Using policy learning to inform health insurance targeting: a case study of Indonesia

Abstract

This paper demonstrates how optimal policy learning can inform the targeted allocation of Indonesia's two subsidised health insurance programmes. Using national survey data, we develop policy rules aimed at minimising "catastrophic health expenditure" among enrollees of APBD or APBN, the two government-funded schemes. Employing a super learner ensemble approach, we use regression and machine learning methods of varying complexity to estimate conditional average treatment effects and construct policy rules to optimise programme benefits, both with and without budget constraints. We find that the financial impact of APBD enrollment over APBN differs with household characteristics, particularly demographic composition, socioeconomic status, and geography. Households assigned to APBD under the policy rule are typically urban-based with better facilities, whereas rural households with less accessible healthcare are assigned to APBN – a pattern intensified under budget constraints. Both constrained and unconstrained optimal policy assignments show lower expected catastrophic expenditure risk than the current assignment strategy. This study contributes to the literature on heterogeneous treatment effects, optimal policy leaning, and health financing in developing countries, showcasing data-driven solutions for more equitable resource allocation in public health insurance contexts.

Keywords— optimal policy learning; heterogeneous treatment effects; health insurance; Indonesia; global health

1 Introduction

Designing effective health insurance policies for low- and middle-income countries (LMICs) is a persistent challenge, exacerbated by limited resources and diverse population needs. Traditional approaches to policy design often fall short in capturing the nuances of individual and household characteristics that influence health insurance efficacy, leaving substantial room for efficiency improvements (Bhattacharya and Dupas, 2012). Recent evidence from LMICs demostrates that health insurance impacts vary significantly across populations, with effectiveness heavily dependent on household characteristics, regional factors, and healthcare access (Fink et al., 2013; Grogger et al., 2015). This heterogeneity suggests the potential value of more targeted approaches to policy design and implementation.

The growing literature on policy learning and targeting in social protection programs has demonstrated that data-driven approaches can significantly improve program effectiveness through better matching of beneficiaries to interventions (Alatas et al., 2016; Hanna and Olken, 2018). In recent years, the concept of optimal policy learning has emerged as a promising framework for assigning interventions based on characteristics that maximise welfare outcomes, such as reducing healthcare expenditures and improving health coverage (Athey and Wager, 2021; Dehejia, 2005; Kitagawa and Tetenov, 2018; Manski, 2004).

This approach is particularly relevant for health insurance programs, where recent studies have documented substantial heterogeneity in insurance value and utilisation patterns across different population subgroups (Finkelstein and Notowidigdo, 2019). Our study demonstrates novel approaches of policy learning in the context of health insurance. Specifically, the study aims to inform the optimal allocation of lower-income households to Indonesia's two subsidised health insurance programmes, by estimating statistical treatment assignment rules. We explore how these data-driven methods can be used to enhance the allocation process, ensuring that households receive the most suitable insurance scheme to maximise financial protection.

Following recent proposals from the optimal policy learning literature, we consider two types of statistical rules: threshold-based rules and tree-based rules. Threshold-based rules assign policies based on whether expected benefits exceed a certain level, offering simplicity, but depending on correctly identifying the direction of the effect (Luedtke and van der Laan, 2016b; van der Laan and Luedtke, 2015). Tree-based rules, conversely, use decision trees to divide the population into interpretable subgroups, providing policymakers with insights into which household types benefit most from each scheme while allowing for flexibility in rule complexity (Amram et al., 2022; Athey and Wager, 2021; Bertsimas et al., 2019).

Our methodological innovation advances the application of the Super Learner (van der Laan et al., 2007), an ensemble machine learning technique. Although traditionally used for prediction (Polley and van der Laan, 2010) and estimating average treatment effects (van der Laan and Rose, 2011), we build upon a recent extension aimed at deriving the optimal policy rule (Luedtke and van der Laan, 2016b; Montoya et al., 2021, 2023; van der Laan and Luedtke, 2015). We demonstrate that in the Super Learner framework, tree- and threshold-based rules can be combined to form an optimal policy allocation rule.

To demonstrate the practical application of these methodological innovations, we focus on Indonesia's two subsidised health insurance programs, PBI-APBD and PBI-APBN, which target lower-income households under the broader Jaminan Kesehatan Nasional (JKN) scheme. Our aim is to construct statistical rules to assign households to either insurance scheme to minimise the burden of catastrophic health expenditures, a key component of financial protection and a Sustainable Development Goal indicator (World Health Organization, 2010). We estimate optimal policy rules under two scenarios: one in which APBD enrolment is resource-constrained (limiting enrolment to 10% of eligible households, reflecting the limited healthcare budget of the state government), and one without constraints, allowing free allocation based solely on expected benefits. Our dataset includes socioeconomic and demographic data from Indonesia's 2017 National Socioeconomic Household Survey (SUSENAS) and the 2018 Village Potential Statistics Census (PODES), which provide a detailed basis for estimating policy impacts on household financial outcomes. We report two main sets of results: first the achievable welfare with the estimated rules, in terms of counterfactual values of catastrophic health expenditure. Second, we further explore how the learned policy rules align with household characteristics, offering insights into potential improvements over the current assignment.

Our work contributes to three distinct but interconnected literatures. First, we advance research on subsidised health insurance's impact on financial protection in developing countries. Recent work has highlighted the complex relationship between insurance coverage and financial outcomes in LMICs (Erlangga et al., 2019; Limwattananon et al., 2012), with evidence suggesting that programme effectiveness varies significantly based on local context and beneficiary characteristics. Second, we contribute to the growing literature on optimal policy learning (Athey and Wager, 2021; Kitagawa and Tetenov, 2018), by providing a practical demonstration of recently proposed methods in the context of a health policy targeting using observational data. We propose a workflow that handles practical challenges such as confounding adjustment and selection of effect modifiers, using sample splitting and cross-validation to guard against overfitting concerns. To our knowledge we are the first to incorporate the tree-based policy learning algorithm proposed by Athey and Wager (2021) in an ensembling machine learning framework. Third, we improve our understanding of heterogeneous treatment effects in social and health interventions, particularly through the use of machine learning (Athey et al., 2019; Kennedy, 2023; Künzel et al., 2019; Nie and Wager, 2021; Shalit et al., 2017). Our study highlights that by leveraging heterogeneous treatment effects, policies can be personalised to achieve better results than arbitrary or one-size-fits-all assignments (Dehejia, 2005; Manski, 2004).

2 The Indonesian policy setting

Indonesia's JKN programme had enrolled 83% of the population by mid-2019, representing significant progress toward universal health coverage since its 2014 inception (Prabhakaran et al., 2019). However, substantial challenges remain, particularly in providing effective coverage for the economically vulnerable through the PBI scheme. These challenges mirror those faced by other LMICs, where health financing reforms must balance ambitious coverage goals with limited resources and complex implementation environments.

