TECHNICAL REPORT
R-125
May, 2025

Beyond the back-door: Probabilities of Identification

Drago Plečko
Columbia University

Domagoj Bradač
ETH Zürich

Matija Bucić
Princeton University

Elias Bareinboim
Columbia University

Abstract

Estimation and identification of causal effects are core tasks of causal inference. Within the potential outcomes framework, the two tasks are usually solved jointly, and a range of estimation techniques have been developed, in which identification of causal effects usually rests upon assessing assumptions encoded in algebraic form. On the other hand, within the framework of graphical causal models, questions of identification and estimation are more clearly delineated, and the framework offers a more systematic approach to effect identification. One of the famous examples is the back-door criterion, which allows one to assess the identification of a causal effect, and also choose an appropriate adjustment set for estimation. Back-door type of adjustments, interestingly, lead to the same estimation methods that are used in the PO framework. The back-door criterion, however, is not complete, and thus more general identification approaches were developed in the literature. These approaches allow one to identify various instances of causal effects not covered by back-door adjustment, offering identification of effects not covered in the PO framework. In light of this, proponents of the PO framework argue that instances that go beyond the back-door identification are unlikely to occur in practice, and that general identification strategies of the graphical approach do not constitute a major advantage. In this work, we take a probabilistic approach to analyzing this issue, focusing on the task of non-parametric identification from observational data. In particular, we consider a setting in which the causal graph $\mathcal{G}$ of a Semi-Markovian model is sampled at random, while the treatment X and the outcome Y are singletons drawn uniformly at random. We compute the probability that the effect of X on Y is (back-door) identifiable from observational data and the graph $\mathcal{G}$, for different sampling models. For the case when both the observed and unobserved structure are sampled from an Erdős–Rényi model, we compute the exact probability of back-door identification, together with upper and lower bounds for complete identification, via dynamic programming. We confirm our theoretical findings empirically for this class of models, and extend the discussion to other random graph models, including uniformly sampled and scale-free graphs, and graphs in which the unobserved confounding structure is generated by projecting out (or hiding) variables. Our results show that for the majority of considered models, more than 85% of all identifiable cases can be identified through the back-door criterion.

1. Introduction

Estimation of causal effects is one of the key building blocks of causal inference. Although randomized control trials (RCTs) (Fisher, 1936) are still considered to be the gold standard for estimating causal effects in many applied sciences, a substantial part of the literature on causal inference is concerned with the estimation of causal effects from observational studies. In such settings, however, the estimation of causal effects may not always be possible, even if an unlimited amount of data was available. Therefore, the question of when and under which assumptions causal effects can be

computed from the data has been studied under the rubric of causal effect *identifiability* (Pearl, 2000), representing an important aspect of causal effect estimation as a broader concept.

There are two main, sometimes perceived as competing, frameworks concerned with the inference of causal effects. The first is known as the *potential outcomes* (PO) framework, associated with the works of Donald Rubin and colleagues (Rubin, 2005), building on pioneering ideas of Jerzey Neyman (Neyman, 1923). Within this framework, assumptions are encoded in a more algebraic form, in terms of conditional independence statements across potential outcomes. The questions of identification and estimation are not entirely separate in the framework, and a range of techniques for estimating the causal effects have been developed, such as inverse-propensity weighting (Rosenbaum and Rubin, 1983), matching (Rosenbaum et al., 2010), and doubly-robust methods (Robins et al., 1994, 1995; Bang and Robins, 2005), just to name a few (see (Imbens, 2004) for a review). Often these estimation techniques rest upon assessing some form of unconfoundedness, usually encoded in ignorability assumptions (Rubin, 1978; Rosenbaum and Rubin, 1983).

The second, alternative framework, rooted in early ideas from the fields of genetics (Wright, 1921b) and econometrics (Wright, 1921a, 1934; Haavelmo, 1944), and nowadays popular in the computer science literature, is the structural-graphical approach to causality, pioneered by Judea Pearl (Pearl, 2000). Unlike the PO framework, this framework employs an explicit graphical representation of the underlying causal model, in which variables are represented as nodes in a directed acyclic graph (DAG). When it comes to causal effect identification, instead of assessing ignorability assumptions, the graphical approach introduces general machinery for analyzing causal queries, including the well-known *do-calculus* (Pearl, 1995). One of the famous identification methods in this context is the *back-door* criterion, in which a set of variables Z causally preceding both the treatment variable X and the target variable Y are selected for adjustment, to block spurious correlations between X and Y. This approach offers a human-understandable way of selecting the adjustment set and can be used even in cases where some of the variables Z are not observed. Interestingly, when performing an estimation after applying the backdoor criterion, the graphical approach builds on the estimation techniques from the PO literature, such as matching, weighting, or doubly robust approaches (Imbens, 2004), while providing a systematic way to choose the exact adjustment set. However, the identification machinery in the graphical approach does not end with the back-door criterion. Examples of DAGs in which a causal effect may be identified, but no back-door adjustment set exists, have been brought forward. The first discovered example of this is the so-called *front-door* adjustment setting (Fig. 1). In this case, the back-door type of adjustment for the effect of X on Y fails, since the confounders between X and Y are unobserved (indicated by the dashed, bidirected arrow), but the effect is still identifiable due to the existence of a mediator Z on a pathway $X \to Y$, which is not affected by the unobserved confounders. As subsequent works demonstrated, the graph in Fig. 1 is not just an idiosyncrasy, but a whole range of identifiable examples exists in which back-door adjustment fails, but the target effect can be identified. This led to the development of a general algorithmic approach to identification that include *complete* algorithms (Tian and Pearl, 2002; Shpitser and Pearl, 2006; Huang and Valtorta, 2012; Bareinboim and Pearl, 2012b; Lee et al., 2020b), which recover identification expressions whenever this is possible. Building on these results, the do-calculus was shown to be complete, meaning that its rules can be used for identifying an effect, whenever this is possible. These developments have closed the open questions of non-parametric identification of interventional distributions from combinations of experimental and observational data within the graphical framework for causality.

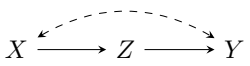


Figure 1: Graphical model of the front-door example.

When considering the identifiable functionals beyond the back-door criterion, proponents of the PO framework argue that graphical models requiring more sophisticated adjustment methods than the back-door are unlikely to occur in practice. For example, (Cox et al., 1995, p.689) argue that

the diagram in Fig. 1 is unrealistic, due to the requirement that the unobserved confounder U which affects X, Y does not affect the mediator Z. More recently, Imbens (2020) also discussed the front-door setting, raising several concerns about its applicability for identifying the effect of smoking on lung cancer[1]. At the same time, however, (Pearl and Mackenzie, 2018, p.231) argues that the front-door approach may eventually become a serious competitor to randomized controlled trials, as it allows one to control for confounders which are unobserved or even entirely unknown. Currently, however, there are few applied studies in the literature that use identification techniques beyond the back-door, with some interesting exceptions such as (Glynn and Kashin, 2018; Piccininni et al., 2023; Inoue et al., 2022).

Interestingly, to the best of our knowledge, few formal mathematical arguments have been given in the above debate, and it seems that the applicability of identification methods beyond the back-door adjustment is a matter of opinion. In this manuscript, we add to this discussion by taking a probabilistic approach to this question. In particular, by putting a probability measure over the space of all possible Semi-Markovian causal graphs, we ask the following types of questions: given a randomly chosen graph $\mathcal{G}$ and two randomly chosen variables X, Y, what is the probability that the causal effect of X on Y is identifiable (i) using back-door adjustment; (ii) using an instrumental variable approach; (iii) using the complete identification algorithm? Such questions motivate the concept of *probabilities of identification*. For instance, we are interested in, writing informally for now, P(BD ID), P(IV ID), and P(ID), and how large the gaps are between them. These quantities, for different random models over the space of causal graphs, will allow us to answer how much there is to be gained from more fine-grained identification approaches compared to only back-door identification, from a probabilistic standpoint. In particular, our contributions are the following:

(i) We introduce the formal notions of probabilities of identification (Def. 12),

(ii) For the case when the observed structure of the graph is drawn from an Erdős-Rényi model $G(n, p)$, and the latent structure of bidirected edges is drawn from an Erdős-Rényi model $G(n, q)$, using dynamic programming we compute:

  (a) the exact probability of identification for the back-door algorithm (Thm. 2, Lems. 1, 2),

  (b) the upper and lower bounds for the probability of complete identification (Thms. 4, 6, Lems. 4, 5),

(iii) We confirm our findings empirically through simulation, and extend the discussion to other random graph models, including uniform DAGs, scale-free DAGs (Bollobás et al., 2003), and graphs in which the unobserved structure is generated by projecting out variables with a fixed probability.

A probabilistic approach to questions in causal inference is not entirely new. For example, Uhler et al. (2013) consider randomly sampled linear structural causal models to answer how likely it is that a faithfulness assumption is satisfied, showing that faithfulness is violated on a measure-zero set, while strong-faithfulness is violated on a set with a positive measure. Furthermore, several works investigate the sizes and properties of (interventional) Markov equivalence classes (Katz et al., 2019) using a random-graphs perspective. Another work, related in spirit, investigates causal discovery in graphs where edges exist probabilistically, and attempts to recover the most probable graph for which a given causal query is identifiable (Akbari et al., 2022). However, to the best of our knowledge, a random graphs approach has not yet been employed for the questions we investigate in this manuscript.

Finally, after clarifying what is discussed in the manuscript, we mention some aspects that go beyond its content. Firstly, as noted earlier, assessing ignorability assumptions does not provide a

1. This discussion is a continuation of previous criticism of the smoking and lung cancer example, found in (Koller and Friedman, 2009).

systematic way of finding valid adjustment sets, as opposed to the structural/graphical approach and the back-door criterion, which allow one to choose adjustment sets that possibly result in improved sample efficiency. However, more fundamentally, having a specified graphical model is required for the latter approach, which may be non-trivial to obtain, especially in cases with a large number of variables. Secondly, we focus on the setting with X, Y variables being singletons, while these variables could in general be sets. Thirdly, the issues explored in this manuscript relate to identifying interventional distributions from observational data, and do not cover other interesting settings where the gap between the two frameworks of causality may exist. Such settings include the identification of interventional or counterfactual distributions from arbitrary combinations of observational and interventional data (Shpitser and Pearl, 2007; Lee et al., 2020a; Correa et al., 2021; Correa and Bareinboim, 2025).

1.1 Organization of the Manuscript

The rest of the manuscript is organized as follows. Sec. 2 introduces important notions from the causal inference literature related to identifiability. In Sec. 3, the concept of probability of identification is formally defined. Then, we compute the exact value for the probability of back-door identification $P(\text{BD ID})$, and upper and lower bounds for probabilities of identification $P(\text{ID})$, using chained sequences of dynamic programs, for the model in which directed edges are drawn independently with a probability p (i.e., from a directed $\overrightarrow{G}(n, p)$ model), and bidirected edges are drawn independently with a probability q (i.e., from a $G(n, q)$ model). These theoretical results are also verified through simulation. In Sec. 4, we extend our simulation results to other generating models, including scale-free and uniformly-sampled graphs, and also cover different ways of generating the confounding structure, by projecting out or "hiding" observed variables from the model. In Sec. 4.1, some of the methods usually associated with the potential outcomes framework (such as instrumental variables and conditional ignorability) are discussed, and we analyze their probabilities of identification. In Sec. 5 we give an interpretation of the results and some concluding remarks.

2. Prior Art

The first basic concept of causal inference we introduce is the structural causal model (SCM):

Definition 1 (Structural Causal Model (SCM) (Pearl, 2000)) *A structural causal model $\mathcal{M}$ is a 4-tuple $\langle V, U, \mathcal{F}, P(u)\rangle$, where*

1. *U is a set of exogenous variables, also called background variables, that are determined by factors outside the model;*
2. *$V = \{V_1, ..., V_n\}$ is a set of endogenous (observed) variables, that are determined by variables in the model (i.e. by the variables in $U \cup V$);*
3. *$\mathcal{F} = \{f_1, ..., f_n\}$ is the set of structural functions determining V, $v_i \leftarrow f_i(\text{pa}(v_i), u_i)$, where $\text{pa}(V_i) \subseteq V \setminus V_i$ and $U_i \subseteq U$ are the functional arguments of f_i;*
4. *$P(u)$ is a distribution over the exogenous variables U.*

□

The SCM contains the "ground truth" about the underlying causal phenomenon. In particular, the assignment mechanisms $\mathcal{F}$ determine how each of the observed variables V_i attains its value, based on other observed variables and the latent variables U. Together with the probability distribution $P(u)$ over the exogenous variables U, it specifies the entire behavior of the underlying phenomenon. In particular, the SCM also specifies the *observational distribution* of the underlying phenomenon, defined through:

Definition 2 (Observational Distribution (Bareinboim et al., 2022)) *An SCM $\mathcal{M}$ that is a 4-tuple $\langle V, U, \mathcal{F}, P(u)\rangle$ induces a joint probability distribution $P(V)$ such that for each $Y \subseteq V$,*

$$P^{\mathcal{M}}(y) = \sum_{u} \mathbb{1}\Big(Y(u) = y\Big) P(u), \tag{1}$$

where $Y(u)$ is the solution for Y after evaluating $\mathcal{F}$ with $U = u$. □

A further important notion building on the concept of the SCM is that of a submodel, which is defined next:

Definition 3 (Submodel (Pearl, 2000)) *Let $\mathcal{M}$ be a structural causal model, X a set of variables in V, and x a particular value of X. A submodel $\mathcal{M}_x$ (of $\mathcal{M}$) is a 4-tuple:*

$$\mathcal{M}_x = \langle V, U, \mathcal{F}_x, P(u)\rangle \tag{2}$$

where

$$\mathcal{F}_x = \{f_i : V_i \notin X\} \cup \{X \leftarrow x\}, \tag{3}$$

and all other components are preserved from $\mathcal{M}$. □

In words, the SCM $\mathcal{M}_x$ is obtained from $\mathcal{M}$ by replacing all equations in $\mathcal{F}$ related to variables X by equations that set X to a specific value x. This corresponds to setting the value of $X = x$ in the model, and is also abbreviated with $\text{do}(X = x)$. Related to the concept of a submodel, we introduce next the important concept of an interventional distribution:

Definition 4 (Interventional Distribution) *Let $X, Y \subseteq V$ be disjoint sets of variables in an SCM $\mathcal{M}$. Then, the interventional distribution $P(Y \mid do(X = x))$ denotes the distribution of Y in the submodel $\mathcal{M}_x$.* □

Finally, there is one more prerequisite notion for our discussion. The mechanisms $\mathcal{F}$ and the distribution over the exogenous variables $P(u)$ are almost never observed. However, to perform causal inference, we need a way of encoding assumptions about the underlying SCM. A common way of doing so is through an object called a causal diagram, which is defined next:

Definition 5 (Causal Diagram (Pearl, 2000; Bareinboim et al., 2022)) *Let an SCM $\mathcal{M}$ be a 4-tuple $\langle V, U, \mathcal{F}, P(u)\rangle$. A graph $\mathcal{G}$ is said to be a causal diagram (of $\mathcal{M}$) if:*

(1) there is a vertex for every endogenous variable $V_i \in V$,

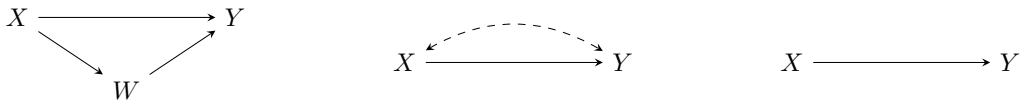
(2) there is an edge $V_i \to V_j$ if V_i appears as an argument of $f_j \in \mathcal{F}$,

(3) there is a bidirected edge $V_i \dashleftarrow\dashrightarrow V_j$ if the corresponding $U_i, U_j \subset U$ are correlated or the corresponding functions f_i, f_j share some $U_{ij} \in U$ as an argument.

□

Note that while the SCM contains explicit information about all structural mechanisms ($\mathcal{F}$) and distribution over the exogenous variables ($P(u)$), the causal diagram requires much weaker information. The directed edges encode information about which functional arguments are possibly used as inputs to the functions in $\mathcal{F}$, while the bidirected edges encode which exogenous variables share information, i.e., where unobserved confounding is present. Based on the causal diagram, following the usual notation, we will write $\text{pa}(V_i), \text{ch}(V_i), \text{an}(V_i)$, and $\text{de}(V_i)$ for the parents, children, ancestors, and descendants of V_i, respectively. [2]

2. We note for readers less familiar with causal inference that bidirected edges are not considered when judging relationships of ancestrality and descendancy.



(a) Diagram of Ex. 1. (b) Bow graph. (c) No-bow graph.

Figure 2: Causal diagrams in Sec. 2.1.

Example 1 (Causal Inference Blocks) *Consider the following structural causal model (SCM):*

$$\mathcal{F}, P(u): \begin{cases} X \leftarrow U_X & (4) \\ W \leftarrow -X + U_W & (5) \\ Y \leftarrow X + W + U_Y. & (6) \\ \\ U_X, U_W, U_Y \sim N(0,1). & (7) \end{cases}$$

The causal diagram associated with this SCM is shown in Fig. 2a. The edge $X \to W$ exists because X is a functional argument of the f_w mechanism in Eq. 5. Similarly, the $X \to Y, W \to Y$ edges exist since X, W are arguments of f_y in Eq. 6. The observational distribution is given by

$$(X, W, Y) \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \right). \tag{8}$$

The interventional distribution can be obtained from the submodel $\mathcal{M}_x$, which is computed by replacing Eq. 4 with $X \leftarrow x$. Thus, we can compute the interventional distribution as

$$(X, W, Y) \mid do(X = x) \sim N\left(\begin{pmatrix} x \\ -x \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \right). \tag{9}$$

□

2.1 Identification

After introducing the basic building blocks of causal inference, We now move onto the task of *identification*. As was noted earlier in the text, a very common quantity of interest in causal inference is known as the average treatment effect, which, for a single binary variable $X \in x_0, x_1$ is written as:

$$\text{ATE}_{x_0,x_1}(y) = \mathbb{E}[Y \mid do(X = x_1)] - \mathbb{E}[Y \mid do(X = x_0)]. \tag{10}$$

In non-parametric estimation from observational data, our ability to compute quantities like the ATE depends on recovering the interventional distribution $P(Y \mid do(X = x))$. Therefore, from this point onwards, we focus on the task of recovering distributions of the form $P(Y \mid do(X = x))$. To this end, we first define what it means for $P(Y \mid do(X = x))$ to be *identifiable*:

Definition 6 (Identification (Pearl, 2000)) *An interventional distribution $P(Y \mid do(X = x))$ is said to be identifiable if it can be uniquely computed from the observational distribution $P(V)$ and*

the causal diagram $\mathcal{G}$. Formally, $P(Y \mid do(X = x))$ is identifiable if, for all $\mathcal{M}_1, \mathcal{M}_2$ compatible with $P(v)$ and $\mathcal{G}$, we have

$$P^{\mathcal{M}_1}(Y \mid do(X = x)) = P^{\mathcal{M}_2}(Y \mid do(X = x)). \tag{11}$$

□

We now ground the notion of identifiability through an example.

Example 2 (Lack of Identification) *Consider the structural causal model (SCM) parameterized by α, ρ, given as follows:*

$$\mathcal{F}, P(u): \begin{cases} X \leftarrow U_X & (12) \\ Y \leftarrow \alpha X + U_Y. & (13) \\ \begin{pmatrix} U_X \\ U_Y \end{pmatrix} \sim N\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \Big). & (14) \end{cases}$$

The parameter α measures the strength of the causal effect, whereas the ρ parameter measures the correlation of the noise terms U_X, U_Y (which in this case one would call hidden confounding). The SCM is compatible with the causal diagram shown in Fig. 2b, known as the bow graph. The bidirected arrow between X and Y indicates that the noise variables may be correlated (which is the case whenever $\rho \neq 0$).

We can further compute that the observational distribution is given by

$$(X, Y) \sim N\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \alpha + \rho \\ \alpha + \rho & 1 + 2\alpha\rho + \alpha^2 \end{pmatrix} \Big), \tag{15}$$

whereas the interventional distribution after setting $X = x$ is given by

$$(X, Y) \mid do(X = x) \sim N\Big(\begin{pmatrix} x \\ \alpha x \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \Big). \tag{16}$$

We are now interested in inferring the distribution in Eq. 16, but we only have access to the observational distribution in Eq. 15. Def. 6 is concerned with the question if this computation can be done uniquely. Suppose that we observed that

$$(X, Y) \sim N\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} \Big). \tag{17}$$

Then, one can find that

$$(\alpha, \rho) = \big(0, \frac{1}{2}\big), \big(\alpha, \rho\big) = \big(1, -\frac{1}{2}\big) \tag{18}$$

both generate the observational distribution in Eq.17, but the corresponding interventional distributions $Y \mid do(X = x)$ are equal to,

$$N(0, 1) \text{ and } N(x, 1), \tag{19}$$

respectively, and thus differ for any $x \neq 0$. Therefore, we say that $P(Y \mid do(X = x))$ is not identifiable. □

In the above example, we have chosen a specific parametric family (a linear model with Gaussian noise) to illustrate the issue of identifiability in a familiar setting, and to leverage the reader's intuition on the non-identifiability of *statistical parameters*. As the example also showed, the bow graph in Fig. 2b may produce non-identifiable instances of interventional distributions.

In general, Def. 6 is concerned with a more involved task of *non-parametric* identification, in which we do not assume a parametric family for the SCM (like we did in Eqs. 12·14) and we do not attempt to infer the true parameters of the SCM. Instead, we are interested in deriving general, non-parametric identification expressions that connect the interventional and the observational distributions by leveraging the knowledge in the causal diagram, as in the following basic example:

Example 3 (Identification) *Consider any SCM compatible with the causal graph in Fig. 2c, and let $P(x, y)$ denote its observational distribution. For any such SCM, the interventional distribution $P(y \mid do(x))$ can be uniquely computed as*

$$P(y \mid do(x)) = P(y \mid x), \tag{20}$$

that is, intervening to set $X = x$ is equal to conditioning on $X = x$. Going back to Ex. 2, the additional information offered by the graph in Fig. 2c compared to Fig. 2b is that the former precludes the possibility of the correlation parameter $\rho \neq 0$. □

As the above example shows, the simple graph in Fig. 2c can never cause issues with respect to identifying interventions. In the following sections, we start introducing general graphical conditions which ensure that an interventional distribution can be uniquely computed from the graph and the observational distribution.

2.2 Back-door Identification

A well-known and widely used strategy for determining whether an interventional distribution is identifiable is the *back-door* criterion. Before we explain the back-door criterion, we introduce the basic idea of d-separation, which allows one test whether a path carries any information over it:

Definition 7 (d-separation) *A path p is said to be blocked (or d-separated) by a set of nodes Z if and only if*

(i) there is a chain along p of the form $i \to m \to j$ or $i \leftarrow m \to j$ such that $m \in Z$,

(ii) there is a chain along p of the form $i \to m \leftarrow j$ with a collider m such that $\{m, \mathrm{de}(m)\} \cap Z = \emptyset$, where $\mathrm{de}(m)$ is the set of descendants of m.

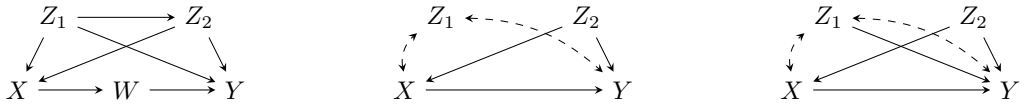
□

There are two ways a path can be blocked: (i) there is a non-collider m which is included in Z; (ii) there is a collider m which is not in Z, and neither are any of its descendants. Intuitively, for paths of the form $i \to m \to j$ or $i \leftarrow m \to j$ conditioning on m will block the flow of information between i, j and make them independent. However, conditioning on a collider m (or any of its descendants) in a structure $i \to m \leftarrow j$ will in fact entangle i, j and make them dependent, whereas if m is ignored, they would be independent. Based on this notion, we can now state the famous back-door criterion:

Definition 8 (Back-door Criterion (Pearl, 2000)) *A set of variables Z satisfies the back-door criterion relative to an ordered pair of variables (X, Y) in a causal diagram $\mathcal{G}$ if:*

(i) no node in Z is a descendant of X,

(ii) Z blocks every path between X and Y that contains an arrow into X.



