The impact of austerity on children: Uncovering effect heterogeneity by political, economic, and family factors in low- and middle-income countries

Citation for the original published paper (version of record):
Daoud, A., Johansson, F. (2024). The impact of austerity on children: Uncovering effect heterogeneity by political, economic,
and family factors in low- and middle-income countries. Social Science Research, 118.
http://dx.doi.org/10.1016/j.ssresearch.2023.102973

Social Science Research 118 (2024) 102973

Contents lists available at ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch

The impact of austerity on children: Uncovering effect heterogeneity by political, economic, and family factors in low- and middle-income countries

Adel Daoud [a,b,c,*], Fredrik D. Johansson [c]

[a] Center for Population and Development Studies, Harvard T.H. Chan School of Public Health, Harvard University, Boston MA, USA
[b] Institute for Analytical Sociology, Linköping University, Sweden
[c] The Division of Data Science and Artificial Intelligence, The Department of Computer Science and Engineering, Chalmers University of Technology, Sweden

ARTICLE INFO

JEL classification:
I32 (Measurement and Analysis of Poverty)
I15 (Health and Economic Development)
I38 (Government Policy
Provision and Effects of Welfare Programs)
J11 (Demographic Trends
Macroeconomic Effects
and Forecasts)
J18 (Public Policy)
J13 (Fertility
Family Planning
Child Care
Children
Youth)
Z13 (Economic Sociology Economic
Anthropology
Language
Social and Economic Stratification)
Z18 (Public Policy)

Keywords:
Child welfare
Causal inference
Demography
Families
Population heterogeneity
Poverty
Austerity
Public policy
Social stratification

ABSTRACT

Which children are most vulnerable when their government imposes austerity? Research tends to focus on either the political-economic level or the family level. Using a sample of nearly two million children in 67 countries, this study synthesizes theories from family sociology and political science to examine the heterogeneous effects on child poverty of economic shocks following the implementation of an International Monetary Fund (IMF) program. To discover effect heterogeneity, we apply machine learning to policy evaluation. We find that children's average probability of falling into poverty increases by 14 percentage points. We find substantial effect heterogeneity, with family wealth and governments' education spending as the two most important moderators. In contrast to studies that emphasize the vulnerability of low-income families, we find that middle-class children face an equally high risk of poverty. Our results show that synthesizing family and political factors yield deeper knowledge of how economic shocks affect children.

* Institute for Analytical Sociology, Linköping University, Sweden.
E-mail address: adel.daoud@liu.se (A. Daoud).

https://doi.org/10.1016/j.ssresearch.2023.102973
Received 20 July 2022; Received in revised form 26 July 2023; Accepted 7 December 2023
Available online 30 December 2023

1. Introduction

Sociologists have long been interested in the effects of economic change on individuals' well-being. A large body of research in family sociology examines how dynamics within households translate economic shocks into material and social consequences (Boss et al. 2016; Conger et al., 1992; Elder 1998; Ferraro et al. 2016; Garfinkel et al. 2016; McLanahan and Percheski 2008; Schafer et al. 2011). A second substantial body of research in political sociology shows how institutional constellations moderate the effects of economic turmoil on the living conditions of populations (Babb 2005; Beckfield 2018; Brady 2019; Burgard and Kalousova 2015; Redbird and Grusky 2016; Shandra et al. 2012). These two traditions—one focused on micro-level processes and the other on macro-level processes—have mainly evolved separately. We argue that they should be brought together because the effects of economic shocks are likely to vary, depending on both family resources and government choices. However, empirical investigations of how such macro and micro processes work together to either exacerbate or alleviate such effects remain limited.

In family sociology, Elder (1998) pioneering study, *Children of the Great Depression*, inspired a stream of research on families under economic stress. He showed how the deep stress that families in California endured after the Great Depression affected children's short- and long-term outcomes. Recent studies have extended this research to other contexts by investigating what happens when an economic shock causes breadwinners to lose their jobs, thereby reducing parents' ability to provide for their children's well-being (e.g., Boss et al., 2016; Garfinkel et al., 2016). Although this research tradition acknowledges the importance of macro-social contexts, it tends to zoom in on the micro-dynamics within households rather than focusing on these societal variations.

In contrast, political sociology takes a broader view by looking at institutional arrangements and policy choices as moderators of the effects of economic turmoil on populations' living conditions (Babb 2005; Beckfield 2018; Burgard and Kalousova 2015; Redbird and Grusky 2016). This research sheds light, for example, on the political economy of crises, with much of it focusing on the policies of international organizations, most notably the International Monetary Fund (IMF) (Babb 2005; Daoud et al., 2017; Daoud and Reinsberg 2018; Halliday and Carruthers 2009; Shandra et al., 2012; Stuckler and Basu 2013). This literature analyzes the content of these policies (Åkerström et al. 2019; Daoud et al., 2019a,b; Kentikelenis et al. 2016) and identifies, for example, how IMF policies affected institutional change in Eastern Europe after the dissolution of the Soviet Union (Hamm et al. 2012). However, macro perspectives tend only to include individual- or household-level outcomes, such as child poverty or mortality, if these outcomes are indicators of social development (Conley and Springer 2001:770).

The relatively independent evolution of micro and macro traditions has left at least two gaps in scholarly understanding of how crises affect vulnerable populations such as children. First, previous research has accentuated one of two analytical units, rather than examining the macro-micro link itself (Brady 2019; Coleman 1987). A synthesis of insights from family and political sociology would allow sociologists to consider a broader range of causal mechanisms through which economic shocks affect families and their children (Burgard and Kalousova 2015; Garfinkel et al., 2016; McLanahan and Percheski 2008).

Second, a combined macro-micro approach would lay a foundation for systematically identifying which subgroups of children are at higher risk of poverty during a crisis (Brady 2019). Because families have different resources and behaviors and are embedded in different sets of institutions, economic shocks will likely produce a large amount of impact heterogeneity. Analyzing the average effect of such shocks on children's risk of poverty—that is, using one summary estimate to cover several subgroups of children—will mask important social differences in how families and their children are likely to deal with such changes.

Our study combines macro-micro theories to ask and answer two empirical questions: How does economic turmoil following IMF programs affect child poverty in low and middle-income countries, and how do macro and micro factors interact to generate impact heterogeneity in the relationship between IMF programs and children's risk of poverty? Using political sociology (Beckfield 2018), we answer the first question by examining IMF austerity programs as a type of economic crisis (Babb 2005). Applying family sociology, we focus on child poverty as an outcome because poverty in the early years strongly predicts educational attainment, adult income, health, and other vital outcomes throughout an individual's life course (Garfinkel et al., 2016). We use a multidimensional indicator of deprivation that measures children's access to necessary resources, including water, food, shelter, sanitation, health care, information, and education (Gordon et al., 2003). These indicators provide a widely agreed standard for the provisioning of basic human needs and human rights, and their definition informs social policy world-wide (Halleröd et al., 2013; Pemberton et al. 2012). In low- and middle-income countries, poverty is more accurately represented by deprivation measures than by income measures.

To answer the second question, we first observe that despite some findings that IMF austerity programs may be most harmful to vulnerable groups (Babb 2005), relatively little is known about how the effects are distributed within populations. Therefore, using both macro and micro data, we identify the distribution of impact across population subgroups, thereby revealing how children's lives are simultaneously influenced by families' capabilities and societal characteristics.

We analyze data from 2000, plus or minus six years, on nearly two million children, living in about 570,000 households and residing in 67 low- and middle-income countries. The data is representative of about half the world's population. Given that plausible theories suggest that many micro and macro characteristics could moderate the effects of IMF economic shocks, we employ machine learning (ML)—a set of algorithms that automatically discover patterns in large population data—to inductively identify which moderators account for impact heterogeneity (Athey et al. 2019; Balgi et al. 2022; Daoud and Dubhashi 2023; Jerzak et al. 2023).

ML allows us to respond to the sociological challenge to move beyond linear models. As Goldthorpe (2015) and others (e.g., Abbott 1988) have emphasized, sociology requires a scalable method and framework "through which variability and heterogeneity can be accommodated and in various ways exploited" (Goldthorpe 2015:31). Traditional quantitative practices rely primarily on techniques that assume linearity between independent variables and an outcome. However, linear models are unsuitable for detecting the complex interactions present in large population data (Abbott 1988; Xie et al. 2012). Using ML for causal inference (Daoud and

Dubhashi 2023), our article shows how this can be achieved.

Several studies have relaxed the linearity assumption. One early study is by Charles Ragin (1987), whose "qualitative comparative analysis"—a method for identifying the empirical validity of a theory in small-sized case studies—has been used in various empirical applications (Rihoux and Ragin 2008). Bail (2008) uses this form of analysis to identify how immigrants and native Europeans classify "us" and "them" from survey data. More recently, Goldberg (2011) developed what he calls "relational class analysis"—a method for identifying similar groups of individuals in multivariate data—to study people's attitudes in a variety of domains, from finding cultural similarities in musical taste to the formation of voters' belief systems (Baldassarri and Goldberg 2014). Latent class analysis (Bonikowski 2015), sequence analysis (), and cluster analysis (Garip 2012, 2016) provide additional methods for using complex data to identify population heterogeneity. Yet while all these analytical methods aim at transcending linearity (see Molina and Garip, 2019, for a general overview), none of them is adapted for disentangling population heterogeneity within a causal or policy evaluation research design.

To fill this population-heterogeneity knowledge gap, our study uses recent methodological advances in the overlap between causal inference and computer science (Pearl and Mackenzie 2018). This burgeoning literature posits that a specific class of ML algorithms is well equipped for identifying causal heterogeneity in population data (Athey and Imbens 2017; Mullainathan and Spiess 2017). While traditional ML algorithms do not differentiate between the exposure of interest (treatment) and other covariates (control variables) to predict the outcome, this new class of algorithms is specifically designed for retaining that difference. Moreover, as these new algorithms can estimate many thousands of models concurrently and use out-of-sample data for predictions—thereby reducing bias due to researcher discretion—they are often more robust than linear models.

Thus, this ML-powered causal approach resonates with recent calls in sociology for evaluating scientific findings on their out-of-sample predictive performance (Watts 2014). Despite Turco and Zuckerman's (2017) concern that interpretability—the ability to draw substantive conclusions from statistical models—might be lost in such computational approaches to sociology, we show how our approach helps us conduct complex sociological theorizing without loss of interpretability. By "interpretability," we mean a method that is both transparent about how it works (Lipton 2017) and that its results serve causal inference (Athey and Imbens 2017; Athey et al., 2019). Section 5.2 delineates how our ML-powered approach is causally interpretable. By comparing our method with commonly used methods, this section also discusses four advantages of ML for causal inference.

Our empirical findings demonstrate the benefits of our research design. First, we show that the economic shockwaves following IMF interventions produce adverse outcomes for children who face an increased probability of poverty by an average of 14 percentage points. Second, our ML algorithm suggests that families' material assets and a government's spending on education are the two most important factors moderating this increased risk of child poverty. For example, we find that middle-class children face at least as high a risk of falling into poverty as the children of low-income families. Our findings also show that children residing in countries with high education spending prior to IMF program implementations are at greater risk of poverty than children in low-education spending.

Third, differences between countries account for about half of the variation in the effects of IMF interventions on children's poverty risk, with household-level differences within populations accounting for the other half. Together, these two portions of variations reveal the deep connection between family and political sociology.

2. Background

The first aim of our analysis is to capture all the impact that an IMF program can yield on children's living conditions. In causal terms, we are interested in the "total effect," which is the sum of all paths (i.e., the direct and indirect effects) connecting the treatment and the outcome (Pearl 2009). Statistically, to estimate the total effect, we do not need to specify all pathways.

Nonetheless, any causal framework needs to propose plausible pathways through which treatment affects an outcome (Imbens and Rubin 2015). Such theoretically described pathways provide a conceptual heuristic of how exposure affects an outcome. We suggest two main pathways through which the total effect travels. While these two pathways are inexhaustive, they provide a heuristic for how IMF policies affect child poverty.

The first causal pathway focuses mainly on parents' resources (Elder 1998). IMF austerity programs produce economic effects similar to that of an economic crisis. IMF programs affect parents mainly through the labor market. Although IMF programs are aimed at stabilizing the economy and supporting long-term economic growth, the policies contained in these programs are likely to generate more adverse than beneficial effects in the short term (Babb 2005; Stuckler and Basu 2013). The implementation of these policies leads to a cascade of effects: They change public policies, thereby affecting labor markets, and shifts in these markets rekindle parents' employment and earning prospects (Boss et al., 2016), in turn affecting how well the parents can care for their children (McLanahan and Percheski 2008).

The second causal pathway refers to reduced government spending, affecting children's risk of poverty. IMF policies are likely to influence governments to reduce their spending to balance their budgets, and reduced public spending erodes the quality of the social programs serving children. Although some studies find no clear link between IMF programs and social spending (Martin and Segura-Ubiergo 2004), other studies show that IMF policies often require governments to reduce their spending on education (Daoud 2021; Rowden 2011) and health care (Stuckler and Basu 2013). These reductions lead, at the very least, to short-term adverse effects on populations (Daoud et al., 2017; Shandra et al., 2012; Stuckler and Basu 2013). Through the government-spending pathway, parents play only a secondary role. Children often have access to complementary resources via the health and education systems, regardless of their parents' resources (Kane 2009).

