A. Details of the FCI based approach

In this section we describe the details of how to utilize the FCI algorithm to make inferences ‘∗’, ‘0’, or ‘?’ as mentioned in Section 5 of the article. Without going into all the details of the vast theory of ancestral graphs (Richardson and Spirtes, 2002; Ali et al., 2009), we note that the output of FCI is a partial ancestral graph (PAG), representing the equivalence class over all maximal ancestral graphs (MAGs) with the same (in)dependencies over the observed variables. It is possible to make the inferences ‘∗’, ‘0’, and ‘?’ from the output PAG of FCI as follows: If there is a directed edge \( x \rightarrow y \) in the PAG, output ‘∗’ with an admissible set \( Z \) as constructed in the following paragraph.\(^1\) If there is no edge between \( x \) and \( y \) or a bidirected edge \( x \leftrightarrow y \) in the PAG, output ‘0’. If there is an edge \( x \leftrightarrow y \) in the PAG, output ‘?’\(^2\).

If we infer ‘∗’, we also need to output an admissible set \( Z \) to estimate the strength of the causal effect. This set can be read off the PAG as follows: First, one obtains a MAG from the equivalence class of the given PAG using Lemma 4.3.6 of Zhang (2006). One then adds a ‘policy’ variable \( p \) to this MAG whose only connection is from \( p \) to \( x \). Using Theorem 6.2 of Spirtes et al. (2000) one then can find a set \( Z' \) which m-separates \( p \) from \( y \) (the equivalence of d-separation in DAGs for MAGs). It turns out that the set \( Z' \) contains, in our specific setting, exactly all the parents of \( y \) in the MAG. Knowing this, we can actually read off this set \( Z' \) directly from the PAG (without requiring any of the MAGs) simply by selecting all those variables \( v \) with either an edge \( v \rightarrow y \), or \( v \leftarrow y \) in the PAG. We thus obtain an admissible set (blocking all back-door paths from \( x \) to \( y \)) as \( Z = Z' \setminus \{ x \} \), since to any back-door path from \( x \) to \( y \) we can concatenate the edge \( p \rightarrow x \) to obtain a path from \( p \) to \( y \) via \( x \) with \( x \) being an active collider, and hence any such path must be blocked by \( Z \).

B. Simulations with 100 covariates

As mentioned in Section 6 of the article, we tested our inference rules on models with 100 observed and 20 hidden covariates using the sampling approach of Section 4. In such large models it is computationally expensive to infer whether there truly exists an admissible set, so for simplicity we merged tasks #1 and #2 (non-zero causal effect of \( x \) on \( y \)), as well as tasks #3 and #4 (no effect of \( x \) on \( y \)). For both settings we generate 100 models as described in Section 6, and generate data with sample sizes of 100, 1000, and 10000, respectively.

In tasks #1 and #2, the method conservatively outputs ‘?’ in 59, 53, and 64 cases (out of the 100 models), for 100, 1000, and 10000 samples, respectively. Even though the cases of wrongly inferring ‘0’ decrease with growing sample size, our method still outputs ‘0’ in 24 cases for the largest sample size. However, as Figure 1 shows, when our approach infers a zero effect from \( x \) on \( y \), the true underlying non-zero effect is rather close to zero. Furthermore, we can see from this figure that for growing sample size the magnitude of the error made in the estimates decrease, on average.

For tasks #3 and #4 our method can reliably detect the zero effect, inferring ‘0’ in 41, 82, and 77 cases (out of the 100 models), for sample sizes 100, 1000, and 10000, respectively, and ‘?’ in almost all other cases, i.e. we rarely make the only wrong decision ‘∗’.
Figure 1: True versus estimated causal effect using 100 models with 100 observed and 20 hidden covariates with a non-zero effect of \( x \) on \( y \), and various sample sizes (100, 1000, and 10000, respectively, from left to right). True effects are on the horizontal axis, estimated effects are on the vertical axis. Estimated effects are only shown for those models for which our method inferred ‘±’ or ‘0’.

C. Proof of Theorem 4

Condition (i) ensures that there is an active path \( p_1 \) from \( w \) to \( x \) not blocked by \( Z \), which is out of \( w \) (by the exogeneity) and into \( x \), i.e. \( w \rightarrow \ldots \rightarrow x \) (this is not necessarily a directed path). Condition (ii) ensures that there is an active path \( p_2 \) from \( w \) to \( y \) not blocked by \( Z \cup \{x\} \), which is out of \( w \) and into \( y \).

If \( x \) is not necessarily needed in the conditioning set of condition (ii), i.e. \( w \not\perp \perp y \mid Z \) holds, then concatenating \( p_1 \) and \( p_2 \) at \( w \) yields an active back-door path from \( x \) to \( y \) not blocked by \( Z \cup \{x\} \), which is out of \( w \) and into \( y \).

On the other hand, if \( x \) needs to be in the conditioning set of condition (ii), then \( x \) is either a collider on \( p_2 \), implying that the subpath from \( x \) to \( y \) in \( p_2 \) is an active back-door path from \( x \) to \( y \) not blocked by \( Z \), or \( x \) is a descendant of a collider on the path \( p_2 \), and concatenating the directed path from the collider to \( x \), with the subpath from the collider to \( y \) on \( p_2 \) yields an active back-door path (once more using Lemma 3.3.1 of Spirtes et al., 2000).

References


