



PDF Download  
3715275.3732008.pdf  
21 January 2026  
Total Citations: 0  
Total Downloads: 2205

 Latest updates: <https://dl.acm.org/doi/10.1145/3715275.3732008>

RESEARCH-ARTICLE

## Time Can Invalidate Algorithmic Recourse

**GIOVANNI DE TONI**, University of Trento, Trento, TN, Italy

**STEFANO TESO**, University of Trento, Trento, TN, Italy

**BRUNO LEPRI**, Bruno Kessler Foundation, Trento, TN, Italy

**ANDREA PASSERINI**, University of Trento, Trento, TN, Italy

Open Access Support provided by:

University of Trento

Bruno Kessler Foundation

Published: 23 June 2025

[Citation in BibTeX format](#)

FAccT '25: The 2025 ACM Conference  
on Fairness, Accountability, and  
Transparency  
June 23 - 26, 2025  
Athens, Greece

# Time Can Invalidate Algorithmic Recourse

Giovanni De Toni

Fondazione Bruno Kessler, University of Trento  
Trento, Italy  
giovanni.detoni@unitn.it

Bruno Lepri

Fondazione Bruno Kessler  
Trento, Italy  
lepri@fbk.eu

Stefano Teso

University of Trento  
Trento, Italy  
stefano.teso@unitn.it

Andrea Passerini

University of Trento  
Trento, Italy  
andrea.passerini@unitn.it

## Abstract

Algorithmic Recourse (AR) aims to provide users with actionable steps to overturn unfavourable decisions made by machine learning predictors. However, these actions often take time to implement (e.g., getting a degree can take years), and their effects may vary as the world evolves. Thus, it is natural to ask for recourse that remains valid in a dynamic environment. In this paper, we study the robustness of algorithmic recourse over time by casting the problem through the lens of causality. We demonstrate theoretically and empirically that (even robust) causal AR methods can fail over time except in the – unlikely – case that the world is *stationary*. Even more critically, unless the world is fully deterministic, *counterfactual* AR cannot be solved optimally. To account for this, we propose a simple yet effective algorithm for temporal AR that explicitly accounts for time under the assumption of having access to an estimator approximating the stochastic process. Our simulations on synthetic and realistic datasets show how considering time produces more resilient solutions to potential trends in the data distribution.

## CCS Concepts

• **Social and professional topics**; • **Computing methodologies**  
→ **Machine learning**;

## Keywords

Recourse, Counterfactual Explanations, Causality, Time Series

### ACM Reference Format:

Giovanni De Toni, Stefano Teso, Bruno Lepri, and Andrea Passerini. 2025. Time Can Invalidate Algorithmic Recourse. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*, June 23–26, 2025, Athens, Greece. ACM, New York, NY, USA, 19 pages. <https://doi.org/10.1145/3715275.3732008>

## 1 Introduction

Machine Learning (ML) models play an increasingly prominent role in high-stakes decision-making tasks such as credit lending [7], bail approval [20] and medical diagnosis [82]. Over the past

years, concerns have risen regarding the safety and fairness of these systems [10, 30], as well as their impact on human dignity and autonomy [39, 40]. Indeed, there have been examples of algorithmic systems providing discriminatory assessments in contexts such as crime justice [4], employment decisions [73] and benefit allocations [34]. Consequently, there is a growing consensus that algorithmic systems must offer users the means to challenge their decisions, thus preserving human agency - a principle now reflected in recent AI legislation such as the GDPR and the European AI Act [1, 23]. *Algorithmic Recourse* (AR) [80] aims to identify counterfactual explanations that users can follow to overturn unfavourable machine decisions. A counterfactual explanation highlights “*the smallest change to the feature values that changes the prediction to a predefined output*” [31]. For example, in a credit lending task, a counterfactual recourse might suggest a user obtain a Master’s degree as this will net them a higher income, thereby improving loan approval prospects.

In order to be of use, suggested recourse must be *actionable* [75] and sufficiently inexpensive for the user to implement [17]. We argue that actionability inherently includes considerations of *timing*. Specifically: (i) implementing and realizing the effects of recourse actions takes time, and (ii) the same action performed at different times may yield different outcomes. For instance, earning a Master’s degree requires several years and only begins to influence salary after some delay. Additionally, pursuing a degree later in life may impose a substantial financial burden while providing a comparatively smaller salary increase, often due to a *ceiling effect* [46]. In realistic mortgage applications data, real features that a counterfactual might suggest modifying include the debt-to-income ratio, acquiring collateral, or addressing insufficient liquidity [9]. Yet again, improving such features demands significant effort and time.<sup>1</sup> Our key insight is that, since *time plays a key role in the effectiveness of recourse*, one has to ensure recourse suggestions should be *robust to temporal factors*. Specifically, recourse should lead to a favourable outcome regardless of when the user acts or, at a minimum, at a user-defined future point. Figure 1a illustrates this concept.

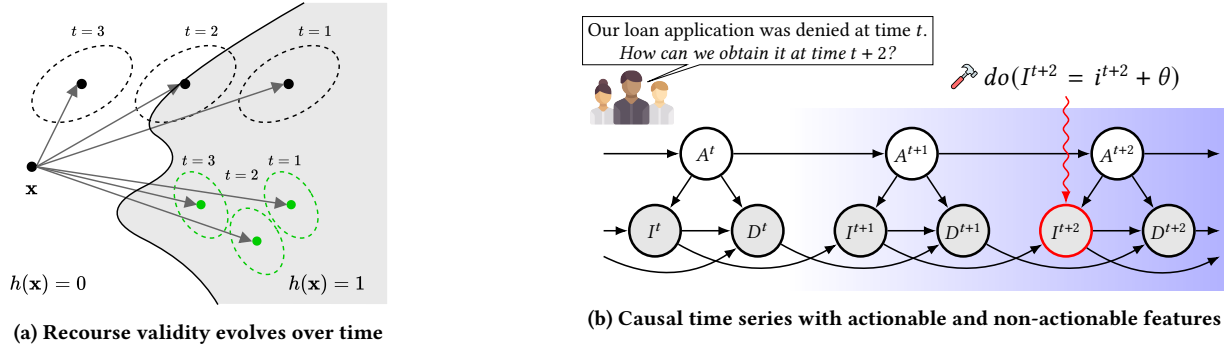
While many studies focus on improving the robustness of recourse in various contexts [38], to the best of our knowledge, little



This work is licensed under a Creative Commons Attribution 4.0 International License. *FAccT '25, Athens, Greece*

© 2025 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-1482-5/25/06  
<https://doi.org/10.1145/3715275.3732008>

<sup>1</sup> While this paper presents examples from the financial sector, similar challenges arise in other domains, such as the labour market [29] and college admissions [36], all of which are increasingly influenced by AI-driven decision systems. These domains also have a potential need for algorithmic recourse recommendations, which must account for trends (e.g., evolving job market skills) and the timing of users’ ability to act on these recommendations.



**Figure 1:** (a) A recourse suggestion may fail to achieve a positive outcome ( $h(x) = 1$ ) over time, even when accounting for the uncertainty surrounding the individual  $x$  requesting recourse (black). In contrast, a temporally robust recourse suggestion ensures a positive classification, regardless of the time  $t$  at which it is applied (green). (b) Illustration of a time series described by a fictional causal graph. Each node represents the value of the  $j$ -th feature at time  $t$ , categorized as either actionable ( $\square I$ : income,  $\square D$ : debt-to-income ratio) or non-actionable ( $\square A$ : age). Arrows depict temporal causal relationships between variables, while nodes showing future timesteps are shown against a progressively bluer background. Similar graphs are used in our simulations (cf. Section 4). In temporal algorithmic recourse (Definition 3.2), user actions are modelled as *robust interventions* ( $\color{red}\curvearrowright$ ), modifying one or more actionable features (red) to ensure recourse at a specific future  $t + \tau$ .

has been done to formalize and study recourse robustness over time explicitly. For instance, Upadhyay et al. [74] examine recourse under model shifts due to retraining, Pawelczyk et al. [58] address scenarios where users imperfectly execute recourse actions, and Dominguez-Olmedo et al. [19] explore robustness against input specification errors. However, explicit consideration of time as a factor in recourse robustness remains limited. Beretta and Cinquini [8] recently incorporated time into the recourse cost but did not address its impact on the validity of recourse recommendations.

In this paper, we study the problem of time in algorithmic recourse through the lens of *causality* [61]. In *causal algorithmic recourse* [8, 19, 44, 50], recourse suggestions are modelled as *interventions* on the user’s features, thus giving a reliable representation of how the features will hypothetically change as the user acts on them to achieve recourse, provided we know the (approximate) *causal model* [44]. We consider a novel setting where we are asked to provide recourse for a causal discrete-time stochastic process characterized by trends. An example of this setting is depicted in Figure 1b. Our theoretical and empirical results challenge the usefulness of the mainstream variants of causal and non-causal AR when extended over time, showing their recommendations *can become invalid* as time progresses.

**Our contributions.** Summarizing, we (i) introduce a sound but intuitive formalization of temporal causal AR (Definition 3.2), based on causality [61] and time-series with independent noise [62], (ii) show theoretically how uncertainty and non-stationarity hinder optimal counterfactual and sub-population recourse (CAR and SAR, [44]) for simple discrete-time stochastic processes, (iii) show how robustifying causal recourse via *uncertainty sets* [19] is not enough to counteract time (Proposition 3.9), (iv) present numerical simulations showcasing the detrimental effects of time on robust (non-)causal AR approaches, and (v) propose a simple time-aware algorithm (Algorithm 1) that mitigates these issues in synthetic and realistic scenarios (Section 4).

## 2 Preliminaries and Related Work

Throughout, we indicate (random) variables  $X$  in upper case, constants  $x$  in lower case, vectors in bold  $\mathbf{x}$ , and sets  $\mathcal{X}$  in italics. We also abbreviate  $\{1, \dots, n\}$  as  $[n]$ .

**Causality.** *Structural Causal Models* (SCMs) [61] allow us to formalize and reason about the causal behaviour of a system. An SCM  $\mathcal{M} = (\mathbf{X}, \mathbf{U}, P, \mathcal{F})$  encompasses endogenous variables  $\mathbf{X} = \{X_i\}_{i=1}^d$ , noise variables  $\mathbf{U} = \{U_i\}_{i=1}^d$  distributed according to  $P(\mathbf{U})$ , and structural assignments  $\mathcal{F}$  of the form  $X_i := f_i(\mathbf{Pa}_i, U_i)$  that describe all causal relationships between variables and their *direct causes* (or parents)  $\mathbf{Pa}_i \subseteq \mathbf{X} \setminus X_i$  e.g., the link between education level and income. An SCM induces a pushforward distribution  $P(\mathbf{X}, \mathbf{U}) = P(\mathbf{X} | \mathbf{U})P(\mathbf{U})$ , where  $P(\mathbf{X} | \mathbf{U})$  is deterministic. *Hard interventions*  $do(\mathbf{X}_{\mathcal{I}} = \theta)$  allow implementing external actions on an SCM to simulate their causal effect on the system, e.g., understanding the impact of getting either a PhD or a Master’s on our salary. They replace a subset of variables  $\mathbf{X}_{\mathcal{I}} \subseteq \mathbf{X}$  with constants  $\theta \in \mathbb{R}^{|\mathbf{X}_{\mathcal{I}}|}$ , detached from their original parents, yielding a new SCM  $\mathcal{M}^{do(\theta)}$  with updated structural assignments  $\mathcal{F}^{do(\theta)}$  and an associated *interventional distribution*  $P^{do(\theta)}(\mathbf{X})$ . Instead, *soft interventions*  $do(\mathbf{X}_{\mathcal{I}} = \mathbf{x}_{\mathcal{I}} + \theta)$  change how the affected variables depend on their parents without detaching them. We shorten both kinds of intervention as  $do(\theta)$  for readability. SCMs also enable us to reason *counterfactually* on what would have happened if the world were different due to an intervention  $do(\theta)$ , all else being equal. For example, we could simulate questions such as “*What would have happened if we had undertaken a PhD rather than a Master’s?*”. Given a realization  $\mathbf{x}$ , the *counterfactual distribution*  $P^{do(\theta), \mathbf{X}=\mathbf{x}}(\mathbf{X})$  is obtained by first *abducting* the exogenous factors  $\mathbf{U}$  in the original SCM and then inferring the state of  $\mathbf{X}$  in the intervened SCM, that is,  $P^{do(\theta), \mathbf{X}=\mathbf{x}}(\mathbf{X}) = P^{do(\theta)}(\mathbf{X} | \mathbf{U})P(\mathbf{U} | \mathbf{X} = \mathbf{x})$  [61, Theorem 7.1.7]. If the structural equations are invertible,  $P(\mathbf{U} | \mathbf{x})$  is deterministic, and so is the counterfactual distribution.

**Causal algorithmic recourse.** In AR, the key quantities of interest are the *user's state*  $\mathbf{x} \sim P(\mathbf{X})$ , which encapsulates the user's features relevant to the decision-making task (e.g., for a loan application: salary, debt-to-income ratio, occupation, etc.), and the *outcome*  $y \sim P(Y | \mathbf{X})$ , representing the predicted event (e.g., whether the user will repay their loan). The SCM underlying  $P(\mathbf{X})$  is assumed to be known or estimated from data [44], enabling us to apply interventions to evaluate the effect of changing the user's state while considering all causal dependencies between variables. Given a (potentially black-box) classifier  $h : \mathbf{x} \mapsto [0, 1]$  approximating  $P(Y | \mathbf{X})$  and a realization  $\mathbf{x}$  yielding an undesirable outcome, i.e.,  $h(\mathbf{x}) < 1/2$ , AR involves finding an intervention  $\theta^*$  that, once implemented by the user, leads in expectation to a more favourable outcome. There are two mainstream approaches to causal AR [44]. *Sub-population recourse* (SAR) provides recourse to users belonging to a specific sub-group, and it is defined as:<sup>2</sup>

$$\theta^* \in \operatorname{argmin}_{\theta \in \mathbb{R}^d} \mathbb{E}_{\hat{\mathbf{x}}} [C(\hat{\mathbf{x}}, \mathbf{x})] \quad \text{s.t.} \quad \mathbb{E}_{\hat{\mathbf{x}}} [h(\hat{\mathbf{x}})] \geq 1/2 \quad (1)$$

Here,  $C$  is a non-negative cost function measuring the user's effort, e.g., the  $\ell_2$ -norm. Selecting a user belonging to a specific subgroup amounts to sampling  $\hat{\mathbf{x}}$  in Eq. (1) by conditioning on a subset of variables  $P^{do(\theta)}(\mathbf{X} | \mathbf{X}_{nd(I)} = \mathbf{x}_{nd(I)})$  where  $nd(I)$  indicates the non-descendants of the intervened upon variables  $I$ . Instead, *counterfactual recourse* (CAR) allows computing *individualized* recourse for the specific individual  $\mathbf{x}$ . It is formulated like Eq. (1), except that the interventional distribution  $P^{do(\theta)}(\mathbf{X} | \mathbf{X}_{nd(I)} = \mathbf{x}_{nd(I)})$  is replaced with the counterfactual distribution  $P^{do(\theta), X=\mathbf{x}}(\mathbf{X})$ . Lastly, since providing optimal recourse (Eq. (1)) is harder for unrestricted SCMs [44], causal AR typically assumes the SCM of  $P(\mathbf{X})$  belongs to an identifiable and invertible class, e.g., Additive Noise Models (ANMs) [19, 41, 43].

**Further related works.** Our work extends the literature on causal algorithmic recourse [19, 44, 50, 79] and it builds upon research in robust algorithmic recourse, counterfactual explanations for time series, and causality. The robustness literature focuses on generating counterfactual explanations that remain stable despite changes in the model  $h$  [25, 32, 52, 55, 60, 74], endogenous dynamics [2], model multiplicity [48, 57], noisy intervention execution [58, 78], and input instance uncertainty [5, 19, 70]. We remark that none of these works explicitly formalize time as a dimension in their computational model. Recently, Fonseca et al. [26] studied empirically algorithmic recourse in a multi-agent setting focusing instead on the *competition* for resources over time, without a causal notion. The literature on counterfactual explanations for multivariate time series provides techniques to generate explanations for stochastic processes [6, 18, 81] by considering *independently manipulable features* (IMF) (i.e., no causal relationships between variables), Pearl causality [61] or Granger causality [28]. Nevertheless, practical causal inference over time remains an open research area, especially its conjunction with algorithmic recourse [13]. Given the breadth and depth of the recourse and explainability fields, our discussion here is necessarily limited. We refer interested readers to existing surveys [31, 38, 42, 77] that cover many other aspects of these frameworks.

<sup>2</sup>The choice of  $\theta$  is often restricted by actionability requirements (e.g., age cannot be changed at will) or other constraints (e.g., monotonicity: age can only increase). We omit this detail for readability.

### 3 Algorithmic Recourse in Time

In real-world applications, users do not implement nor complete suggested interventions immediately. For instance, improving our debt-to-income ratio to a satisfactory level can take months. This is problematic because  $P(\mathbf{X}, Y)$  might change over time due to, e.g., inflation rates, seasonality of loan interests and classifier updates, meaning that *recourse produced by existing approaches could become ineffective or even counterproductive in the future*. Fig. 2 shows the empirical average validity (% of interventions achieving recourse) of state-of-the-art robust (non-)causal recourse methods on different binary decision problems, where  $P(\mathbf{X}, Y)$  is a stochastic process exhibiting a non-linear trend. Unfortunately, *current robust (non-)causal recourse methods are increasingly fragile to time proportionally to the trend's strength*. This section aims to provide a theoretical description and analysis of this issue. We present a summary of our findings in Table 1, along with the characteristics of the (non-)causal recourse methods considered.

