Identifying the Causal Effects of Cross-World Policies

Noah Weber[*]
Johns Hopkins University
nweber6@jhu.edu

Levi Boyles
Microsoft Ads
leboyles@microsoft.com

Shuayb Zarar
Microsoft Ads
shuayb@microsoft.com

Abstract

In this work, we introduce the notion of a counterfactual response to what we call a *Cross World Policy*. Cross World policies are defined as a type of dynamic treatment regime which assigns treatments based on a fixed function of the naturally observed value of causally prior covariates, including the treatment itself. Cross World policies share commonalities with treatment effects on the treated (albeit in a dynamic treatment regime setting) and generalize the idea of shift interventions on the treated (SITs) developed in Sani et al. (2020). We give examples of potential queries of interest which may be described as Cross World policies and complete identification criteria for estimation from observed data.

1 Introduction

Policies (also called dynamic treatment regimes) define decision rules which use the values of previously observed covariates, $[Z_0, ..., Z_i] = \bar{Z}_{:i}$, and treatments, $[A_0, ..., A_{i-1}] = \bar{A}_{:i-1}$ to pick treatment assignment A_i at time step i. Evaluating the effect that deploying a particular policy has on some outcome Y is a general problem encountered across a wide array of settings. Of course, evaluating this effect by actually deploying a the policy is often infeasible. In this case, a policy maker is forced to rely on two things to reason about the effects of a hypothetical policy f: (1) the available observed data, generated from a possibly unknown *natural* policy f_n, and (2) their assumptions about the problem setting. Determining when (ie under what assumptions) and how this effect can be estimated directly from observed data is a key topic in causal inference (Robins, 1986, 1997; Tian, 2008; Young et al., 2014; Nabi et al., 2018). If, given a set of causal assumptions, the effect of running a different policy can be estimated from the observed data, the effect is said to be identified.

Policies are typically defined with respect to a single world. That is to say, if we let $X_i(f)$ stand for the value that the variable X_i would take had we ran the policy f, the past history that a policy relies on is defined as $(\bar{Z}_{:i}(f), \bar{A}_{i-1}(f))$; so all variables the policy depends on are evaluated with respect to the world in which we had ran the policy. For policies defined in this way, complete algorithms for identification have been developed (Tian, 2008; Shpitser & Sherman, 2018).

There are also policies of interest that may be defined with respect to more than one world. As an example, a regime considered in section 5.1 of Richardson & Robins (2013) intervenes at step i to enforce a mandatory twenty minuets of exercise if a person would have, in the absence of an intervention at step i, exercised for less than twenty minuets; the regime does nothing if they would have naturally exercised for more than twenty minuets. In contrast to the kinds of policies described above, this policy additionally relies on the value that A_i would have taken in a world in which the policy was stopped at step i. Policies of this form were studied in (Robins et al., 2004; Young et al., 2014), and termed as *shift intervention policies* (SIPs) in Sani et al. (2020), who also provided complete identification criterion. One can also define policies which depend not just on the value A_i had we stopped the policy at step i, but also on the value of A_i had we we not run the policy at all (ie

[*] Work done during internship at Microsoft

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

the *natural* observed value of A_i) in a way analogous to treatment effects on the treated. Policies of this kind were also studied in Sani et al. (2020), who termed them as shift interventions on the treated (SITs).

In this paper, we generalize the notion of SITs defined in Sani et al. (2020) to policies that may depend not only on the natural value of A_i had we not run the policy, but also potentially on the values of causally prior covariates from either world (the 'natural' observed world or the counterfactual world in which we had run the policy). We define these types of policies as *Cross-World policies* due to their potential cross world dependences. To motivate the problem, we first give examples of possible queries of interest which may be described through cross world policy effects. We then provide complete identification criteria for cross-world policy effects.

1.1 Instances of Cross-World Policies

The task of formalizing (or reading another's formalization of) a type of causal question benefits greatly if it corresponds to a question someone might actually want to ask. Here we give several examples of kinds of queries that correspond to the evaluation of cross-world policy effects (from here on refered as CWP):

- **Thresholding and Shifting Policies with Covariates:** Much like how SITs allow thresholding policies on the natural treatment value (policies of the form: *Do X if the natural value that treatment would take is above/below a threshold.*) and shifting policies (policies which modify the natural value of a treatment) CWPs may extend similar policies to permit the input of covariates in the policy function. **Possible Questions:** *What would the chance of a patients recovery be had we doubled the dosage on the days where their systolic blood pressure rose had risen above 150 and kept it the same on all other days?*
- **Replacement Policies**: In fields such as digital advertising or robotics, quantifying the effects of policies that replace or drop instances of one action with another suitable action may be relevant for answering questions of importance attribution. **Possible Questions:** *What would the chance of a user visiting a website be if, on days in which the user had seen the ad A we instead showed them an alternate ad B still relevant to their current query?*
- **Policies with different behaviours on subpopulations defined by original treatment values**: Similar to how the effect of treatment on the treated calculates an effect of an intervention on the subpopulation of those who originally would receive treatment, CWPs can also be used to formulate the effects of policies whose behaviour is different for units in certain subpopulations, where the subpopulation is defined by a specific setting of original treatment value. **Possible Questions:** *What would be the effect of a new policy which which runs an alternate regime A on units that otherwise, up the current time, would have not received treatment under the default policy, and runs the default policy on all other units?*.

2 Preliminaries

In this section we will layout some preliminaries. First some notation. We will denote a set of variables and values with a bolded uppercase $\boldsymbol{V}$ and $\boldsymbol{v}$ respectively. Single variables and values will be likewise but unbolded. We will be working with models defined on graphs, $\mathcal{G}$, with variables $\boldsymbol{V}$ as nodes. We will denote subgraphs containing only nodes in the set $\boldsymbol{U}$ as $\mathcal{G}_{\boldsymbol{U}}$. We will also use the following graphically defined sets: parents, children, descendants, non-descendants, and ancestors of a variable. These will be written as $\text{Pa}_{\mathcal{G}}$, $\text{Ch}_{\mathcal{G}}$, $\text{de}_{\mathcal{G}}$, $\text{nd}_{\mathcal{G}}$, and $\text{an}_{\mathcal{G}}$[2]. For all graphs we work with in here, we assume the Non Parametric Structural Equation Model with Independent Errors (Pearl, 2009) as our underlying causal model. See the aforementioned reference for further details.

2.1 DAGs, ADMGs, and CADMGs

Acyclic Directed Mixed Graphs (ADMGs) are a graphical formalism similar to directed acyclic graphs (DAG) that permit for a more parsimonious representation of hidden variables. Given a DAG $\mathcal{G}(\boldsymbol{V} \cup \boldsymbol{H})$, where $\boldsymbol{V}$ are observed variables and $\boldsymbol{H}$ are hidden, one can define an equivalent ADMG $\mathcal{G}(\boldsymbol{V})$, with edges taking the following meanings: An edge $A \rightarrow B$ in an ADMG indicates a directed

[2] Unless otherwise noted, we will use the same definition for these sets as Shpitser & Sherman (2018).

path from A to B in $\mathcal{G}(\boldsymbol{V} \cup \boldsymbol{H})$ with all intermediate nodes on the path in $\boldsymbol{H}$. A $A \leftrightarrow B$ edge indicates a path with no colliders from A to B in the DAG with the first edge on the path being of the form $A \leftarrow$, the last being of the form $\rightarrow B$, and all other nodes on the path being in $\boldsymbol{H}$.