2.1. *Effect heterogeneity emanates from three sourcestification strategy for*

The second aim of our analysis is to identify the main pretreatment factors moderating IMF's impact on child poverty. While a pathway captures all the mechanisms connecting the treatment to the outcome (i.e., mediation), effect heterogeneity captures pretreatment factors such that the effect of the treatment on the outcome varies for different levels of those factors (i.e., moderation) (Pearl 2009). For example, although several scholars argue that an IMF program will likely adversely affect the entire population of a country (i.e., mediation via austerity policies), they also hold that such a program will likely affect low-income families more than wealthy ones (i.e., moderation by families' socio-economic status) (Babb 2005). Nonetheless, whether this reasoning also implies that family wealth ranks as the most important moderator among the set of all potential moderators remains unclear.

In the following subsection, we discuss key factors moderating the causal relationship between IMF programs and child poverty. To do so, we use David Brady's (2019) framework that classifies theories on poverty into three types: behavioral, structural, and political. Behavioral theories, which represent mechanisms related to an individual's actions and choices, include the social-psychological foundation of family dynamics (McLanahan and Percheski 2008). Structural theories focus on macro-level demographic and economic constraints conditioning poverty outcomes. Political theories identify a population's collective choices as to how to distribute public resources and that distributions effect poverty (Beckfield 2018). Given that we adopt an inductive (data-driven) approach to identifying effect heterogeneity, we do not endorse any source of variation or hypothesis but instead offer a list of potential sources that produce heterogeneities.

2.1.1. *First heterogeneity source: Families' capabilities to protect their children*

Theoretically, the scholarly consensus is that IMF programs will likely affect low-income families and their children more than other social strata (Babb 2005; Stiglitz 2003). Low-income families may be the least able to weather economic stresses by (e.g., drawing on savings or support from others in their social networks) (Stuckler and Basu 2013). As low-income families tend to cluster in deprived neighborhoods, their social networks have fewer resources for them to mobilize in times of crisis.

Empirically, however, studies on economic turmoil—beyond the context of the IMF—show the plausibility of middle-class families being affected at least as severely as low-income families, if only because middle-class families have more to lose (Burgard and Kalousova 2015). According to this perspective, many low-income families might already be in long-term unemployment and social exclusion, so whether their government has agreed to implement an IMF program or not matters little to their already dire situation. Because most studies do not use household surveys, the literature has sparse empirical evidence on which families are most affected by IMF programs (Burgard and Kalousova 2015). While studies using micro data tend to examine the effects of macroeconomic changes on families within countries, they commonly focus on only one country at a time (Boss et al., 2016).

2.1.2. *Second heterogeneity source: demographic and economic conditions*

The effect of IMF programs on children may vary with a country's demographic and economic conditions (Brady 2019), such as labor market composition and the dependency ratio of children and the elderly to the working-age population. When the IMF demands pension cuts, older people will have to work more or rely on their adult children for support. If these adult children also conceive their own children, then these breadwinners must start supporting both older and younger family members. Thus families with a higher dependency ratio will likely experience a more pronounced adverse IMF effect than families with a lower ratio.

The impact of IMF programs will likely amplify geographical differences within countries. As IMF and government officials seek to streamline fiscal spending, they are likely to roll back public services in low-population density areas (Stuckler and Basu 2013), thereby creating the financial incentive to scale back services in rural areas, at the cost of eroding the quality of health and educational services. Thus, in the face of spending cuts, rural children are less likely to be vaccinated (Daoud and Reinsberg 2018) or educated (Rowden 2011). Despite theoretical reasons for believing that rural families are more adversely affected than urban ones, only a few studies have tested this difference empirically (Daoud et al., 2017).

2.1.3. *Third heterogeneity source: Policy, politics, and institutions*

As the effects of economic shocks depend on the political and institutional structures of a country (Beckfield 2018), the form of government affects the IMF's negotiation posture. Rickard and Caraway (2014) find that democratic countries with stronger labor unionization receive less intrusive IMF labor market conditions. In such countries, IMF programs likely affect working-class families and their children less than in nondemocratic and less unionized countries.

Public spending is not only an important causal pathway during an IMF program but also an important moderator before that IMF program has even been designed. Many public policy studies show that governments that spend more on health and education also have less inequality. Low inequality tends to give families more equal chances of raising healthy and educated children across all social classes than countries with high inequality (Schneider et al. 2018). However, high government expenditures may present especially attractive targets for IMF cuts—aimed at stabilizing a fiscal deficit. If such cuts were implemented, the adverse effect would likely be larger for a child living in high-spending countries than one living in low-spending ones. Although the IMF tends to discourage universal social protection, it has supported limited and targeted social policies mainly in low-income countries (IMF 2015). All these policies may interact in unexpected ways, leaving "which children have more to win" as an open empirical question.

3. Conceptual framework: Identification of IMF treatment effect on child poverty

While the previous section covered both substantive theories and previous findings on how IMF programs affect child poverty, in this section, we translate these substantive insights into a formal conceptual framework. This framework fulfills two purposes: First, it encodes our substantive assumptions about the causal system that regulates how IMF affects child poverty (structural causal assumptions). Second, these assumptions allow us to derive an estimator of this causal effect, which we formulate by using the potential outcomes framework (identification of causal effect). Thus, the effect we will identify is only causal in so far that these structural causal assumptions are fulfilled.

3.1. Structural causal assumptions

Fig. 2 illustrates the causal assumptions underlying our conceptual framework. Each node in our directed acyclic graph (DAG) represents a causal factor: An arrow starting at node *X* and pointing at *Y* indicates that *X* is a cause of *Y* (i.e., the graph is "directed"). A circle around a node means that the node is unobserved. A node is unobserved either because we lack measurements for it. However, we still include these non-modeled nodes in our DAG because they are central to our research design and our theoretical framework (Pearl 2009).

We observe children and their families at time *t*, a year after an IMF intervention (*t-1*), and capture political and structural characteristics one year before the intervention (*t-2*). The superscripts *P, S,* and *F* indicate political, structural, and family characteristics, respectively. These indicators cover the three previously discussed sources of heterogeneity (Brady 2019).

The exposure, AUSTERITY OF IMF PROGRAM, occurs at time *t-1*, and the outcome of interest, CHILD POVERTY$_t$, occurs the year after. The outcome is placed at the bottom of our DAG. Children's material well-being is affected by both their households' LIVING CONDITIONS$_t$ and by PARENT INCOME$_t$. We assume that while LIVING CONDITIONS$_t$ are not immediately affected by IMF programs in the year of implementation; they are affected by the state of the economy at ECONOMY$_{t-2}$. The node LIVING CONDITIONS captures slow-moving household characteristics, whereas PARENT INCOME$_t$ captures fast-moving changes in parents' resources.

As previously discussed, the quantity of interest is the total effect. Yet we suggested two non-exhaustive causal pathways mediating the total effect as a conceptual heuristic. These pathways are highlighted by the red arrows. Along the first pathway lie mechanisms that affect parents' income via the labor market: AUSTERITY OF IMF PROGRAM$_{t-1}$ → PUBLIC POLICY$_t$ → LABOR MARKET$_t$ → PARENT INCOME$_t$ → CHILD POVERTY$_t$. Along the second pathway lie mechanisms that directly affect children's level of deprivation via their contact with health and education services: AUSTERITY OF IMF PROGRAM$_{t-1}$ → PUBLIC SPENDING$_t$ → PUBLIC SERVICES$_t$ → CHILD POVERTY$_t$. Because our aim is to estimate the total causal effect, we do not control for any of the nodes along the two mediating pathways.

3.2. Identification of causal effects

To identify the total causal effect of IMF intervention on child poverty, we need to account for the selection process represented by the node SELECTION INTO IMF$_{t-1}$. The assumed causal system (our DAG) enables us to infer the conditions under which we can identify the effect of IMF treatment on child poverty. In this context, *identification* means that we have fulfilled a set of statistical conditions that enable us to calculate the quantity of interest (i.e., the total effect) from observational data. To define these conditions, we use the Neyman-Rubin potential outcomes framework (Imbens and Rubin 2015). We define IMF treatment as a binary country-level variable, $D_i = 1$, indicating the starting year of a program ($D_i = 0$ otherwise). Children living in the same country experience the same treatment. In addition to each child's observed poverty outcome Y_i, as given by the data, we assume that each child *i* has two potential outcomes. For each child, the assignment of treatment results in our observing one of two potential poverty outcomes: one when they live under an IMF program, $Y_i(D_i = 1)$, and one without such a program, $Y_i(D_i = 0)$.

Our analysis has two quantities of interest. The first is the "conditional average treatment effect" (CATE) that we calculate using a set of attributes *X* representing characteristics of a child, their household, and country. Formally, the CATE for a child *i* with covariates x_i is defined as $\tau(x_i) = E[Y(1) - Y(0)|X = x_i]$. In our case, *X* makes up the adjustment set of variables that we use to control for confounding. Our method estimates CATE by imputing potential outcomes. This method finds comparable *groups* of children, using children's observed covariates x_i.

We are interested in learning how structural, political, and family-level variables moderate the IMF program's impact on groups of children (Brady 2019). To this end, we study CATEs with respect to variables in each of these categories, defined as follows: $E[\tau(X) \mid F = f_i]$ where *F* is a set of family-level variables; $E[\tau(X) \mid S = s_i]$, with *S* being a set of structural variables; and $E[\tau(X) \mid P = p_i]$, with *P* representing political variables. A combined CATE is a function, $\tau(f_i, p_i, s_i) = E[\tau(X) \mid F = f_i, P = p_i, S = s_i]$, capturing how combinations of variables from all categories moderate the strength of the treatment effect. These covariates are subsets of the adjustment set *X*, extracted from our DAG (mathematically, $F, S, P \subseteq X$). In the language of interactions, this function, $\tau(f_i, p_i, s_i)$, captures all possible two-, three-, four-, and *n*-th level interactions between the IMF program, child poverty, and the moderators. In this work, we first estimate the granular-level CATE, $\tau(x_i)$, for each child *i,* after which—in the post-estimation stage—we can calculate the coarser CATE, $\tau(f_i, p_i, s_i)$. Our method section provides a detailed explanation of this estimation procedure.

The second quantity of interest is the "average treatment effect" (ATE) of economic shocks caused by IMF programs, $E[\tau(X)]$. This quantity captures the effect across all groups of children. We also quantify the "average treatment effect on the treated" (ATT), $E[\tau(X)|D = 1]$ and the "average treatment effect on the control" (ATC), $E[\tau(X)|D = 0]$.

Our identification strategy for ATE, ATT, ATC, and CATE relies on our ability to eliminate (block) confounding bias by using an adjustment set. This set consists of covariates, *X*, is considered a valid adjustment set if it fulfills the backdoor criterion—evaluated by scanning off our DAG (Pearl 2009). A statistical model fulfills the backdoor criterion if it blocks all backdoor paths. A "backdoor path" is a path consisting of a sequence of variables starting with an arrow pointing to the outcome node and ending with an arrow pointing to the treatment node (e.g., Y ← **C** → D, where *C* represents one or several confounding variables) (Pearl 2009). A backdoor path is considered blocked if a statistical model *controls* for one or more variables along the path (Pearl 2009). If the backdoor criterion holds for a set of variables *X*, this set is a valid adjustment set, and confounding bias will be eliminated by controlling for *X*. Statistically, satisfying this criterion means that the potential outcomes are independent of the treatment conditional on the adjustment set, denoted as $Y(1), Y(0) \perp\!\!\!\perp D|X$. This independence assumption is called "conditional ignorability" with respect to *X*.

Our DAG, depicted in Fig. 2, has four backdoor paths that our identification strategy has to block. We can block three of these paths by controlling for observed variables on the path. We block the first backdoor path, selection into IMF$_{t-1}$ ← polity$_{t-2}$ ← political will$_{t-2}$ → child poverty$_{t}$, by conditioning on the observed polity$_{t-2}$ node. We block the second backdoor path, selection into IMF$_{t-1}$ ← public spending$_{t-2}$ ← political will$_{t-2}$ → child poverty$_{t}$, by conditioning on the observed public spending$_{t-2}$ node. We block the third backdoor path, Selection into imf$_{t-1}$ ← economy$_{t-2}$ → living conditions$_{t}$ → child poverty$_{t}$, by controlling for economy$_{t-2}$ and living conditions$_{t}$. Because we control economy$_{t-2}$, we ensure that our statistical estimate distinguishes the IMF-program effect from other forms of economic shocks.

However, the fourth backdoor path, selection into IMF$_{t-1}$ ← political will$_{t-2}$ → child poverty$_{t}$, challenges any identification strategy that aims at controlling for only observed variables. As we cannot observe political will$_{t-2}$ in our data, we cannot directly block this fourth path. Yet we can block it indirectly. We block this backdoor path by using a Heckman selection model to account for governments' willingness to collaborate with the IMF (Heckman 1979). An important advantage of the Heckman model (over, e.g., instrumental variable methods) is that it produces a proxy variable, *U*, that we can then test in the ML model for its importance and estimate CATE. The intuition underlying the Heckman model is that it produces a proxy for the unobserved factor, political will$_{t-2}$. (Technically, this proxy is called "the inverse Mills ratio" which we denote as *U*.) The intuition of the Heckman model is related to a propensity score model because the model's estimates are the residual of a propensity model. This residual, *U*, captures the unobserved factor Political Will$_{t-2}$. The Method delineates the Heckman procedure, with additional details in Appendix C.

