Causal Modeling for Fairness in Dynamical Systems

Elliot Creager*, David Madras, Toniann Pitassi, Richard Zemel

University of Toronto

Vector Institute

ABSTRACT

In this work, we present causal directed acyclic graphs (DAGs) as a unifying framework for the recent literature on fairness in dynamical systems. We advocate for the use of causal DAGs as a tool in both designing equitable policies and estimating their impacts. By visualizing models of dynamic unfairness graphically, we expose implicit causal assumptions which can then be more easily interpreted and scrutinized by domain experts. We demonstrate that this method of reinterpretation can be used to critique the robustness of an existing model/policy, or uncover new policy evaluation questions. Causal models also enable a rich set of options for evaluating a new candidate policy without incurring the risk of implementing the policy in the real world. We close the paper with causal analyses of several models from the recent literature, and provide an in-depth case study to demonstrate the utility of causal DAGs for modeling fairness in dynamical systems.

KEYWORDS

causal modeling, fairness, machine learning, dynamical systems, counterfactual generation, off-policy evaluation
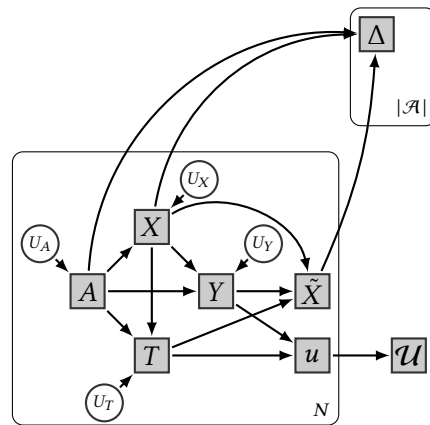

Figure 1: A structural causal model (SCM) formulation of the one-step model from Liu et al. [39]. We argue that formulating models of fairness in dynamical systems as causal DAGs enables a range of new questions about the model, in turn allowing more effective interpretation and communication, facilitating a range of model critiques and extensions, and yielding more robust policy evaluation. See Section 5 for further explanation of this figure, and Table 1 in Appendix B for symbol legend.

1 INTRODUCTION

How do we design equitable policies for complex, evolving societies? A wide range of work in the social sciences aims to understand the long-term consequences of decisions and events [2, 15, 17, 18, 33, 58]. Recently, the literature on fairness in dynamical systems (a.k.a. "feedback loops") has begun exploring the role of algorithmic systems in shaping their environments over time [3, 9, 14, 21, 24, 31, 40, 46]. The key insight from this literature is that the repeated application of algorithmic tools in a changing environment can have long-term fairness impacts distinct from their short-term impacts. Each paper in this literature proposes a dynamics model for a particular domain (e.g. lending [46], hiring [24], recommendations [3]), exposing unfairness that arises from long-term reapplication of a baseline policy, and then proposes a "fair" policy to mitigate some of these biases. These works have moved forward in moving the fair machine learning literature towards more realistic modeling of a dynamic, changing world.

However, the study of fairness in dynamical systems is, at this point, quite disparate, with little overlap existing between various works in terms of modeling choices, goals, or sets of assumptions. Part of this is to be expected: context is important when considering fairness, and different problem setups will lead to different proposed

solutions. However, a common language for expressing these models would nonetheless be helpful for simplifying the process of building on and synthesizing past work.

In this work, we propose causal DAGs (directed acyclic graphs) [7, 50, 53] as just such a common language (see Figure 1). We claim that the field of fairness in dynamical systems can be unified through the lens of causal DAGs, and advocate for this form of modeling when approaching this class of problems. While causal DAGs have been used to study one-shot fair decision-making [36, 37], they are uncommon in the fairness settings involving *sequential* decisions. In this context, we argue causal DAGs provide two distinct types of advantages: they provide a tool for interfacing with policymakers, enabling communication of frequently implicit causal exposition; and they enable technologists to investigate new candidate policies and evaluate their merit. Our contributions are:

- We show that causal DAGs are a unifying framework for the literature on fairness in dynamical systems. Using examples, we show that many dynamics models and proposed "fair" policies from this literature can be reformulated equivalently as *structural causal models* and *policy interventions*, respectively.
- We show that this framework enables inquiry (inspection and evaluation) into a rich set of further policy questions which were not explicit in each model's original formulation.

- Through a detailed case study of Liu et al. [39], we show how to refactor an existing model as a causal DAG. To demonstrate the flexibility of our proposed framework, we then extend the model to the fully dynamic multi-step scenario, and highlight novel interventions that capture interactions between multiple actors.

This paper proceeds as follows. In Section 2, we discuss background literature. In Section 3, we introduce our main argument regarding the advantages of using causal DAGs for these modeling problems. In Section 4, we present a simple example of SCM design and policy evaluation. In Section 5, we review this literature through a causal DAG lens, reformulating several previously proposed models in this manner. In Section 6, we provide a case study of one such model, using off-policy evaluation and SCMs to explore this model's causal assumptions and implications. Finally, in Section 7, we discuss various extra topics and future work.

2 BACKGROUND & RELATED WORK

2.1 Fairness

Dynamical Systems. There has been work on modeling the long-term dynamics of fairness in a range of potential domains. Recently, the first paper to bring these issues to light was Lum and Isaac [40], discussing the bias feedback loops which could arise in predictive policing systems, with follow-up work by Ensign et al. [9]. Domains such as hiring [24], loans [46], and recommender systems [3, 21] have also been explored in this way. Other related explorations have dealt with short-term dynamics [39] and strategic actions [25, 44]. There is also a line of work studying the long-term effects of affirmative action, with some classic works from the economics literature [6, 12], and more recent computer science focused work [31]. On the theoretical side, several general algorithms for improved fairness in sequential decision-making have been characterized, with work discussing bandits [29], reinforcement learning [27], and importance sampling estimators [8].

Causality. Causal modeling has been used in a variety of non-dynamic fair machine learning approaches. Work on counterfactual fairness [37] has considered fairness definitions which encourage models to treat examples similarly to hypothetical situations where they were from the other group. Some other works focus on learning fair policies [36, 42, 48] or on learning decision rules which follow only causal paths deemed to be non-discriminatory [5, 32]. Outside of fairness, Everitt et al. [10] propose using influence diagrams as a framework for understanding AI safety systems.

2.2 Structural Causal Models

There are several ways of encoding causal assumptions in DAG form (see Section 7 for a discussion of alternatives). In this paper, we use one specific type of causal DAG, structural causal models (SCMs) [50], which we overview here[1].

SCMs are similar to probabilistic graphical models (PGMs) [35]. They consist of nodes (random variables representing entities in the world) and edges (relationships between those entities). However, whereas PGMs only specify a set of conditional independence
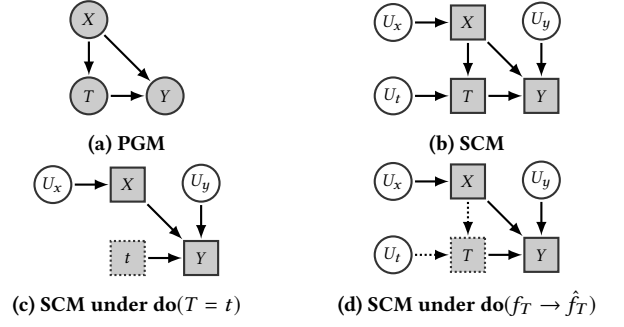
---

[1] Other overviews of various levels of detail can be found elsewhere [4, 42, 50]

(a) PGM **(b) SCM**

(c) SCM under do(T = t) **(d) SCM under do($f_T \to \hat{f}_T$)**

Figure 2: Treatment model expressed as PGM (2a) and SCM (2b). We also show the SCM under atomic intervention (2c) and policy intervention (2d).

relationships, SCMs specify a unique data generating process (analogously, a *particular* probability factorization, as opposed to the multiple isomorphic factorizations available in a PGM).

There are two types of nodes in SCMs: endogenous and exogenous variables. Endogenous variables are entities which are part of the model. Exogenous variables are external to the model; they represent all the relevant inputs to the endogenous variables. Exogenous variables are stochastic and are the exclusive source of randomness in the observations. The edges between variables are deterministic functions called *structural equations*. Hence, a setting of the exogenous variables corresponds to exactly one setting of the endogenous variables. In Figure 2b, the dark squares are endogenous variables, representing specific entities such as a credit score, a medical treatment, a sensitive attribute, or an employee's wages. Each endogenous variable is the output of a structural equation, e.g. $T = f_T(X, U_T)$, $Y = f_Y(T, X, U_Y)$. The light circles are exogenous variables, corresponding to stochasticity in the environment. Finally, *plates* sometimes enclose a set of endogenous and exogenous nodes, which denotes that multiple instances of the enclosed variables are generated i.i.d. from the same joint distribution[2].

We can calculate causal quantities under a particular SCM by using the *do*-operator. Given the probability distribution implied by the SCM in Figure 2b (call the model $\mathcal{M}$ and the implied joint distribution p), we may wish to ask – "What would be the expected value of Y if T was set to 1?" The corresponding estimand can be notated $\mathbb{E}_{p^{do(T=1)}}[Y]$. This differs from the more straightforward conditional probability $\mathbb{E}_p[Y|T = 1]$. The expression $\mathbb{E}_{p^{do(T=1)}}[Y]$ means we wish to calculate the expected value of Y under a modified SCM which is specified by $do(T = 1)$ (call it $\mathcal{M}^{do(T=1)}$, with the associated probability distribution $p^{do(T=1)}$). $\mathcal{M}^{do(T=1)}$ is intended to simulate a randomized experiment — if the true data-generating process is represented by $\mathcal{M}$, what would happen to the observed data if we forcibly change the data-generating process, so that $T = 1$ always? Graphically, $\mathcal{M}^{do(T=1)}$ is created by starting with

---

[2] The "sample size" of a plate appears in its lower-right corner, and specifies the number of i.i.d. draws. E.g., in Figure 1, the tuple $(U_A, A, U_X, X, U_Y, Y, U_T, T, \tilde{X}, u)$ is sampled N times (once for each of the N individuals), Δ is sampled $|\mathcal{A}|$ times (once for each group). There is only instance of $\mathcal{U}$. If a variable is inside two overlapping plates with sample sizes m and n, it is sampled mn times. A parent-to-child arrow between two non-overlapping plates (e.g. the arrow into Δ in Figure 1) indicates *each* child sample depends on *all* of the parent samples.