A Conditional ADMG (CADMG) is defined in a similar fashion to a ADMG with its variables partitioned into two sets: *random variables*: $\boldsymbol{V}$, and *fixed variables*: $\boldsymbol{W}$. In a CADMG, $\mathcal{G}(\boldsymbol{V}, \boldsymbol{W})$, the variables in $\boldsymbol{V}$ are treated the same as before, whereas the nodes in $\boldsymbol{W}$ as instead fixed to constant values, with all incoming arrows to them removed.

An important structure in ADMGs and CADMGs are *districts*. A District is a maximal set of nodes in a graph connected to eachother by paths made up of only $\leftrightarrow$ edges. In a CADMG, districts are defined only over nodes in $\boldsymbol{V}$. The set of districts in a graph is noted as $\mathcal{D}(\mathcal{G})$, while the district that a node V_i belongs to in $\mathcal{G}$ is denoted as $\text{Dis}_{\mathcal{G}}(V_i)$. See Richardson et al. (2017) for details on ADMGs and their properties.

2.2 Kernels and Fixing

A *kernel*, $q_V(V|W)$, is defined as a map from values in $\boldsymbol{W}$ to normalized densities over $\boldsymbol{V}$. Kernels may be seen as a generalized version of conditional probability distributions (though they need not follow properties such as Bayes rule, etc). Marginalization and conditioning in kernels are defined in a similar fashion in the sense that:

$$q(A|W) \equiv \sum_{V \setminus A} q(V|W); q(V \setminus A|A, W) \equiv \frac{q(V|W)}{q(A|W)}$$

Note that the full joint distribution, $P(\boldsymbol{V})$ can also be seen as a kernel (with $\boldsymbol{W} = \emptyset$).

A variable $V \in \boldsymbol{V}$ is said to be *fixable* in a CADMG $\mathcal{G}(\boldsymbol{V}, \boldsymbol{W})$ if $\text{de}_{\mathcal{G}}(\boldsymbol{V}) \cap \text{Dis}_{\mathcal{G}}(\boldsymbol{V}) = \emptyset$. In words, a variable is fixable in a graph if none of its descendants are found in the same district. If a variable $V \in \boldsymbol{V}$ is fixable one may define a fixing operator, $\phi_V(q(\boldsymbol{V}|\boldsymbol{W}), \mathcal{G})$ on the CADMG $\mathcal{G}(\boldsymbol{V}, \boldsymbol{W})$ and its respective kernel $q(\boldsymbol{V}|\boldsymbol{W})$ which yields a new kernel:

$$\phi_V(q(\boldsymbol{V}|\boldsymbol{W}), \mathcal{G}(\boldsymbol{V}, \boldsymbol{W})) = \frac{q(\boldsymbol{V}|\boldsymbol{W})}{q(V|\text{nd}_{\mathcal{G}}(V) \cup W)}$$

The fixing operation also returns a new CADMG, $\mathcal{G}(\boldsymbol{V} \setminus \{V\}, \{V\} \cup \boldsymbol{W})$, which fixes the node V and removes all of its incoming arrows.

A distribution $p(\boldsymbol{V})$ that follows the nested Markov factorization (Richardson et al., 2017) with respect to an ADMG $\mathcal{G}$ may be factorized as:

$$p(\boldsymbol{V}) = \prod_{D \in \mathcal{D}(\mathcal{G}_{\boldsymbol{V}})} \phi_{V \setminus D}(p(\boldsymbol{V}), \mathcal{G})$$

and for any fixable set $\boldsymbol{S}$, fixing the nodes in $\boldsymbol{S}$ yields a kernel that factorizes as:

$$\phi_{\boldsymbol{S}}(p(\boldsymbol{V}), \mathcal{G}) = \prod_{D \in \mathcal{D}(\phi_{\boldsymbol{S}}(\mathcal{G}))} \phi_{V \setminus D}(p(\boldsymbol{V}), \mathcal{G})$$

This factorization forms the backbone of much work in nonparametric identification theory, and is at the heart of the ID algorithm (Shpitser & Pearl, 2006; Richardson et al., 2017); a complete algorithm for the identification of interventional effects.

2.3 Edge and Path Interventions

Edge interventions (Shpitser & Tchetgen, 2016) are a generalization of the above node interventions which, in addition to cutting the incoming edges of the intervened on variable V, can also set V to different constant values for each of its outgoing edges. As an example, for a graph $X \leftarrow Z \rightarrow Y$, a node intervention $\text{do}(Z = a)$ can set Z to a constant a leading to a joint factorization where Z is set to a wherever it appears $p(X|Z = a)p(Y|Z = a)$. An edge intervention may instead intervene to set Z to a constant a for the edge $(ZX)_{\rightarrow}$ and a different constant b for the edge $(ZY)_{\rightarrow}$, giving a joint factorization $p(X|Z = a)p(Y|Z = b)$. Path interventions further generalize edge interventions, allowing one to set Z to a constant value along a specified directed path starting at Z. As shown in Shpitser & Tchetgen (2016), only path interventions that are expressible as edge interventions are identified.

Edge interventions play an important role here as our query of interest, the effect of a CWP, may be seen as a path specific effect. Here we will give notation for the policy edge/path interventions that will be of importance here. Let α be a subset of edges in our graph, then define a edge specific policy as $\boldsymbol{f}_\alpha = \{f_A^{(AX)_\rightarrow} | (AX)_\rightarrow \in \alpha\}$, where each $f_A^{(AX)_\rightarrow}$ assigns a value to A along edge $(AX)_\rightarrow$. The potential outcome of an edge policy intervention is then written as $Y(\boldsymbol{f}_\alpha)$. For path interventions, let $\pi_{\boldsymbol{X},\boldsymbol{Y}}$ be a subset of paths that start at some $X_i \in \boldsymbol{X}$ and end at some $Y \in \boldsymbol{Y_i}$ and do not intersect $\boldsymbol{X} \setminus X_i \cup \boldsymbol{Y} \setminus Y_i$. Let $\boldsymbol{f}_{\pi_{\boldsymbol{X},\boldsymbol{Y}}}$ be the set of functions which assign values to the sources of paths in $\pi_{\boldsymbol{X},\boldsymbol{Y}}$. If this is identified, it is expressible as the edge policy $\{f_A^{(AX)_\rightarrow} | (AX)_\rightarrow$ is prefix of a path in $\pi_{\boldsymbol{X},\boldsymbol{Y}}\}$.