#### 3.1 Formalizing Temporal Causal AR

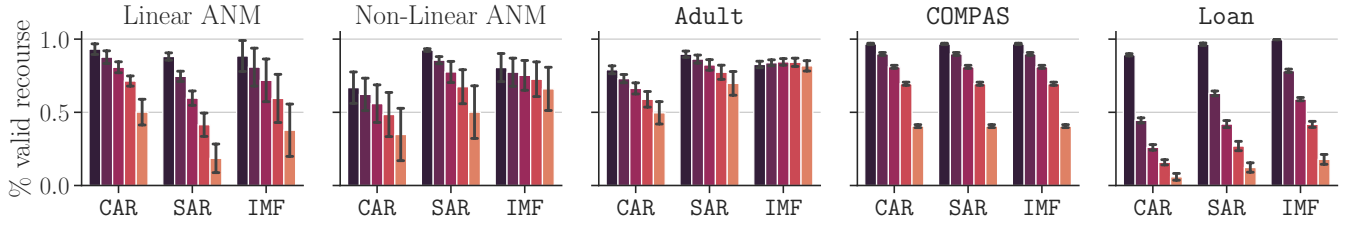
Before proceeding, we need to extend causal AR with a time dimension. We do so by considering a stochastic process  $P(\mathbf{X}^t, Y^t)$  capturing the evolution of the user's state and its relationship to the target variable over time  $t \in \mathbb{N}$ . In the following, we assume that  $P$  is induced by an SCM over the same variables in which the parents of each variable lie in the past, and specifically within a fixed (but otherwise arbitrary) horizon  $\rho \geq 1$ , i.e.,  $\mathbf{Pa}_{X_i^t} \subseteq \bigcup_{\delta=0}^{\rho} \mathbf{X}^{t-\delta}$ , and similarly for  $Y^t$ . We also assume that the outcome  $Y^t$  only depends on the user's states  $\mathbf{X}^1, \dots, \mathbf{X}^t$  up to time  $t$ . We investigate the effect of time on recourse by considering stochastic processes where  $P(\mathbf{X}^t)$  might not be stationary, while the conditional distribution  $P(Y^t | \mathbf{X}^t)$  is unchanged, similarly to *covariate shift* [68]. For this reason, we can assume the classifier  $h^t$  is fixed.<sup>3</sup>

**How do we describe time series with causality?** In the remainder we assume the stochastic process is a *Time series Model with Independent Noise* (TiMiNo), adapted from [62]:

*Definition 3.1 (TiMiNo for Algorithmic Recourse).*  $P(\mathbf{X}^t, Y^t)$  satisfies TiMiNo if it causally factorizes as  $X_i^t = f_{X_i}^t(\mathbf{Pa}_{X_i^t}) + U_{X_i}^t$  and  $Y^t = f_Y^t(\mathbf{X}^t) + U_Y^t$ , for all  $i \in [d]$ , where  $U_{X_i}^t, U_Y^t$  are jointly independent and identically distributed for all  $i \in [d]$  and  $t \in \mathbb{N}$ .

Under appropriate conditions, TiMiNo SCMs are *invertible*, allowing us to apply causal reasoning to infer the counterfactual distribution when computing AR, and can be identified from observational data. Specifically, under appropriate choices of the family of  $f_{X_i}^t, f_Y^t$  – which still allow them to be non-linear – and  $P(U^t)$  we are guaranteed to identify both the *summary graph* and *full-time graph* [62]. Moreover, Peters et al. [62] provides a causal discovery procedure for *TiMiNo* time series that avoids drawing wrong causal conclusions in the presence of confounders. Throughout the paper, we assume the full-time graph is *sufficient* (e.g., there are no unobserved confounders [63]). We will show how, even in this

<sup>3</sup>If this is not the case, one option is to leverage existing techniques for addressing changes due to retraining, such as Upadhyay et al. [74] and Pawelczyk et al. [60]. These works and ours are complementary and studying their interplay is a promising avenue for future work.



**Figure 2: Algorithmic recourse is not robust in time. Empirical average validity and standard deviation of robust counterfactual (CAR), sub-population (SAR) and non-causal recourse (IMF) at time  $t = 50$  for synthetic (Section 4.1) and realistic (Section 4.2) time series with a non-linear trend. We report the validity varying the strength  $\alpha$  of the trend. Legend: ■ 0 ■ 0.3 ■ 0.5 ■ 0.7 ■ 1.0.**

optimistic scenario, our negative results for temporal recourse still hold.

**How do we represent user interventions in time?** Intuitively, the validity of recourse can only be compromised by changes occurring *after* the time  $t$  at which the recourse is issued, if these changes influence the distribution of the future state  $\mathbf{X}^{t+\tau}$ . In the unlikely scenario where recourse is framed as a hard intervention,  $do(\mathbf{X}^{t+\tau} = \theta)$ , affecting *all* variables  $\mathbf{X}$ , the future state  $\mathbf{X}^{t+\tau}$  becomes entirely independent of the past states. This is because hard interventions *detach* all variables from their causal parents, effectively overriding temporal dependencies and eliminating potential temporal effects [19]. Consequently, recourse in this case remains valid *by construction*. In practice, however, recourse: (i) is typically *restricted to a subset of variables  $\mathcal{I}$* , due to constraints such as actionability or cost; and (ii) may be implemented as a *soft intervention*, represented as  $do(\mathbf{X}_{\mathcal{I}}^t = \mathbf{x}_{\mathcal{I}}^t + \theta)$ , without fully detaching  $\mathbf{X}^t$  from its causal structure [19]. In this work, we focus on this more realistic setting of soft interventions, where only a limited subset of variables is targeted, and temporal dependencies persist.

The next challenge lies in capturing the *temporal nature* of actions that decision subjects can realistically perform within the causal framework. Consider the following example: at time  $t$ , Alice is denied a loan application, and the recourse action is “get a degree.” Suppose it takes Alice  $\tau$  timesteps to complete the degree. She begins studying at  $t$  (e.g., enrolling in university) and obtains the degree at  $t + \tau$ . The action “get a degree” represents a process initiated at time  $t$  (enrolling) but whose causal effect materializes at  $t + \tau$  (degree completion). To model this, we simplify the causal framework by treating the “enrol/get a degree” process as a single intervention  $\theta$  applied at  $t + \tau$ , assuming its causal effects are fully realized at that point. Thus, since all user actions require time to implement, we define any action as an intervention occurring at a future  $t + \tau$ , where  $\tau > 0$ .

For the rest of the paper, in the context of temporal algorithmic recourse, we will use the  $do(\theta)$  notation to represent an intervention **always** applied at time  $t + \tau$ , unless specified otherwise. Additionally, this formulation assumes our SCM describing the time series must exhibit *instantaneous effects* [62]. As we will demonstrate in the remainder of this paper, even under this simplified intervention model, the validity of algorithmic recourse is highly sensitive to potential trends or changes in the underlying stochastic process. Now, equipped with the previous definitions, we can formulate the Temporal Causal AR optimization problem as follows:

**Definition 3.2 (Temporal Causal AR).** Consider a stochastic process  $P(\mathbf{X}^t, Y^t)$ , a cost function  $C(\cdot, \cdot)$  (e.g., the  $\ell_2$  norm), a constant classifier  $h$  and a user  $\mathbf{x}^t$  such that  $h(\mathbf{x}^t) < 1/2$ . We want to find the cheapest intervention  $do(\mathbf{X}_{\mathcal{I}}^{t+\tau} = \mathbf{x}_{\mathcal{I}}^{t+\tau} + \theta)$  that achieve recourse at time  $t + \tau$ , in expectation, assuming the causal effects of the action fully manifest at time  $t + \tau$ :

$$\theta^* \in \min_{\theta \in \mathbb{R}^d} \mathbb{E} [C(\hat{\mathbf{x}}^{t+\tau}, \mathbf{x}^{t+\tau})] \quad \text{s.t.} \quad \mathbb{E} [h(\hat{\mathbf{x}}^{t+\tau})] \geq 1/2 \quad (2)$$

where  $\hat{\mathbf{x}}^{t+\tau}$  can be sampled from either the interventional (Eq. (3)), or the counterfactual distribution (Eq. (4)):

$$\hat{\mathbf{x}}^{t+\tau} \sim P^{do(\mathbf{X}_{\mathcal{I}}^{t+\tau} = \mathbf{x}_{\mathcal{I}}^{t+\tau} + \theta)}(\mathbf{X}^{t+\tau} \mid \mathbf{X}_{nd(\mathcal{I})}^{t+\tau} = \mathbf{x}_{nd(\mathcal{I})}^{t+\tau}) \quad (3)$$

$$\hat{\mathbf{x}}^{t+\tau} \sim P^{do(\mathbf{X}_{\mathcal{I}}^{t+\tau} = \mathbf{x}_{\mathcal{I}}^{t+\tau} + \theta); \mathbf{X}^{t+\tau} = \mathbf{x}^{t+\tau}}(\mathbf{X}^{t+\tau}) \quad (4)$$

In contrast,  $\mathbf{x}^{t+\tau}$  is sampled from  $P(\mathbf{X}^{t+\tau} \mid \mathbf{X}^t = \mathbf{x}^t)$ , conditioned on the initial user state  $\mathbf{x}^t$ .

As with non-temporal causal recourse, Definition 3.2 describes both *temporal subpopulation causal AR* (T-SAR), for a user belonging to a specific sub-group, and individualized *temporal counterfactual causal AR* (T-CAR) except that now the passage of time plays a central role in the optimization problem. We remark that, while practical solutions for T-SAR can be devised (see Section 3.6), T-CAR is intrinsically more challenging (as we discuss in Section 3.2).

**Naïvely recomputing recourse is not possible.** A seemingly straightforward “solution” to address the impact of time is to allow users to recompute recourse at time  $t + \tau$ , using their updated state  $\mathbf{x}^{t+\tau}$ . However, this approach is simply not possible because the implementation of the newly suggested action may itself require additional time,  $\tau' > \tau$ . For instance, even if Alice were prepared to act on the updated recourse suggestion (e.g., obtain a degree) at  $t + \tau$ , it is unrealistic for her to achieve this instantaneously. By the time Alice completes the action—at some later time  $t + \tau'$  where  $\tau' > \tau$ —her state may no longer satisfy the updated requirements for recourse. In such a scenario, we might need to propose yet another action, such as “raise yearly income by 10%” potentially leading to the same negative outcome.

### 3.2 Uncertainty Over Time Compromises Counterfactual Recourse

We begin by studying temporal counterfactual AR (T-CAR). It provides *individualized* suggestions, thus achieving the true optimal intervention for a given user [44]. According to Definition 3.2,

**Table 1: Overview of the theoretical results and (non-)causal recourse methods. We summarize the characteristics of the (non-)causal algorithmic recourse methods we considered in this work and the related theoretical results (Section 3). T-SAR is the only approach robust to time (upon possessing an estimator of the stochastic process). IMF is a non-causal method, so it assumes features can be manipulated *independently from each other*, which is seldom the case in applications.**

Method	Causality	Recourse	Robust to time
IMF (Wachter et al. [80])	-	Individualized	✗ (by Corollary 3.7 and Proposition 3.9)
CAR (Karimi et al. [44])	Counterfactual	Individualized	✗ (by Corollary 3.7 and Proposition 3.9)
SAR (Karimi et al. [44])	Interventional	Sub-population	✗ (by Corollary 3.7 and Proposition 3.9)
T-CAR (Definition 3.2)	Counterfactual	Individualized	✗ (by Proposition 3.3 and Corollary 3.4)
T-SAR (Definition 3.2)	Interventional	Sub-population	✓ (upon having an estimator $\tilde{P}(X^t)$ )

it must be computed considering the *future* counterfactual distribution  $p^{do(\theta); X^{t+\tau}=\mathbf{x}^{t+\tau}}(X^{t+\tau})$ . Unfortunately, this turns out to be problematic. The following proposition shows that this distribution cannot be recovered exactly except under strong assumptions.

**PROPOSITION 3.3.** *Let  $P(X^t, Y^t)$  satisfy TiMINo. Given a realization  $\mathbf{x}^t$  and an intervention  $\theta \in \mathbb{R}^d$ , we can recover the counterfactual distribution over the future  $p^{do(\theta); X^{t+\tau}=\mathbf{x}^{t+\tau}}(X^{t+\tau})$  if and only if, for all  $t > 0$ ,  $\text{Var}(U^t) = 0$  and  $\mathbb{E}[U^t]$  are constant.*

All proofs can be found in Appendix A. In words, given  $\mathbf{x}^t$ , one cannot know the true counterfactual distribution at time  $t + \tau$  unless all exogenous factors have zero variance, that is  $P(X^{t+\tau}, \dots, X^{t+1} | \mathbf{x}^t)$  is *deterministic*. Proposition 3.3 has profound consequences for recourse because, recalling the central role of the counterfactual distribution in Eq. (1), it entails that *for non-deterministic processes we cannot solve T-CAR optimally*.

**COROLLARY 3.4 (INFORMAL).** *Let  $P(X^t)$  satisfy TiMINo and consider a constant injective classifier  $h$ . Given a realization  $\mathbf{x}^t$ , a counterfactual recourse  $\theta \in \mathbb{R}^d$  applied at time  $t + \tau$ , with  $\tau > 0$ , cannot be optimal unless exogenous factors have zero variance.*

We remark that Proposition 3.3 also holds for non-TiMINo stochastic processes as long as they admit performing abduction (e.g., the structural equations are invertible), and so does Corollary 3.4. We explore this issue empirically in Section 4.

### 3.3 Sub-Population AR Deteriorates in a Non-Stationary World

Given the inherent limitations of temporal counterfactual AR, in the remainder, we focus on temporal *sub-population* AR (T-SAR), which is generally regarded as the most *plausible* form of recourse [44]. The next proposition shows that, insofar as  $P(X^t, Y^t)$  is *stationary*<sup>4</sup> and the classifier  $h$  is constant and injective, recourse that is optimal for *static* sub-population recourse (Eq. 1) remains optimal over time (Eq. 2).

**PROPOSITION 3.5.** *Consider a stationary stochastic process  $P(X^t)$  and a constant injective classifier  $h$ . Any optimum  $\theta^*$  of Eq. (1) is also optimal for Eq. (2) for any time lag  $\tau \in \mathbb{N}$ .*

<sup>4</sup>A discrete stochastic process  $\{X^t\}_{t \in \mathbb{N}}$  is (weak-sense) stationary when it satisfies the following properties:  $\mathbb{E}[X^{t+\tau} - X^t] = 0$  and  $K(t + \tau, t) = K(\tau, 0)$  for all  $t, \tau \in \mathbb{N}$  where  $K(p, q) = \mathbb{E}[(X^p - \mathbb{E}[X^p])(X^q - \mathbb{E}[X^q])]$  is the autocovariance and  $\mathbb{E}[|X^t|^2] < \infty$  for all  $t \in \mathbb{N}$ .

Despite this positive result, the issue is that *stationarity is seldom satisfied in practice*: many real-world processes exhibit trends (e.g., inflation rate, seasonality of loan interests, etc.). The next example shows how recourse can become invalid if the  $P(X^t, Y^t)$  is not stationary, even for a simple one-dimensional *trend-stationary*<sup>5</sup> stochastic process. Full derivations are in Appendix A.

**Example 3.6.** Consider a trend-stationary stochastic process defined by these structural equations:

$$\begin{aligned} X^t &= \alpha X^{t-1} + m(t) + U_X^t, & U_X^t &\sim \mathcal{N}(\mu_X, \sigma_X) \\ Y^t &= \beta X^t + U_Y^t, & U_Y^t &\sim \mathcal{N}(0, 1) \end{aligned} \quad (5)$$

for all  $t$ , where  $\alpha \in (0, 1)$  and  $\beta \in \mathbb{R}$ . The function  $m(t) : \mathbb{R} \rightarrow \mathbb{R}$  represents a *trend* independent of  $X^t$  and  $Y^t$ . We consider a linear trend  $m(t) = -ct + U_m^t$ , where  $U_m^t \sim \mathcal{N}(\mu_m, \sigma_m)$  and  $c \in \mathbb{R}^+$ . Consider the fixed classifier  $h(X^t) = \sigma(Y^t | X^t)$  where  $\sigma(x) = 1/(1 + e^{-x})$ . We have that the optimal intervention  $\theta^{t+\tau} \in \mathbb{R}$  for which we have  $\mathbb{E}[h(X^{t+\tau} + \theta)] \geq 1/2$  can be expressed as:

$$\theta^{t+\tau} = -\alpha^{\tau+1} x^{t-1} - \sum_{i=0}^{\tau} \alpha^{\tau-i} (-c(t+i) + \mu_m + \mu_X) \quad (6)$$

Since  $m(t)$  is monotonically decreasing for  $c > 0$ , we have the optimal interventions satisfy  $\theta^t \leq \theta^{t+\tau}$ , implying that *a recourse issued at time  $t$  becomes invalid as time passes*. Following Proposition 3.5, we can state the following general corollary regarding our ability to provide optimal subpopulation recourse for general stochastic processes:

**COROLLARY 3.7 (INFORMAL).** *Consider a discrete-time process  $P(X^t)$  and a constant injective classifier  $h$ . Unless  $P(X^t)$  is stationary, the optimal intervention  $\theta^*$  achieving recourse can vary depending on  $t, \tau \in \mathbb{N}$ .*

### 3.4 Robust Algorithmic Recourse Is Not Enough To Counteract Time

Given the previous results, we could imagine to *robustify* the recourse procedure to account for non-stationarity of  $P(X^t, Y^t)$ . For example, a common solution to robustify CAR and SAR is to provide an intervention  $\theta$  achieving recourse within a causal uncertainty set  $B(X; \Delta)$  [19]. In this section, we show how *set-based robust causal recourse method falls short when dealing with time*. We first start by defining robust causal AR:

<sup>5</sup>A stochastic process  $\{X^t\}_{t \in \mathbb{N}}$  is trend-stationary when it can be expressed as  $X^t = f(t) + e^t$ , where  $f(t)$  is (non-)linear trend function and  $e^t$  is a stationary stochastic process.

