we hold $\{X_1, ..., X_{p-1}, X_{p+1}, ..., X_n\}$ constant.

C Direct effect estimation

Frequently, a treatment variable X causes a variable Y both directly and indirectly through a set of mediating variables. However, in some scenarios, we are only interested in the direct effect of X on Y. For simplicity and without loss of generality, let us suppose that a treatment variable X causes Y directly and indirectly through the mediating variable Z. Intuitionally, in order to find the direct effect of X on Y, we should hold the mediating variable Z steady. Thus, any change in the value of Y would have to be due to the direct effect of X.

Traditionally, statistics has held the variable Z steady by conditioning on it. However, although conditioning on Z blocks the indirect causal path $X \rightarrow Z \rightarrow Y$, it may unblock a non-causal path $X \rightarrow Z \rightarrow ... \rightarrow Y$, leading to spurious results. Fortunately, Causal reasoning gives us a tool to hold the mediating variable steady without introducing spuriousness in our results. Indeed, to keep the value of a mediating variable constant, we can intervene on it. Thanks to interventions, we can define the Controlled Direct Effect(CDE) on Y of changing the value of X from x to x' as:

$$CDE = P(Y = y \mid do(X = x), do(Z = z)) - P(Y = y \mid do(X = x'), do(Z = z))$$

Note that to fully understand the direct effect of X on Y, we need to compute the value of the CDE for all relevant values of Z. The CDE's formula ensures that we can compute the direct effect of X on Y whenever we can identify the interventional probabilities from the observed data. This can be done from scratch via Do-calculus. Alternatively, we can use the Backdoor criterion to find a set of variables S we can adjust for, such that $S = S_1 \cup S_2$, where:

- S_1 is a set of variables that blocks all backdoor paths from Z to Y.
- S_2 is a set of variables that blocks all backdoor paths from X to Y, after deleting all incoming edges in Z.

D Structural Causal Models

A Structural Causal Model(SCM) is a model that describes the causal mechanism through which a certain data set has been generated. In other words, a SCM tries to model all the relevant variables that participate in a given phenomenon and how they interact with each other. Given that the causal mechanism behind a data set is known, it is possible to generate new data which exhibits the same patterns. Thus, SCM may be used to simulate new data as well.

Mathematically, a SCM is a three-tuple (U, V, F) where U is the set of exogenous variables, V is the set of endogenous variables, and F is the set of functions that assign each variable in V a value based on the values of the other variables in the model.

The set of exogenous variables U contains all those variables whose causes are not explained by the model. Such causes may be irrelevant to the problem or even unknown. Exogenous variables are crucial in a SCM as they stand in for any unknown or random effects that may alter the relationship between the endogenous variables. This holds because, in most natural phenomena, we only know a subset of causes for a given variable(the endogenous variables). Thus, to model the remaining subset of unknown causes, we need to fall back to the exogenous variables, which are commonly treated as random variables with an underlying distribution.

On the other hand, the endogenous variables' causes are delineated by the model. These causes may either be exogenous variables, endogeneous variables or a combination of both. If the value of every exogenous variable in a SCM is known, the functions in F enable the determination of each endogenous variable.

By means of Structural Causal Models, it is possible to define causation formally: **a variable X is a direct cause of a variable Y if Y's function is directly defined in terms of X. X is a cause of Y if it is a direct cause of Y, or of any cause of Y.**

It turns out that every SCM has an associated graphical causal model; the next Appendix will give more details about such models.

E Graph causal models

A graph causal model is a directed graph (L, E) that graphically models the relevant variables that participate in a given phenomenon and how they interact with each other. Every SCM can be mapped to a graph causal model through a function $T : (U, V, F) \rightarrow (L, E)$, where (L, E) denotes a graph with vertices L and edges E. The function $T(\cdot)$ is defined as follows:

$$T(U, V, F) = (U \cup V, \{(x, y) | x \text{ is an argument of the function} f_y \in F)\}$$

Substantially, given as input a SCM M, the function $T(\cdot)$ maps M to a graph where the nodes represent the exogenous and endogenous variables and the directed edges between the nodes represent the functions in F. It is important to notice that $T(\cdot)$ is not injective; thus, multiple SCMs have the same underlying graph causal model.