After we obtain a proxy of political will$_{t-2}$ (i.e., *U*), we can block the last backdoor path. Our adjustment set now satisfies the backdoor criterion, meaning that *X* and *U* are jointly a valid adjustment set for the effect of *D* on *Y*. By satisfying this criterion, the potential outcomes of children are independent of the IMF treatment conditional on the adjustment set (i.e., $Y(1), Y(0) \perp\!\!\!\perp D|X, U$), with respect to *X* and *U*. This ignorability assumption enables us to identify the causal effect (i.e., a difference between potential outcomes) from our observational sample (i.e., the observed difference between the treated and the control groups). This identification implies the following, formal relationship,

$$E[Y(1) - Y(0)|X, U] = E[Y|D = 1, X, U] \; - \; E[Y|D = 0, X, U]$$

in sum, the conceptual framework encoded in our DAG explicitly states our causal assumptions. This DAG enables us to connect the three aspects of our research design: substantive theories (section 2), empirical material (section 4), and causal estimates (section 5).

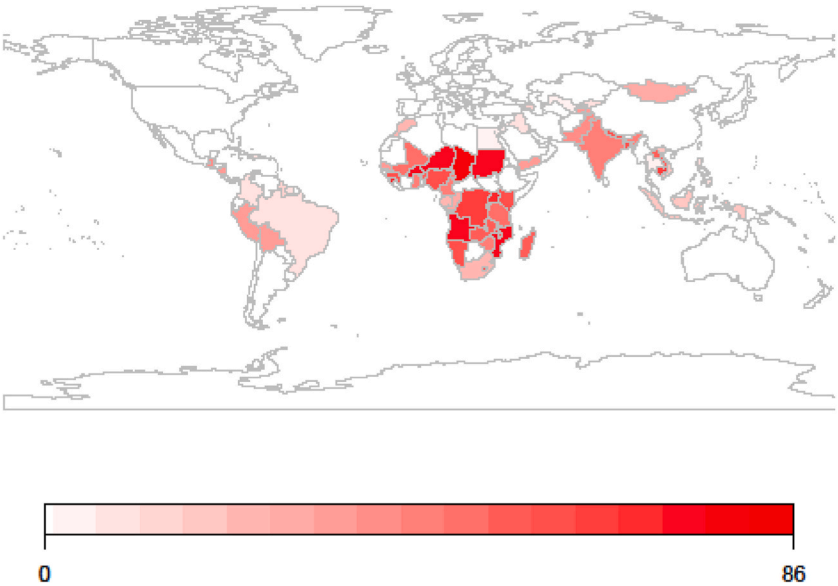

Fig. 1. *World distribution of child poverty.* Notes: We calculated country averages from the DHS and MICS micro data. Authors' calculations of baseline child poverty for each of the 67 countries, is included in the sample. White-colored countries are not included in the sample.

4. Data and method

4.1. Data

4.1.1. Micro: families and their children's living condition

We use household data from the Demographic and Health Survey (DHS) and the Multiple Indicator Cluster Survey (MICS), both cross-sectional and nationally representative surveys conducted mostly in low- and middle-income countries. The response rate for these studies, which use face-to-face interviews, tends to exceed 90 percent, and sample sizes range from 4000 to 30,000 households. One advantage of these two data sources is that the interviewers are rigorously trained, ensuring quality; another advantage is that the data collection procedure is standardized across countries and years. MICS and DHS also collaborate to ensure that their surveys are comparable across core modules. This comparability allows us to pool the two data sets.

We choose data from countries that were sampled within plus-or-minus five years of 2000, just before the large health vaccination programs launched by the Gates foundation (Lancet 2009). Our pooled data spans from 1995 through 2005. Our sample covers 1,940, 734 children in 567,344 families in 67 countries, primarily in low- and middle-income countries. This sample is representative of about 2.8 billion households or roughly 50 percent of the world's population. Fig. 1 presents a world map of the sample's geographical distribution of child poverty.

We extract several family- and child-level variables from the pooled DHS and MICS samples. In this and the following subsection, we signify the nodes of our DAG with small capitals, and we use italics for the statistical variables that measure these nodes. We measure each node by one or multiple variables. Appendix A shows additional details (e.g., sources) of the statistical variables.

To define CHILD POVERTY, we use a deprivation-based approach. This type of approach measures children's direct access to material resources, rather than measuring household income, which only indirectly captures deprivation. We use the Bristol method because it focuses on child poverty and is designed to measure children's living conditions globally (Gordon et al., 2003). The sample size and the timing of our samples are in Appendix A. This method measures *child poverty* in seven (binary) deprivation dimensions: water, malnutrition, shelter, sanitation, health, information, and education. Table 1 presents these dimensions, their definitions, and sample frequencies. The Bristol method defines a child that is deprived of two or more of these seven dimensions to be poor (Gordon et al., 2003). This outcome operationalizes CHILD POVERTY in our DAG.

While the Bristol method is the most appropriate measure of child poverty in low and middle-income countries, there are some

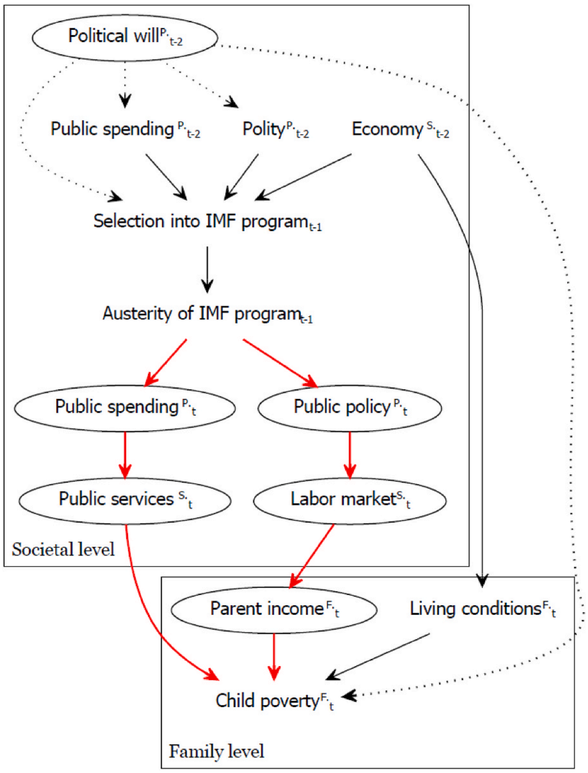

Fig. 2. *Causal assumptions pictured in a directed acyclic graph (DAG).* Notes: Each node represents a causal factor. (See method section for the empirical measure of each concept.) A directed arrow shows a causal link. The directions of the arrows are acyclic, meaning that each causal path starting from any given node cannot terminate at the same node. Dashed lines show that the causal path is unobserved; a circle around a node shows that the node is unobserved. The superscripts, *P., S.,* and *F.* indicate political, structural, and family characteristics, respectively.

Table 1
Definition of child poverty and deprivations.

Child Deprivation	*n* (age-filtered sample)	Proportion of *n* deprived
Water: Children who only have access to surface water (e.g., rivers) for drinking or who lived in households where the nearest source of water was more than 15 min away. Children <18 years old.	1,941,734	0.24
Malnutrition: Children whose heights and weights for their age were more than −3 standard deviations below the median of the international reference, that is, severe anthropometric failure. Children <5 years old.	559,418	0.15
Shelter: Children in dwellings with more than five people per room and/or with no flooring material. Children <18 years old.	1,926,435	0.51
Sanitation: Children who had no access to a toilet of any kind in the vicinity of their dwelling, that is, no private or communal toilets or latrines. Children <18 years old.	1,940,599	0.28
Health: Children who has not been immunized against diseases or young children who had a recent illness involving diarrhea and had not received any medical advice or treatment. [polio, measles, DPT (diphtheria, pertussis, and tetanus), tuberculosis (Bacillus Calmette-Guérin) recommended by the WHO (cf. 2013)]. Children <5 years old	559,418	0.22
Information: Children with no access to radio, television, telephone, or newspaper at home. Children 3–17 years old.	1,837,578	0.18
Education: Children who had never been to school and were not currently attending school, in other words, no professional education of any kind. Children 7–17 years old.	1,150,711	0.14
Absolute poverty: Children who suffers from two or more deprivations, as defined above. Children <18 years old.	1,941,734	0.48

Notes: "Proportion of *n* deprived" signifies the proportion of deprived children relative to the age-filtered sample, not relative to the total sample.

concerns regarding the impact of random noise on the validity of that measure (Loken and Gelman 2017; Catalán and Héctor, 2019; Catalán et al., 2020). To account for potential measurement error, we conduct a sensitivity analysis (Smith et al. 2021), presented in Appendix D.

We measure children's LIVING CONDITIONS$_t$ using eight variables. We use the *head of household's educational attainment.* This person tends to be the individual leading the household's family affairs, for example, the mother, father, or oldest person in the family. Educational attainment is a three-valued ordinal variable: no education, only primary, and at least secondary. We use the DHS and MICS asset index to measure *family material wealth.* Surveyors sample the assets of a household by adding the family's ownership of cars, bicycles, televisions, and other valuable commodities, and apply principal component analysis to these asset indicators to construct a new index. The ranked position of each family in the national distribution of this index provides a proxy for the family's material wealth.

We record the *location* of the household (urban or rural), and count the *number of adults* dwelling in the household, with more adults suggesting an increased likelihood of additional breadwinners. We count the *number of children* because more children increase a family's economic burden. As younger children are more dependent on their parents than older ones, we include variables for the *child's age* (up to 17 years) and the *child's sex*, to evaluate if boys or girls are affected differently. Moreover, to adjust for time variations, we control for the *year of the interview*.

4.1.2. Macro: economy, society, and public policy

We operationalize the node IMF AUSTERITY PROGRAMS$_{t\text{-}1}$ as the treatment, measured as a binary statistical variable, *IMF program*. As our conceptual framework shows, while selection into IMF program$_{t\text{-}1}$ captures the economic context motivating governments to select an IMF program, IMF AUSTERITY PROGRAMS$_{t\text{-}1}$ represents the government's implementation of such a program. Statistically, this process (i.e., selection into IMF program$_{t\text{-}1}$) represents our *first-stage Heckman model.*

Our model includes a set of macro measures to block the backdoor paths in our DAG. The node ECONOMY$_{t\text{-}2}$ captures structural and demographic variables, and we measure it with eight statistical variables (Vreeland 2003). We use *GDP growth* to measure how fast countries are growing, as fast-growing countries are less likely to ask for IMF support than slow-growing ones. Moreover, we include an indicator for *negative growth*, because countries with a shrinking economy are more likely to ask for IMF assistance. Beyond economic growth, we control for Log *GDP per capita,* because both the rate of change and the level of economic activity matter for the risk of economic turmoil. We control for *current account balance*, because the higher the fiscal imbalance, the more likely the country is to ask for IMF help. In addition, because countries well integrated into international trade have a lower probability of asking for IMF support than those with low integration, we condition on *trade* and *economic globalization index*. We include an indicator for *high inflation* because rapidly increasing prices for common goods and services suggest structural economic imbalances. We control for the *dependency ratio*, which measures the size of the population not in the labor force (i.e., ages 0 to 14 and 65 and older), divided by the size of those in the labor force. A higher ratio suggests a higher demand for public services and thus more financial stress.

PUBLIC SPENDING$_{t\text{-}2}$ consists of three measures. We measure spending in three ways: *education spending* as a percentage of total government spending, *health spending* as a percentage of total government spending, and *total government spending,* as a percentage of GDP. These measures capture how much resources governments allocate to public services. Lower spending suggests an increased likelihood of children falling into poverty.

POLITY$_{t\text{-}2}$ comprises political and institutional measures. We include an indicator for *democracy*, because autocratic regimes can solicit IMF support with lower public relations costs than countries that are more democratic. We control for *political terror* as a proxy for how likely social movements are to mobilize resistance to IMF policies through protests, without the risk of political retaliation (Stiglitz 2003). To measure how effective a government is in implementing IMF policies, we condition on *government efficiency*. We also account for *corruption*, because the more corrupt a government is, the more likely it is to steal public resources. We also control for both

Table 2
Descriptive demographics stratified by IMF treatment status.