3 Identification of Cross World Policies

Here we provide complete identification criteria for CWPs. The proofs will build upon the theory presented in Sani et al. (2020) and Shpitser & Sherman (2018), though some additional subtlety arises due to two issues: (1) the cross world dependencies are not readily displayable in a graphical format, and (2) depending on the inputs of the hypothetical policy we want to evaluate, new dependencies between variables may open up which lead to the query being unidentified. At a high level, our proof strategy involves handling these two issues in turn, allowing us to build off prior results to prove completeness.

3.1 Cross World Policies: Formal Definitions

Before we begin, we first define cross world policy responses using the substitution construction of counterfactuals. Fix a set of functions on each treatment variable: $\boldsymbol{f} \equiv \{f_i : \mathcal{X}_{W_i} \mapsto \mathcal{X}_{A_i} | A_i \in A\}$ where $\mathcal{X}_{A_i}$ gives the domain of possible values that A_i may be assigned to, and W_i is the domain for the input of f_i, with the restriction that this input be causally prior to A_i.

A Cross World Policy is defined in a similar fashion to typical policy/dynamic regimes with the distinguishing feature being the type of input that functions in $\boldsymbol{f}$ can take. Let $\mathcal{B}(A_i)$ be variables in $\boldsymbol{V}$ causally prior to A_i. Then for each $V_i \in \mathcal{B}(A_i)$, f_i may potentially take one of two versions of this variable. The first version of this variable, $V_i(\boldsymbol{f})$, is the value that the variable would take had we been running the set of policies $\boldsymbol{f}$ from the start. This is generally the default semantics of policy arguments when specifying dynamic treatment regimes. The second version of this variable is the natural value the variable would take if we didn't run $\boldsymbol{f}$. We will denote this as $V_i(\boldsymbol{f}_\mathfrak{N})$ (where $\boldsymbol{f}_\mathfrak{N}$ is the "natural" policy) when we need to distinguish it from references to the variable V_i.

Note for a potential input $V_i \in \mathcal{B}(A_i)$, f_i may take either $V_i(\boldsymbol{f})$ or $V_i(\boldsymbol{f}_\mathfrak{N})$ as input, but not both, as $(V_i(\boldsymbol{f}), V_i(\boldsymbol{f}_\mathfrak{N}))$ is unidentified even from experimental data[3]. For a policy f_i, write its domain as $W_i = W_i^f \cup W_i^\mathfrak{N}$, where W_i^f are inputs of the form $V_i(\boldsymbol{f})$ and $W_i^\mathfrak{N}$ are inputs of the form $V_i(\boldsymbol{f}_\mathfrak{N})$. For the purpose of simplifying proofs, we will assume for the rest of the paper that for each $f_i \in \boldsymbol{f}$ for a cross world policy $\boldsymbol{f}$ we have $W_i^\mathfrak{N} \neq \emptyset$.

Our with the above definition of W_i, a Cross World Policy is defined as:

$$Y(f) = Y(\{f_{A_i}(W_i) | A_i \in \text{pa}_\mathcal{G}(Y) \cap \boldsymbol{A}\}, \{V_i(f) | V_i \in \text{pa}_\mathcal{G}(Y) \setminus \boldsymbol{A}\}) \qquad (1)$$

As an example, the shift intervention on the treated defined in Sani et al. (2020) may be considered a case of the above where for each f_i, $W_i = W_i^\mathfrak{N} = A_i(\boldsymbol{f}_\mathfrak{N})$.

As done similarly in the proofs of Sani et al. (2020), one can represent such a quantity above as the response to the *ETT path policy intervention*[4]. For treatment set $\boldsymbol{A}$ and outcome set $\boldsymbol{Y}$, the ETT path intervention is defined as $\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}$, where $\boldsymbol{f}$ is a set of policy functions for each path, $\pi_{\boldsymbol{X},\boldsymbol{Y}}$ is defined as in Section 2.3, and $\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}$ sets the sources of paths in $\pi_{\boldsymbol{X},\boldsymbol{Y}}$ by their corresponding function in $\boldsymbol{f}$.

[3] An exception to this is the case in which $V_i(f_\mathfrak{N})$ and $V_i(f)$ are always equivalent, such as the case when they are an ancestor for all variables in $\boldsymbol{A}$. In this case, one can arbitrarily pick either variable kind as input.

[4] See section 4 of Shpitser & Tchetgen (2016)

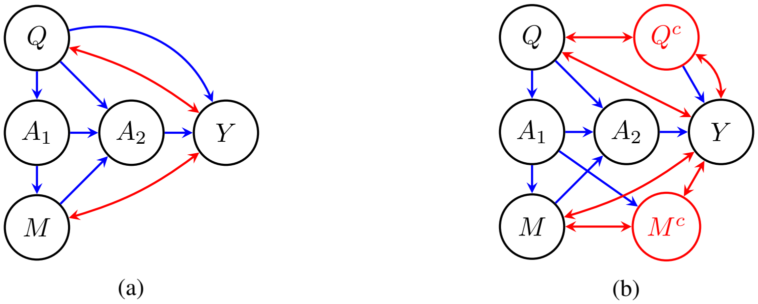

Figure 1: The graph (b) gives the CWPG constructed for graph (a), with A_1 and A_2 as the treatment variables, using the procedure given in Section 3.2. The "copy" nodes (ie the nodes in the set $\boldsymbol{C}$) are colored red with variable names superscripted by c. Note that the node Q in (b) no longer has a directed edge to Y in accordance with the last step of the CWPG construction procedure.

Note that this path intervention implicitly sets the sources of paths not in $\pi_{\boldsymbol{X},\boldsymbol{Y}}$ to the natural value they would have taken under no intervention.

From here we will rephrase the counterfactual (1) in terms of the path intervention $\boldsymbol{Y}(\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}})$. One can look at this rephrasing as follows: for any counterfactual response of $Y(A = a)$ conditioned on the original observed value of $A = b$ (eg an ETT computation), what we wish to evaluate is what happens when all causal paths from A to Y behave like we intervened on $A = a$, while all other covariates take on values as if A was still set to its original value b.