Because of the relationship between SCMs and graph models, we can give a graph definition of causation: **if in a graph model, the variable X is the parent of a variable Y, X is a direct cause of Y. Similarly, if the variable X is an ancestor of the variable Y, X is a likely cause of Y.**
If X is an ancestor of a variable Y, it does not necessarily imply that X is a cause of Y because the causation relation is not transitive. Therefore, X is said to be a likely cause of Y. Given the above definition of graph causation, it follows that the exogenous variables are root nodes in a graph causal model.

Although graph causal models are less informative than SCMs as they lack the quantitative aspect of causal relationships, they provide an intuitive understanding of the patterns of dependencies among the variables. Therefore, we can learn which variables in the data set are independent of each other and which are independent of each other conditional on other variables by merely looking at their graph causal model. The tool that allows understanding the dependency relations in a graph causal model of any complexity is called **d-separation**. d-separation builds upon a set of three rules that are derived from a set of three special graph causal models: chains, forks and colliders.

F Structure identifiability

F.1 Introduction

Structure identifiability comprises all those processes that aim to identify an underlying SCM from a joint distribution $P(X, Y)$. That is, given a joint distribution $P(X, Y)$, these methods try to understand whether the underlying causal mechanism is a SCM from X to Y or from Y to X.

F.2 Non-uniqueness of graph structures without additional assumptions

Without any additional assumptions about the relations between probability distributions and causality, the structure of the underlying causal mechanism cannot be identified from a given joint distribution $P(X, Y)$. The result proving this claim is known as *Non-uniqueness of graph structures*, and it states:

Given any joint distribution $P(X, Y)$ of two real-valued variables, there exists a SCM $Y = f_y(X)+N_Y$, where f_Y is a measurable function, N_Y is a real-valued noise variable and $X \perp\!\!\!\perp N_Y$.

The result holds both when $X = C, Y = E$ and $X = E, Y = C$, where C denotes the cause and E denotes the effect. For this reason, the causal direction between two observed variables cannot be inferred from passive observations without any further assumptions.

F.3 Structure identifiability under non-Gaussian acyclic models

The distinction between $X \to Y$ and $Y \to X$ becomes feasible from observational data alone under the assumption of linear non-Gaussian acyclic models. In other words, given a cause C and an effect E, the underlying SCM is assumed to be

$$E = \alpha C + N_E$$

where $N_E \perp\!\!\!\perp C$ and N_E is not a gaussian noise. This assumption is sufficient for structure identification because of the following theorem:

Identifiability of linear non-Gaussian models Assuming that $P(X, Y)$ admits the linear model

$$Y = \alpha X + N_Y, \quad X \perp\!\!\!\perp N_Y$$

where X, N_Y, Y are continuous random variables. Then, if and only if N_Y and X are gaussian, there exists a $\beta \in R$ and a random variable N_X such that

$$X = \beta Y + N_X, \quad N_X \perp\!\!\!\perp Y$$

Hence, it is sufficient that C or N_E are non-Gaussian to render the causal direction identifiable.

F.4 Independence of residuals test

Here, we illustrate a test, based on Additive Noise Models(ANMs), to identify the causal direction from a given joint distribution $P(X, Y)$. This test is called *Independence of residuals test.* The steps to follow in order to identify the causal direction from a joint distribution $P(X, Y)$ are the following:

- Regress Y on X
- Test whether the residuals are independent of X
- Regress X on Y
- Test whether the residuals are independent of Y
- If the independence is accepted for one direction and rejected for the other, then the former is the right causal direction