(a) Graph with a back-door. (b) Graph with M-bias. (c) Graph with no back-door.

Figure 3: Causal graphs from Ex. 4

□

Based on the definition of the back-door criterion, the following important result can be obtained:

Proposition 1 (Back-door Adjustment (Pearl, 2000)) *If a set of variables Z satisfies the back-door criterion relative to (X, Y), then the distribution $P(Y \mid do(X = x))$ is identifiable from observational data.* □

Example 4 (Back-door ID) *Consider the diagram in Fig. 3a. Then, the set $\{Z_1, Z_2, W\}$ does not satisfy the back-door criterion, since it includes $W \in \mathrm{de}(X)$. Furthermore, the set $\{Z_1\}$ also does not satisfy the back-door criterion, because the path $X \leftarrow Z_2 \rightarrow Y$ remains open. Finally, one can check that the set $\{Z_1, Z_2\}$ satisfies the back-door criterion.*

Based on the above, one may be tempted to think that including all variables which causally precede X is a good idea. Consider now the graph in Fig. 3b. The set $\{Z_1, Z_2\}$ is not a back-door set, since Z_1 is a collider on the path $X \dashleftarrow\dashrightarrow Z_1 \dashleftarrow\dashrightarrow Y$ and we are conditioning on it. However, the set $\{Z_2\}$ is a back-door set.

Of course, graphs with no back-door sets also exist. Consider the graph in Fig. 3c, in which neither of $\emptyset, \{Z_1\}, \{Z_2\}, \{Z_1, Z_2\}$ satisfy the back-door criterion. In this case, the effect of X on Y is not identifiable via the back-door. □

In general graphs, there may be a large number of sets Z that could possibly satisfy the back-door criterion for identifying the causal effect of X on Y, seemingly creating a problem for verifying whether a valid back-door set exists. However, the following result helps us avoid searching over a large number of possible adjustment sets:

Proposition 2 (One-shot Back-door Identification) *Let (X, Y) be an ordered pair of variables in a causal graph $\mathcal{G}$, and let F denote the set of variables that lie on any directed path $X \rightarrow \cdots \rightarrow Y$. If a back-door set for (X, Y) exists, then the set $\mathrm{an}(X \cup Y) \setminus F$ is also a back-door set for (X, Y).* □

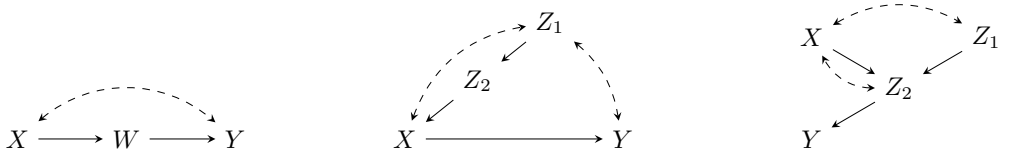
The above proposition can be seen as very useful from a computational point of view. Instead of having to check every possible back-door set, we can simply verify whether $\mathrm{an}(X \cup Y) \setminus F$ is a back-door set.

We remark here that back-door types of adjustment are similar to identification performed based on the ignorability assumption in the potential outcomes literature. However, judging the ignorability assumption with a causal graph may be difficult in practice (as the example in Fig. 3b illustrates). Nonetheless, for the purposes of our discussion, we assume that the back-door identification strategy is in common to both the graphical approach and the potential outcomes approach.

We come back to the back-door identification in Sec. 3, but for now, move to more general identification strategies, which go strictly beyond the identification strategies of the PO literature.

2.3 Identifying $P(V \setminus X \mid do(X = x))$

In this manuscript, we are interested in identification of interventional distributions $P(Y \mid do(X = x))$ where X, Y variables are singletons. However, in the literature, there is a useful strategy for



(a) Front-door graph. (b) Napkin graph. (c) Graph with a hedge.

Figure 4: Causal graphs from Ex. 4

identifying the causal effects of the form $P(V \setminus X \mid do(X = x))$, that is, the interventional distribution of all other variables (and not just Y) when setting $X = x$. We note that if one can recover $P(V \setminus X \mid do(X = x))$, then the distribution $P(Y \mid do(X = x))$ follows easily, since we can simply marginalize out all the variables $V \setminus \{X, Y\}$, i.e.

$$P(y \mid do(x)) = \sum_{v \setminus \{x,y\}} P(v \setminus x \mid do(x)). \tag{21}$$

The first preliminary notion for building such stronger identification strategies is that of a C-component.

Definition 9 (Confounded Component (Tian and Pearl, 2002)) *Let $\mathcal{G}$ be a causal graph of a Semi-Markovian model. A confounded component (C-component) C of $\mathcal{G}$ is any maximal subset of the observables V in which any two variables $V_i, V_j \subseteq C$ are connected through a path consisting of only bidirected edges.* □

It is not difficult to show that the set of C-components forms a partition of the set of observables V. We further ground the notion through an example:

Example 5 (C-component) *Consider the graph in Fig. 3a that has no bidirected edges. Then, the C-components are all the singleton variables. In Fig. 3b the C-components are $\{X, Z_1, Y\}$ and $\{Z_2\}$, which is also the case in Fig. 3c.* □

Based on the notion of a C-component, Tian and Pearl (2002) showed a seminal result on identifiability of the queries of the form $P(V \setminus X \mid do(X = x))$:

Proposition 3 ($P(V \setminus X \mid do(X = x))$ Identification (Tian and Pearl, 2002)) *The interventional distribution $P(V \setminus X \mid do(X = x))$ is identifiable from observational data if and only if there is no bidirected path in $\mathcal{G}$ connecting X to any of its children.* □

We now revisit an example from the beginning of the manuscript:

Example 6 (Front-door Identification) *Consider the front-door graph in Fig. 4a. Due to the bidirected edge $X \dashleftarrow\dashrightarrow Y$, we note that no set Z can be back-door for X, Y. Thus, back-door identification fails in this instance. However, observe that there is no bidirected path from X to W (its only child), and thus by Prop. 3 the distribution $P(W, Y \mid do(X = x))$ is identifiable from observational data.* □

As the above example illustrates, we now have a second strategy for verifying identification of causal effects, which is more powerful than the back-door criterion.

Towards a lower bound. The graphical criterion in Prop. 3 provides another easily verifiable strategy for determining whether causal effects are identifiable. As it turns out, the identification of $P(Y \mid do(X = x))$ is unaffected if we restrict our attention only to the subgraph consisting of nodes $\text{an}(Y)$. Thus, our strategy will be to:

1. Consider the subgraph over the nodes $\text{an}(Y)$, labeled $\mathcal{G}[\text{an}(Y)]$,
2. Check if there exists a bidirected path in $\mathcal{G}[\text{an}(Y)]$ between X and any of its children.

However, as witnessed by the following example, this identification strategy is still not perfect:

Example 7 (Lack of completeness) *Consider the causal graph in Fig. 4b. The bidirected path $X \dashleftarrow\dashrightarrow Z_1 \dashleftarrow\dashrightarrow Y$ connects X to its child Y. Therefore, Prop. 3 cannot guarantee that the distribution $P(Y \mid do(X = x))$ is identifiable. However, as it turns out, there exists an expression connecting the interventional and the observational distributions, given by*

$$P(y \mid do(x)) = \frac{\sum_{z_1} P(x, y \mid z_1, z_2)P(z_1)}{\sum_{z_1} P(x \mid z_1, z_2)P(z_1)}. \tag{22}$$

□

We omit the proof of Eq. 22, as it requires tools of the celebrated do-calculus (Pearl, 1995), which is beyond the scope of our discussion. The main takeaway for the reader is that our second identification strategy which looks for a bidirected path between X and its children, can in fact be improved upon. Thus, our strategy provides a "lower bound" for general identification, or a sufficient but not necessary condition for identification. We will make this notion formal shortly, after discussing the complete identification strategy.

2.4 Complete Identification of $P(Y \mid do(X = x))$

We now move onto describing a general and complete criterion for identification. We first consider definitions of C-trees and C-forests:

Definition 10 (C-tree and C-forests (Shpitser and Pearl, 2006)) *A graph $\mathcal{G}$ is called a C-tree if its nodes form a C-component, each node has at most one child, and all nodes are ancestors of a single root node. A graph $\mathcal{G}$ is called a C-forest if its nodes form a C-component, and each node has at most one child.* □

These structures are further grounded in the following example:

Example 8 (C-trees and C-forests) *Consider the graph in Fig. 3c. The set $\{X, Y, Z_1\}$ forms a C-tree rooted in Y, since it is a C-component and every node has at most one child. It is also a C-forest, since any C-tree is also a C-forest.*

Consider the graph in Fig. 4b. The graph over nodes $\{X, Y, Z_1, Z_2\}$ does not form a Y-rooted C-tree, since Z_2 is not in the same C-component as $\{X, Y, Z_1\}$. However, $\{X, Y, Z_1\}$ form a C-forest, where each of the nodes X, Y, and Z_1 are roots. □

Based on the notion of C-forests, the lack of identifiability of a causal effect will be witnessed by the existence of a structure known as a hedge:

Definition 11 (Hedge (Shpitser and Pearl, 2006)) *Let X, Y be sets of variables in $\mathcal{G}$. Let F_0, F_1 be two C-forests in $\mathcal{G}$ with the set of root nodes R, such that:*

(i) F_0 is a subgraph of F_1 ($F_0 \subsetneq F_1$),

(ii) X and F_0 are disjoint ($X \cap F_0 = \emptyset$),

(iii) X and F_1 are not disjoint ($X \cap F_1 \neq \emptyset$),

(iv) $R \in \text{an}_{\mathcal{G}_{\overline{X}}}(Y)$, where $\mathcal{G}_{\overline{X}}$ denotes the graph where incoming arrows into X are deleted.

Then F_0, F_1 are said to form a hedge for the distribution $P(Y \mid do(X = x))$. □

Now, building on the notion of hedges, we can state the general identification result for interventional distributions, which is known to be complete:

Proposition 4 (Identification Criterion (Shpitser and Pearl, 2006)) *The interventional distribution $P(Y \mid do(X = x))$ is identifiable from observational data $P(v)$ and the graph $\mathcal{G}$, written $P(Y \mid do(X = x)) \vdash_{id} \mathcal{G}$, if and only if no two subgraphs F_0, F_1 exist which form a hedge for $P(Y \mid do(X = x))$.* □

Two things are needed to check if an effect of X on Y is identifiable. First, we need to find a smaller C-forest, rooted in $\text{an}(Y)$, that does not contain X. Then, we need to "grow" this forest further (while retaining the same set of roots), until we have a larger forest that does contain X. If we manage to do so, then the effect is not identifiable. Conversely, if we fail in this exercise, then the effect is identifiable, i.e., the condition is complete.

Example 9 (Hedges and Identifiability) *Consider the causal graph in Fig. 4b. Recall that, since a bidirected path between X its child Y exists ($X \dashleftarrow\dashrightarrow Z_1 \dashleftarrow\dashrightarrow Y$), we cannot guarantee identifiability of $P(Y \mid do(X = x))$ based on Prop. 3. Thus, we now attempt to find a hedge for this query.*

Firstly, we note that the only possible C-forest rooted in $\text{an}(Y)$ that does not contain X is just $\{Y\}$. Thus, the smaller forest of a possible hedge is fixed, and it remains to check whether we can extend this forest to find a larger forest that does include X. The candidates for the larger forest are considered in order. Firstly, the set $\{X, Y\}$ does not form a C-forest, since X, Y are not connected with a bidirected graph in the subgraph induced by X, Y. Further, $\{X, Y, Z_2\}$ also do not form a C-forest, since they are not connected. Finally, $\{X, Y, Z_2, Z_1\}$ is still not a C-forest, since it is not a single C-component. Thus, we conclude $P(Y \mid do(X = x))$ is identifiable in this graph.

We next consider the causal graph in Fig. 4c. Then, take $\{Z_2\}$ as a Z_2-rooted C-forest, and note it can be extended to $\{Z_2, Z_1, X\}$, thus giving a hedge, and showing non-identification of the query. Similarly, in the bow graph from Fig. 2b, we can take $\{Y\}$ and $\{X, Y\}$ as the smaller and larger C-forests, and show non-identification of the effect. □

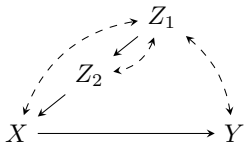
Hedges provide us with a necessary and sufficient graphical criterion for determining if an interventional distribution is identifiable. However, the number of sets which could form a hedge could in principle be large, and navigating this space may be a non-trivial exercise.

Towards an upper bound. For our previous two identification strategies, the back-door and the $P(V \setminus X \mid do(X = x))$ strategy, there was a simple test for verifying whether the strategy can guarantee identification. For finding hedges, no such criterion is known to exist. For purposes of our investigation, we are therefore interested in a graphical structure which ensures that a hedge exists but is not necessary for it, and that can be easily verified from the graph. This is encapsulated in the following theorem.

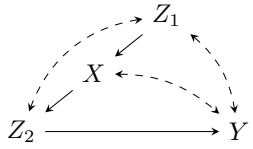
Theorem 1 (0^{th} and 1^{st} Order Hedges – Upper Bounds) *Let $\text{an}_{\mathcal{G}_{\overline{X}}}(Y)$ be the set of ancestors of Y in the graph $\mathcal{G}_{\overline{X}}$ obtained by removing all the incoming arrows into X. Then, if there is a bidirected path in $\mathcal{G}[\text{an}_{\mathcal{G}_{\overline{X}}}(Y)]$ from X to any of its children, there exists a hedge for the query $P(Y \mid do(X = x))$. We say this is a 0^{th} order hedge.*

Furthermore, let $\text{pa}^*(X)$ be the parents of X that are not in $\text{an}_{\mathcal{G}_{\overline{X}}}(Y)$. Let $\text{pa}^{**}(X)$ be the subset of $\text{pa}^*(X)$ such that $\{X, \text{pa}^{**}(X)\}$ form a C-component in the induced subgraph $\mathcal{G}[\{X, \text{pa}^*(X)\}]$. Then, if there exists a bidirected path between $\{X, \text{pa}^{**}(X)\}$ and any child of X in the graph $\mathcal{G}[\{\text{an}_{\mathcal{G}_{\overline{X}}}(Y), \text{pa}^{**}(X)\}]$, then there exists a hedge for the query $P(Y \mid do(X = x))$. We say this is a 1^{st} order hedge. □

The above theorem gives two ways of ensuring that a hedge exists. The first way is to add X to a bidirected path in $\mathcal{G}[\text{an}_{\mathcal{G}_{\overline{X}}}(Y) \setminus X]$ that originates in some $\text{ch}(X)$ (what we call a a 0^{th} order hedge). The second strategy is to add not just X, but rather X and some of its select parents $\text{pa}^{**}(X)$ (what



(a) Napkin graph first variation.



(b) Napkin graph second variation.

Figure 6: Causal graphs from Ex. 10.

we call a a 1^{st} order hedge). We now give the proof of the theorem (see the mental picture shown in Fig. 5 for an illustration of the setting).

Proof Suppose that $S_1 \in \text{ch}(X)$ is such that a bidirected path exists from S_1 to X only through $\mathcal{G}[\text{an}_{\mathcal{G}_{\overline{X}}}(Y)]$. Let $S_1 \dashleftarrow\dashrightarrow \ldots \dashleftarrow\dashrightarrow S_k \dashleftarrow\dashrightarrow X$ be this path. Notice that $\{S_1, \ldots, S_k\}$ form a C-component (since they lie on a bidirected path), and thus they form a $\{S_1, \ldots, S_k\}$-rooted C-forest. We can then add X to $\{S_1, \ldots, S_k\}$, to obtain a larger C-forest with the same set of roots that contains X, thus guaranteeing a hedge. We call this the 0-th order approximation. The second part is similar but slightly more involved. Instead of having a path starting in S_1 and entering from S_k directly to X, we now allow it to enter into a parent of X, which then has a subsequent bidirected path to X only along $\text{pa}^*(X)$. Then, instead of adding just X to $S_1, \ldots, S_k$ as in the 0-th order approximation, we add X and all of its parents that are in the same C-component in the subgraph induced by $X, \text{pa}^*(X)$, which were labelled $\text{pa}^{**}(X)$ in Thm. 1. Once again, we can see that this structure forms a hedge. ■

The above theorem provides an easily verifiable criterion that guarantees the existence of a hedge for an identification query of interest. Therefore, whenever the condition from Thm. 1 is satisfied, the query cannot be identified. In other words, the criterion cannot give us any false negatives with respect to identification – there is no query that is identifiable while Thm. 1 claims it is not identifiable. However, the theorem provides an upper bound, since it allows for the possibility of false positives with respect to identification – there may be queries that are not identifiable but the graph does not contain a 1^{st} order hedge. This is investigated in the following example:

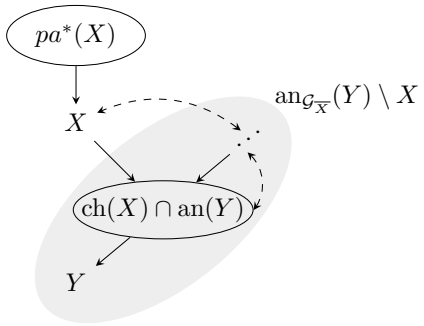


Figure 5: Mental picture of Thm. 1.

Example 10 (Hedges Upper Bound) *Consider the Napkin diagram in Fig. 4b and the query $P(Y \mid do(X = x))$. As discussed in Ex. 9 there is no hedge for this query and therefore it is identifiable. When applying Thm. 1, we first find that $\text{pa}^*(X) = Z_2$, and $\text{pa}^{**}(X) = \emptyset$. Therefore, the graph $\mathcal{G}[\{\text{an}_{\mathcal{G}_{\overline{X}}}(Y), \text{pa}^{**}(X)\}]$ is just $X \to Y$, and there is no bidirected path from X to Y in this graph, making the query identifiable according to the theorem.*

Now, consider a variation of the Napkin graph in Fig. 6a. Note that $F_0 = \{Y\}, F_1 = \{Y, X, Z_2, Z_1\}$ is a hedge for the $P(Y \mid do(X = x))$ query, making the query non-identifiable. However, when applying Thm. 1, we obtain the same result as before, namely $\text{pa}^*(X) = Z_2, \text{pa}^{**}(X) = \emptyset$ and $\mathcal{G}[\{\text{an}_{\mathcal{G}_{\overline{X}}}(Y), \text{pa}^{**}(X)\}]$ equal to $X \to Y$, which means there is no bidirected path from X to its child Y. Here, Thm. 1 would mistakenly imply the query is identifiable, giving a "false positive".

Finally, consider another variation of the Napkin graph in Fig. 6b. Again, $F_0 = \{Y\}, F_1 = \{Y, X, Z_2, Z_1\}$ form a hedge for the $P(Y \mid do(X = x))$ query. However, with Thm. 1 we obtain that $\text{pa}^*(X) = Z_1, \text{pa}^{**}(X) = \emptyset$, and thus $\mathcal{G}[\{\text{an}_{\mathcal{G}_{\overline{X}}}(Y), \text{pa}^{**}(X)\}]$ is $X \to Z_2 \to Y$ with a bidirected edge

$X \dashleftarrow\dashrightarrow Y$. *In this graph, there is no bidirected path between* X *and its child* Z_2*, and thus the query would be deemed identifiable according to Thm. 1, even though a hedge exists. This offers another example of a "false positive" with respect to identification produced by Thm. 1.* □

Towards Probabilities of Identification. Based on the identification strategies discussed so far, we can motivate the notion of a probability of identification (defined formally in Sec. 3). The key idea is that we will think of sampling causal diagrams $\mathcal{G}$ *at random*, from a sampling model. Furthermore, the treatment variable X and the outcome variable Y will be chosen at random from the set of observed variables V. Under such random sampling, we will quantify how likely different strategies for identification are to identify the query $P(Y \mid do(X = x))$. In particular, we are interested in probabilities of back-door identification $P(\text{BD ID})$, joint identification $P(\text{Joint ID})$, complete identification $P(\text{ID})$, and 1^{st} order hedge identification $P(1^{st}\text{ OH ID})$. Another important strategy we mention is that of 0 ID, which checks if there are any causal paths from X to Y in the graph $\mathcal{G}$ – if there are none, the strategy identifies the query $P(Y \mid do(X = x))$ as $P(Y)$, and gives a negative answer otherwise. We note that, strictly speaking, back-door identification does not encompass 0 ID, since back-door type of adjustment always provides an identification expression of the form

$$P(Y \mid do(X = x)) = \sum_{z} P(Y \mid X = x, Z = z)P(Z = z), \tag{23}$$

which, semantically, cannot produce the answer $P(Y)$ obtained from 0 ID (unless additional assumptions are made). However, for simplicity of the exposition, when we write BD ID, we assume that both 0 ID and back-door ID are included in the strategy. Importantly, the following set of relations holds for different probabilities of identification, regardless of the sampling model chosen:

$$P(0\text{ ID}) \leq P(\text{BD ID}) \leq P(\text{Joint ID}) \leq P(\text{ID}) \leq P(1^{st}\text{ OH ID}). \tag{24}$$

A visualization is given in Fig. 7, which shows the space of graph-query pairs, defined by the triplet $(\mathcal{G}, X, Y)$. Different ID strategies form a hierarchy, in the sense that 0 ID is encompassed by 0 + Backdoor ID, which is encompassed by Joint ID of $P(V \setminus X \mid do(X = x))$, which is encompassed by complete ID, further encompassed by 1^{st} OH ID. These inclusion relations imply the inequalities in Eq. 24.

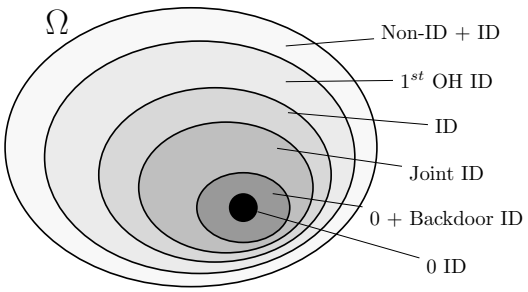


Figure 7: Hierarchy of identification strategies.

We can now summarize our strategy for the remainder of the manuscript, as shown in Tab. 1. In Sec. 2.2 we discussed the back-door identification strategy, and in Prop. 2 we stated a one-shot verification for the existence of a back-door set. In the sequel, in Thm. 2 we will compute exactly (via dynamic programming) the probability that the query $P(Y \mid do(X = x))$ is identifiable in a randomly sampled causal graph, where the directed part is drawn from $\overrightarrow{G}(n, p)$, and the bidirected part from $G(n, q)$. Further, in Sec. 2.3 we discussed a criterion for checking the identification of $P(V \setminus X \mid do(X = x))$, which further allows us to compute our query of interest (Prop. 3). We will compute in Thm. 4 the exact probability for this to occur, which will serve as a lower bound for the probability of complete identification. Finally, in Sec. 2.4 we discussed the graphical criterion called a hedge, which represents a necessary and sufficient condition for verifying identifiability. However, the exact probability of this event may be difficult to compute, so we will instead compute its upper bound (see Thm. 6 in the sequel), based on the notion of a 1^{st} order hedge from Thm. 1.

ID Strategy	Probability of ID	Section	Key Result	P(event)
Back-door	P(BD ID)	Sec. 2.2	Prop. 2	Thm. 2
$P(V \setminus X \mid do(X = x))$ (Lower Bound)	P(Joint ID)	Sec. 2.3	Prop. 3	Thm. 4
Hedges (Complete)	P(ID)	Sec. 2.4	Prop. 4	✘
1^{st} Order Hedge (Upper Bound)	$P(1^{st}$ OH ID)	Sec. 2.4	Thm. 1	Thm. 6

Table 1: Summary of the paper organization.

