Counterfactual Analysis of the Impact of the IMF Program on Child Poverty in the Global-South Region using Causal-Graphical Normalizing Flows

Sourabh Balgi 1 Jose M. Peña 1 Adel Daoud 2

1 Department of Computer and Information Science (IDA), Linköping University, Linköping, Sweden
2 Institute for Analytical Sociology (IAS), Linköping University, Linköping, Sweden
sourabh.balgi@liu.se 1 jose.m.pena@liu.se 1 adel.daoud@liu.se 2

Abstract
This work demonstrates the application of a particular branch of causal inference and deep learning models: causal-Graphical Normalizing Flows (c-GNFs). In a recent contribution, scholars showed that normalizing flows carry certain properties, making them particularly suitable for causal and counterfactual analysis. However, c-GNFs have only been tested in a simulated data setting and no contribution to date have evaluated the application of c-GNFs on large-scale real-world data. Focusing on the AI for social good, our study provides a counterfactual analysis of the impact of the International Monetary Fund (IMF) program on child poverty using c-GNFs. The analysis relies on a large-scale real-world observational data: 1,941,734 children under the age of 18, cared for by 567,344 families residing in the 67 countries from the Global-South. While the primary objective of the IMF is to support governments in achieving economic stability, our results find that an IMF program reduces child poverty as a positive side-effect by about 1.2 ± 0.24 degree (‘0’ equals no poverty and ‘7’ is maximum poverty). Thus, our article shows how c-GNFs further the use of deep learning and causal inference in AI for social good. It shows how learning algorithms can be used for addressing the untapped potential for a significant social impact through counterfactual inference at population level (ACE), sub-population level (CACE), and individual level (ICE). In contrast to most works that model ACE or CACE but not ICE, c-GNFs enable personalization using ‘The First Law of Causal Inference’.

1. Introduction
As poverty hinders individuals to flourish, many studies have deepened our understanding of the causes and consequences that push vulnerable groups such as children into poverty (Banerjee et al., 2011). However, there is lacking knowledge about how to optimally tailor public policies to alleviate poverty for such vulnerable groups in times of macroeconomic volatility (Halleröd et al., 2013; Kino et al., 2021). A key question is then, to what extent can policymakers tailor policies to each child’s circumstances and to what extent is it viable to articulate one policy for a population. One of the most powerful international organization today is the International Monetary Fund (IMF), as it is responsible to support macroeconomic stability in the global system. The IMF is part of the United Nations, and hence the mandate of almost all countries in the world. Yet the effects of its programs on children are disputed. Thus, more research is required not only to determine the average affect of IMF on child poverty, but also to develop methods to personalize public policies to children’s circumstances (Daoud et al., 2017; Daoud & Reinsberg, 2018; Daoud et al., 2019; Åkerström et al., 2019).

In this article, we analyze the impact of the IMF program on child poverty at an individual child level by considering the microeconomic factors such as family conditions along with the macroeconomic factors impacted by the austerity of the IMF program. This idea of individualization or personalization follows from a policy vision of tailoring interventions to an individual’s needs and context. For example, Banerjee et al. (2011) states,

“We have to abandon the habit of reducing the poor to cartoon characters and take the time to really understand their lives, in all their complexity and richness.”

In other words, instead of the conventional Average Causal Effect (ACE) policymaking that ascribes the same treatment to an entire population, researchers need to develop more sensitive frameworks. While articulating policies based on conditional ACE (CACE) is an improvement over ACE policymaking, CACE is still insufficient as it analyzes groups of individuals and hence ignores the individual complexity within-groups (Künzel et al., 2019). Thus, CACE still sug-
gests a sub-optimal ‘one-size-fits-all’ policy, and therefore, it is critical to move to the finest level of granularity and perform an Individual Causal Effect (ICE) analysis. Such personalized analysis relies on the tools of causal inference and, more specifically, counterfactual inference (Pearl, 2009a; b). Currently, to combat poverty, government officials use mainly an individual’s income and similar characteristics to identify vulnerable population and then assess if they are eligible for social welfare programs. Then, a simple rule is often applied: if an individual is eligible, that person will likely receive a fixed ‘one-size-fits-all’ public policy; if the individual is ineligible, no policy is ascribed. Although such public-policy making is highly transparent as it applies what is best on average for a population (i.e., ACE) or group (i.e., CACE) it lacks an adaptability to that individual’s needs that is necessary to efficiently combat poverty, ill-health, and other social ills (Daoud, 2018). For effective combating, government officials and others require methods that are able to personalize these policies (Kino et al., 2021). That is, personalized public policy analysis (P³A) requires methods that can move beyond ACE estimation, and into counterfactual inference.

Although there exists a range of ACE and CACE estimation methods powered by machine-learning algorithms (Künzel et al., 2019), there is a lack of similar ICE methods. If the statistical (functional) form among the outcome, treatment, and covariates are known, one could in principle estimate ICE parametrically (Pearl, 2009a). However, in observational studies, functional forms are rarely known due to the complexity of reality and, thus, causal estimation would benefit if machine-learning methods can learn these functions in a data-driven way. To learn these functional forms, recently Balgi et al. (2022) developed an ICE estimation method, called causal-Graphical Normalizing Flows (c-GNFs). This method is a flow-based deep-learning model, that encapsulates the underlying SCM. In other words, c-GNFs enable a scholar to stipulate the causal connections among variables in an SCM, but without making any functional form assumption. The c-GNF is able to approximate these functional forms efficiently in a data-driven manner. Most importantly, because c-GNFs are invertible (they can propel forward and backward through a SCM) and thus calculate noise variables for each individual, it has been shown that c-GNFs enable ICE estimation using ‘The First Law of Causal Inference’ (Pearl, 2009a).