	IMF = 0	IMF = 1	p-test
Sample size (children)	985805	955929	
Health deprivation	0.14 (0.35)	0.13 (0.34)	<0.001
Water deprivation	0.22 (0.42)	0.27 (0.44)	<0.001
Malnutrition deprivation	0.13 (0.33)	0.09 (0.29)	<0.001
Education deprivation	0.11 (0.32)	0.13 (0.34)	<0.001
Shelter deprivation	0.49 (0.50)	0.54 (0.50)	<0.001
Sanitation deprivation	0.29 (0.45)	0.28 (0.45)	<0.001
Information deprivation	0.17 (0.38)	0.19 (0.39)	<0.001
Child age	8.27 (5.06)	8.08 (5.07)	<0.001
Child sex = Male (%)	501117 (50.8)	483457 (50.6)	<0.001
Urban household = rural (%)	613439 (62.2)	603594 (63.1)	<0.001
Family wealth (%)			<0.001
1	256662 (26.0)	223967 (23.4)	
2	192713 (19.5)	192630 (20.2)	
3	187835 (19.1)	190872 (20.0)	
4	170723 (17.3)	173794 (18.2)	
5	177872 (18.0)	174666 (18.3)	
Nr of children per household	3.77 (2.44)	4.04 (3.21)	<0.001
Nr of adults per household	2.91 (1.87)	2.97 (2.42)	<0.001
Head of household's education (%)			<0.001
No education	314602 (31.9)	273742 (28.6)	
Primary	360369 (36.6)	330722 (34.6)	
Secondary and above	310834 (31.5)	351465 (36.8)	
Year of household interview	2003.49 (3.68)	2003.35 (2.72)	<0.001
Health_spending	8.32 (5.61)	10.03 (3.51)	<0.001
Democracy	5.12 (3.23)	5.89 (2.01)	<0.001
Trade	66.07 (39.99)	56.73 (27.38)	<0.001
Government's expenses balance	−2.29 (6.31)	−3.40 (2.79)	<0.001
Economic development	6.51 (0.86)	6.09 (0.85)	<0.001
War	0.13 (0.80)	0.21 (0.90)	<0.001
Dependency ratio	76.07 (11.95)	80.31 (12.89)	<0.001
Negative growth	6.99 (6.82)	5.04 (2.58)	<0.001
Inflation	0.10 (0.29)	0.05 (0.22)	<0.001
Political will	−0.42 (0.49)	0.27 (0.20)	<0.001
Government's education spending	12.80 (5.97)	15.15 (5.06)	<0.001
Economic globalization	41.26 (13.59)	39.88 (13.12)	<0.001
Government effectivness	−0.53 (0.60)	−0.66 (0.36)	<0.001
Political terror	3.64 (0.91)	3.22 (0.97)	<0.001
Corruption	−0.66 (0.47)	−0.80 (0.46)	<0.001
Child labor law	0.51 (0.50)	0.62 (0.49)	<0.001
Government's public spending	24.35 (8.44)	22.12 (8.22)	<0.001
Countries with IMF	57.84 (8.84)	61.09 (6.70)	<0.001
Government's UN vote with G7	0.62 (0.04)	0.63 (0.05)	<0.001

Notes: The p-values pertains to a χ^2-test of groupwise comparison. Mean values are presented in each column with standard deviations in paratheses.

international war and *civil conflict*, because the IMF disengages from any country entangled in open hostilities. To capture whether children are protected from child labor, we include an indicator for *minimum age labor law*. During economic turmoil, families in poor societies are more likely to take their children out of school and place them in paid labor.

4.2. A methodological recap of machine learning for causal inference

To estimate the impact of IMF programs on child poverty, we combine covariate adjustment using *the generalized random forest* (GRF) (Athey et al., 2019). This subsection discusses the benefits and limitations of ML in general (Daoud and Dubhashi 2021, 2023; Kino et al., 2021), and the GRF in particular.

This subsection presents an overview of what ML is and how it works. ML is the computer science sub-discipline that studies the question of how to build algorithms that "learn from experience" without explicit programming (Hastie et al. 2009). An algorithm is a set of instructions defining how this learning takes place, and no explicit programming means that the algorithm has no predefined instructions for how the statistical relationship between an outcome (Y) and a set of variables (X) appears. "Learning" in ML means that the algorithm improves its identification of this relationship by predicting the outcome from the set of variables. The algorithm improves through experience, which entails exposure to increasing amounts of data. The better the algorithm's predictions for new data, the more reliable the algorithm is.

The ML toolkit contains hundreds of different algorithms (Hastie et al., 2009), several of which work well for policy evaluation (Künzel et al., 2018). We use one of these algorithms, the GRF, for our analyses (Athey et al., 2019). The GRF belongs to the family of regression estimators, which are models that use a set of covariates to predict the outcome (Hastie et al., 2009). However, the GRF has certain desirable properties that ordinary least squares (OLS) and other commonly used methods lack. One main property is that the GRF is designed to search through many combinations of effect heterogeneity (interactions) and present those that are potentially important for further social theorizing (Athey et al., 2019).

This design makes the GRF particularly valuable for analyzing the research questions and data of this paper. As the two million children in our global data set live in different families, countries, and regions, they will likely react differently to IMF policies. Instead of manually specifying, testing, and searching through many interaction models, with the risk of missing some essential interactions, the GRF conducts this process for us.

The GRF obtains its desired properties by implementing nonparametric estimation for causal effects, building on the "random forest" algorithm (Breiman 2001), which grows a collection of regression trees to predict an outcome (in our case, child poverty). An ML forest comprises a great many trees. A single tree consists of splitting rules applied to a collection of covariates that jointly predict the outcome. A "splitting rule" divides a covariate into two smaller pieces. The algorithm selects the splitting rule that optimizes the variance explained in that split. A tree combines several splitting rules and applies them to several covariates in a sequence, thereby creating an interaction.

For example, a tree with two sequences could consist of the following variables and splitting rules: In the first sequence, the algorithm applies the splitting rule "*child age* > 5" to divide the covariate child age into a first part (ranging from 6 to 18) and a second part (ranging from 0 to 5). In the second sequence, for the first part, the algorithm applies the splitting rule "*family wealth* > average," while for the second part, it applies the rule "*public spending* > the first quintile." The more sequences a tree has, the taller (i.e., more complex) it is. The random forest algorithm— called "random" because it randomly samples both variables and cases to find an optimal splitting rule—automatically finds the optimal depth.

For statistical efficiency, the random forest grows a large number of trees, with each tree predicting the same outcome. For example, if a forest consists of 1000 trees for predicting child poverty, then each tree produces a potential prediction for each child in our data set. The forest thus has 1000 predictions for that one child. By calculating the average of all those predictions, the forest produces one estimate for each child. Because the random forest averages over a great many trees, it diminishes its reliance on any single combination of splitting rules. For our analysis, we grow forests of 4000–8000 trees, varying the forest size to test the sensitivity of our results.

The GRF has the unique capability of elevating the status of the treatment variable over the other covariates, a capability that the simple random forest (i.e., not generalized) lacks. For each splitting rule applied to a covariate, the GRF also balances cases from both the treated and the control groups (Athey et al., 2019). In this way, conditional on the covariates, the average difference between the predicted outcomes of these two groups is the estimated causal effect, $\tau(x_i)$. This effect is the CATE.

One strength of the random-forest algorithm is that it has a potentially limitless capacity to fit data: it is a universal function approximator (Breiman 2001). This algorithm finds the optimal function by optimizing the predictive accuracy of new data by trading off bias and variance by regularizing the growth and depth of trees (Hastie et al., 2009). "Regularization" is a procedure present in all ML algorithms that penalizes model complexity. On the one hand, a smaller random forest, which has fewer and shorter trees, tends to generalize better in predicting outcomes on new data. On the other hand, such a forest runs the risk of having a larger bias and thus generalizing poorly. Every random forest, including the GRF, balances this tradeoff when the forest fits well with both training data and new data. Regularization is the key mechanism that identifies this balance.

4.2.1. Four advantages of ML for causal inference

The question is how an ML approach in general and the GRF in particular, provides a more comprehensive estimation strategy for our causal analysis than more widely used methods. At the core of these causally adapted algorithms lies the assumption of an experimental (ignorability by randomization) or quasi-experimental design (conditional ignorability). When combined with causal-inference techniques, ML offers four capabilities that traditional estimation techniques lack (Athey and Imbens 2017; Künzel et al., 2018).

First, as Section 3.2 discusses, GRF imputes the treatment effect for each child, that is $\tau(x_i)$, based on the children's group similarity as defined by their covariates x_i (Künzel et al., 2018). As long as the treated and control populations are comparable, imputing the treatment effect provides a flexible method for evaluating the impact of IMF policies on child poverty (Johansson et al., 2019). These imputed values make the GRF an interpretable model, as these values are the causal estimates of interest. When summing all these values across groups of children, these values converge to the (conditional) average treatment effect of IMF programs (Lipton 2017). These imputed values constitute a necessary step for the second capability.

Second, the GRF inductively identifies effect heterogeneity instead of making the researcher define it deductively. Whereas more common methods require researchers to explicitly specify potential interactions between moderating variables and treatment, the GRF does not, because it can search through a vast number of interactions and present the most noticeable ones. For example, in a linear model, if researchers hypothesized that the effect of IMF programs (*D*) varied with children's age (*X*), they would specify an interaction *D*X* in the model for the observed outcome: $Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 D \bullet X + e$. They would then test this model against the data to verify whether β_3 is significantly different from zero.

The GRF finds interactive-variable threshold rules that capture as much of the variance (heterogeneity) in the causal effect as possible (Athey et al., 2019). The algorithm probes many possible combinations of variable interactions, records those interactions that contain substantial effect heterogeneity, and demotes those that do not. As it systematically scouts for all the data, it has the potential to capture all intricate non-linear interactions between variables. It will record both conspicuous patterns and complex patterns hidden in remote regions of the joint distribution. For example, although the sex of a child might not be a relevant moderator when evaluating our entire data, it might still be statistically and substantively significant for explaining IMF treatment heterogeneity among wealthy households (Daoud and Puaca 2011). Therefore, based on the GRF, our CATE quantity (defined in section 3) encodes a vast number of combinations in effect heterogeneity that IMF programs induce in child poverty,

$$\tau(f_i, p_i, s_i) = E[(\tau(X) | F = f_i, P = p_i, S = s_i]$$

This equation shows that the IMF effect varies with three heterogeneity sources (F for family, P for political, and S for structural variables). These sources correspond to Brady's (2019) poverty framework, which we synthesized in section 2. This CATE quantifies all possible interactions in our data. As our covariate data contains 25 variables, the function $\tau(f_i, p_i, s_i)$ can slice the estimated treatment effect $\tau(X)$ in all combinations of these variables and their strata. The GRF provide summary statistics of the effect-heterogeneity importance for each single variable (Hastie et al., 2009). We rely on a type of statistic that measures the number of times the algorithm used each variable. The more the algorithm uses a variable, the more important this variable is in accounting for effect heterogeneity, $\tau(f_i, p_i, s_i)$. Therefore, we use this CATE logic and this list to disentangle effect heterogeneity. Appendix B described additional mechanics of the variable importance metric.

Third, while more common methods force researchers to commit to one specific statistical relationship—often a linear parametric model and thereby imposing this linearity on the data (deductive approach)—between the treatment, outcome, and controls, the GRF approximates its functional form to fit the data non-parametrically (inductive approach). This inductive approach gives us the advantage of not committing to how the treatment is statistically related to the outcome, conditional on covariates. The algorithm will find the appropriate statistical relationship between these variables in a data-driven way. This inductive-approximation property follows from the random forest algorithm is a flexible estimator (Breiman 2001), and this property enables us to relax what Abbott (1988) calls the dominance of general linear reality in sociology.

Fourth, ML methods are more robust than common parametric methods. Common methods are prone to overfitting because they tend to use all the data to estimate the model and evaluate its predictive power (Watts 2014). "Overfitting" means that a model becomes so expressive that while it will fit a given sample perfectly, it will predict poorly for new data. In contrast, as discussed previously, the estimates of the GRF are more robust because the GRF relies on regularization and out-of-sample evaluation (Athey et al., 2019). Additionally, regularization and out-of-sample evaluation provide another benefit; they guard against the influence of outliers, p-hacking, and cherry-picking results (Watts 2014). Although sample splitting does not completely eradicate researcher discretion, it renders such bias less influential.

4.3. *Estimation*

Fig. 3 summarizes our estimation strategy. The first stage uses the Heckman procedure to obtain a proxy of governments' political will to implement an IMF program—Appendix E presents those results. This procedure comprises first estimating a probit model to estimate the propensity that each government enrolls in an IMF program. This model regresses all the country-level variables specified in section 4.1. To align this probit model with our DAG, we measure these variables two years, $t-2$, prior to measuring child poverty at t. After estimating our probit model, we use it to calculate the inverse Mills ratio, denoted by λ. All the equations are delineated in Appendix C. In the second stage of our estimation strategy, we use λ in the same way as the other covariates when we train a GRF to estimate the CATE, $\tau(x_i)$. As Fig. 3 shows, this GRF imputes $\tau(x_i)$ for each child in our data.

In the third stage, we stratify the imputed effects according to the variables that the algorithm suggests are important in explaining the treatment effect heterogeneity. Our stratification analysis can then focus on those variables. As previously mentioned, while the GRF estimates the IMF impact, it ranks the importance of each variable in accounting for effect heterogeneity. When using parametric models, such as OLS regression, one usually inspects the coefficients to interpret what covariates were most important for explaining the outcome. In contrast, nonparametric models such as random forests have no fixed set of parameters for inspecting. Therefore, we use a variable-importance statistic that records the proportion of times for which the GRF selected a variable for growing trees. The higher the proportion, the more important that variable is in moderating effect heterogeneity (Athey et al., 2019). Through this ranking, the GRF algorithm thus signals to us which variables are likely containers of treatment heterogeneity.

4.4. *Estimating clustered standard errors*

The original random forest algorithms estimate confidence and prediction intervals through resampling, which fits each tree to a different random subset of the data (Breiman 2001). The algorithm then calculates the variance of these predictions. Using "half--sampling" (i.e., half the total sample), the GRF implements a variant of the classical bootstrap, referred to as "bootstrap of little bags" (Athey et al., 2019). In this variant, the algorithm divides the set of trees into small sets (bags) and assigns each bag a random subset of training data (i.e., each subset contains half the total sample). First, the algorithm fits each tree in the same bag to the samples assigned to that bag. Second, it estimates the total variance by computing the average squared deviation between individual trees and the full forest, and then, subtracting the average within-bag variance (Athey et al., 2019).