$\mathcal{M}$ (Fig. 2b), removing from the graph all the incoming arrows to T (in this case, arrows originating from X and U_T), and setting $T = 1$ (yielding Fig. 2c). This is referred to as an *intervention*. Under certain conditions [50], we can identify $\mathbb{E}_{p^{do(T=1)}}[Y]$ by using observational data generated by p to simulate sampling from $p^{do(T=1)}$. Intervening on the value of T in this way is an *atomic* intervention. Alternatively, we can intervene directly on the structural equation governing T (Fig. 2d), resulting in model $\mathcal{M}^{do(f_T \rightarrow \hat{f}_T)}$ with distribution $p^{do(f_T \rightarrow \hat{f}_T)}$ (a *policy* intervention). We discuss the implications of this type of intervention in Section 4.

3 ON THE ADVANTAGES OF CAUSAL DAGS

We assert that causal DAGs (for the purposes of this paper, SCMs) are a productive way to formulate models of dynamic fairness. We argue they enable three essential tasks:

(1) **Visualization:** Laying out a complex mathematical model graphically as an SCM is beneficial for compactly exposing the structure of the model and communicating its content to others, particularly stakeholders who may not have mathematical expertise.
(2) **Introspection:** Causal DAGs make explicit a number of implicit causal assumptions made by the modelers. This means that inspection of an already existing model by domain experts can yield candidates for further policies to consider, or alternatively allow for criticism of a model's robustness to changes in its assumptions.
(3) **Evaluation:** Interpreting a mathematical model as the structural equations governing an SCM allows for causal reasoning. Causal reasoning is useful for off-policy evaluation, which is essential for understanding the impact of policies in complex, high-stakes domains.

3.1 Visualization

Graphical models have been used to visualize algorithmic processes since the early days of programming. Their usage goes back at least as far as Goldstine and Von Neumann [16], who use "flow diagrams" to visualize the control flow of a program. Hand-drawn and automatically generated flowcharts were also frequently used in practice as a visualization tool [20, 34]. In the following decades, a range of work demonstrated the utility of this approach for program visualization [47]. Diagrammatic representations of programs were shown to help with understanding written code [43], algorithms [54], and written (non-programmatic) instructions [30], particularly for more complex algorithms [59].

3.2 Introspection

Beyond visualization, formalizing mathematical models in an SCM framework enables a range of additional tools for analyzing the models and their implications, which we call *introspection*. Causal models are useful for their ability to simulate counterfactual("what-if") scenarios, which can let users explore potential flaws in their model. Recent work [49] provides modeling tools for extending this procedure, and discusses its utility for detecting model mismatch through isolating cases where counterfactual outcomes are

not as expected. Indeed, the management science literature recommends using scenario-based planning for policy-making [55], as a method of making decisions which are robust to a range of possible outcomes [38, 45].

3.3 Evaluation

When analyzing the fairness of various policies in dynamical systems, the conclusions could be seen as concrete policy proposals intended to mitigate bias towards disadvantaged groups. However, at least one major obstacle stands in the way of implementing these policies: evaluation. Causal modeling enables *off-policy evaluation*, a method which we argue is critical for the domains the fairness literature is usually concerned with. Off-policy evaluation has been studied extensively in the causal inference literature from both economics [1, 22, 23] and computer science perspectives [4, 49, 56], particularly in reinforcement learning [28, 51, 57].

In off-policy evaluation, the goal is to use data generated by (possibly biased) historical policies to evaluate prospective new policies, without running the new policies in the world. This contrasts with more straightforward forms of evaluation: *experimentation*, where one runs the policy in the real world on a random sample of the population (i.e. A/B testing), and *simulation*, where one runs the policy in a simulator believed to be an accurate representation of the real world. Being able to perform off-policy evaluation is essential, since neither of these methods are compatible in most domains where fairness is a concern — experimentation is risky or ethically dubious in most high-stakes scenarios, and accurate simulation of complex social phenomena is difficult.

4 POLICY INTERVENTIONS IN SCMS

Having motivated the usage of SCMs for modeling decision-making processes, we present a simple example: comparing various policies in a contextual bandit, presented as an SCM in Figure 3.

In a contextual bandit, a *context* U_c is first sampled, and an action A is taken (potentially stochastically, as defined by U_a). Following the action, some feedback O is observed (i.e. a reward signal), also affected by the context and a scenario U_o (e.g. measurement noise). A *policy* π represents the key decision function, choosing actions based on contexts. In this section, we walk through the process of policy evaluation in this SCM, using the simplest off-policy evaluation method, model-based policy evaluation.
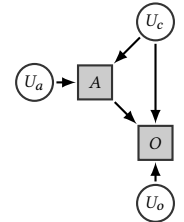
Figure 3: SCM for a contextual bandit with unobserved context U_c, action A, feedback O and scenario U_o [4].

4.1 Policy Interventions

First we introduce the relevant causal machinery for policy evaluations. In Section 2, we discussed how to intervene with the *do*-operation to set a variable's value to some constant (e.g. $do(T) = t$). However, off-policy evaluation requires a slightly different tool,

which we will call a *policy intervention*. We can think of a policy intervention as intervening not on a node's value, but on a structural equation's value. Consider the SCM in Figure 3, with action A defined by the structural equation $A = f_A(U_c, U_a)$. To estimate the expected value of O under an action value of $A = 1$, we specify our intervention with $do(A = 1)$, which replaces the structural equation f_A with a constant value of 1, cutting all arrows into A in the SCM (analogous to Figure 2c). However, to estimate the expected value of O under a new treatment *policy* $\hat{f}_A$ (say, $A = 1$ iff $U_c < \frac{1}{2}$), then we specify our intervention with $do(f_A \to \hat{f}_A)$, and the estimand would be $\mathbb{E}_{p^{do(f_A \to \hat{f}_A)}}[Y]$ (analogous to Figure 2d). In general, to denote the expected value of a variable $\mathcal{U}$ under a target policy $\hat{\pi}$ which intervenes on a variable V, we write $\mathbb{E}_{p^{do(f_V \to \hat{\pi})}}[\mathcal{U}]$.

4.2 Policy Evaluation

We aim to evaluate three policies π, each mapping contexts U_c to binary actions A. The policies are: Policy 1 (noisy optimal), Policy 2 (nearly optimal on "easy" examples, very noisy otherwise), and Policy 3 (optimal). To evaluate these policies, we will perform *model-based policy evaluation*. This is the simplest form of off-policy evaluation: we evaluate our policy on simulated data from the model we pro-

Figure 4: Model-based policy evaluation of three policy interventions in the SCM from Figure 3. Policy 3 is the optimal policy.

pose. The estimand is $\mathbb{E}_{p^{f_A \to \pi}}[O]$ — the average reward (O) observed when we generate data from the model distribution (p), after the policy intervention where we replace f_A (the structural equation for A) with the target policy (π). If our model is a perfect representation of the world, this will give us an unbiased evaluation. If we have model mismatch (e.g. measurement error [11, 13]), then this estimate will be biased. There are alternative forms of off-policy evaluation which attempt to mitigate the risk of mismatch; see Section 7 and Appendix D for a discussion and experiments in this vein.

Figure 4 shows model-based evaluations of the three policy interventions (see Appendix D for experimental details). We see that Policy 3, which is optimal, is indeed evaluated to be the best. This is not surprising, since there is no model mismatch in this experiment. However, this section simply serves as an introduction to policy evaluation in SCMs in a simple setting. In Sections 5 and 6, we examine more complex SCMs and policy interventions.

5 CAUSAL INTERPRETATIONS FOR THREE DYNAMICAL FAIRNESS MODELS

As argued in Section 3, SCMs are helpful for visualization (understanding the high level structure of a model's assumptions), introspection (manipulating the model to ask a range of further
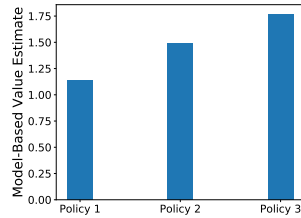
questions), and evaluation (estimating the value of a hypothetical policy without actually implementing it). In this section, we will provide evidence for this argument with a brief review of the field of fairness in dynamical systems through the lens of SCMs.

We will discuss three models from the recent fairness in dynamical systems literature, ordered by increasing complexity (in Appendix A, we display a further selection of SCMs representing other papers in the literature):

(1) Liu et al. [39]'s lending model outlines one time step of a loan application.
(2) Hashimoto et al. [21]'s repeated classification model follows the dynamics of a changing population's preferences with unobserved sensitive attributes.
(3) Hu and Chen [24]'s hiring model is quite complex, representing the long-term dynamics of a two-stage labour market.

5.1 Lending

We begin with the model from Liu et al. [39], which examines threshold-based classification in general, but with specific focus on the lending setting. Our SCM formulation of this model can be seen in Figure 1. In this model, a person with group membership (a.k.a. sensitive attribute) A receives a credit score X, and applies to a bank for a loan. The bank makes a binary decision T about whether to award the loan using the policy f_T. The binary potential outcome Y is realized, which is converted to institutional profit or loss only if $T = 1$[3]. Finally, the applicant's credit score is modified to $\tilde{X}$ (increased on repayment, decreased on default, static if $T = 0$)[4].

Utility is measured through the bank's profit $\mathcal{U}$ (a sum over the individual profits u) as well as the expected score change Δ_j, representing the average change in credit score after one time-step among members of group $A = j$. Liu et al. [39] consider the effect of various (group-specific) thresholds for loan assignment under this model: what are the expected values of $\mathcal{U}$ and Δ for some group-specific thresholds $\tau \triangleq (\tau_0, \tau_1)$ that offer loans to applicants of group j with score X if and only if their credit score $X > \tau_j$. They show that different thresholds satisfy different criteria: maximum profit (MaxProf), demographic parity (DemPar), and equal opportunity (EqOpp). In the language of our paper, this is simply a policy evaluation. Define π_τ as the threshold policy described above with thresholds τ. Then, the conclusions of the paper can be phrased with the tool of policy intervention: we are looking to evaluate the policy π_τ by estimating the quantities $\mathbb{E}_{p^{do(f_T \to \pi_\tau)}}[\mathcal{U}]$ and $\mathbb{E}_{p^{do(f_T \to \pi_\tau)}}[\Delta_j] \; \forall j$, for various τ computed under different fairness criteria. If the reader wants to follow this example all the way through they should skip ahead to Section 6, where we provide a detailed case study of this SCM, including several extensions to the model and novel policy evaluations.