The existing development of c-GNFs relies purely on statistical proofs and simulation studies. In Balgi et al. (2022), scholars developed c-GNFs and benchmarked them against well established methods such as Inverse Probability Weighting (Rosenbaum & Rubin, 1983; Hernán & Robins, 2009), Regression-With-Residuals (Wodtke, 2020), and Meta-Learners (Künzel et al., 2019) under controlled simulated experimental settings where the true causal effect is known. Thus, a c-GNFs real-world applicability for P³A remains uncertain.

In this article we evaluate c-GNFs in a large-scale real-world observational study. Based on substantive assumptions encoded in an SCM, we encapsulate the causal mechanism identifying the impact of the IMF program on child poverty. The data used are observational and thus carries the usual caveats for observational studies. Our study shows how c-GNFs are likely to provide insightful conclusions that are hidden in such large scale data and that are significant for P³A specifically and AI for social good, generally.

Our article supplies at least three contributions. First, it shows how a synthesis of causal inference and deep-neural networks, through c-GNF can further AI for Social Good. This contribution focuses on tailoring macroeconomic policies, minding children’s health and living conditions. Our analysis shows how this tailoring can be done empirically and with different personalized-treatment strategies at different population-level personalization granularity: ACE for the planetary (Global-South) level, CACE for the country level, and ICE for individual (child) level.

Second, in contrast to previous work (Daoud & Johansson, 2019), our findings reveal that the IMF program has beneficial average (ACE) effect. Our CACE finding indicates that the program is most beneficial for India and Bangladesh, and least beneficial for Iraq, Pakistan, Rwanda, and Zambia. Third, as virtually all previous IMF studies focus on ACE or CACE for evaluating the effect of macroeconomic policies (Daoud et al., 2017; Daoud & Reinsberg, 2018), our research breaks new ground by being the first in evaluating ICE on children. Our individual child level analysis finds that IMF program is beneficial to 61.22±4.89% and harmful to 7.25±4.89% of a sample of 1,941,734 children from the Global-South population.

Although we apply c-GNFs to the setting of IMF and children, c-GNFs generalize to other settings and P³A. There is a large application space in AI for social good, covering many social challenges in criminal justice, public health, education, and social welfare (Lakkaraju et al., 2015; Potash et al., 2015; Sankaran et al., 2017; Ghani, 2018; Ye et al., 2019; Daoud & Dubhashi, 2020; Shiba et al., 2021a; Kino et al., 2021). Thus, one important future task is to evaluate the strengths and weaknesses of c-GNFs, and continue refining their capacities.

2. Notation, Problem Definition and Assumptions

In this section, we discuss the formulation of the causal problem focusing on the relationship between the treatment (IMF program) and the outcome (child poverty). We define our notations and discuss key assumptions. As our analy-
sis relies on observational data, we are making conditional independence assumptions: the treatment assignment and the outcome are independent conditional on a set of confounders. These confounders have to block all back-door paths (Pearl, 2009b). To enforce conditional independence, we rely on existing social-science studies on IMF and child poverty (Daoud et al., 2017; Daoud & Reinsberg, 2018; Daoud & Johansson, 2019). Fig. 1(a) depicts a conceptual causal-Directed Acyclic Graph (c-DAG) identifying the impact of the IMF program on the degree of child poverty derived from Daoud & Johansson (2019). In Fig. 1(a), A denotes the IMF program as the treatment (exposure, action or intervention) and Y denotes the degree of child poverty as the outcome of causal interest. Moreover, C denotes the observed confounders of A and Y, and O are other causes of Y but not of A (macroeconomic factors). Specifically, C includes macroeconomic factors such as country’s economy, polity, political will, public spending, trade, and governance. These macroeconomic factors are the main confounders. O includes microeconomic and family related factors such as child’s age, sex and family wealth. Although these microeconomic variables are unnecessary for adjusting, they are critical to individualize the treatment. The c-DAG used in this work involves 32 random variables and it is presented in Fig. 5 of Appendix C. We note that the c-DAG is constructed with care with the help of social scientists who are the domain experts well aware of the socio-economic literature. However careful the c-DAG is considered, there might be some confounding variables which might be missing and hence, in the assumptions below, we explicitly state the unconfoundedness assumption. Unfortunately, unconfoundedness is an untestable assumption (Rubin, 1990; Robins & Hernán, 2008) and identifiability of the causal effect is only achievable under unconfoundedness assumption (Pearl, 2012).

Our main objective in this work is to estimate the causal effect of the treatment (IMF program) at any given population level and personalize treatments depending on the causal effects at that population level or personalization granularity. Depending on the personalization granularity, the instantiation of the causal effect might vary, i.e., ACE defines the causal effect when personalization granularity is the entire population level (Global-South), CACE defines the causal effect when personalization granularity is a sub-population level (e.g., country, age, gender, etc.), and ICE defines the causal effect when personalization granularity is the individual child level. Therefore, in the personalization granularity spectrum, we have ACE that offers no personalization and ICE that offers highest personalization at the extremes, with CACE offering intermediate personalization. Formally, ACE, CACE and ICE are defined as follows.

\[
\text{ACE} = E[(1-A)Y_0] \\
\text{CACE}(c) = E[(1-A)Y_c] \\
\text{ICE}_c = (1-Y_0)(C=c,O=0) \\
\text{ICE}_{a,c} = (1-Y_0)(C=c,O=0) \\
\]

where \( Y_0 \) denotes the potential outcome of \( Y \) under the intervention \( A=a \). Eqs. (3) and (4) differ in that the former considers the actual degree of child poverty, whereas the latter considers the threshold \( > 1 \) to obtain an indicator of child poverty, i.e., degrees of 2 to 7 indicate poor and degrees of 0 and 1 indicate not poor, a simplifying assumption considered in Daoud & Johansson (2019).