Our data is clustered in two key analytical units of analyses, and consequently, our ATE estimate requires clustered standard errors. As discussed previously, the first unit of analysis is the child level, where the poverty outcome is measured; the second unit of analysis is countries, where the treatment assignment occurs.[1] In other words, children's poverty outcomes are nested within their respective

[1] Observed, that there is a third unit of analysis, which is households. However, as household variables are secondary to our analysis and that our GRF does not estimate parameters for these variable in the traditional sense, they do not require statistical clustering.

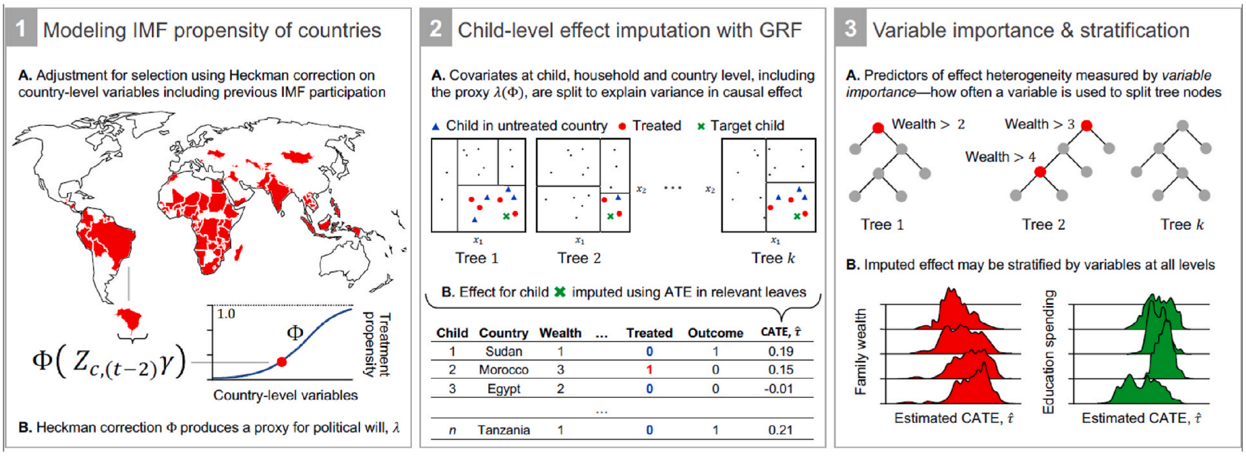

Fig. 3. *Overview of the estimation workflow*. Notes: In the first analytical stage, we impute a proxy for political will, λ, using a probit Heckman selection model based on country-level variables alone. We use this proxy to account for unobserved confounders in the estimation. In the second stage, we fit a GRF to impute the conditional average treatment effect of IMF interventions on child poverty. For each child, the imputed effect is the average of the predictions of a number of regression trees. In each tree, the prediction for a target child is the difference in mean outcome between treated and untreated children placed in the same leaf as the target. Leaves are constructed by recursively splitting the covariate space to maximize the variance in estimated treatment effect across leaves. In the third stage, we inspect the GRF model by examining variables that were important in explaining heterogeneity and by stratifying and averaging its predictions.

countries where IMF policy occurs. Statistically accounting for this clustering is critical because the statistical efficiency of the data is merely equal to the number of countries in our sample and not equal to the number of children (Snijders and Bosker 1999). In traditional regression models, clustering is handled either by applying clustered standard errors or fitting a multilevel model (Abadie et al., 2020). Such models enable a myriad of other estimation possibilities (e.g., varying coefficients by country), but for our estimation purposes, clustered standard errors are sufficient.

Our GRF produces robust clustered standard errors for the ATE estimate to handle clustering. To estimate this variance, the GRF uses the cluster-robust bootstrap (Cameron et al. 2008), which applies the half-sampling procedure at the country (cluster) level instead of the individual level (Athey and Wager 2019). In other words, the algorithm fits each bag of trees to individuals, all of whom belong to a random half of the countries (Tibshirani et al. 2020).

5. Results

5.1. The average impact of IMF programs on child poverty

Our GRF estimates an ATE of IMF programs on child poverty. Fig. 4 shows the effect: IMF programs increase children's risk of poverty by 14 percentage points. The 95% confidence interval (CI) of the ATE spans the probability scale of 0.03–0.24.

Policymakers frequently want to know the effect of a program on those who actually receive it. The average treatment effect on the treated (ATT), $E[\tau(x_i)|\, D=1]$, is larger than the ATE. In countries with an actual IMF program (i.e., implying observed in our sample), children's risk of poverty increases by 24 percentage points (95% CI: 0.02–0.46). Given the fewer number of countries with IMF programs ($n_1 = 28$) than for the full sample of countries ($n = 67$), the CI of ATT is wider than the CI of ATE. The average treatment effect on the control group (ATC), $E[\tau(x_i)|\, D=0]$, is still adverse, even though the point estimate is smaller than the ATE and the 95% CI overlaps zero. Thus, for countries not implementing IMF programs ($n_0 = 39$), our model suggests that if they were to implement such programs, these programs would increase children's risk of poverty by 7 percentage points (95% CI: −0.05–0.18).

We find a great deal of heterogeneity in IMF effects. Fig. 5 shows the estimated effect for each child's $\tau(x_i)$ in our sample (n = 1,940,734). Children of the same group (defined as having the same covariate values x_i) have identical effects. The blue line marks the ATE. These individual-level effects range from a beneficial impact—i.e., a reduced probability of poverty by 10 percentage points—to an increased probability of poverty by 45 percentage points. About one million of the children in our sample have a larger effect than the ATE value. The next subsection investigates which factors—family, structural, or political—account most for this impact variation.

5.2. Effect heterogeneity

A key feature of ML is that its algorithms help identify subgroups that react differently to policy changes. As discussed in the method section, the GRF algorithm is specifically designed for quantifying such heterogeneity (Athey et al., 2019).

While Fig. 5 shows the variation in impact among all children, Fig. 6 shows the variation within and between countries. The estimated IMF effect in the most adversely affected countries—such as Lao People's Democratic Republic, Sudan, and Tanzania—is approximately twice as high as the ATE (blue line). The effect size in these countries is about 23 percentage points. In contrast, children

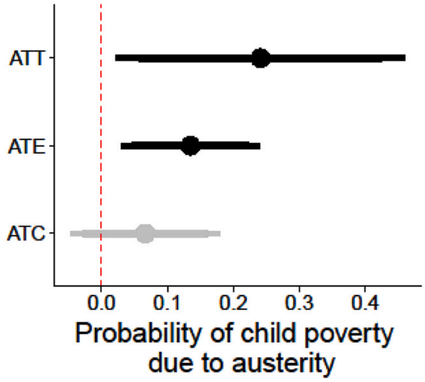

Fig. 4. *Quantities of interest.* Notes: The figure shows the impact of IMF programs on child poverty. The same GRF estimated all three point estimates, i.e., the average treatment effect on the treated (ATT), $E[\tau(x_i)| D = 1]$; the average treatment effect (ATE), $E[\tau(x_i)]$ and; the average treatment effect on the control (ATC), $E[\tau(x_i)| D = 0]$. The 95% CI are based on robust clustered standard errors computed from the empirical predictive distribution. The results are based on data described in Table 2 and using a GRF as described in the Methods. Note that because the GRF is a non-parametric model, only an estimate of the treatment effect is provided but none for the controls (unlike a parametric model).

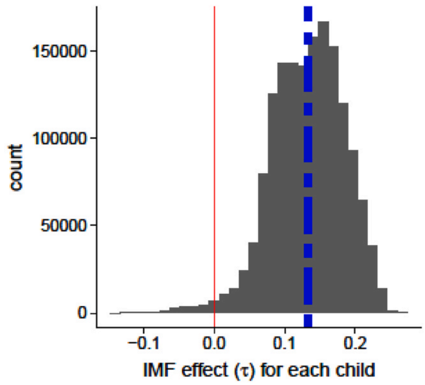

Fig. 5. *Histogram of individual-level treatment effect – IMF effect heterogeneity*. Note: The histogram displays the estimated IMF effect for each child's $\tau(x_i)$ in our sample (n = 1,940,734). The blue line signifies the ATE of 0.14. The x-axis displays the change in probability of poverty, where positive values represent an increase in poverty. The results are based on data described in Table 2 and using a GRF as described in the Methods. Note that because the GRF is a non-parametric model, only an estimate of the treatment effect is provided but none for the controls (unlike a parametric model). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

living in Iraq, Lesotho, and Mongolia have a much lower IMF probability of falling into poverty, with an effect size of about 5 percentage points. Thus, this density plot indicates the extent to which macro factors (structural and political) drive effect heterogeneity.

Fig. 6 also reveals substantial within-country variations. Given that effect heterogeneity between countries arises mainly from macro-factor differences, within-country variations mainly reflect differences in families' ability to cope with IMF programs. In countries with large variations in impact, families have a wide range of experiences with how IMF programs affect their children, whereas families' experiences are more homogenous across social strata in countries with less variation.

To quantify how much micro (family) or macro (structural and political) factors can account for this variation, we partitioned the variance in impact. Using a three-level multilevel random intercept null model—children nested in families and families nested in countries—we find that 47 percent of the total impact variance comes from between-country differences, with 53 percent coming from differences within and between families. Thus family-level and country-level characteristics are about equally important in accounting for how economic shocks affect families and their children.

The GRF algorithm inductively identifies and ranks which macro and micro factors moderate most of the IMF effect heterogeneity. Fig. 7 shows this ranking.

Our GRF identifies one PUBLIC SPENDING$_{t\text{-}2}$ measure (government education expenses) and one LIVING CONDITION$_t$ measure (family wealth) both of which moderate most of the effect heterogeneity. Family wealth has a variable importance value of 0.11, indicating that the GRF selected this variable in 11 percent of all tree branches. Fig. 8 shows CATE by family wealth quintile. The group averages of these five distributions show that children in the wealthiest quintile have the lowest probability of falling into poverty: 9 percentage points (95% CI: 0.02–0.16). Children in the three middle quintiles of family income tend to experience an IMF effect as harsh as those in the poorest quintile. The first quintile, which is the poorest, has an effect estimate of 11 percentage points (95% CI: 0.08–0.14),

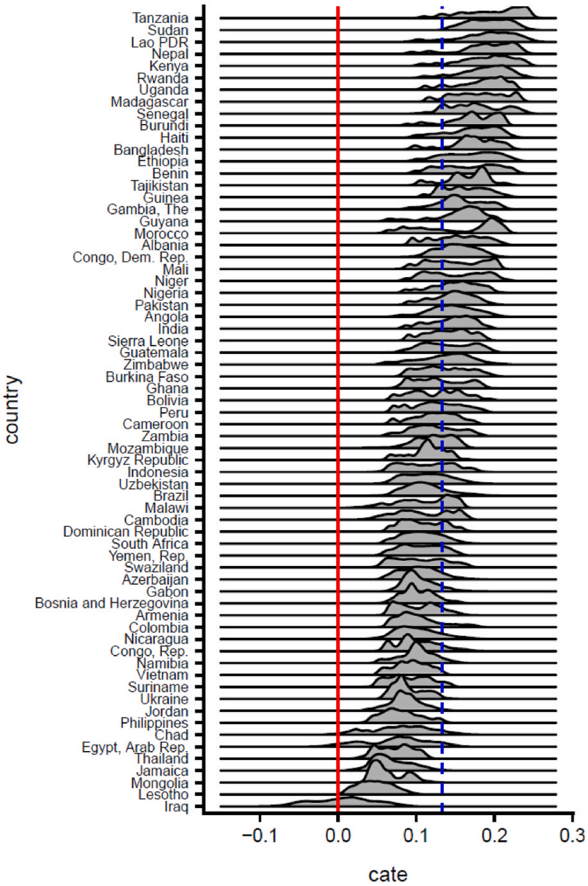

Fig. 6. *IMF effect heterogeneity by society*. Notes: The densities show the impact variation within and between societies. The blue line signifies the ATE of 0.14. The x-axis displays the change in probability of poverty, where positive values represent an increase in poverty. The results are based on data described in Table 2 and using a GRF as described in the Methods. Note that because the GRF is a non-parametric model, only an estimate of the treatment effect is provided but none for the controls (unlike a parametric model). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

followed by 16 (95% CI: 0.11–0.19), 18 (95% CI: 0.12–0.24), 15 (95% CI: 0.8–0.22) for the following quintiles. As the densities overlap for all five groups, the effect difference is not statistically significant.

These results challenge a sociological expectation that the most vulnerable tend to be those who struggle the most in times of economic turmoil (Babb 2005). This finding resonates with the theoretical explanation that middle-class families—which are better integrated into the economy before a crisis erupts—will likely be affected at least as severely as low-income families (Burgard and Kalousova 2015; Redbird and Grusky 2016). Although these effects travel along with the mechanisms of the first causal pathway (parents' earnings), the IMF effect appears to travel along the second causal pathway as well.