5.2 Repeated Classification

We now turn to the repeated classification setting discussed by Hashimoto et al. [21], presented in SCM form in Figure 5. The

---

[3] Therefore this model does not capture a notion of opportunity loss for *not* extending a loan to applicants who *are* qualified.
[4] Likewise, the applicant's score does not change in the absence of a loan; this assumption may be inaccurate, since *not* receiving a loan could create additional financial issues for a real-world applicant.
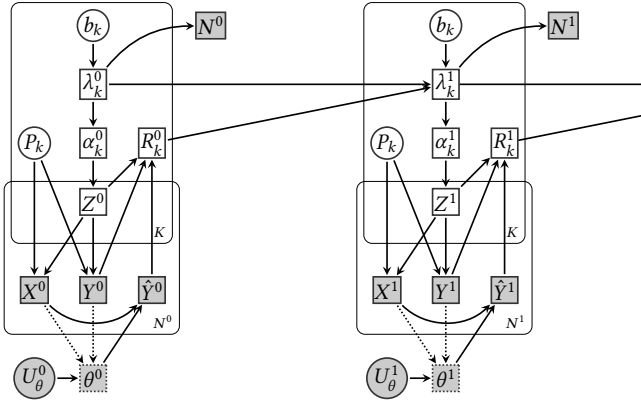
Figure 5: SCM for the repeated loss minimization model discussed by Hashimoto et al. [21]. The dotted arrows highlight the policy as the learning algorithm that produces parameters θ^t, which in term affect the predictions $\hat{Y}_i^t$. See Table 2 for explanation of all symbols and text for description.

model is fairly general, and the authors discuss several domains where it could apply (e.g. speech recognition, text auto-completion). A binary classifier with parameters θ (Hashimoto et al. [21] use a logistic regression classifier) is repeatedly trained on a population of individuals with features X and labels Y. The population distribution is a mixture of components $P = \sum_k \alpha_k P_k$, where each of the k demographic groups has proportion α_k and a unique distribution over the input-output pairs $P_k(X, Y)$. Group memberships (i.e. cluster assignments) $Z \in [1 \ldots k]$ are not observed.

The key idea is that the group distributions P_k remain static over time, but their relative proportions α_k change dynamically in response to the classifier performance on the k-th group. At the t-th step, the classifier is trained on the overall population $\{(X_i^t, Y_i^t)\}$, yielding classifier parameter θ^t and predictions[5] $\hat{Y}^t$. At each step, some subjects choose to stay in the population, some choose to leave, and some new subjects are added to the pool. In particular, the Poisson parameter λ_k (proportional to mixing coefficient α_k) is computed as a function of the per-group risk R_k, which itself is a simple function of the predictions, e.g. classification error. Misclassified subjects are more likely to leave, so under-served groups shrink over time. The authors coin this phenomenon as *disparity amplification*. Interestingly, disparity amplification can *improve* the overall loss/accuracy since the shrinking minority group contributes less to these global metrics as time proceeds. To mitigate disparity amplification, Hashimoto et al. [21] propose a robust optimization technique that seeks low loss for worst-case group assignments Z (assuming a minimum group size).

The SCM framework suggests several interesting interventions.

(1) Intervention on *latent dynamics*: $\text{do}(f_\lambda \rightarrow \hat{f}_\lambda)$ represents an intervention on the population dynamics, which we could test how a given policy responds to environment changes or misspecified structural equations. $\text{do}(b_k = \hat{b}_k)$ is a simple atomic intervention of a similar flavor, which changes the

expected number of individuals entering each group at a given time step.

(2) Intervention on *group distributions*: $\text{do}(P_k \rightarrow \hat{P}_k)$ represents a static change to the group-k distribution over input-output pairs, which could be carried out at one or every time step. Modeling group distributions that change dynamically would require a modification to the SCM.

5.3 Hiring

Finally, we examine the hiring market model proposed in Hu and Chen [24] (Figure 6). Figure 6a shows the higher-level structure of the model: a global state of the hiring market Θ progresses through time, a cohort of workers are initialized at each time step with attributes Φ set by the current global state, and the cohorts progress through time, feeding back into the global state at each step.

Figure 6b shows the structure of each individual/cohort's journey through the labour market. At the top of Figure 6b, we see the variables which constitute the global state Θ: wages w, reputation Π_μ of group μ, and the proportion of "good" workers on the permanent labour market in group μ, g_μ. The bottom plate of Figure 6b shows the variables which are part of Φ and which correspond to attributes of an individual worker's experience.

In this model, a worker is born into a sensitive group μ with "ability" level θ; θ and μ are not correlated. The first choice they must make, which defines their working life and hiring potential, is how much to invest, defined by η e.g. how much money to spend on education or training. Workers from disadvantaged groups will face higher costs of investment; workers with higher ability will face lower costs. A worker chooses to invest more if their expected costs are lower than their expected wages. This investment affects their qualification status ρ; however, this qualification status does not effect hiring. Rather, a worker is hired onto the first stage of the labour market ($h = 1$), the *temporary labour market* (TLM), when their investment η is high enough.

Once hired onto the TLM, the worker must choose how hard to work. The cost of effort e is determined by qualification level and ability: highly-qualified and high-ability workers can exert high effort with lower cost. Workers will exert high effort when its cost is lower than the expected wage increase from doing so.

At each step, workers respond to the current wage signal w^t by exerting effort ϵ^t at their job (regardless of whether they are in the TLM or PLM) resulting in an outcomes o^t — exerting high effort and having good qualifications yield a better chance of "good" outcomes. Using an average of the last T steps of their outcome history $H^{t-T:t}$, a worker builds a reputation π, which will affect the overall reputation Π_μ of their group μ. The worker will be hired into the second stage, the *permanent labour market* (PLM), if their reputation is good enough. If hired, this affects g, the overall quality of workers on the PLM, which in turn affects the wages w — an increased supply of good workers will lower wages.

Hu et al. [25] argue that the unconstrained dynamics of this systems produce inequality. By having higher costs of investment, workers from disadvantaged groups are less likely to invest, leading to worse outcomes for those workers. These worse outcomes will lower the reputation π_μ of that group, raising the cost of investment even further. However, Hu et al. [25] propose a solution: by

---

[5] Using held-out data for the predictions is expressible via a small change to the SCM.
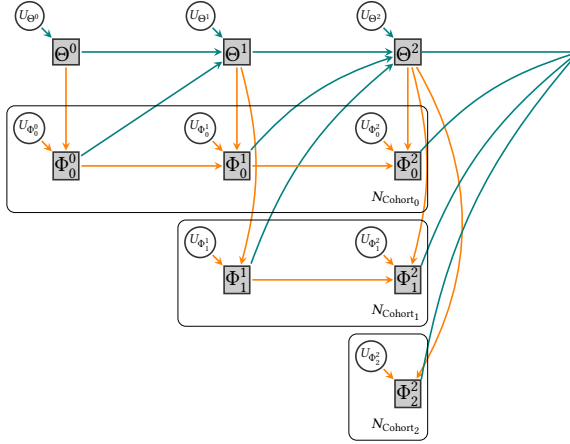
(a) Macro-level DAG showing how market state Θ^t and worker cohorts Φ_j^t dynamically affect one another
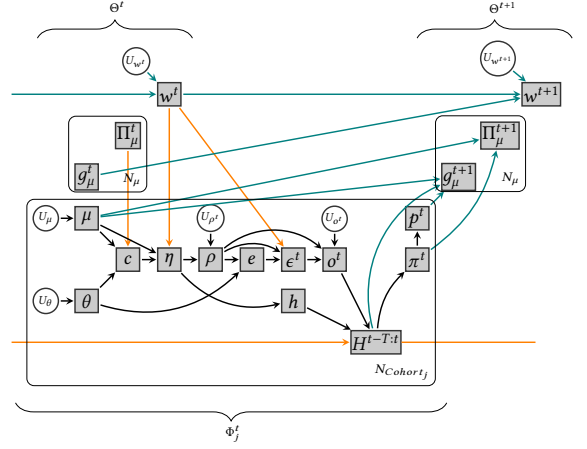
(b) Micro-level DAG isolating how market state Θ^t affects investment and effort levels of a single worker cohort, and how worker choices affect market state at the next step

Figure 6: SCM for the hiring model from Hu and Chen [24]. 6a shows macro-level causal assumptions. At step t the global state Θ^t of the PLM affects the choices of all cohorts of workers (a cohort denotes workers that enter the market at the same step) via wage signals (6b). The choices of investment and effort and resulting outcomes in turn affect the workers themselves in terms of hiring decisions, and the global state of the market in terms of average group reputation and performance per group. Teal arrows denote structural functions going into the global state. Orange arrows denote structural functions going into the cohort state. Black arrows denote structural functions within the cohort state. See Table 3 in Appendix B for explanation of all symbols, and Section 5 for description of the dynamics.

constraining TLM firms to hire equal numbers of workers from each group μ, they show the PLM will in turn have an equal number of workers from each group.

This model is the most intricate of the three we study, resulting in a number of equilibria analyzed by Hu and Chen [24], and providing a rich set of possible alternative policy interventions to consider.

(1) Intervention on *TLM hiring policy*: suggested by Hu and Chen [24], described above. Under this model, it can also be seen as an intervention on investment strategy, since all workers with $\eta > 0$ get hired in this model.

(2) Intervention on *cost of investment*: this is the primary variable through which bias enters the system (a fact easier to see through the DAG visualization). We could consider lowering costs of investment for under-privileged groups, which would encourage them to invest more heavily.

(3) Intervention on *wages*: we could ensure that all workers will be compensated at a higher rate for their labour by raising the minimum wage $\underline{w}$. This could indirectly reduce

inequality if one group would otherwise be paid below that rate more frequently, giving that group higher incentive to invest and exert high effort.

(4) Intervention on *cost of effort*: We could lower costs of effort for workers from disadvantaged groups e.g. providing subsidized childcare makes it cheaper for primary caregivers (who are disproportionately women) to work overtime if they so

choose. This would indirectly reduce inequality since workers with low qualifications are more likely to be from disadvantaged groups. Workers more likely to exert high effort will have improved outcomes and better group reputation.