*Definition 3.8 (Adapted from Dominguez-Olmedo et al. [19]).* Consider a realization  $\mathbf{x} \in \mathbb{R}^d$ , a norm  $\|\cdot\|$ , and a tolerance  $\epsilon > 0$ . We define a causal uncertainty set  $B(\mathbf{x}; \Delta) = \{\mathbf{x}' \sim p_{do(\Delta), \mathbf{X}=\mathbf{x}}(\mathbf{X}) : \|\Delta\| \leq \epsilon\}$  as the collection of the causal counterfactuals under small additive perturbations  $\Delta \in \mathbb{R}^d$ . We want to find the cost-minimizing intervention  $\theta$  achieving recourse in *all* the region defined by  $B(\mathbf{x}; \Delta)$ . Thus, the optimization objective for robust recourse becomes:

$$\theta^* \in \operatorname{argmin}_{\theta \in \mathbb{R}^d} \mathbb{E}[C(\hat{\mathbf{x}}, \mathbf{x})] \quad \text{s.t.} \quad \mathbb{E}[h(\hat{\mathbf{x}})] \geq 1/2 \quad \forall \hat{\mathbf{x}} \sim B(\mathbf{x}; \Delta) \quad (7)$$

where  $\hat{\mathbf{x}}$  is distributed according to either the counterfactual or interventional distribution.

A robust intervention might still obtain recourse for later time steps depending on the tolerance  $\epsilon$ . Intuitively, by asking the user to perform a more difficult action (e.g., increase your income by \$1000, instead of \$100), we can provide interventions that are less susceptible to potential dynamics. However, if the intervention is applied too late, we will not achieve recourse, as shown by the following proposition:

**PROPOSITION 3.9.** *Consider a fixed  $\epsilon > 0$ , a trend-stationary process  $P(\mathbf{X}^t)$ , a constant injective classifier  $h$  and realization  $\mathbf{x}^t$  where  $h(\mathbf{x}^t) < 1/2$ . Let us assume we have an optimal  $\epsilon$ -robust intervention  $\theta$  for timestep  $t$ . There always exists a trend  $m : \mathbb{N} \rightarrow \mathbb{R}$  and a positive  $\tau$ , such that  $\mathbb{E}_{\mathbf{x}^{t+\tau} \sim p_{do(\theta)}(\mathbf{X}^{t+\tau} | \mathbf{X}_{nd(I)}^{t+\tau} = \mathbf{x}_{nd(I)}^{t+\tau}, \mathbf{X}^t = \mathbf{x}^t)} [h(\mathbf{x}^{t+\tau})] < 1/2$ .*

### 3.5 On the Stability of Recourse Over Time

In Sections 3.2 to 3.4, we showed how recourse validity can be compromised by the uncertainty and non-stationarity of the stochastic process  $P(\mathbf{X}^t, Y^t)$ , and we also showed how set-based robustness techniques fail over time. However, decision subjects might be willing to accept recourses that slowly become less effective rather than performing more challenging interventions. Thus, we now characterize instead the *rate* at which our recourse suggestion validity decreases, while assuming a non-stationary  $P(Y^t | \mathbf{X}^t)$  approximated by a sequence of classifiers  $h^t$ , one for each  $t \in \mathbb{N}$ .

*Definition 3.10 (Temporal recourse invalidation rate).* Consider a discrete-time stochastic process  $P(\mathbf{X}^t, Y^t)$ , and any classifier  $h^t$  approximating  $P(Y^t | \mathbf{X}^t)$  for each  $t \in \mathbb{N}$ . Given a realization  $\mathbf{x}^t$  and an intervention  $\theta$  such that  $\mathbb{E}[h(\hat{\mathbf{x}}^t)] \geq 1/2$ , where  $\hat{\mathbf{x}}^t \sim p_{do(\theta)}(\mathbf{X}^t | \mathbf{X}_{nd(I)}^t = \mathbf{x}_{nd(I)}^t)$ , we define the temporal invalidation rate after a time-lag  $\tau > 0$  as:

$$\Delta h(\theta; \tau) = \mathbb{E} \left[ \left| h^{t+\tau}(\hat{\mathbf{x}}^{t+\tau}) - h^t(\hat{\mathbf{x}}^t) \right| \right] \quad (8)$$

where  $\hat{\mathbf{x}}^{t+\tau} \sim p_{do(\theta)}(\mathbf{X}^{t+\tau} | \mathbf{X}_{nd(I)}^{t+\tau} = \mathbf{x}_{nd(I)}^{t+\tau})$ .

Let us now consider the setting in which we have a bounded stochastic process  $P(\mathbf{X}^t, Y^t)$ , where  $-k \leq X_i^t \leq k$  for some  $k \in \mathbb{R}^+$  for all  $t \in \mathbb{N}$ . If we have access to a dataset  $\mathcal{D}^t \sim P(\mathbf{X}^t, Y^t)$ , we can use it to train a classifier  $h^t$  via empirical risk minimization. Let us consider a *linear* classifier  $h^t(\mathbf{x}) = \langle \boldsymbol{\beta}^t, \mathbf{x}^t \rangle$  with bounded weights  $-k \leq \beta_i^t \leq k$  e.g., trained via Bounded Least-Squares (BLS) [72]. Then, given an intervention  $\theta$ , we can derive the following upper bound on the recourse instability within a time interval  $(t, t + \tau)$ .

**THEOREM 3.11 (UPPER-BOUND INVALIDATION RATE).** *Consider a discrete-time stochastic process  $P(\mathbf{X}^t, Y^t)$  and a sequence of linear*

*classifiers  $h^t(\mathbf{x}^t) = \langle \boldsymbol{\beta}^t, \mathbf{x}^t \rangle$  approximating  $P(Y^t | \mathbf{X}^t)$ . We assume  $-k \leq \beta_i, X_i \leq k$  for  $k \in \mathbb{R}^+$ . The temporal invalidation rate is upper-bounded as follows:*

$$\Delta h(\theta; \tau) \leq k\sqrt{d} \cdot \mathbb{E} \left[ \|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\| \right] + \mathbb{E} \left[ \|\hat{\mathbf{x}}^{t+\tau} - \hat{\mathbf{x}}^t\| \right] \quad (9)$$

where  $\|\cdot\|$  is the  $\ell_2$ -norm and the expectation is over  $\mathcal{D}^t \sim P(\mathbf{X}^t, Y^t)$  and the training process.

Theorem 3.11 shows how the recourse instability is upper bounded by how much the world varies between  $t$  and  $t + \tau$  in terms of the data distribution  $P(\mathbf{X}^t)$ , and the classifier. Moreover, the size of the problem  $d$  concurs by increasing the worst-case error at a sublinear rate. The upper bound can be useful for non-linear classifiers  $h$  if we consider a linear function approximating locally [69] their decision function close to a realization  $\mathbf{x}^t$ , such as LIME [67]. If our stochastic process is *trend-stationary* as in Example 3.6, we can derive the following upper bound:

**COROLLARY 3.12.** *Consider a discrete-time trend-stationary stochastic process  $P(\mathbf{X}^t)$ , where  $m_i : \mathbb{N} \rightarrow \mathbb{R}$  represents the trend for  $X_i$  and the classifiers  $h^t(\mathbf{x}^t) = \langle \boldsymbol{\beta}^t, \mathbf{x}^t \rangle$ . Let us define  $m^*(t) = \max_{i \in [d]} m_i(t)$  as the largest trend for  $t \in \mathbb{N}$ . Then, we have the upper bound:*

$$\Delta h(\theta; \tau) \leq k \left( \sqrt{d} \mathbb{E} \left[ \|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\| \right] + d(m^*(t+\tau) - m^*(t)) \right) \quad (10)$$

These results assume that the cost function remains constant over time. In Appendix G, we show an analogous result when this is not the case, as in *personalized cost functions* [17, 22].

### 3.6 Accounting for Time in Practice

In Sections 3.2 to 3.5, we showed how recourse validity is hindered by the *uncertainty* and *non-stationarity* of the stochastic process. Given a factual instance  $\mathbf{x}$  and a recourse  $\theta$ , Theorem 3.11 also implies the upper bound can grow quickly depending on the difference of the induced interventional distributions and classifier  $h$  between  $t$  and  $t + \tau$ . Unfortunately, for any model  $h$  and tolerance  $\epsilon$ , there can exist *multiple* (robust) recourses  $\theta$  to choose from [57]. Since (robust) CAR and SAR have *no means to differentiate between recourses*, they might end up suggesting interventions which rapidly become invalid. An alternative is to settle for classical robust AR, which *can* provide some amount of safety w.r.t. time depending on the chosen epsilon  $\epsilon$  (cf. Section 4). The issue with this is that, as shown by Proposition 3.9, choosing  $\epsilon$  without considering how the world changes can be dramatically suboptimal, i.e., robust AR might recommend expensive actions that risk becoming invalid.

Luckily, for T-SAR, we can mitigate these issues, as long as we have access to an estimator of the stochastic process. We present a simple algorithm (Algorithm 1) for temporal *sub-population* algorithmic recourse (Eq. (2)) drawing inspiration from *adversarially robust recourse* [19]. Following the results of Section 3.4, we argue that, instead of providing robust recourse for an arbitrary uncertainty set with a fixed  $\epsilon$ , we need to provide a robust  $\theta$  for a *forecasted* region of the feature space. We do so by extending the notion of uncertainty set  $B(\mathbf{x}^t; \tau)$  to consider the distribution entailed by the TiMINo SCM conditioned on the observed realization,

**Algorithm 1** Generate robust recourse solutions for T-SAR given a future time-lag  $\tau$ ,  $\gamma < 1$ ,  $\lambda > 1$ , a differentiable classifier  $h$ , an estimator  $\hat{P}(X^t)$  and a subset of intervened nodes  $\mathcal{I} \subseteq [d]$ . We denote with  $\ell$  the *binary cross-entropy loss*.

---

**Require:**  $\mathbf{x}^t$ , individual;  $N > 0$ ;  $\lambda > 0$ ;  $\eta > 0$ ;

- 1:  $B(\mathbf{x}^t; \tau) \leftarrow \{\mathbf{x}' \sim \hat{P}(X^{t+\tau} \mid X^t = \mathbf{x}^t)\}$
- 2: **for**  $epochs = 1$  to  $N$  **do**
- 3:     **while**  $\exists \mathbf{x}' \in B(\mathbf{x}^t; \tau)$  s.t.  $ER(\mathbf{x}', \tau; \theta) < 1/2$  **do**
- 4:         **if** it converges **then**
- 5:             **return**  $\theta$
- 6:              $\mathbf{x}^* \leftarrow \operatorname{argmax}_{\mathbf{x}' \in B(\mathbf{x}^t; \tau)} \ell(ER(\mathbf{x}', \tau; \theta), 1)$
- 7:              $\mathcal{L} \leftarrow \ell(ER(\mathbf{x}^*, \tau; \theta), 1) + \lambda \|\theta\|$
- 8:              $\theta \leftarrow \theta - \eta \nabla \mathcal{L}$
- 9:          $\lambda \leftarrow \gamma \lambda$
- 10: **return**  $\theta$

---

$P(X^{t+\tau} \mid X^t = \mathbf{x}^t)$ , after a time lag  $\tau > 0$ :

$$B(\mathbf{x}^t; \tau) = \{\mathbf{x}' \sim P(X^{t+\tau} \mid X^t = \mathbf{x}^t)\} \quad (11)$$

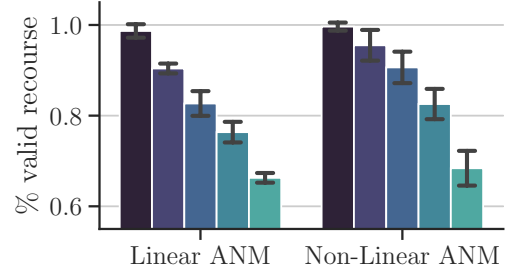
Such a region does not depend on a fixed  $\epsilon$ , thus sidestepping the issue shown by Proposition 3.9. Algorithm 1 assumes to have access to a constant and *differentiable* classifier  $h$ , and to an estimator  $\hat{P}(X^t)$  of the stochastic process. We define  $ER(\mathbf{x}, \tau; \theta) = \mathbb{E}[h(\hat{\mathbf{x}})]$  where  $\hat{\mathbf{x}}$  is sampled from the interventional distribution conditioned on the non-descendant  $nd(\mathcal{I})$  of the intervened upon nodes  $\mathcal{I}$ . Similarly to Dominguez-Olmedo et al. [19], we approximate  $B(\mathbf{x}^t; \tau)$  by sampling a finite number of instances from  $P(X^{t+\tau} \mid X^t = \mathbf{x}^t)$ . As usual, users can control the trade-off between cost and robustness by varying the  $\lambda$  factor (line 5, Algorithm 1). In practice, to solve the optimization problem in Definition 3.2, we follow the causal recourse approach of Dominguez-Olmedo et al. [19] and use projected gradient descent over the (feasible) actions and projected gradient ascent over the temporal uncertainty set (line 6 and 4 Algorithm 1, respectively).

**Computational complexity.** The running time of Algorithm 1 depends on (i) the number of epochs  $N$ , and (ii) an upper bound on the number of iterations  $K$  of the inner loop (lines 3-6). Considering all potential variable subsets  $\mathcal{I} \subset [d]$ , the complexity is  $O(NK2^d)$ . Luckily, not all features are actionable, and the general wisdom is to provide sparse solutions (e.g., by considering only  $|\mathcal{I}| \leq m$  sets) so we have  $O(NK \binom{d}{m})$  if  $m \ll d$ . Lastly, the time-lag  $\tau$  has no impact on the running time of the algorithm, but we would need to run Algorithm 1 for each  $\tau$  specified by the user.

**Relationship between Algorithm 1 and other causal methods.** Algorithm 1 subsumes existing causal recourse methods. If we replace  $B(\mathbf{x}^t; \tau)$  with the uncertainty set in Definition 3.8, we obtain the *robust* counterfactual (CAR) or sub-population (SAR) recourse method [19], depending on how we define the distribution over the expectation in  $ER(\mathbf{x}, \tau; \theta)$ . Moreover, if we do not consider  $\tau$  such as  $ER(\mathbf{x}; \theta) = h(\mathbf{x} + \theta)$  and  $B(\mathbf{x}^t; \Delta) = \{\mathbf{x}^t + \Delta : \|\Delta\| \leq \epsilon\}$ , we obtain (robust) non-causal recourse (IMF, [80]).

## 4 Empirical Evaluation

In this section, we empirically study the effect of time on recourse validity in synthetic and realistic settings taken from the literature,



**Figure 3: Effect of uncertainty on counterfactual AR. Empirical average validity and standard deviation over 10 runs of robust counterfactual algorithmic recourse (CAR) at  $t = 50$ . We vary the variance  $\sigma_U$  of the exogenous factors of the stochastic process. Legend ( $\sigma_U$ ): ■ 0 ■ 0.3 ■ 0.5 ■ 0.7 ■ 1.0.**

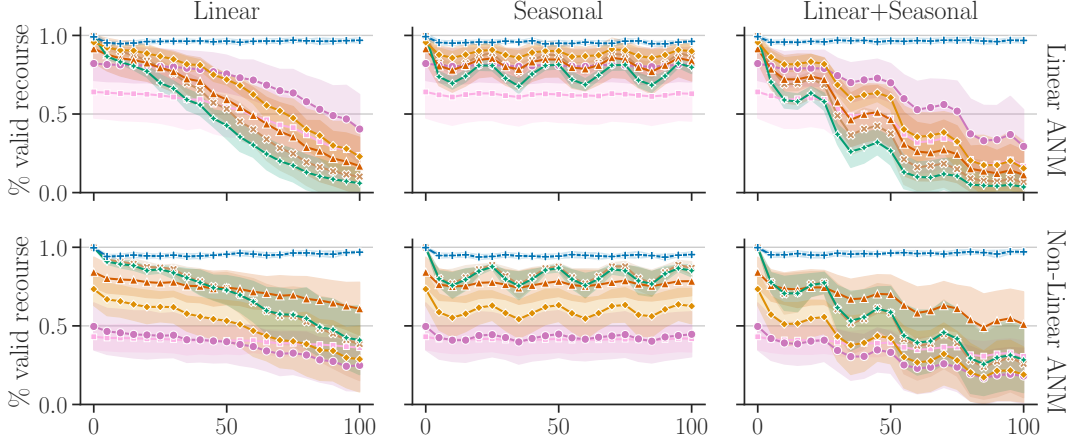
by comparing Algorithm 1 against several robust (non-)causal AR methods. The source code, datasets and raw results are available on GitHub under a permissive license<sup>6</sup>. See Appendix B for a detailed explanation of the experimental setting, techniques, and training setup.

### 4.1 Experiments with Synthetic Time-Series

**Experimental setup.** First, we consider the linear and non-linear 3-variable synthetic ANMs from Karimi et al. [43] representing a binary decision problem (e.g., loan granted/denied). We adapt them to describe a trend-stationary stochastic process by adding an additive trend function  $m(t) = \alpha \cdot (\beta_l \cdot l(t) + \beta_s \cdot s(t))$  to the structural equations, where  $l(t)$  and  $s(t)$  are the *linear* and *seasonal* components. The parameter  $\alpha \in (0, 1)$  governs the strength of the trend. We consider three types of trends: *linear* ( $\beta_l > 0, \beta_s = 0$ ), *seasonal* ( $\beta_l = 0, \beta_s > 0$ ) and *linear+seasonal* ( $\beta_l > 0, \beta_s > 0$ ). Then, we sample a time series for each ANM with 10000 individuals for  $t \in [0, 100]$  timesteps. We split the time series into training (8000) and testing (2000) and train a fixed 3-layer MLP to approximate  $P(Y^t \mid X^t)$  using *only the training data at time  $t = 0$* . We pick 250 individuals negatively classified ( $h(\mathbf{x}) < 1/2$ ) by the MLP from the test set at time  $t = 0$ , and we compute recourse suggestions, by considering the  $\ell_1$ -norm as a cost function, with *robust* counterfactual and sub-population recourse (CAR and SAR, [19]), *robust* non-causal recourse (IMF, [80]) and time-aware sub-population recourse (T-SAR, Algorithm 1). We empirically choose a smaller and larger  $\epsilon \in \{3, 5\}$  maximizing the robust methods' validity at  $t = 0$ . To simulate the user implementing the suggested intervention at a later time, we vary the time lag  $\tau$  and compute the *empirical average validity* (% of interventions achieving recourse) for each method at time  $\tau$ . We repeat the procedure 10 times. In these experiments, we assume to know the true causal graph and structural equations.