As we will witness, the displayed identification strategies are in fact, probabilistically, rather close together. Furthermore, the upper and lower bounds for complete identification are remarkably similar.

3. Probabilities of Identification

In this section, we move to the theoretical analysis of the questions of identifiability, when the causal graph $\mathcal{G}$, and variables X, Y are chosen at random. We denote by $\mathcal{G} = (\mathcal{V}, E_d, E_b)$ the causal graph, with $\mathcal{V}$ being the set of vertices, E_d the set of directed edges, and E_b the set of bidirected edges. For simplicity, from now on, we assume that the vertices of $\mathcal{G}$ are labelled $1, \ldots, n$, and the numbering also represents a fixed ordering over the vertices. By $\mathbb{G}_n^{\text{SM}}$ we denote the space of all Semi-Markovian models over n vertices. The set $\{1, \ldots, i\}$ is abbreviated with $[i]$, and the subgraph of $\mathcal{G}$ induced by the vertices $\mathcal{V}'$ is denoted by $\mathcal{G}[\mathcal{V}']$. The random variable I_X taking values in $[n]$ denotes the index of the treatment variable, that is $X = V_{I_X}$. Similarly, I_Y denotes the index of the outcome variable, such that $Y = V_{I_Y}$. Further, we label the interventional distribution $P(Y \mid do(X = x))$ with $Q(I_X, I_Y)$, with $Q(i, j)$ indicating the interventional distribution $P(V_j \mid do(V_i = v_i))$. We write $Q(I_X, I_Y) \vdash_{id}^{\mathcal{A}} \mathcal{G} = 1$ if the query is identifiable from observational data and the graph $\mathcal{G}$, where $\mathcal{A}$ indicates an identification strategy. Otherwise $Q(I_X, I_Y) \vdash_{id}^{\mathcal{A}} \mathcal{G} = 0$.

Different identification strategies are listed in Tab. 1. We first compute the probability of back-door identification, for which we use the notation $\vdash_{id}^{bd}$ (and abbreviate the strategy BD ID). Then, we will compute the probability of identifying $P(V \setminus X \mid do(X = x))$ in the subgraph $\mathcal{G}[\text{an}(I_Y)]$, which is abbreviated as Joint ID (since the joint distribution $V \setminus X$ is identified in this strategy). This strategy provides a lower bound for complete identification, and we use the notation $\vdash_{id}^{jid}$. Similarly, we will also compute the probability of the 1^{st} order hedge criterion from Thm. 1, and use the notation $\vdash_{id}^{1oh}$ (the strategy is abbreviated as 1^{st} OH ID). Note that this strategy provides an upper bound for complete identification (meaning that $Q(I_X, I_Y) \vdash_{id} \mathcal{G} = 1 \implies Q(I_X, I_Y) \vdash_{id}^{1oh} \mathcal{G} = 1$). The probability of complete identification, for which we simply use $\vdash_{id}$, will not be computed explicitly, but we will instead resort to its upper and lower bounds.

Equipped with the necessary notation, we begin by providing a general definition of a *probability of identification*:

Definition 12 (Probability of Identification) *Let μ be a probability measure over the space of Semi-Markovian models $\mathbb{G}_n^{SM}$. Let I_X, I_Y be two randomly drawn subsets of $\{1, \ldots, n\}$ according to a probability measure λ. Further, let $\mathcal{A}$ denote an identification strategy. We then define the probability of identification with respect to $(\mu, \lambda, \mathcal{A})$ as*

$$P_{\lambda,\mu}(\mathcal{A}\ ID) \triangleq \mathbb{P}_{\lambda,\mu}(Q(I_X, I_Y) \vdash_{id}^{\mathcal{A}} \mathcal{G} = 1). \tag{25}$$

□

(a) P = $(1-p)(1-q)$. (b) P = $p(1-q)$. (c) P = $(1-p)q$. (d) P = pq.

Figure 8: Causal diagrams in Sec. 2.

The provided definition allows for very general probabilities of identification, and some simplification is in order. Firstly, for the rest of the manuscript, we focus on identifying the causal effects of singleton variables X on singleton variables Y. In particular, since for the identification of $P(Y \mid \text{do}(X))$ the non-ancestors of Y can be ignored, we assume that Y is the last variable in the graph, meaning that $I_Y = n$, and X is chosen uniformly over the remaining $n-1$ variables. Furthermore, for the theoretical analysis of this section, we restrict the measure μ such that edges E_d are drawn independently with probability p, and edges E_b independently with probability q. Therefore, our choice of λ, μ yields:

(a) restricting λ so that $I_Y = \{n\}$ with probability 1, and I_X chosen uniformly over $\{1, \ldots, n-1\}$, i.e.,

$$\mathbb{P}(I_X = i) = \frac{1}{n-1} \; \forall i \in \{1, \ldots, n-1\}, \tag{26}$$

$$\mathbb{P}(I_Y = n) = 1. \tag{27}$$

(b) restricting μ so that

$$\mathbb{P}((i,j) \in E_d) = p, \tag{28}$$

$$\mathbb{P}(\{(i,j),(j,i)\} \in E_b) = q, \tag{29}$$

independently for each i, j such that $i < j$.

Eq. 28 corresponds to sampling edges from a directed Erdős–Rényi model $\overrightarrow{G}(n,p)$, whereas Eq. 29 corresponds to sampling from an undirected Erdős–Rényi model $G(n,q)$. We now give an illustration for graphs over two nodes:

Example 11 (Two Node Graphs) *Consider the 4 possible graphs over two nodes in Fig. 8 The empty graph (a) occurs when the directed edge is not sampled (with probability $1-p$), and the bidirected edge is not sampled either (with probability $1-q$). The graph (b) occurs with probability $p(1-q)$ since the directed edge is sampled, and the bidirected is not. Similarly for graph (c). Finally, in graph (d), both edges are sampled, which happens with probability pq.* □

For the remainder of this section, we are concerned with computing the values of $P_{\lambda,\mu}$(BD ID), $P_{\lambda,\mu}$(Joint ID) and $P_{\lambda,\mu}(1^{st}$ OH ID) for the specific choice of μ and λ given above. In Section 4, we consider other interesting choices of μ. The main question we are trying to answer is the following: if we sample a causal graph at random, how likely are different identification strategies to work?

3.1 Computing $P_{\lambda,\mu}$(BD ID)

In this section, we show how to efficiently compute the probability of identification for the back-door criterion for the above-described random graph model. We begin by proving the following theorem, which gives a way to compute this quantity, although it is not immediately clear whether the computation can be performed efficiently. In the subsequent lemmas following the theorem, we show that this is indeed the case.

Theorem 2 (Probability of Back-door Identification) *Let $\mathcal{G}, I_X, I_Y$ be sampled according to Eq. 26-29. Further, define the random variables $S_1 := |\mathrm{an}(n) \cap [i-1]|, S_2 := |\mathrm{an}(n) \cap \mathrm{de}(i)|, S_3 := |\mathrm{an}(n) \cap \{i, \ldots, n\}|$, and let $P(S_1, S_2, S_3)$ denote their joint distribution. Then,*

$$\mathbb{P}(Q(i,n) \vdash_{id}^{bd} \mathcal{G} = 1) = \sum_{\ell=0}^{i-1} \sum_{j=1}^{n-i+1} \sum_{k=j}^{n-i+1} \sum_{c=0}^{\ell+k-j} \Big[\binom{\ell+k-j}{c} p_{c+1,q} (1-q)^{(c+1)(\ell+k-j-c)} (1-q)^{(c+1)(j-1)} \cdot P(S_1 = \ell, S_2 = j, S_3 = k) \Big] + P(S_2 = 0), \tag{30}$$

where $p_{c,q}$ is the probability that a randomly drawn undirected Erdős-Renyi graph $G(c,q)$ is connected. Therefore, the probability of back-door identification can be computed as

$$\mathbb{P}(Q(I_X, n) \vdash_{id}^{bd} \mathcal{G} = 1) = \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{P}(Q(i,n) \vdash_{id}^{bd} = 1). \binom{\ell+k-j}{c} \tag{31}$$

□

Proof According to Prop. 2, for identification of the effect of i on n we need to verify whether the set $Z = \mathrm{an}(i \cup n) \setminus F$ satisfies the back-door criterion, where F are the variables on causal paths $i \to \cdots \to n$. Therefore, F is the set $\mathrm{an}(n) \cap \mathrm{de}(i) \setminus \{i, n\}$. First, note that if there are no variables in $\mathrm{an}(n) \cap \mathrm{de}(i)$, then there is no causal path from i to n, and the effect is trivially identifiable (this case corresponds to the term $P(S_2 = 0)$ in Eq. 30). Further, assume now that $S_2 > 0$, i.e., there is a causal path from i to n. Notice that variables that are not in $\mathrm{an}(n)$ can be ignored since they cannot lie on an open back-door path from i to n, or a causal path from i to n. Now, let $F_{+n} = F \cup \{n\}$. Suppose there is a bidirected path from i to any element of F_{+n}, going through $\mathrm{an}(n) \setminus F_{+n}$ only. Let this path be denoted by $i \dashleftarrow\dashrightarrow s_1 \dashleftarrow\dashrightarrow \ldots \dashleftarrow\dashrightarrow s_t$, where $s_t \in F_{+n}$. Then, since all the nodes $s_1, \ldots, s_{t-1}$ are in the Z set, and there is a directed path from s_t to n (or $s_t = n$), this creates an open path not blocked by Z. Therefore, Z is not a back-door set for (i,n) (and hence by Prop. 2 there is no back-door set). Conversely, suppose that there is no bidirected path from i to F_{+n}, going through $\mathrm{an}(n) \setminus F_{+n}$ only. Take any path from i to F_{+n}, going through nodes $i, s_1, \ldots, s_t$, where $s_t \in F_{+n}$, $\{s_1, \ldots, s_{t-1}\} \in \mathrm{an}(n) \setminus F_{+n}$. Since the path $i, s_1, \ldots, s_t$ cannot consist of only bidirected edges, it must contain a directed edge, between s_{t_0} and s_{t_0+1} say. However, both s_{t_0} and s_{t_0+1} are in the Z set, and at least one of these is not a collider on the path, so the path is necessarily blocked (one may also additionally check then special cases when the directed edge is the edge $i \leftarrow s_1$ or $s_{t-1} \to s_t$, where in the former case s_1 is a non-collider in the Z set, while in the latter it is s_{t-1}). Therefore, any back-door path between i and F_{+n} is blocked by the Z set, in turn implying that every back-door path from i to n is also blocked.

Now, we condition on the event $S_1 = \ell, S_2 = j, S_3 = k$. We wish to compute the probability that i can reach F_{+n} via nodes $\mathrm{an}(n) \setminus F_{+n}$ using bidirected edges. Let C denote the set of nodes in $\mathrm{an}(n) \setminus \mathrm{de}(i)$ that can be reached from i in the subgraph $\mathcal{G}[\mathrm{an}(n) \setminus \mathrm{de}(i)]$ using only bidirected edges, and let $|C| = c$. There are $\binom{\ell+k-j}{c}$ different choices for the set C, and the probability that i can reach C and no other node in $\mathcal{G}[\mathrm{an}(n) \setminus \mathrm{de}(i)]$ by bidirected edges equals

$$\underbrace{p_{c+1,q}}_{C,\{i\} \text{ connected}} \times \underbrace{(1-q)^{(c+1)(\ell+k-j-c)}}_{\text{other } \ell+k-j-c \text{ nodes not reached}}. \tag{32}$$

Since exactly c nodes can be reached from i, it remains to ensure that none of the nodes $C \cup \{i\}$ have a bidirected edge into F_{+n}, which satisfies $|F_{+n}| = j-1$, yielding an additional term $(1-q)^{(c+1)(j-1)}$. Summing over different values of c, ℓ, j, k yields Eq. 30. ■

Following the approach of (Gilbert, 1959), the probability $p_{c,q}$ can be computed recursively for any

c as

$$1 - p_{c,q} = \sum_{r=1}^{c-1} \binom{c-1}{r-1} p_{r,q} (1-q)^{r(c-r)}, \tag{33}$$

with the initial value $p_{1,q} = 1$. To see why this holds, consider conditioning on the size of the connected component of some node in a graph drawn from $G(c,q)$ (component size is denoted by r). Then, the set of nodes in the connected component can be chosen in $\binom{c-1}{r-1}$ ways, $p_{r,q}$ guarantees that the r vertices in the connected component are indeed connected, and $(1-q)^{r(c-r)}$ ensures there are no nodes from the connected component of size r to any of the remaining $c-r$ nodes. Then, Eq. 33 follows from the law of total probability.

Therefore, Thm. 2 allows us to compute the probability of back-door identification, provided we can recover the joint distribution over $S_1 = |\mathrm{an}(i)|, S_2 = |\mathrm{an}(n) \cap \mathrm{de}(i)|, S_3 = |\mathrm{an}(n)|$. To do so, we first find the joint distribution over the $S_2 = |\mathrm{an}(k) \cap \mathrm{de}(1)|, S_3 = |\mathrm{an}(k)|$ for a graph of size k, which can be computed in polynomial time.

Lemma 1 (Recovering $P(S_2, S_3)$) *Let $\mathcal{G}$ be drawn from $\overrightarrow{G}(n,p)$. Define the random variables $S_2 := |\mathrm{an}(n) \cap \mathrm{de}(1)|, S_3 := |\mathrm{an}(n)|$. Then, the joint distribution $P(S_2, S_3)$ over S_2, S_3, can be computed in polynomial time with complexity $O(n^5)$.* □

Proof Denote by $\mathcal{G}[[i]]$ the graph consisting of all edges with both endpoints at most i and by $\mathcal{G} - \mathcal{G}[[i]]$ the graph of all edges touching $\{i+1, \ldots, n\}$. We shall calculate the distribution of (S_2, S_3) using dynamic programming, whereby in step i we expose the edges from vertices $1, \ldots, i-1$ to vertex i, thus after i steps we have exposed $\mathcal{G}[[i]]$. The crucial idea is, after exposing the edges to vertex i, to also "decide" whether or not i will be in $\mathrm{an}(n)$. Then, after i steps, we distinguish four types of vertices. Each vertex $j \in [i]$, can either be:

- unreachable if $j \notin \mathrm{de}(1)$,
- mandatory if $j \in \mathrm{de}(1)$ and in $\mathcal{G} - \mathcal{G}[[i]]$ there must be a path from j to n,
- ensured if $j \in \mathrm{de}(1)$ and j has an edge to a vertex j' with $j < j' \leq i$ for which we have decided that j' will be in $\mathrm{an}(n)$ (implying that j' is now either mandatory or ensured),
- or forbidden if $j \in \mathrm{de}(1)$, but we have decided that j will not be in $\mathrm{an}(n)$.

Informally speaking, after exposing the edges to vertex i, if it is in $\mathrm{de}(1)$, we make two separate guesses: one that i will be in $\mathrm{an}(n)$ and one that i will not be in $\mathrm{an}(n)$. While running the dynamic program, we only keep the sets of guesses which are consistent with $\mathcal{G}[[i]]$, the graph exposed so far. A set of guesses becomes invalidated if we have guessed that i will be in $\mathrm{an}(n)$ and there is an edge from a forbidden vertex $j < i$ to i. Additionally, if a vertex $j \leq i$ is ensured, then by definition, there is a vertex j' such that $j \to j'$ and j' will be in $\mathrm{an}(n)$, hence it is irrelevant whether or not j can reach n in $\mathcal{G} - \mathcal{G}[[i]]$. Finally, when we reach $i = n$, we must make sure that all our guesses were correct, i.e. every vertex which is still mandatory, must have an edge toward n while no forbidden vertex can have an edge toward n.

To devise a polynomial time algorithm, what is left to observe is that at some point i, we only need to keep track of the number of vertices of each type and not the precise sets of vertices of each type.

Now, we formally describe the algorithm. To this end, given the entire graph $\mathcal{G}$ and a vertex $i \in [n]$ contained in a path from 1 to n, let $\mathrm{nxt}(i)$ denote the smallest $j > i$ such that $i \to j$ and j can reach n, that is j is the smallest possible vertex after i in a path from i to n.

Let $f(i, U, M, E, F)$ with $i = U + M + E + F$ denote the probability that the following holds:

- there are precisely U vertices in $[i]$ which are not reachable from 1;

- there are precisely M vertices $j \in [i]$ that are contained in a path from 1 to n and have $\mathrm{nxt}(j) > i$;
- there are precisely E vertices $j \in [i]$ that are contained in a path from 1 to n and have $\mathrm{nxt}(j) \leq i$;
- there are precisely F vertices in $[i]$ that are reachable from 1 but cannot reach n;

conditioned on the probability that for given specific disjoint subsets S_M and S_F of $[i]$ containing M and F vertices, respectively, in the graph $\mathcal{G} - \mathcal{G}[[i]]$, every vertex in S_M can reach n, while no vertex in S_F can reach n. The variables U, M, E, F correspond to the number of unreachable, mandatory, ensured and forbidden vertices after step i, respectively.

In the definition, we are implicitly using that the probability we condition on is the same for any choice of S_M and S_F.

We calculate $f(i, U, M, E, F)$ using dynamic programming with the start case $f(1, 0, 1, 0, 0) = f(1, 0, 0, 0, 1) = 1$ and $f(1, 1, 0, 0, 0) = f(1, 0, 0, 1, 0) = 0$.

The transitions are as follows. At each step, we decide whether the new vertex i is reachable from 1 and if it is whether it can reach n. Then, expose the edges from $[i-1]$ to i and this gives us the new values of U, M, E, F. To avoid writing edge cases, we define $f(i, U, M, E, F)$ to equal 0 for any undefined values where one of the parameters is negative or when $i \neq U + M + E + F$. We then formally have for any $i \geq 2$:

$$\begin{aligned} f(i, U, M, E, F) = & f(i-1, U-1, M, E, F) \cdot \mathbb{P}[\mathrm{Bin}(M+E+F, p) = 0] + \\ & f(i-1, U, M, E, F-1) \cdot \mathbb{P}[\mathrm{Bin}(M+E+F-1, p) > 0] + \\ & f(i-1, U, M-1, E, F) \cdot \mathbb{P}[\mathrm{Bin}(M, p) = 0] \cdot \mathbb{P}[\mathrm{Bin}(F, p) = 0] \cdot \mathbb{P}[\mathrm{Bin}(E, p) > 0] + \\ & \sum_{k=1}^{E} f(i, U, M+k-1, E-k, F) \cdot \mathbb{P}[\mathrm{Bin}(M, p) = k] \cdot \mathbb{P}[\mathrm{Bin}(F, p) = 0]. \end{aligned}$$

The first line corresponds to vertex i not being reachable from 1; the second line to vertex i being reachable from 1 but it cannot reach n; the third line to vertex 1 being on a path from 1 to n with no incoming edges from the set of vertices j with $\mathrm{nxt}(j) > i-1$ (hence it must have an edge from one of the E vertices with $\mathrm{nxt}(j) \leq i-1$); the fourth line corresponds to i having $k \geq 1$ edges from vertices $j < i$ with $\mathrm{nxt}(j) > i-1$.

Let $S_2 := |\mathrm{an}(n) \cap \mathrm{de}(1)|$ denote the number of vertices which lie on some path from 1 to n in $\mathcal{G}$. Let $S_3' := |\mathrm{de}(1)|$, and $S_3 = |\mathrm{an}(n)|$. Then, for any $m \leq n$:

$$P(S_2 = 0, S_3' = m) = \sum_{U} f(n-1, U, 0, 0, m)(1-p)^m. \tag{34}$$

For any $1 \leq j \leq m$ and $0 \leq m \leq n$:

$$P(S_2 = j, S_3' = j+m) = \sum_{U, M, E : M+E = j-1} f(n-1, U, M, E, m) p^M (1-p)^m. \tag{35}$$

Finally, due to the symmetry of the ancestrality and descendency in the problem, we have that $P(S_2 = j, S_3' = k) = P(S_2 = j, S_3 = k)$, concluding the proof. ■

After recovering the joint distribution $P(S_2, S_3)$ in polynomial time, we can now reconstruct the distribution $P(S_1, S_2, S_3)$ appearing in Thm. 2 using the $P(S_2, S_3)$ distribution as a boundary condition:

Lemma 2 (Recovering $P(S_1, S_2, S_3)$) *Let $\mathcal{G}$ be drawn from $\overrightarrow{G}(n, p)$. Fix $i \in [n-1]$, and define the random variables $S_1 = |\text{an}(i)|, S_2 = |\text{an}(n) \cap \text{de}(i)|, S_3 = |\text{an}(n)|$. The distribution $P(S_1, S_2, S_3)$ can be recovered in polynomial time with complexity $O(n^4)$, given the distribution $P_{\overrightarrow{G}(k,p)}(S_2, S_3)$ as a boundary condition for various values of $k \leq n$.* □

Proof We start from the subgraph over the nodes $i, \ldots, n$, and expose the outgoing edges from nodes $i-1, i-2, \ldots, 1$ in that order. Let $f(t, a, b, c)$ denote the probability that $S_1 = a, S_2 = b, S_3 = c$ in the graph with the outgoing edges from all nodes with an index $\geq t$ exposed. Then, note that we have the following transition:

$$f(t, a, b, c) = \underbrace{f(t-1, a-1, b, c) \cdot (1 - (1-p)^{a-1+c})}_{t \in \text{an}(n)} + \underbrace{f(a-1, a, b, c) \cdot (1-p)^{a+c}}_{t \notin \text{an}(n)}, \tag{36}$$

and the boundary condition $f(i, 0, b, c) = P_{\overrightarrow{G}(n-i+1,p)}(S_2 = b, S_3 = c)$, with the complexity of the dynamic program of $O(n^4)$. ■

We can now tie together the results of Thm. 2, Lem. 1, and Lem. 2, to complete the theoretical discussion on probabilities of back-door identification. The formal result is given in the following corollary.

Corollary 3 (Efficient Computation of Probability of Back-door Identification) *Let $\mathcal{G}$, I_X, and I_Y be sampled according to Eq. 26-29. Then, the probability of back-door identification $\mathbb{P}(Q(I_X, I_Y) \vdash_{id}^{bd} \mathcal{G} = 1)$ can be computed in polynomial time, with computational complexity of $O(n^5)$.* □

Proof First, based on Thm. 2, the expression for $\mathbb{P}(Q(I_X, I_Y) \vdash_{id}^{bd} \mathcal{G} = 1)$ given in Eq. 31 has $O(n)$ terms, and each of terms can be computed in $O(n^4)$ (see Eq. 30) steps provided $p_{c,q}$ and $P(S_1, S_2, S_3)$ and are available. This computation thus takes $O(n^5)$ steps. Furthermore, based on the recursion in Eq. 33, probabilities $p_{c,q}$ can be computed in $O(n^2)$ for all $c \leq n$. Furthermore, in Lem. 1 we showed how the joint distributions $P(S_2, S_3)$ for all graph sizes $n_0 \leq n$ can be computed in $O(n^5)$ steps. Using the distributions $P(S_2, S_3)$, Lem. 2 shows how $P(S_1, S_2, S_3)$ can be computed in $O(n^4)$. Therefore, it follows that probabilities of back-door identification can be computed efficiently, with a computational cost of $O(n^5)$. ■

3.2 Computing $P_{\lambda,\mu}$(Joint ID) (Lower Bound)

Our next task is to efficiently compute the lower bound for the probability of complete identification. We follow a similar approach as in Sec. 3.1. First, we give the general theorem which allows us to compute the lower bound for the probability of identification, and then show via subsequent lemmas that this can be done efficiently.