3.2 Graphical Representations of Cross World Policies

To facilitate the proof we develop a graphical representation which explicitly represents the natural (ie $V_i(\boldsymbol{f}_{\mathfrak{N}})$) and counterfactual versions (ie $V_i(\boldsymbol{f})$) of the covariates. For lack of a better name, we will refer to them here as Cross-World Path Graphs (CWPG). A CWPG is built with respect to an ADMG, $\mathcal{G}$, a fixed set of outcome variables, $\boldsymbol{Y} \subset \boldsymbol{V}$, and of treatment variables, $\boldsymbol{A} \subset \boldsymbol{V} \setminus \boldsymbol{Y}$. A CWPG, $\mathcal{G}'(\mathcal{G}, \boldsymbol{Y}, \boldsymbol{A})$ is created by extending $\mathcal{G}$ as follows:

1. For each variable $V_i \notin \boldsymbol{A} \cup \boldsymbol{Y}$ create a copy version of that variable, V_i^c, which shares the same functional mechanism and all exogenous or unobserved error terms. Add V_i^c to the graph.
2. For each new variable V_i^c:
   (a) Let V_i be the variable V_i^c was copied from.
   (b) For each $X \in \text{Pa}_{\mathcal{G}}(V_i)$, if $X \in \boldsymbol{A} \cup \boldsymbol{Y}$ then V_i^c inherits X as a parent. If $X \notin \boldsymbol{A} \cup \boldsymbol{Y}$ then V_i^c inherits the copied version X^c as a parent.
   (c) For each $X \in \text{Ch}_{\mathcal{G}}(V_i)$, if $X \in \boldsymbol{Y}$ then V_i^c inherits X as a child. If $X \notin \boldsymbol{A} \cup \boldsymbol{Y}$ then V_i^c inherits the copied version X^c as a child. V_i^c does not inherit as children any variables in $\boldsymbol{A}$
   (d) If V_i has any $\to$ edge to an element of $\boldsymbol{Y}$, remove it.

The representation can be seen as a special case of the counterfactual graph representations developed in Shpitser & Pearl (2007). The returned graph $\mathcal{G}'$ will also be an ADMG defined over a set of variables $\boldsymbol{V'} \setminus \boldsymbol{C} \cup \boldsymbol{C}$, where the variables $\boldsymbol{C} \subset \boldsymbol{V'}$ denotes the "copy" variables in the graph; the variables V_i^c described above. Since G' is also a functional causal model, it also implies a distribution $p'(\boldsymbol{V'})$ over its variables. Given a graph $\mathcal{G}$ and a CWPG built from it, $\mathcal{G}'(\mathcal{G}, \boldsymbol{Y}, \boldsymbol{A})$, each observed variable $V_i \in \boldsymbol{V}$ will have a corresponding version of itself, $V_i' \in \boldsymbol{V'} \setminus \boldsymbol{C}$ with the same functional mechanism. The variable V_i' may also have copies of itself in the set $\boldsymbol{C}$ (these copies will also have the same functional mechanism as V_i). Then V_i', (and its possible copy $V_i^c \in \boldsymbol{C}$) are the *counterparts* of V_i in $\mathcal{G}'$, denoted as $\text{CP}_{\mathcal{G}'}(V_i)$. Define the vice-versa direction, $\text{CP}_{\mathcal{G}}(V_i')$, similarly for $V_i' \in \boldsymbol{V'}$. Nodes that are counterparts share the same functional mechanism and exogenous/unobserved noise. For a CWPG, we have the following useful properties:

Proposition 1. *For a CWPG $\mathcal{G}'(\mathcal{G}, \boldsymbol{Y}, \boldsymbol{A})$ built from $\mathcal{G}$ with observed variables $\boldsymbol{V}$, we have:*

1. *For any $V'_i \in \boldsymbol{V'}$, $|CP_{\mathcal{G}}(V'_i)| = 1$*
2. *For any $V'_i \in \boldsymbol{V'}$, $|Pa_{\mathcal{G}'}(V'_i)| = |Pa_{\mathcal{G}}(CP_{\mathcal{G}}(V'_i))|$*
3. *For any $V_i \in \boldsymbol{V}$ and any $V'_i \in CP_{\mathcal{G}'}(V_i)$ we have that $\bigcup_{V'_j \in Pa_{\mathcal{G}'}(V'_i)} CP_{\mathcal{G}}(V'_j) = Pa_{\mathcal{G}}(V_i)$*

The above Propositions give us the guarantee of a one to one function from the counterparts of the parents of some variable V'_i in $\mathcal{G}'$, and the parents of $\text{CP}_{\mathcal{G}}(V'_i)$, a fact that is important to keep in mind when proving functional equivalences.

As we will show, the construction of $\mathcal{G}'$ will allow us to rephrase any cross world policy intervention from the original model $\mathcal{G}$ in terms of a specific edge policy intervention on $\mathcal{G}'$. After this reconstrual, we can use existing theory from Shpitser & Sherman (2018) to develop complete identification criteria.

In order to get to this point, we must first prove that identification in one model implies identification in the other. The following two lemmas will prove useful towards this (see Appendix for all proofs):

Lemma 1. *Let $p(\boldsymbol{V})$ be a distribution nested Markov relative to an ADMG $\mathcal{G}$. Fix a set of outcome variables, $\boldsymbol{Y} \subset \boldsymbol{V}$, and treatment variables, $\boldsymbol{A} \subset \boldsymbol{V} \setminus \boldsymbol{Y}$, and build a CWPG $\mathcal{G}'(\mathcal{G}, \boldsymbol{Y}, \boldsymbol{A})$. Choose an assignment $\boldsymbol{v}$ for the variables in V, and let copy$(\boldsymbol{v})$ be an assignment to all variables in $\boldsymbol{V'}$ such that $V'_i = \boldsymbol{v}_{CP_{\mathcal{G}}(V'_i)}$ for all $V'_i \in \boldsymbol{V'}$. Then for all values of $\boldsymbol{v}$ we have:*

$$p(\boldsymbol{V} = \boldsymbol{v}) = p'(\boldsymbol{V'} = copy(\boldsymbol{v})) = \sum_{\boldsymbol{C}} p'(\boldsymbol{V'} \setminus \boldsymbol{C} = \boldsymbol{v}, \boldsymbol{C})$$

Lemma 2. *For a graph $\mathcal{G}$ and its CWPG $\mathcal{G}'$ with variables $\boldsymbol{V'}$, if an ETT path specific policy intervention, $\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}$, on $\mathcal{G}$ for treatments $\boldsymbol{A}$ and outcomes $\boldsymbol{Y}$ is expressible as an edge specific policy intervention $\boldsymbol{f}_\alpha$, we have:*

$$p(\boldsymbol{Y}(\boldsymbol{f}_\alpha)) = p'(\boldsymbol{Y'}(\boldsymbol{f}'_{\alpha'}))$$

where $\boldsymbol{Y}'$ are the counterparts of $\boldsymbol{Y}$ in $\mathcal{G}'$, α is the set of all edges in $\mathcal{G}$ starting a path in $\pi_{\boldsymbol{X},\boldsymbol{Y}}$, α' is the set $\{(A'X)'_{\rightarrow} | X \in \boldsymbol{C} \cup \boldsymbol{Y}'\}$ in $\mathcal{G}'$, and $f_A^{(AX)\rightarrow} \in \boldsymbol{f}$ is functionally equivalent to $f_{A'}^{(A'C)'_{\rightarrow}} \in \boldsymbol{f'}$ when $CP_{\mathcal{G}'}(A) = A'$

The above allows us to easily establish the following :