Much like Figs. 8 and 9 shows how much government education expenses moderate effect heterogeneity. The effect for those that spend in the fifth (spending most) to the second quintiles are similar, about 16 percentage points (95% CI: 0.14–0.18). While those in the first quintile, those spending less, have an effect size of 5 percentage points (95% CI: 0.02–0.08. Despite that it may appear puzzling that governments that spend more on education before implementing IMF programs (quintiles 2 to 5), also have the greatest increase in child poverty as a result of IMF programs (compared to governments that spent the least, i.e., quintile 1), there is a logical explanation. As previous research finds, when the IMF finds that countries have high public spending such as on education, it will seek to lower that spending to balance fiscal spending (Babb 2005; Daoud 2021; Daoud et al., 2017; Rowden 2011; Stuckler and Basu 2013). One of the results of such lowered education spending is the increase of vulnerable children. Nonetheless, although several studies find that these programs adversely affect the education system (Moosa and Moosa 2019), more research is required if scholars are to fully disentangle and quantify the relationship between IMF policies, education spending, and their joint effect on child poverty.

5.3. Sensitivity analysis

Although we estimated the confounding effect of a government's political will using a Heckman selection model, one may question the degree of bias political will inject at baseline. To evaluate the sensitivity of our analysis to various thought-experimental scenarios

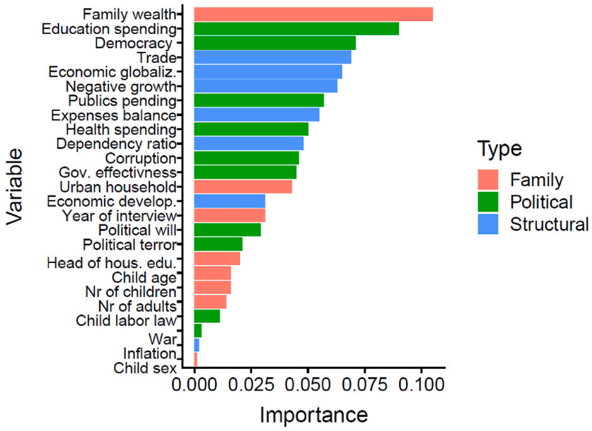

Fig. 7. *Variable importance moderating effect heterogeneity*. Notes: The bars highlight variable importance ranking. This importance metric identifies how often our GRF selected each variable to grow a tree. For each branch (node) split of a tree that the GRF grows, the GRF evaluates how much effect heterogeneity each covariate induces. The GRF then selects the covariate that maximizes this heterogeneity in that split. The variable importance metric calculates the proportion of time our GRF selected a variable, counting all the branch splits. A variable importance metric of 0.10 means that the GRF used that variable in? for? one-tenth of all splits across all trees. The results are based on data described in Table 2 and using a GRF as described in the Methods. Note that because the GRF is a non-parametric model, only an estimate of the treatment effect is provided but none for the controls (unlike a parametric model).

of unobserved confounding, we produced Fig. 10, showing a contour graph across such scenarios (Ding and VanderWeele 2016). The y-axis represents the risk difference between high and low degrees of political will on the probability (risk) of child poverty. This axis is the confounder-outcome association. The higher the y-axis value, the higher is the difference in political will, and thereby, increasing the risk of child poverty. Had political will been unrelated to child poverty, then a government's political will could not be confounding because it would not be associated with the outcome of interest. However, as discussed in the Background section, political will is plausibly associated with child poverty, where more poverty motivates a higher degree of political will to enroll in an IMF program.

The x-axis represents the risk difference between the high and low degree of political will on the probability (propensity) of enrolling in an IMF program—that is, the confounder-treatment association. Here, this risk difference represents the degree of difference in political will between those children exposed to IMF programs and those that are not. The larger this difference is, the more bias is injected.

The confounder-outcome association and the confounder-treatment associations must be at least 0.13 (in terms of risk difference), to explain away all the IMF effects on child poverty. This value is what is named the *explain-away value*—or E-value—in the sensitivity analysis (Ding and VanderWeele 2016; Peña 2022; Sjölander and Hössjer 2021). The dashed line in Fig. 10 highlights this E-value, and is the closest point that shifts the sign from IMF producing adverse effects on children's risk of poverty to instead producing beneficial effects. The higher prior difference is in terms of political will due to poverty (i.e., higher y-axis values), and the higher prior difference in political will in enrolling in IMF programs, the more baseline bias we have in our data. In those scenarios, the IMF effect is traversing from the grey area in Fig. 10 (representing adverse effects) to the white area (representing beneficial IMF effects).

While our sensitivity analysis shows under what confounding conditions our results can be explained away, we conducted additional sensitivity analyses, incorporating potential confounding and measurement error in the outcome. These analyses are presented in the first section of Appendix D. Additionally, we performed a series of robustness checks that evaluates model uncertainty. These robustness results are presented in Appendix C and D, showing relative stability.

6. Discussion

Despite a consensus among sociologists that economic conditions and individual well-being are closely connected, this strand of research has evolved by focusing on either the micro or macro level, not on the interactions between them. Our study evaluated these interactions in the context of how a form of economic shock—IMF programs—affects child poverty in low- and middle-income countries. To fully disentangle these complex interactions, we applied a machine-learning (ML) approach, which revealed an average adverse impact and several important heterogeneities.

Through both our novel method and our findings, our study contributes to the literature on public policy and poverty in the following three ways. First, building on Brady's (2019) framework on the causes of poverty, we apply a novel method that synthesizes insights from family and political sociology to examine how characteristics of both households and countries interact to affect children's risk of deprivation in the face of economic shocks (Beckfield 2018; Conger et al., 1992). We use a DAG to explicate our theoretical understanding of how macro and micro characteristics form pathways through which the shockwaves of austerity reverberate. Using causal modeling, we estimate that, on average, IMF programs increase the probability of child deprivation by 0.14, or 14 percentage points (95% CI: 0.03–0.24). This effect is larger still for children living in countries that have IMF programs.

Second, we also make two applied methodological contributions. One is that sociologists have long debated the relative merits of

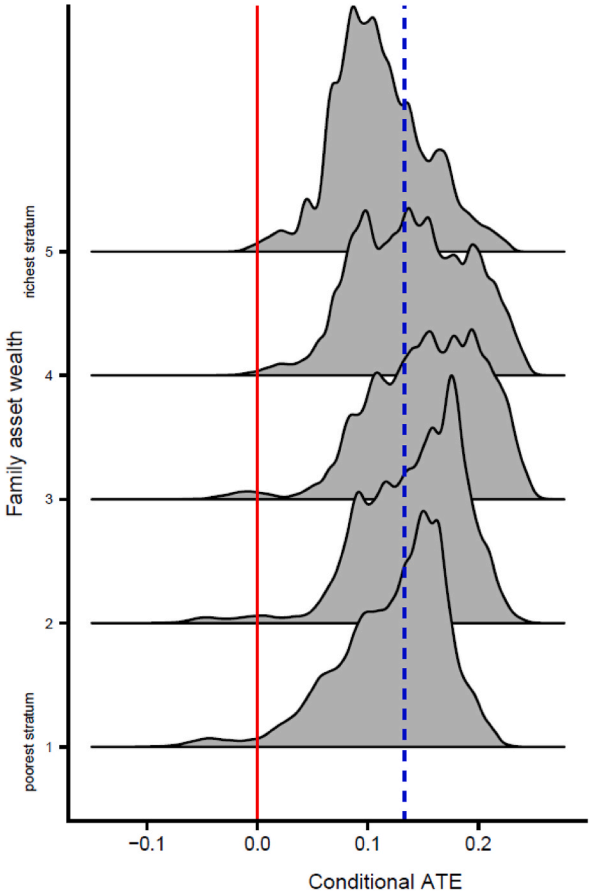

Fig. 8. *Family wealth moderating effect heterogeneity by quintile.* Notes: The densities show the impact variation within and between the quintiles of the DHS wealth asset variable. The blue line signifies the ATE of 0.14. The x-axis displays the change in probability of poverty, where positive values represent an increase in poverty. The results are based on data described in Table 2 and using a GRF as described in the Methods. Note that because the GRF is a non-parametric model, only an estimate of the treatment effect is provided but none for the controls (unlike a parametric model). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

predictive and explanatory models (Aldridge 1999; Hedström and Ylikoski 2010; Watts 2014). Common predictive models take the shape of a black box, with low interpretability, whereas many explanatory linear models have high interpretability but are limited to an in-sample analysis. Our research design shows that scholars can combine both types of modeling logic into a hybrid (Daoud and Dubhashi 2021, 2023). This design offers sociologists a new path for combining social theorizing and robust causal modeling.

The other new applied methodological contribution is showing how social scientists can combine social theorizing with novel methodologies from computer science. With this combination, sociologists can inductively identify impact heterogeneity in large population data (Athey and Imbens 2017; Grimmer et al. 2017; Mullainathan and Spiess 2017; Pearl and Mackenzie 2018). In so doing, we respond to Goldthorpe's call for a methodological framework that can capture population heterogeneity (2015:31).

Recent sociological studies aimed at capturing heterogeneity have focused on applying mostly unsupervised ML techniques to a variety of social and cultural settings (Molina and Garip 2019) and combining them with linear regression, social network analysis, and text analysis (Baldassarri and Goldberg 2014; Daoud et al. 2019a,b; DiMaggio et al. 2013; Goldberg 2011; McFarland et al., 2013). Our study extends this emergent research to causal inference for poverty research. We demonstrate how ML algorithm finds impact heterogeneity instead of merely testing it. These algorithms, especially tree-based techniques, provide a ranked list of the importance of each measure in moderating IMF impact. Such a list reveals an extensive picture of which measures are *the most* and *the least* important in modifying the relationship between a treatment and an outcome. Commonly used methods cannot generate such an extensive picture of variable interactions. Therefore, using ML for policy evaluation enables researchers to discover which social groups gain or lose the most from a specific event, making this type of ML particularly useful for scholars researching social stratification, economic sociology, social demography, public policy, and related fields.

Third, using the generalized random forest (GRF) algorithm, which is one of many ML algorithms, we discover three important empirical findings. One is the substantial amount of treatment heterogeneity distributed around the average IMF effect, with roughly half of the variation in impact taking place between countries and the other half within countries. This balanced variation suggests that both macro and micro factors have approximately equal weight in accounting for impact heterogeneity. Although several factors

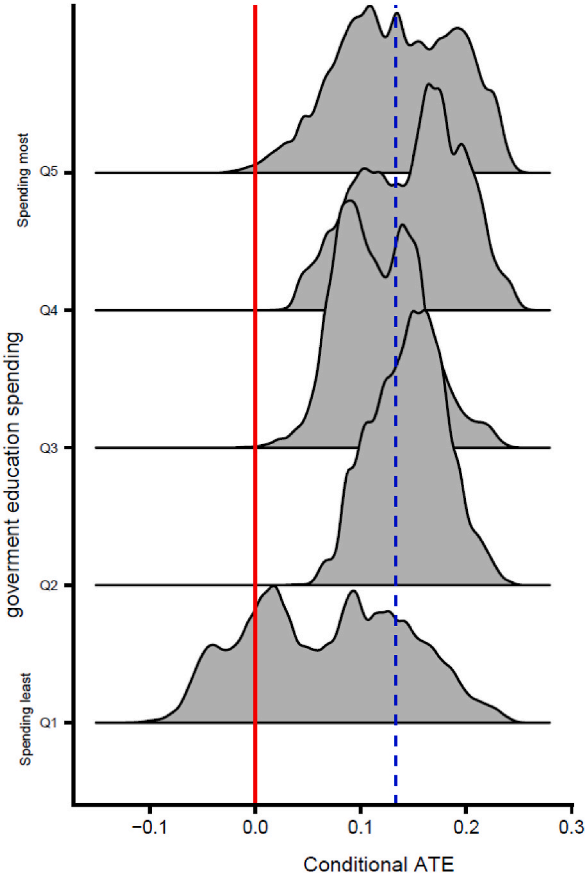

Fig. 9. *Education spending moderating effect heterogeneity by quintile.* Notes: The densities show the impact variation within and between the quintiles of the government education spending variable. The blue line signifies the ATE of 0.14. The x-axis displays the change in probability of poverty, where positive values represent an increase in poverty. The results are based on data described in Table 2 and using a GRF as described in the Methods. Note that because the GRF is a non-parametric model, only an estimate of the treatment effect is provided but none for the controls (unlike a parametric model). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

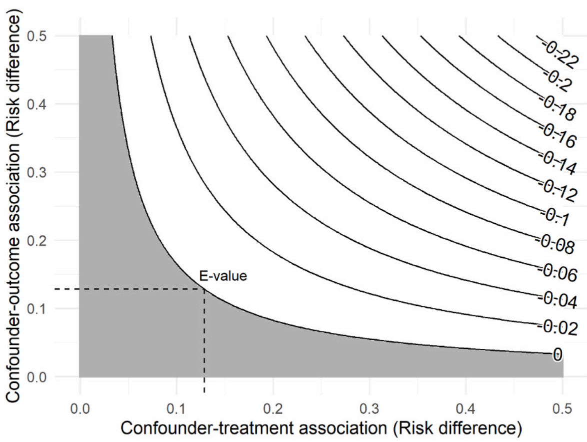

Fig. 10. *Sensitivity analysis*. They grey area represents IMF adverse effects on child poverty, and white area where this effect is beneficial. The dashed lines represent the explain-away value (E-value) of the joint association between confounder-outcome and confounder-treatment that would be required to explain away the estimated ATE.

moderate the IMF effect, we focused on the two-top ranking.