A key concern remains in JKN's effectiveness in preventing catastrophic health expenditures, a challenge frequently observed in similar LMIC health insurance schemes (El-Sayed et al., 2018; Pratiwi et al., 2021). Research on health insurance impacts in developing countries reveals mixed results regarding financial protection and healthcare access (Bernal et al., 2017; Erlangga et al., 2019; Maulana et al., 2022), with effectiveness often moderated by local healthcare infrastructure and socioeconomic conditions. For example, Grogger et al. (2015) report substantial variation in the impact of Mexico's public health insurance on catastrophic expenditures across income groups and regions, while Fink et al. (2013) demonstrate how the effectiveness of community health insurance in Burkina Faso varies by household wealth and distance to health facilities. These examples underscore the potential value of targeted approaches in programme design and implementation to address such heterogeneity.

In Indonesia, most JKN enrollees are in the PBI scheme, where the government subsidises premiums. Table 1 presents data from SUSENAS 2017, showing the utilisation of insurance for outpatient and inpatient care among PBI-insured household members.

Within the PBI scheme, there are two programmes available. The APBN scheme, managed by the central government, targets the poorest households using the Unified Database (UDB) and indicators such as household assets, income, and consumption, prioritising those below the national poverty line. On the other hand, the APBD scheme, managed by regional governments, relies on regional databases, community assessments, and local vulnerability criteria, often extending coverage to households near or just above the poverty line.

Among APBD enrollees, 38% did not use insurance for outpatient care and 48% for inpatient care. The corresponding figures for APBN enrollees were 11% and 34%, respectively. These disparities suggest that differences in the two schemes may lead to varying levels of financial protection and utilisation of health services, with some individuals continuing to pay out-of-pocket due to concerns about care quality, discrimination, and wait times (Harimurti et al., 2013).

JKN has struggled to engage the large informal working sector, with low uptake driven by immediate health needs and financial constraints (Dartanto et al., 2020; Vilcu et al., 2016). While JKN aimed to unify all subsidised programmes, local schemes like APBD remain crucial, particularly in rural and remote areas where local governments have been more effective in enrolling the uninsured (Kruse et al., 2012; Sparrow et al., 2017). Local initiatives and targeted enrolment strategies could enhance health coverage access for these groups (Dutta et al., 2020).

Table 1: Demand for inpatient and outpatient health care in 2017

	N	**Inpatient demand (N)**			**Outpatient demand (N)**		
		Accessed care	Used APBD	Used APBN	Accessed care	Used APBD	Used APBN
PBI-APBD	120,346	3,096	1,926 (62%)	-	13,271	6,857 (52%)	-
PBI-APBN	212,334	8,744	-	7,786 (89%)	29,970	-	19,841 (66%)

Note: Table reports the number of household members that used inpatient and outpatient health care, by PBI insurance scheme, and whether they used insurance to pay. Source: SUSENAS 2017.

3 The policy allocation problem

Our study aims to demonstrate how optimal policy learning approaches can help allocate eligible households among the two subsidised health insurance schemes, APBD and APBN. Given JKN's universal coverage goals, we do not focus on evaluating the benefits of insurance compared to no insurance, but on determining the most suitable scheme for a given household. By focusing exclusively on PBI enrollees, we mitigate self-selection bias, and ensure overlap between the two compared programmes, as APBN and APBD enrollees tend to share comparable socioeconomic profiles due to similar social protection targeting criteria, while both observed and unobserved differences between the insured and uninsured populations may be substantial. Although our focus on insured households may appear to bypass under-enrolment challenges, especially within the informal sector, this methodological choice reflects several considerations from recent research. Studies from Thailand, for example, show that analysing insurance value among current beneficiaries can inform strategies to expand coverage (Limwattananon et al., 2012). Similarly, Alatas et al. (2016) demonstrate that improving the efficiency of existing programs can free resources for expanding services to uninsured populations.

Our identification strategy leverages the distinct targeting mechanisms of the two schemes, controlling for criteria used by both schemes, and accounting for regional implementation differences. In selecting variables for policy rules, we align with policy decision-making criteria, reflecting the targeting approaches employed for PBI enrolment. The lists of variables for confounding adjustment and for inclusion in the potential policy rules are presented in detail in Section 4.

Table 2: Targeting variables included in our optimal policy learning model

Household member characteristics and demography
Household head marital status (1 if married, 0 otherwise)
Number of household members
Number of productive household members (aged 15-64)
Number of children at school (aged 5-14)
Socioeconomic status
Number of household members in work
Household head education level (1 if completed compulsory education, 0 otherwise)
Number of household members that completed compulsory education
Dwelling characteristics
Homeowner status (1 if homeowner, 0 otherwise)
Wall type (1 if concrete, 0 otherwise)
Roof type (1 if concrete/roof tile, 0 otherwise)
Lighting source (1 if electricity, 0 otherwise)
Drinking water source (1 if protected, 0 otherwise)
Toilet facility (1 if private/shared, 0 otherwise)
Cooking fuel (1 if electricity/gas, 0 otherwise)
Asset ownership
Refrigerator (1 if owns, 0 otherwise)
Gas canister >5.5kg (1 if owns, 0 otherwise)
Number of cellular phones in household
Car/motorcycle (1 if owns, 0 otherwise)
Gold/jewellery >10g (1 if owns, 0 otherwise)

Note: Targeting variables are based on the original variables included in the PMT model specification to determine the PBI selection criteria.

4 Data

SUSENAS is an annual national socioeconomic household survey that collects data on approximately 300,000 households, and its members, across 514 districts (Badan Pusat Statistik, 2017). A two-stage sampling design is used to ensure that households are representative at the district-level, introducing clustering, and frequency weights are provided to reflect the population. The survey comprises two modules that gather information on basic household characteristics, such as occupation, education, health, housing, asset ownership and consumption expenditure. We use cross-sectional data from the 2017 survey to construct a household-level dataset, and restrict the sample to households where all members reported having the same insurance type at the time of survey (that is, either PBI-APBD or PBI-APBN). Households where at least one member reported having alternative health insurance cover, either through their employers or private plans, are excluded from the analysis. We generate a binary indicator for our catastrophic expenditure outcome, which equals 1 if health care expenditure, as a proportion of non-food household expenditure, exceeds 10%, and 0 otherwise (Wagstaff and Lindelow, 2008; Wagstaff et al., 2007; Wagstaff and van Doorslaer, 2003; Xu et al., 2003).

We refer to the related literature on health insurance and catastrophic expenditure, and control for a rich set of variables that may influence both the selection into treatment and the outcome. Using SUSENAS, we control for head-of-household characteristics since they are representative of household well-being (Bookwalter et al., 2006). These include age, sex, marital status and socioeconomic status (which encompasses literacy, level of education, occupation and use of technology). We additionally capture broader household characteristics by controlling for the number of household members that are educated, in work, of a productive working age (between 15 and 64), or currently at school. To further capture households' socioeconomic status, we include a selection of variables that reflect asset ownership and housing characteristics, including household amenities (e.g. types of water supply, toilet facilities and electricity), small households assets (e.g. fridge, television, computer), and other household characteristics (e.g. number of rooms, home ownership).[1]

Finally, we control for certain village-level characteristics using Potensi Desa data, PODES 2018, which collects information on villages' availability of natural and human resources, and can be merged with SUSENAS at the district-level (Badan Pusat Statistik, 2018). We include information on whether access to all types of health care facilities (that is, primary and secondary care, community health care, maternity care and pharmacies) is considered to be easy.