**Uncertainty invalidates counterfactual recourse (CAR) over time.** First, we consider a *stationary* version of the synthetic ANMs with exogenous noise  $U_i \sim \mathcal{N}(0, \sigma_U)$  for all  $i \in [3]$  and vary the variance of the exogenous factors  $\sigma_U \in \{0, 0.3, 0.5, 0.7, 1.0\}$ , where  $\sigma_U = 0$  means that the ANM is deterministic. For each

<sup>6</sup><https://github.com/unitn-sml/temporal-algorithmic-recourse>



**Figure 4: Causal algorithmic recourse on diverse time series. Empirical average validity and standard deviation (10 runs) for the robust ( $\epsilon \in \{3, 5\}$ ) and time-aware causal recourse methods for the synthetic ANMs under different trends ( $\alpha = 1.0$ ). Legend: ■ T-SAR ■ CAR ( $\epsilon = 3$ ) ■ SAR ( $\epsilon = 3$ ) ■ IMF ( $\epsilon = 3$ ) and ■ CAR ( $\epsilon = 5$ ) ■ SAR ( $\epsilon = 5$ ) ■ IMF ( $\epsilon = 5$ ).**

value of  $\sigma_U$ , we compute the recourse suggestions using robust counterfactual recourse (CAR). Fig. 3 displays how validity decreases as variance increases showing that the validity over time of (robust) CAR recommendations is strongly impacted by the exogenous noise, as per Proposition 3.3, even in an ideal case in which the SCM is stationary and known. Appendix C shows extended results for  $t \in \{0, 100\}$ .

**Incorporating time is beneficial in causal algorithmic recourse.** Fig. 4 shows how T-SAR (Algorithm 1) achieves superior validity over time than robust (non-)causal methods on the synthetic settings considering the diverse type of trends  $m(t) \in \{\text{Linear, Seasonal, Linear+Seasonal}\}$ . Interestingly, robust causal recourse methods depend highly on the chosen hyperparameters ( $\epsilon$ ,  $\eta$  and  $\lambda$ ) while T-SAR requires less tuning. For example, both CAR and IMF show worse validity on the non-linear ANM when increasing  $\epsilon$ . Lastly, in Appendices D and E we provide further experiments on the *tradeoff* between *validity over time* and *cost*, and on the *sparsity* of recourses, respectively.

## 4.2 Experiments With Realistic Time-Series

In Section 4.1, we assume perfect knowledge of the causal graph and structural equations governing the stochastic process. We now relax these assumptions by learning the structural equations in a data-driven manner, using a simple generative model, on three datasets.

**Experiments setup.** We consider three real-world datasets concerning high-risk decision tasks: recidivism prediction (COMPAS [3]), and loan approval (Adu1t [21] and Loan [43]). They involve categorical and continuous features, some of which are not actionable (e.g., age, ethnicity, etc.). We use the causal graphs defined by Nabi and Shpitser [54] for COMPAS and Adu1t, and Karimi et al. [43] for Loan. We extend these datasets by adding linear+seasonal trends simulating real-world phenomena (e.g., income can fluctuate depending on the job market or individual expenses). Full details are available in Appendix B. Given the ground truth SCMs, we sample an additional

separate time series with 2000 individuals for  $t \in [0, 100]$ , and we use all the samples up to  $t = 50$  to learn an approximate SCM for each real-world dataset. We approximate the structural equations using a CVAE-like generative model [71]. In the experiment, all methods use the same approximate SCMs to compute recourse. Lastly, we perform the same evaluation procedure used for the synthetic experiments.

**Incorporating time is beneficial also with approximate SCMs.** Table 2 shows how T-SAR can provide more robust recourse recommendations than the counterparts by exploiting an estimator  $\hat{P}(X^t)$ . In both Adu1t and Loan, T-SAR shows comparable or better validity than the non-temporal methods. For example, in Loan, T-SAR achieve almost twice the expected average validity ( $\sim 72\%$ ) of the best non-temporal approach SAR ( $\sim 39\%$ ) for  $t = 100$ , albeit exhibiting larger variance. Understandably, T-SAR’s performance hinges on the quality of the underlying estimator. In COMPAS, while T-SAR has good performance for  $t \in \{50, 60, 70\}$ , it gracefully degrades afterwards. This occurs because the trend estimator underestimates the trend impact on the features. In fact, Appendix F shows that, if we employ a perfect estimator  $P(X^t)$ , T-SAR outperforms all non-temporal methods (Fig. 12). As in Section 4.1, the non-temporal methods are sensitive to the hyperparameters, e.g., for COMPAS, CAR provides higher validity for  $\epsilon = 0.05$  rather than  $\epsilon = 0.5$ , while T-SAR needs less tuning.

**Accounting for time ensures more targeted interventions.** Lastly, Fig. 5 shows how T-SAR suggests interventions on feature sets  $\mathcal{I}$  counteracting the effect of the trend more effectively than non-temporal methods. We excluded COMPAS from the analysis since it has a single actionable feature. In Loan, we have two actionable features  $\{\text{income, savings}\}$ , with *income* subject to a trend. T-SAR provide recourse including the trend variable, while the other methods exclude it from the recommendation, thus yielding a lower validity. In Adu1t, we again have two actionable features  $\{\text{education, work-hours-per-week}\}$ , where *work-hours-per-week* is

**Table 2: Effect of time on realistic datasets. Empirical average validity and standard deviation (10 runs) for the robust ( $\epsilon \in \{0.05, 0.5\}$ ) and time-aware causal recourse methods for the realistic datasets under a non-linear trend for  $t \in \{50, 60, 70, 100\}$ . We denote the best results in bold with a shaded background.**

	$t$	CAR		SAR		IMF		T-SAR
		$\epsilon = 0.05$	$\epsilon = 0.5$	$\epsilon = 0.05$	$\epsilon = 0.5$	$\epsilon = 0.05$	$\epsilon = 0.5$	(Algorithm 1)
AduIt	50	0.86 $\pm$ 0.05	0.89 $\pm$ 0.02	0.98 $\pm$ 0.05	0.54 $\pm$ 0.33	<b>0.99 <math>\pm</math>0.02</b>	<b>0.99 <math>\pm</math>0.03</b>	0.98 $\pm$ 0.01
	60	0.77 $\pm$ 0.07	0.86 $\pm$ 0.04	<b>0.97 <math>\pm</math>0.05</b>	0.54 $\pm$ 0.33	0.88 $\pm$ 0.02	0.94 $\pm$ 0.02	0.95 $\pm$ 0.02
	70	0.78 $\pm$ 0.08	0.86 $\pm$ 0.04	<b>0.97 <math>\pm</math>0.05</b>	0.54 $\pm$ 0.32	0.87 $\pm$ 0.03	0.94 $\pm$ 0.02	0.95 $\pm$ 0.02
	100	0.78 $\pm$ 0.08	0.86 $\pm$ 0.05	<b>0.97 <math>\pm</math>0.05</b>	0.54 $\pm$ 0.33	0.88 $\pm$ 0.03	0.93 $\pm$ 0.03	0.96 $\pm$ 0.02
COMPAS	50	<b>1.00 <math>\pm</math>0.00</b>	0.65 $\pm$ 0.35	<b>1.00 <math>\pm</math>0.00</b>	0.80 $\pm$ 0.39	<b>1.00 <math>\pm</math>0.00</b>	<b>1.00 <math>\pm</math>0.00</b>	<b>1.00 <math>\pm</math>0.00</b>
	60	0.85 $\pm$ 0.01	0.64 $\pm$ 0.35	0.94 $\pm$ 0.01	0.80 $\pm$ 0.39	0.56 $\pm$ 0.01	0.64 $\pm$ 0.01	<b>1.00 <math>\pm</math>0.00</b>
	70	0.91 $\pm$ 0.01	0.65 $\pm$ 0.35	0.96 $\pm$ 0.01	0.80 $\pm$ 0.39	0.75 $\pm$ 0.01	0.80 $\pm$ 0.01	<b>0.99 <math>\pm</math>0.02</b>
	100	0.74 $\pm$ 0.02	0.63 $\pm$ 0.34	<b>0.87 <math>\pm</math>0.02</b>	0.79 $\pm$ 0.39	0.43 $\pm$ 0.01	0.51 $\pm$ 0.01	0.78 $\pm$ 0.17
Loan	50	0.86 $\pm$ 0.11	<b>0.92 <math>\pm</math>0.07</b>	<b>0.92 <math>\pm</math>0.05</b>	0.78 $\pm$ 0.24	<b>0.92 <math>\pm</math>0.07</b>	0.91 $\pm$ 0.08	0.82 $\pm$ 0.06
	60	0.39 $\pm$ 0.02	0.65 $\pm$ 0.07	0.54 $\pm$ 0.12	0.60 $\pm$ 0.21	0.46 $\pm$ 0.07	0.54 $\pm$ 0.08	<b>0.89 <math>\pm</math>0.20</b>
	70	0.42 $\pm$ 0.02	0.66 $\pm$ 0.06	0.54 $\pm$ 0.13	0.61 $\pm$ 0.21	0.50 $\pm$ 0.08	0.57 $\pm$ 0.07	<b>0.84 <math>\pm</math>0.29</b>
	100	0.07 $\pm$ 0.03	0.24 $\pm$ 0.08	0.15 $\pm$ 0.09	0.39 $\pm$ 0.25	0.13 $\pm$ 0.06	0.17 $\pm$ 0.07	<b>0.72 <math>\pm</math>0.41</b>

subject to a trend. In this case, T-SAR suggests acting on  $\{\text{work-hours-per-week}\}$  only. Robust sub-population methods (SAR) will instead ask the user to act on both  $\{\text{education, work-hours-per-week}\}$  because they have to robustify on both variables since they cannot forecast how they will change. Non-causal methods (IMF) act on all actionable features but achieve a lower validity since they cannot account for trend effects.

## 5 Discussion and Limitations

We now discuss some limitations of our work, which open up interesting avenues for future work.

**Feasibility of temporal recourse and its evaluation.** Algorithm 1 depends on the quality of the estimator  $\hat{P}(X^t)$  and, if the estimator is flawed, Algorithm 1 could provide sub-optimal recourse. Our experiments on synthetic and realistic datasets confirm that time presents a non-trivial challenge for AR and show how an estimator approximating  $P(X^t)$  can still be useful in some scenarios. However, practical temporal recourse needs reliable time series forecasting, which is well-known to be challenging in various settings [51], because of issues like concept drift [27]. Additionally, it requires effective implementations of causal inference over time [13] which is an ongoing research topic. In practice, we could not fully evaluate the effectiveness of Algorithm 1 in real-world situations as this requires temporal datasets for recourse, which are currently not available or hardly obtainable. For instance, in the case of lending applications, while banks hold temporal data on loan applicants - e.g., records of the same individual reapplying for a loan after some time – these datasets are usually proprietary and not publicly accessible [12, 15]. Indeed, the scarcity of suitable data is a well-known issue affecting the evaluation of AR approaches at large [42, 43, 50].

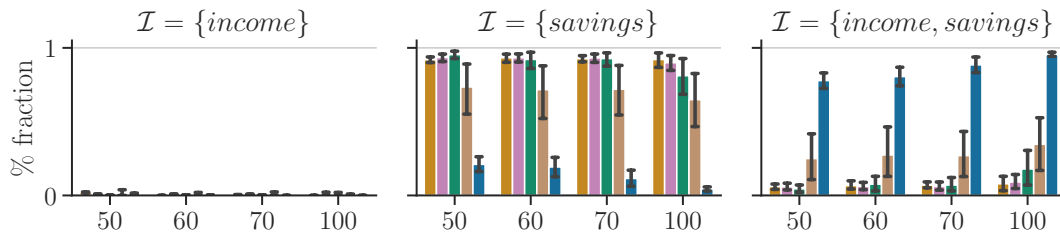
Suitable datasets for temporal recourse tasks, in particular, should at a minimum include the features used by decision-makers

to train classifiers. For instance, in credit lending, financial institutions may rely on both frequently updated data (e.g., credit card spending patterns) and more static attributes (e.g., collateral assets). To capture evolving trends in data distributions, these datasets will likely need to incorporate not only internal institutional records but also external sources such as governmental statistics (e.g., census data). Moreover, certain features may be considered "fixed" and thus not require temporal modeling. In addition, regulatory frameworks could facilitate access to richer datasets; for example, Article 40 of the European DSA<sup>7</sup> permits vetted researchers to access data from large online platforms (e.g., Facebook, Twitter) for research purposes. Extending such provisions to other domains could significantly benefit algorithmic recourse research. However, privacy and ethical concerns remain central, necessitating rigorous data handling and anonymization practices.

**Causal models, trends and interventions.** Our formalization assumes that the stochastic process adheres to the TiMiNo framework [62], but it leaves room for exploring more complex Structural Causal Models (SCMs). Addressing such complexities would necessitate more versatile models for causal inference in time series, such as integrating classical approaches like VAR [83]. Additionally, we do not examine trend models for the classifier  $h$  or the cost function  $C(\cdot)$ , even though such trends may coexist with those of  $P(X^t)$ . Lastly, our task also aligns with alternative paradigms, such as the *dynamic treatment regime* [53], which emphasize constructing a *sequential policy* rather than a single counterfactual to ensure a desired effect at time  $t + \tau$ . We believe our theoretical results can extend to these settings, and developing their connection with AR represents a potential future direction.

**Further ethical considerations.** Our work seeks to achieve *algorithmic contestability* [49] through actionable counterfactual explanations, empowering users to challenge decisions made by

<sup>7</sup><https://eur-lex.europa.eu/eli/reg/2022/2065/oj>



**Figure 5: Effect of time on actionable features. Distribution of the intervention sets  $\mathcal{I}$  over the actionable features achieving recourse on Loan for different  $t$ . Legend: ■ T-SAR ■ CAR ( $\epsilon = 0.05$ ) ■ SAR ( $\epsilon = 0.05$ ) and ■ CAR ( $\epsilon = 0.5$ ) ■ SAR ( $\epsilon = 0.5$ )**

machine learning models. We believe recourse can offer tangible benefits to decision subjects. For instance, in credit lending, recourse actions could help *scarred borrowers* [14], individuals who self-exclude from (re-)applying due to a perceived negative response from banks, by providing tools that enhance their understanding of the decision-making process. However, AR introduces both technical and ethical challenges [76]. For example, counterfactuals must avoid exhibiting *adversarial characteristics* [47] and should aim to improve the user’s ground truth classification.<sup>8</sup> Additionally, fairness considerations must be addressed [79] before implementing recourse solutions in real-world settings. Related research has explored fairness in sequential decision-making within dynamical systems [37, 65], and extending these principles to recourse actions represents a critical step for realistic deployment. Lastly, recourse methods have been shown to pose risks of *private data leakage* [59], potentially exposing sensitive user information to malicious entities. To be effective in the wild, AR applications must take into account all these factors to ensure their effectiveness.

## 6 Conclusions

In this paper, we have investigated the impact of time on algorithmic recourse. Our formalization of temporal causal recourse extends both counterfactual and sub-population causal AR by modelling the world as a (possibly non-stationary) causal stochastic process  $P(X^t, Y^t)$ . It allows us to theoretically demonstrate how *standard and robust AR approaches are fragile*, as their solutions become invalid in the presence of trends and future uncertainty. We also show that a simple algorithm, leveraging an estimator of the stochastic process fitted on historical data, can deliver more robust solutions. Our experiments with causal and non-causal approaches support our findings. With this work, we aim to highlight the negative impact of time on existing AR approaches while demonstrating how these challenges can be at least partially mitigated by leveraging historical data.

## Acknowledgments

The authors thank Martin Pawelczyk for the helpful feedback and discussions on an initial version of the manuscript. The work was

<sup>8</sup>Consider an imaginary university application screening system that evaluates the number of books owned by an applicant’s family, based on the assumption that this metric correlates with academic performance [24]. While purchasing a crate of 1000 books might artificially improve the applicant’s chances of acceptance, it does not enhance their actual academic abilities (unless, of course, they read the books). Such an action would represent an example of an *adversarial recourse suggestion*.

partially supported by the following projects: MUR PNRR project FAIR - Future AI Research (PE00000013) funded by the NextGenerationEU (AP and BL), Horizon Europe Programme, grant #101120237-ELIAS (AP and BL) and grant #101120763-TANGO (ST, AP and BL). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

## References

- [1] EU AI Act. 2021. Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. *Proposal for a regulation of the European parliament and of the council* (2021).
- [2] Patrick Altmeyer, Giovan Angela, Aleksander Buszydlak, Karol Dobiczek, Arie van Deursen, and Cynthia CS Liem. 2023. Endogenous Macrodynamics in Algorithmic Recourse. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. IEEE, IEEE Computer Society, Los Alamitos, CA, USA, 418–431.
- [3] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias: There’s software used across the country to predict future criminals. *And it’s biased against blacks*. *ProPublica* 23 (2016), 77–91.
- [4] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2022. Machine bias. In *Ethics of data and analytics*. Auerbach Publications, 254–264.
- [5] André Artelt, Valerie Vaquet, Riza Velioglu, Fabian Hinder, Johannes Brinkhoff, Malte Schilling, and Barbara Hammer. 2021. Evaluating robustness of counterfactual explanations. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 01–09.
- [6] Emre Ates, Burak Aksar, Vitus J. Leung, and Ayse K. Coskun. 2021. Counterfactual Explanations for Multivariate Time Series. In *2021 International Conference on Applied Artificial Intelligence (ICAAI)*. IEEE, 1–8.
- [7] Luca Barbaglia, Sebastiano Manzan, and Elisa Tosetti. 2021. Forecasting Loan Default in Europe with Machine Learning\*. *Journal of Financial Econometrics* 21, 2 (07 2021), 569–596. doi:10.1093/jfinc/nbab010 arXiv:https://academic.oup.com/jfec/article-pdf/21/2/569/49676594/nbab010.pdf
- [8] Isacco Beretta and Martina Cinquini. 2023. The Importance of Time in Causal Algorithmic Recourse. *arXiv preprint arXiv:2306.05082* (2023).
- [9] Neil Bhutta, Steven Laufer, and Daniel R Ringo. 2017. Residential mortgage lending in 2016: Evidence from the Home Mortgage Disclosure Act data. *Fed. Res. Bull.* 103 (2017), 1.
- [10] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. ‘It’s Reducing a Human Being to a Percentage’: Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI ’18*). Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3173574.3173951
- [11] Lucius EJ Bynum, Joshua R Loftus, and Julia Stoyanovich. 2023. Counterfactuals for the Future. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 14144–14152.
- [12] Alessandro Castelnovo, Riccardo Crupi, Giulia Del Gamba, Greta Greco, Aisha Naseer, Daniele Regoli, and Beatriz San Miguel Gonzalez. 2020. Befair: Addressing fairness in the banking sector. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 3652–3661.
- [13] Martina Cinquini, Isacco Beretta, Salvatore Ruggieri, and Isabel Valera. 2024. A Practical Approach to Causal Inference over Time. *arXiv preprint arXiv:2410.10502* (2024).