All of the reasoning in the above list is significantly easier to perform with an SCM formulation at hand. The SCM is compactly specifies the inherent modeling assumptions and relationships, and facilitates evaluation of changes to these relationships. Hence, it is an effective tool for analyzing rich models such as the one presented by Hu and Chen [24], potentially discovering a range of new policy options implicit in the assumptions being made. We note that the reasoning we present here is not a policy recommendation or an assertion of efficacy; rather, it is an example of the type of analysis which is available using these causal tools.

6 CASE STUDY: INDIVIDUAL DYNAMICS

Here we present a case study whereby SCMs are used to evaluate fair policies in a lending setting. We begin by detailing our reinterpretation of the model proposed by Liu et al. [39] as a one-step causal DAG, including the relevant structural equations. To demonstrate the flexibility of the SCM framework, we then extend their model to compute a variety of new policy evaluations. The modularity of the SCM formulation facilitates a multi-step extension and, and novel interventions on the SCM can be used to estimate how the bank's policy interacts with the policies of other actors.

6.1 SCM Analysis of Liu et al. [39]

As briefly discussed above, Liu et al. [39] propose a one-step feedback model for a decision-making setting then analyze several candidate policies—denoted by the structural equation f_T in our analysis—by simulating one step of dynamics to compute the institution's profit and group outcomes for each policy. Figure 1 shows our SCM formulation of this dynamics model. Here we provide expressions for the specific structural equations used.

To sample over $p(X, A)$ we start with Bernoulli sampling of A, parameterized SCM-style like

$$U_{A_i} \sim \text{Bernoulli}(U_{A_i}|\theta); \quad A_i = f_A(U_{A_i}) \triangleq U_{A_i} \tag{1}$$

where $\theta \in [0, 1]$ is the proportion of the $A = 1$ group.

We then sample scores by the inverse CDF trick[6]. Given an inverse cumulative distribution function CDF_j^{-1} for each group $j \in \{0, 1\}$, we can write

$$U_{X_i} \sim \text{Uniform}(U_{X_i}|[0, 1]) \tag{2}$$

$$X_i = f_X(U_{X_i}, A_i) \triangleq \text{CDF}_{A_j}^{-1}(U_{X_i}) \tag{3}$$

Liu et al. [39] discuss implementing threshold policies for each group $j \in \{0, 1\}$, which are parameterized by thresholds c_j and tie-breaking Bernoulli probabilities γ (for simplicity of exposition we assume the tie-breaking probability is shared across groups). The original expression was

$$\mathbb{P}(T = 1|X, A = j) = \begin{cases} 1 & X > c_j \\ \gamma & X = c_j \\ 0 & X < c_j. \end{cases} \tag{4}$$

Then, after denoting by $\mathbb{1}(\cdot)$ the indicator function, we can rephrase this distribution in terms of a structural equation governing treatment:

$$U_{T_i} \sim \text{Bernoulli}(U_{T_i}|\gamma) \tag{5}$$

$$T_i = f_T(U_{T_i}, X_i, A_i)$$

$$\triangleq 1^{\mathbb{1}(X_i > c_{A_i})} \cdot U_{T_i}^{\mathbb{1}(X_i = c_{A_i})} \cdot 0^{\mathbb{1}(X_i < c_{A_i})}. \tag{6}$$

A policy f_T (which itself may or may not satisfy some fairness criteria) is evaluated in terms of whether loans were given to creditworthy individuals, and in terms of whether each demographic group successfully repaid any allocated loans on average. To capture the notion of *creditworthiness*, we introduce a *potential outcome* Y (repayment if the loan were given) for each individual, which is drawn[7] from $p(Y|X, A)$[8]. By convention $T = 1$ as the "positive" treatment (e.g., got loan) and $Y = 1$ as the "positive" outcome (e.g., would have repaid loan if given) Note that Y is independent of T given X,

---

meaning Y is really an indicator of *potential* success. Formally, the potential outcome Y is distributed as $Y_i \sim \text{Bernoulli}(Y_i|\boldsymbol{\rho}(X_i, A_i))$ for some function $\boldsymbol{\rho} : X \times A \to [0, 1]$. We reparameterize this as a structural equation using the Gumbel-max trick[9] [19, 41]:

$$U_{Y_i} \sim \text{Uniform}(U_{Y_i}|[0, 1]) \tag{7}$$

$$Y_i = f_Y(U_Y, X_i, A_i)$$

$$\triangleq \mathbb{1}\left(\log \frac{\boldsymbol{\rho}(X_i, A_i)}{1 - \boldsymbol{\rho}(X_i, A_i)} + \log \frac{U_Y}{1 - U_Y} > 0.5\right). \tag{8}$$

The institutional utility u_i and the updated individual score $\tilde{X}_i$ are deterministic functions of the outcome Y_i and the treatment T_i, and the original score X_i:

$$u_i = f_u(Y_i, T_i) \triangleq \begin{cases} u_+^{\mathbb{1}(Y_i)=1} \cdot u_-^{\mathbb{1}(Y_i)=0} & \text{if } T_i = 1 \\ 0 & \text{else} \end{cases}, \tag{9}$$

$$\tilde{X}_i = f_{\tilde{X}}(X_i, Y_i, T_i) \triangleq \begin{cases} X_i + c_+^{\mathbb{1}(Y_i)=1} \cdot c_-^{\mathbb{1}(Y_i)=0} & \text{if } T_i = 1 \\ X_i & \text{else} \end{cases}. \tag{10}$$

As mentioned in Section 5, $\{u_+, u_-, c_+, c_-\}$ are fixed parameters that encode expected gain/loss in utility/score based on payment/default of loan.

There are two *global* quantities of interest. Firstly, the institution cares about its overall utility at the current step (ignoring all aspects of the future), expressed as

$$\mathcal{U} = f_{\mathcal{U}}(u_{1\ldots N}) \triangleq \frac{1}{N} \sum_{i=1}^{N} u_i. \tag{11}$$

Secondly, society (and possibly the institution) might care the average per-group score change induced by the policy, expressed for group $A = j$ as

$$\Delta_j = f_{\Delta_j}(X_{1\ldots N}, \tilde{X}_{1\ldots N}, A_{1\ldots N}) \triangleq \frac{1}{N_{A_j}} \sum_{i=1}^{N} (\tilde{X}_i - X_i)^{\mathbb{1}(A_i=j)}, \tag{12}$$

with $N_{A_j} \triangleq \sum_{i'} \mathbb{1}(A_{i'} = j)$ is the size of the $A_j = 1$ group.

6.2 Extensions: multi-step and multi-actor

Laying out the original model as an SCM begs several interesting policy evaluation questions. We show how simple modifications to the SCM allow us to scrutinize the causal assumptions implicit in the original model by analyzing long-term effects, interventions by third-party actors, and robustness of a policy to structural equation misspecifications.

Longer-term impacts. Given a policy whose one-step effect is purportedly fair, what can we say about its longer-term impacts? The modularity of the SCM formulation allows us to easily estimate these effects.[10] For example, the structural equation $\tilde{X}_i = f_{\tilde{X}}(X_i, Y_i, T_i)$ can be modified to the recursive update $X_i^{t+1} = f_{X^{t>0}}(X_i^t, Y_i^t, T_i^t)$. Note that X_i^t (which does not directly depend on A) is only computed in this way for steps $t > 0$, since the original scores X^0 are sampled from $p(X^0|A)$ via Equation 3. On the other hand, since

---

[6] This standard trick is used for sampling from distributions with know densities. Recalling that $\text{CDF}_p : \mathcal{X} \to [0, 1]$ is a monotonic (invertible) function representing $\text{CDF}_p(X') = \int_{-\infty}^{X'} dX p(X < X')$. Then to sample $X' \sim p$ we first sample $U \sim \text{Uniform}(U|[0, 1])$ then compute $X' = \text{CDF}_p^{-1}(U)$.

[7] The authors denoted by $\boldsymbol{\rho}(x)$ the probability of potential success at score X. Various quantities were then computed, e.g., $\boldsymbol{u}(x) = u_+ \boldsymbol{\rho}(x) + u_-(1 - \boldsymbol{\rho}(x))$. We observe that this is equivalent to marginalizing over potential outcomes $\boldsymbol{u}(x) = \mathbb{E}_{p(Y|X)}[u_+ Y + u_-(1 - Y)]$; in our simulations we compute such expectations via Monte Carlo sampling with values of Y explicitly sampled.

[8] The authors use $\boldsymbol{\rho}(X) = p(Y|X)$ in their analysis (suggesting that potential outcome is independent of group membership conditioned on score) but $\boldsymbol{\rho}(X, A) = p(Y|X, A)$ in the code, i.e. the potential outcome depends differently on score for each group. The SCM as expressed in Figure 1 represents the codebase version.

[9] This trick reparameterizes a Categorical or Bernoulli sample as a deterministic transformation of a Uniform sample. See Oberst and Sontag [49] for discussion of how to perform counterfactual inference for SCMs with Categorical random variables.