G Exercise 1 figures

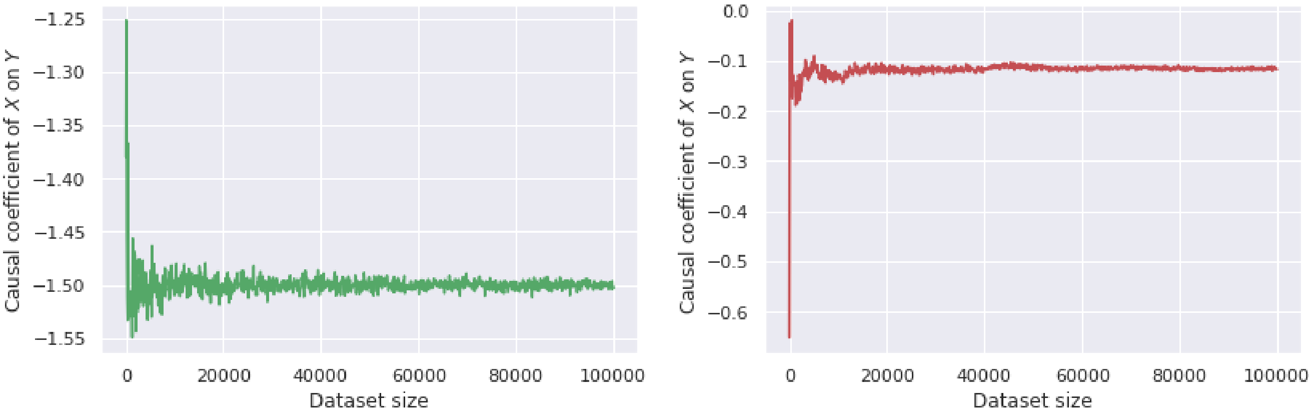

Figure 3: (Left) The plot shows that as the dataset size gets larger and larger, the total causal coefficient of X on Y estimated through the backdoor adjustment approaches -1.5, which is the actual total causal coefficient of X on Y. (Right) The plot illustrates that as the dataset size becomes increasingly larger, the total causal coefficient of X on Y estimated by regressing Y solely on X approaches roughly -0.1, which is not the actual total causal coefficient of X on Y.

H Exercise 2 comments on additional interaction term

The main change stemming from the inclusion of the interaction term $w_{xz}XZ$ in Y's function is the alteration of Y's interaction with its parents. Indeed, while Y is affected in a linear fashion when the interaction term is absent, when the latter is present, both X and Z cause Y non-linearly. As a consequence, the total or direct causal coefficient of either Z or X on Y cannot be computed by merely looking at the path coefficients as these can only be used for linear SCMs. Alternative techniques, like the one used in the exercise, need to be used to figure out the causal effect of X or Z on Y.