- [14] Marc Cowling, Weixi Liu, and Raffaella Calabrese. 2022. Has Previous Loan Rejection Scarred Firms from Applying for Loans during Covid-19? 59, 4 (2022), 1327–1350. doi:10.1007/s11187-021-00586-2 pmid:38624828
- [15] Riccardo Crupi, Alessandro Castelnovo, Daniele Regoli, and Beatriz San Miguel Gonzalez. 2024. Counterfactual explanations as interventions in latent space. *Data Mining and Knowledge Discovery* 38, 5 (2024), 2733–2769.
- [16] Giovanni De Toni, Bruno Lepri, and Andrea Passerini. 2023. Synthesizing explainable counterfactual policies for algorithmic recourse with program synthesis. *Machine Learning* 112, 4 (2023), 1389–1409.
- [17] Giovanni De Toni, Paolo Viappiani, Stefano Teso, Bruno Lepri, and Andrea Passerini. 2024. Personalized Algorithmic Recourse with Preference Elicitation. *Transactions on Machine Learning Research* (2024). <https://openreview.net/forum?id=8sg219zXgO>
- [18] Eoin Delaney, Derek Greene, and Mark T. Keane. 2021. Instance-Based Counterfactual Explanations for Time Series Classification. In *Case-Based Reasoning Research and Development*, Antonio A. Sánchez-Ruiz and Michael W. Floyd (Eds.). Springer International Publishing, Cham, 32–47.
- [19] Ricardo Dominguez-Olmedo, Amir H Karimi, and Bernhard Schölkopf. 2022. On the adversarial robustness of causal algorithmic recourse. In *International Conference on Machine Learning*. PMLR, 5324–5342.
- [20] Julia Dressel and Hany Farid. 2018. The accuracy, fairness, and limits of predicting recidivism. *Science advances* 4, 1 (2018), ea05580.
- [21] Dheeru Dua and Casey Graff. 2017. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml>
- [22] Seyedehdelaram Esfahani, Giovanni De Toni, Bruno Lepri, Andrea Passerini, Katya Tentori, and Massimo Zancanaro. 2024. Preference Elicitation in Interactive and User-centered Algorithmic Recourse: an Initial Exploration. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (Cagliari, Italy) (UMAP '24)*. Association for Computing Machinery, New York, NY, USA, 249–254. doi:10.1145/3627043.3659556
- [23] EP EU Council. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008,(EU) No 167/2013,(EU) No 168/2013,(EU) 2018/858,(EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU,(EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) Off. J. Eur. Union 50 (2024), 202.
- [24] M. D. R. Evans, Jonathan Kelley, and Joanna Sikora. 2014. Scholarly Culture and Academic Performance in 42 Nations. *Social Forces* 92, 4 (06 2014), 1573–1605. doi:10.1093/sf/sou030 arXiv:<https://academic.oup.com/sf/article-pdf/92/4/1573/6889358/sou030.pdf>
- [25] Andrea Ferrario and Michele Loi. 2022. The robustness of counterfactual explanations over time. *IEEE Access* 10 (2022), 82736–82750.
- [26] João Fonseca, Andrew Bell, Carlo Abrate, Francesco Bonchi, and Julia Stoyanovich. 2023. Setting the right expectations: Algorithmic recourse over time. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–11.
- [27] João Gama, Indrè Zliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. 2014. A survey on concept drift adaptation. *ACM computing surveys (CSUR)* 46, 4 (2014), 1–37.
- [28] Clive WJ Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* (1969), 424–438.
- [29] Francis Green and Golo Henseke. 2021. Europe’s Evolving Graduate Labour Markets: Supply, Demand, Underemployment and Pay. 55, 1 (2021), 2. doi:10.1186/s12651-021-00288-y
- [30] Nina Grgic-Hlaca, Elissa M. Redmiles, Krishna P. Gummadi, and Adrian Weller. 2018. Human Perceptions of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction. In *Proceedings of the 2018 World Wide Web Conference (Lyon, France) (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 903–912. doi:10.1145/3178876.3186138
- [31] Riccardo Guidotti. 2024. Counterfactual explanations and how to find them: literature review and benchmarking. *Data Mining and Knowledge Discovery* 38, 5 (2024), 2770–2824.
- [32] Faisal Hamman, Erfan Noorani, Saumitra Mishra, Daniele Magazzeni, and Sanghamitra Dutta. 2023. Robust counterfactual explanations for neural networks with probabilistic guarantees. In *International Conference on Machine Learning*. PMLR, 12351–12367.
- [33] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. 2020. Array programming with NumPy. *Nature* 585, 7825 (Sept. 2020), 357–362. doi:10.1038/s41586-020-2649-2
- [34] Melissa Heikkilä. 2022. Dutch scandal serves as a warning for Europe over risks of using algorithms. <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/>. Accessed: 19-12-2024.
- [35] Hans Hofmann. 1994. Statlog (German Credit Data). UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5NC77>.
- [36] Don Hossler, Emily Chung, Jihye Kwon, Jerry Lucido, Nicholas Bowman, and Michael Bastedo. 2019. A study of the use of nonacademic factors in holistic undergraduate admissions reviews. *The Journal of Higher Education* 90, 6 (2019), 833–859.
- [37] Yaowei Hu and Lu Zhang. 2022. Achieving Long-Term Fairness in Sequential Decision Making. 36, 9 (2022), 9549–9557. Issue 9. doi:10.1609/aaai.v36i9.21188
- [38] Junqi Jiang, Francesco Leofante, Antonio Rago, and Francesca Toni. 2024. Robust Counterfactual Explanations in Machine Learning: A Survey. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, Kate Larson (Ed.). International Joint Conferences on Artificial Intelligence Organization, 8086–8094. doi:10.24963/ijcai.2024/894 Survey Track.
- [39] Anna Jobin, Marcello Lenca, and Efy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature machine intelligence* 1, 9 (2019), 389–399.
- [40] Margot E Kaminski. 2018. Binary governance: Lessons from the GDPR’s approach to algorithmic accountability. *S. Cal. L. Rev.* 92 (2018), 1529.
- [41] A. Karimi, G. Barthe, B. Balle, and I. Valera. 2020. Model-agnostic counterfactual explanations for consequential decisions. In *AISTATS*. PMLR, 895–905.
- [42] Amir-Hossein Karimi, Gilles Barthe, Bernhard Schölkopf, and Isabel Valera. 2022. A Survey of Algorithmic Recourse: Contrastive Explanations and Consequential Recommendations. *ACM Comput. Surv.* 55, 5, Article 95 (dec 2022), 29 pages. doi:10.1145/3527848
- [43] Amir-Hossein Karimi, Bernhard Schölkopf, and Isabel Valera. 2021. Algorithmic recourse: from counterfactual explanations to interventions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 353–362.
- [44] Amir-Hossein Karimi, Julius Von Kügelgen, Bernhard Schölkopf, and Isabel Valera. 2020. Algorithmic recourse under imperfect causal knowledge: a probabilistic approach. *Advances in Neural Information Processing Systems* 33 (2020), 265–277.
- [45] Ralph L. Keeney and Howard Raiffa. 1993. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press. doi:10.1017/CBO9781139174084
- [46] Finn Lattimore, Daniel M Steinberg, and Anna Zhu. 2023. The Economic Effect of Gaining a New Qualification Later in Life. *arXiv preprint arXiv:2304.01490* (2023).
- [47] Tobias Leemann, Martin Pawelczyk, Bardh Prenkaj, and Gjergji Kasneci. 2024. Towards Non-adversarial Algorithmic Recourse. In *World Conference on Explainable Artificial Intelligence*. Springer, 395–419.
- [48] Francesco Leofante, Elena Botoeva, and Vineet Rajani. 2023. Counterfactual explanations and model multiplicity: a relational verification view. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*, Vol. 19. 763–768.
- [49] Henrietta Lyons, Eduardo Velloso, and Tim Miller. 2021. Conceptualising contestability: Perspectives on contesting algorithmic decisions. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–25.
- [50] Ayan Majumdar and Isabel Valera. 2024. CARMA: A practical framework to generate recommendations for causal algorithmic recourse at scale. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1745–1762.
- [51] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. 2020. The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting* 36, 1 (2020), 54–74.
- [52] Anna P Meyer, Dan Ley, Suraj Srinivas, and Himabindu Lakkaraju. 2023. On minimizing the impact of dataset shifts on actionable explanations. In *Uncertainty in Artificial Intelligence*. PMLR, 1434–1444.
- [53] Susan A Murphy. 2003. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 65, 2 (2003), 331–355.
- [54] Razieh Nabi and Ilya Shpitser. 2018. Fair inference on outcomes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.
- [55] Duy Nguyen, Ngoc Bui, and Viet Anh Nguyen. 2023. Distributionally Robust Recourse Action. In *The Eleventh International Conference on Learning Representations*.
- [56] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019).
- [57] Martin Pawelczyk, Klaus Broelemann, and Gjergji Kasneci. 2020. On Counterfactual Explanations under Predictive Multiplicity. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI) (Proceedings of Machine Learning Research, Vol. 124)*, Jonas Peters and David Sontag (Eds.). PMLR, 809–818. <https://proceedings.mlr.press/v124/pawelczyk20a.html>
- [58] Martin Pawelczyk, Teresa Datta, Johannes van den Heuvel, Gjergji Kasneci, and Himabindu Lakkaraju. 2022. Probabilistically robust recourse: Navigating the trade-offs between costs and robustness in algorithmic recourse. *arXiv preprint arXiv:2203.06768* (2022).

- [59] Martin Pawelczyk, Himabindu Lakkaraju, and Seth Neel. 2023. On the Privacy Risks of Algorithmic Recourse. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research, Vol. 206)*, Francisco Ruiz, Jennifer Dy, and Jan-Willem van de Meent (Eds.), PMLR, 9680–9696. <https://proceedings.mlr.press/v206/pawelczyk23a.html>
- [60] Martin Pawelczyk, Tobias Leemann, Asia Biega, and Gjergji Kasneci. 2022. On the Trade-Off between actionable explanations and the right to be forgotten. *arXiv preprint arXiv:2208.14137* (2022).
- [61] Judea Pearl. 2009. *Causality*. Cambridge university press.
- [62] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. 2013. Causal inference on time series using restricted structural equation models. *Advances in neural information processing systems* 26 (2013).
- [63] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. 2017. *Elements of causal inference: foundations and learning algorithms*. The MIT Press.
- [64] Gabriella Pigozzi, Alexis Tsoukias, and Paolo Viappiani. 2016. Preferences in artificial intelligence. *Annals of Mathematics and Artificial Intelligence* 77 (2016), 361–401.
- [65] Miriam Rateike, Isabel Valera, and Patrick Forré. 2024. Designing Long-term Group Fair Policies in Dynamical Systems. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) (FAccT '24). Association for Computing Machinery, New York, NY, USA, 20–50. doi:10.1145/3630106.3658538
- [66] Kaivalya Rawal and Himabindu Lakkaraju. 2020. Beyond individualized recourse: Interpretable and interactive summaries of actionable recourses. *Advances in Neural Information Processing Systems* 33 (2020), 12187–12198.
- [67] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.
- [68] Hidetoshi Shimodaira. 2000. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference* 90, 2 (2000), 227–244.
- [69] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034* (2013).
- [70] Dylan Slack, Anna Hilgard, Himabindu Lakkaraju, and Sameer Singh. 2021. Counterfactual Explanations Can Be Manipulated. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.), Vol. 34. Curran Associates, Inc., 62–75. [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/009c434cab57de48a31f6b669e7ba266-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/009c434cab57de48a31f6b669e7ba266-Paper.pdf)
- [71] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. 2015. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems* 28 (2015).
- [72] Philip B Stark and Robert L Parker. 1995. Bounded-variable least-squares: an algorithm and applications. *Computational Statistics* 10 (1995), 129–129.
- [73] Bill Turque. 2012. "Creative ... motivating" and fired. [https://www.washingtonpost.com/local/education/creative--motivating-and-fired/2012/02/04/gIQAwzZpVR\\_story.html](https://www.washingtonpost.com/local/education/creative--motivating-and-fired/2012/02/04/gIQAwzZpVR_story.html). Accessed: 19-12-2024.
- [74] Sohini Upadhyay, Shalmali Joshi, and Himabindu Lakkaraju. 2021. Towards robust and reliable algorithmic recourse. *Advances in Neural Information Processing Systems* 34 (2021), 16926–16937.
- [75] B. Ustun, A. Spangher, and Y. Liu. 2019. Actionable recourse in linear classification. In *FAT\**. 10–19.
- [76] Suresh Venkatasubramanian and Mark Alfano. 2020. The philosophical basis of algorithmic recourse. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 284–293.
- [77] Sahil Verma, Varich Boonsanong, Minh Hoang, Keegan E Hines, John P Dickerson, and Chirag Shah. 2020. Counterfactual explanations and algorithmic recourses for machine learning: a review. (2020).
- [78] Marco Virgolin and Saverio Fracaros. 2023. On the robustness of sparse counterfactual explanations to adverse perturbations. *Artificial Intelligence* 316 (2023), 103840.
- [79] Julius von Kügelgen, Amir-Hossein Karimi, Umang Bhatt, Isabel Valera, Adrian Weller, and Bernhard Schölkopf. 2022. On the fairness of causal algorithmic recourse. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 9584–9594.
- [80] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2017. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL & Tech.* 31 (2017), 841.
- [81] Jingquan Yan and Hao Wang. 2023. Self-interpretable time series prediction with counterfactual explanations. In *International Conference on Machine Learning*. PMLR, 39110–39125.
- [82] Tae Keun Yoo, Ik Hee Ryu, Geunyoung Lee, Youngnam Kim, Jin Kuk Kim, In Sik Lee, Jung Sub Kim, and Tyler Hyungtaek Rim. 2019. Adopting machine learning to automatically identify candidate patients for corneal refractive surgery. *NPJ digital medicine* 2, 1 (2019), 1–9.

- [83] Eric Zivot and Jiahui Wang. 2006. Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS®* (2006), 385–429.