[10] See Appendix C for code snippets demonstrating ease-of-implementation.
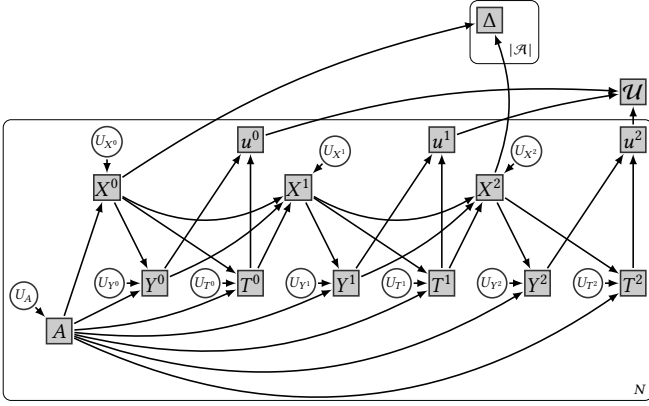
Figure 7: Phrasing the model from Liu et al. [39] as an SCM enables a multi-step extension for measuring long-term impacts, e.g., in the two-step version shown here.
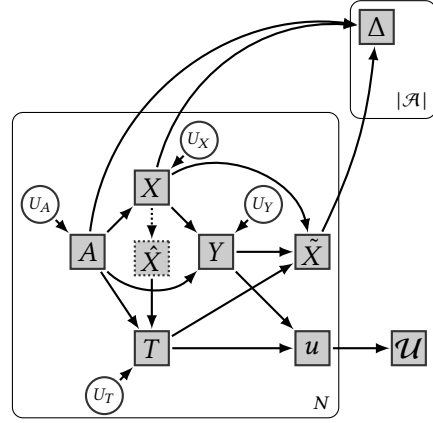

Figure 8: This reparameterization of the SCM from Liu et al. [39] makes explicit the role of the credit bureau via the structural equation $\hat{X}_i = f_{\hat{X}}(X_i)$ (with X_i and $\hat{X}_i$ representing individual features and scores), then carries out a credit bureau intervention $\mathbf{do}(f_{\hat{X}} \rightarrow \hat{f}_{\hat{X}})$. Policy estimates under intervention (e.g. $\mathbb{E}_{p^{\mathbf{do}(f_{\hat{X}} \rightarrow \hat{f}_{\hat{X}})}} [\mathcal{U}]$) can then be computed.

the structural equations for T_i^t and Y_i^t (Equations 6 and 8) do not change in the multi-step setting, we see that group membership does indeed have a long-term influence[11] on the outcomes and score trajectories for individuals.

Intervention by credit bureau. Liu et al. [39] carry out policy evaluation based on statistics of FICO credit scores assigned by the credit bureau TransUnion [52]. We note that these credit score decisions themselves constitute a policy; and moreover, the language of interventions in the SCM framework allows us to characterize decisions made by the *credit bureau* (rather than the bank) in terms of the same fairness and profit metrics as before. Mechanically, this new actor is incorporated into the SCM by reinterpreting X_i as *features* related to creditworthiness of an individual, then introducing $\hat{X}_i = f_{\hat{X}}(X_i)$ as a *score* that is deterministically computed by the agency from the features. This modification is equivalent to the original model if we assume $f_{\hat{X}}$ is the identity function.

We could additionally characterize the interaction between the policies of multiple actors. Policy evaluation under double intervention $\mathcal{M}^{\mathrm{do}(f_T \rightarrow \hat{f}_T, f_{\hat{X}} \rightarrow \hat{f}_{\hat{X}})}$ estimates the sensitivity of the bank's decisions to the decisions of the credit bureau (and vice versa), potentially leading to a game-theoretic analysis.

Intervention by government. Beyond the bank and credit agency, the government is another actor which can effect outcomes. Consider the effect of a new and far-reaching government policy such as a tax code change or economic stimulus legislation. We can capture the effect of such an intervention in the current SCM via the intervention $\mathrm{do}(f_Y \rightarrow \hat{f}_Y)$, representing a change to the potential repayment function. A new policy with uniformly positive impacts on repayment can be expressed as $\hat{f}_Y(\cdot) = f_Y(\cdot) + b$ for some global

(or perhaps group-specific) bias term b. The SCM allows us to estimate the effect of this policy, possibly under composition with interventions by the other two actors (the bank and credit bureau). The results of this evaluation could in turn be used by a domain expert to improve the causal assumptions of the model. In other words, if policy evaluations under the current SCM are judged by the domain expert as failing to capture how multi-actor policies would interact in the real world, then the expert could recommend the SCM be revised. For example, if the aforementioned uniformly positive intervention $\hat{f}_Y$ yields policy evaluations that are unrealistic, then this could motivate a new SCM (as in Hu and Chen [24]) that expresses effect of large-scale policies in terms of market signals that in term affect the *choices* made by individuals.

Robustness to model mismatch. We can also express policy robustness using the interventional language of SCMs by formulating model mismatch[12] in terms of a class of possible interventions on the structural equations. As an example, consider the potential outcomes Y, which could be modeled as independent of the sensitive attribute A given the score, or dependent on both A and X. We can estimate the robustness of our policy to this potential model mismatch—$p(Y|X)$ vs. $p(Y|X, A)$, in two ways. This intervention is not meant to evoke a particular real-world event, but instead is meant to measure sensitivity of the policy estimate to model assumptions — in this case, the assumption that the score X is well calibrated for both values of A (an assumption which is difficult to validate given that Y is only observed when $T = 1$).

6.3 Results

Figure 9 shows the effect on the average utility $\mathbb{E}[\mathcal{U}]$ and average per-group score change $\mathbb{E}[\Delta_j]$ of a simple policy intervention by

---

[11] This multi-step model is related to the one proposed by Mouzannar et al. [46], which instead assumed static individual behavior but dynamic group behavior through the per-group score distributions $p(X|A)$ (See Appendix A for an SCM analysis of their model). Whereas Mouzannar et al. [46] focused their analysis on the limit as $t \rightarrow \infty$, we instead emphasize finite-time horizons, which may be more relevant to policy makers.

[12] "Mismatch" refers here to structural equations with misspecified functional forms, not incorrect causal assumptions.
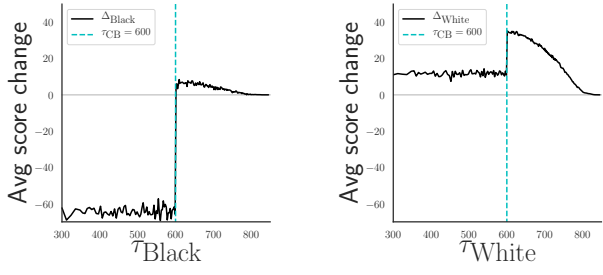
(a) Score change, min. group.

(b) Score change, maj. group.
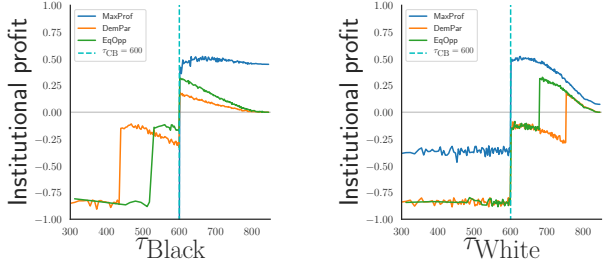
(c) Profit as fn. of min. thresh.

(d) Profit as fn. of maj. thresh.

Figure 9: Policy evaluation under credit bureau intervention $\hat{f}_{\hat{X}}(X) = \min(X, \tau_{CB})$ **with** $\tau_{CB} = 600$. **Group score change—formally** $\mathbb{E}_{p^{do(f_{\hat{X}} \to \hat{f}_{\hat{X}}, f_T \to \hat{f}_T)}}[\Delta_j] \forall j \in \{$**Black, White**$\}$— **and institutional profits—formally** $\mathbb{E}_{p^{do(f_{\hat{X}} \to \hat{f}_{\hat{X}}, f_T \to \hat{f}_T)}}[\mathcal{U}]$**—are shown as functions of the two group thresholds** $\{\tau_j\}$. **Bank profits depend on its fairness criteria.**

the credit bureau. The intervention involves the bureau setting the minimum score to 600 for all applicants via the structural equation $\hat{f}_{\hat{X}}(X) = \min(X, 600)$. This intervention is unlikely in the real world because it contradicts the profit incentives of the bureau, which encourage well-calibrated scores. Nevertheless, it coarsely captures a potential scenario where an actor besides the bank seeks to encourage fair outcomes in a group-blind way, since under the new scoring policy minority applicants are more likely to receive loans. However, we see in Figure 9a that the average group outcome for Black applicants is negative when the bank's group threshold τ_{Black} is below 600, since in this case its policy offers loans to individuals who have good scores on paper but are unlikely to repay the loans. Interestingly, the expected profit (Figure 9b) under credit bureau intervention differs depending on the fairness criteria of the bank. This is because each fairness criteria differently constrains the relationship between the two thresholds $\{\tau_{Black}, \tau_{White}\}$, so the choice of fairness criteria implicitly sets how many many applicants with boosted scores ($X < 600$, thus $\hat{X} = 600$) are selected for loans. We observe that the EqOpp policy is less sensitive than DemPar to credit bureau intervention. This is because the DemPar constraint is more strict, and results in offering loans to more applicants with boosted scores (who are are unlikely to repay, and disproportionately belong to the minority group).
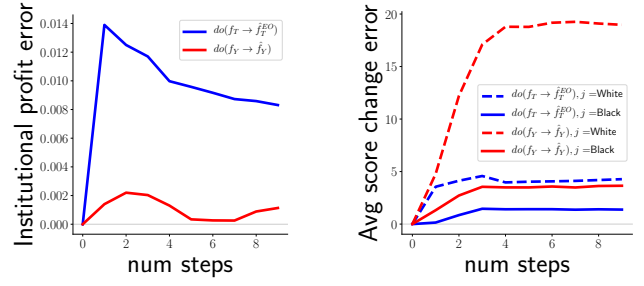
(a) Group improvement.

(b) Institutional profit.

Figure 10: Evaluating multi-step policy robustness to distribution shift for various choice of intervention distribution q. **Sensitivity of institutional utility—formally** $|\mathbb{E}_q[\mathcal{U}] - \mathbb{E}[\mathcal{U}]|$—**and sensitivity of group avg. score change—formally** $|\mathbb{E}_q[\Delta_j] - \mathbb{E}[\Delta_j]|$—**are shown as a function of steps. Expected profit is relatively robust to both interventions, whereas the expected per-group score changes are relatively more sensitive to these interventions.**

Figure 10 estimates robustness of the EqOpp policy to two forms of model mismatch described above. $\text{do}(f_T \to \hat{f}_T^{EO})$ recomputes the per-group thresholds under the EqOpp constraint, but using *incorrect* statistics from the credit bureau. In particular, the marginal $p(Y|X)$ was used for both group's repayment probabilities rather than the correct $p(Y|X, A)$. The second intervention $\text{do}(f_Y \to \hat{f}_Y)$ is more severe, where $p(Y|X)$ is used to *sample* potential outcomes Y rather than just set the thresholds within f_T. We measure error under each intervention relative to the "correct" baseline where the correct potential outcome distributions are used to set thresholds and sample data. We repeat this procedure over a varying number of steps to determine how errors compound over time. In Figure 10a we observe the institutional profits are surprisingly robust to both forms of intervention (compare the Y-axis with Figure 9c), while in Figure 10b we observe the per-group outcomes are more sensitive to these interventions, especially to $\text{do}(f_Y \to \hat{f}_Y)$.