I Exercise 2 figures

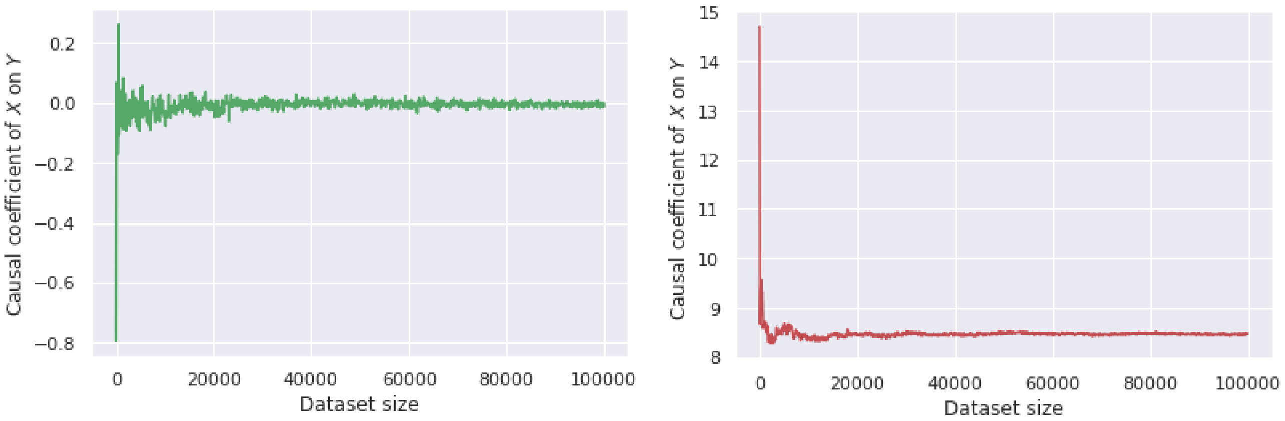

Figure 4: (Left) The plot shows that as the dataset size gets larger and larger, the total causal coefficient of X on Y estimated through the backdoor adjustment approaches 0.0, which is the actual total causal coefficient of X on Y. (Right) The plot illustrates that as the dataset size becomes increasingly larger, the total causal coefficient of X on Y estimated by regressing Y solely on X approaches roughly 8.5, which is not the actual total causal coefficient of X on Y.

J Exercise 3 figures

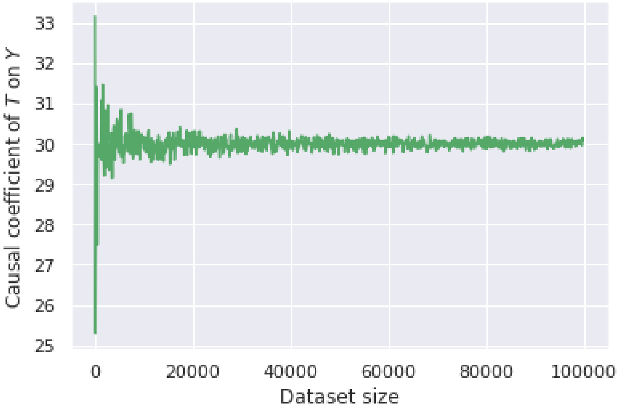

Figure 5: The plot shows that as the dataset size gets larger and larger, the total causal coefficient of T on Y estimated through the front-door adjustment approaches 30.0, which is the actual total causal coefficient of T on Y.

K Exercise 4 experiments intuitive explanations

Backdoor adjustment outperforms front-door adjustment when the noise level in Y is incremented. The reason for this is explained in what follows. In front-door adjustment, the total causal coefficient of T on Y is estimated by multiplying the estimates of the total causal coefficients of M on Y and T on M. Even though the latter of these estimates is not affected by the added noise, the former is altered. Estimating the total causal coefficient of M on Y involves adjusting for T; thus, the regression coefficients of T and M need to compensate for the increased noise in Y and for the absence of the variable Z in the regression equation. On the other hand, the backdoor adjustment involves adjusting for Z, which is a direct cause of Y. Therefore, the regression coefficients need only to compensate for the increase in the noise in Y. To conclude, while in the backdoor adjustment, the regression coefficients need only to compensate for the increase in the noise in Y, the regression coefficients in the front-door adjustment need to compensate for both the increase in the noise in Y and the absence of a direct cause in the regression equation, Z. This is why backdoor outperforms front-door.

Front-door adjustment outperforms backdoor adjustment when the standard deviation of U_M is increased due to the following reason. In front-door adjustment, the total causal coefficient of T on Y is estimated by multiplying the estimates of the total causal coefficients of M on Y and T on M. Only the latter of these estimations gets affected by the increased noise in M. In contrast, when applying the backdoor adjustment, we regress Y on T and Z. As a consequence, the estimation of the regression coefficient of T depends on the path $T \rightarrow M \rightarrow Y$, which is highly affected by the increased variance in the noise in M. Thus, front-door outperforms backdoor because while the former gives a good estimate for the path $M \rightarrow Y$ and a bad estimate for the path $T \rightarrow M$, the latter produces a poor estimate for the whole path $T \rightarrow M \rightarrow Y$.

The reason why increasing the noise in the variable Z only slightly affects the estimates of the backdoor and front-door adjustment relies on the nature of Z. Indeed, Z is a root node, which means that Z only produces causal effects, but it is not the effect of any other variable. Consequently, the causal generation process is not affected by an increase in noise in Z, leading to only slight changes in the total causal coefficient of T on Y recovered through the regression processes.