## A Proofs

PROOF. Consider a TiMINo stochastic process  $\{X_t\}_{t \in \mathbb{N}}$  with structural equations:

$$X_i^t = f(\mathbf{Pa}_i^{t-p}, \dots, \mathbf{Pa}_i^t) + U_i^t, \quad U_i^t \sim \mathcal{N}(\mu_X, \sigma_X), \quad \mu_X, \sigma_X > 0 \quad (12)$$

where  $U_i^t$  have positive mean and variance. We prove the proposition by contradiction by looking at the value of the time lag  $\tau$ . For the sake of clarity, we state the proof for  $p = 1$  (the proof for  $p > 1$  is similar). Assume that we observed a sequence of realizations  $\mathbf{x}^{t-p:t}$  and we can compute the counterfactual distribution  $p_{do}(\mathbf{X}^{t+\tau} = \mathbf{x}^{t+\tau} + \theta), \mathbf{X}^{t-p:t} = \mathbf{x}^{t-p:t}(\mathbf{X}^{t+\tau})$  for any  $\tau \geq 0$  and  $\theta \in \mathbb{R}^d$ . If  $\tau = 0$ , we can immediately recover the value of all exogenous factors via  $u_i^t = x_i^t - f(\mathbf{Pa}_i^{t-p}, \dots, \mathbf{Pa}_i^t)$ , as we know  $\mathbf{x}^{t-p:t}$ . Let us assume we can also recover the exogenous factors for a time lag  $\tau > 0$  even if we did not observe future realizations  $\{\mathbf{x}^t, \dots, \mathbf{x}^{t+\tau}\}$ . Since we did not observe such realizations, the classical abduction step to recover the exogenous factors within the  $[t, t + \tau]$  interval is impossible. Recently, Bynum et al. [11] introduced *forward-looking counterfactuals* to overcome this challenge. They postulate we can still compute counterfactuals over unseen realizations by *propagating* the latest exogenous factors we were able to abduce [11, Section 3]. Similarly, we assume we can *propagate* the latest exogenous factors we can recover ( $u^t$ ) into the future, e.g.,  $U_i^{t+\tau} = u_i^t$  for any  $\tau > 0$  and  $i \in [d]$ . Since we assume we can compute the exact counterfactual distribution, the only setting where  $U_i^{t+\tau} = u_i^t$  holds is if  $\sigma_X = 0$  and  $\mu_X = 0$ . However, we initially stated that  $\mu_X, \sigma_X > 0$ , so we have a contradiction.  $\square$

### A.1 Proof of Proposition 3.5

PROOF. Consider a discrete-time stationary time series  $\{(\mathbf{x}, y)^t\}_{t \in \mathbb{N}}$  and wlog consider a fixed linear classifier  $h(\mathbf{x}) = \beta^\top \mathbf{x}$  that is *injective*, that is,  $\mathbf{x} \neq \mathbf{x}' \Rightarrow h(\mathbf{x}) \neq h(\mathbf{x}')$ . Given a realization  $\mathbf{x}^t$ , let us assume  $\theta^*$  is the optimal intervention for Eq. (2) at time  $t$ , but not at time  $t + \tau$ , with  $\tau \in \mathbb{N}$ . Thus, either our recourse becomes invalid or exceedingly expensive. We focus on the former since we assume a fixed cost function such as the  $\ell_1$ -norm of the intervention  $C(\theta) = \|\theta\|$ . As a consequence, we have  $\mathbb{E}[h(\hat{\mathbf{x}}^{t+\tau}) - h(\hat{\mathbf{x}}^t)] \neq 0$  where  $\hat{\mathbf{x}}^{t+\tau} \sim p_{do}(\theta^*)(\mathbf{X}^{t+\tau} \mid \mathbf{X}_{nd(I)}^{t+\tau} = \mathbf{x}_{nd(I)}^{t+\tau})$  and  $\hat{\mathbf{x}}^t \sim p_{do}(\theta^*)(\mathbf{X}^t \mid \mathbf{X}_{nd(I)}^t = \mathbf{x}_{nd(I)}^t)$ . We denote with  $I \subseteq [d]$  the intervention set of the successful intervention  $\theta^*$ , and with  $do(\theta^*) = do(\mathbf{X}_I^t = \mathbf{x}_I^t + \theta^*)$  the corresponding soft intervention. Consider the case in which our intervention does not achieve recourse after  $\tau$  time steps. We can decompose the previous expectation as follows:

$$\mathbb{E}[h(\hat{\mathbf{x}}^{t+\tau}) - h(\hat{\mathbf{x}}^t)] = \mathbb{E}[\beta^\top \hat{\mathbf{x}}^{t+\tau} - \beta^\top \hat{\mathbf{x}}^t] \quad (13)$$

$$= \beta^\top \mathbb{E}[\hat{\mathbf{x}}^{t+\tau} - \hat{\mathbf{x}}^t] \quad (14)$$

Thus, since  $\theta^*$  is not optimal for  $t + \tau$ , we must have  $\mathbb{E}[\hat{\mathbf{x}}^{t+\tau} - \hat{\mathbf{x}}^t] \neq 0$ . This, however, is a contradiction since we assume the time series is stationary. It follows that, since  $\tau$  is arbitrary, for any  $t$ , the corresponding optimal solution  $\theta^*$  is also optimal for all  $\tau > 0$ . Now, this also means that  $\theta^*$  is also an optimal solution for the classical

recourse optimization problem Eq. (1) as long as we optimize Eq. (1) by considering  $P(X, Y) = P(X^t, Y^t)$  for any time steps  $t \in \mathbb{N}$ .  $\square$

## A.2 Full derivations for Example 3.6

PROOF. Consider a trend-stationary stochastic process defined by these structural equations:

$$\begin{aligned} X^t &= \alpha X^{t-1} + m(t) + U_X^t, & U_X^t &\sim \mathcal{N}(\mu_X, \sigma_X) \\ Y^t &= \beta X^t + U_Y^t, & U_Y^t &\sim \mathcal{N}(0, 1) \end{aligned} \quad (15)$$

for all  $t \in \mathbb{N}$ ,  $\alpha \in (0, 1)$  and  $\beta \in \mathbb{R}$ . The function  $m(t) : \mathbb{R} \rightarrow \mathbb{R}$  represents a *trend* independent of  $X^t$  and  $Y^t$ . We consider a linear trend  $m(t) = -ct + U_m^t$ , where  $U_m^t \sim \mathcal{N}(\mu_m, \sigma_m)$  and  $c \in \mathbb{R}^+$ . Given a realization  $x^{t-1}$ , the state of  $X^{t+\tau}$  admits the closed-form expression:

$$X^t = \alpha x^{t-1} + m(t) + U_X^t \quad (16)$$

$$X^{t+1} = \alpha^2 x^{t-1} + \alpha m(t) + \alpha U_X^t + m(t+1) + U_X^{t+1} \quad (17)$$

$$\vdots \quad (18)$$

$$X^{t+\tau} = \alpha^{\tau+1} x^{t-1} + \sum_{i=0}^{\tau} \alpha^{\tau-i} \left( m(t+i) + U_X^{t+i} \right) \quad (19)$$

Hence, the expectation of  $Y^{t+\tau}$  with respect to the interventional distribution  $P^{do(\theta)}(X^t, Y^t)$  is:

$$\begin{aligned} \mathbb{E}[Y^{t+\tau}] &= \mathbb{E}[\beta X^{t+\tau} + U_Y^{t+\tau}] \\ &= \beta \mathbb{E}[X^{t+\tau}] + \mathbb{E}[U_Y^{t+\tau}] \\ &\stackrel{(i)}{=} \beta \mathbb{E}[\alpha^{\tau+1} x^{t-1} + \sum_{i=0}^{\tau} \alpha^{\tau-i} (m(t+i) + U_X^{t+i})] \\ &\stackrel{(i)}{=} \beta \alpha^{\tau+1} x^{t-1} + \sum_{i=0}^{\tau} \alpha^{\tau-i} (\mathbb{E}[m(t+i)] + \mathbb{E}[U_X^{t+i}]) \\ &= \beta (\alpha^{\tau+1} x^{t-1} + \sum_{i=0}^{\tau} \alpha^{\tau-i} (-c(t+i) + \mu_m + \mu_X)) \end{aligned}$$

Here, (i) follows by construction since  $\mathbb{E}[U_Y^{t+\tau}] = 0$  for all  $t, \tau$  and  $i \in [d]$ . We now consider the following fixed classifier  $\sigma(Y^t | X^t)$  where  $\sigma(x) = 1/(1 + e^{-x})$ . Thus, the expectation over the classifier output becomes:

$$\mathbb{E}[h(X^{t+\tau})] = \sigma \left( \beta (\alpha^{\tau+1} x^{t-1} + \sum_{i=0}^{\tau} \alpha^{\tau-i} (-c(t+i) + \mu_m + \mu_X)) \right) \quad (20)$$

Given that we are considering soft interventions, we consider the cost function  $C(\hat{x}^{t+\tau}, x^{t+\tau}) = \hat{x}^{t+\tau} - x^{t+\tau} = \theta$  since  $\hat{x}^{t+\tau} = x^{t+\tau} + \theta$ . Given that  $\sigma(x) \geq 1/2$  if and only if  $x \geq 0$ , we have that the optimal intervention  $\theta^{t+\tau} \in \mathbb{R}$  for which we have  $\mathbb{E}[h(X^{t+\tau} + \theta)] \geq 1/2$  can be expressed as:

$$\theta^{t+\tau} = -\alpha^{\tau+1} x^{t-1} - \sum_{i=0}^{\tau} \alpha^{\tau-i} (-c(t+i) + \mu_m + \mu_X) \quad (21)$$

$\square$

## A.3 Proof of Proposition 3.9

We can prove Proposition 3.9 by showing how we can *always* find a simple trend invalidating *any* (robust) intervention.

PROOF. Let us consider a trend-stationary stochastic process  $P(X^t, Y^t)$  and fixed injective classifier  $h$  approximating  $P(X^t | Y^t)$ . We denote with  $\mathbf{m}(t) : \mathbb{N}^d \rightarrow \mathbb{R}^d$  the  $d$ -variate trend function where  $m_i(t)$  is the trend component for a single random variable  $X_i^t$  for any  $i \in [d]$  and  $t \in \mathbb{N}$ . Given a negatively classified instance

$\mathbf{x}^t$ , assume  $\theta$  is the optimal robust intervention for a fixed  $\epsilon > 0$  and for the timestep  $t$ .

Consider the following trend function  $\mathbf{m}(t) = \mathbf{1}\{t \geq \tau\}(-\theta)$  which is adding the inverse of the optimal intervention if an only if  $t \geq \tau$ . Specifically, we define each trend component as  $m_i(t) = \mathbf{1}\{t \geq \tau\}(-\theta_i)$ . If our stochastic process exhibits such a trend, then, for any fixed  $\tau > 0$ , the robust intervention is invalid e.g.,  $\mathbb{E}[h(\hat{\mathbf{x}}^{t+\tau})] < 1/2$ . Moreover, we can always build such a trend for any  $\theta$  and any unrestricted trend-stationary stochastic process.  $\square$

## A.4 Proof of Theorem 3.11

PROOF. We first apply the following substitutions (a)  $\beta' = \beta^{t+\tau}$  and  $\mathbf{x}' = \hat{\mathbf{x}}^{t+\tau}$  (b)  $\beta = \beta^t$  and  $\mathbf{x} = \hat{\mathbf{x}}^t$ , to improve the clarity of the proof. Further, consider  $\Delta h = \mathbb{E}[|h^{t+\tau}(\hat{\mathbf{x}}^{t+\tau}) - h^t(\hat{\mathbf{x}}^t)|]$ . Then, the proof is the following:

$$\begin{aligned} \Delta h &= \mathbb{E}[|\langle \beta', \mathbf{x}' \rangle - \langle \beta, \mathbf{x} \rangle|] \\ &= \mathbb{E}[|\langle \beta', \mathbf{x}' \rangle \\ &\quad + \langle \beta, \mathbf{x}' \rangle - \langle \beta, \mathbf{x}' \rangle - \langle \beta, \mathbf{x} \rangle + \langle \beta, \mathbf{x}' \rangle - \langle \beta, \mathbf{x}' \rangle|] \\ &= \mathbb{E}[|\langle \beta' - \beta, \mathbf{x}' \rangle + \langle \beta, \mathbf{x}' \rangle - \langle \beta, \mathbf{x} \rangle - \langle \beta, \mathbf{x}' \rangle|] \\ &\leq \mathbb{E}[|\langle \beta' - \beta, \mathbf{x}' \rangle|] + \mathbb{E}[|\langle \beta, \mathbf{x}' - \mathbf{x} \rangle|] \\ &\stackrel{(i)}{\leq} \mathbb{E}[\|\beta' - \beta\| \cdot \|\mathbf{x}'\|] + \mathbb{E}[\|\beta\| \cdot \|\mathbf{x}' - \mathbf{x}\|] \\ &\stackrel{(ii)}{\leq} k\sqrt{d} \cdot \mathbb{E}[\|\beta' - \beta\|] + k\sqrt{d} \cdot \mathbb{E}[\|\mathbf{x}' - \mathbf{x}\|] \\ &= k\sqrt{d} \cdot (\mathbb{E}[\|\beta^{t+\tau} - \beta^t\|] + \mathbb{E}[\|\hat{\mathbf{x}}^{t+\tau} - \hat{\mathbf{x}}^t\|]) \end{aligned}$$

where (i) follows from the Cauchy-Schwarz inequality and (ii) from the bounds we placed on  $\mathbf{X}$  and  $\beta$  (e.g., since  $-\mathbf{k} \leq \beta \leq \mathbf{k}$  and since  $|\beta| = d$  we have  $\max_{\beta} \|\beta\| = k\sqrt{d}$ ). Moreover, since  $h^t$  is trained over a fixed dataset  $\mathcal{D}^t$ , we have that  $\beta^t \perp X^t | \mathcal{D}^t$  for all  $t \in \mathbb{N}$ . We stress that the bounds placed on  $\mathbf{X}$  and  $\beta$  enable us to constrain the  $\Delta h(\theta; \tau)$  variation. In the unbounded case, where  $k \rightarrow \infty$ , clearly no upper bound is possible.  $\square$

## A.5 Proof of Corollary 3.12

We begin the Proof of Corollary 3.12 by starting from the previous proof for Theorem 3.11 (given in Appendix A.4). Please recall that a stochastic process  $P(X^t)$  is trend-stationary when it can be expressed as  $X^t = \mathbf{m}(t) + \mathbf{e}^t$ , where  $\mathbf{m}(t)$  is a (non-)linear trend function and  $\mathbf{e}^t$  is a stationary stochastic process. In the following, we denote with  $\tilde{\mathbf{x}}$  the stationary part of the stochastic process, and we consider *deterministic* trend functions.

PROOF. Let us assume that each random variable  $X_i^t$  can be described as a trend-stationary univariate stochastic process. Thus, let us define with  $\mathbf{m}(t) = \{m_i(t)\}_{i=1}^d$  the trend function, where  $m_i(t)$  the trend component for the  $i$ -th variable. Then, we define as  $m^* = m_j$  the trend with the largest difference  $j = \operatorname{argmax}_{i \in [d]} m_i(t+\tau) - m_i(t)$  for  $t \in \mathbb{N}$  and  $\tau > 0$ . Further, consider  $\Delta h = \mathbb{E}[|h^{t+\tau}(\hat{\mathbf{x}}^{t+\tau}) - h^t(\hat{\mathbf{x}}^t)|]$ . The derivation for the upper bound

is:

$$\begin{aligned}
\Delta h &\leq k\sqrt{d} \left( \mathbb{E} [\|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\|] + \mathbb{E} [\|\tilde{\mathbf{x}}^{t+\tau} - \tilde{\mathbf{x}}^t\|] \right) \quad (\text{Theorem 3.11}) \\
&\stackrel{(i)}{=} k\sqrt{d} \left( \mathbb{E} [\|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\|] + \mathbb{E} [\|\tilde{\mathbf{x}}^{t+\tau} + \mathbf{m}(t+\tau) - \tilde{\mathbf{x}}^t - \mathbf{m}(t)\|] \right) \\
&= k\sqrt{d} \left( \mathbb{E} [\|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\|] + \mathbb{E} [\|\tilde{\mathbf{x}}^{t+\tau} - \tilde{\mathbf{x}}^t\|] + \mathbb{E} [\|\mathbf{m}(t+\tau) - \mathbf{m}(t)\|] \right) \\
&\stackrel{(ii)}{=} k\sqrt{d} \left( \mathbb{E} [\|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\|] + \mathbb{E} [\|\mathbf{m}(t+\tau) - \mathbf{m}(t)\|] \right) \\
&= k\sqrt{d} \left( \mathbb{E} [\|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\|] + \sqrt{\sum_{i=1}^d (m_i(t+\tau) - m_i(t))^2} \right) \\
&\stackrel{(iii)}{\leq} k\sqrt{d} \left( \mathbb{E} [\|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\|] + \sqrt{\sum_{i=1}^d (m^*(t+\tau) - m^*(t))^2} \right) \\
&= k\sqrt{d} \left( \mathbb{E} [\|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\|] + \sqrt{d(m^*(t+\tau) - m^*(t))^2} \right) \\
&= k\sqrt{d} \left( \mathbb{E} [\|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\|] + \sqrt{d} (m^*(t+\tau) - m^*(t)) \right) \\
&= k\sqrt{d} \mathbb{E} [\|\boldsymbol{\beta}^{t+\tau} - \boldsymbol{\beta}^t\|] + kd (m^*(t+\tau) - m^*(t))
\end{aligned}$$

where (i) and (ii) follows from the definition of a *trend-stationary* stochastic process. Then, (iii) follows since we can substitute each univariate trend  $m_i(t)$  with the maximum  $m^*(t)$  for the time step  $t$  within the  $l_2$ -norm.  $\square$

## B Implementation Details for the Experiments With Synthetic and Real Data

In this section, we describe the synthetic and realistic stochastic processes we used in our experiments (cf. Section 4). We also describe the training steps and generative model used to approximate the SCMs. Lastly, we report several technical implementation details.

### B.1 Synthetic Additive Trend Function

We want our causal stochastic process to describe an environment where the changes to obtain a positive classification fluctuate over time. For example, as an individual ages, it will become less likely for her to repay her loan in the case of a loan application (based on the survival rate of the population). We define an additive trend function  $m(t) : \mathbb{N} \rightarrow \mathbb{R}^+$ , which is a linear combination between a *linear* and *seasonal* trend. We control the mixture of each component with two parameters,  $\beta_l \in \mathbb{R}^+$  and  $\beta_s \in \mathbb{R}^+$ , respectively. We also consider an additional parameter  $\alpha \in [0, 1]$  controlling the trend's strength over the stationary causal process.

$$m(t) = \alpha \cdot (\beta_l \cdot \min(0.05 \cdot t, 10) + \beta_s \cdot |\sin(0.5 \cdot t)|) \quad (22)$$

In our experiments, we set  $\beta_l \in \{0, 1\}$  and  $\beta_s \in \{0, 1.5\}$  for the linear ANM, and  $\beta_l \in \{0, 2\}$  and  $\beta_s \in \{0, 5\}$  for the non-linear ANM. For the realistic experiments, we set  $\beta_l, \beta_s \in \{0, 1\}$  for Adult,  $\beta_l \in \{0, 0.3\}$  and  $\beta_s \in \{0, 1\}$  for COMPAS, and  $\beta_l \in \{0, 0.5\}$  and  $\beta_s \in \{0, 5\}$  for Loan.

### B.2 Causal Graphs for the Experiments

For the synthetic experiments, we considered the synthetic 3-variables additive noise models (ANMs) from [44]. We extended them by transforming them into trend-stationary stochastic processes, by adding an autoregressive component and the trend  $m(t)$

on the 3rd feature. If  $\alpha = 0$ , both ANMs give rise to stationary time series.

#### Linear Additive Noise Model.

$$\begin{aligned}
X_1^t &= 0.5 \cdot X_1^{t-1} + U_1^t \\
X_2^t &= 0.5 \cdot X_2^{t-1} - 0.25 \cdot X_1^t + U_2^t \\
X_3^t &= 0.5 \cdot X_3^{t-1} + 0.05 \cdot X_1^t + 0.25 \cdot X_1^t - m(t) + U_3^t
\end{aligned} \quad (23)$$

where  $U_1^t \sim \text{MoG}(\mathcal{N}(-1, 1.5), \mathcal{N}(1, 1))$ ,  $U_2^t \sim \mathcal{N}(0, 0.1)$  and  $U_3^t \sim \mathcal{N}(0, 1)$ .