Theorem 1. *Let α' be the set of edges $\{(A'X)'_{\rightarrow} | A \in \boldsymbol{A}'; X \in \boldsymbol{C} \cup \boldsymbol{Y}'\}$ in $\mathcal{G}'$ and let $\boldsymbol{f}'_{\alpha'}$ be an edge policy intervention in $\mathcal{G}'$ on edges in α'. Assume that $\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}$ is expressible as an edge intervention $\boldsymbol{f}_\alpha$ and $f_{A_i} \in \boldsymbol{f}$ is functionally equal to $f'_{A'_i} \in \boldsymbol{f'}$ when $CP_{\mathcal{G}'}(A_i) = A'_i$. Then $p(\boldsymbol{Y}(\boldsymbol{f}_\alpha))$ is identified if and only if $p'(\boldsymbol{Y}'(\boldsymbol{f}'_{\alpha'}))$ is.*

3.3 Complete Identification Criteria for Cross World Policies

In the previous section we reduced the problem of cross world policy identification to one of identification of an edge specific policy $p'(\boldsymbol{Y'}(f_{\alpha'}))$ on $\mathcal{G}'$ (given the path intervention can be expressed as such). For the following theorems, let $\mathcal{G}' = \mathcal{G}'(\mathcal{G}, \boldsymbol{Y}, \boldsymbol{A})$ be the CWPG built from $\mathcal{G}$. Let $\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}$ be our CWP formulated as a path intervention, $\boldsymbol{f}_\alpha$ its respective edge intervention (if applicable), and $\boldsymbol{f}'_{\alpha'}$ the corresponding edge intervention on $\mathcal{G}'$. Furthermore, let $\mathcal{G}'_{\boldsymbol{f}'_{\alpha'}}$ be a graph created from $\mathcal{G}'$ as follows: For each A'_i assigned by function f_i with input variables $W_i = W_i^f \cup W_i^{\mathfrak{N}}$, add an edge from $V'_k \in \boldsymbol{V}' \setminus \boldsymbol{C}$ to A'_i if $V_k \in W_i^{\mathfrak{N}}$ (ie are natural valued) and add an edge from $V_k^c \in \boldsymbol{C}$ to A'_i if $V_k^c \in W_i^f$. Define $\boldsymbol{Y}^* = \text{an}_{\mathcal{G}'_{\boldsymbol{f}'_{\alpha'}}}(\boldsymbol{Y}')$. Then we have the following:

Theorem 2. *The Cross World Policy response $p(Y(\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}))$ is identified from $p(\boldsymbol{V})$ if and only if the following hold:*

- *$\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}$ is expressible as an edge intervention $\boldsymbol{f}_\alpha$ on $\mathcal{G}$*
- *$Ch_{\mathcal{G}'_{\boldsymbol{Y}^*}}(A'_i) \cap Dis_{\mathcal{G}'_{\boldsymbol{Y}^*}}(A'_i) = \emptyset$ for all $A'_i \in \boldsymbol{A}'$,*
- *No districts $D \in \mathcal{D}((\mathcal{G}'_{\boldsymbol{f}'_{\alpha'}})_{\boldsymbol{Y}^*})$ contain both a variable $V'_i \in \boldsymbol{V}' \setminus (\boldsymbol{C} \cup \boldsymbol{Y}')$ and a variable $V'_j \in \boldsymbol{C} \cup \boldsymbol{Y}'$*

If these hold, then the identifying formula is:

$$\sum_{\boldsymbol{Y}^* \setminus \boldsymbol{Y}'} \prod_{D \in \mathcal{D}(\mathcal{G}'_{\boldsymbol{Y}^*})} \phi_{\boldsymbol{V}' \setminus D}(p(\boldsymbol{V}'), \mathcal{G}')\Big|_{\{A'_i = f'_{A_i}(W'_i) | A'_i \in \boldsymbol{A}' \cap Pa^Y(D)\}} \tag{2}$$

Where $Pa^Y(D)$ are parents of D along the edges $\{(A'X)'_{\rightarrow} | A' \in \boldsymbol{A}'; X \in \boldsymbol{C} \cup \boldsymbol{Y}'\}$ and W'_i are the inputs to the policy f'_{A_i}.

4 Conclusion

In this work we defined counterfactual responses to Cross-World policies, a type of dynamic treatment regime whose input dependencies may be cross world, generalizing prior work done in Sani et al. (2020). We gave examples of the types of questions that may be answered as Cross-World policy responses. We showed how the subtleties of dealing with cross world dependencies could be dealt with via a graphical construction that makes explicit these dependencies. With this tool, we developed complete non parametric identification criteria for responses to Cross-World policies.

References

Daniel Malinsky, Ilya Shpitser, and Thomas Richardson. A potential outcomes calculus for identifying conditional path-specific effects. *AISTATS 2019 - 22nd International Conference on Artificial Intelligence and Statistics*, 89, 2020.

Razieh Nabi, Phyllis Kanki, and Ilya Shpitser. Estimation of personalized effects associated with causal pathways. In *Uncertainty in artificial intelligence: proceedings of the... conference. Conference on Uncertainty in Artificial Intelligence*, volume 2018. NIH Public Access, 2018.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Thomas S Richardson and James M Robins. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128(30):2013, 2013.

Thomas S Richardson, Robin J Evans, James M Robins, and Ilya Shpitser. Nested markov properties for acyclic directed mixed graphs. *arXiv preprint arXiv:1701.06686*, 2017.

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

James M Robins. Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality*, pp. 69–117. Springer, 1997.

James M Robins, Miguel A Hernán, and Uwe Siebert. Effects of multiple interventions. *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors*, 1:2191–2230, 2004.

Numair Sani, Jaron J R Lee, and Ilya Shpitser. Identification and Estimation of Causal Effects Defined by Shift Interventions. *UAI*, 2020.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference, AAAI-06/IAAI-06*, pp. 1219–1226, 2006.

Ilya Shpitser and Judea Pearl. What counterfactuals can be tested. In *23rd Conference on Uncertainty in Artificial Intelligence, UAI 2007*, pp. 352–359, 2007.

Ilya Shpitser and Eli Sherman. Identification of personalized effects associated with causal pathways. *34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018*, 1:530–539, 2018. ISSN 1525-3384.

Ilya Shpitser and Eric Tchetgen Tchetgen. Causal inference with a graphical hierarchy of interventions. *Annals of statistics*, 44(6):2433, 2016.

Jin Tian. Identifying dynamic sequential plans. *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, UAI 2008*, pp. 554–561, 2008.

Jessica G Young, Miguel A Hernán, and James M Robins. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic methods*, 3(1):1–19, 2014.