Increasing the noise in T produces only slight changes in the backdoor and front-door adjustment estimates for the total causal coefficient of T on Y. The reason for this is that although the causal path $Z \rightarrow T$ is affected by the increased noise, the latter is not a path utilised to estimate the total causal effect of T on Y.

L Exercise 4 figures

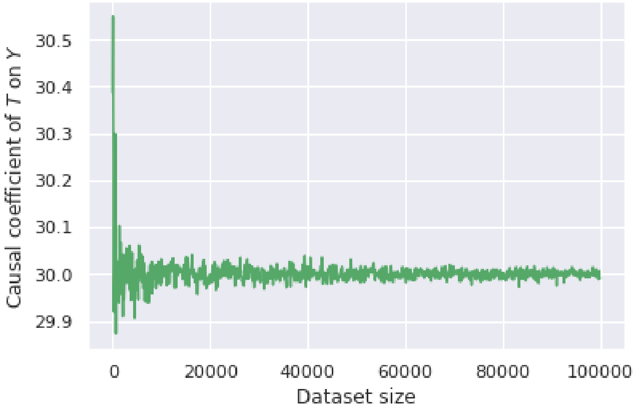

Figure 6: The plot shows that as the dataset size gets larger and larger, the total causal coefficient of T on Y estimated through the backdoor adjustment approaches 30.0, which is the actual total causal coefficient of T on Y.

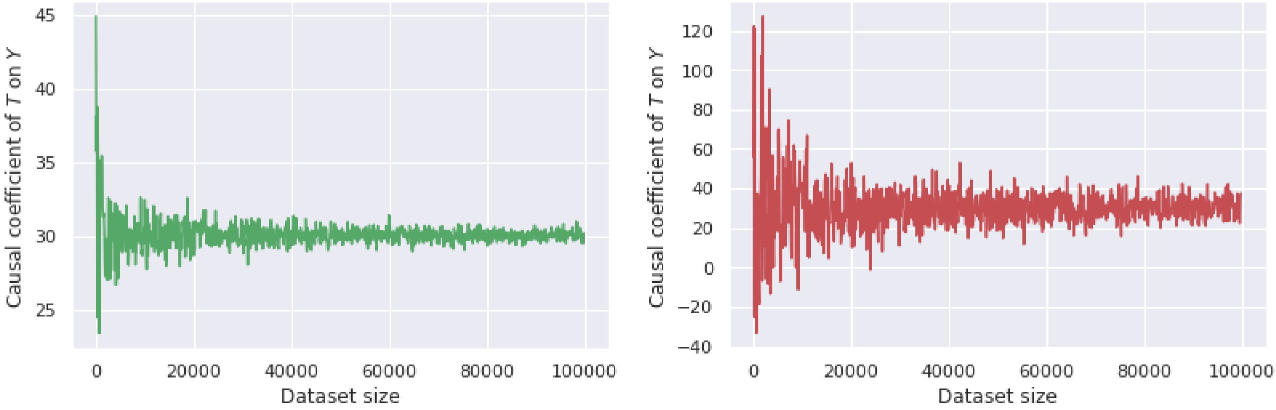

Figure 7: The plots show the estimation of the total causal coefficient of T on Y, when the standard deviation of the exogenous variable U_Y is increased to 100.0, versus the dataset size. While the plot on the left uses the backdoor adjustment, the plot on the right makes use of the front-door adjustment. Since the actual total causal coefficient of T on Y is 30.0, it is noticeable that front-door adjustment estimates are poorer than backdoor ones, as the dataset size increases, because they have wider oscillations