#### Non-linear Additive Noise Model.

$$\begin{aligned}
X_1^t &= 0.5 \cdot X_1^{t-1} + U_1^t \\
X_2^t &= 0.5 \cdot X_2^{t-1} - 1 \frac{3}{1 + e^{-2X_1^t}} + U_2^t \\
X_3^t &= 0.5 \cdot X_3^{t-1} + 0.05 \cdot X_1^t + 0.25 \cdot (X_1^t)^2 - m(t) + U_3^t
\end{aligned} \quad (24)$$

where  $U_1^t \sim \text{MoG}(\mathcal{N}(-2, 1.5), \mathcal{N}(1, 1))$ ,  $U_2^t \sim \mathcal{N}(0, 0.1)$  and  $U_3^t \sim \mathcal{N}(0, 1)$ .

**Label function.** We also consider the following conditional distribution  $P(Y^t | \mathbf{X}^t)$ , again taken from [44], which produces roughly balanced groups:

$$Y^t \sim \text{Binomial} \left( 1 / (1 + \exp(-2.5 \cdot (X_1^t + X_2^t + X_3^t) / \rho)) \right) \quad (25)$$

where  $\rho$  is the empirical mean of  $X_1^t + X_2^t + X_3^t$  for  $t = 0$ . We use the label function only to train the classifier  $h$  at time  $t = 0$ , then it is discarded and we rely only on  $h$  for our experiments.

### B.3 Causal Graphs for the Realistic Experiments

We now describe the design choices for the realistic datasets. We tried to find a balance between realism and simplicity, to provide scenarios close to potential real-world situations, that, however, we can easily control for our experiments. Therefore, some of the design choices might not represent faithfully how the system can evolve in real life.

**Adult [21].** We use the features and causal graph defined by [54]. Eq. (26) shows the full causal graph. We have the following features:  $S$  (sex),  $A$  (age),  $US$  (resident of the United States of America),  $M$  (married),  $E$  (education level) and  $H$  (working hours per week). The features  $S$ ,  $US$ , and  $M$  are categorical variables, the others are represented as continuous random variables. We assume  $S$ ,  $A$ ,  $US$  and  $M$  remain fixed over time. The only actionable features are the education level  $E$ , and working hours per week  $H$ . We apply a decreasing trend to  $H$ . We employ non-linear structural equations  $f_M$ ,  $f_E$  and  $f_H$  by using pre-trained 3-layer MLPs, trained on the original dataset, from [19]. Moreover, as a label function, we also employ a classifier  $h$  taken by [19] to label the examples at  $t = 0$ .

$$\begin{aligned}
S^t &= \mathbf{1}\{t > 0\} \cdot S^{t-1} + \mathbf{1}\{t = 0\} \cdot U_S \\
A^t &= \mathbf{1}\{t > 0\} \cdot U^{t-1} + \mathbf{1}\{t = 0\} \cdot U_A^t \\
US^t &= \mathbf{1}\{t > 0\} \cdot US^{t-1} + \mathbf{1}\{t = 0\} \cdot U_{US} \\
M^t &= \mathbf{1}\{\sigma(f_M(S^t, A^t, US^t)) > 1/2\} \\
E^t &= 0.5 \cdot E^{t-1} + f_E(S^t, A^t, US^t, M^t) + U_E \\
H^t &= 0.5 \cdot H^{t-1} + f_H(S^t, A^t, US^t, M^t) - m(t) + U_H
\end{aligned} \quad (26)$$

where the exogenous factors are  $U_S \sim \text{Binomial}(0.9)$ ,  $U_A^t \sim \mathcal{N}(0, 1)$ ,  $U_{US} \sim \text{Binomial}(0.9)$ ,  $U_E \sim \mathcal{N}(0, 1)$  and  $U_H \sim \mathcal{N}(0, 1)$ .

**COMPAS [3].** We use the features and causal graph defined by [54]. Eq. (27) shows the full causal graph. We have the following features:  $S$  (sex),  $A$  (age),  $C$  (ethnicity, if caucasian or not), and  $P$  (prior counts). The feature  $S$  is categorical, and the others are represented as continuous random variables. We assume  $S$ ,  $A$ , and  $C$  remain fixed over time, and that the only actionable feature is the prior count  $P$ . We apply an increasing trend to  $P$ . As we did for Adult, we obtain the pre-trained non-linear structural equations  $f_C$  and  $f_P$  and label function from [19].

$$\begin{aligned} S^t &= \mathbf{1}\{t > 0\} \cdot S^{t-1} + \mathbf{1}\{t = 0\} \cdot U_S \\ A^t &= \mathbf{1}\{t > 0\} \cdot U^{t-1} + \mathbf{1}\{t = 0\} \cdot U_A^t \\ C^t &= \mathbf{1}\{t > 0\} \cdot C^{t-1} + \mathbf{1}\{t = 0\} \cdot (\sigma(f_C(S^t, A^t)) + U_C^t) \\ P^t &= 0.5 \cdot P^{t-1} + f_P(S^t, A^t, C^t) + m(t) + U_P^t \end{aligned} \quad (27)$$

where the exogenous factors are  $U_S \sim \text{Binomial}(0.8)$ ,  $U_A^t \sim \text{Poisson}(1)$ ,  $U_C^t \sim \mathcal{N}(0, 1)$ , and  $U_P^t \sim \mathcal{N}(0, 1)$

**Loan [44].** We use the causal graph defined by [44] presenting a semi-synthetic loan approval scenario inspired by the German Credit dataset [35]. For the structural equations, we use the one adapted by [19] in their paper. Eq. (28) shows the full causal graph. We have the following features:  $G$  (gender),  $A$  (age),  $E$  (education level),  $L$  (loan amount),  $D$  (loan duration),  $I$  (income) and  $S$  (savings).  $G$  is a categorical variable, while the rest are considered continuous. Moreover,  $G$  remains fixed over time. We assume the only actionable features are  $S$  and  $I$ . We apply a trend to the income  $I$ .

$$\begin{aligned} G^t &= \mathbf{1}\{t > 0\} \cdot G^{t-1} + \mathbf{1}\{t = 0\} \cdot U_G^t \\ A^t &= 0.5 \cdot A^{t-1} - 35 + U_A^t \\ E^t &= 0.5E^{t-1} - 0.5 + \left(1 + e^{-\left(-1+0.5G^t + (1+e^{-0.1A^t})^{-1} + U_E\right)}\right)^{-1} \\ L^t &= 0.5 \cdot L^{t-1} + 1 + 0.01(A^t - 5)(5 - A^t) + G^t + U_L \\ D^t &= 0.5 \cdot D^{t-1} - 1 + 0.1A^t + 2G^t + L^t + U_D \\ I^t &= 0.5 \cdot I^{t-1} - 4 + 0.1(A^t + 35) + 2G^t + G^t E^t + U_I - m(t) \\ S^t &= 0.5 \cdot S^{t-1} - 4 + 1.5 \cdot \mathbf{1}\{I^t > 0\} \cdot I^t + U_S \end{aligned} \quad (28)$$

where the exogenous factors are  $U_G^t \sim \text{Binomial}(0.5)$ ,  $U_A^t \sim \text{Gamma}(10, 3.5)$ ,  $U_E \sim \mathcal{N}(0, \sqrt{0.25})$ ,  $U_L \sim \mathcal{N}(0, 2)$ ,  $U_D \sim \mathcal{N}(0, 3)$ ,  $U_I \sim \mathcal{N}(0, 2)$ , and  $U_S \sim \mathcal{N}(0, 5)$ . We sample the labels from the function defined by [44]:

$$Y^t \sim \text{Bernoulli}\left(\left(1 + e^{-0.3(-L^t - D^t + I^t + S^t + IS^t)}\right)^{-1}\right) \quad (29)$$

## B.4 On Learning an Approximate SCM

We now describe the simple generative model we used in the experiment in Section 4.2. Similarly to [44] and [19], we *approximate* the structural equations in a data-driven manner. We assume we can represent the actionable features  $X_i$  as Gaussian random variables  $\mathcal{N}(\mu_i^t, 1)$  with constant variance and time-dependent  $\mu_i^t$ . For each random variable, we define the mean as the output of a regressor

$f_i$  taking as input: the autoregressive component  $X_i^{t-1}$ , the parents  $\text{Pa}_i^t$  and the time  $t$ . Thus, we obtain the following structural equations:

$$X_i^t = \mu_i^t + U_i^t \quad U_i^t \sim \mathcal{N}(0, 1) \quad \mu_i^t = f_i(X_i^{t-1}, \text{Pa}_i^t, t) \quad (30)$$

Similarly to a conditional VAE [71], we can both sample new instances from the approximate SCMs, but we can also compute the interventional or counterfactual distributions.

## B.5 Technical Details and Code

We based our implementation of CAR, SAR, IMF and T-SAR following *adversarial robust algorithmic recourse* [19]. Namely, we leveraged their implementation<sup>9</sup> and we adapted their code to work with time-based uncertainty sets (cf. Section 3.6). Moreover, we also used their causal graph implementations, pre-trained models, and preprocessing steps as a starting point for building our stochastic processes. In the case of the synthetic datasets, we instead looked at [44] original implementation<sup>10</sup>. Our code and experimental results will be released on GitHub with a permissive license<sup>11</sup>. Lastly, we run our experiments on a Linux machine (Ubuntu 22.04, 4 LTS) with 32 CPU cores and 125 GB of RAM. Our implementation is written in Python 3.10, using standard scientific and deep learning libraries such as numpy [33] and PyTorch [56]. The various hyperparameters are duly specified in the source code, but we report the most important here for clarity.

**Algorithm 1 hyperparameters.** We approximate  $B(x^t; \tau)$  by sampling only 20 instances for all datasets and we set the number of *epochs* to  $N = 30$ . As a penalty, we set  $\lambda = 1$  for the Lagrangian (line 5, Algorithm 1). Please notice that [19] uses a *decaying rate* to reduce the impact of the cost function on the loss  $\mathcal{L}$  after each epoch (we kept the original hardcoded value of 0.02). As learning rate, we set  $\eta = 0.5$  for the synthetic experiments, and  $\eta = 3$  for the realistic datasets. The learning rate is the same for all the methods. We did not perform a full grid search over the parameter space, since we found empirically our chosen hyperparameters were giving satisfactory performances.

**Classifiers  $h$ .** For each setting, we trained a 3-layered MLP approximating  $P(Y^t \mid x^t)$ , via *empirical risk minimization* by sampling a given dataset for  $t = 0$ . We use stochastic gradient descent (SGD) to minimize the binary cross entropy loss (e.g., torch.nn.BCELoss<sup>12</sup>)

$$\mathcal{L} = -\frac{1}{N} \sum_{x^0, y^0 \in \text{batch}} (y^0 \log h(x^0) + (1 - y^0) \log (1 - h(x^0))) \quad (31)$$

where  $y^0$  is the ground truth label. In our experiments, we set the batch size to 100, the number of epochs to 15 and the learning rate to 0.001, for all datasets. The accuracy of the models for a single seed are: 0.847 (Linear ANM), 0.963 (Non-Linear ANM), 0.817 (Adult), 0.645 (COMPAS) and 0.842 (Loan).

**Approximate structural equations.** In our experiments, we consider only linear  $f_i$ . We train each  $f_i$  via *empirical risk minimization* following the procedure outlined in Section 4.2. For each feature  $i \in [d]$ , and for each epoch, we consider a batch  $\{(x_i^t, x_i^{t-1}, \text{Pa}_i^t, t)_j\}$

<sup>9</sup><https://github.com/RicardoDominguez/AdversariallyRobustRecourse>

<sup>10</sup><https://github.com/charmlab/recourse>

<sup>11</sup><https://github.com/unitn-sml/temporal-algorithmic-recourse>

<sup>12</sup><https://pytorch.org/docs/stable/generated/torch.nn.BCELoss.html>

$\}_{j=1}^b$  and we minimize the *mean squared error* between the ground truth  $x_i^t$  and the model output  $f_i(X_i^{t-1}, \mathbf{Pa}_i^t, t)$  with stochastic gradient descent. We fix the batch size to 100, the learning rate to 0.001 and the number of epochs to 15 for all settings. We report here the *mean squared error* over 50 timesteps (2000 individuals) for the approximate SCMs we used in Section 4.2. We compute the MSE for each feature for each timestep, and then we average. The empirical average MSE and standard deviation over 10 runs is:  $1.162 \pm 0.005$  (Adult),  $258.226 \pm 7.328$  (COMPAS) and  $10.447 \pm 0.037$  (Loan).

### C Additional Empirical Results on the Effect of Uncertainty on Counterfactual Recourse Over Time

We present the extended results of the experiment measuring empirically the impact of the uncertainty in CAR (Section 3.2). The experimental setting and evaluation procedure are the same as Section 4.1. Fig. 6 shows the empirical average validity of CAR’s recourse over time  $t \in \{0, 100\}$ . The plot shows how uncertainty heavily impacts the recourse validity from the initial time steps as  $\sigma_U$  grows, even when  $P(X^t)$  is stationary.

### D Analysis of the Trade-off Between Validity and Cost

We present an analysis of the trade-off between the validity and cost of recourse generated by Algorithm 1. Using the same experimental setup and evaluation procedure detailed in Sections 4.1 and 4.2, we evaluate the impact of varying the  $\lambda$  parameter, which governs the strength of the cost constraint (line 5, Algorithm 1). For this analysis, we focus on time series with a complex linear and seasonal trend. The results, shown in Fig. 7, demonstrate a clear cost-validity trade-off across all experimental scenarios: **cheaper interventions result in lower validity over time**. As the weight of the cost constraint ( $\lambda$ ) increases during optimization, the interventions become more affordable but exhibit reduced validity. These findings align with prior research on cost-validity trade-offs in non-temporal settings [60], extending the observations to temporal scenarios. For example, in Adult, we observe a reduction in the validity over time ( $\sim 0.05$ ), but a decreased cost as shown by the lighter dots. In COMPAS, the trade-off presents a smoother behaviour since we have only one actionable feature. Lastly, we observe how the reduction in validity is not consistent across time series: we suspect this might depend on the quality of the estimator  $\hat{P}(X^t)$  and on the decision boundary of the classifier  $h$ .

### E Further Analysis on the Empirical Cost and Sparsity

In this section, we report further analysis and results when considering the empirical average *cost* and *sparsity* (# of  $I$  achieving recourse) of the valid interventions. They are both common metrics in the algorithmic recourse literature [42, 77].

**Sparsity.** Figs. 8 and 9 shows the empirical average sparsity of the causal recourse methods for both synthetic and realistic stochastic processes. We do not report values for IMF since it always acts on *all* the actionable features ( $|I| = 3$ ). In Fig. 8, T-SAR presents

a similar or lower sparsity than other methods. However, as Section 4.1 shows, T-SAR is the only method achieving good validity over time. Thus, these results suggest that incorporating time might not increase the sparsity of the solutions. In the case of approximate SCMs, Fig. 9 shows how T-SAR provides sparse interventions for Adult, but increasingly larger interventions for Loan. COMPAS has only one actionable feature, thus the sparsity is equal for all approaches. As highlighted in Section 4.2, T-SAR performance is also dependent on the quality of the estimator  $\hat{P}(X^t)$ .

**Cost.** Figs. 10 and 11 shows the empirical average cost for the users for which all (non-)causal recourse methods found a valid intervention. We consider only the top-3 methods achieving recourse for each time step  $t$ . In realistic and synthetic settings, T-SAR can provide *cost-adaptive* interventions which follow the underlying trend. For example, we can observe this phenomenon in both COMPAS and Loan (Fig. 11). It is also visible for  $m(t) \in \{\text{Linear}, \text{Linear+Seasonal}\}$  in the Non-linear and Linear ANMs, respectively. We also notice how T-SAR seems to provide costlier recourses than the standard robust methods. However, in Adult, T-SAR produces cheaper interventions than the other approaches. We can explain this behaviour for Adult by looking at the analysis of the successful intervention sets  $I$  in Section 4.2.

In conclusion, by incorporating an estimator of the stochastic process, we can provide *sparse* interventions more resilient to time. These interventions tend to be costlier and the cost varies with the time lag  $\tau$  when they will be applied. However, robust (non-)causal methods achieve dissimilar validity, thus making them not fully comparable to each other by measuring their cost. Nevertheless, we believe the analysis has merit since it hints at a tradeoff between sparsity, cost and validity.

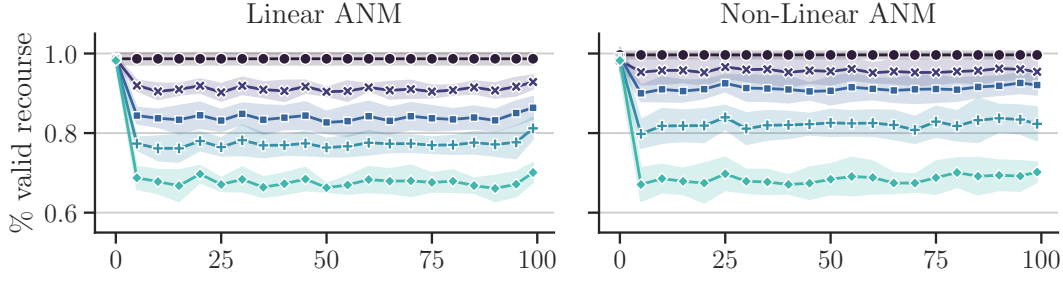
### F Further Experiments With a Perfect Estimator

We replicated the experiments in Section 4.2 by using instead the perfect estimator  $\hat{P}(X^t) = P(X^t)$  of the stochastic process for each dataset. Fig. 12 shows how T-SAR offers superior performances in terms of validity than robust (non-)causal algorithmic recourse methods. In COMPAS and Loan, T-SAR achieves now perfect validity over all time steps. These results highlight the importance of relying on a good estimator of the stochastic process and, as outlined in Section 6, we argue it is also a mandatory requirement for realistic applications of the proposed method.