For any given personalization granularity, the treatment assignment condition identifies the optimal treatment to prescribe at the given personalization granularity. In our analysis, we consider six treatment strategies at three per-
sonalization granularities as summarized in Table 1 of Appendix B, i.e., Global-South (no personalization), country (intermediate personalization), and individual child (finest personalization) levels. The six treatment strategies from our study are described as follows:

1. $TS_0$: The IMF program is discouraged ($A=0$) for the entire Global-South. Then, there is no personalization at any granularity.

2. $TS_1$: The IMF program is encouraged ($A=1$) for the entire Global-South. Then, there is no personalization at any granularity.

3. $TS_{0k}$: The IMF program is encouraged ($A=1$) or discouraged ($A=0$) for the entire country based on the observed treatment, i.e., the countries treated get encouraged and the rest do not.

4. $TS_C$: The IMF program is encouraged ($A=1$) or discouraged ($A=0$) or neither for the entire country based on the country’s CACE in Eq. (2). Then, there is personalization at country level.

5. $TS_1$: The IMF program is encouraged ($A=1$) or discouraged ($A=0$) or neither based on the child’s ICE in Eq. (3). Then, there is personalization at child level.

6. $TS_{1l}$: The IMF program is encouraged ($A=1$) or discouraged ($A=0$) or neither based on the child’s ICE in Eq. (4). Then, there is personalization at child level.

We note that the expectations in Eqs. (1) and (2) are with respect to the samples from the interventional distributions $P(Y_a)$. However, the fundamental problem of causal inference dictates that the counterfactual outcomes under the unobserved interventions are always missing, i.e., the samples from $P(Y_a)$ are not readily available to compute the expectations in Eqs. (1) and (2). These missing values can be computed using the observational data under the following six main assumptions: (a) unconfoundedness or no unobserved confounders (Rubin, 1990), (b) positivity or overlap or common support or extrapolation (Rosenbaum & Rubin, 1983), (c) conditional ignorability or exchangeability (Rosenbaum & Rubin, 1983), (d) consistency (Robins, 1986; Cole & Frangakis, 2009; VanderWeele, 2009), (e) no interference or Stable Unit-Treatment Value Assumption (SUTVA) (Cox, 1958), and (f) modularity or independent mechanisms or autonomy or invariance (Pearl, 2009b; Peters et al., 2017). As our study relies on observational data, some of these assumptions, in particular unconfoundedness and positivity, are untestable. This limitation applies to all observational study. Therefore, any observational study results should be evaluated, jointly with the plausibility of these causal identification assumptions. If one believes the assumptions lack plausibility, then it is possible to make improvement to the unconfoundedness and positivity assumptions by changing the c-DAG and collecting more data. Even though these improvements might change the results, the general mechanics of applying c-GNF to P$^A$ specifically and AI for social good, generally, remains the same.

For the c-DAG in Fig. 1(a), the observational data can at best help modeling the observational joint distribution as

$$P(O, C, A, Y) = P(Y|A, C, O)P(A|C)P(C)P(O).$$

As the c-DAG considered assumes unconfoundedness, the $do$-calculus (Pearl, 2012) provides a way to identify a valid adjustment set and expresses $P(Y_a)$ from the components of the observational joint distribution in (5) as,

$$P(Y_a) = \sum_{\{c,o\}} P(Y|A=a, C=c, O=o)1_{(A=a)}P(C=c)P(O=o),$$

where the probability term corresponding to the independent mechanism of the intervened variable is replaced by the indicator function $1_{(A=a)}$ denoting the intervention $A=a$ (modularity assumption). Even with the identified adjustment set $\{C\}$ and the non-parametric expression for the interventional distribution in Eq. (6), the computation of ACE or CACE in Eqs. (1)-(2) still requires a complex and potentially time-consuming summation over the adjustment set variable $\{C\}$. Even though $do$-calculus identifies a non-parametric expression for ACE or CACE, $do$-calculus still lacks the estimation of ICE. However, in the next section, we see that it is possible to easily draw the samples from the distributions in Eqs. (5) and (6) using causal-Graphical Normalizing Flows (c-GNFs) without the need of the adjustment set or Eq. (6), using just ‘The First Law of Causal Inference’ (Pearl, 2009b; Pearl & Mackenzie, 2018). Hence easily estimating ACE, CACE and ICE in Eqs. (1)-(4) using Monte-Carlo expectation estimation from the samples drawn from the c-GNF.

3. SCM, encapsulated-SCM and c-GNF

Here, we briefly present the c-GNFs and describe their connection to SCMs. An SCM consists of a set of assignment equations describing the causal relations between the random variables of a causal system represented by a causal-Directed Acyclic Graph (c-DAG), such as

$$X_i := f_i(PA_i,U_i) \quad \text{with} \quad i=1, \ldots, d,$$

where $\{X_i\}_{i=1}^d$ represents the set of observed endogenous random variables, $PA_i$ represents the set of variables that are connoting parents of $X_i$, $\{U_i\}_{i=1}^d$ denotes the set of exogenous noise random variables, and $\{f_i\}_{i=1}^d$ denotes the set of functions (independent mechanisms) that generate the the endogenous variable $X_i$ from its observed causes $PA_i$ and noise $U_i$. An SCM is also
referred as Functional Causal Model (FCM) due to the functional mechanisms \( \{ f_i \}_{i=1}^d \). An SCM can be seen as a transformation \( F: U \rightarrow X \) such that \( X = F(U) \) where \( U = [U_1, \ldots, U_d]^{\mathbb{R}^d} \) and \( X = [X_1, \ldots, X_d]^{\mathbb{R}^d} \). Let \( \mathcal{G} \) represent a c-DAG with \( \{ X_1, \ldots, X_d \} \) as the set of nodes and adjacency matrix \( A_{\mathcal{G}} \in \{0, 1\}^{d \times d} \). Let \( P_X(X) \) represent the joint distribution over the endogenous variables \( X \), which factorizes according to \( \mathcal{G} \) as