Theorem 4 (Probability of Identification Lower Bound) *Let $\mathcal{G}, X, Y$ be sampled according to Eq. 26-29. Fix some $i \in [n-1]$ and define the random variables $T_1 := |\text{an}(n)|, T_2 := |\text{ch}(i) \cap \text{an}(n)|$. Let $P(T_1, T_2)$ denote their joint distribution. Then, we have that:*

$$\mathbb{P}(Q(i, n) \vdash_{id}^{lb} \mathcal{G} = 1) = \sum_{k=1}^{n-i-1} \sum_{j=k}^{n} \sum_{r=0}^{j-k-1} \Big[\binom{j-k-1}{r} p_{r,k,q} (1-q)^{(r+k)(j-k-r)} \cdot P(T_1 = j, T_2 = k) \Big] + P(T_2 = 0), \tag{37}$$

where $p_{r,k,q}$ is the probability that a set of k nodes is connected to the remaining r nodes in a graph sampled from the undirected Erdős-Rényi model $G(r+k, q)$. It follows that the lower bound for the

probability of identification can be computed as:

$$\mathbb{P}(Q(I_X, n) \vdash_{id}^{lb} \mathcal{G} = 1) = \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{P}(Q(i, n) \vdash_{id}^{lb} = 1). \tag{38}$$

□

Proof [Thm. 4] First, note that if $T_2 = 0$, the query $Q(i, n)$ is trivially identifiable. Now, we fix $i \in [n-1]$ and condition on the event $T_1 = j, T_2 = k$. Based on Prop. 3, the query $Q(i, n)$ is identifiable if and only if there is no bidirected path through $\mathrm{an}(n)$ from i to any of its k children that are in $\mathrm{an}(n)$. Now, let the set R denote the vertices in $\mathrm{an}(n) \setminus \mathrm{ch}(i)$ that are reached from the k children of i in $\mathrm{an}(n)$ via bidirected paths within the subgraph over $\mathrm{an}(n) \setminus \{i\}$ (as before, vertices not in $\mathrm{an}(n)$ can be ignored), and let $C = \mathrm{an}(n) \setminus \{i, \mathrm{ch}(i)\}$. We further condition on $|R| = r$, and note that the r vertices can be chosen in $\binom{j-k-1}{r}$ ways, since the vertices are chosen from C which satisfies $|C| = j - k - 1$. Then, the term $p_{r,k,q}$ ensures the k sources are connected to R, while the term $(1-q)^{(r+k)(j-k-1-r)}$ ensures that the k sources and the set R are not connected to any other node in C. An additional term $(1-q)^{r+k}$ is necessary to ensure that there is no bidirected edge from i to $\mathrm{ch}(i) \cap \mathrm{an}(n)$ or R, which guarantees identifiability. Summing over the values of r, j, k yields the result in Eq. 37. ■

The first step towards using Thm. 4 to efficiently compute the lower bound for the probability of identification is to recover the probabilities $p_{r,k,q}$, as shown in the following supporting lemma:

Lemma 3 *The probability $p_{r,k,q}$ defined in Thm. 4 satisfies the recursion*

$$1 - p_{r,k,q} = \sum_{r_0=0}^{r-1} \binom{r}{r_0} p_{r_0,k,q} (1-q)^{r_0(r-r_0)} (1-q)^{k(r-r_0)}, \tag{39}$$

with $p_{0,k,q} = 1$. □

Proof Let r_0 denote the number of points reached from the k sources, which can be chosen in $\binom{r}{r_0}$ ways, and are connected to k sources with probability $p_{r_0,k,q}$. Further, term $(1-q)^{r_0(r-r_0)}$ ensures there are no connections between the remaining $r - r_0$ points and the r_0 reached points, and $(1-q)^{k(r-r_0)}$ ensures no edges between the j sources and $r - r_0$ remaining points. The law of total probability yields Eq. 39. ■

Given the above lemma, it remains to recover the joint distribution $P(T_1, T_2)$ from Thm. 4. This is achieved as follows:

Lemma 4 (Recovering $P(T_1, T_2)$) *Let $\mathcal{G}$ be drawn from $\overrightarrow{G}(n, p)$. Fix $i \in [n-1]$ and define the random variables $T_1 := |\mathrm{an}(n)|, T_2 := |\mathrm{ch}(i) \cap \mathrm{an}(n)|$. Then, the joint distribution $P(T_1, T_2)$, can be computed in polynomial time with time complexity $O(n^3)$.* □

Proof We will use a multiple exposure argument in which we will gradually expose the outgoing edges from nodes $n-1, n-2, \ldots, 1$ in that order. Let $f(t, a, b)$ denote the probability that $T_1 = a, T_2 = b$ in the graph in which we exposed only the outgoing edges from the vertices with an index $\geq t$. Then, we obtain the following transitions:

$$f(t,a,b) = \begin{cases} f(t+1, a-1, 0) \cdot (1 - (1-p)^{a-1}) + f(t+1, a, 0) \cdot (1-p)^a & \text{if } t > i, b = 0, \\ f(t+1, a, 0) \cdot (1-p)^a & \text{if } t = i, b = 0, \\ f(t+1, a-1, 0) \cdot \binom{a-1}{b} p^b (1-p)^{a-b-1} & \text{if } t = i, b > 0, \\ f(t+1, a-1, b) \cdot (1 - (1-p)^{a-1}) + f(t+1, a, b) \cdot (1-p)^a & \text{if } t < i, \end{cases} \tag{40}$$

and the boundary condition $f(n, 1, 0) = 1$. The complexity of the dynamic program is $O(n^3)$. ■

Equipped with the above lemmas, we are now able to efficiently compute the probability of identification for the model in which the observed structure is sampled from $\overrightarrow{G}(n,p)$, while the unobserved structure is sampled from a $G(n,q)$ model:

Corollary 5 (Efficient Computation of Lower Bound for Probability of Identification) *Let $\mathcal{G}$, I_X, and I_Y be sampled according to Eq. 26-29. Then, the lower bound for the probability of identification, $\mathbb{P}(Q(I_X,I_Y)\vdash_{id}^{lb}\mathcal{G}=1)$, can be computed in polynomial time, with computational complexity of $O(n^4)$.* □

Proof First, based on Thm. 4, the expression for $\mathbb{P}(Q(I_X,I_Y)\vdash_{id}=1)$ given in Eq. 38 has $O(n)$ terms, and each of terms can be computed in $O(n^3)$ steps (see Eq. 37) provided $p_{r,k,q}$ and $P(T_1,T_2)$ are available. This computation thus takes $O(n^4)$ steps. Furthermore, based on the recursion in Eq. 39, probabilities $p_{r,k,q}$ can be computed in $O(n^3)$ for all $r,k\leq n$. Furthermore, in Lem. 4, we showed how the joint distribution $P(T_1,T_2)$ can be recovered in $O(n^3)$ steps. Therefore, it follows that the lower bound for the probability of identification can be computed efficiently, with a computational cost of $O(n^4)$. ■

3.3 Computing $P_{\lambda,\mu}(1^{st}$ OH ID$)$ (Upper Bound)

After computing the lower bound for the probability of identification, we now move onto computing the upper bound. We proceed similarly, and prove the expression for its value in the following theorem:

Theorem 6 (Probability of Identification Upper Bound) *Let $\mathcal{G},I_X,I_Y$ be sampled according to Eq. 26-29. Fix some $i\in[n-1]$ and define the random variables $M_1:=|\mathrm{an}_{G_{\overline{i}}}(n)\setminus\{i\}|, M_2:=|\mathrm{ch}(i)\cap\mathrm{an}(n)|, M_3:=|\mathrm{pa}^{**}(i)|$, where $\mathrm{pa}^{**}(i)$ follows the definition from Thm. 1. Let $P(M_1,M_2,M_3)$ denote their joint distribution. Then, we have that:*

$$\mathbb{P}(Q(i,n)\vdash_{id}^{ub}\mathcal{G}=1)=\sum_{k=1}^{n-i-1}\sum_{j=k}^{n-1}\sum_{r=0}^{j-k}\sum_{\ell=0}^{i-1}\Big[\binom{j-k}{r}p_{r,k,q}(1-q)^{(r+k)(j-k-r)}(1-q)^{(r+k)(\ell+1)}\cdot P(M_1=j,M_2=k,M_3=\ell)\Big]+P(M_2=0),\tag{41}$$

where $p_{r,k,q}$ is the probability that a set of k nodes is connected to the remaining r nodes in a graph sampled from the undirected Erdős-Rényi model $G(r+k,q)$. It follows that the probability of identification can be computed as:

$$\mathbb{P}(Q(I_X,n)\vdash_{id}^{ub}\mathcal{G}=1)=\frac{1}{n-1}\sum_{i=1}^{n-1}\mathbb{P}(Q(i,n)\vdash_{id}^{ub}=1).\tag{42}$$

□

Proof [Thm. 6] First, note that if $M_2=0$ (i has no children that are ancestors of n), the query $Q(i,n)$ is trivially identifiable. Now, we fix $i\in[n-1]$ and condition on the event $M_1=j,M_2=k,M_3=\ell$. Let $A=\mathrm{an}_{G_{\overline{i}}}(n)\setminus\{i\}$ be the set of ancestors of n in the graph where incoming arrows into i are removed (and we further remove i from this set). A visualization is shown in Fig. 5. Now, denote by R the set of nodes in A that are not in $\mathrm{ch}(i)\cap\mathrm{an}(n)$ that can be reached by bidirected edges from the k nodes in $\mathrm{ch}(i)\cap\mathrm{an}(n)$. Say $r=|R|$, and note that these r nodes can be chosen in $\binom{j-k}{r}$ ways. Note that conditioning on the event $M_3=\ell$ involves bidirected edges, and hence some care needs to be taken for the remainder of the proof. However, this event depends only on the incoming directed edges from $\mathrm{pa}^*(i)$ (the set of parents of i which are not $\mathrm{an}(n)$) into i, and the bidirected

edges within the set $\{i, \mathrm{pa}^*(i)\}$. This, in turn, implies that conditioning on $M_3 = \ell$ is independent of the bidirected edges touching A. Thus, the term $p_{r,k,q}$ ensures that all of the nodes in R are reached from $\mathrm{ch}(i) \cap \mathrm{an}(n)$ while staying only in the set A. Furthermore, the term $(1-q)^{(r+k)(j-k-r)}$ ensures that the remaining $j-k-r$ nodes in A are not reached from $\mathrm{ch}(i) \cap \mathrm{an}(n)$. Finally, if any node in $\{i, \mathrm{pa}^{**}(i)\}$ can be reached from either R or $\mathrm{ch}(i) \cap \mathrm{an}(n)$ then the procedure returns a negative result (according to Thm. 1). Thus, to avoid any such edge, we add an additional term of $(1-q)^{(r+k)(\ell+1)}$, since $\ell + 1$ is the size of the set $\{i, \mathrm{pa}^{**}(i)\}$. Summing over the values of r, j, k yields the result in Eq. 41. ■

To complete our argument, we need to show how the joint distribution $P(M_1, M_2, M_3)$ can be recovered efficiently:

Lemma 5 (Recovering $P(M_1, M_2, M_3)$) *Let $\mathcal{G}$ be drawn from $\overrightarrow{G}(n,p)$. Fix $i \in [n-1]$ and define the random variables $M_1 := |\mathrm{an}_{G_{\bar{i}}}(n) \setminus \{i\}|, M_2 := |\mathrm{ch}(i) \cap \mathrm{an}(n)|, M_3 := |\mathrm{pa}^{**}(i)|$. Then, the joint distribution $P(M_1, M_2, M_3)$, can be computed in polynomial time with time complexity $O(n^4)$.* □

Proof We will again use a multiple exposure argument and gradually expose the outgoing edges from nodes $n-1, n-2, \ldots, 1$ in that order. Let $M_3' = \mathrm{pa}^*(i)$, that is the set of parents of i which have a directed path to n only through i. We first compute the joint distribution of $P(M_1, M_2, M_3')$ and then transform it to $P(M_1, M_2, M_3)$. Let $f(t, a, b, c)$ denote the probability that $M_1 = a, M_2 = b, M_3' = c$ in the graph in which we exposed only the outgoing edges from the vertices with an index $\geq t$. Then, we obtain the following transitions:

$$f(t,a,b,c) = \begin{cases} f(t+1, a-1, 0, 0) \cdot (1-(1-p)^{a-1}) + f(t+1, a, 0, 0) \cdot (1-p)^a & t > i, b = c = 0, \\ f(t+1, a, 0, 0) \cdot \binom{a}{b} p^b (1-p)^{a-b} & t = i, c = 0 \\ f(t+1, a-1, b, 0) \cdot (1-(1-p)^{a-1}) + f(t+1, a, b, 0) \cdot (1-p)^{a+1} & t < i, c = 0, \\ f(t+1, a-1, b, c) \cdot (1-(1-p)^{a-1}) + & \\ f(t+1, a, b, c-1) \cdot (1-p)^a p + f(t+1, a, b, c) \cdot (1-p)^{a+1} & t < i, c > 0. \end{cases} \tag{43}$$

and the boundary conditions $f(n, 1, 0, 0) = 1$ and $f(t, a, b, c) = 0 \ \forall \ t \geq i; c > 0; a, b \geq 0$. The complexity of the dynamic program is $O(n^4)$. Finally, we need to compute $f^*(a, b, d) = P(M_1 = a, M_2 = b, M_3 = d)$ from the obtained f. This can be done as follows:

$$f^*(a,b,d) = \sum_{c=d}^{i-1} f(1,a,b,c) \binom{c}{d} p_{d,1,q} (1-q)^{(d+1)(c-d)} \ \ \forall a, b, d. \tag{44}$$

The term $\binom{c}{d}$ chooses d elements out of $c = |\mathrm{pa}^*(i)|$ that end up in $\mathrm{pa}^{**}(i)$. The term $p_{d,1,q}$ ensures that i has a path to all the d nodes using bidirected edges (in the induced subgraph $\mathcal{G}[\{i, \mathrm{pa}^*(i)\}]$), and the term $(1-q)^{(d+1)(c-d)}$ ensures that the remaining $c-d$ nodes are not in $\mathrm{pa}^{**}(i)$. This additional computation also has a complexity of $O(n^4)$, so recovering $P(M_1, M_2, M_3)$ requires $O(n^4)$ steps altogether. ■

Equipped with the above lemma, we are now able to efficiently compute the upper bound for the probability of identification for the model in which the observed structure is sampled from $\overrightarrow{G}(n,p)$, while the unobserved structure is sampled from a $G(n,q)$ model:

Corollary 7 (Efficient Computation of Upper Bound for Probability of Identification) *Let $\mathcal{G}$, X, and Y be sampled according to Eq. 26-29. Then, the upper bound for the probability of identification, $\mathbb{P}(Q(I_X, I_Y) \vdash_{id}^{ub} \mathcal{G} = 1)$, can be computed in polynomial time, with the computational complexity of $O(n^5)$.* □

Proof Based on Thm. 6, the expression for $\mathbb{P}(Q(I_X, I_Y) \vdash_{id}^{ub} = 1)$ given in Eq. 42 has $O(n)$ terms, and each of terms can be computed in $O(n^4)$ steps (see Eq. 41) provided $p_{r,k,q}$ and $P(M_1, M_2, M_3)$ are available. This computation thus takes $O(n^5)$ steps. Furthermore, based on the recursion in Eq. 39, probabilities $p_{r,k,q}$ can be computed in $O(n^3)$ for all $r, k \leq n$. Furthermore, in Lem. 5, we showed how the joint distribution $P(M_1, M_2, M_3)$ can be recovered in $O(n^4)$ steps. Therefore, it follows that the upper bound for the probability of identification can be computed efficiently, with a computational cost of $O(n^5)$. ■

3.4 Empirical Verification

We now verify the theoretical results obtained in Sec. 3.1, 3.2 and 3.3 through simulation. The generating model for our random graphs is specified by the parameters (p, q, n). Thus, for each choice of (p, q, n), we sample $M = 10^4$ Semi-Markovian graphs at random, and sample the values of I_X, I_Y as described in Sec. 3 (i.e., $I_Y = n$, $I_X \in [n-1]$ uniformly). For each randomly sampled graph and values of I_X, I_Y, we check whether the causal effect of X on Y is back-door identifiable, and whether it is identifiable with the general ID algorithm. Let $(\mathcal{G}_m, I_{X_m}, I_{Y_m})_{m=1}^M$ denote the samples. Our empirical estimators are then given by

$$\widehat{P}_{p,q,n}(\text{BD ID}) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}(Q(I_{X_m}, I_{Y_m}) \vdash_{id}^{bd} \mathcal{G}_i), \tag{45}$$

$$\widehat{P}_{p,q,n}(\text{Joint ID}) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}(Q(I_{X_m}, I_{Y_m}) \vdash_{id}^{jid} \mathcal{G}_i), \tag{46}$$

$$\widehat{P}_{p,q,n}(1^{st} \text{ OH ID}) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}(Q(I_{X_m}, I_{Y_m}) \vdash_{id}^{1oh} \mathcal{G}_i). \tag{47}$$

Notice that the empirical estimators are averages of independent Bernoulli samples, so we can obtain confidence intervals for the true values of $P(\text{BD ID}), P(\text{Joint ID})$, and $P(1^{st}$ OH ID) using a Gaussian approximation. The exact true values can be computed via the dynamic programs introduced in Sec. 3.1, 3.2, and 3.3 and are labeled $P_{p,q,n}(\mathcal{A}$ ID) for identification strategy $\mathcal{A}$. In Fig. 9, we plot the 99% confidence intervals obtained through simulation, and also the exact values computed by the dynamic programs, over a range of values of (p, q, n), thereby verifying the correctness of our theoretical approach. Furthermore, we also test the correctness of our approach formally. For each combination of (p, q, n), and a strategy $\mathcal{A}$, we know that

$$\frac{\widehat{P}_{p,q,n}(\mathcal{A} \text{ ID}) - P_{p,q,n}(\mathcal{A} \text{ ID})}{\sqrt{P_{p,q,n}(\mathcal{A} \text{ ID})(1 - P_{p,q,n}(\mathcal{A} \text{ ID}))\frac{1}{M}}} \sim N(0, 1), \tag{48}$$

which allows us to compute a p-value associated with each pair $\widehat{P}_{p,q,n}(\mathcal{A}$ ID$), P_{p,q,n}(\mathcal{A}$ ID), label $v_{p,q,n}$. Under the null hypothesis that $\widehat{P}_{p,q,n}(\mathcal{A}$ ID) is an average of Bernoulli variables with a parameter $P_{p,q,n}(\mathcal{A}$ ID), the p-values $v_{p,q,n}$ should be uniformly distributed. Using Fisher's method of aggregating p-values (Fisher, 1928), we know that $-2\sum_{p,q,n} \log v_{p,q,n} \sim \chi^2_{2T}$, where T is the number of hypothesis tests performed. This allows us to compute an overall p-value for whether our dynamic programs are correct, and this value equals 0.11 meaning the null-hypothesis is not rejected.

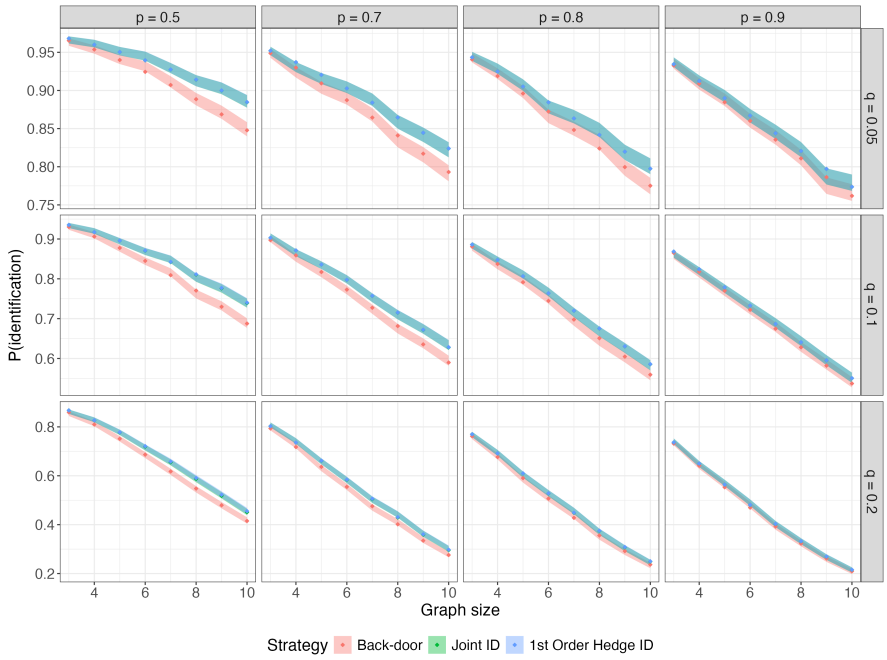


Figure 9: Empirical verification of the theoretical results from Sec. 3. The shaded area represents the 99% confidence intervals for the true value obtained from 10^4 samples, whereas the points are the true values obtained from the dynamic programming approach from Sec. 3. Values of $P(\text{Joint ID})$ and $P(1^{st}\text{ OH ID})$ overlap.

4. Further Models

In the previous section, we successfully computed the exact probabilities of identification for the combination of $\overrightarrow{G}(n,p)$ and $G(n,q)$ models. However, there are also other random graph models that are possibly interesting to investigate. In this section, we extend the empirical analysis of Sec. 3.4 to cover some other models. In particular, we investigate three different ways of generating the observed structure, and two ways of generating the latent, unobserved structure. We consider each combination of generating the observed and unobserved structure, resulting in six different random graph models from which the Semi-Markovian graphs are sampled. In particular, for the observed structure, we consider the following three models:

(A) **directed Erdős–Rényi model**, labeled $\overrightarrow{G}(n,p)$, already introduced in Sec. 3,

(B) **Uniform DAG model**, in which the graph is chosen uniformly across the space of all DAGs on n vertices, labeled $\text{UG}(n)$,

(C) **Price model** (Price, 1965), which generates scale-free random graphs on n vertices using linear preferential attachment (Albert and Barabási, 2002), parameterized by the value m corresponding to the number of edges added in each step. For constructing the graphs, we initialize the procedure with the fully connected graph on $2m+1$ nodes, as this ensures the average out-degree at each step is m (Evans et al., 2020). This model is labeled $\text{PM}(n,m)$.

After specifying the models determining the observed structure, we consider two ways of generating the unobserved structure:

Confounding / Adjacency	$G(n,q)$	IP(q)
$\overrightarrow{G}(n,p)$	$\mathcal{H}_M = \{p \in [0,1], q \in [0,1]\}$	$\mathcal{H}_M = \{p \in [0,1], q \in [0,1]\}$
UG(n)	$\mathcal{H}_M = \{q \in [0,1]\}$	$\mathcal{H}_M = \{q \in [0,1]\}$
PM(n,m)	$\mathcal{H}_M = \{m \leq 2n+1 \in \mathbb{N}, q \in [0,1]\}$	$\mathcal{H}_M = \{m \leq 2n+1 \in \mathbb{N}, q \in [0,1]\}$

Table 2: Overview of models M and the associated hyperparameter ranges $\mathcal{H}_M$.

(1) **Erdős–Rényi model**, labeled $G(n,q)$, (see Sec. 3 for more details),

(2) **Independent Projection model**, labeled IP(q), which takes as an input a DAG on n vertices, and independently with probability q marks each of the variables in the graph as unobserved. The unobserved variables then generate the structure of bidirected edges, following the graphical rules of projecting out variables (see (Tian, 2002) for details on projection rules).