See Appendix A for a full list of variables included in our model for confounding adjustment.

In Indonesia, central government programs use proxy means testing (PMT) models to identify households in poverty (Alatas et al., 2012; Vilcu et al., 2016). In 2011, the government collected socioeconomic data from 24.5 million of the poorest households, creating the UDB, which was updated in 2015. This PMT model identifies household assets and demographic characteristics associated with consumption, informing PBI eligibility (National Team for the Acceleration of Poverty Reduction (Indonesia), 2015). We base the targeting variables used in the policy rules on the UDB, ensuring that they are accessible by decision makers for real world policy targeting. These include the household size, number of household members who are employed, and various indicators of asset ownership. The complete list of targeting variables is presented in Table 2.

5 Methods

5.1 Notation and setup

We construct an observational dataset of households $i = 1, \ldots, N$ that are characterised by the tuple (X_i, Y_i, D_i), where X_i is a vector of confounders and effect modifiers; $Y_i \in \{0, 1\}$ is a

[1] We do not control for health care need since we are unable to identify whether health status is a confounder or a mediator.

binary outcome variable, which takes the value of 1 if the household's health expenditure as a proportion of total expenditure exceeds 10%; and $D_i \in \{0,1\}$ is a binary treatment variable, which is equal to 0 if all household members are enrolled into the nationally funded PBI scheme, PBI-APBN (the "control"), and 1 if all household members are enrolled into the locally funded PBI scheme, PBI-APBD (the "treatment"). Individual level treatment effects are defined as $\tau_i = Y_i(1) - Y_i(0)$, where $Y_i(1)$ and $Y_i(0)$ are potential outcomes under the treatment and the control.

To learn a policy rule π, we first define our targeting criteria $V \subseteq X$, where V could include all variables in X, or a subset. The policy rule maps V into a binary decision, $\pi : V \to \{0,1\}$. We specify a policy class Π that encodes any possible constraints on the policymaker's objective function (for example, resource or functional form restrictions) that the policy rule must satisfy, where $\pi \in \Pi$. We encode a resource constraint, a fraction $\kappa \in [0,1]$, which represents the maximum fraction of households that can receive treatment (APBD). We consider two scenarios: an unconstrained environment where $\kappa = 1$, and a constrained environment where $\kappa = 0.1$, to reflect the proportion of the state health care budget that is distributed to local governments.[2] The counterfactual outcome for a household if treatment is assigned using a policy rule is denoted as $Y_i(\pi(V_i))$. Since smaller values of Y correspond to better outcomes (we want to minimise catastrophic expenditure), the optimal policy rule π^* can be defined as a minimiser of the expected counterfactual mean outcomes over all candidate rules $\pi \in \Pi$:

$$\pi^* = \arg\min_{\pi \in \Pi} E[Y_i(\pi(V_i))] \text{ s.t. } E[\pi(V_i)] \leq \kappa, \tag{1}$$

where $E[Y_i(\pi(V_i))]$ is the policy value: our target causal parameter. The utility of implementing an optimal policy rule π^* can be measured by estimating its policy value relative to a given reference rule $\pi^r(V_i)$. As reference rules, we include static rules that assign either APBD or APBN to all households. We denote with $U(\pi^*)$ the following contrast between the outcome under π^* and $\pi^r(V_i)$:

$$U(\pi^*) = E[Y_i(\pi^*(V_i)) - Y_i(\pi^r(V_i))]. \tag{2}$$

Note that the causal target parameter in Equation 1 and 2 is written in terms of potential outcomes. In order to write our target causal parameter in terms of the observed data (known as the statistical target parameter or estimand) we need the following assumptions to hold:

(a) Unconfoundedness: $Y_i(d) \perp D_i | X_i = x, \forall d \in D, \forall x \in \mathcal{X}$

(b) Overlap: $0 < e(x) \equiv P[D_i = 1 | X_i = x] < 1, \forall x \in \mathcal{X}$

(c) Stable Unit Treatment Value Assumption (SUTVA): $Y_i = Y_i(D_i)$

[2] The imposed resource constraint is mainly for demonstrative purposes to model the Minister of Finance's Regulation No. 78/PMK.02/2020, which stipulates that 10% of APBD recipients' contributions will be funded by the central government through the state budget. Our model does not take into account that the remaining 90% will be funded through other sources.

We refer to $e(x)$ as the propensity score. If above listed Assumptions (a)-(c) are satisfied, our causal target parameter $E[Y_i(\pi(V_i))]$ (and $U(\pi^*)$) can be written in terms of the observed data. We emphasise that Assumptions (a)-(c) are standard assumptions made in the causal inference literature.

The optimal policy rule π^* can be learned using the CATE function $\tau(v)$, which is defined as the expected difference in the potential outcomes, as a function of the households' observed covariate profile $V_i \in X_i$:

$$\tau(v) = E[Y_i(1) - Y_i(0)|V_i = v]. \tag{3}$$

Analogous to the identification result for the causal target parameter $E[Y_i(\pi(V_i))]$, we can write $\tau(v)$ in terms of the observed data as

$$\begin{align} \tau(v) &= E[Y_i|V_i = v, D_i = 1] - E[Y_i|V_i = v, D_i = 0] \tag{4} \\ &= \mu_1(v) - \mu_0(v), \tag{5} \end{align}$$

where the expectation is taken with respect to the outcome and all covariates in X not in V. If V constitutes all available covariates X, we simply write $\mu_d(x)$ as the counterfactual response surface defined as $E[Y_i|X_i = x, D_i = d]$. In the remainder of the manuscript, we use $\tau(v)$ to denote the statistical target parameter written in terms of the observed data. We also denote estimated values from data with the $\hat{()}$ symbol, so that, for example, an estimate of the optimal policy rule π^* is $\hat{\pi}^*$.

5.2 Learning optimal policy rules

Below we introduce an important building block of policy learning, double-robust scores. We outline how these scores are used to learn rules within the two main policy classes we consider here: threshold based rules and tree-based rules. We then describe the meta-learning approach we suggest to select among candidate rules.