\[
P_X(X) = \prod_{i=1}^d P(X_i | PA_i), \quad (8)
\]

where \( PA_i = \{ X_j : A_{\mathcal{G}, i,j} = 1 \} \) denotes the set of parents of the vertex/node \( X_i \) in \( \mathcal{G} \). As we see in the next section, the factorized joint distributions in Eqs. (5) and (8) can be modeled using an autoregressive model parameterized by a deep-neural-network \( \theta \), which we denote by \( P_X(X; \theta) \). Such a model is what is called as c-GNF.

### 3.1. Causal-Graphical Normalizing Flows (c-GNFs)

Normalizing Flows (NFs) (Tabak & Vanden-Eijnden, 2010; Tabak & Turner, 2013; Rezende & Mohamed, 2015; Kobyzev et al., 2020; Papamakarios et al., 2021) are a family of generative models with tractable distributions where both sampling and density evaluation can be efficient and exact. A NF is a flow-based model with transformation \( T: X \rightarrow Z \) such that \( Z = T(X) \), where \( Z = [Z_1, \ldots, Z_d]^{\mathbb{R}^d} \) represents the base random variable of the flow-model with the base distribution \( P_Z(Z) \) that is usually a \( d \)-dimensional standard normal distribution for computational convenience and ease of density estimation. Hence the name normalizing flow. The defining properties of \( T \) are: (i) \( T \) must be invertible with \( T^{-1} \) as the inverse generative flow such that \( T^{-1} Z \rightarrow X \), and (ii) \( T \) and \( T^{-1} \) must be differentiable, i.e., \( T \) must be a \( d \)-dimensional diffeomorphism (Milnor & Weaver, 1997). Under these properties, from the change of variables formula, we can express the endogenous joint distribution \( P_X(X) \) in terms of the assumed base distribution \( P_Z(Z) \) as

\[
P_X(X) = P_Z(TX) | \det J_T(X) \mid. \quad (9)
\]

Calculating the joint density of \( P_X(X) \) requires the calculation of the determinant of the Jacobian of \( T \) with respect to \( X \), i.e., \( \det J_T(X) \). It is, therefore, advantageous for computational reasons to choose \( T \) to have an autoregressive structure such that \( J_T(X) \) is a lower-triangular matrix and \( \det J_T(X) \) is just the product of the diagonal elements. Autoregressive Flows (AFs) (Kingma et al., 2016; Papamakarios et al., 2017; Huang et al., 2018) are NFs that model an autoregressive structure. AFs are composed of two components, the transformer and the conditioner (Papamakarios et al., 2021). Under the assumption of a strictly monotonic transformer, AFs are universal density estimators (Huang et al., 2018). Among these AFs, Graphical Normalizing Flows (GNFs) (Wehenkel & Louppe, 2021) facilitate the use of a desired DAG representation as opposed to an arbitrary autoregressive structure. For causal inference, it is crucial that the DAG has a causal interpretation. In this work, we assume the true c-DAG \( \mathcal{G} \) for the GNFs, hence the name causal-GNFs (c-GNFs). In c-GNFs, we use Unconstrained Monotonic Neural Network (UMNN) (Wehenkel & Louppe, 2019), a strictly monotonic integration based transformer along with \( \mathcal{G} \) for the graphical conditioner to ensure c-GNF is an universal density estimators (Huang et al., 2018).

As Markov equivalent DAGs induce equivalent factorizations of the observational joint distributions, GNFs that use Markov equivalent DAGs represent the same observational joint distribution. This is problematic for causal and counterfactual inference, as different GNFs may represent different interventional joint distributions. In other words, to perform exact causal inference as we do in this work, the GNF needs to use causal-DAG \( \mathcal{G} \). Note that only a GNF with the true c-DAG \( \mathcal{G} \) encapsulates the true SCM, thus satisfying the modularity assumption, which is necessary for correct causal and counterfactual inference. This is formalized as

\[
X = F(U) = F(H(Z)) = F.H(Z) = \tilde{F}(Z) = T^{-1}(Z) \quad (10)
\]

where \( H: Z \rightarrow U \) is an auxiliary transformation such that \( U_i = h_i(Z_i) \) for \( i = 1, \ldots, d \), i.e., \( U = H(Z) \) without loss of generality. It follows from Eq. (10) that the c-GNF \( T^{-1} = F.H \) encapsulates the true SCM \( F \) as \( T^{-1} \) both encode \( \mathcal{G} \) in the graphical conditioner, thus providing a way to indirectly model \( F \) without making assumptions on \( U \) or the functional causal mechanisms \( F \) or the auxiliary transformation \( H \). This sets apart the c-GNFs from most other models for causal inference.