Family wealth is the highest-ranking impact moderator of all 25 covariates (7 structural, 10 political, and 8 family-related). In contrast to theories suggesting that economic crises harm lower-class families the most (Babb 2005; Burgard and Kalousova 2015), we find that children from middle-class families are equally at risk of falling into poverty after the implementation of an IMF program. The adverse effect on the middle class is comparable to the mechanisms found in stratification research for industrialized countries plagued by economic crises (Garfinkel et al., 2016; Grusky et al., 2008; Redbird and Grusky 2016; Stuckler and Basu 2013). This research informs us that although middle-class families possess better socio-economic defenses to protect their children from short-term shocks than those living in lower-class families, a strong enough shock will overpower family defenses. Additionally, if these adverse effects specifically hit those parts of society that tend to benefit middle-class families more—such as the education system—than low-income families, then the adverse effects amplify further for the middle-class. As specified in our DAG, while the first part of this IMF effect likely moves via the labor market pathway, the second part travels via the public policy path. Consequently, an important question for future research is to identify, based on mediation analysis, via which pathways a crisis affects middle-class families.

The magnitude of government education spending prior to an IMF program ranks as the second most important impact moderator. Our statistical analysis reveals that countries spending most on education before enrolling in an IMF program experience an increase in child poverty due to program implementation. Although this finding is likely partly associative (confounded), our archival analysis shows that this finding has a causal aspect. Several governments—such as Bolivia in 1999 and Tanzania in 2003—had to agree to immediate and severe reductions in education spending. The IMF's reasons for imposing on countries such as austerity policies range from "reducing fiscal burden" to "redistributing from the middle-class to the poor." Such policies show that IMF policies causally affect education spending.

Our findings are in line with previous research showing that IMF austerity policies will likely adversely affect education spending (Nielsen 2006). Sometimes, the IMF has encouraged charging tuition fees for primary education in other countries as well (Daoud 2021; Moosa and Moosa 2019; Rowden 2011). With such fees placing additional economic burdens on families, parents in low-income societies may keep children at home or make them useful by putting them to work at a young age, instead of sending them to school (Alexander 2001; Kane 2009). Whereas most previous research on how IMF affects government spending has focused on health and public spending (Clements et al. 2013; Stubbs et al., 2017), our findings show that more research is necessary for identifying the causal link between IMF, education spending, and child poverty (Buchmann 1996; Rowden 2011).

Our study has several limitations. First, countries are not randomly assigned to IMF programs. As our directed acyclic graph (DAG) shows, we control extensively for characteristics of countries and families that may be associated with both IMF intervention and child poverty. Nonetheless, conclusive identification of causal effects remains challenging with observational data (Morgan and Winship 2014). In our case, the Heckman selection model relies on the excludability of instruments and the normality of error terms, both of which are untestable assumptions (Heckman 1979). However, after conducting several checks, we find that our results are robust.

Second, our study focuses on estimating the reduced form of the IMF total effect and thus does not estimate the complete mediation paths (structural form) specified in our DAG. As the primary aim of the study is to analyze impact heterogeneity, we selected the newest ML algorithm—the GRF—for this task. However, this algorithm is not designed for mediation analysis. A GRF for mediation would need to optimize both the regularization of the treatment variable and the mediating mechanism (Athey et al., 2019; VanderWeele 2015). Moreover, with small samples, ML may struggle to balance the bias-variance trade-off, which results in "regularization bias" in estimated treatment effects (Nie and Wager 2018). This bias arises because the minimization of predictive error is invariant in distinguishing between a treatment variable and other covariates.

Researchers have developed several solutions to the problem of regularization bias, such as orthogonalization (Chernozhukov et al., 2018; Laan and Rubin 2006)., which decouples the treatment variable from covariates by separately modeling the propensity and the causal effects. Another solution is the R-learner paradigm, whereby the conditional average treatment effect is estimated according to the residuals of separate models of the treatment propensity and the marginal effect (Nie and Wager 2018). Thus future research needs to quantify any potential regularization bias in small samples.

Third, our study assumed that IMF programs contain a somewhat homogenous package of policies. By using a binary exposure variable, we estimate a single effect of IMF programs on child poverty, even though IMF policy treatments may vary across countries (Babb and Carruthers 2008; Caraway et al. 2012; Daoud et al., 2019a,b; Kentikelenis et al., 2016). As our complementary qualitative analysis shows, IMF programs contain different policies targeting different areas of a country's educational system. All these variations in IMF policy induce "heterogeneity in the treatment" (how differences in the content of the treatment effect an outcome) (Grimmer et al., 2017). While impact heterogeneity (how a homogenous treatment induces variations across subpopulations) is the focus of our study, future work should investigate what specific types of policies induce changes in child poverty (Daoud and Reinsberg 2018).

Moreover, while the implementation of IMF programs constitutes an important type of economic crisis—as they rely on policy-makers' political intentions—we do not know whether other types of economic turmoil—such as currency crises, bank crises, and sovereign debt crises—have similarly heterogeneous effects on child poverty (Daoud 2020; Redbird and Grusky 2016; Reinhart and Rogoff 2008). Future research needs to compare the impact of these crises on child poverty.

Although governments may stabilize their macroeconomic problems under the IMF's supervision, our empirical findings show that such stability comes at the high price of a greatly increased number of children falling into poverty. These adverse effects of economic shocks are population-wide, affecting both low- and middle-class families. Given that children's early development is crucial for their flourishing as adults, the difficult challenge facing international bodies and national governments is to find solutions for countries with macroeconomic problems—solutions that do not place the future of children in jeopardy (Daoud et al. 2022).

Authors' contribution

Conceived the research topic and led the research: AD. Identified the research design: AD and FJ. Collected the data: AD. Performed the research: AD. Wrote the methods section: AD and FJ. Wrote the rest of the manuscript: AD. Revised the manuscript: AD and FJ.

Data availability statements

The data that support the findings of this study are available from the Demographic and Health Survey, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the Demographic and Health Survey (www.measuredhs.com).

Funding statement

1. Royal Swedish Academy of Letters, History and Antiquities Research Fellow. 2. The Swedish Research Council Consolidator Grant.

Acknowledgments

Christopher Winship, Jason Beckfield, Lisa Berkman, Sebastian Kohl, Elias Nosrati, Bernhard Reinsberg, Leslie Adams, Emilie Courtin, Angela Dixon, Dick Durevall, Arne Bigsten, We would like to extend a special acknowledgment to Beth Truesdale and Natalie Reid for reading and providing extensive edits on several versions of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssresearch.2023.102973.

References

Abadie, Alberto, Athey, Susan, Imbens, Guido W., Wooldridge, Jeffrey M., 2020. Sampling-based versus design-based uncertainty in regression analysis. Econometrica 88 (1), 265–296. https://doi.org/10.3982/ECTA12675.

Abbott, Andrew, 1988. Transcending general linear reality. Socio. Theor. 6 (2), 169–186. https://doi.org/10.2307/202114.

Åkerström, Joakim, Daoud, Adel, Johansson, Richard, 2019. Natural Language processing in policy evaluation: extracting policy conditions from IMF loan agreements. In: The 22nd Nordic Conference on Computational Linguistics, vol. 5.

Aldridge, Alan, 1999. Prediction in sociology: prospects for a devalued activity. Socio. Res. Online 4 (3), 1–6. https://doi.org/10.5153/sro.333.

Alexander, Nancy C., 2001. Paying for education: how the world bank and the international monetary Fund influence education in developing countries. Peabody J. Educ. 76 (3–4), 285–338. https://doi.org/10.1080/0161956X.2001.9682002.

Athey, Susan, Imbens, Guido W., 2017. The state of applied econometrics: causality and policy evaluation. J. Econ. Perspect. 31 (2), 3–32. https://doi.org/10.1257/jep.31.2.3.

Athey, Susan, Wager, Stefan, 2019. Estimating Treatment Effects with Causal Forests: an Application. *ArXiv:1902.07409 [Stat]*.

Athey, Susan, Tibshirani, Julie, Wager, Stefan, 2019. Generalized random forests. Ann. Stat. 47 (2), 1148–1178. https://doi.org/10.1214/18-AOS1709.

Babb, Sarah, 2005. The social consequences of structural adjustment: recent evidence and current debates. Annu. Rev. Sociol. 31, 199–222.

Babb, Sarah L., Carruthers, Bruce G., 2008. Conditionality: forms, function, and history. Annu. Rev. Law Soc. Sci. 4 (1), 13–29. https://doi.org/10.1146/annurev.lawsocsci.4.110707.172254.

Bail, Christopher A., 2008. The configuration of symbolic boundaries against immigrants in Europe. Am. Socio. Rev. 73 (1), 37–59. https://doi.org/10.1177/000312240807300103.

Baldassarri, Delia, Goldberg, Amir, 2014. Neither ideologues nor agnostics: alternative voters' belief system in an age of partisan politics. Am. J. Sociol. 120 (1), 45–95. https://doi.org/10.1086/676042.

Balgi, Sourabh, Peña, Jose M., Daoud, Adel, 2022. Personalized public policy analysis in social sciences using causal-graphical normalizing flows. Proc. AAAI Conf. Artif. Intell. 36 (11), 11810–11818. https://doi.org/10.1609/aaai.v36i11.21437.

Beckfield, Jason, 2018. Political Sociology and the People's Health. Oxford University Press, Oxford.

Bonikowski, Bart, 2015. The promise of bourdieusian political sociology. Theor. Soc. 44 (4), 385–391. https://doi.org/10.1007/s11186-015-9246-7.

Boss, Pauline, Bryant, Chalandra M., Mancini, Jay A., 2016. Family Stress Management: A Contextual Approach. Sage Publications.

Brady, David, 2019. Theories of the causes of poverty. Annu. Rev. Sociol. 45 (1), 155–175. https://doi.org/10.1146/annurev-soc-073018-022550.

Breiman, Leo, 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Buchmann, Claudia, 1996. The debt crisis, structural adjustment and women's education. Int. J. Comp. Sociol. 37 (1), 5–30. https://doi.org/10.1163/002071596X00208.

Burgard, Sarah A., Kalousova, Lucie, 2015. Effects of the great recession: health and well-being. Annu. Rev. Sociol. 41 (1), 1–21. https://doi.org/10.1146/annurev-soc-073014-112204.

Cameron, A. Colin, Gelbach, Jonah B., Miller, Douglas L., 2008. Bootstrap-based improvements for inference with clustered errors. Rev. Econ. Stat. 90 (3), 414–427. https://doi.org/10.1162/rest.90.3.414.

Caraway, Teri L., Rickard, Stephanie J., Anner, Mark S., 2012. International negotiations and domestic politics: the case of IMF labor market conditionality. Int. Organ. 66 (1), 27–61.

Catalán, Nájera, Héctor, E., 2019. Reliability, population classification and weighting in multidimensional poverty measurement: a Monte Carlo study. Soc. Indicat. Res. 142 (3), 887–910. https://doi.org/10.1007/s11205-018-1950-z.

Catalán, Nájera, Héctor, E., David, Gordon, 2020. The importance of reliability and construct validity in multidimensional poverty measurement: an illustration using the multidimensional poverty index for Latin America (MPI-LA). J. Dev. Stud. 56 (9), 1763–1783. https://doi.org/10.1080/00220388.2019.1663176.