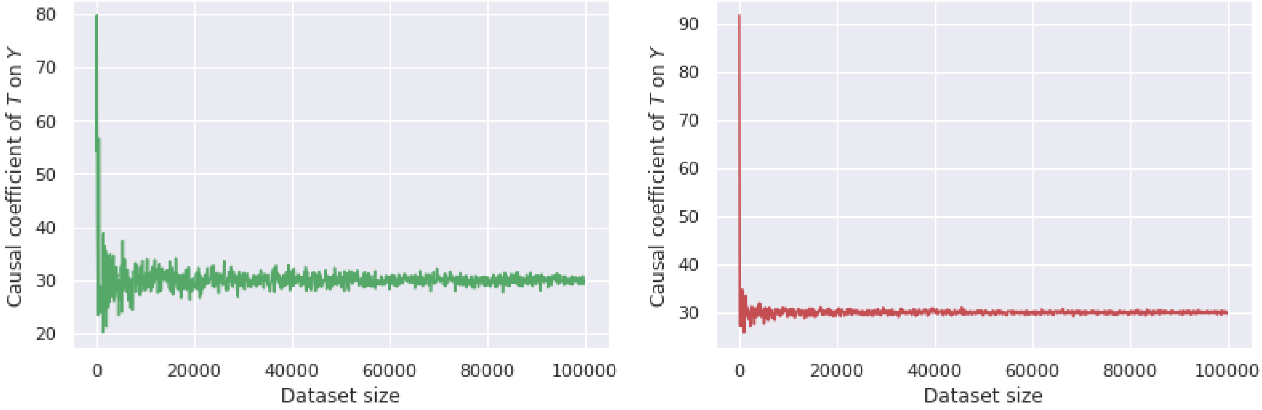

Figure 8: The plots show the estimation of the total causal coefficient of T on Y, when the standard deviation of the exogenous variable U_M is increased to 100.0, versus the dataset size. While the plot on the left uses the backdoor adjustment, the plot on the right makes use of the front-door adjustment. Since the actual total causal coefficient of T on Y is 30.0, it is noticeable that backdoor adjustment estimates are poorer than front-door ones, as the dataset size increases, because they have wider oscillations.

M Exercise 5 Independence of residuals tests

M.1 Independence of residuals test for first linear function

Here, we employ the *Independence of residuals test* to give an alternative demonstration to the fact that the causal mechanism underlying D_1 cannot be identified from the observed joint distribution.

The first two steps of this test consist of regressing Y_1 on X_1 and checking whether the resulting residuals are independent of X_1. Here, we do not use sophisticated high-order statistical tools to verify the independence relation, but instead, we limit ourselves to drawing a scatter plot for this aim. Such a scatter plot is represented in Figure 9. The latter clearly shows that the resulting residuals are independent of X_1. This holds because both X_1 and the resulting residuals follow a uniform distribution (Figure 10), and their joint scatter plot has the classic shape of a bivariate Gaussian consisting of two independent Gaussian random variables.

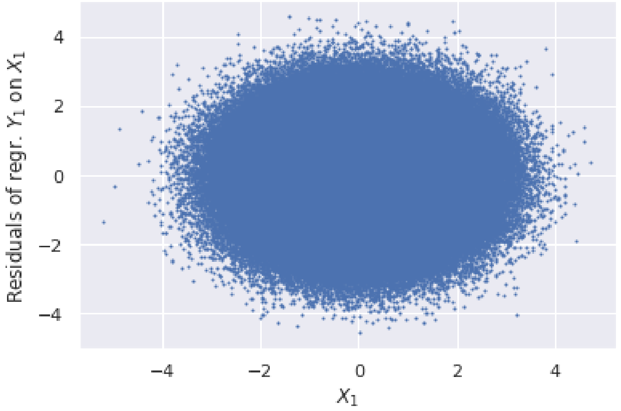

Figure 9: Scatter plot of X_1 versus the residuals resulting from regressing Y_1 on X_1