Training a c-GNF amounts to training the deep neural networks that parameterize the transformers and conditioners. This is typically done by maximizing the log-likelihood of the training dataset \( \{ X^{\ell} \}_{\ell=1}^{N_{\text{train}}} \), which is expressed as shown below, by using Eq. (9) for the summation term \( P_X(X^{\ell}; \theta) \)

\[
\mathcal{L}(\theta) = \sum_{\ell=1}^{N_{\text{train}}} \log P_X(X^{\ell}; \theta), \quad (11)
\]

where \( \theta \) denotes the parameters of the deep neural networks of the UMNN transformer and graphical conditioner, optimized using stochastic gradient descent. We utilize the Gaussian quantization trick from Balgic et al. (2022) presented in Algorithm 1 of Appendix E to model discrete variables into NFs using the fact that NFs are strongest in modeling Gaussian distributions seamlessly. Fig. 1(b) shows the c-GNF architecture for causal and counterfactual inference using \( T \) and \( T^{-1} \).
3.2. Counterfactual Inference and ‘The First Law of Causal Inference’

Since a c-GNF encapsulates the true SCM, this enables us to compute the potential outcomes under all the treatment strategies described in Section 2 using ‘The First Law of Causal Inference’ (Pearl, 2009a,b; Pearl & Mackenzie, 2018). This law essentially provides a framework to identify the unit level potential outcomes using an SCM, thereby addressing the fundamental missing value problem of counterfactual inference. The law states that the missing/unseen potential outcome $Y$ for a unit $Z^\ell$ under the treatment strategy $TS_a$, denoted as $Y_{TS_a}(Z^\ell)$, can be computed as

$$Y_{TS_a}(Z^\ell) = Y_{\tilde{F}_{TS_a}}(Z^\ell),$$

where $Y_{\tilde{F}_{TS_a}}(Z^\ell)$ represents the actual outcome $Y$ observed from the mutilated-SCM $\tilde{F}_{TS_a}$ that is obtained by removing the causes of $A$ from the encapsulated-SCM $\tilde{F}$, and setting $A$ to the value dictated by $TS_a$. Essentially, this law involves the following three steps.

1. **Abduction:** For a given observed unit-individual evidence $X^\ell$, the respective exogenous noise $Z^\ell$ corresponding to $X^\ell$ is recovered from the encapsulated-SCM $\tilde{F}$. In the case of a c-GNF, we have $Z^\ell = TX^\ell = \tilde{F}^{-1}X^\ell$ where the encapsulated-SCM noise $Z^\ell$ represents the unobserved yet unique identifier of an individual (e.g., DNA) of the $\ell$th individual with observed evidence $X^\ell$.

2. **Action:** The action or intervention on $A$ dictated by the treatment strategy $TS_a$ is performed after mutilating the structural equation corresponding to $A$ in the encapsulated-SCM $\tilde{F}$ (from the modularity assumption), which results in the mutilated-SCM $\tilde{F}_{TS_a}$.

3. **Prediction:** The recovered noises $Z^\ell$ from abduction are re-propagated through $\tilde{F}_{TS_a}$ and the outcomes $Y_{\tilde{F}_{TS_a}}(Z^\ell)$ observed are nothing but the potential outcomes $Y_{TS_a}(Z^\ell)$ we were interested to begin with.

4. Experimental Results and Discussion

In our experiments, we consider the IMF dataset used in Daoud & Johansson (2019); Daoud et al. (2017); Halleröd et al. (2013) with 1,941,734 children under the age of 18, cared for by 567,344 families residing in 67 countries from the Global-South. Fig. 4 in Appendix A shows the country-wise proportions of the observational samples in the IMF dataset. The benefit of this dataset is that it represents 2.8 billion samples (50%) of the world’s population by the year 2000. Fig. 5 in Appendix C shows the c-DAG with 32 variables such as economy, polity, public spending, living condition, etc. along with the observed IMF program treatment at a country level and the degree of the child poverty under the observed treatment considered in our study. The degree of the child poverty in the dataset is calculated based on 7 individual dimensions of poverty for each of the child based on 1) education, 2) health, 3) information, 4) malnutrition, 5) sanitation, 6) shelter, 7) water, resulting in an aggregate degree of child poverty between 0 and 7 with 0 indicating no poverty and 7 indicating severe poverty.

Due to the sensitive nature and the accompanying ethical considerations, the IMF dataset is not made publicly available and can be requested upon from the original authors of Halleröd et al. (2013); Daoud et al. (2017); Daoud & Johansson (2019). Since deep neural networks are prone to overfit, we split the data into 1,922,316 training, 9,709 validation and 9,709 test samples. We run multiple simulations for the hyperparameter selection and select the best hyperparameters with the best held-out validation loss. We strictly use only the training set samples for training and use the held-out validation set for early stopping to get the model with best validation loss. We further validate the generalization of the best validation loss model on the held-out test set. We use three fully-connected layers with [40, 30, 20] hidden units for the graphical conditioner and three fully-connected layers with [15, 10, 5] hidden units for the monotonic UMN transformer. We implement c-GNFs in Pytorch (Paszke et al., 2017) using GNF baseline code and AdamW (Loshchilov & Hutter, 2019) optimizer with learning-rate=3e−4 and a batch-size of 1024 (4GB of GPU memory) for all our experiments.

Fig. 2(a) shows the child poverty statistics sorted degree-wise for the entire Global-South observed across multiple treatment strategies for multiple simulations. The invariability of the boxplot corresponding to the strategy $TS_{Ob}$ (i.e., naturally observed treatment) validates the consistency assumption of the c-GNF model, as the same outcome is observed for the same observed treatments across multiple simulations. Thus, the c-GNF model satisfies consistency by construction. More importantly, Fig. 2(b) shows the IMF program advisability statistics or the proportions of personalization across multiple treatment strategies. Since $TS_I$ provides the finest personalization at the individual child level, we observe that IMF program is harmful (discouraged) for 7.25±4.89%, beneficial (encouraged) for 61.22±4.08%, and neutral for the rest. Contrast to $TS_I$, $TS_C$ provides intermediate personalization at country level, hence we observe that IMF program is harmful (discouraged) for 10.8±7.1%, beneficial (encouraged) for 88.1±6.04%, and neutral for the rest. It should noted that the invariability of the $TS_0$, $TS_I$ and $TS_C$ is due to the fact that the treatments for these strategies are fixed beforehand.