5.2.1 Doubly robust scores

We assign doubly robust scores Γ_i to each household, according to the augmented inverse probability of treatment weighted (AIPTW, or doubly robust) estimator for the ATE:[3]

[3] The AIPTW estimator for the ATE is formed by averaging the doubly robust scores: $\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N} \hat{\Gamma}_i(X_i)$

$$\hat{\Gamma}_i = \underbrace{\left[\hat{\mu}_1(X_i) + \frac{D_i}{\hat{e}(X_i)}(Y_i - \hat{\mu}_1(X_i))\right]}_{\hat{\Gamma}_i(1)} - \underbrace{\left[\hat{\mu}_0(X_i) + \frac{1-D_i}{1-\hat{e}(X_i)}(Y_i - \hat{\mu}_0(X_i))\right]}_{\hat{\Gamma}_i(0)} \tag{6}$$

$$= \hat{\mu}_1(X_i) - \hat{\mu}_0(X_i) + \frac{D_i - \hat{e}(X_i)}{\hat{e}(X_i)(1-\hat{e}(X_i))}(Y_i - \hat{\mu}_D(X_i)), \tag{7}$$

where $\hat{\mu}_D(X_i)$ is the conditional expectation function for the observed outcome under the treatment actually received. Doubly robust scores can be used (beyond using them to estimate the ATE) to evaluate the value of a given policy rule, and to select the optimal policy rule within a policy class. The empirical solution to Equation (1) minimises the counterfactual mean outcome under the rule using the doubly robust scores:

$$\hat{\pi}^* \in \arg\min_{\pi \in \Pi} \frac{1}{n}\sum_{i=1}^{n} \hat{\Gamma}_i(\pi(V_i)) \quad \text{s.t.} \quad \frac{1}{n}\sum_{i=1}^{n} \pi(V_i) < \kappa, \tag{8}$$

where $\hat{\Gamma}_i(\pi(V_i)) = \pi(V_i)\hat{\Gamma}_i(1) + (1-\pi(V_i))\hat{\Gamma}_i(0)$. Here, $\hat{\Gamma}_i(1)$ and and $\hat{\Gamma}_i(0)$ represent the doubly robust score for household (i) under treatment and control, respectively. We also use a modified version of the doubly robust scores, proposed by Athey and Wager (2021), that use an initial CATE estimate and apply a doubly robust adjustment (more details in Section 5.2.3).

5.2.2 Threshold-based rules

Threshold-based rules follow the intuition that treatment should only be assigned when it is effective. In our policy learning problem, this means assigning treatment to households with negative $\hat{\tau}(v)$). In recent years, a number of flexible estimators for $\tau(v)$ have been proposed, including relying on fully nonparametric models (examples include Athey et al. (2019); Hahn et al. (2020); Kennedy (2023); Künzel et al. (2019); Nie and Wager (2021); Shalit et al. (2017)). In particular, van der Laan and Luedtke (2014) and Luedtke and van der Laan (2016b) propose the two-stage doubly robust estimator, where the first stage requires the construction of doubly robust scores $\hat{\Gamma}_i$ as in (7), and the second stage involves modelling $\hat{\Gamma}_i$ as a function of $V_i \subseteq X_i$ using any regression-based approach. The first stage takes care of the confounding adjustment by controlling for the full covariate vector X_i in the outcome and treatment models, so that the second stage CATE estimation can focus on the target criteria $V_i \subseteq X_i$. The advantages of threshold-based rules include their simplicity and the ability to incorporate resource restrictions, where there is a maximum proportion of the population that can be assigned to treatment (Luedtke and van der Laan, 2016a). This constrained minimisation problem can be solved by sorting the observed data according to their estimated CATEs in increasing order, and assigning treatment to those with the lowest estimates until the constraint is met (see Appendix C for further details). A limitation of threshold-based rules is their lack of interpretability, as they do not reveal which variables influence assignment decisions. To enhance interpretability, threshold-based rules can incorporate interpretable algorithms, such as decision trees or linear models, which clarify variable importance and facilitate understanding of the decision-making process. This approach can also be strengthened by employing a discrete super learner[4] (described in detail later) limited to a library of interpretable algorithms, ensuring that the decision rules remain both effective and transparent.

5.2.3 Tree-based rules

Tree-based rules are an alternative method for policy learning that uses fixed-depth decision trees to directly optimise net policy benefits (Athey and Wager, 2021; Zhou et al., 2022). Decision trees recursively split observations into subgroups, in a way that maximises expected outcomes, until they end up in a tree leaf that is associated with a policy decision (i.e., APBD or APBN). Essentially, the optimiser takes $\hat{\Gamma}_i$ and V_i as inputs, and searches through the space of all candidate trees to identify the one that solves the minimisation problem in Equation (8).[5]. To construct $\hat{\Gamma}_i$, Athey and Wager (2021) propose a modified version of Equation (7) that replaces the first component (that is, the difference between the response functions) with a causal forest estimate of the CATE function $\hat{\tau}^{cf}(v)$:

$$\hat{\Gamma}_i^{cf} = \hat{\tau}^{cf}(V_i) + \frac{D_i - \hat{e}(X_i)}{\hat{e}(X_i)(1 - \hat{e}(X_i))}(Y_i - \hat{m}(X_i) - (D_i - \hat{e}(X_i))\hat{\tau}^{cf}(V_i)), \tag{9}$$

where $\hat{m}(X_i) = E[Y_i|X_i = x]$.[6,7] Causal forests find neighbourhoods of observations with similar CATEs by regressing the Y-residual on the D-residual, and recursively partitioning the data into leaves to maximise the within-leaf heterogeneity in treatment effects, thus forming a causal tree. Each observation is dropped down the tree and assigned a weight based on how frequently it is used to estimate $\hat{\tau}(V_i)$ at $V_i = v$. This process is repeated across many causal trees using bootstrapped samples to grow a forest. The CATE estimator is constructed as:

$$\hat{\tau}^{cf}(v) = \frac{\sum_{i=1}^{N} w_i(v)(D_i - \hat{e}(X_i))(Y_i - \hat{m}(X_i))}{\sum_{i=1}^{N} w_i(v)(D_i - \hat{e}(X_i))^2}, \tag{10}$$

[4] The discrete SuperLearner selects the single best-performing model from a library of candidate algorithms, based on cross-validated performance. In contrast, the continuous SuperLearner combines multiple models by assigning weights to each, producing an ensemble that aims to improve predictive accuracy by leveraging strengths across algorithms (Polley and van der Laan, 2010)

[5] Further details on the tree-search algorithm can be found in Sverdrup et al. (2020) and Zhou et al. (2022)

[6] Note that $\hat{m}(X_i)$ is different to the outcome regression $\hat{\mu}_D(X_i)$ in (7) as it is marginalised over the treatment D_i.

[7] The intuition behind constructing $\hat{\Gamma_i}$ using (9) rather than (7) is to produce a (possibly) better score since it requires a single estimate of the CATEs $\tau^{cf}(V_i)$ rather than two separate estimates of nuisance models $\mu_1(X_i)$ and $\mu_0(X_i)$.

where $w_i(v)$ are the weights derived from the forest splitting on the vector-valued gradient of $\tau^{cf}(v)$.[8,9] The main advantage of tree-based rules is their interpretability since the policy rule can depend on a small number of variables in V, if a shallow depth decision tree is specified. The downside, however, is that shallow trees may not capture all heterogeneity in treatment effects. Furthermore, the direct reliance on the CATE estimates can introduce additional complexities and potential biases. However, recent work demonstrates that model-agnostic doubly robust scores can be used for policy trees (instead of CATEs), thereby enhancing the flexibility of tree-based learning (Hatamyar and Kreif, 2023).