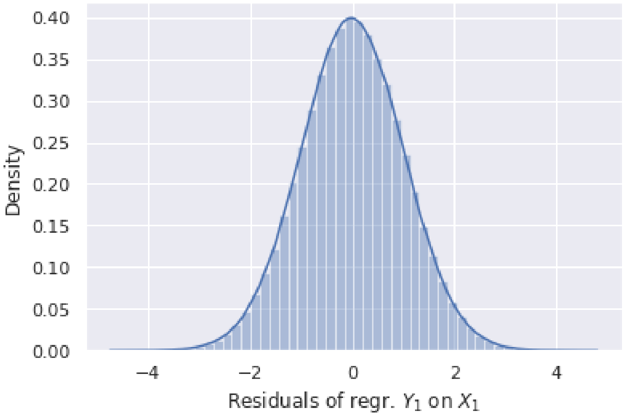

Figure 10: Density plot of the residuals resulting from regressing Y_1 on X_1

The last two steps of this test consist of regressing X_1 on Y_1 and checking whether the resulting residuals are independent of Y_1. Their joint scatter plot is provided in Figure 11. The latter clearly shows that the resulting residuals are independent of Y_1. This holds because both Y_1 and the resulting residuals follow a uniform distribution (Figure 12), and their joint scatter plot has the classic shape of a bivariate Gaussian consisting of two independent Gaussian random variables.

Given that the residuals resulting from regressing Y_1 on X_1 are independent of X_1 and the residuals resulting from regressing X_1 on Y_1 are independent of Y_1, then the underlying causal direction cannot be identified.

M.2 Independence of residuals test for second linear function

Here, we employ the *Independence of residuals test* to give an alternative demonstration to the fact that the causal mechanism underlying D_2 can be identified from the observed joint distribution.

The first two steps of this test consist of regressing Y_2 on X_2 and checking whether the resulting residuals are independent of X_2. Here, we do not use sophisticated high-order statistical tools to verify the independence relation, but instead, we limit ourselves

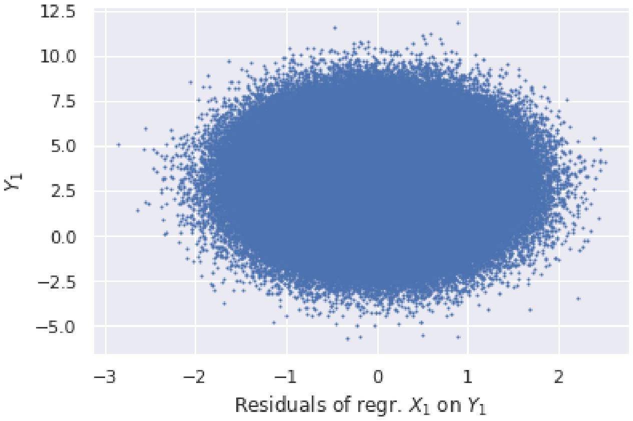

Figure 11: Scatter plot of the residuals resulting from regressing X_1 on Y_1 versus Y_1

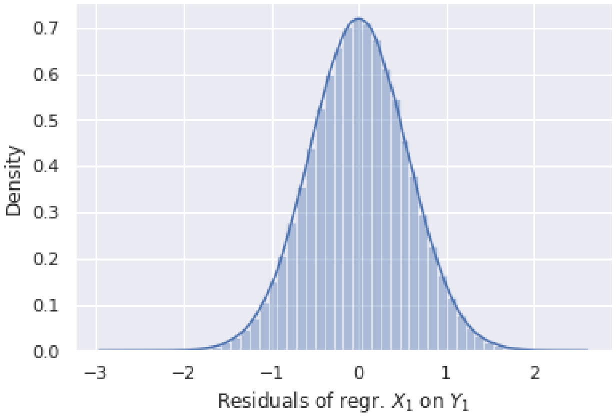

Figure 12: Density plot of the residuals resulting from regressing X_1 on Y_1

to drawing a scatter plot for this aim. Such a scatter plot is represented in Figure 13. The latter clearly shows that the resulting residuals are independent of X_2. This holds because both X_2 and the resulting residuals are uniform distributions (Figure 14), so if their joint plot is rectangle-shaped, then their joint distribution can be factorised into two independent distributions.