As for $TS_{II}$, our findings also indicate the importance of considering all the seven poverty degrees in the analysis.
Figure 2. All the boxplot results are reported from 5 different random seeded simulations to indicate the variability in the c-GNF results across multiple simulations. (a) presents the number of samples in each of the 8 possible degrees of child poverty across different treatment strategies in the entire Global-South. (b) indicates the advisability of the IMF program for each treatment strategy in the entire Global-South. (c) indicates the average degree of child poverty for each treatment strategy in the entire Global-South. (d) indicates the ACE estimated in the entire Global-South.

Figure 3. The region-wise average causal effect estimates from 5 random seeded simulations plotted along with the zero-ACE line.

In contrast to Daoud & Johansson (2019) that considers only the indicator of child poverty, as the IMF program may help children move from severe to moderate poverty. This may get obscured if a binary indicator of poverty (poor vs non-poor) is used, leading to the erroneous conclusion that personalization is irrelevant. The fundamental oversight from over-simplification by grouping poverty degrees 2 to 7 neglects the improvements within the group, e.g., 7 to 2 is a significant improvement for which the IMF program will be rightly encouraged in $TS_1$. However, $TS_{11}$ considers no change in the indicator of the poverty and hence wrongly assumes that the IMF program is neutral for resource optimization. In other words, $TS_{11}$ wrongly values a change of degree from 2 to 1 more than a change of degree from 7 to 2. This is specifically seen in the advisability plots of $TS_{11}$ where, compared to $TS_1$, there is an increase in the neutral advisability over the encouragement. This important finding of ours reinforces the need for the radical rethinking of Banerjee et al. (2011)’s Nobel Prize winning work stated previously in Section 1 that suggests one should not...
resort to over-simplification of the poor into cartoon characters by ignoring their individual characteristics in all their rich complexities and relative improvements in their living conditions.

Fig. 2(c) shows the variation of the average degree of child poverty for all the treatment strategies obtained from averaging Fig. 2(a). From Fig. 2(c), the treatment strategies can be sorted in the decreasing order of their expected degree of child poverty $E[Y_{TS_a}]$: $TS_0 > TS_{1t} > TS_{Oh} > TS_1 > TS_C > TS_I$. This indicates that the IMF program is beneficial for the Global-South ($TS_0 > TS_{Oh} > TS_1$). Fig. 2(c) shows that the personalization at the country level due to the treatment strategy $TS_C$ (purple) represents a significant reduction in child poverty over the ‘one-size-fits-all’ treatment strategy $TS_1$ (blue) and the sub-optimal naturally observed treatment strategy $TS_{Oh}$ (orange). Moreover, personalization at the individual child level is even more beneficial ($TS_C > TS_I$).

Fig. 2(d) reconfirms the beneficial nature of the IMF program across multiple simulations as the average degree of the child poverty is observed to be reduced by $1.2 \pm 0.24$ degree. Similarly, Fig. 3 shows that most countries experience a reduction in the average degree of child poverty from the IMF program, and hence the program has beneficial effects on the Global-South as a whole. Figs. 6 in Appendix D provides a qualitative visual analysis of the country-wise average degree of child poverty in the form of the six potential (interventional) worlds under each of the six treatment strategies from one of the 5 simulations. From these results, we see that it is indeed possible to perform P3A, under the assumptions in Section 2, to effectively combat social ills in contrast to the ‘one-size-fits-all’ approaches.

4.1. On interpretability and explanability of the c-GNF

Since the c-GNF is able to provide the encapsulated-SCM, this enables to go beyond total causal effect estimation and into the mediation analysis dealing with natural/controlled direct/indirect effects (Pearl, 2009b). The probabilistic nature of the c-GNF further enables one to analyse important causal quantities such as probabilities of causation (Pearl, 1999) to dive deep into the causal aspects of the social system under study and enable policymakers with personalized policies. Finally, the c-GNF/encapsulated-SCM is able to successfully navigate across all the three rungs of the ‘Ladder of Causation’ (Pearl & Mackenzie, 2018), i.e., (i) association, (ii) intervention, and (iii) counterfactuals, to be able to answer every causal and counterfactual questions using ‘The First Law of Causal Inference’ (Pearl, 2009b; Pearl & Mackenzie, 2018). Of all the 16 macroeconomic variables that cause child poverty, we observed that political will is the most impactful as it defines the way of the nation and hence impacts in a macroeconomic shock that trickles down to the individual child level, via microeconomic family-level factors. Observing the c-DAG in Fig. 5 of Appendix C, it is evident that the political will of a country determines other macroeconomic factors such as public spending, polity, governance, and also trickles down to individual child status via intermediate microeconomic factors. For example, consider Pakistan or Iraq where a negative political will is observed and hence the IMF programs are seen to be harmful (recall Fig. 3) due to the tendency of the monetary funds being misused for state-sponsored terrorism instead of the primary objective of economic stability. In contrast, countries such as India and Bangladesh, where positive political will is observed, are seen to be benefiting the most from the IMF program both economy and child poverty-wise. Similarly, for instance, the most number of current CEOs of tech-giants such as Google, Microsoft, Adobe, IBM, Twitter can loosely be attributed to the positive effects of positive political will in India. It was also observed that the macroeconomic shocks on child poverty can be isolated if the macroeconomic factors at the family-level such as family wealth, family head education, number of children/adults in the family are well conditioned. This strengthens the belief of personalizing the treatments/policies to effectively combat/isolate the macro-level shocks trickling down to individual family or child. In summary, our major observations are two-folds: (i) political will is an important macroeconomic factor that defines both macro- and micro- socio-economic status of a nation, (ii) the IMF program coupled with a positive political will is seen to be beneficial for the child poverty status of the country.