Appendix

Proofs

Proposition 2. *For a CWPG $\mathcal{G}'(\mathcal{G}, \boldsymbol{Y}, \boldsymbol{A})$ built from $\mathcal{G}$ with observed variables $\boldsymbol{V}$, we have:*

1. *For any $V_i' \in \boldsymbol{V'}$, $|CP_{\mathcal{G}}(V_i')| = 1$*
2. *For any $V_i' \in \boldsymbol{V'}$, $|Pa_{\mathcal{G}'}(V_i')| = |Pa_{\mathcal{G}}(CP_{\mathcal{G}}(V_i'))|$*
3. *For any $V_i \in \boldsymbol{V}$ and any $V_i' \in CP_{\mathcal{G}'}(V_i)$ we have that $\bigcup_{V_j' \in Pa_{\mathcal{G}'}(V_i')} CP_{\mathcal{G}}(V_j') = Pa_{\mathcal{G}}(V_i)$*

Proof. (1) follows directly from the construction of a CWPG. (2) follows from steps 2c and 2d in the CWPG construction, which enforces that for any $V_i' \in \boldsymbol{V'}$ that has a copied version $V_i^c \in \boldsymbol{C}$, $\text{Ch}_{\mathcal{G}'}(V_i') \cap \text{Ch}_{\mathcal{G}'}(V_i^c) = \emptyset$; so the addition of the copy nodes do not add any "extra" incoming edges to any variable in the graph. (3) follows from the construction of the CWPG, together with (1) and (2). □

The above Propositions give us the guarantee of a one to one function from the counterparts of the parents of some variable V_i' in $\mathcal{G}'$, and the parents of $\text{CP}_{\mathcal{G}}(V_i')$, a fact that is important to keep in mind when proving functional equivalences.

Lemma 3. *Let $p(\boldsymbol{V})$ be a distribution nested Markov relative to an ADMG $\mathcal{G}$. Fix a set of outcome variables, $\boldsymbol{Y} \subset \boldsymbol{V}$, and treatment variables, $\boldsymbol{A} \subset \boldsymbol{V} \setminus \boldsymbol{Y}$, and build a CWPG $\mathcal{G}'(\mathcal{G}, \boldsymbol{Y}, \boldsymbol{A})$. Choose an assignment $\boldsymbol{v}$ for the variables in V, and let copy$(\boldsymbol{v})$ be an assignment that copies the assignments of $\boldsymbol{v}$ to the corresponding variables and copy variables in $\boldsymbol{V'} = \boldsymbol{V} \setminus \boldsymbol{C} \cup \boldsymbol{C}$. Then for all values of $\boldsymbol{v}$ we have:*

$$p(\boldsymbol{V} = \boldsymbol{v}) = p'(\boldsymbol{V'} = copy(\boldsymbol{v})) = \sum_{\boldsymbol{C}} p'(\boldsymbol{V'} \setminus \boldsymbol{C} = \boldsymbol{v}, \boldsymbol{C})$$

Proof. Since we assume a functional causal model, the settings of the variables are deterministic functions of the exogenous error variables $\boldsymbol{\epsilon}$. By construction of $\mathcal{G}'$ shares the same distribution over the error variables, $p(\boldsymbol{\epsilon})$. Set $\boldsymbol{\epsilon}$ to the same value in both $\mathcal{G}$ and $\mathcal{G}'$ models. Now pick an observed variable $V_i' \in \boldsymbol{V'} \setminus \boldsymbol{C}$ which has no observed parents; the existence of such a variable follows from the acyclicity of $\mathcal{G}'$. By construction, their also exists a variable V_i in $\mathcal{G}$ which has the same assignment function, $\mathcal{F}_i$ as V_i'. Since V_i and V_i' share the same assignment function that only depends on the fixed error terms, and the error terms are fixed to the same values in both graphs, it follows that $V_i = V_i'$ whenever the error variables have the same values for both models. Set both variables to the value they take, and add them to the set F of fixed variables. If V_i' has a copy in $\boldsymbol{C}$ then it also takes on the same value due to functional equivalence. Set it to this value and add it to F.

Now pick a new observed variable $V_j' \in \boldsymbol{V'} \setminus \boldsymbol{C}$ which either has no observed parents, or all of its parents are in F.

We can prove the existence of such a variable as follows: Assume no such variable exists, and not all variables in $\mathcal{G}'$ have been added to F. Then there must exist some variable V_k' in $\mathcal{G}'$ which has a parent not in F. This parent must also have a parent not in F. Continuing this recursion, we will eventually reach a node with either no parents, or with all parent in F (by the finiteness of our variable set and acyclicity), which contradicts the initial assumption. Thus such a variable must exist, or all variables in the graph are in F.

Thus the variable V_j', as well as its analogue in $\mathcal{G}$, V_j, both have parents in F set to the same value. From the shared assignment function between V_j' and V_j, and the equal setting of error variables across both models, it follows that $V_j' = V_j$. We can add these variables to F, along with any copied version of V_j', and repeat this process until all variables are in F.

From this it follows that whenever the error variables are set the same in $\mathcal{G}$ and $\mathcal{G}'$, then $\boldsymbol{V} = \boldsymbol{v}$ implies $\boldsymbol{V'} = \text{copy}(\boldsymbol{v})$. Since the distribution over error variables is the same in both models, it follows that $p(\boldsymbol{V} = \boldsymbol{v}) = p'(\boldsymbol{V'} = \text{copy}(\boldsymbol{v}))$. That $p'(\boldsymbol{V'} = \text{copy}(\boldsymbol{v})) = \sum_{\boldsymbol{C}} p'(\boldsymbol{V'} \setminus \boldsymbol{C} = \boldsymbol{v}, \boldsymbol{C})$ follows from the fact that the nodes in C must take on the same value as the node they are a copy of in $\boldsymbol{V'} \setminus \boldsymbol{C}$, so all other assignment have zero probability mass in the observed distribution. □

Lemma 4. *For a graph $\mathcal{G}$ and its CWPG $\mathcal{G}'$ with variables $\boldsymbol{V'}$, if an ETT path specific policy intervention, $\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}$, on $\mathcal{G}$ for treatments $\boldsymbol{A}$ and outcomes $\boldsymbol{Y}$ is expressible as an edge specific policy intervention $\boldsymbol{f}_\alpha$, we have:*

$$p(\boldsymbol{Y}(\boldsymbol{f}_\alpha)) = p'(\boldsymbol{Y'}(\boldsymbol{f}'_{\alpha'}))$$

where $\boldsymbol{Y}'$ are the counterparts of $\boldsymbol{Y}$ in $\mathcal{G}'$, α is the set of all edges in $\mathcal{G}$ starting a path in $\pi_{\boldsymbol{X},\boldsymbol{Y}}$, α' is the set $\{(A'X)'_{\rightarrow} | X \in \boldsymbol{C} \cup \boldsymbol{Y}'\}$ in $\mathcal{G}'$, and $f_A^{(AX)_\rightarrow} \in \boldsymbol{f}$ is functionally equivalent to $f_{A'}^{(A'C)'_\rightarrow} \in \boldsymbol{f}'$ when $CP_{\mathcal{G}'}(A) = A'$