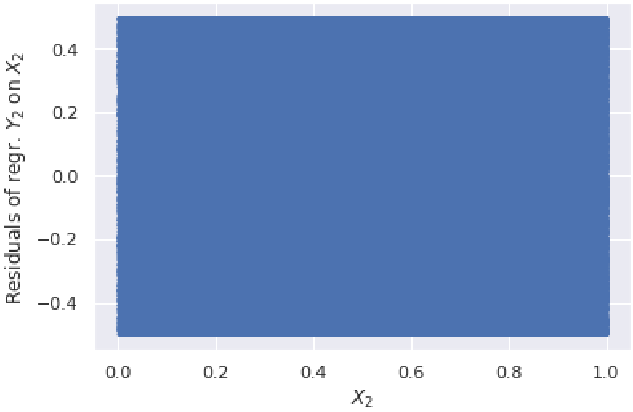

Figure 13: Scatter plot of X_2 versus the residuals resulting from regressing Y_2 on X_2

The last two steps of this test consist of regressing X_2 on Y_2 and checking whether the resulting residuals are independent of Y_2. Their joint scatter plot is provided in Figure 15. The latter clearly shows a pattern of dependence between Y_2 and the resulting residuals. This holds because given that the distribution of the resulting residuals is the one illustrated in Figure 16, then their joint plot cannot be represented by a joint distribution factorised into two independent distributions.

Given that the residuals resulting from regressing Y_2 on X_2 are independent of X_2 and the residuals resulting from regressing X_2 on Y_2 are dependent of Y_2, then the underlying causal direction is $X_2 \to Y_2$.

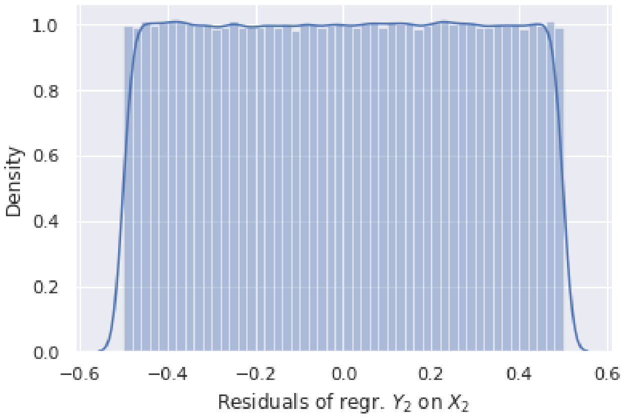

Figure 14: Density plot of the residuals resulting from regressing Y_2 on X_2

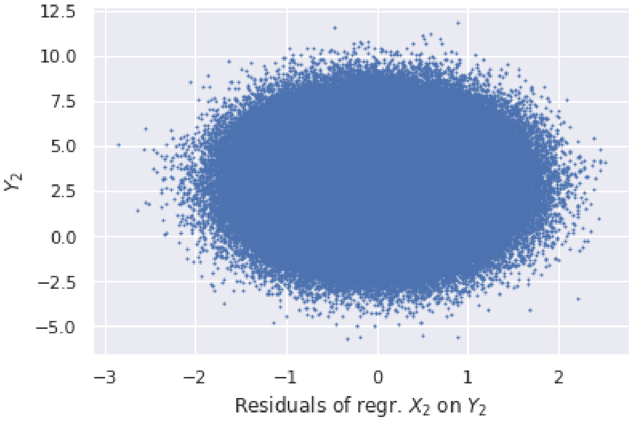

Figure 15: Scatter plot of the residuals resulting from regressing X_2 on Y_2 versus Y_2.

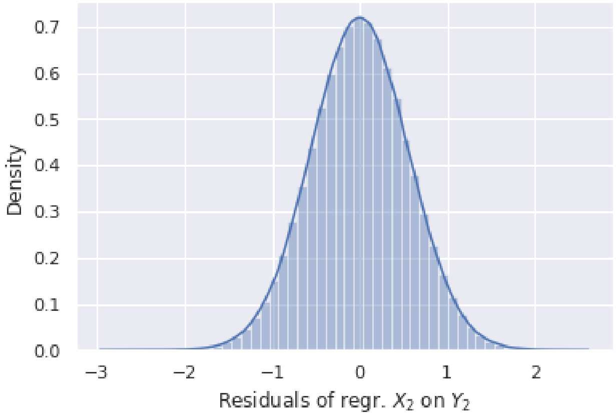

Figure 16: Density plot of the residuals resulting from regressing X_2 on Y_2