In this section, we adopt an informal notation for the models. For example, we write $\overrightarrow{G}(n,p)+G(n,q)$ to indicate the model M in which a directed $\overrightarrow{G}(n,p)$ model is used for the observed structure, and an undirected $G(n,q)$ for the unobserved structure. Other models are indicated similarly. Furthermore, for each model, we denote by $\mathcal{H}_M$ the set of valid hyperparameters for the model. For example, the model $\overrightarrow{G}(n,p) + G(n,q)$ has hyperparameters p, q, with $\mathcal{H}_M$ being $p \in [0,1], q \in [0,1]$ (for ease of notation, graph size n is not considered as a hyperparameter). Therefore, each choice $h \in \mathcal{H}_M$ is associated with a measure μ_h over the space of Semi-Markovian models $\mathbb{G}_n^{SM}$. [3] Each measure μ_h is further associated with the values of $P_{\mu_h}(Q(I_X, I_Y) \vdash_{id}^{\mathcal{A}} \mathcal{G})$ for different identification strategies $\mathcal{A}$, where P_{μ_h} indicates that $\mathcal{G}$ is sampled according to μ_h.

Several ways of quantifying the gap between the identification strategies exist (visually, we are interested in determining how large of an area is contained between consecutive ellipses in Fig. 7). We focus on two quantification approaches. We start by computing the probabilities of identification over a range of parameters for different models M. Throughout, we focus mostly on smaller graph sizes, and in places run experiments for up to $n = 50$. [4] After probabilities of identification, we discuss the so-called coverage probabilities which tell us the proportion of all identifiable cases that a specific strategy $\mathcal{A}$ can identify.

Pointwise Probabilities of Identification. In our first experiment, the probabilities of identification are computed over a range of parameter values and graph sizes for each model M. We estimate the values of $P(\text{BD ID}), P(\text{ID})$ by sampling 10^4 graphs $\mathcal{G}$ and variables X, Y, for each value on a grid in $\mathcal{H}_M$. A representative set of results for different model choices and graph sizes $5 \leq n \leq 10$ is presented in Fig. 10. We also provide an exhaustive set of experiments in Appendix A. The second column of Tab. 3 provides a summary of figures that report the results on probabilities of identification across different models. Two main observations ensue from these results. Firstly, graphs with a more dense observed and unobserved structure have lower probabilities of identification, for any method. This reflects the well-known fact that a dense structure of edges implies that causal diagrams encode few assumptions. Fewer assumptions, of course, lead to fewer identifiable instances. Secondly, the results also demonstrate that the gap between back-door and complete identification does not seem large in any of the models under investigation. This aspect is explored explicitly in the sequel.

3. We note that when the IP(q) model of confounding is used, the associated measure μ_h is over the space $\cup_{i=0}^{n} \mathbb{G}_i^{SM}$ of all Semi-Markovian graphs of size less than or equal to n.
4. Generally, in this manuscript, we are less interested in the asymptotic behavior as the number of nodes $n \to \infty$.

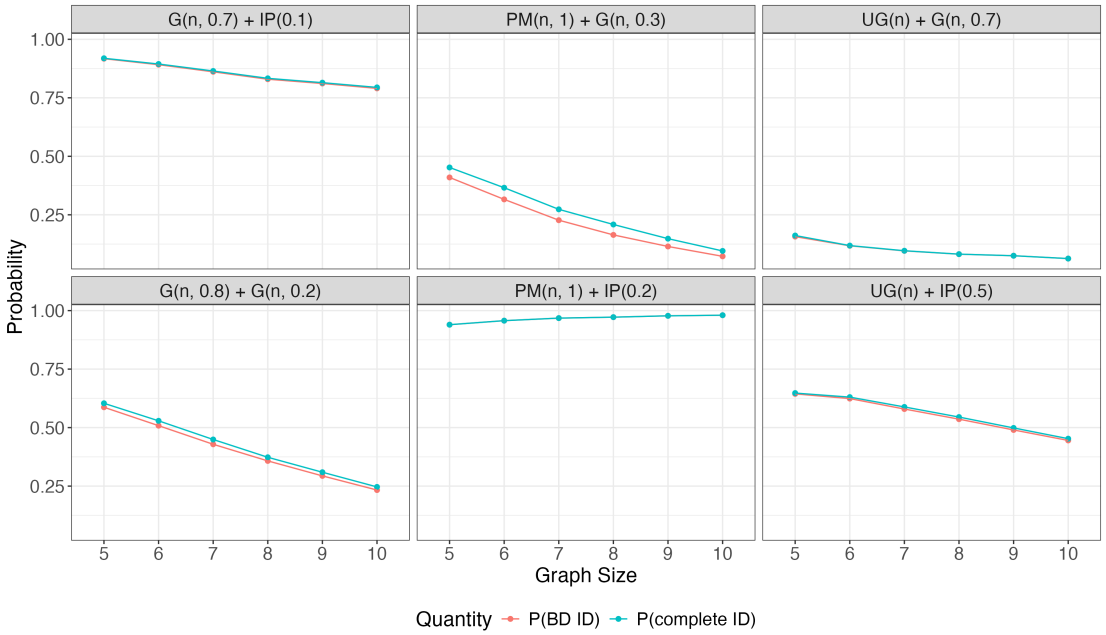


Figure 10: Probabilities of identification over six different graph-generating models for graph size n in the range $5 \leq n \leq 10$.

Model	Probabilities of ID	Coverage Probabilities
$\overrightarrow{G}(n,p) + G(n,q)$	Fig. 18	Fig. 24
$\overrightarrow{G}(n,p) + \text{IP}(q)$	Fig. 19	Fig. 25
$\text{UG}(n) + G(n,q)$	Fig. 20	Fig. 11
$\text{UG}(n) + \text{IP}(q)$	Fig. 21	Fig. 26
$\text{PM}(n,m) + G(n,q)$	Fig. 22	Fig. 27
$\text{PM}(n,m) + \text{IP}(q)$	Fig. 23	Fig. 28

Table 3: Summary of the empirical analyses across models.

Back-door and Lower Bound Coverage. We now introduce a measure for summarizing the difference between identification strategies. The measure we introduce is called strategy coverage, and is defined as follows:

Definition 13 (Strategy Coverage) *Let M be a random model of Semi-Markovian graphs. Then, the strategy coverage is defined as*

$$\mathcal{A}\text{-}coverage_M = \frac{P(\mathcal{A}\ ID)}{P(ID)}. \tag{49}$$

Similarly, the definition of zero-ID-adjusted coverage of a strategy is given by:

$$\mathcal{A}\text{-}coverage^{\text{ZA}}_M = \frac{P(\mathcal{A}\ ID) - P(0\ ID)}{P(ID) - P(0\ ID)}. \tag{50}$$

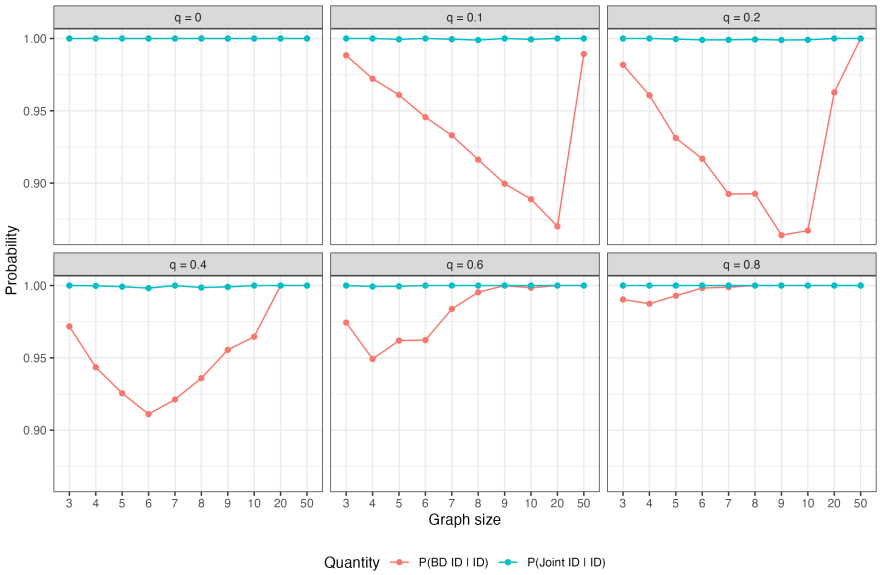


Figure 11: Coverage of Strategies for the UG(n) + $G(n, q)$ model over a range of parameters.

□

The intuition behind the strategy coverage measure is simple. We wish to know what proportion of instances that are identifiable through a complete algorithm will also be identifiable using a simpler strategy. Additionally, we may be interested in removing all instances in which there is no causal path between X and Y since these instances are trivially identifiable (and therefore such graphs are unlikely to be the object of interest in the first place). The third column of Tab. 3 provides a summary of the figures that contain the empirical results. In the main text, we report the results for the model UG(n)+$G(n, q)$, shown in Fig. 11. From the figure, we see that the back-door coverage is always above 70% and often much larger than that. This observation is also true for other models, see Appendix B. Interestingly, the lower bound covers $> 99\%$ of all identifiable cases, uniformly across all models and graph sizes that we explored.

4.1 Expanding to Potential Outcomes Methods

The discussion so far revolved around the graphical notions of identification, such as back-door identification, identification of $P(V \setminus X \mid do(X = x))$, and complete identification. In this section, we try to find further comparisons between the two frameworks of causality. In particular, we focus on two of the most common ways for inferring causal effects in the PO literature: via ignorability assumptions and via instrumental variables.

Ignorability. We mentioned in the introduction that methods of causal effect inference in the PO framework (such as matching, inverse propensity weighting, etc.) can often be used for effect estimation when identification is licensed by the back-door criterion. However, in the PO literature, the back-door criterion is rarely used to license identification. Instead, the criterion known as conditional ignorability is more commonly invoked:

Definition 14 (Ignorability) *Let $P(V)$ be a probability distribution, and denote by Y_x the potential outcome of Y subject to $X = x$. Conditional ignorability is said to hold if:*

$$Y_x \perp\!\!\!\perp X \mid Z. \tag{51}$$

□

The above definition is a well-known building block in the PO literature, and many debates over this definition have ensued between the proponents of the PO framework and those of the graphical framework. However, for brevity we do not expand on this well-known discussion, since this topic deserves a dedicated manuscript (Bareinboim and Plecko, 2024). Instead, we focus on suggesting a mental construct in the graphical framework that is most similar to the notion of ignorability in the PO framework. We begin by taking a deeper look at how ignorability statements are assessed in practice. Firstly, three distinct blocks are evoked in the analysis, replicated below for visual convenience:

$$\underbrace{Y_x}_{\text{block } B_1} \perp\!\!\!\perp \underbrace{X}_{\text{block } B_2} \mid \underbrace{Z}_{\text{block } B_3}.$$

The separation of the ignorability statement in blocks is useful to better understand the mental construct evoked by scientists when judging the plausibility of such statements. The block B_1 considers the variable Y under the hypothetical regime $do(X = x)$, B_2 the treatment variable X itself, and B_3 the set of covariates Z. The key question under investigation is whether block B_1 is independent of block B_2 given the block B_3. Based on this mental picture, we propose the construct called graphical ignorability:

Definition 15 (Graphical Ignorability (Bareinboim and Plecko, 2024)) *Let $\mathcal{M}$ be an SCM over variables V, and suppose a topological order over the variables is fixed. Let* $\text{pre}(X)$ *denote all variables that precede X in the topological ordering, $V_i < X$. Denote* $\text{pre}(X)$ *by Z, and consider the causal diagram constructed in the following way:*

(i) Project out all variables that do not precede X that are different from Y,

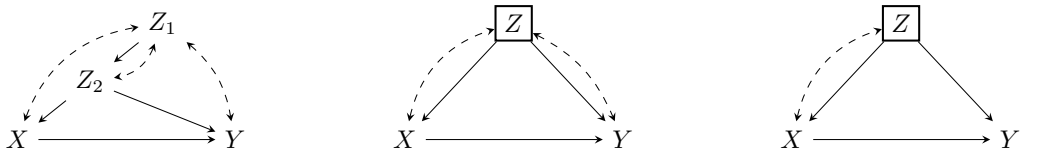
(ii) Group all variables in Z to a single variable.

Let CG_3 be the causal diagram obtained in this way, with nodes Z, X, Y. Graphical ignorability (G-ignorability, for short) is said to hold for a pair (X, Y) if Z is back-door for X, Y in CG_3. □

The graphical ignorability criterion from the above definition attempts to mimic the ignorability construct in a graphical setting. Firstly, all variables that come after the treatment X are projected out (or unobserved). After this, all pre-treatment variables Z are put into a single group.[5] Following this, a causal diagram with three nodes, corresponding to X, Y and Z is constructed. For assessing whether Z is a valid adjustment set, we use the back-door criterion in this simplified graph, called CG_3. It is important to note that the CG_3 diagram may be much easier to construct than the initial causal diagram $\mathcal{G}$ over all variables. In other words, this diagram is more easily attainable than the full diagram, and requires fewer assumptions.

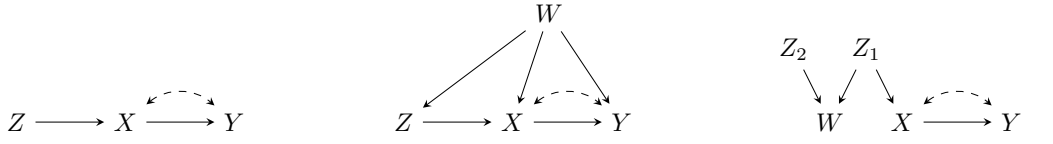
Example 12 (Graphical Ignorability) *Consider the causal diagram $\mathcal{G}$ in Fig. 12a. The CG_3 diagram corresponding to this example is shown in Fig. 12b. For the directed arrows, note that from the set Z there are edges into both X, Y. For the bidirected edges, note that $X \dashleftarrow\dashrightarrow Z_1$ implies an edge $X \dashleftarrow\dashrightarrow Z$ in CG_3. Similarly, $Z_1 \dashleftarrow\dashrightarrow Y$ implies an edge $Z \dashleftarrow\dashrightarrow Y$ in CG_3. Therefore, since the back-door criterion is not satisfied for the graph in Fig. 12b, the pair (X, Y) does not satisfy graphical ignorability given the set Z.*

5. Formally, this construction represents a *cluster* causal diagram. For details, we refer the reader to (Anand et al. 2023).



(a) G-Ignorability example $\mathcal{G}$. (b) CG_3 of $\mathcal{G}$. (c) CG_3 of $\mathcal{G} \setminus \{Z_1 \dashleftarrow\dashrightarrow Y\}$.

Figure 12: Causal diagrams from Ex. 12



(a) Classic IV setting. (b) Conditional IV setting. (c) Non-ancestral IV setting.

Figure 13: Causal diagrams for instrumental variable examples.

Consider now the causal diagram in Fig. 12a with the edge $Z_1 \dashleftarrow\dashrightarrow Y$ removed, denoted by $\mathcal{G} \setminus \{Z_1 \dashleftarrow\dashrightarrow Y\}$. The construction of the CG_3 graph corresponding to this graph remains similar as above, with the difference in the edge $Z \dashleftarrow\dashrightarrow Y$ that is absent, since there is no bidirected edge between $\{Z_1, Z_2\}$ and Y. In this CG_3, the back-door criterion is satisfied, and the pair (X, Y) satisfies graphical ignorability given Z. □

The following proposition, given without proof, justifies the construct of graphical ignorability:

Proposition 5 (G-Ignorability $\implies$ Ignorability) *Let $\mathcal{M}$ be an SCM over variables V. Graphical ignorability of (X, Y) with respect to Z implies conditional ignorability $Y_x \perp\!\!\!\perp X \mid Z$.* □

For the purpose of our discussion, we investigate how G-ignorability compares to the back-door criterion in terms of its coverage (Def. 13), that is, what proportion of identifiable cases can be solved using G-ignorability. A summary of the exhaustive set of results can be found in the third column of Tab. 3. Similar to the back-door coverage, we once again observe that a simple strategy such as G-ignorability has a rather large coverage compared to complete identification across a range of models and their parameter values.

Instrumental Variables. The second strategy for causal effect inference popular in the PO literature is that of instrumental variables (IVs). The usage of IV dates back to (Wright, 1928). The typical IV graph is shown in Fig. 13a. In this graph, the causal effect of X on Y is not identifiable using any of the standard methods (e.g., no back-door adjustment set exists). Under the assumption of linearity, however, one can regress X on Z, and compute the fitted values of this regression, $\hat{X}$. These fitted values $\hat{X}$ are "deconfounded", since as a function of Z, they are not affected by the latent variable U affecting X, Y. In the second stage, one can obtain the correct causal coefficient of X on Y by regressing Y onto $\hat{X}$. This well-known method is referred to in the literature as the two-stage least squares approach.

In our context, we will focus on the conditional version of instrumental variables. Here, we make use of the graphical definition of conditional IV as proposed by (Pearl, 2000):

Definition 16 (Conditional IV (Pearl, 2000)) *Z is said to be a conditional instrument relative to $X \to Y$ if there exists a set W such that:*

(a) Z correlates with X conditional of W,

(b) W d-separates Z and Y in $\mathcal{G} \setminus \{X \to Y\}$,

(c) W consists of non-descendants of Y.

□

Example 13 (Conditional IV) *Consider the example in Fig. 13a. We can verify that Z correlates with X (due to the $Z \to X$ arrow). Further, the empty set $\emptyset$ d-separates Z and Y in $\mathcal{G} \setminus \{X \to Y\}$. Therefore, Z is a conditional IV for $X \to Y$ given $W = \emptyset$.*

Consider the example in Fig. 13b. We can verify that Z correlates with X (due to the $Z \to X$ arrow). Further, the set W d-separates Z and Y in $\mathcal{G} \setminus \{X \to Y\}$. Therefore, Z is a conditional IV for $X \to Y$ given W, which consists of non-descendants of Y. □

Finding conditional IVs turns out to be a non-trivial matter. For this reason, Van der Zander et al. (2015) propose a slightly constrained version of the conditional IV:

Definition 17 (Ancestral Conditional IV (Van der Zander et al., 2015)) *Z is said to be an ancestral instrument relative to $X \to Y$ if there exists a set W such that:*

(a) Z correlates with X conditional of W,

(b) W d-separates Z and Y in $\mathcal{G} \setminus \{X \to Y\}$,

(c) W consists of ancestors of Y or Z which are non-descendants of Y.

□

The usefulness of the ancestral IV concept stems from the following proposition:

Proposition 6 ($\exists$ Conditional IV $\implies$ $\exists$ Ancestral IV (Van der Zander et al., 2015)) *A conditional IV Z for the pair (X, Y) exists if and only if an ancestral IV Z' exists for (X, Y).* □

Example 14 (Ancestral and Non-ancestral Conditional IVs) *Consider the causal diagram in Fig. 13b, in which Z is a conditional IV for (X, Y) given W. Since $W \in \text{an}(Z)$ and $W \in \text{an}(Y)$, Z is an ancestral IV with respect to the pair (X, Y) and the conditioning set W.*

Consider the causal diagram in Fig. 13c. Here, Z_2 correlates with X conditional on W. Furthermore, W d-separates Z_2 from Y in the graph where $X \to Y$ edge is removed. W is also a non-descendant of Y, and thus Z_2 is a conditional IV for (X, Y) given W. However, $W \in \text{de}(Z_2)$ and therefore Z_2 is not an ancestral IV for (X, Y) given W. According to Prop. 6, the existence of a conditional IV implies the existence of an ancestral IV. One can easily verify that Z_1 given $\emptyset$ is an IV for (X, Y), and that this IV is also ancestral. □

Prop. 6 shows that our attention can be turned to searching for ancestral IVs, since one is guaranteed to exist whenever a conditional IV exists. Van der Zander et al. (2015) also introduce an efficient algorithm for finding ancestral IVs, which we implement. This allows us to compute the probabilities of identification for the conditional IV strategy.

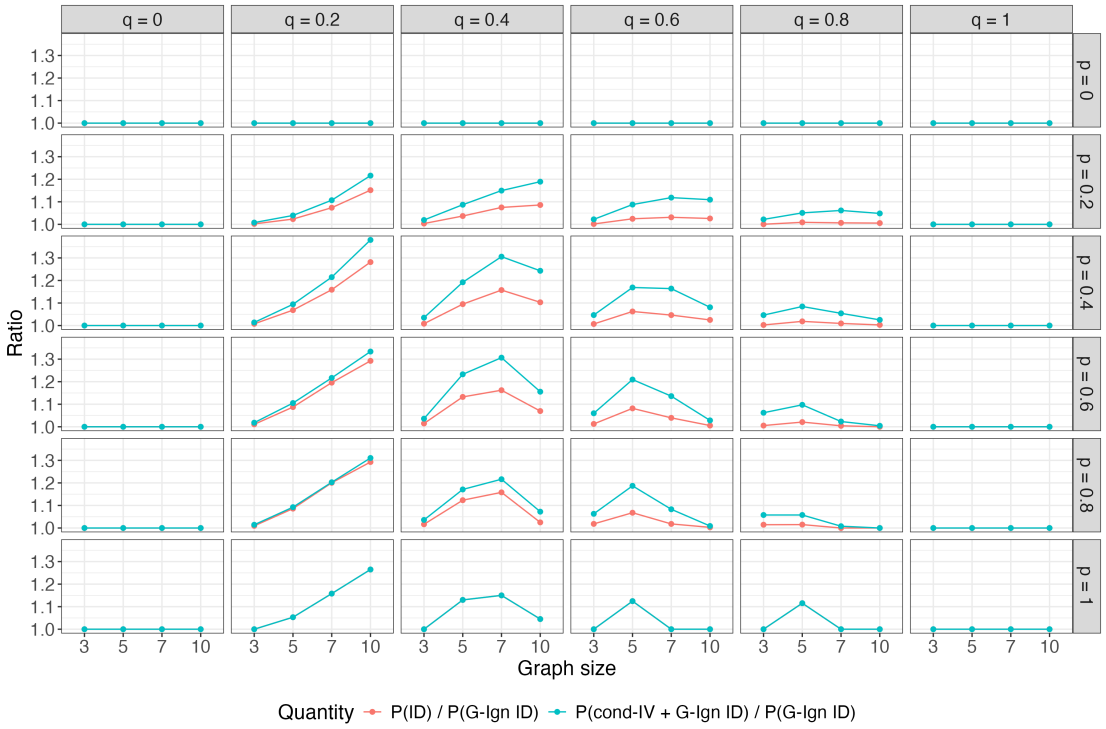


Figure 14: Comparison on how useful complete and conditional IV identification strategies are compared to the basic ignorability strategy in the $\overrightarrow{G}(n,p) + G(n,q)$ model.

Extension Ratios. Using probabilities of conditional IV identification, we can perform the following experiment. We first compute, across different parameter values of a graph-generating model, and different graph sizes, how much the complete identification strategy extends the number of identifiable cases compared to the basic graphical ignorability criterion from Def. 15. The quantity we compute is given by:

$$\frac{P(\text{ID})}{P(\text{G-Ign ID})}. \tag{52}$$

In words, this quantity tells us how many more cases (in terms of proportion) an analyst using complete identification can solve compared to an analyst using only the ignorability criterion. Crucially, we can compare this quantity to the quantity:

$$\frac{P(\text{cond-IV + G-Ign ID})}{P(\text{G-Ign ID})}. \tag{53}$$

Similarly, this quantity tells us how much the conditional IV approach improves upon the basic ignorability strategy. The comparison of these two quantities for the $\overrightarrow{G}(n,p) + G(n,q)$ model is shown in Fig. 14. Interestingly, the conditional IV criterion seems to provide a somewhat larger improvement than complete identification, when compared to the ignorability criterion. However, the improvements resulting from conditional IV and complete identification are comparable.