Proof. Under the recursive substitution definition for counterfactuals, $\boldsymbol{Y'}(f_{\alpha'})$ is equal to:

$$Y'(f_{\alpha'}) = Y'(\{f_{A_i'}^{(A_i'Y')_\rightarrow}(W_i) | A_i' \in \text{Pa}_{\mathcal{G}'}(Y') \cap \boldsymbol{A'}\}, \{V_i(f_{\alpha'}) | V_i \in \text{pa}_{\mathcal{G}}(Y') \setminus \boldsymbol{A'}\})$$

Similarly, the counterfactual $Y(f_{\alpha^*})$ is defined as:

$$Y(f_\alpha) = Y(\{f_{A_i}^{(A_iY)_\rightarrow}(W_i) | A_i \in \text{Pa}_{\mathcal{G}}(Y) \cap \boldsymbol{A}\}, \{V_i(f_\alpha) | V_i \in \text{pa}_{\mathcal{G}}(Y) \setminus \boldsymbol{A}\})$$

From Proposition 1, and step (d) in the CWPG construction, one can see that $|\text{Pa}_{\mathcal{G}}(Y_i)| = |\text{Pa}_{\mathcal{G}'}(Y_i')|$, and that each parent of Y_i has a one counterpart in $\text{Pa}_{\mathcal{G}'}(Y_i')$. As such, to complete the proof we will show that the terms appearing in the recursive substitution definition of $Y'(f_{\alpha'})$ are functionally equivalent to their counterparts in $Y(f_\alpha)$. In $\mathcal{G}'$, the parents of Y_i' may be of one of three kinds of variables: a treatment variable in $\boldsymbol{A}'$, a copy variable in $\boldsymbol{C}$, or an outcome variable in $\boldsymbol{Y}'$. Since the elements of $\boldsymbol{Y}'$ are fixed for the purpose of evaluating a probability, the parents of Y_i' that are elements of $\boldsymbol{Y'}$ are vacuously equivalent to their counterparts in $\mathcal{G}$.

For the variables in A', we have by construction that they are set according to policy functions $f_{A_i'}^{(A_i'Y)_\rightarrow}(W_i')$, and their counterparts which are parents of Y in $\mathcal{G}$ are set by equivalent functions $f_{A_i}^{(A_iY)_\rightarrow}(W_i)$. Thus, the assignments to parents belonging in the treatment set will be identical across graphs as long as their inputs, W_i and W_i', are the same.

We now look at variables in $\boldsymbol{C}$ which are parents of $\boldsymbol{Y}'$. By recursive substitution, all variables $C \in \text{Pa}(Y') \cap \boldsymbol{C}$ are set to $C(f_{\alpha'})$. For some C, if it does not have parents, then it is functionally equivalent to its counterpart in $\mathcal{G}$ (since its only dependence is on the exogenous noise terms which are the same across graphs). If C has parents its parents are either: (1) a variable $A_i' \in \boldsymbol{A'}$ whose value is set by a policy function $f_{A_i'}^{(A_i'Y)_\rightarrow}(W_i')$, (2) an element of Y' which is fixed and identical across graphs, or (3) another variable in $\boldsymbol{C}$. This nesting ends only in the case of no parents, or in the case of (1) or (2). Due to acyclicity and finiteness, all terms in the recursive substitution definition of $C(f_{\alpha'})$ will eventually reach either the no parents case, case (1), or (2). Since assignments in the no parents case and in case (2) are functionally equivalent across graphs, we will have that $C(f_{\alpha'})$ is functionally equivalent to its counterpart in $\mathcal{G}$ as long as policy assignments to variables in $\boldsymbol{A'}$ are equivalent to there counterparts. Since the counterpart for a treatment variable shares an assignment function $f_{A_i'} = f_{A_i}$, this will be the case if we can show that the inputs to these functions are the same across graphs. If we show this, then the parents of Y' are functionally equivalent to their counterparts in $\mathcal{G}$ under the edge policy intervention, and the counterfactuals $Y_i(f_\alpha)$ and $Y_i'(f_{\alpha'})$ are thus functionally equivalent.

Recall that the inputs to a policy may either be the original, naturally occurring value of the variable V_i, or the counterfactual version of it under the policy $V_i(f)$. In $\mathcal{G}'$ under our assumed edge intervention, the former is represented explicitly by variables in $\boldsymbol{V}' \setminus \boldsymbol{C}$, while the later is represented explicitly by variables in $\boldsymbol{C}$. If the input corresponds to the naturally occurring value of the variable, then the input is functionally equivalent across graphs by the proof of Lemma 1.

If the input is a counterfactual, then look at a treatment variable A'_i in $\mathcal{G}'$ which has no other treatment variable as an ancestor (that such a variable exists comes from the acyclicity of the graph). The inputs to the policy function that assigns treatments to A'_i must all be natural valued variables, or counterfactuals that are equivalent to the natural value of the variable. This follows from rule 3 of the PO calculus (Malinsky et al., 2020). Thus, the output of the policy function on A'_i for edges going to nodes in $\boldsymbol{C}$ is functionally equal the output of the policy function on its counterpart in $\mathcal{G}$. We can now reuse the proof of Lemma 1 to pick a new variable A'_j which, if it has a treatment variable as an ancestor, it is a treatment variable whose policy inputs have been shown equivalent across graphs, which proves input equivalence across graphs for the assignment policy on A'_j. From this, input equivalence for all treatment assignment policies follows.

Thus, $Y_i(f_\alpha)$ and $Y'_i(f_{\alpha'})$ are functionally equivalent for any Y_i and counterpart Y'_i. From the equivalence of the error terms and their respective distribution, we thus have that

$$p(\boldsymbol{Y}(f_\alpha)) = p'(\boldsymbol{Y'}(f_{\alpha'}))$$

□

Theorem 3. *Let α' be the set of edges $\{(AX)_{\rightarrow} | A \in \boldsymbol{A'}; X \in \boldsymbol{C} \cup \boldsymbol{Y'}\}$ in $\mathcal{G}'$ and let $\boldsymbol{f}'_{\alpha'}$ be an edge policy intervention in $\mathcal{G}'$ on edges in α'. Assume that $\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}$ is expressible as an edge intervention $\boldsymbol{f}_\alpha$ and $f_{A_i} \in \boldsymbol{f}$ is functionally equal to $f'_{A'_i} \in \boldsymbol{f'}$ when $CP_{\mathcal{G}'}(A_i) = A'_i$. Then $\boldsymbol{f}_\alpha$ is identified if and only if $\boldsymbol{f}'_{\alpha'}$ is.*

Proof. Assume that $\boldsymbol{f}_\alpha$ is not identified. Then there exists two parameterizations M_1 and M_2 for $\mathcal{G}$ such that $p_1(\boldsymbol{V}) = p_2(\boldsymbol{V})$ but $p_1(\boldsymbol{Y}(\boldsymbol{f}_\alpha)) \neq p_2(\boldsymbol{Y}(\boldsymbol{f}_\alpha))$, where p_i is the distribution under model M_i. Create a CWPG $\mathcal{G}'$ with variables $\boldsymbol{V}'$. Then by Lemma 1, we have $p'_1(\boldsymbol{V}') = p'_2(\boldsymbol{V}')$. By Lemma 2 we have $p'_1(\boldsymbol{Y}'(\boldsymbol{f}'_{\alpha'})) = p_1(\boldsymbol{Y}(\boldsymbol{f}_\alpha))$ and $p'_2(\boldsymbol{Y}'(\boldsymbol{f}'_{\alpha'})) = p_2(\boldsymbol{Y}(\boldsymbol{f}_\alpha))$ and thus $p'_1(\boldsymbol{Y}'(\boldsymbol{f}'_{\alpha'})) \neq p'_2(\boldsymbol{Y}'(\boldsymbol{f}'_{\alpha'}))$. Thus $\boldsymbol{f}'_{\alpha'}$ is not identified.