5.2.4 An ensemble learner for optimal policy rules

Ensemble learners can consider alternative models of varying complexity within a single framework, and enable the selection of the best performing policy rule in a data-adaptive way. The ensemble learner we consider using in this study is the super learner; a cross-validation based ensemble prediction algorithm (van der Laan et al., 2007). In the policy learning setting, the objective of the super learner is to improve the estimation of the policy rule by constructing an ensemble learner that, among a library of candidate estimators of the optimal rule, chooses the optimal weighted convex combination of candidates (known as the "continuous" super learner) (Luedtke and van der Laan, 2016b). One of its favourable properties is that, in large enough samples, it performs at least as well as the best performing candidate model (Dudoit and van der Laan, 2005; van der Laan and Dudoit, 2003; van der Laan et al., 2006, 2007; van der Vaart et al., 2006). In the policy learning setting, implementing the super learner requires the researcher to define a so-called library of candidate estimators of the policy rule, the type of method (or meta-learner) to combine candidate estimates, and the loss function to evaluate candidate performance. We explain these in details in the following section.

5.3 The workflow of policy learning

In this section, we explain our workflow for estimating and evaluating optimal policy rules in relation to our PBI assignment problem (see Figure 1 for an accompanying graphic diagram). We closely refer to van der Laan and Luedtke (2015) and Luedtke and van der Laan (2016b), which describe the theory behind the super learner framework for policy learning, and Montoya et al. (2021) and Montoya et al. (2023), which provide the implementation and interpretation.

[8] The CATEs are estimated "out-of-bag", meaning that for each observation dropped down the tree, a prediction is made using trees that did not use this observation during the training process. Out-of-bag prediction produce CATE estimates $\hat{\tau}^{cf}(V_i)$ without the need for explicit data splitting techniques (Athey et al., 2019).

[9] In theory, the doubly robust scores $\hat{\Gamma}_i^{cf}$ in (12) could be regressed onto V_i to estimate CATEs as per the method described in the previous section. However, since the causal forest directly produces CATE estimates $\tau^{cf}(V_i)$, the additional regression step is not required.

We rely on the following software packages in `R`, and modify code where necessary to suit our implementation: `SL.ODTR`[10,11], `SuperLearner`[12], `grf`[13] and `policytree`[14].

5.3.1 Selection of effect modifiers

We select $V \subseteq X$ as the list of targeting variables included in Table 2. These variables represent decision makers' selection criteria for assigning subsidised health insurance (PBI), and we assume that they are also relevant for our policy learning problem. We denote this vector of covariates as $V1$. To further reduce the large set of potential effect modifiers, we follow the doubly robust adaptive LASSO approach to construct a second, smaller covariate vector $V2 \in V1$ (Bahamyirou et al., 2022).[15] We do this after dividing the full data into two parts with a 30:70 split, using the smaller partition (≈28,000 households) to learn $V2$, leaving the larger partition (≈65,000 households) to learn policy rules. We implement the adaptive LASSO on the smaller 30% data partition using the following steps:

1. We fit super learners to estimate the nuisance parameters – $e(X_i)$, $\mu_1(X_i)$, $\mu_0(X_i)$ and $\mu_D(X_i)$ – and construct $\hat{\Gamma}_i$. See section 5.3.2 for details on candidate algorithms included in the super learners.

2. We regress $\hat{\Gamma}_i$ on $V1_i$ to obtain $\tilde{\beta}_l$, the estimated coefficient of $V1_{(l)}$ for $l = 1, \ldots, p$ (where l indexes the candidate algorithms and p is the total number of candidate algorithms in the super learner), and construct coefficient-specific weights $\hat{w}_l = \frac{1}{|\tilde{\beta}_l|^\gamma}$ (we set $\gamma = 1$) so that the regularisation penalises more those coefficients with lower estimates in the initial linear regression.

3. We fit a LASSO regression of $\hat{\Gamma}_i$ on $V1_i$ (again on the full 30% data partition) using $\hat{w}_l$ as the penalty factor associated with each coefficient [16]. The tuning parameter λ is selected using cross-validation. The final selection of covariates in $V2$ are those with non-zero coefficients.

We do not use this 30% data partition again and from here on, we refer to the remaining 70% data partition as the full data.

[10] https://github.com/lmmontoya/SL.ODTR

[11] We modify the original code in `SL.ODTR` to incorporate tree-based rules into the candidate library and to enable minimisation problems.

[12] https://github.com/ecpolley/SuperLearner

[13] https://github.com/grf-labs/grf

[14] https://github.com/grf-labs/policytree

[15] The adaptive LASSO is a regularisation method based on the traditional LASSO. The algorithm has the oracle property, in that it consistently selects the right subset of variables, and has an optimal estimation rate. It also uses coefficient-specific weights in the regularisation so that true-zero coefficients are less likely to be selected than in the traditional LASSO. See Zou (2006) for further details.

[16] We use the same 30% data partition to perform both regressions in the adaptive LASSO, but in theory, the data could be further partitioned to separate the data used for both tasks.

5.3.2 Specification of candidate algorithms

We specify J candidate algorithms for estimating the optimal policy rule $\pi^*(V_i)$ for $j = 1, \ldots, J$. These include:

- Threshold-based rules using CATE estimates generated in the following ways: (1) various parametric and nonparametric regression specifications of $\hat{\Gamma}_i$ on V_i.[17]; and (2) causal forests.
- Tree-based rules (i.e. shallow decision trees)[18] using doubly robust scores that incorporate causal forest CATE estimates $\hat{\tau}^{cf}(V_i)$.
- Static rules that assign the same policy to all households.

See Table 3 for a full list of candidate algorithms included in our library. Note that all candidate tree-based and threshold-based estimators are separately fitted on $V1$ and $V2$.

Table 3: Candidate estimators included in the super learner library

Estimator	**Description**	**Inputs**
Threshold-based rules		
GLM	Generalised linear model	$\hat{\Gamma}_i, V_i$
GLMi	Generalised linear model with interactions	$\hat{\Gamma}_i, V_i$
GBM	Generalised boosted model (depth 2)	$\hat{\Gamma}_i, V_i$
PM	Multivariate adaptive polynomial spline regression	$\hat{\Gamma}_i, V_i$
NN	Neural network	$\hat{\Gamma}_i, V_i$
SVM	Support vector machines	$\hat{\Gamma}_i, V_i$
CF	Causal forest	$\hat{\tau}^{cf}(V_i)$
Tree-based rules		
PT1	Policy tree (depth 1)	$\hat{\Gamma}_i^{cf}, V_i$
PT2	Policy tree (depth 2)	$\hat{\Gamma}_i^{cf}, V_i$
PT3	Policy tree (depth 3)	$\hat{\Gamma}_i^{cf}, V_i$
Static rules		
Treat all	Assign PBI-APBD to all households	-
Treat none	Assign PBI-APBN to all households	-

Note: Inputs refer to the parameters required to learn the policy rule. All threshold- and tree-based rules are separately fitted on $V1$ and $V2$. The super learner library for the constrained policy rule only includes threshold-based rules.

[17] Including a diverse range of candidate regressions in our library, from simple parametric models to more flexible, data-adaptive algorithms, hedges against the possibility that the optimal policy rule could be best modelled by a simple (rather than a more complex) estimator.