7 DISCUSSION

7.1 Off-Policy Evaluation Methods

In Section 6, we discuss model-based policy evaluation. However, there are other methods for off-policy evaluation, which can use historical data to overcome some of the weaknesses of model-based policy evaluation. We discuss three basic approaches to off-policy evaluation of a target policy $\hat{\pi}$ given a dataset of trajectories[13] $\{\tau\}$ that were generated by the historical policy π^{obs} interacting with the environment, and a model $\mathcal{M}$ of the environment.

(1) Importance sampling (IS): ignore the model and use historical data. Compute a reweighted expected reward, where the rewards from each trajectory τ in the sample are reweighted according to density ratio $\frac{p^{\text{do}(\pi \to \hat{\pi})}(\tau)}{p^{\text{do}(\pi \to \pi^{\text{obs}})}(\tau)}$.

---

[13] Each trajectory is one tuple comprising a sequence of all observations over all the time steps specified by the SCM.

(2) Model-based policy evaluation (MB-PE, discussed in Section 4): ignore the historical data and use the model. Given a model M, sample exogenous noise from the priors $p(U)$, produce trajectories $\{\tau\}$ by running the model M along with the target policy $\hat{\pi}$, and compute the expected reward.

(3) Counterfactual-based policy evaluation (CF-PE) [4]: Use historical data and model together. The same as MB-PE, except sample exogenous noise from the posterior $p(U|\tau)$ (conditioned on a historical trajectory τ) rather than the prior $p(U)$. When there is no model mismatch, counterfactual policy evaluation is equivalent to model-based [4].

Using the contextual bandit (Section 4), we conduct a comparison between these three methods of off-policy evaluation under model mismatch. We evaluate three policies using these off-policy methods, and show results in Figure 11, finding that counterfactual-based policy evaluation — the method leveraging causal reasoning most fully — proves the most robust to model mismatch. We defer the details of this experiment to

Figure 11: Mean average error in contextual bandit (Figure 3) policy evaluation for several off-policy evaluation methods under model mismatch.

Appendix D. These experiments provide a proof-of-concept that causal reasoning can be used for more robust off-policy evaluation. If some historical data is available, this technique could be applied to any of the methods proposed in the fairness in dynamical systems literature, possibly providing a more complete picture of their performance and potential impacts.

7.2 On Limitations of SCMs

There has been significant criticism of the graphical approach to causal inference for social science, well-encapsulated by Imbens [26], who states that "the [SCM] literature has not shown much evidence of the benefits for empirical practice in settings that are important in economics." This point is well-taken. We do not claim that graphical approaches are always superior for causal effect estimation, as they can suffer from model mismatch and make many assumptions which can be difficult to ensure.

The method we propose is not intended as a silver bullet to solve policy evaluation, which remains as a challenging problem. However, we believe the field of fairness in dynamical systems is concerned with a set of problems that are well-represented by SCMs; in Section 5 we show there is frequently an equivalency between the equations governing a model of dynamic unfairness and the structural equations of an SCM. So to the degree that one would take seriously conclusions drawn from any of the mathematical models cited by this paper, one should approach conclusions drawn from SCM-enabled policy evaluation similarly. Additionally, these models lends themselves to graphical approaches due to the number

of variables included in them (the SCMs in this paper are an order of magnitude larger than canonical SCM examples [50]). Hence, SCMs provide a principled approach for drawing conclusions about candidate models and policies, possibly directing which policies will receive further investigation.

7.3 Different Types of Causal DAGs

Throughout this paper, we have focused on SCMs as our causal DAG formalism of choice. However, several exist. Here we discuss two: the major conclusions of this paper are compatible with both.

Single-World Intervention Graphs (SWIGs) [53]: SWIGs are a graphical formalism which uses counterfactual outcomes as a modeling primitive, and therefore is more consistent with the potential outcome theory of causal inference. They make slightly weaker assumptions than SCMs as they are able to express conditional independence relationships with specific counterfactual outcomes (e.g. "weak ignorability"). The graph surgery procedure representing intervention in SWIGs is done through "node-splitting" rather than removing arrows as in the *do*-operation.

Augmented DAGs/Influence diagrams [7]: Augmented DAGs (a form of influence diagrams) are a graphical formalism for causal inference with the advantage of not requiring graph surgery to represent interventions. Rather, "intervention variables" are added to the graph as parents of every node where an intervention is possible. Then, an intervention is represented by setting the value of the intervention node, and various causal conditions can be read off of the graph using standard rules for PGMs.

7.4 Future Work

A range of future work can be found at the intersection of existing fields:

- **Reinforcement learning and fairness:** Finding off-policy estimation methods better for low data, high-stakes regimes
- **Causal inference and dynamical systems:** Characterizing identifiability of long-term effects of policy interventions in terms of graphical criterion
- **Reinforcement learning and causal inference:** Developing methods for sensitivity analysis to estimate uncertainty of policy evaluations under confounding
- **Causal inference and visualization:** Visualizing complex, many-variable graphical models of policy problems
- **Fairness and decision science:** Integrating theoretical models of fairness in into scenario-based planning procedures

By advocating for this modeling approach, we hope to lay the groundwork in this paper for future research and practice in this area. We encourage researchers to formulate their dynamics models as causal DAGs and to consider running an off-policy evaluation of proposed models with historical data, where available. We hope that policymakers or stakeholders without mathematical backgrounds can find models of fairness in dynamical systems more accessible when presented as DAGs, and that they can become a useful tool in the researcher's and policymaker's toolbox.
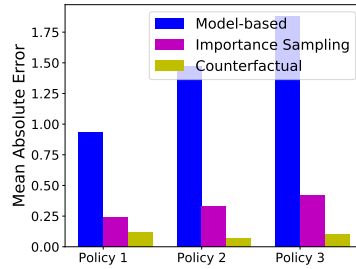
REFERENCES

[1] Susan Athey. 2015. Machine learning and causal inference for policy evaluation. In *Proceedings of the 21st ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 5–6.

[2] Lisa F Berkman, Ichirō Kawachi, and M Maria Glymour. [n. d.]. *Social epidemiology*.

[3] Dimitrios Bountouridis, Jaron Harambam, Mykola Makhortykh, Mónica Marrero, Nava Tintarev, and Claudia Hauff. 2019. SIREN: A Simulation Framework for Understanding the Effects of Recommender Systems in Online News Environments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, 150–159.

[4] Lars Buesing, Theophane Weber, Yori Zwols, Sebastien Racaniere, Arthur Guez, Jean-Baptiste Lespiau, and Nicolas Heess. 2019. Woulda, coulda, shoulda: Counterfactually-guided policy search. In *International Conference on Representation Learning*.

[5] Silvia Chiappa. 2019. Path-specific counterfactual fairness. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7801–7808.

[6] Stephen Coate and Glenn C Loury. 1993. Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review* (1993), 1220–1240.

[7] A Philip Dawid. 2002. Influence diagrams for causal modelling and inference. *International Statistical Review* 70, 2 (2002), 161–189.

[8] Shayan Doroudi, Philip S Thomas, and Emma Brunskill. 2017. Importance Sampling for Fair Policy Selection.. In *Uncertainty in Artificial Intelligence (UAI)*.

[9] Danielle Ensign, Sorelle A Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. 2018. Runaway Feedback Loops in Predictive Policing. In *Conference on Fairness, Accountability and Transparency*. 160–171.

[10] Tom Everitt, Pedro A Ortega, Elizabeth Barnes, and Shane Legg. 2019. Understanding agent incentives using causal influence diagrams, Part I: single action settings. *arXiv preprint arXiv:1902.09980* (2019).

[11] Claes Fornell and David F Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research* 18, 1 (1981), 39–50.

[12] Dean P Foster and Rakesh V Vohra. 1992. An economic argument for affirmative action. *Rationality and Society* 4, 2 (1992), 176–188.

[13] Wayne A Fuller. 2009. *Measurement error models*. Vol. 305. John Wiley & Sons.

[14] Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2018. Predictably unequal? the effects of machine learning on credit markets. *The Effects of Machine Learning on Credit Markets (November 6, 2018)* (2018).

[15] Dora Gicheva and Jeffrey Thompson. 2015. The effects of student loans on long-term household financial stability. *Student loans and the dynamics of debt* (2015), 287–316.

[16] Herman Heine Goldstine and John Von Neumann. 1947. Planning and coding of problems for an electronic computing instrument. (1947).

[17] Bridget J Goosby and Chelsea Heidbrink. 2013. The Transgenerational Consequences of Discrimination on African-American Health Outcomes. *Sociology compass* 7, 8 (2013), 630–643.

[18] John Grin, Jan Rotmans, and Johan Schot. 2010. *Transitions to sustainable development: new directions in the study of long term transformative change*. Routledge.

[19] Emil J Gumbel and Julius Lieblein. 1954. Some applications of extreme-value methods. *The American Statistician* 8, 5 (1954), 14–17.

[20] Lois M Haibt. 1959. A program to draw multilevel flow charts. In *Papers presented at the the March 3-5, 1959, Western Joint Computer Conference*. ACM, 131–137.

[21] Tatsunori B Hashimoto, Megha Srivastava, Hongseok Namkoong, and Percy Liang. 2018. Fairness without demographics in repeated loss minimization. In *International Conference on Machine Learning*.

[22] James J Heckman. 2000. Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics* 115, 1 (2000), 45–97.

[23] James J Heckman and Edward J Vytlacil. 2007. Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. *Handbook of econometrics* 6 (2007), 4779–4874.

[24] Lily Hu and Yiling Chen. 2018. A short-term intervention for long-term fairness in the labor market. In *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, 1389–1398.

[25] Lily Hu, Nicole Immorlica, and Jennifer Wortman Vaughan. 2019. The disparate effects of strategic manipulation. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, 259–268.

[26] Guido Imbens. 2019. *Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics*. Technical Report. National Bureau of Economic Research.

[27] Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. 2017. Fairness in reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 1617–1626.

[28] Nan Jiang and Lihong Li. 2016. Doubly Robust Off-policy Value Evaluation for Reinforcement Learning. In *International Conference on Machine Learning*. 652–661.

[29] Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth. 2016. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*. 325–333.