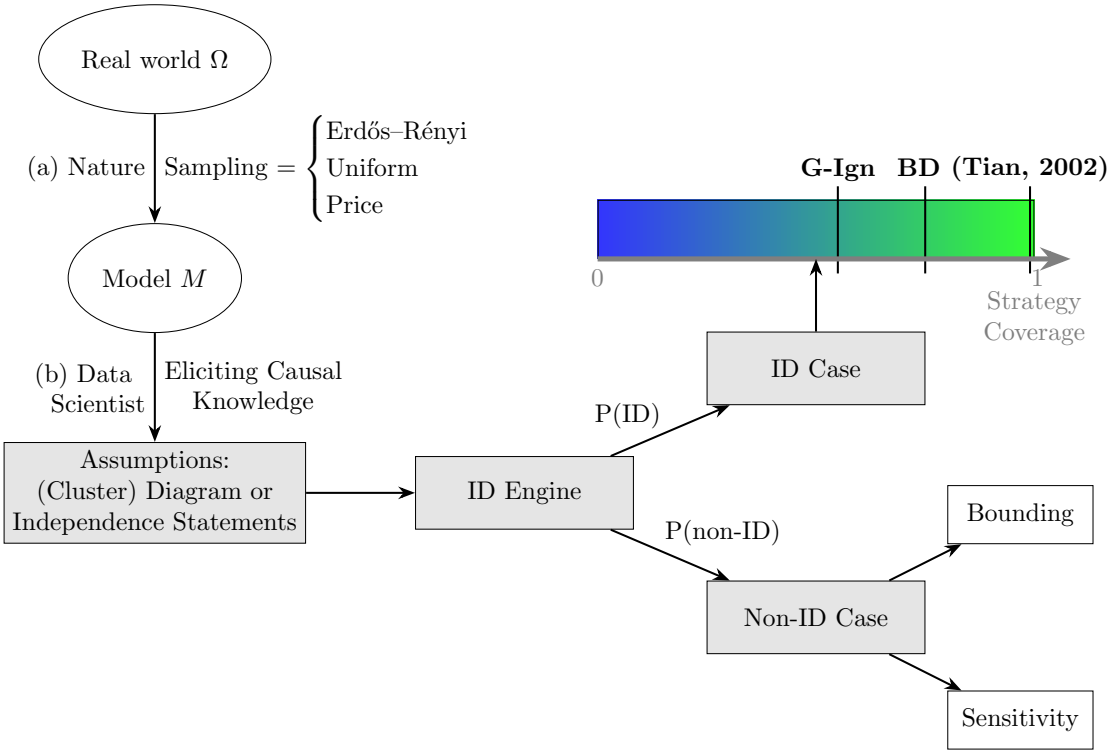


Figure 15: Schematic representation of the discussion (Sec. 5).

5. Discussion & Conclusion

In this manuscript, we investigate the question of identifying causal effects through the lens of probabilistic modeling. A schematic overview of our approach is shown in Fig. 15, and the figure provides the structure for the discussion that follows. The space of all possible causal models is labeled Ω, corresponding to all causal models that exist in the real world. In practice, when one formulates a research question and focuses on a certain aspect of reality, Nature can be thought of as "sampling" from the space Ω and producing an instance of a model, labeled M, which the scientist seeks to understand by eliciting assumptions and drawing inferences. While the underlying sampling model from Ω to specific instances M is unknown (that is, it is not know a priori how Nature generates models), in this work, we consider different sampling schemes that randomly generate a model represented through a Semi-Markovian causal diagram. For the directed structure of the causal diagram, we considered (a) the directed Erdős–Rényi model, (b) the uniform model across DAGs, and (c) the scale-free Price model. For generating the unobserved structure in a causal diagram, we considered (1) an undirected Erdős–Rényi model that generates bidirected edges, and (2) a projection model in which the unobserved structure is generated by "unobserving" or projecting out variables from an initial directed graph. Of course, these are just some possible sampling strategies, and we do not know if Nature samples the models of reality in some other, specifically fine-tuned way. [6] For each of these graph-generating models (six combinations in total), we considered whether

6. There is ample discussion in the physical sciences about fine-tuned constants of the universe, such as the gravitational constant, Planck's constant, the fine-structure constant, and the cosmological constant, whose precise values appear to be necessary for the emergence of complex structures, from atomic stability to galaxy formation (Barrow and Tipler, 1988; Carr et al., 1979). Planck's constant, for example, determines the scale of quantum effects, setting fundamental limits on precision and influencing everything from atomic energy levels to the struc-

the interventional distribution $P(Y \mid do(X = x))$ (query) was identifiable in a randomly sampled graph $\mathcal{G}$, with the singleton indices I_X, I_Y of X, Y sampled uniformly at random from the set of all nodes. In practice, after specifying the query under investigation, scientists aim to elicit causal knowledge in order to determine whether the query can be identified (that is, uniquely computed from the assumptions and the available data).

As the output of the identification decision problem, a binary outcome is obtained – a positive or a negative answer, depending on the identification strategy employed. In the discussion that follows, we analyze these two cases separately. The first case, which we call the *Identifiable Case*, occurs when the effect is identifiable, meaning that one of the complete identification algorithms will be able to decide favorably when the causal diagram is provided. This case occurs with a probability $P(\text{ID})$, which depends on the underlying sampling process. The complement of this case is the *Non-Identifiable Case*, in which a complete identification algorithm returns a negative answer (meaning that there are two models that could equally well explain the observed data, are compatible with the specified causal diagram, but have different estimates of the effect measure). This case occurs with a probability $P(\text{non-ID}) = 1 - P(\text{ID})$. In the discussion that follows, we first discuss the values of the $P(\text{ID})$ probability for different sampling processes. Then, we discuss the implications of our results for the ID Case, in particular focusing on how the probability of complete identification $P(\text{ID})$ compares to simpler strategies such as graphical ignorability (G-Ign, for short) or the back-door (BD) criterion. Finally, we also discuss the implications of our results for the Non-ID Case.

On the Probabilities of Identification. We start with the probabilities of identification, which represent the proportion of causal effect queries that would be identifiable, assuming a probabilistic sampling model. The second column of Tab.3 provides a summary of figures reporting the results on probabilities of identification across different models and graph sizes (see Figs. 18-23). Our results confirm some well-known facts and introduce new ones to the literature. First, note that causal assumptions are encoded in the absence of edges, not in their presence. Any edge could encode any effect, including zero, while a missing edge guarantees that the effect is zero. Therefore, fewer edges mean that a causal diagram implies stronger assumptions, which, in turn, increases the probability of identification. This trend is evident in the empirical results. For instance, as the parameter p, which determines the probability of adding a directed edge in a $\overrightarrow{G}(n, p)$ model, the probability of identification decreases monotonically (Fig. 18). Similarly, as the q parameter, which governs the probability of sampling a bidirected edge in a $G(n, q)$ model, increases, the probability of identification decreases monotonically (Fig. 18). Therefore, in graph-generating settings where the observed and unobserved structures are sparse, the probabilities of identification are relatively high; conversely, in settings where the observed and unobserved structures are dense, the probabilities of identification are significantly lower. Naturally, our ability to identify causal effects depends on the strength of the assumptions available in our causal models and the proportion of relevant variables recorded in the data. A larger discrepancy between our understanding of the underlying structure and what is available in the data (reflected in a larger number of unobserved variables) reduces the likelihood of identification. For instance, if only the treatment and outcome variables X, Y are observed, it is unlikely that the effect would be identifiable. Conversely, if all relevant variables are observed – making the system essentially Markovian – effects are always identifiable. These observations can be seen as a special case of the Causal Hierarchy Theorem Bareinboim et al. (2022) [Thm. 1], which states that the true underlying collection of mechanisms can almost never (i.e., with measure zero) be recovered from observational or interventional data alone, which includes causal effects.

ture of matter itself. This fine-tuning has been widely explored in the context of fundamental physics, including quantum mechanics, general relativity, and the Standard Model of particle physics, raising deep questions about the underlying principles that govern these constants (Weinberg 1987 Susskind 2005). This may be interpreted as Nature not merely sampling randomly but rather following an underlying structure that governs how reality emerges. In our discussion, we are oblivious to this fact, and pursue a purely probabilistic argument.

5.1 The Identifiable Case

On the Ignorability & Back-door Coverage. We now discuss the so-called coverage probabilities (Def. 13) of different identification strategies. Coverage is defined as the proportion of all cases identifiable by a strategy $\mathcal{A}$ among all identifiable cases. A back-door coverage of 90% indicates that when selecting queries and graphs at random, and finding an identifiable query, 90% of such cases will be identifiable using the back-door adjustment. A graphical ignorability coverage of 90% has an analogous interpretation. The third column of Tab. 3 provides a summary of figures that report the results on coverage probabilities across different models and graph sizes (see Figs. 11, 24-28). These figures highlight a key finding: the probabilities of back-door and conditional ignorability coverage exceed 70% across all considered models and sometimes surpass 90%, given our samplig scheme. This suggests that when identification is possible, it is likely achievable through back-door or graphical ignorability type of adjustment. In other words, if a query is identifiable, it is often also identifiable using simple strategies. Additionally, since some of these strategies do not require a full causal diagram, constructing relaxed causal diagrams, such as cluster diagrams (Anand et al., 2023; Bareinboim and Plecko, 2024), may serve as a powerful practical tool. However, these findings need to be interpreted with care, since the probability $P(\text{ID})$ can be rather low in some settings. In such cases, graphical ignorability or back-door adjustment would also fail (see Sec. 5.2 for a more detailed discussion). In other words, even though the backdoor may cover 90% of the identifiable cases in a particular setting, perhaps only 10% of the overall cases are identifiable, implying that the backdoor is still unlikely to work. This implies that a more comprehensive approach to evaluating identification is still needed.

On the Gap Between Joint and Complete Identification. Another interesting aspect of our analysis is the comparison of the joint identification strategy (lower bound from Sec. 3.2) with a complete identification strategy. The joint identification strategy starts by identifying $P(V \setminus X \mid do(X = x))$ in the ancestral graph $\mathcal{G}[an(Y)]$, a seemingly more demanding task than identifying the marginal distribution $P(Y \mid do(X = x))$. This strategy also provides a much simpler graphical criterion for determining identification (Prop. 3) compared to the complete criterion (Prop. 4). This is also reflected in simpler identification expressions. In fact, the observational distribution can be factorized into the so-called C-factors (corresponding to C-components from Def. 9), and can be written as

$$P(v) = \prod Q_i, \tag{54}$$

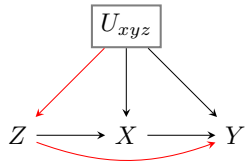
where each Q_i is a C-factor corresponding to a C-component (Tian and Pearl, 2002). If the distribution $P(v \mid do(x))$ can be recovered by this strategy, then its expression is given by

$$P(v \mid do(x)) = \frac{P(v)}{Q^X} \sum_x Q^X. \tag{55}$$

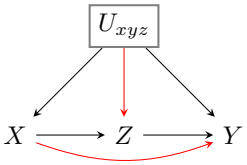
The final expression for $P(y \mid do(x))$ can be obtained by marginalizing out all variables $V \setminus \{X, Y\}$:

$$P(y \mid do(x)) = \sum_{v \setminus \{x,y\}} P(v \mid do(x)). \tag{56}$$

Interestingly, from a theoretical viewpoint, the gap between recovering $P(v \mid do(x))$ and $P(y \mid do(x))$ may seem substantive. However, the coverage of the joint identification strategy is extremely high across experiments with different random models, as we did not find any models where it was less than 99%. This may suggest that the practical gap between the lower bound and the complete identification strategy may be rather small. Although the theoretical gap between the lower bound strategy and complete identification exists (see, e.g., the discussion around the napkin graph in Ex. 9), based on the sampling models considered in this manuscript, the lower bound strategy could be described as "nearly" complete from a probabilistic viewpoint.



(a) IV setting with U_{xyz}.



(b) Front-door setting with U_{xyz}.

Figure 16: Instrumental variable (IV) and the front-door settings.

On the Usefulness of Techniques Beyond the Back-door. Coverage probabilities offer a possible way to assess the benefit of using the complete identification approach compared to simpler identification strategies. When taking a probabilistic perspective on this question, graphical ignorability coverage exceeded 70% across settings, given our sampling model. Therefore, a natural question remains: how much practical advantage does the full set of tools in the graphical approach to causality offer compared to a more modest strategy, such as graphical ignorability?

There are different techniques that can be applied beyond the back-door approach. One of these methods is known as the front-door identification method (Pearl, 2000), which is just the first setting learned after the back-door, and technique within the larger class of methods covered by do-calculus (Pearl, 1995). Since these methods are still relatively unknown to a large portion of researchers, there is a noticeable absence of applied papers on the topic (with some interesting exceptions such as (Glynn and Kashin, 2018; Piccininni et al., 2023; Inoue et al., 2022)). One important consideration is that they require a causal model to be explicitly articulated, whereas the current practice in many applied fields often involves articulating back-door/ignorability assumptions without a proper model. If one technique is perceived as requiring additional modeling effort while the weaknesses of alternative methods remain unknown, it is natural that the simpler approach will prevail.

On the Comparison with Instrumental Variables. One point that corroborates this argument – low awareness about the method implies less reported use cases – is the pervasiveness of a technique known as instrumental variables (IVs). We now compare the IV methodology with the complete identification algorithms, first through a probabilistic lens, and then theoretically.

In Sec. 4.1, we investigated how much IV methods add to identification, compared to only using the graphical ignorability criterion. For this, we used the extension ratio in Eq. 53. For instance, an extension ratio of 1.3 tells us that 30% more of cases become identifiable when adding the conditional IV strategy to the standard graphical ignorability criterion. For comparison, we also computed the extension ratio comparing complete identification with graphical ignorability. These extension ratios, for the setting where both directed and undirected edges are sampled from an Erdős–Rényi model, are summarized in Fig. 14. Perhaps surprisingly, Fig. 14 illustrates that the proportion of extra cases identified by IVs is somewhat similar to the proportion of extra cases gained using complete identification methods. From the viewpoint of probabilistic modeling used in this paper, one would perhaps expect the two approaches to be equally useful.

Another possible way to compare methods is through the lens of the assumptions that are required. The IV method requires assumptions about missing edges, similar to methods for complete identification. Consider the representation of the IV setting shown in Fig. 16a, where the variable U_{xyz} is unobserved. The required edge absences for the method are highlighted in red. The exclusion restriction is represented by the absence of the edge $Z \to Y$, which in counterfactual notation can be written as $Y_{xz} = Y_x$. Furthermore, the exogeneity of the instrument is represented through the missing edge $U_{xyz} \to Z$, which implies the counterfactual independence $Y_z \perp\!\!\!\perp Z$. Assuming linearity, these two conditions imply that the causal effect of X on Y can be identified from observational data. It is interesting to draw a comparison with the front-door setting, represented in Fig. 16b. Analogous to the IV setting, the edge $X \to Y$ must be absent, which implies the exclusion constraint

$Y_{xz} = Y_z$. Further, the absence of the $U_{xyz} \to Z$ edge is required, which implies the counterfactual independencies $Z_x \perp\!\!\!\perp X$ and $Y_z \perp\!\!\!\perp Z_x$. These two edge absences imply that the causal effect of X on Y is identifiable from observational data (in Appendix A, we provide a more detailed comparison of the IV and front-door settings). Therefore, when taking this perspective, the IV and front-door settings seem to require equally stringent assumptions from a theoretical viewpoint.

However, at the same time, the IV approach has been widely adopted and shown useful in many practical applications, such as (Angrist, 1990; Card, 2001; Acemoglu et al., 2015; Levitt, 1996) (to name a few), while front-door identification is still largely missing from the applied literature. This observation suggests that more detailed identification strategies could be useful once practitioners become more familiar with such methods and attempt to apply them in practice. Furthermore, the observation may also suggest that the sampling of causal models from the underlying reality happens in specifically fine-tuned ways (see also footnote 6). Importantly, however, we remark that the results discussed so far do not suggest that a graphical model is unnecessary for identifying causal effects, as we discuss in the next section.

5.2 The Non-Identifiable Case

The results from the first part of our discussion on the *ID Case* indicate that whenever an identification expression can be derived, it can likely also be obtained using a simple identification strategy such as the back-door criterion or graphical ignorability. The latter approach is characterized by conditioning on the set of all pre-treatment covariates. However, our results do not indicate that conditioning on pre-treatment covariates always provides the correct answer for identification, as the probability of non-identification P(non-ID) may be high. For the same reason, our results also do not imply that a graphical model – an encoder of structural assumptions – can be sidestepped when assessing identification. In particular, whenever the query is non-identifiable (which may occur with a large probability depending on the context), applying pre-treatment conditioning may lead to misleading results. We next discuss the implications of our findings for the *Non-ID Case*.

Need for systematic assessment of identifiability. The probability of identification depends on the strength of the available assumptions, or the density of edges in the underlying causal diagram. Our results show that, in some contexts, the P(ID) may be rather small, and consequently, the P(non-ID) may be quite large. Therefore, it is important to develop a language that systematically allows one to assess whether a query of interest is identifiable. In this work, we discussed several strategies for identification, including the back-door criterion, complete ID algorithms, and graphical ignorability. All of these strategies are *model-based*, meaning that the evaluation of identifiability is based on the causal diagram. Interestingly, the traditional approach in the potential outcomes (PO) literature aims to assess ignorability statements using a model-free approach, in absence of a graphical model. Recent works, however, show that assessing ignorability in a model-free way is far more challenging than commonly assumed (Bareinboim and Plecko, 2024). In light of this, and following (Bareinboim and Plecko, 2024), instead of considering model-free ignorability statements, we discussed the concept of graphical ignorability in Sec. 4.1, which provides a graph-based criterion for verifying if the set of all pre-treatment covariates provides a valid adjustment set. While G-Ign is, in spirit, related to the ignorability notion of the PO framework, it is nonetheless a model-based approach and should therefore not be equated with traditional usage of ignorability. In summary, while the ignorability notion suggests a useful strategy of pre-treatment conditioning (which is likely to work if a query is identifiable), its operationalization in the PO framework may not be the best tool for determining when pre-treatment conditioning actually fails.

Can we always condition on pre-treatment covariates? An important question to consider is whether simply conditioning on all the pre-treatment covariates is sufficient for causal effect adjustment in general. Counterfactually speaking, if this were the case, it would also imply that observational studies can replace experimental studies, such as randomized controlled trials. Our

results, however, indicate that across models the probabilities of identification vary substantially, and in models with dense directed and confounded structure, none of the identification strategies are likely to work (that is, no strategy provides a positive identification answer). In such instances, conditioning on all pre-treatment variables would lead to biased estimates, and possibly erroneous conclusions. This once again emphasizes the need for a proper language for assessing identification. In such non-identifiable settings, one must resort either to partial identification techniques (also known as bounding), or to the usage of external data (such as data from different environments, or experimental studies) (Bareinboim and Pearl, 2016). The fact that the probability of non-identification P(non-ID) can be high, together with the observation that coverage probabilities of simpler strategies are often high, emphasizes the importance of research in sensitivity analysis (VanderWeele and Ding, 2017; Cinelli and Hazlett, 2020; Cinelli et al., 2019) and partial identification (Zhang et al., 2022; Zhang and Bareinboim, 2021). Such methods may help address settings where assumptions do not license point identification of the causal query of interest, and allow researchers to perform partial inference. The discussion in this paper is based on the task of identifying an interventional distribution from observational data, which is a prototypical and simplest type of causal inference. A more realistic and similar discussion could be considered for various other settings, including when considering identification from a combination of observational and interventional distributions (Bareinboim and Pearl, 2012b; Lee et al., 2020b; Correa et al., 2019a), non-atomic interventions (Correa and Bareinboim, 2020), counterfactual distributions (Correa et al., 2021), selection bias (Bareinboim and Pearl, 2012a; Correa et al., 2018, 2019b), to cite a few.

Acknowledgments. We thank Judea Pearl for his thoughtful comments. Drago Plecko and Elias Bareinboim were supported in part by the NSF, ONR, AFOSR, DoE, Amazon, JP Morgan, and the Alfred P. Sloan Foundation. Matija Bucic was supported in part by the Oswald Veblen Fund and the NSF grant 2349013. Domagoj Bradac was supported in part by SNSF grant 200021_196965.

References

Daron Acemoglu, Camilo García-Jimeno, and James A Robinson. State capacity and economic development: A network approach. American economic review, 105(8):2364–2409, 2015.

Sina Akbari, Fateme Jamshidi, Ehsan Mokhtarian, Matthew J Vowels, Jalal Etesami, and Negar Kiyavash. Causal discovery in probabilistic networks with an identifiable causal effect. arXiv preprint arXiv:2208.04627, 2022.

Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. Reviews of modern physics, 74(1):47, 2002.

Tara V Anand, Adele H Ribeiro, Jin Tian, and Elias Bareinboim. Causal effect identification in cluster dags. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, pages 12172–12179, 2023.

Joshua D Angrist. Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records. The american economic review, pages 313–336, 1990.

Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. Biometrics, 61(4):962–973, 2005.

E. Bareinboim and J. Pearl. Controlling selection bias in causal inference. In N. D. Lawrence and M. Girolami, editors, Proceedings of the 15th International Conference on Artificial Intelligence and Statistics, volume 22 of Proceedings of Machine Learning Research, pages 100–108, La Palma, Canary Islands, Apr 2012a. PMLR.

E. Bareinboim and J. Pearl. Causal inference and the data-fusion problem. In R. M. Shiffrin, editor, Proceedings of the National Academy of Sciences, volume 113, pages 7345–7352. National Academy of Sciences, 2016.

Elias Bareinboim and Judea Pearl. Causal inference by surrogate experiments: z-identifiability. In Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence, UAI'12, page 113–120, Arlington, Virginia, USA, 2012b. AUAI Press. ISBN 9780974903989.

Elias Bareinboim and Drago Plecko. On the basis of structural ignorability. arXiv preprint arXiv:XYZ, 2024.

Elias Bareinboim, Juan D. Correa, Duligur Ibeling, and Thomas Icard. On pearl's hierarchy and the foundations of causal inference. In Probabilistic and Causal Inference: The Works of Judea Pearl, page 507–556. Association for Computing Machinery, New York, NY, USA, 1st edition, 2022.

J.D. Barrow and F.J. Tipler. The Anthropic Cosmological Principle. ISSR library. Oxford University Press, 1988. ISBN 9780192821478. URL https://books.google.com/books?id=uSykSbXklWEC.

Béla Bollobás, Christian Borgs, Jennifer T Chayes, and Oliver Riordan. Directed scale-free graphs. In SODA, volume 3, pages 132–139, 2003.

David Card. Estimating the return to schooling: Progress on some persistent econometric problems. Econometrica, 69(5):1127–1160, 2001.

Bernard J. Carr, Bernard J. Carr, and Martin J. Rees. The anthropic principle and the structure of the physical world. Nature, 278:605–612, 1979. URL https://api.semanticscholar.org/CorpusID:4363262.

Carlos Cinelli and Chad Hazlett. Making sense of sensitivity: Extending omitted variable bias. Journal of the Royal Statistical Society Series B: Statistical Methodology, 82(1):39–67, 2020.

Carlos Cinelli, Daniel Kumor, Bryant Chen, Judea Pearl, and Elias Bareinboim. Sensitivity analysis of linear structural causal models. In International conference on machine learning, pages 1252–1261. PMLR, 2019.

J. Correa and E. Bareinboim. A calculus for stochastic interventions: Causal effect identification and surrogate experiments. In Proceedings of the 34th AAAI Conference on Artificial Intelligence, New York, NY, 2020. AAAI Press.

J. Correa, J. Tian, and E. Bareinboim. Generalized adjustment under confounding and selection biases. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence, pages 6335–6342, New Orleans, LA, 2018. AAAI Press.

J. Correa, J. Tian, and E. Bareinboim. Adjustment criteria for generalizing experimental findings. In K. Chaudhuri and R. Salakhutdinov, editors, Proceedings of the 36th International Conference on Machine Learning, volume 97, pages 1361–1369, Long Beach, CA, 2019a. PMLR.

J. Correa, J. Tian, and E. Bareinboim. Identification of causal effects in the presence of selection bias. In Proceedings of the 33rd AAAI Conference on Artificial Intelligence, pages 2744–2751, Honolulu, Hawaii, 2019b. AAAI Press.

Juan Correa, Sanghack Lee, and Elias Bareinboim. Nested counterfactual identification from arbitrary surrogate experiments. Advances in Neural Information Processing Systems, 34:6856–6867, 2021.