[18] Athey and Wager (2021) advise fitting shallow trees (that is, an interaction depth of 2 or 3) to prevent overfitting and for computational efficiency, although cross-validation can also be used to choose the optimal depth.

5.3.3 Trimming the sample to address practical positivity violations

Although we do not find significant violations in overlap between the covariate distributions of the APBD and APBN populations, we trim the sample to address practical positivity violations caused by extreme propensity scores near 1 or 0 – that is, almost always assigned to APBD or never assigned to APBD, conditional on observed characteristics (Stürmer et al., 2021). Extreme scores can affect inverse probability of treatment weights, thus creating bias and excessive variance in the treatment effect estimators (Li et al., 2019).[19] We fit a super learner on the full data, and following Crump et al. (2009), we remove approximately 4,000 households (6% of the sample) with $\hat{e}(X_i)$ outside of the range [0.1,0.9]. We refer to the remaining data as the trimmed data.

5.3.4 Learning the policy rules

To prevent overfitting, we divide the trimmed data randomly into two equal subsamples, denoted s_1 and s_2; s_1 is used to train the candidate models for estimating the policy rule, and s_2 is used to test the models and make predictions on new, unseen data. To account for a potential loss in efficiency from sample splitting, and to achieve better finite-sample performance with good asymptotic properties, we use 2-fold cross-fitting that swaps the roles of s_1 and s_2 and recreates the trimmed dataset by pooling the test predictions of the nuisance models, CATEs and policy rules (Chernozhukov et al., 2018a; Kennedy, 2023; Newey and Robins, 2018; Zhou et al., 2022). We describe the following procedure on the iteration where s_1 and s_2 are the respective training and testing data.

K-fold cross-validation on the training data

We split s_1 into K folds (we choose $K = 2$ based on the effective sample size, as per Phillips et al. (2022)) to perform the following steps using the K-fold cross-validation framework, where $k = 1, \ldots, K$ is the validation fold and $K - k$ is the training fold.

1. Generate candidate estimates

   Using the training fold $K - k$, we estimate the nuisance parameters – $e(X_i)$, $\mu_1(X_i)$, $\mu_0(X_i)$, $\mu_D(X_i)$ and $m(X_i)$[20] – using the super learner and construct doubly robust scores.[21,22,23] We train $j = 1, \ldots, J$ candidate estimators using the respective inputs

[19] Trimming the sample effectively changes the causal estimand to the expected policy value for households with sufficient overlap.

[20] We require predictions of $m(X_i)$ to estimate $\hat{\tau}^{cf}$ and to construct $\hat{\Gamma}_i^{cf}$.

[21] The doubly robust scores include both $\hat{\Gamma}_i$ and $\hat{\Gamma}_i^{cf}$

[22] Fitting the metalearner and predicting both the nuisance parameters and the doubly robust scores on the same training data is a form of "nested" cross-validation called "revere" (Coyle, 2017).

[23] Our candidate library for the nuisance models include a generalised linear model, LASSO regression and a generalised boosted model (with an interaction depth of 2).

detailed in Table 3. Using the validation fold k, and for each candidate j, we use the trained models to predict CATEs $\tau_j^k(V_i)$ and policy rules $\pi_j^k(V_i) = \mathbb{I}\{\tau_j^k(V_i) < 0\}$ for each household $i \in n$.

2. Meta-learning of policy rules using the super learner
The meta-learner finds weighted convex combinations of candidate estimators of the CATEs $\tau_\alpha(V_i)$ or candidate estimators of the policy rules $\pi_\alpha(V_i)$, defined as:

$$\tau_\alpha(V_i) = \sum_j \alpha_j \tau(V_i), \qquad \alpha_j \geq 0 \,\forall\, j, \sum_j \alpha_j = 1 \tag{11}$$

$$\pi_\alpha(V_i) = \mathbb{I}[\sum_j \alpha_j \pi_j(V_i) > 0.5], \qquad \alpha_j \geq 0 \,\forall\, j, \sum_j \alpha_j = 1, \tag{12}$$

where the weight vectors associated with a given convex combination $\alpha_1, \ldots, \alpha_J$ are non-negative and sum to one.[24,25] The unconstrained policy rules (where $\kappa = 1$) use (12) since the candidate library includes non-threshold-based estimators, and the constrained policy rules (where $\kappa = 0.1$) use (11) since the candidate library can only include threshold-based estimators in order to rank households by their estimated CATEs. Each candidate convex combination of algorithms is evaluated by estimating the counterfactual mean outcome under the candidate rule $L_{E[Y(\pi_j(V_i))]}$ (i.e. the loss) using the AIPTW estimator. We estimate the loss for each validation fold k, and average the estimates across all folds K to produce a single estimated loss for each candidate convex combination. The one with the lowest cross-validated loss is selected as the optimal candidate weighting α^*.

Constructing final policy rules using test data

We re-train each candidate estimator of the CATEs $\tau_j(V_i)$ and the optimal policy rule $\pi_j(V_i)$ on the full training data s_1, and generate predictions $\hat{\tau}_j(V_i)$ and $\hat{\pi}_j(V_i)$ on the testing data s_2. Using the optimal weights α^* from the previous step, we combine the candidate predictions to yield the super learner estimate of the optimal policy rule, where $\hat{\pi}^*_{\tau(V_i)}(V_i) = \mathbb{I}[\tau_{\alpha^*}(V_i) > 0]$ (that is, the rule based on the sign of the optimal weighted convex combination of candidate CATEs) and $\hat{\pi}^*_{\pi(V_i)}(V_i) = \hat{\pi}_{\alpha^*}(V_i)$ (that is, the rule that corresponds to the optimal weighted convex combination of candidate rules).