Chernozhukov, Victor, Chetverikov, Denis, Demirer, Mert, Duflo, Esther, Hansen, Christian, Newey, Whitney, Robins, James, 2018. Double/debiased machine learning for treatment and structural parameters. Econom. J. 21 (1), C1–C68. https://doi.org/10.1111/ectj.12097.
Clements, Benedict, Gupta, Sanjeev, Nozaki, Masahiro, 2013. What happens to social spending in IMF-supported programmes? Appl. Econ. 45 (28), 4022–4033. https://doi.org/10.1080/00036846.2012.744136.
Coleman, James, 1987. Microfoundations and macrosocial behavior. In: Alexander, J., Giesen, B., Munch, R., Smelser, N. (Eds.), The Micro-macro Link.
Conger, Rand D., Katherine, J. Conger, Elder, Glen H., Lorenz, Frederick O., Simons, Ronald L., Whitbeck, Les B., 1992. A family process model of economic hardship and adjustment of early adolescent boys. Child Dev. 63 (3), 526–541. https://doi.org/10.1111/j.1467-8624.1992.tb01644.x.
Conley, Dalton, Springer, Kristen W., 2001. Welfare state and infant mortality. Am. J. Sociol. 107 (3), 768–807. https://doi.org/10.1086/338781.
Daoud, Adel, 2020. The Wealth of Nations and the Health of Populations: A Quasi-Experimental Design of the Impact of Sovereign Debt Crises on Child Mortality. *ArXiv Preprint ArXiv:2012.14941*.
Daoud, Adel, 2021. The International Monetary Fund's Intervention in Education Systems and its Impact on Childrens Chances of Completing School. *ArXiv: 2201.00013 [Econ, q-Fin]*.
Daoud, Adel, Dubhashi, Devdatt, 2021. Melting together prediction and inference. Observational Studies 7 (1), 1–7.
Daoud, Adel, Dubhashi, Devdatt, 2023. Statistical modeling: the three cultures. Harvard Data Science Review 5 (1). https://doi.org/10.1162/99608f92.89f6fe66.
Daoud, Adel, Puaca, Goran, 2011. An organic view of students' want formation: pragmatic rationality, habitus and reflexivity. Br. J. Sociol. Educ. 32 (4), 603–622. https://doi.org/10.1080/01425692.2011.578440.
Daoud, Adel, Reinsberg, Bernhard, 2018. Structural adjustment, state capacity and child health: evidence from IMF programmes. Int. J. Epidemiol. dyy251, 1–10. https://doi.org/10.1093/ije/dyy251.
Daoud, Adel, Elias, Nosrati, Reinsberg, Bernhard, Kentikelenis, Alexander E., Stubbs, Thomas H., King, Lawrence P., 2017. Impact of international monetary Fund programs on child health. Proc. Natl. Acad. Sci. USA 114 (25), 6492–6497. https://doi.org/10.1073/pnas.1617353114.
Daoud, Adel, Kim, Rockli, Subramanian, S.V., 2019a. Predicting women's height from their socioeconomic status: a machine learning approach. Soc. Sci. Med. 238, 112486 https://doi.org/10.1016/j.socscimed.2019.112486.
Daoud, Adel, Reinsberg, Bernhard, Kentikelenis, Alexander E., Stubbs, Thomas H., King, Lawrence P., 2019b. The international monetary fund's interventions in food and agriculture: an analysis of loans and conditions. Food Pol. 83, 204–218. https://doi.org/10.1016/j.foodpol.2019.01.005.
Daoud, Adel, Anders, Herlitz, Subramanian, S.V., 2022. IMF fairness: calibrating the policies of the international monetary Fund based on distributive justice. World Dev. 157, 105924 https://doi.org/10.1016/j.worlddev.2022.105924.
DiMaggio, Paul, Nag, Manish, Blei, David, 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: application to newspaper coverage of U.S. Government arts funding. Poetics 41 (6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004.
Ding, Peng, VanderWeele, Tyler J., 2016. Sensitivity analysis without assumptions. Epidemiology (Cambridge, Mass 27 (3), 368–377. https://doi.org/10.1097/EDE.0000000000000457.
Elder, Glen H., 1998. Children of the Great Depression, 25th Anniversary, Updated edition. Westview Press. 25th Anniversary Edition.
Ferraro, Kenneth F., Schafer, Markus H., Wilkinson, Lindsay R., 2016. Childhood disadvantage and health problems in middle and later life: early imprints on physical health? Am. Socio. Rev. 81 (1), 107–133. https://doi.org/10.1177/0003122415619617.
Garfinkel, Irwin, McLanahan, Sara, Wimer, Christopher, 2016. Children of the Great Recession. Ruseel Sage Foundation, New York.
Garip, Filiz, 2012. Discovering diverse mechanisms of migration: the Mexico–us stream 1970–2000. Popul. Dev. Rev. 38 (3), 393–433. https://doi.org/10.1111/j.1728-4457.2012.00510.x.
Garip, Filiz, 2016. On the Move: Changing Mechanisms of Mexico-U.S. Migration. Princeton University Press, Princeton, NJ.
Goldberg, Amir, 2011. Mapping shared understandings using relational class analysis: the case of the cultural omnivore reexamined. Am. J. Sociol. 116 (5), 1397–1436. https://doi.org/10.1086/657976.
Goldthorpe, John H., 2015. Sociology as a Population Science. Cambridge University Press.
Gordon, David, Nandy, Shailen, Pantazis, Christina, Pemberton, Simon, Townsend, Peter, 2003. Child Poverty in the Developing World. Polity Press, Bristol, UK.
Grimmer, Justin, Solomon, Messing, Westwood, Sean J., 2017. Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods. Polit. Anal. 25 (4), 413–434. https://doi.org/10.1017/pan.2017.15.
Grusky, David B., Ku, Manwai C., Szelényi, Szonja (Eds.), 2008. Social Stratification: Class, Race, and Gender in Sociological Perspective, third ed. Westview Press, Boulder, Colo.
Halleröd, Björn, Rothstein, Bo, Daoud, Adel, Nandy, Shailen, 2013. Bad governance and poor children: a comparative analysis of government efficiency and severe child deprivation in 68 low- and middle-income countries. World Dev. 48 (0), 19–31. https://doi.org/10.1016/j.worlddev.2013.03.007.
Halliday, Terence C., Carruthers, Bruce G., 2009. Bankrupt: Global Lawmaking and Systemic Financial Crisis. Stanford University Press, Stanford.
Hamm, Patrick, King, Lawrence P., Stuckler, David, 2012. Mass privatization, state capacity, and economic growth in post-communist countries. Am. Socio. Rev. 77 (2), 295–324. https://doi.org/10.1177/0003122412441354.
Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome, 2009. The Elements of Statistical Learning, 2 edition. Springer, New York.
Heckman, James J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–161.
Hedström, Peter, Ylikoski, Petri, 2010. Causal mechanisms in the social sciences. Annu. Rev. Sociol. 36 (1), 49–67. https://doi.org/10.1146/annurev.soc.012809.102632.
Imbens, Guido W., Rubin, Donald B., 2015. Causal Inference for Statistics, Social, and Biomedical Sciences: an Introduction, 1 edition. Cambridge University Press, New York.
IMF, 2015. Factsheet: Protecting the Most Vulnerable under IMF-Supported Programs. Retrieved (www.imf.org.
Jerzak, Connor, Johansson, Fredrik, Daoud, Adel, 2023. Image-Based Treatment Effect Heterogeneity. Causal Learning and Reasoning.
Johansson, Fredrik D., Wei, Dennis, Oberst, Michael, Gao, Tian, Gabriel, Brat, Sontag, David, Varshney, Kush R., 2019. Characterization of Overlap in Observational Studies. *ArXiv:1907.04138 [Cs, Stat]*.
Kane, June, 2009. What the economic crisis means for child labour. Global Soc. Pol. 9 (1 Suppl. l), 175–196. https://doi.org/10.1177/1468018109106890.
Kentikelenis, Alexander, Stubbs, Thomas H., King, Lawrence P., 2016. IMF conditionality and development policy space, 1985–2014. Rev. Int. Polit. Econ. 23 (0), 543–582. https://doi.org/10.1080/09692290.2016.1174953.
Kino, Shiho, Hsu, Yu-Tien, Shiba, Koichiro, Chien, Yung-Shin, Mita, Carol, Kawachi, Ichiro, Daoud, Adel, 2021. A scoping review on the use of machine learning in research on social determinants of health: trends and research prospects. SSM - Population Health 15, 100836. https://doi.org/10.1016/j.ssmph.2021.100836.
Künzel, Sören R., Sekhon, Jasjeet S., Bickel, Peter J., Yu, Bin, 2018. Meta-Learners for Estimating Heterogeneous Treatment Effects Using Machine Learning. *ArXiv: 1706.03461 [Math, Stat]*.
Laan, Mark J. van der, Rubin, Daniel, 2006. Targeted maximum likelihood learning. Int. J. Biostat. 2 (1) https://doi.org/10.2202/1557-4679.1043.
Lancet, The, 2009. What has the Gates foundation done for global health? Lancet 373 (9675), 1577. https://doi.org/10.1016/S0140-6736(09)60885-0.
Lipton, Zachary C., 2017. The Mythos of Model Interpretability. *ArXiv:1606.03490 [Cs, Stat]*.
Loken, Eric, Gelman, Andrew, 2017. Measurement error and the replication crisis. Science 355 (6325), 584–585. https://doi.org/10.1126/science.aal3618.
Martin, Ricardo, Segura-Ubiergo, Alex, 2004. Social Spending in IMF-Supported Programs. *IEO Background Paper* (BP/04/1).
McFarland, Daniel A., Ramage, Daniel, Chuang, Jason, Heer, Jeffrey, Manning, Christopher D., Jurafsky, Daniel, 2013. Differentiating Language usage through topic models. Poetics 41 (6), 607–625. https://doi.org/10.1016/j.poetic.2013.06.004.
McLanahan, Sara, Percheski, Christine, 2008. Family structure and the reproduction of inequalities. Annu. Rev. Sociol. 34 (1), 257–276. https://doi.org/10.1146/annurev.soc.34.040507.134549.
Molina, Mario, Garip, Filiz, 2019. Machine learning for sociology. Annu. Rev. Sociol. 45 (1), 27–45. https://doi.org/10.1146/annurev-soc-073117-041106.
Moosa, Imad A., Moosa, Nisreen, 2019. The Effects of IMF Operations on Social Expenditure. Springer, 111–34 in Eliminating the IMF.

Morgan, Stephen L., Winship, Christopher, 2014. Counterfactuals and Causal Inference: Methods and Principles for Social Research, 2 edition. Cambridge University Press, New York, NY.
Mullainathan, Sendhil, Spiess, Jann, 2017. Machine learning: an applied econometric approach. J. Econ. Perspect. 31 (2), 87–106. https://doi.org/10.1257/jep.31.2.87.
Nie, Xinkun, Wager, Stefan, 2018. Quasi-Oracle Estimation of Heterogeneous Treatment Effects. *ArXiv:1712.04912 [Econ, Math, Stat]*.
Nielsen, H. Dean, 2006. From Schooling Access to Learning Outcomes: an Unfinished Agenda: an Evaluation of World Bank Support to Primary Education. World Bank Independent Evaluation Group, Washington D.C.
Pearl, Judea, 2009. Causality: Models, Reasoning and Inference, second ed. Cambridge University Press, Cambridge, U.K. ; New York.
Pearl, Judea, Mackenzie, Dana, 2018. The Book of Why: the New Science of Cause and Effect, 1 edition. Basic Books, New York.
Pemberton, Simon, Gordon, David, Nandy, Shailen, 2012. Child rights, child survival and child poverty: the debate. In: Global Child Poverty and Well-Being. Policy Press, pp. 19–38.
Peña, Jose M., 2022. Simple yet sharp sensitivity analysis for unmeasured confounding. J. Causal Inference 10 (1), 1–17. https://doi.org/10.1515/jci-2021-0041.
Ragin, Charles C., 1987. The Comparative Method : Moving beyond Qualitative and Quantitative Strategies. University of California Press, Berkeley.
Redbird, Beth, Grusky, David B., 2016. Distributional effects of the great recession: where has all the sociology gone? Annu. Rev. Sociol. 42 (1), 185–215. https://doi.org/10.1146/annurev-soc-073014-112149.
Reinhart, Carmen M., Rogoff, Kenneth S., 2008. This Time Is Different: A Panoramic View of Eight Centuries of Financial Crises. National Bureau of Economic Research.
Rickard, Stephanie J., Caraway, Teri L., 2014. International negotiations in the shadow of national elections. Int. Organ. 68 (3), 701–720. https://doi.org/10.1017/S0020818314000058.
Rihoux, Benoît, Ragin, Charles C., 2008. Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques (Applied Social Research Methods). Sage Publications, Inc.
Rowden, Rick, 2011. Impacts of IMF Policies on National Education Budgets and Teachers. Education International Research Institute, Brussels.
Schafer, Markus H., Ferraro, Kenneth F., Mustillo, Sarah A., 2011. Children of misfortune: early adversity and cumulative inequality in perceived life trajectories. Am. J. Sociol. 116 (4), 1053–1091. https://doi.org/10.1086/655760.
Schneider, Daniel, Hastings, Orestes P., Joe, LaBriola, 2018. Income inequality and class divides in parental investments. Am. Socio. Rev. 83 (3), 475–507. https://doi.org/10.1177/0003122418772034.
Shandra, Carrie L., Shandra, John M., Bruce, London, 2012. The international monetary Fund, structural adjustment, and infant mortality: a cross-national analysis of sub-saharan africa. J. Poverty 16 (2), 194–219. https://doi.org/10.1080/10875549.2012.667059.
Sjölander, Arvid, Hössjer, Ola, 2021. Novel bounds for causal effects based on sensitivity parameters on the risk difference scale. J. Causal Inference 9 (1), 190–210. https://doi.org/10.1515/jci-2021-0024.
Smith, Louisa H., Mathur, Maya B., VanderWeele, Tyler J., 2021. Multiple-bias sensitivity analysis using bounds. Epidemiology (Cambridge, Mass 32 (5), 625–634. https://doi.org/10.1097/EDE.0000000000001380.
Snijders, Tom A.B., Bosker, Roel J., 1999. Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling. Sage Publications, Thousand Oaks.
Stiglitz, Joseph E., 2003. Globalization and its Discontents, 1 edition. W. W. Norton & Company.
Stubbs, Thomas, Alexander, Kentikelenis, Stuckler, David, McKee, Martin, King, Lawrence, 2017. The impact of imf conditionality on government health expenditure: a cross-national analysis of 16 west african nations. Soc. Sci. Med. 174, 220–227. https://doi.org/10.1016/j.socscimed.2016.12.016.
Stuckler, David, Basu, Sanjay, 2013. The Body Economic: Eight Experiments in Economic Recovery, from Iceland to Greece. Allen Lane, London.
Tibshirani, Julie, Athey, Susan, Wager, Stefan, 2020. The GRF Algorithm. Retrieved. https://grf-labs.github.io/grf/REFERENCE.html#cluster-robust-estimation. (Accessed 15 May 2020).
Turco, Catherine J., Zuckerman, Ezra W., 2017. Verstehen for sociology: comment on Watts. Am. J. Sociol. 122 (4), 1272–1291. https://doi.org/10.1086/690762.
VanderWeele, Tyler, 2015. Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford University Press, Oxford, New York.
Vreeland, James Raymond, 2003. The IMF and Economic Development. Cambridge University Press, New York.
Watts, Duncan J., 2014. Common sense and sociological explanations. Am. J. Sociol. 120 (2), 313–351. https://doi.org/10.1086/678271.
Xie, Yu, Brand, Jennie E., Ben, Jann, 2012. Estimating heterogeneous treatment effects with observational data. Socio. Methodol. 42 (1), 314–347. https://doi.org/10.1177/0081175012452652.