[30] Richard Kammann. 1975. The comprehensibility of printed instructions and the flowchart alternative. *Human factors* 17, 2 (1975), 183–191.

[31] Sampath Kannan, Aaron Roth, and Juba Ziani. 2019. Downstream effects of affirmative action. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, 240–248.

[32] Niki Kilbertus, Mateo Rojas Carulla, Giambattista Parascandolo, Moritz Hardt, Dominik Janzing, and Bernhard Schölkopf. 2017. Avoiding discrimination through causal reasoning. In *Advances in Neural Information Processing Systems*. 656–666.

[33] Eric I Knudsen, James J Heckman, Judy L Cameron, and Jack P Shonkoff. 2006. Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences* 103, 27 (2006), 10155–10162.

[34] Donald E Knuth. 1963. Computer-drawn flowcharts. *Commun. ACM* 6, 9 (1963), 555–563.

[35] Daphne Koller and Nir Friedman. 2009. *Probabilistic graphical models: principles and techniques*. MIT press.

[36] Matt Kusner, Chris Russell, Joshua Loftus, and Ricardo Silva. 2019. Making Decisions that Reduce Discriminatory Impacts. In *International Conference on Machine Learning*. 3591–3600.

[37] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual fairness. In *Advances in Neural Information Processing Systems*. 4066–4076.

[38] Robert J Lempert, David G Groves, Steven W Popper, and Steve C Bankes. 2006. A general, analytic method for generating robust strategies and narrative scenarios. *Management science* 52, 4 (2006), 514–528.

[39] Lydia Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. 2018. Delayed Impact of Fair Machine Learning. In *International Conference on Machine Learning*. 3156–3164.

[40] Kristian Lum and William Isaac. 2016. To predict and serve? *Significance* 13, 5 (2016), 14–19.

[41] Chris J Maddison, Daniel Tarlow, and Tom Minka. 2014. A* sampling. In *Advances in Neural Information Processing Systems*. 3086–3094.

[42] David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel. 2019. Fairness through causal awareness: Learning causal latent-variable models for biased data. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, 349–358.

[43] Richard E Mayer. 1975. Different problem-solving competencies established in learning computer programming with and without meaningful models. *Journal of educational psychology* 67, 6 (1975), 725.

[44] Smitha Milli, John Miller, Anca D Dragan, and Moritz Hardt. 2019. The Social Cost of Strategic Classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, 230–239.

[45] M Granger Morgan, Milind Kandlikar, James Risbey, and Hadi Dowlatabadi. 1999. Why conventional tools for policy analysis are often inadequate for problems of global change. *Climatic Change* 41, 3 (1999), 271–281.

[46] Hussein Mouzannar, Mesrob I Ohannessian, and Nathan Srebro. 2019. From fair decision making to social equality. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, 359–368.

[47] Brad A Myers. 1990. Taxonomies of visual programming and program visualization. *Journal of Visual Languages & Computing* 1, 1 (1990), 97–123.

[48] Razieh Nabi and Ilya Shpitser. 2018. Fair inference on outcomes. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[49] Michael Oberst and David Sontag. 2019. Counterfactual Off-Policy Evaluation with Gumbel-Max Structural Causal Models. In *International Conference on Machine Learning*. 4881–4890.

[50] Judea Pearl et al. 2009. Causal inference in statistics: An overview. *Statistics surveys* 3 (2009), 96–146.

[51] Doina Precup. 2000. Eligibility traces for off-policy policy evaluation. *Computer Science Department Faculty Publication Series* (2000), 80.

[52] US Federal Reserve. 2007. Report to the Congress on Credit Scoring and its Effects on the Availability and Affordability of Credit. *Washington, DC: Board of Governors of the Federal Reserve System* (2007).

[53] Thomas S Richardson and James M Robins. 2013. 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.

[54] David A. Scanlan. 1989. Structured flowcharts outperform pseudocode: An experimental comparison. *IEEE software* 6, 5 (1989), 28–36.

[55] Peter Schwartz. 1996. *The art of the long view: paths to strategic insight for yourself and your company*. Crown Business.

[56] Adith Swaminathan, Akshay Krishnamurthy, Alekh Agarwal, Miro Dudik, John Langford, Damien Jose, and Imed Zitouni. 2017. Off-policy evaluation for slate recommendation. In *Advances in Neural Information Processing Systems*. 3632–3642.

[57] Philip Thomas and Emma Brunskill. 2016. Data-efficient off-policy policy evaluation for reinforcement learning. In *International Conference on Machine Learning*.

2139–2148.

[58] Debra Umberson, Camille B Wortman, and Ronald C Kessler. 1992. Widowhood and depression: Explaining long-term gender differences in vulnerability. *Journal of health and social behavior* (1992), 10–24.

[59] Patricia Wright and Fraser Reid. 1973. Written information: some alternatives to prose for expressing the outcomes of complex contingencies. *Journal of Applied Psychology* 57, 2 (1973), 160.

A OTHER SCMS

Here we provide some SCMs for some additional papers from the literature:

- Figure 12 describes the multi-step loan setting discussed by Mouzannar et al. [46]. Their model is similar to the one discussed in Section 6.1. The main difference is that Mouzannar et al. [46] describes dynamics that unfold *exclusively* at the population level, where decisions rendered by the institution do not affect the future well-being of the individuals themselves.
- Figure 13 corresponds to the news recommender simulator discussed by Bountouridis et al. [3]. The goal of this simulator was to understand the long-term effects of recommender algorithms on news consumption behaviors.
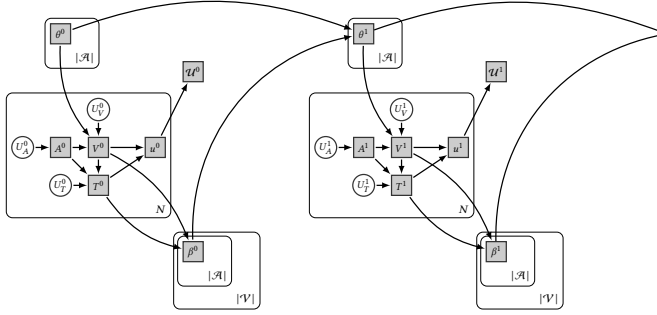

Figure 12: SCM for the group dynamics model proposed by Mouzannar et al. [46]. See Table 4 for a description of each symbol.

B SYMBOL LEGENDS

Here we provide the following symbol decoders for SCMs expressed in the main body of text:

- Table 1 decodes the symbols used in Figure 1
- Table 2 decodes the symbols used in Figure 5
- Table 3 decodes the symbols used in Figure 6
- Table 4 decodes the symbols used in Figure 12
- Table 5 decodes the symbols used in Figure 13

C CODE SNIPPETS

One upside of SCM programming is modularity. Figures 14 and 15 show the single- and multi-step SCM simulation of a loan setting. The single-step code implements the model described by Liu et al. [39], while the multi-step scenario is realized by a straightforward extension of the single-step SCM implementation.
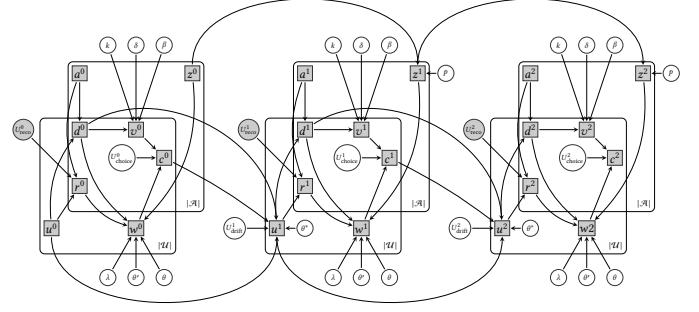

Figure 13: SCM for the news recommendation simulator model proposed by Bountouridis et al. [3]. The key dynamic modeling is in the user vectors in topic space, which drift over time towards the articles that are consumed (these in turn partially depend on the recommendations). Articles are also modeled as decaying in popularity in time. See Table 5 for explanation of all symbols.

Symbol	Meaning
N	Number of individuals
$|\mathcal{A}|$	Number of demographic groups
A_i	Sensitive attribute for individual i
U_{A_i}	Exogenous noise on sensitive attribute for individual i
X_i	Score for individual i
U_{X_i}	Exogenous noise on score for individual i
Y_i	Potential outcome (loan repayment/default) for individual i
U_{Y_i}	Exogenous noise on potential outcome for individual i
T_i	Treatment (institution gives/withholds loan) for individual i
U_{T_i}	Exogenous noise on treatment for individual i
u_i	Utility of individual i (from the institution's perspective)
Δ_i	Expected improvement of score for individual i
$\tilde{X}_i$	Score for individual i after one time step
$\mathcal{U}$	Global utility (from institution's perspective)
Δ_j	Expected change in score for group j

Table 1: Symbol legend for Figure 1

D EXPERIMENTAL DETAILS: OFF-POLICY EVALUATION IN CONTEXTUAL BANDITS

In this paper, we present two experimental results with contextual bandits: a model-based policy evaluation experiment in Section 4, with results presented in Figure 4; and a comparison of various off-policy evaluation methods referenced in Section 7, with results presented in Figure 11. In this appendix, we provide details on these experiments.

In the Section 4 experiment, the SCM of the bandit is defined as follows with $\sigma = 5$:

$$U_c \sim U(\frac{1-\sigma}{2}, \frac{1+\sigma}{2})$$
$$U_o \sim N(0,1) \tag{13}$$
$$O = A(1 - U_c) + (1 - A)U_c$$

The structural equation for A is not specified — this will be defined by our historical policy π_0.