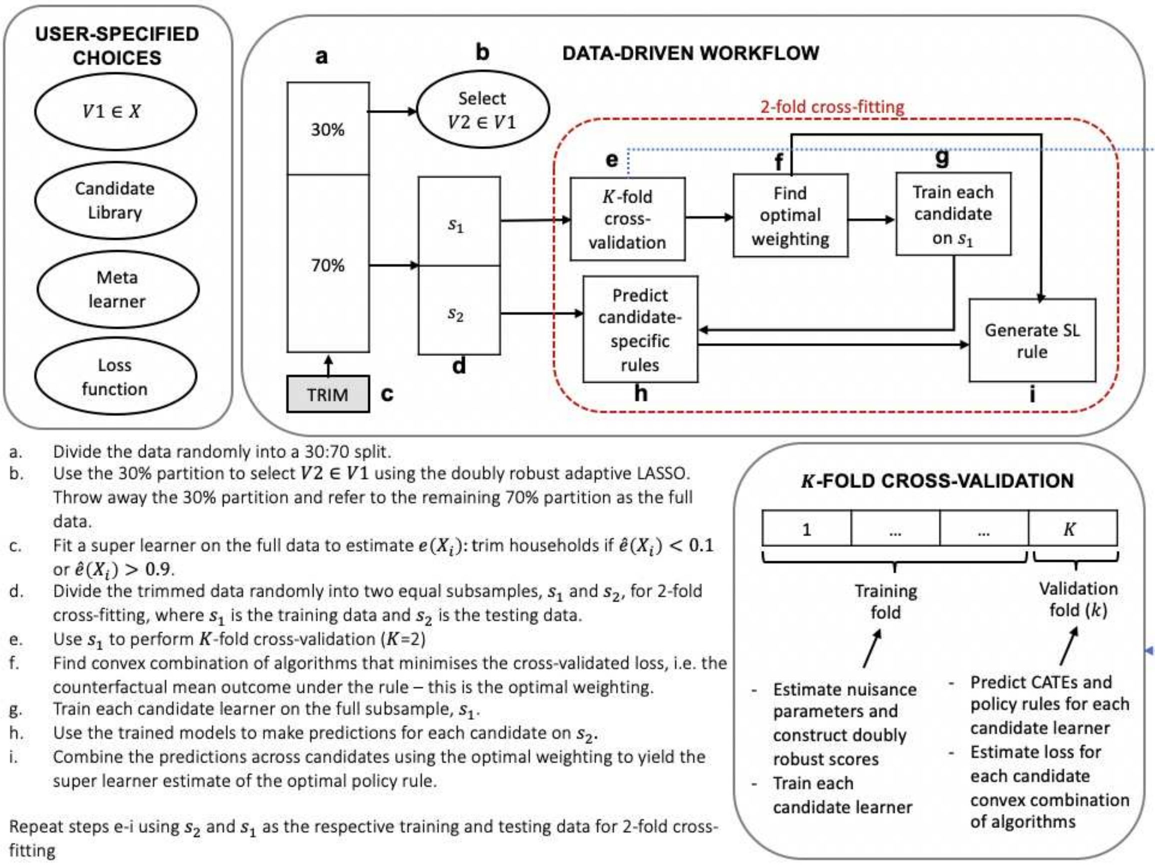
Validation strategy and robustness to overfitting For causal estimands, the lack of a ground truth for CATEs or optional policy rules makes direct validation infeasible. To ensure robustness and generalisability, we use two complementary validation techniques. First, we use 2-fold cross-fitting, which ensures that all predictions – including nuisance parameters,

[24] Convex combinations are found using the simplex method, which represents a linear programming problem as a system of linear equations, and defines an algorithm for finding the solution to this system of linear equations.

[25] Note from (12) that a given convex combination of candidate policy rules is made using a weighted majority vote, meaning that if the weighted average of candidate rules is greater than 0.5, $\pi_\alpha(V_i)$ is equal to one; and zero otherwise.

CATEs, and policy rules – are made on data unseen during training. This approach mimics out-of-sample testing and is widely recognised as a rigorous validation method in causal inference. Second, we apply K-fold cross-validation within the super learner framework to evaluate and adaptively weight candidate models based on their cross-validated performance. Together, these techniques provide safeguards against overfitting and ensure sufficient reliability of the learned policy rules.

Figure 1: Our policy learning workflow.


USER-SPECIFIED CHOICES
$V1 \in X$
Candidate Library
Meta learner
Loss function
DATA-DRIVEN WORKFLOW
a
30%
70%
TRIM
c
b
Select $V2 \in V1$
s_1
s_2
d
2-fold cross-fitting
e
K-fold cross-validation
f
Find optimal weighting
g
Train each candidate on s_1
Predict candidate-specific rules
h
Generate SL rule
i
K-FOLD CROSS-VALIDATION
1
...
...
K
Training fold
Validation fold (k)
- Estimate nuisance parameters and construct doubly robust scores
- Train each candidate learner
- Predict CATEs and policy rules for each candidate learner
- Estimate loss for each candidate convex combination of algorithms

a. Divide the data randomly into a 30:70 split.

b. Use the 30% partition to select $V2 \in V1$ using the doubly robust adaptive LASSO. Throw away the 30% partition and refer to the remaining 70% partition as the full data.

c. Fit a super learner on the full data to estimate $e(X_i)$: trim households if $\hat{e}(X_i) < 0.1$ or $\hat{e}(X_i) > 0.9$.

d. Divide the trimmed data randomly into two equal subsamples, s_1 and s_2, for 2-fold cross-fitting, where s_1 is the training data and s_2 is the testing data.

e. Use s_1 to perform K-fold cross-validation (K=2)

f. Find convex combination of algorithms that minimises the cross-validated loss, i.e. the counterfactual mean outcome under the rule – this is the optimal weighting.

g. Train each candidate learner on the full subsample, s_1.

h. Use the trained models to make predictions for each candidate on s_2.

i. Combine the predictions across candidates using the optimal weighting to yield the super learner estimate of the optimal policy rule.

Repeat steps e-i using s_2 and s_1 as the respective training and testing data for 2-fold cross-fitting

5.3.5 Evaluating the policy rules

First, we analyse the estimated CATEs from our threshold-based candidates, to determine whether there is any value in learning rules that exploit treatment effect heterogeneity in assigning APBD over APBN. We plot Targeting Operating Characteristic (TOC) curves that compare the ATE of assigning treatment (APBD) to increasing fractions of households q (households are sorted according to $\hat{\tau}_j(X_i)$ in decreasing order) (Yadlowsky et al., 2021). Here, q represents the fraction of households exposed to APBD during the evaluation process, similar to κ, but used specifically for exploring treatment effects. While the primary target causal parameter is the value under the optimal policy rule, exploring the ATEs using TOC curves allows us to illustrate the marginal improvements in treatment effects as we increase the fraction of treated households. We also separately plot sorted group average treatment effects (GATEs) that stratify $\hat{\tau}_j(X_i)$ into quintiles ordered in terms of treatment benefit, and estimate $\hat{\tau}_j$ for each quintile. This approach helps to identify and quantify treatment effect heterogeneity across different subgroups. Significant differences between the bottom and higher quintiles provide evidence of varying treatment impacts (Chernozhukov et al., 2018b). Second, we compare the counterfactual mean outcomes under the estimated optimal policy rules to the static rules and the actual policy assignment, to determine whether personalised policy rules outperform the status quo. The counterfactual mean outcomes under the various rules are estimated using the AIPTW estimator, as well as the targeted minimum loss-based estimator (TMLE) for robustness.[26] We additionally report the counterfactual mean outcomes under each candidate policy rule, as well as their weighted contributions to the super learner, to infer their relative importance in the optimal policy rules. Third, we identify the subgroups formed from the optimal policy rules by characterising households that are assigned to APBD and APBN under the optimal policy rule, and comparing them to households under the actual policy assignment. Lastly, we plot visual representations of the tree-based rules that are made up of splitting nodes (that is, the covariates to split on and their associated splitting values) and leaf nodes (that is, the policy decisions), for additional subgroup analyses.