Juan D Correa and Elias Bareinboim. Counterfactual Graphical Models: Constraints and Inference. In Proceedings of the 42nd International Conference on Machine Learning, 2025.

David R Cox, Nanny Wermuth, AP Dawid, SE Fienberg, C Glymour, P Spirtes, D Freedman, GW Imbens, DB Rubin, JM Robins, et al. Discussion - causal diagrams for empirical research. Biometrika, 82(4):669–710, 1995.

Tim S Evans, Lucille Calmon, and Vaiva Vasiliauskaite. The longest path in the price model. Scientific reports, 10(1):10503, 2020.

Ronald Aylmer Fisher. Statistical methods for research workers. Oliver and Boyd, 1928.

Ronald Aylmer Fisher. Design of experiments. British Medical Journal, 1(3923):554, 1936.

Edgar N Gilbert. Random graphs. The Annals of Mathematical Statistics, 30(4):1141–1144, 1959.

Adam N Glynn and Konstantin Kashin. Front-door versus back-door adjustment with unmeasured confounding: Bias formulas for front-door and hybrid adjustments with application to a job training program. Journal of the American Statistical Association, 113(523):1040–1049, 2018.

Trygve Haavelmo. The probability approach in econometrics. Econometrica: Journal of the Econometric Society, pages iii–115, 1944.

Yimin Huang and Marco Valtorta. Pearl's calculus of intervention is complete. arXiv preprint arXiv:1206.6831, 2012.

Guido W Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. Review of Economics and statistics, 86(1):4–29, 2004.

Guido W Imbens. Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. Journal of Economic Literature, 58(4):1129–1179, 2020.

Kosuke Inoue, Beate Ritz, and Onyebuchi A Arah. Causal effect of chronic pain on mortality through opioid prescriptions: Application of the front-door formula. Epidemiology, 33(4):572–580, 2022.

Dmitriy Katz, Karthikeyan Shanmugam, Chandler Squires, and Caroline Uhler. Size of interventional markov equivalence classes in random dag models. In The 22nd International Conference on Artificial Intelligence and Statistics, pages 3234–3243. PMLR, 2019.

Daphne Koller and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.

Sanghack Lee, Juan D Correa, and Elias Bareinboim. General identifiability with arbitrary surrogate experiments. In Uncertainty in artificial intelligence, pages 389–398. PMLR, 2020a.

Sanghack Lee, Juan D Correa, and Elias Bareinboim. Identifiability from a combination of observations and experiments. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 13677–13680, 2020b.

Steven D Levitt. The effect of prison population size on crime rates: Evidence from prison overcrowding litigation. The quarterly journal of economics, 111(2):319–351, 1996.

Jerzy Neyman. On the application of probability theory to agricultural experiments. essay on principles. Ann. Agricultural Sciences, pages 1–51, 1923.

Judea Pearl. Causal diagrams for empirical research. Biometrika, 82(4):669–688, 1995.

Judea Pearl. Causality: Models, Reasoning, and Inference. Cambridge University Press, New York, 2000. 2nd edition, 2009.

Judea Pearl and Dana Mackenzie. The book of why: the new science of cause and effect. Basic books, 2018.

Marco Piccininni, Tobias Kurth, Heinrich J Audebert, and Jessica L Rohmann. The effect of mobile stroke unit care on functional outcomes: an application of the front-door formula. Epidemiology, 34(5):712–720, 2023.

Derek J de Solla Price. The scientific foundations of science policy. Nature, 206:233–238, 1965.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. Journal of the American statistical Association, 89(427): 846–866, 1994.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the american statistical association, 90(429):106–121, 1995.

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

Paul R Rosenbaum, P Rosenbaum, and Briskman. Design of observational studies, volume 10. Springer, 2010.

Donald B Rubin. Bayesian inference for causal effects: The role of randomization. The Annals of statistics, pages 34–58, 1978.

Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association, 100(469):322–331, 2005.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In Proceedings of the National Conference on Artificial Intelligence, volume 21/2, page 1219. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

Ilya Shpitser and Judea Pearl. What counterfactuals can be tested. In Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence, UAI'07, page 352–359, Arlington, Virginia, USA, 2007. AUAI Press. ISBN 0974903930.

Leonard Susskind. The Cosmic Landscape: String Theory and the Illusion of Intelligent Design. Little, Brown, New York, 2005. ISBN 978-0316155793.

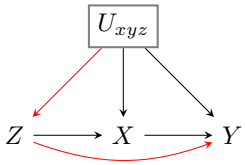
Jin Tian. Studies in causal reasoning and learning. University of California, Los Angeles, 2002.

Jin Tian and Judea Pearl. A general identification condition for causal effects. In Aaai/iaai, pages 567–573, 2002.

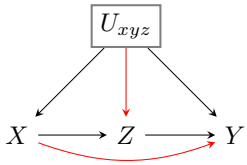
Caroline Uhler, Garvesh Raskutti, Peter Bühlmann, and Bin Yu. Geometry of the faithfulness assumption in causal inference. The Annals of Statistics, pages 436–463, 2013.

Benito Van der Zander, Johannes Textor, and Maciej Liskiewicz. Efficiently finding conditional instruments for causal inference. In Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.

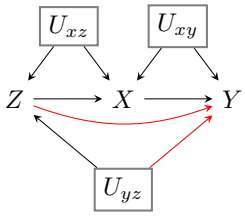
Tyler J VanderWeele and Peng Ding. Sensitivity analysis in observational research: introducing the e-value. Annals of internal medicine, 167(4):268–274, 2017.



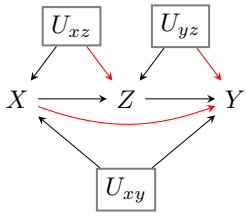
(a) IV setting with U_{xyz}.



(b) Front-door setting with U_{xyz}.



(c) IV setting with U_{xy}, U_{xz}, U_{zy}.



(d) Front-door setting with U_{xy}, U_{xz}, U_{zy}.

Figure 17: Instrumental variable (IV) and the front-door settings.

Steven Weinberg. Anthropic bound on the cosmological constant. Physical Review Letters, 59(22): 2607–2610, 1987. doi: 10.1103/PhysRevLett.59.2607.

Philip Green Wright. The tariff on animal and vegetable oils. Macmillan, 1928.

Sewall Wright. Correlation and causation. Journal of agricultural research, 20(7):557, 1921a.

Sewall Wright. Systems of mating. i. the biometric relations between parent and offspring. Genetics, 6(2):111, 1921b.

Sewall Wright. The method of path coefficients. The annals of mathematical statistics, 5(3):161–215, 1934.

J. Zhang and E. Bareinboim. Bounding causal effects on continuous outcome. In Proceedings of the 35th AAAI Conference on Artificial Intelligence, number R-61, Vancouver, Canada, Feb 2021. AAAI Press.

Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In International Conference on Machine Learning, pages 26548–26558. PMLR, 2022.

Appendix A. Instrumental Variables & the Front-Door Criterion

In this appendix, we discuss the strength of the causal assumptions required for effect identification in (i) the setting of instrumental variables (IV) and (ii) the front-door identification setting. In particular, for both settings, our goal is to determine the number of edge absences (corresponding to assumptions) in a causal diagram that are sufficient to yield the target effect identifiable. In this way, we can make a comparison of the IV and the front-door settings. Throughout, we are interested in identifying the effect of treatment X on the outcome Y.

There are two possible ways to count the number of missing edges required for identification, depending on the structure of the latent variables. The first option is to consider a single latent

confounding variable U_{xyz} that could possibly influence X, Y, Z. The IV graph corresponding to this representation is shown in Fig. 17a. Required edge absences are highlighted in red. The exclusion restriction is represented by the absence of the edge $Z \to Y$, which in counterfactual notation can be written as $Y_{xz} = Y_x$. Furthermore, the exogeneity of the instrument is represented through the missing edge $U_{xyz} \to Z$, which implies the counterfactual independence $Y_z \perp\!\!\!\perp Z$. Assuming linearity, these two conditions imply that the causal effect of X on Y can be identified from observational data.

The front-door setting is represented in Fig. 17b. Analogous to the IV setting, the edge $X \to Y$ must be absent, which implies the exclusion constraint $Y_{xz} = Y_z$. Further, the absence of the $U_{xyz} \to Z$ edge is required, which implies the counterfactual independencies $Z_x \perp\!\!\!\perp X$ and $Y_z \perp\!\!\!\perp Z_x$. These three conditions (resulting from two edge absences) imply that the causal effect of X on Y is identifiable from observational data. Therefore, when taking this perspective, the IV and front-door settings seem to require equally stringent assumptions.

The second option is to consider a latent confounding variables U_{xy}, U_{xz}, U_{yz} for each pair of observed variables X, Y, Z. The IV graph corresponding to this representation is shown in Fig. 17c. As before, the edge $Z \to Y$ must be absent, which implies that $Y_{xz} = Y_x$. The edge $U_{yz} \to Y$ must be absent, which implies that $Y_z \perp\!\!\!\perp Z$. Again, assuming linearity, these two conditions (resulting from two edge absences) imply identification of the target effect. We note here that edges $U_{xz} \to Z$ and $U_{xz} \to X$ are allowed to be present, since the condition $X_z \perp\!\!\!\perp Z$ is not required for identification.

When considering latent variables for each pair of observables, the front-door setting can be represented as in Fig. 17d. The absence of the $X \to Y$ edge implies that $Y_{xz} = Y_z$. Further, the absence of the $U_{xz} \to Z$ edge is required, which implies that $Z_x \perp\!\!\!\perp X$. The absence of $U_{yz} \to Y$ edge is also required, implying that $Y_z \perp\!\!\!\perp Z$. These three conditions (resulting from the absence of three edges) imply the identification of the target effect. When viewed from a graphical perspective, the IV setting appears to require slightly weaker assumptions (two) compared to the front-door setting (three), in addition to linearity or other parametric assumptions when considering the IV model. In terms of counterfactual assumptions, both interpretations lead to the same number of algebraic constraints.

A.1 Further Considerations

To complete the analysis, some remarks beyond the graphical structure are provided.

Relevance and Linearity. The IV setting requires a structural assumption of linearity, which may be a strong requirement. Furthermore, in finite samples, the IV setting also requires the so-called relevance assumption, which means that the instrument Z needs to have a correlation with the treatment X sufficiently different from zero. At the same time, the front-door setting does not require such structural assumptions.

Possibility of Randomization. Another aspect worth considering is the possibility of aiding identification by means of randomization. As pointed out by Imbens (2020), in the IV setting, if one were to randomize the instrument Z, the assumption $Y_z \perp\!\!\!\perp Z$ would be satisfied by design, alleviating the strength of the assumptions required for identification. At the same time, no such analogous randomization of the mediators Z is possible in the front-door setting.

In conclusion, the assumptions required for the IV setting seem to be comparable in strength to those for the front-door setting. However, while the IV approach has been widely applied in practice, the front-door setting remains underutilized. This disparity raises an important question: would the front-door method prove more practical if data scientists were more familiar with it and trained to recognize opportunities for its application.

Appendix A. Probabilities of Identification – Supplementary Figures

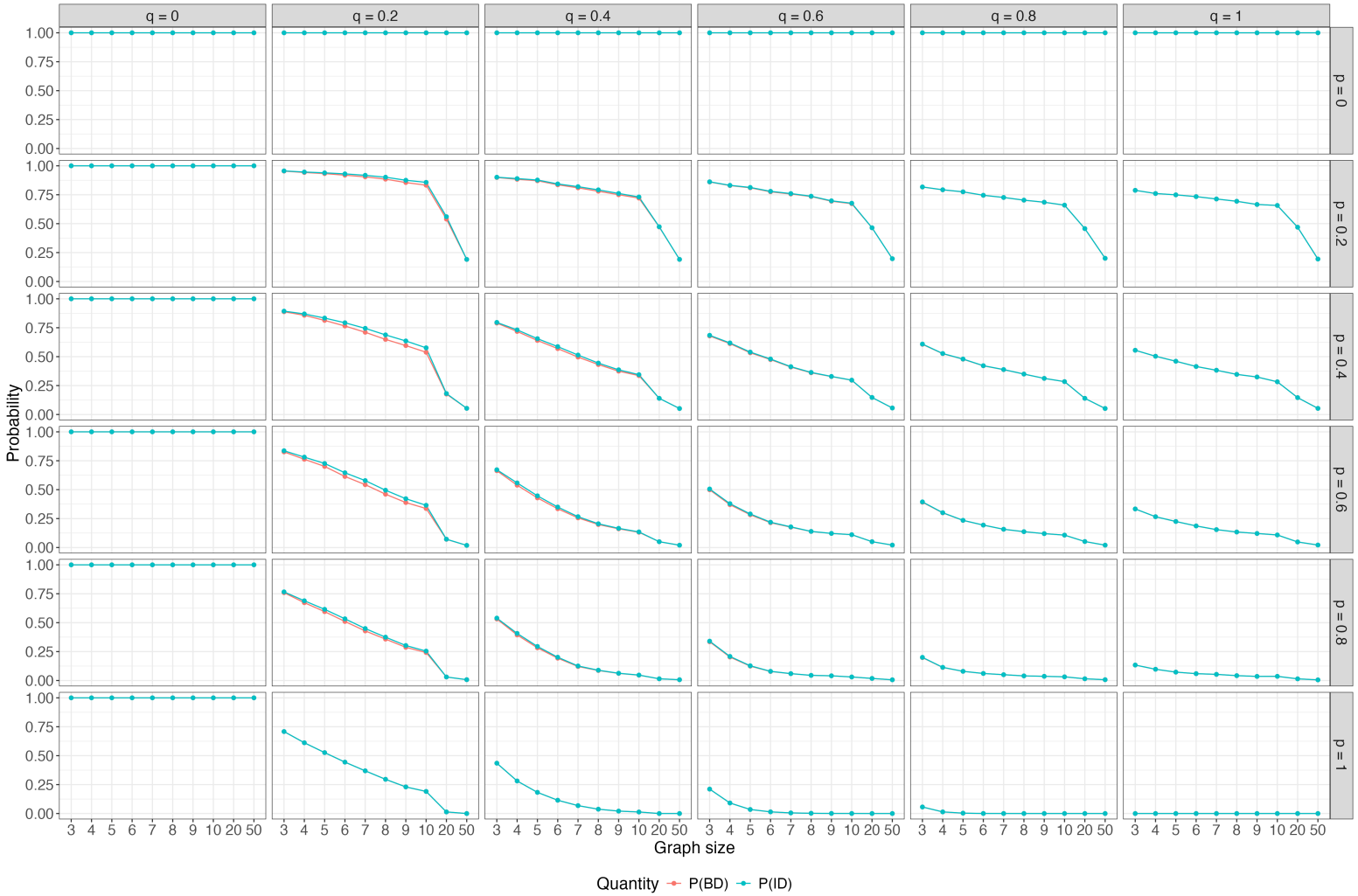


Figure 18: Probabilities of Identification for the $\overrightarrow{G}(n,p) + G(n,q)$ model over a range of parameters.

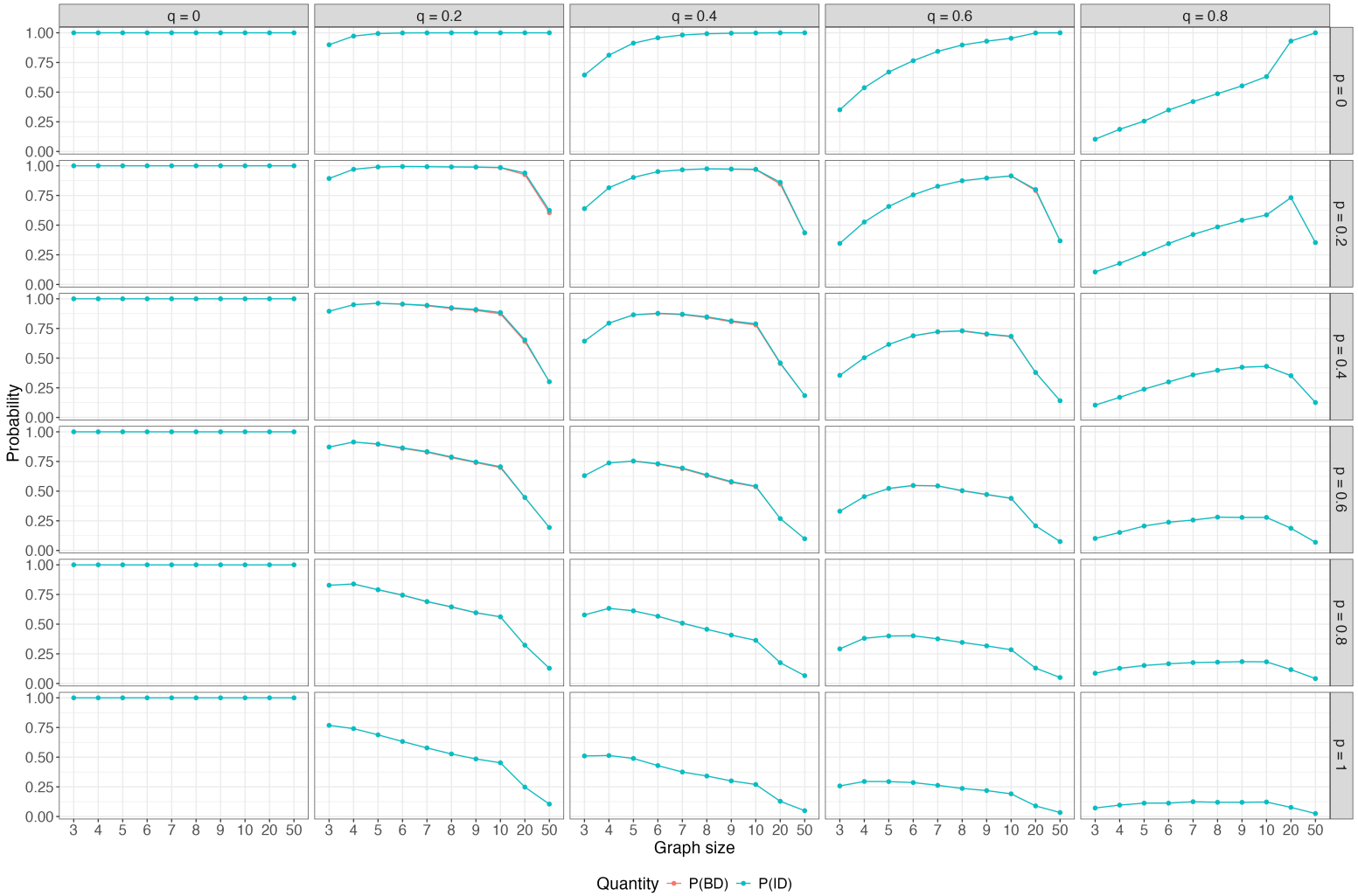


Figure 19: Probabilities of Identification for the $\overrightarrow{G}(n,p)$ + IP(q) model over a range of parameters.

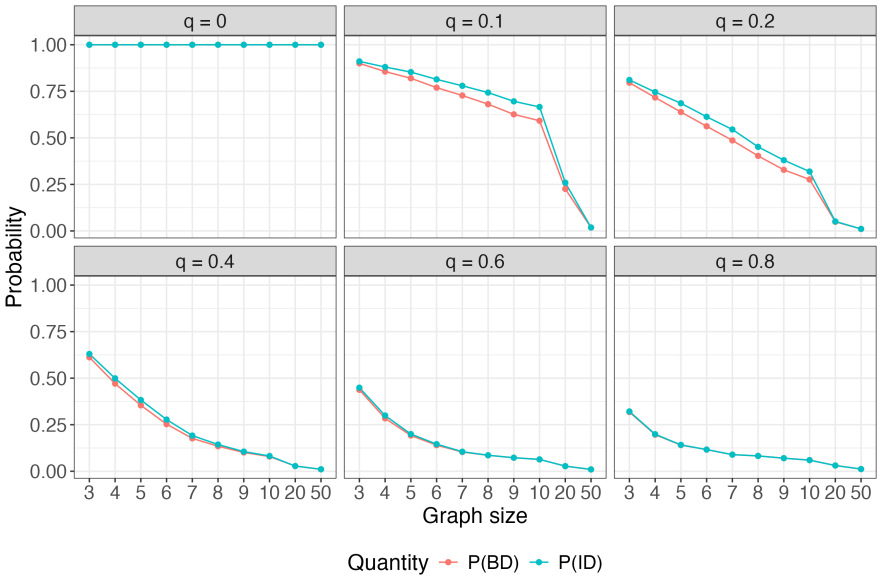


Figure 20: Probabilities of Identification for the UG(n) + $G(n, q)$ model over a range of parameters.

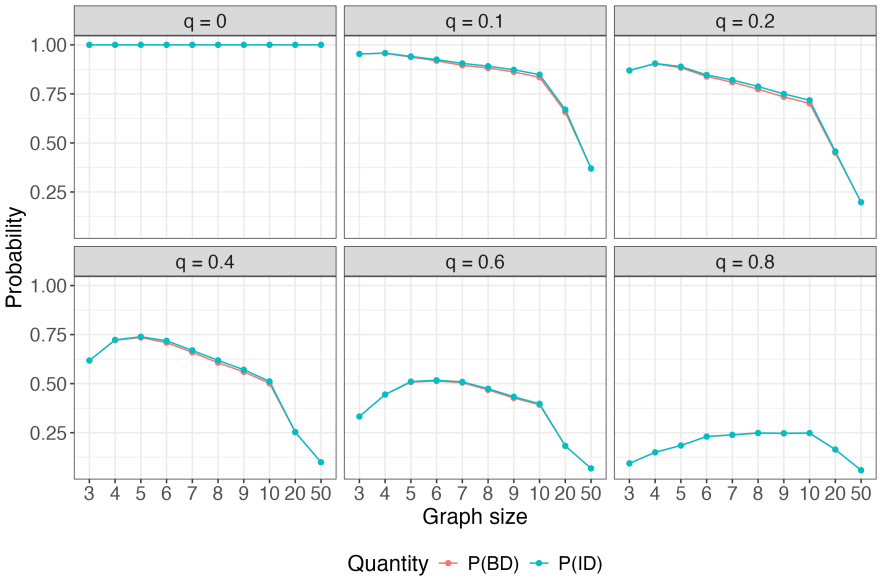


Figure 21: Probabilities of Identification for the UG(n) + IP(q) model over a range of parameters.