Now assume that $\boldsymbol{f}'_{\alpha'}$ is not identified. Then follow a similar proof above to show that this implies $\boldsymbol{f}_\alpha$ is not identified. From this, our result follows. □

Theorem 4. *The Cross World Policy response $p(Y(\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}))$ is identified from $p(\boldsymbol{V})$ if and only if the following hold:*

- *$\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}$ is expressible as an edge intervention $\boldsymbol{f}_\alpha$ on $\mathcal{G}$*
- *$Ch_{\mathcal{G}'_{\boldsymbol{Y}^*}}(A'_i) \cap Dis_{\mathcal{G}'_{\boldsymbol{Y}^*}}(A'_i) = \emptyset$ for all $A'_i \in \boldsymbol{A'}$,*
- *No districts $D \in \mathcal{D}((\mathcal{G}'_{\boldsymbol{f}'_{\alpha'}})_{\boldsymbol{Y}^*})$ contain both a variable $V'_i \in \boldsymbol{V'} \setminus (\boldsymbol{C} \cup \boldsymbol{Y'})$ and a variable $V'_j \in \boldsymbol{C} \cup \boldsymbol{Y'}$*

If these hold, then the identifying formula is:

$$\sum_{\boldsymbol{Y}^* \setminus \boldsymbol{Y}'} \prod_{D \in \mathcal{D}(\mathcal{G}'_{\boldsymbol{Y}^*})} \phi_{\boldsymbol{V}' \setminus D}(p(\boldsymbol{V'}), \mathcal{G}')\Big|_{\{A'_i = f'_{A_i}(W'_i) | A'_i \in \boldsymbol{A}' \cap Pa^Y(D)\}} \tag{3}$$

Where $Pa^Y(D)$ are parents of D along the edges $\{(A'X)'_{\rightarrow} | A' \in \boldsymbol{A'}; X \in \boldsymbol{C} \cup \boldsymbol{Y'}\}$

Proof. By Theorem 3, proving identification for the Cross World Policy response $p(Y(\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}))$ in $\mathcal{G}$ can be done by proving identification for $p'(Y'(\boldsymbol{f}'_{\alpha'}))$ on the graph $\mathcal{G}'$.

The first thing to note is the construction of the graph $\mathcal{G}'_{\boldsymbol{f}'_{\alpha'}}$ which represents the graph under the policy intervention and will determine the variables belonging in $\boldsymbol{Y}^*$. Contrary to usual constructions, no edges are removed in the creation of this graph, and some may be added. To see that this will always be the correct construction for $\mathcal{G}'_{\boldsymbol{f}'_{\alpha'}}$ note that (1) We can construe $\boldsymbol{f}'_{\alpha'}$ as additionally defining policies on edges $\{(A'X)'_{\rightarrow} | A' \in \boldsymbol{A'}; X \notin \boldsymbol{C} \cup \boldsymbol{Y'}\}$ which set A' to its natural value, and (2) we assume for all $f'_i \in \boldsymbol{f'}$ we have $W_i^{\mathfrak{N}} \neq \emptyset$. From this we can see that all nodes will have some edge with a policy which assigns the natural value. Since the inputs to this "policy" are simply the original inputs to A' in $\mathcal{G}'$, we do not remove any edges to A' in the original graph (though edges may be added).

Given this, we first show that the three conditions of Theorem 4 imply identification of $\boldsymbol{f}'_{\alpha'}$. Since our expression is originally a path intervention a necessary condition for identification is for the first point be true; that $\boldsymbol{f}_{\pi_{\boldsymbol{A},\boldsymbol{Y}}}$ be expressible as an edge intervention $\boldsymbol{f}_{\alpha}$ (Theorem 5.2 of Shpitser & Tchetgen (2016)). To prove that $p'(Y'(\boldsymbol{f}'_{\alpha'}))$ (and hence $p(Y(\boldsymbol{f}_{\alpha}))$) is identified, we use Theorem 3 from Shpitser & Sherman (2018), which implies identification for edge policies if (1) $\boldsymbol{Y}^*(\boldsymbol{A}' = \boldsymbol{a}')$ is identified for any assignment $\boldsymbol{a}'$, and (2) the edge assignments to the districts in $\mathcal{G}'_{\boldsymbol{f}'_{\alpha'}}$ are consistent (there exist no recanting districts).

If the second point of Theorem 4 is true, then this will imply that $\boldsymbol{Y}^*(\boldsymbol{A}')$ is identified, using the same observation used in Theorem 3 of Sani et al. (2020) (that Theorem 60 of Richardson et al. (2017), the one line ID algorithm, is valid even when $\boldsymbol{A}$ and $\boldsymbol{Y}$ are not disjoint).

If the third condition of Theorem 4 holds then it will follow that there will be no districts $D \in \mathcal{D}((\mathcal{G}'_{\boldsymbol{f}'_{\alpha'}})_{\boldsymbol{Y}^*})$ with inconsistent edge assignments. In our setting, only edges from some A' to an element of $\boldsymbol{C} \cup \boldsymbol{Y}'$ will have an assignment different from the natural value, hence, only the existence of a variable $V'_i \in \boldsymbol{V}' \setminus (\boldsymbol{C} \cup \boldsymbol{Y}')$ and $V'_j \in \boldsymbol{C} \cup \boldsymbol{Y}'$ in the same district will lead to an inconsistent edge assignment. Thus the conditions of Theorem 4 imply identification of $p'(Y'(\boldsymbol{f}'_{\alpha'}))$, with the identifying formula Equation 3 following from Theorem 60 in Richardson et al. (2017).

We now show the other direction. If this first condition is violated, we are unidentified by Theorem 5.2 in Shpitser & Tchetgen (2016). If the second condition is violated, then we can take advantage of the fact that shift interventions on the treated (SITs) are a special case of the cross world policies defined here, and use the construction used in the proof of Theorem 4 in Sani et al. (2020) to show non-identification for SITs when this condition is violated. If the third condition is violated then by Theorem 7 of Shpitser & Sherman (2018), the edge policy intervention is not identified. Thus, violation of any of the conditions of Theorem 4 implies non identification. □