5. Conclusion

In this article, we deployed causal-Graphical Normalizing Flows (c-GNFs) to draw insights from the real-world observational data on the impact of the IMF program on child poverty. Our findings in terms of ACE indicated the IMF program to be beneficial for the Global-South, in expectation. Apart from considering the traditional ‘one-size-fits-all’ treatment, we also proposed an empirical framework to formulate different personalized treatment strategies at different population granularity levels by computing ACE, country-wise CACE, and child-wise ICE. Our findings reinforced the radical thinking proposed by Banerjee et al. (2011) to address the problem of child poverty without oversimplification of the poor into cartoon characters. Even though we demonstrated the personalized treatment strategy formulation framework in a social science setting, our framework can be extended to the field of personalized medicine as well. The transparent nature of the c-GNF that is able to provide the encapsulated-SCM enables one to successfully navigate across all the three rungs of the ladder of causation to provide an end-to-end causal inference tool.
References

Åkerström, J., Daoud, A., and Johansson, R. Natural language processing in policy evaluation: Extracting policy conditions from IMF loan agreements. In Nordic Conference on Computational Linguistics, pp. 316–320, 2019.

Balgi, S., Peña, J. M., and Daoud, A. Personalized public policy analysis in social sciences using causal-graphical normalizing flows. In AAAI conference on Artificial Intelligence : AI for Social Impact track (A Virtual Conference), 2022.

Banerjee, A., Banerjee, A. V., and Duflo, E. Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty. Public Affairs, 2011.

Cole, S. R. and Frangakis, C. E. The consistency statement in causal inference: A definition or an assumption? Epidemiology, 20(1):3–5, 2009.

Cox, D. R. Planning of Experiments. New York: Wiley, 1958.

Daoud, A. Unifying studies of scarcity, abundance, and sufficiency. Ecological Economics, 147:208–217, 2018.

Daoud, A. and Dubhashi, D. Statistical modeling: the three cultures. arXiv preprint arXiv:2012.04570, 2020.

Daoud, A. and Johansson, F. Estimating treatment heterogeneity of international monetary fund programs on child poverty with generalized random forest. SocArXiv, 2019.

Daoud, A. and Reinsberg, B. Structural adjustment, state capacity and child health: evidence from IMF programmes. Epidemiology, 48(2):445–454, 2018.

Daoud, A., Nosrati, E., Reinsberg, B., Kentikelenis, A. E., Stubbs, T. H., and King, L. P. Impact of international monetary fund programs on child health. Proceedings of the National Academy of Sciences (PNAS), 114(25):6492–6497, 2017.

Daoud, A., Reinsberg, B., Kentikelenis, A. E., Stubbs, T. H., and King, L. P. The international monetary fund’s interventions in food and agriculture: An analysis of loans and conditions. Food Policy, 83:204–218, 2019.

Ghani, R. Data Science for Social Good and Public Policy: Examples, Opportunities, and Challenges. In ACM Special Interest Group in Information Retrieval (SIGIR), pp. 3–3, 2018.

Halleröd, B., Rothstein, B., Daoud, A., and Nandy, S. Bad governance and poor children: A comparative analysis of government efficiency and severe child deprivation in 68 low-and middle-income countries. World Development, 48:19–31, 2013.

Hernán, M. A. and Robins, J. M. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC, 2009.

Huang, C., Krueger, D., Lacoste, A., and Courville, A. C. Neural autoregressive flows. In International Conference on Machine Learning (ICML), pp. 2083–2092, 2018.

Kingma, D. P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., and Welling, M. Improved variational inference with inverse autoregressive flow. In Neural Information Processing Systems (NeurIPS), pp. 4743–4751, 2016.

Kino, S., Hsu, Y.-T., Shiba, K., Chien, Y.-S., Mita, C., Kawachi, I., and Daoud, A. A scoring review on the use of machine learning in research on social determinants of health: Trends and research prospects. Social Science & Medicine - Population Health (SSM-PH), 15:100836–100855, 2021.

Kobyzev, I., Prince, S., and Brubaker, M. Normalizing flows: An introduction and review of current methods. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), pp. 1–1, 2020. doi: 10.1109/TPAMI.2020.2992934.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences (PNAS), 116(10):4156–4165, 2019.

Kakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., and Addison, K. L. A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes. In ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD), pp. 1909–1918, 2015.

Koshchilov, I. and Hutter, F. Decoupled weight decay regularization. In International Conference on Learning Representations (ICLR), 2019.

Milnor, J. and Weaver, D. W. Topology from the Differentiable Viewpoint. Princeton University Press, 1997.

Papamakarios, G., Murray, I., and Pavlakou, T. Masked autoregressive flow for density estimation. In Neural Information Processing Systems (NeurIPS), pp. 2338–2347, 2017.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. Journal of Machine Learning Research (JMLR), 22(57):1–64, 2021.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic Differentiation in PyTorch. NeurIPS Workshops, 2017.
Pearl, J. Probabilities of causation: Three counterfactual interpretations and their identification. Synthese, 121(1): 93–149, 1999.

Pearl, J. Causal inference in statistics: An overview. Statistics Surveys, 3:96–146, 2009a.

Pearl, J. Causality: Models, Reasoning and Inference. Cambridge University Press, 2009b.

Pearl, J. The do-calculus revisited. In Uncertainty in Artificial Intelligence (UAI), pp. 3–11, 2012.

Pearl, J. and Mackenzie, D. The Book of Why: The New Science of Cause and Effect. 2018.