The three policies we consider are:

Symbol	Meaning
k	indexes groups
P_k	distribution over (X, Y) for group k
b_k	expected group-k baseline population growth at each step
λ_k^t	expected population for group k at time t
α_k^t	mixing coeff for group k at time t
N^t	Total population at time t
Z_k^t	indicator of individual belonging to k-th group
X^t	input features for an individual at time t
Y^t	label for an individual at time t
U_θ^t	Exogenous noise in learning algo. (e.g., random seed)
θ^t	Estimated classifier parameters at time t
$\hat{Y}^t$	Predicted label for an individual at time t
R_k^t	Classification error for group k at time t (unobserved)

Table 2: Symbol legend for Figure 5

Symbol	Meaning
t	indexes time
i	indexes individuals
j	indexes cohorts
w^t	wages at time t
g_μ^t	proportion "good" group-μ workers in PLM
Π_μ^t	group μ reputation at time t
μ_i	group membership for worker i
θ_i	individual i ability
c_i	cost of investment for individual i
η_i	investment level for individual i
ρ_i	qualification level for individual i
e_i	individual-i cost of effort
ϵ_i^t	individual-i actual effort exerted at time t
o_i^t	individual-i outcome at time t
h_i	was individual hired to TLM following education?
$H_i^{t-\tau:t-1}$	individual-i τ-recent history (outcomes and TLM/PLM status)
π_i^t	individual i reputation at time t
p_i^t	was individual hired to PLM at step t?

Table 3: Symbol legend for Figure 6

Symbol	Meaning
A_i	Sensitive attribute for individual i
U_{A_i}	Exogenous noise on sensitive attribute for individual i
$|\mathcal{A}|$	Number of demographic groups
V_i	Qualification for individual i
U_{V_i}	Exogenous noise on qualification for individual i
$|\mathcal{V}|$	Number of qualification levels
θ_j^t	Bernoulli parameter of qualifications of group j at time t
N	Number of individuals
T_i	"Treatment" (whether the institution gives loan) for individual i
U_{T_i}	Exogenous noise on treatment for individual i
u_i	Utility of individual i (from the institution's perspective)
$\beta_{j,v}^t$	Selection rate for group j members with qual. v at step t
$\mathcal{U}$	Global institutional utility

Table 4: Symbol legend for Figure 12

```python
import torch
import structural_equations as se
StructuralEqn = se.StructuralEqn  # has similar API to torch.nn.Module

class OneStepSimulation:
    """Runs simulation for one step of dynamics under Liu et al 2018 SCM."""
    def __init__(self,
            f_A: StructuralEqn,  # stochastic SE for group membership
            f_X: StructuralEqn,  # stochastic SE for indiv scores
            f_Y: StructuralEqn,  # stochastic SE for potential repayment
            f_T: StructuralEqn,  # SE for threshold loan policy
            f_Xtilde: StructuralEqn,  # SE for indiv score change
            f_u: StructuralEqn,  # SE for individual utility
            f_Umathcal: StructuralEqn,  # SE for avg instit. utility
            f_Deltaj: StructuralEqn,  # SE per-group avg score change
            ) -> None:
        self.f_A = f_A
        self.f_X = f_X
        self.f_Y = f_Y
        self.f_T = f_T
        self.f_Xtilde = f_Xtilde
        self.f_u = f_u
        self.f_Deltaj = f_Deltaj
        self.f_Umathcal = f_Umathcal

    def run(self, num_steps: int, num_samps: int) -> Dict:
        """Run simulation forward for num_steps and return all observables."""
        if num_steps != 1:
            raise ValueError('Only one-step dynamics are currently supported.')
        blank_tensor = torch.zeros(num_samps)
        A = self.f_A(blank_tensor)
        X = self.f_X(A)
        Y = self.f_Y(X, A)
        T = self.f_T(X, A)
        Xtilde = self.f_Xtilde(X, Y, T)
        u = self.f_u(Y, T)
        Deltaj = self.f_Deltaj(X, Xtilde, A)
        Umathcal = self.f_Umathcal(u)
        return_dict = dict(
            A=A,
            X=X,
            Y=Y,
            T=T,
            u=u,
            Xtilde=Xtilde,
            Deltaj=Deltaj,
            Umathcal=Umathcal,
        )
        return return_dict

    def intervene(self, **kwargs):
        """Update attributes via intervention."""
        for k, v in kwargs.items():
            setattr(self, k, v)
```

Figure 14: Example Python implementation of single-step loan dynamics as SCM.

	Symbol	Meaning
User	u_i^t	i-th user topic vector at step t
	θ	Awareness decay with user-article distance
	θ'	Awareness decay with article prominence
	λ	Prominent vs proximity in awareness computation
	w	Max awareness pool size for any user
	k	i-th user's sensitivity to article proximity in awareness computation
	θ_i^*	i-th user's sensitivity to article proximity in drift computation
	s	number of articles read per user per step
	$U_{\text{drift},i}^t$	Exogenous noise on user i's drift at step t
	$|\mathcal{U}|$	Number of users
Article	a_j	j-th article topic vector
	z_j^0	initial prominence of article j (possibly shared across topic)
	z_j^t	prominence of article j at step t
	p	prominence (linear) decay factor
	$|\mathcal{A}|$	Number of articles
User-article	$d_{i,j}^t$	distance between user u_i and article a_j at step t
	$v_{i,j}^t$	computed step-t distance (user u_i, article a_j) in awareness computation
	$c_{i,j}^t$	computed step-t choice of user u_i about article a_j
	$U_{\text{choice},i,j}^t$	Exogenous noise on user i's choice of article j at step t
Recommender	m	Number of articles recommended to each user at each step
	$\kappa_{i,j}^t$	Rank of recommendation of article j to user i at step t
	δ	Base amount of salience boost induced by a recommendation
	β	Rank-decay of salience induced by a recommendation
	U_{reco}^t	Exogenous (possibly observed) noise in the recommender algo at step t
	d	Number of steps in the simulation

Table 5: Symbol legend for Figure 13

```python
import torch
from simulation import OneStepSimulation

class MultiStepSimulation(OneStepSimulation):
    """Runs simulation for multiple step of dynamics."""

    def run(self, num_steps: int, num_samps: int) -> Dict:
        """Run simulation forward for num_steps and return all observables."""
        blank_tensor = torch.zeros(num_samps)
        A = self.f_A(blank_tensor)
        Xs, Ys, Ts, us, Umathcals = [], [], [], [], []
        Xinit = self.f_X(A)
        X = Xinit
        for _ in range(num_steps):
            Xs.append(X)
            Y = self.f_Y(X, A)
            Ys.append(Y)
            T = self.f_T(X, A)
            Ts.append(T)
            u = self.f_u(Y, T)
            us.append(u)
            Xtilde = self.f_Xtilde(X, Y, T)
            X = Xtilde
            Umathcal = self.f_Umathcal(u)
            Umathcals.append(Umathcal)
        Deltaj = self.f_Deltaj(Xinit, Xtilde, A)  # compute final improvements
        Umathcal = torch.mean(torch.stack(Umathcals, dim=0))
        return_dict = dict(
            A=A,
            X=Xs,
            Y=Ys,
            T=Ts,
            u=us,
            Deltaj=Deltaj,
            Umathcal=Umathcal,
            )
        return return_dict
```

Figure 15: Implementation that extends dynamics across multiple time steps.

- Policy 1: if $U_c > 0.5$, choose $A = 1$ with probability (w.p.) 0.25; else, choose $A = 1$ w.p. 0.75
- Policy 2: if $U_c > 0.75$, choose $A = 1$ with probability (w.p.) 0.1; if $U_c < 0.25$, choose $A = 1$ w.p. 0.9; else, choose $A = 1$ w.p. 0.5
- Policy 3 (optimal): if $U_c > 0.5$, choose $A = 1$; else, choose $A = 0$

We run model-based evaluation of each policy with 5000 samples.

In the Section 7 experiment, the SCM is defined the same as in the Section 4 experiment. The three policies we consider are also the same. For each ordered pair π_0, π_1 of policies, using π_0 as the historical policy to collect data and π_1 as the target policy for off-policy evaluation, we compare model-based evaluation, importance sampling, and counterfactual evaluation.

In this experiment, we consider model mismatch — specifically, a misspecified prior distribution over U_c. we iterate over eight misspecified priors for U_c. These are defined by 4 values of $\sigma = 1, 3, 7, 9$, and 2 functional forms of the prior: Uniform ($U(\frac{1-\sigma}{2}, \frac{1+\sigma}{2})$) and Normal ($N(0, \sigma^2)$). The displayed results are averaged over every pair π_0, π_1 of the three above policies, and each of the misspecified priors. This comes to 216 total off-policy evaluation runs (8 priors, 9 historical/target policy pairs, 3 policy evaluation methods). The evaluation metric is absolute error.

We also ran a second model mismatch experiment whose results figure we did not include in the main body of the paper since they were qualitatively similar. The second form of model mismatch we tested is omitted variable bias, which occurs when a variable is left out of the model. Here this is an unknown hidden confounder, depicted in Figure 16 — what if the confounding context contains two variables (U_c and U_h), but we only model it as containing one?

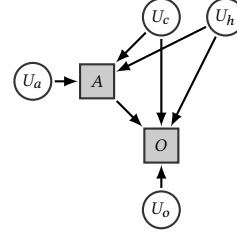This can be seen as a special type of misspecified prior in this specific example.

Figure 16: SCM for a contextual bandit with unobserved context U_c, omitted context U_h, action A, feedback O and scenario U_o.

The SCM for the omitted variable bias experiment is defined by:

$$
\begin{aligned}
U_c &\sim U(0, 1) \\
U_h &\sim U(-3, 3) \\
U_o &\sim N(0, 1) \\
O &= A(1 - (U_c + U_h)) + (1 - A)(U_c + U_h)
\end{aligned} \tag{14}
$$

However, our model of the world does not include U_h — equivalently, assuming $U_h = 0$ always. The policies we use are the same as the previous experiment, but with ($U_c + U_h$) as the input to A rather than U_c.

We display the results of the two model mismatch experiments in Figure 17.

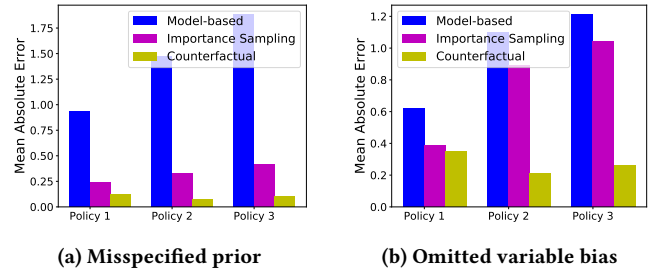(a) Misspecified prior   (b) Omitted variable bias

Figure 17: Mean average error in policy evaluation for importance sampling, model-based and counterfactual policy evaluation methods under various forms of model mismatch — misspecified prior distribution over contexts (Figure 17a) and omitted variable bias (Figure 17b).