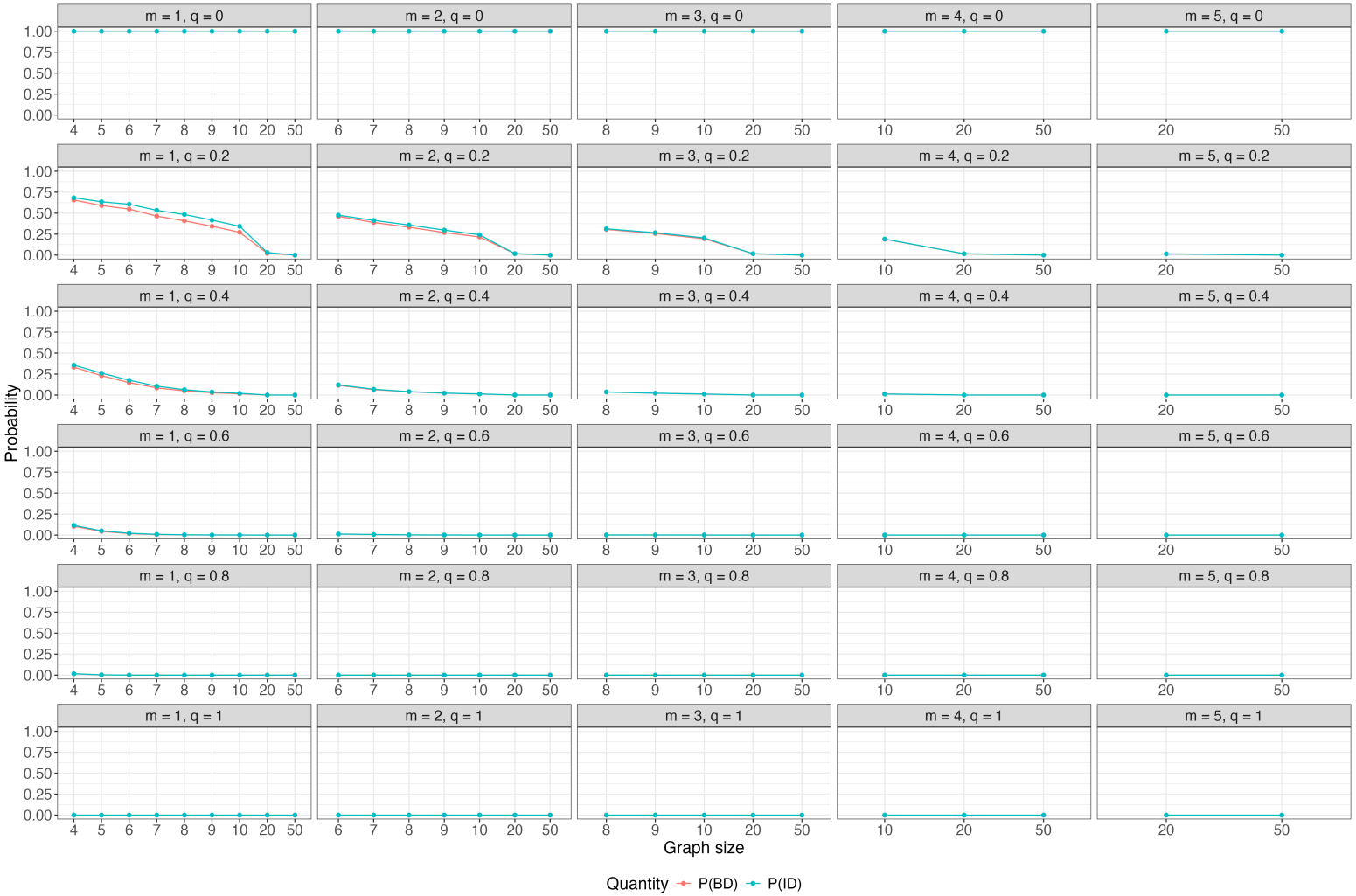


Figure 22: Probabilities of Identification for the PM(n, m) + $G(n, q)$ model over a range of parameters.

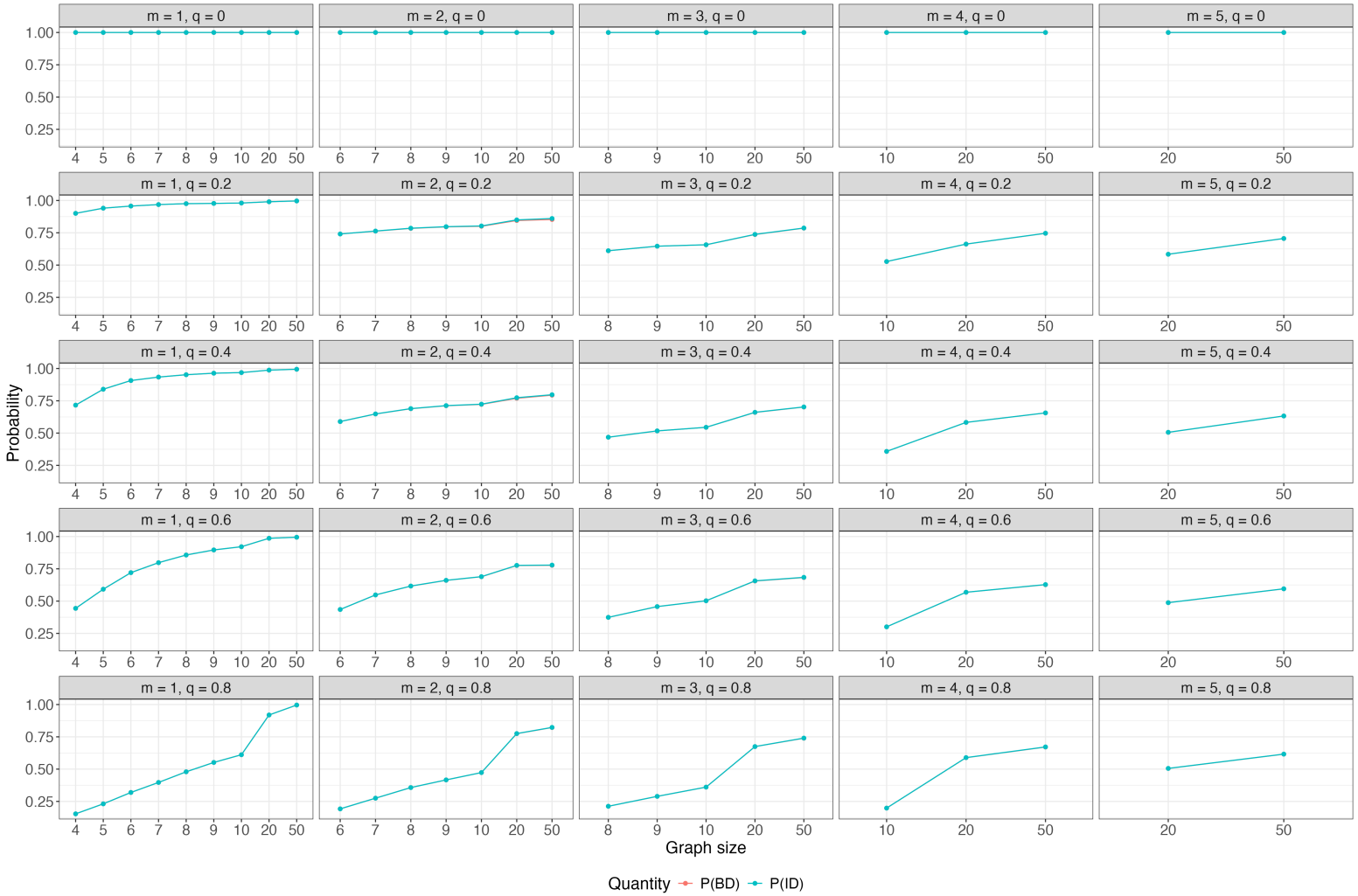


Figure 23: Probabilities of Identification for the PM(n, m) + $G(n, q)$ model over a range of parameters.

Appendix B. Coverage Probabilities – Supplementary Figures

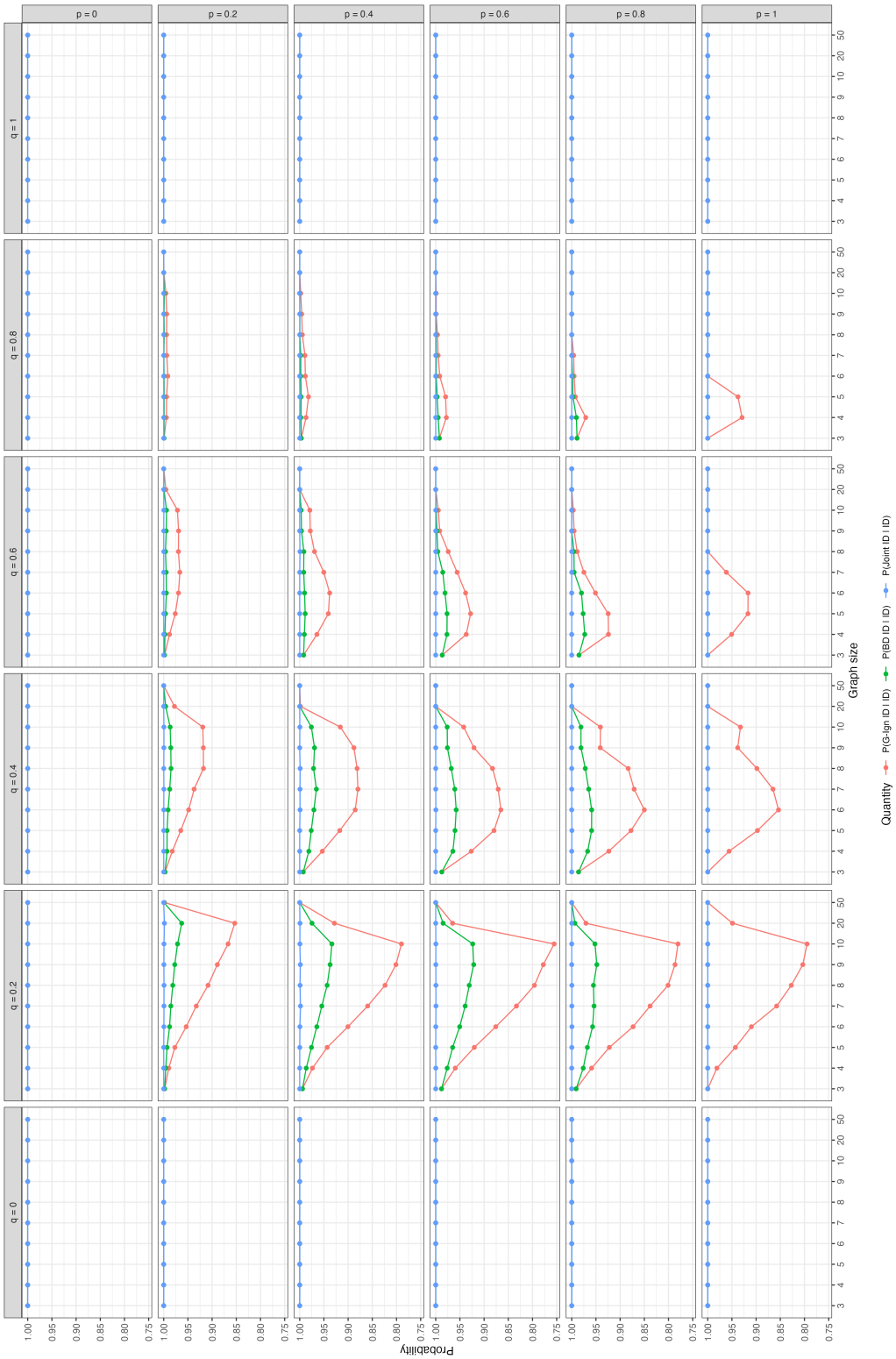


Figure 24: Coverage of Strategies for the $\overrightarrow{G}(n,p) + G(n,q)$ model over a range of parameters.

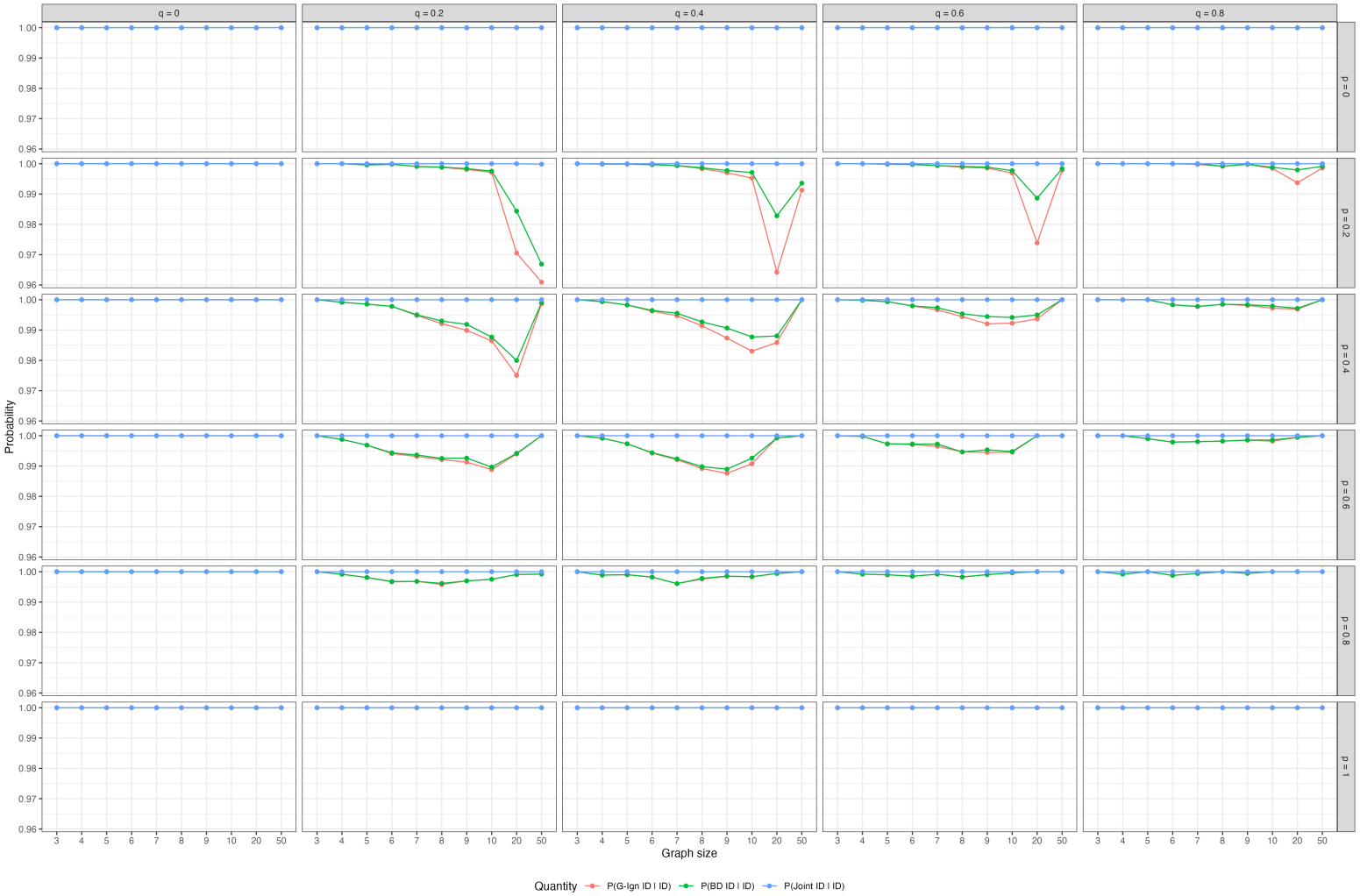


Figure 25: Coverage of Strategies for the $\overrightarrow{G}(n,p)$ + IP(q) model over a range of parameters.

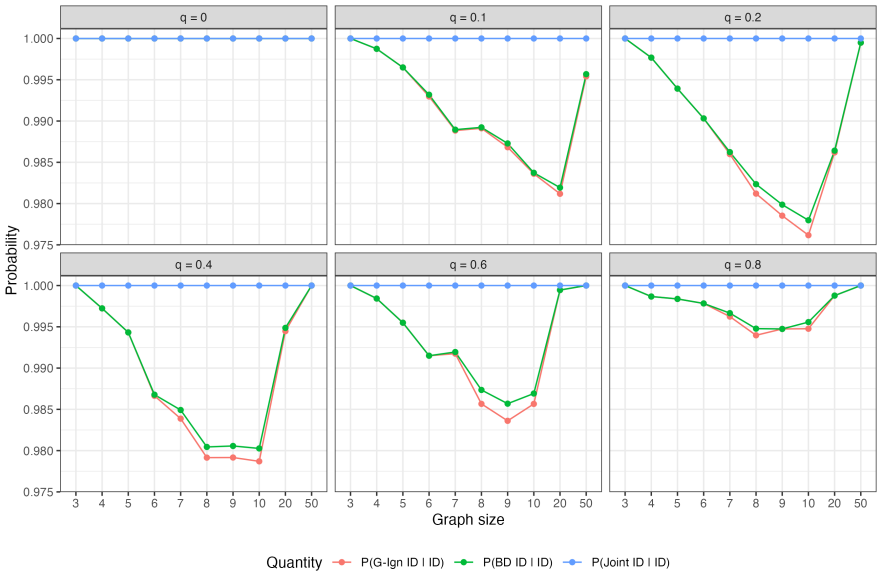


Figure 26: Coverage of Strategies for the UG(n) + IP(q) model over a range of parameters.

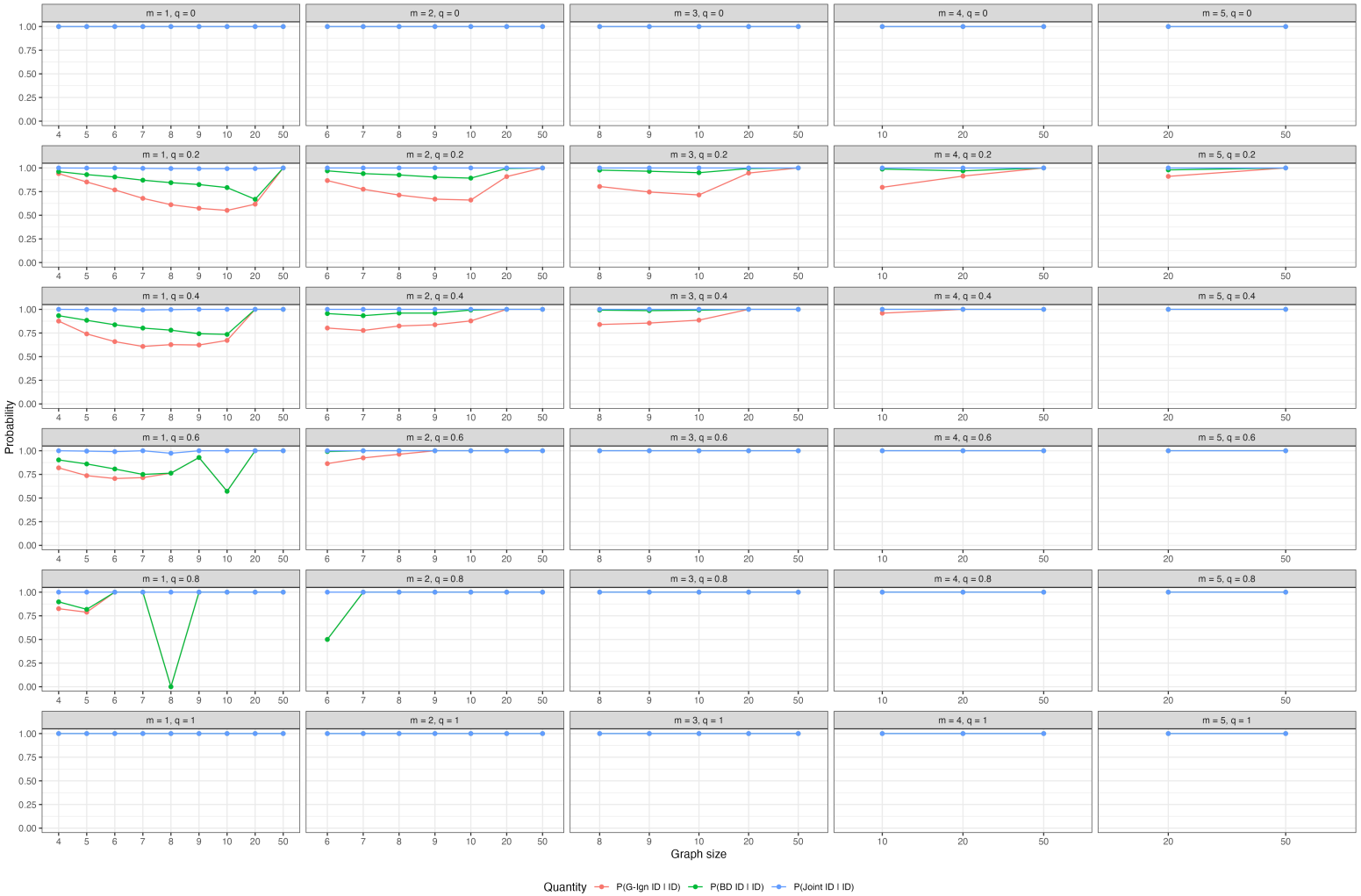


Figure 27: Coverage of Strategies for the PM(n, m) + $G(n, q)$ model over a range of parameters.

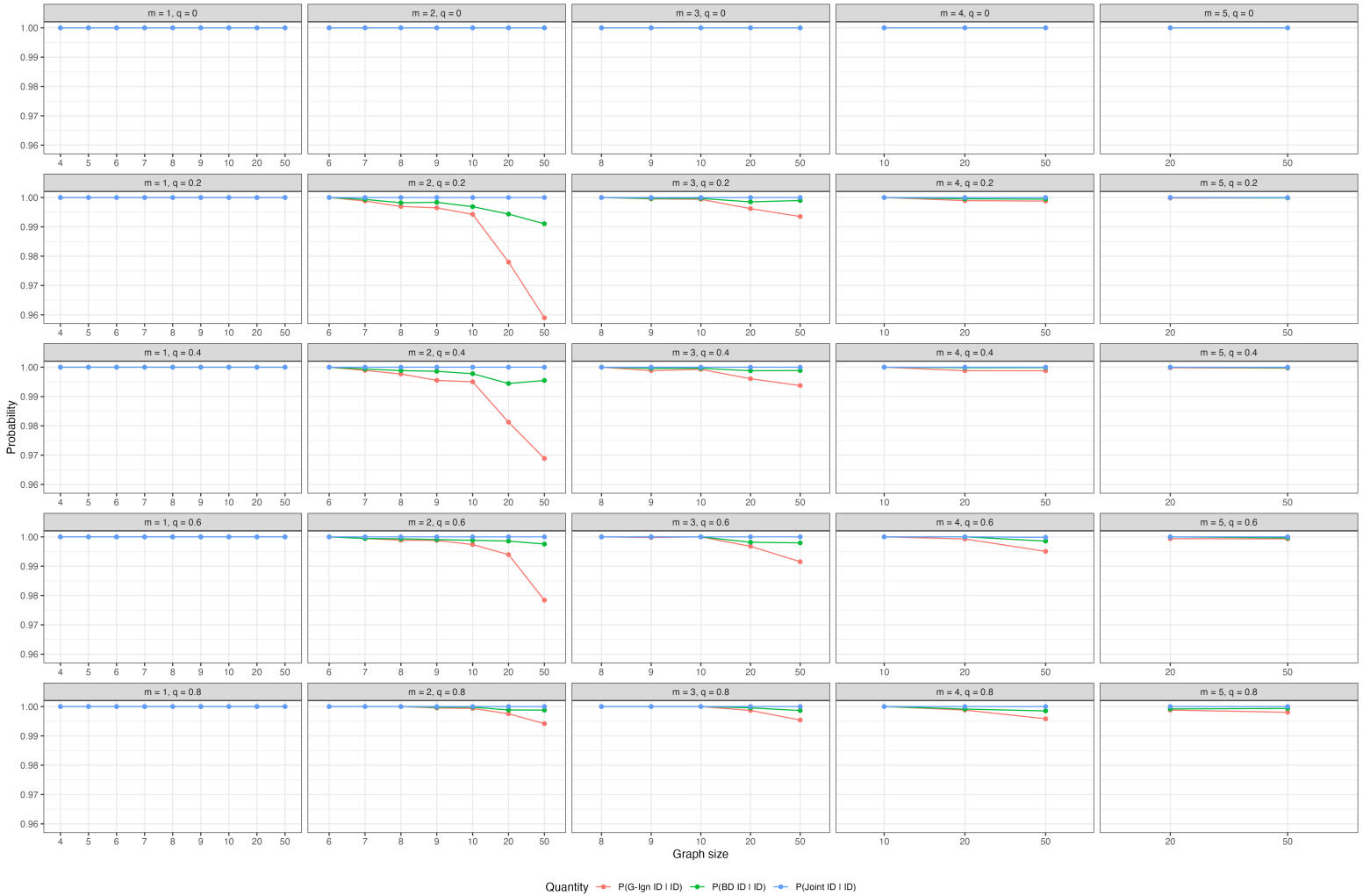


Figure 28: Coverage of Strategies for the PM(n, m) + IP(q) model over a range of parameters.