Peters, J., Janzing, D., and Schlkopf, B. Elements of Causal Inference: Foundations and Learning Algorithms. The MIT Press, 2017.

Potash, E., Brew, J., Loewi, A., Majumdar, S., Reece, A., Walsh, J., Rozier, E., Jorgenson, E., Mansour, R., and Ghani, R. Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. In ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD), pp. 2039–2047, 2015.

Rezende, D. and Mohamed, S. Variational inference with normalizing flows. In International Conference on Machine Learning (ICML), pp. 1530–1538, 2015.

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

Robins, J. M. and Hernán, M. A. Estimation of the causal effects of time-varying exposure. Longitudinal Data Analysis (LDA), pp. 553–599, 2008.

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

Rubin, D. B. Formal mode of statistical inference for causal effects. Journal of Statistical Planning and Inference (JSPI), 25(3):279–292, 1990.

Sankaran, K., Garcia-Olano, D., Javed, M., Alcala-Durand, M. F., De Unáuñe, A., van der Boor, P., Potash, E., Avalos, R. S., Encinas, L. I. A., and Ghani, R. Applying Machine Learning Methods to Enhance the Distribution of Social Services in Mexico. arXiv:1709.05551 [stat], 2017.

Shiba, K., Daoud, A., Hikichi, H., Yazawa, A., Aida, J., Kondo, K., and Kawachi, I. Heterogeneity in cognitive disability after a major disaster: A natural experiment study. Science Advances, 2021a.

Shiba, K., Torres, J. M., Daoud, A., Inoue, K., Kanamori, S., Tsuji, T., Kamada, M., Kondo, K., and Kawachi, I. Estimating the impact of sustained social participation on depressive symptoms in older adults. Epidemiology, 2021b.

Tabak, E. and Vanden-Eijnden, E. Density estimation by dual ascent of the log-likelihood. Communications in Mathematical Sciences, 8(1):217–233, 2010.

Tabak, E. G. and Turner, C. V. A family of nonparametric density estimation algorithms. Communications on Pure and Applied Mathematics, 66(2):145–164, 2013.

Textor, J., van der Zander, B., Gilthorpe, M. S., Liškiewicz, M., and Ellison, G. T. Robust causal inference using directed acyclic graphs: the r package ‘dagitty’. Epidemiology, 45(6):1887–1894, 2016.

VanderWeele, T. J. Concerning the consistency assumption in causal inference. Epidemiology, 20(6):880–883, 2009.

Wehenkel, A. and Louppe, G. Unconstrained monotonic neural networks. In Neural Information Processing Systems (NeurIPS), pp. 1545–1555, 2019.

Wehenkel, A. and Louppe, G. Graphical normalizing flows. In International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 37–45, 2021.

Wodtke, G. T. Regression-based adjustment for time-varying confounders. Sociological Methods & Research (SMR), 49(4):906–946, 2020.

Ye, T., Johnson, R., Fu, S., Copeny, J., Donnelly, B., Freeman, A., Lima, M., Walsh, J., and Ghani, R. Using machine learning to help vulnerable tenants in New York city. In ACM Computing and Sustainable Societies (COMPASS), pp. 248–258, 2019.
A. IMF dataset country-wise observational samples

Country-wise number of observational samples ($N_c$) out of total samples ($N = \sum_{c=1}^{67} N_c = 1941734$) from the Global-South

B. Summary of the six treatment strategies

| #  | Personalization Granularity levels | Treatment assignment condition | Treatment assigned ($A$) | Treatment strategy |
|----|----------------------------------|-------------------------------|--------------------------|-------------------|
| 1  | None, i.e., entire Global-South level | 0 | 0 (D) | $TS_0$ (in red) |
| 2  | None, i.e., entire Global-South level | 1 | 1 (E) | $TS_1$ (in blue) |
| 3  | Intermediate, i.e., Country level | Naturally observed treatment | 0 (D) or 1(E) | $TS_{Ob}$ (in orange) |
| 4  | Intermediate, i.e., Country level | $\text{CACE}(C=c)$ | 0 (D) or 1 (E) or 0/1 (N) | $TS_C$ (in purple) |
| 5  | Finest, i.e., Individual child level | $\text{ICE}_i(C=c, O=o)$ | 0 (D) or 1 (E) or 0/1 (N) | $TS_I$ (in green) |
| 6  | Finest, i.e., Individual child level | $\text{ICE}_{it}(C=c, O=o)$ | 0 (D) or 1 (E) or 0/1 (N) | $TS_{It}$ (in olive) |

Table 1. Six IMF treatment strategies considered in our real-world study with the IMF dataset.
C. Hypothesized causal-DAG for the current social system under study

Figure 5. The detailed causal-DAG involving the 32 variables hypothesized for the SCM, created using Dagitty (http://www.dagitty.net/dags.html) (Textor et al., 2016).
D. Counterfactual worldmaps for different treatment strategies

Figure 6. Six potential (interventional) worlds under the six treatment strategies indicating the country-wise average degree of child poverty.
E. Algorithm for Gaussian dequantization

Algorithm 1 Gaussian Dequantization

**Input:** Discrete variables \{D^\ell\}_{\ell=1}^b
Generate \( \hat{D}^\ell \sim N(\mu = D^\ell, \sigma^2 = 1/36) \).
**Output:** Dequantized / continuous variables \{\hat{D}^\ell\}_{\ell=1}^b

Algorithm 2 Gaussian Quantization

**Input:** Continuous variables \{\hat{D}^\ell\}_{\ell=1}^b
Generate \( D^\ell = \text{clamp}(\text{round}(\hat{D}^\ell), \text{min}=0, \text{max}=N-1) \)
**Output:** Quantized / discrete variables \{D^\ell\}_{\ell=1}^b