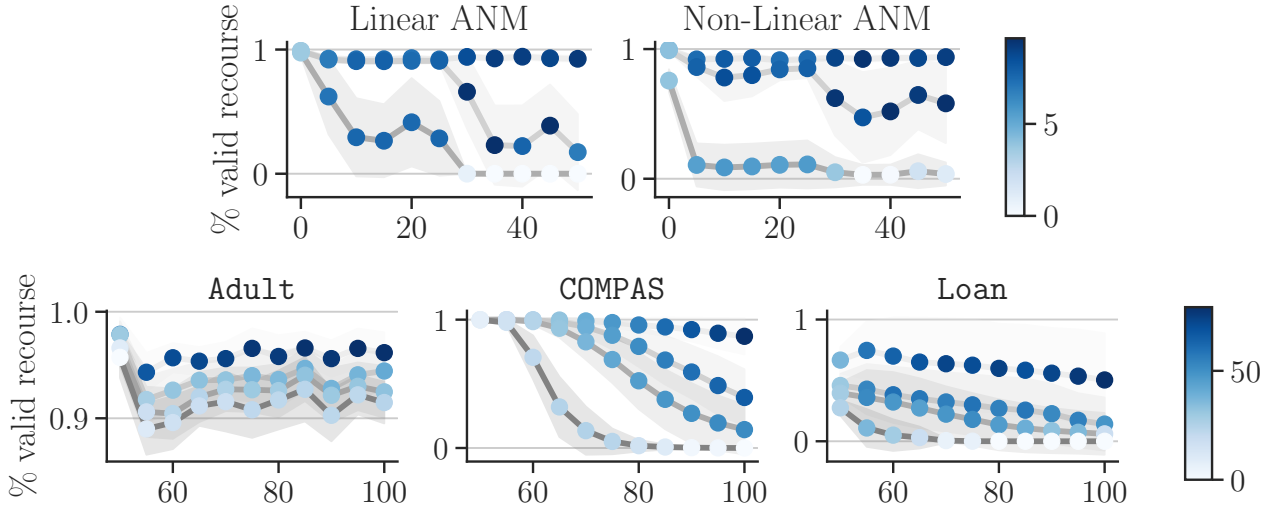
### G Additional Theoretical Results on Cost Stability

In this section, we provide an upper bound on the cost stability of recourse suggestions under a more expressive cost function.

Previous research has shown how providing recourse without considering the user’s preferences can lead to sub-optimal interventions [17]. This is why *personalized* AR, in line with multi-attribute decision making [45, 64], models the cost function as an *additive independence model*  $C(\hat{\mathbf{x}}, \mathbf{x}) = \mathbf{w}^\top |\hat{\mathbf{x}} - \mathbf{x}|$ , where the weights  $\mathbf{w} \in \mathbb{R}^d$  encapsulate the user’s preferences [16, 17]. We assume we can learn these weights, either from historical data, *e.g.*, surveys and interviews [66], or by interacting with the end-user [17]. In the following, we explicitly consider the evolution of the user’s preferences  $\mathbf{W}$ , as



**Figure 6: Effect of uncertainty on counterfactual AR over time. Empirical average validity and standard deviation over 10 runs of robust counterfactual algorithmic recourse (CAR) for  $t \in \{0, 100\}$ . We vary the variance  $\sigma_U$  of the exogenous factors of the stochastic process. Legend ( $\sigma_U$ ): ■ 0 ■ 0.3 ■ 0.5 ■ 0.7 ■ 1.0.**



**Figure 7: Trade-off between cost and validity for realistic and synthetic datasets. We report the empirical average validity for Algorithm 1 under the linear+seasonal trend ( $\alpha = 1.0$ ) when varying the  $\lambda$  for (Top) synthetic and (Bottom) realistic time series. We consider a non-linear classifier  $h$  (3-layer neural network) for each setting. Each dot represents the empirical average cost of the interventions achieving recourse. A darker dot implies a larger cost. We represent the validity for each  $\lambda$  as a grey line and the standard deviation over 10 runs as a shaded area. Legend ( $\lambda$ ): ■ 1 ■ 3 ■ 5 (Synthetic time series), ■ 0.01 ■ 0.05 ■ (Realistic time series).**

it also impacts the effectiveness of recourse, although doing so can be avoided for non-personalized AR approaches.

We assume the user preferences can be represented as a stochastic process  $P(W^t)$ . We do not put any prior assumption on how  $P(W^t)$  factorizes, since it is not relevant for our results. In line with previous work [17], we could imagine  $P(W^t)$  follows a causal model. Then, we can provide the following upper bound:

**THEOREM G.1.** Consider the discrete-time stochastic processes  $P(X^t, Y^t)$ ,  $P(W^t)$  and a parametrized cost function  $C(\hat{x}, x; w) = \langle \hat{x} - x, w \rangle$  with bounded  $-k \leq w_i^t, X_i^t \leq k$  for  $k \in \mathbb{R}^+$ . Given a realization  $x^t$  and user's preferences  $w^t$ , the variation of the cost of

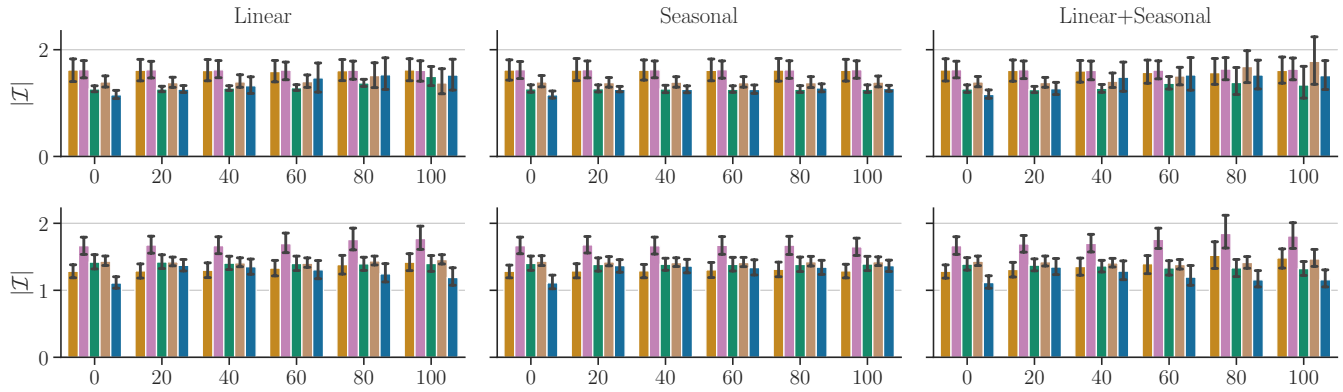
an intervention  $\theta$  is upper bounded by:

$$\begin{aligned} & \mathbb{E} [|C(\hat{x}^{t+\tau}, x^{t+\tau}; w^{t+\tau}) - C(\hat{x}^t, x^t; w^t)|] \\ & \leq k\sqrt{d} \cdot \mathbb{E} [\|w^{t+\tau} - w^t\| + \|\hat{x}^{t+\tau} - x^{t+\tau}\| - \|\hat{x}^t - x^t\|] \end{aligned}$$

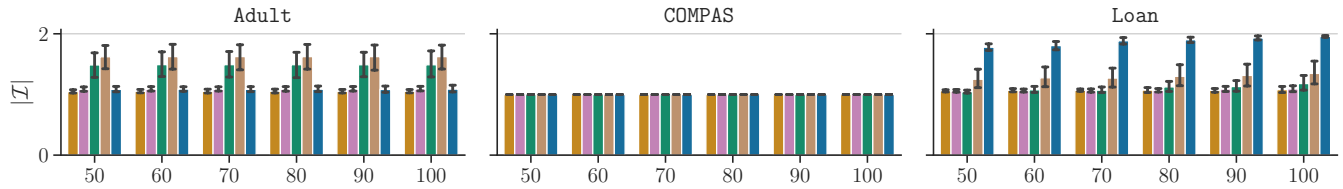
where  $\hat{x}^t \sim \text{pdo}(\theta)(X^t \mid X_{nd(I)}^t = x_{nd(I)}^t)$  and  $\hat{x}^{t+\tau} \sim \text{pdo}(\theta)(X^{t+\tau} \mid X_{nd(I)}^{t+\tau} = x_{nd(I)}^{t+\tau})$ .

Theorem G.1 shows how the recourse cost changes based on how the users' preferences evolve, and also over the relative difference of the proposed changes given the starting value.

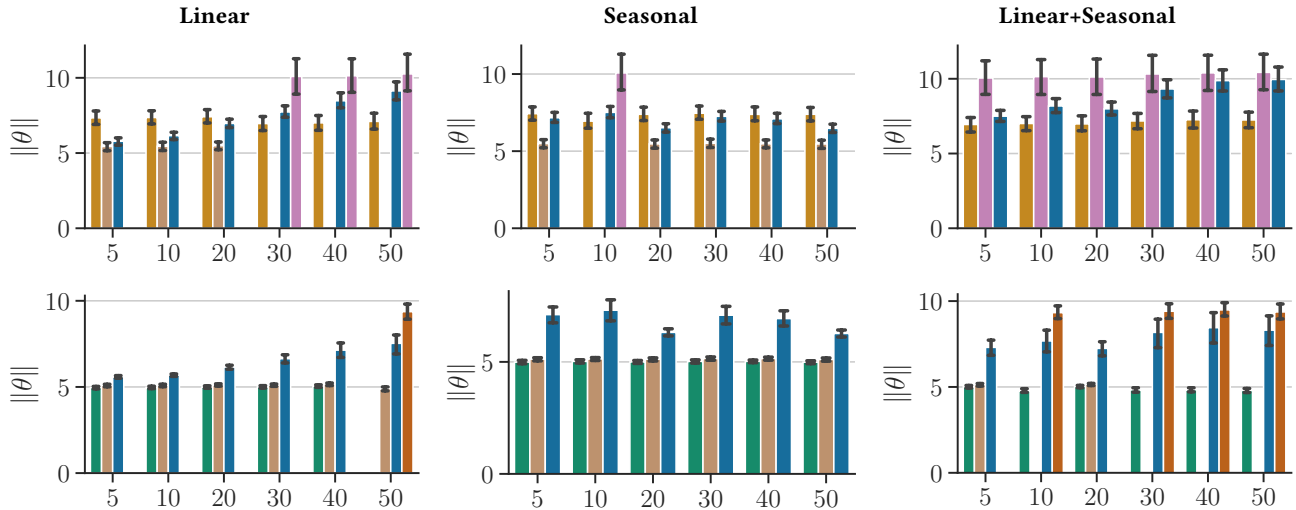
**PROOF.** We first apply the following substitutions (a)  $w' = w^{t+\tau}$  and  $\hat{x}' = \hat{x}^{t+\tau}$  (b)  $x' = x^{t+\tau}$   $\hat{x}'' = \hat{x}^t$  (c)  $w = w^t$  and  $x'' = x^t$ , to improve the clarity of the proof. Further, consider



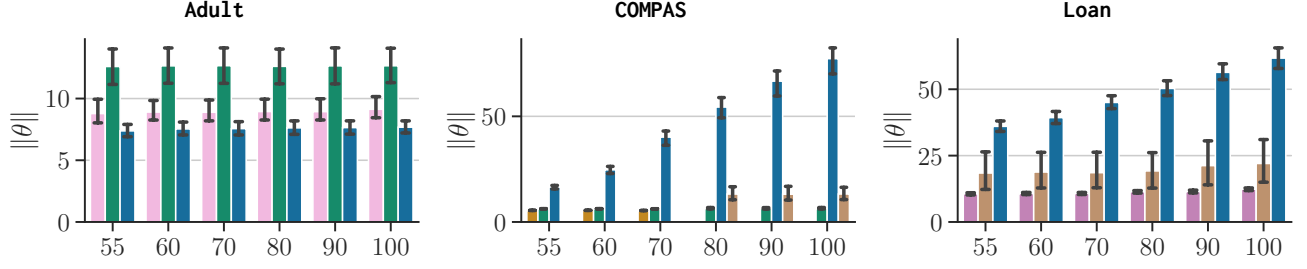
**Figure 8: Empirical average sparsity and standard deviation of interventions achieving recourse for all causal recourse methods in the Linear (top) and Non-Linear (bottom) ANMs. We report the results for all the available trends  $m(t) \in \{\text{Linear, Seasonal, Linear+Seasonal}\}$  and for some time steps  $t$ . Legend: T-SAR CAR ( $\epsilon = 3$ ) SAR ( $\epsilon = 3$ ) and CAR ( $\epsilon = 5$ ) SAR ( $\epsilon = 5$ ).**



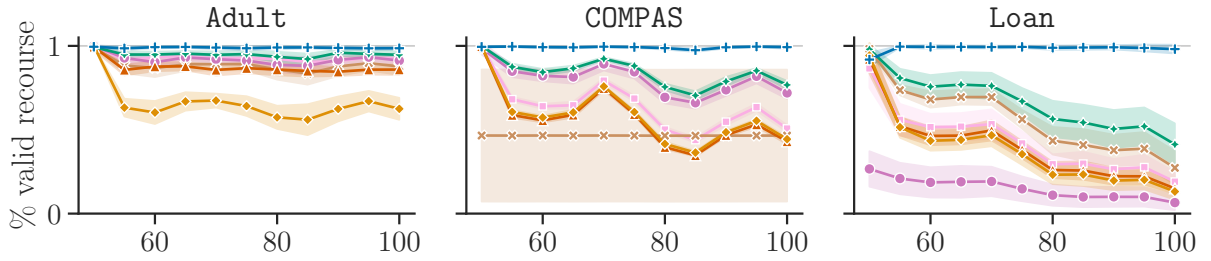
**Figure 9: Empirical average sparsity and standard deviation of interventions achieving recourse for all causal recourse methods in the realistic datasets. We report the results for  $m(t) = \text{Linear+Seasonal}$  and for some time steps  $t$ . Legend: T-SAR CAR ( $\epsilon = 3$ ) SAR ( $\epsilon = 3$ ) and CAR ( $\epsilon = 5$ ) SAR ( $\epsilon = 5$ ).**



**Figure 10: Empirical average cost and standard deviation for the top-3 methods achieving recourse in the Linear (top) and Non-Linear (bottom) ANMs. We report the results for all the available trends  $m(t) \in \{\text{Linear, Seasonal, Linear+Seasonal}\}$ . Legend: T-SAR CAR ( $\epsilon = 3$ ) SAR ( $\epsilon = 3$ ) IMF ( $\epsilon = 3$ ) and CAR ( $\epsilon = 5$ ) SAR ( $\epsilon = 5$ ) IMF ( $\epsilon = 5$ ).**



**Figure 11: Empirical average cost and standard deviation for the top-3 methods achieving recourse in the realistic datasets under a non-linear trend. Legend:** ■ T-SAR ■ SAR ( $\epsilon = 0.05$ ) and ■ CAR ( $\epsilon = 0.5$ ) ■ SAR ( $\epsilon = 0.5$ ) ■ IMF ( $\epsilon = 0.5$ ).



**Figure 12: Effect of time on realistic datasets. Empirical average validity and standard deviation (10 runs) for the robust ( $\epsilon \in \{0.05, 0.5\}$ ) and time-aware causal recourse methods for the realistic datasets under a non-linear trend. Legend:** ■ CAR ( $\epsilon = 0.05$ ) ■ SAR ( $\epsilon = 0.05$ ) ■ IMF ( $\epsilon = 0.05$ ) and ■ CAR ( $\epsilon = 0.5$ ) ■ SAR ( $\epsilon = 0.5$ ) ■ IMF ( $\epsilon = 0.5$ ).

$\Delta C = \mathbb{E} [|C(\mathbf{x}, \mathbf{x}'; \mathbf{w}') - C(\mathbf{x}, \mathbf{x}; \mathbf{w})|]$ . Then, the proof is the following:

$$\begin{aligned}
 \Delta C &= \mathbb{E} [|\langle \mathbf{x}' - \mathbf{x}, \mathbf{w}' \rangle - \langle \mathbf{x} - \mathbf{x}, \mathbf{w} \rangle|] \\
 &= \mathbb{E} [|\langle \hat{\mathbf{x}}' - \mathbf{x}' - |\hat{\mathbf{x}}'' - \mathbf{x}''|, \mathbf{w}' \rangle + \langle \mathbf{w}' - \mathbf{w}, |\hat{\mathbf{x}}'' - \mathbf{x}''| \rangle|] \\
 &\stackrel{(i)}{\leq} \mathbb{E} [|\|\hat{\mathbf{x}}' - \mathbf{x}' - |\hat{\mathbf{x}}'' - \mathbf{x}''|\|| \cdot \|\mathbf{w}'\|] \\
 &\quad + \mathbb{E} [|\|\mathbf{w}' - \mathbf{w}\| \cdot \|\hat{\mathbf{x}}'' - \mathbf{x}''\||] \\
 &\stackrel{(ii)}{\leq} k\sqrt{d} \cdot \mathbb{E} [|\|\hat{\mathbf{x}}' - \mathbf{x}' - |\hat{\mathbf{x}}'' - \mathbf{x}''|\||] \\
 &\quad + k\sqrt{d} \cdot \mathbb{E} [|\|\mathbf{w}' - \mathbf{w}\||] \\
 &= k\sqrt{d} \cdot \mathbb{E} [|\|\hat{\mathbf{x}}' - \mathbf{x}' - |\hat{\mathbf{x}}'' - \mathbf{x}''|\|| + \|\mathbf{w}' - \mathbf{w}\|] \\
 &= k\sqrt{d} \cdot \mathbb{E} [|\|\mathbf{w}^{t+\tau} - \mathbf{w}^t\| + |\|\hat{\mathbf{x}}^{t+\tau} - \mathbf{x}^{t+\tau}| - |\hat{\mathbf{x}}^t - \mathbf{x}^t|||]
 \end{aligned}$$

where (i) follows from the Cauchy-Schwarz inequality, and (ii) from the bounds we placed on  $\mathbf{X}$  and  $\mathbf{W}$ . On the last step, we reorder the terms and we substitute the temporary variables with the original values.  $\square$