6 Results

6.1 Descriptive statistics

Table 4 compares the average characteristics of households that are assigned to PBI-APBD and PBI-APBN under the actual policy assignment D_i for a selection of observed covariates from X_i. Although the two samples are largely comparable, we find some small differences. Compared to their APBN counterparts, the heads of APBD households are more likely to be in the 25-44 age bracket, married and in employment, particularly in the primary sector. On average, their households tend to be larger in size (average of 3.7 members compared to 3.5 for APBN), although with fewer members that are educated. There are also fewer members that use the internet or have cellular phones, and the household is slightly less likely to have basic household amenities such as electricity and drinking water. Both samples report having easy access to all

[26] TMLE is a doubly robust estimator that aims to debias our target parameter in question by estimating a "clever" covariate $H_i = \frac{\mathcal{I}[D_i=\pi(V_i)]}{e(X_i)}$, and updating the initial estimate of the outcome regression under the rule $\hat{\mu}_{\pi(V_i)}(X_i)$ with the intercept term (estimated using maximum likelihood estimation) from a generalised linear model that regresses Y_i on an offset term logit$\{\hat{\mu}_{\pi(V_i)}(X_i)\}$ and H_i. See van der Laan and Luedtke (2015); van der Laan and Rose (2018) for more details.

types of health care facilities, although accessibility is significantly better for APBN households, especially for hospitals. On average, community care centres are the most accessible health facility across all households, and hospitals are the least accessible. Geographically, the average household enrolled into PBI is rurally based. This is especially true for APBD households, who, according to the data, are more likely than APBN households to live in Sumatera, Kalimantan and Maluku-Papua, which represent a few of the regions in the country with high percentages of rural populations (Mardiansjah et al., 2021). Figure 2 displays balance statistics on the full covariate vector X, showing that after reweighting the trimmed sample using the inverse of the propensity score, all covariate means are balanced (absolute SMD < 0.1) across the two subpopulations.[27] For any small remaining imbalances, we address these through our doubly robust approach.

6.2 Variable selection

Figure B.2 plots the coefficients for $V1$, as a function of the $\log(\lambda)$ values used in the adaptive LASSO model. We select the covariates with non-zero coefficients that are associated with the value of $\log(\lambda)$ that minimises the cross-validated mean squared error. Our model selects 10 (out of 19) household-level variables relating to demographics, basic amenities and asset ownership – the number of members aged between 15 and 64, the number of educated members, the number of cellular phones, home ownership status, indicators for whether the household has electricity, drinking water, electric/gas cooking fuel and concrete walls, and indicators for whether the household owns a gas canister, vehicle (car or motorcycle) and gold – that form $V2$.

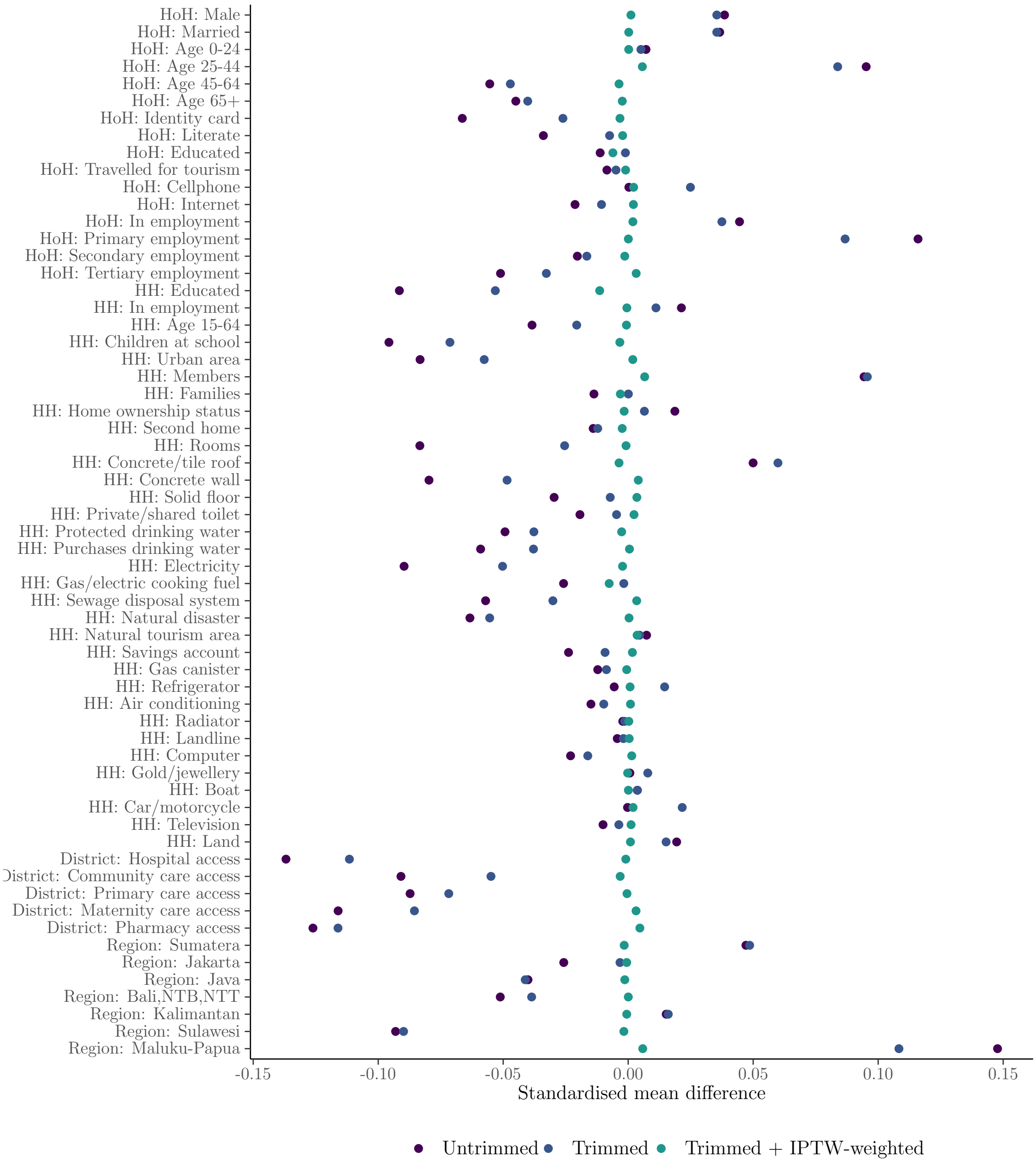
6.3 Exploring treatment effect heterogeneity

Figure 3 plots TOC curves showing estimated improvements in the ATE (compared to the overall ATE) from assigning APBD to increasing fractions of the population, ranked according to their candidate-specific CATE estimates. The area above the TOC curve provides some evidence of treatment effect heterogeneity since the ATEs for highly ranked subpopulations outperform the overall ATEs. We find that, in general, the CATE models that consider $V1$ display more heterogeneity than those that consider $V2$. In both instances, generalised boosted models find very little variation in treatment effects, whereas neural networks and support vector machines find more heterogeneity. The notable differences in effect estimates across the candidates algorithms highlight the potential limitations of relying on a single CATE-based candidate estimator of the policy rule, as opposed to an ensemble.

Figure 4 plots sorted GATEs for quintiles of CATEs estimated by each threshold-based candidate estimator. Differences in sorted GATEs between the lowest quintile (Q1) and the higher quintiles (Q2-Q5) are reported in Table B.1. Our findings support those from Figure 3, in that

[27] See Figure B.1 for a density plot of the predicted propensity scores.

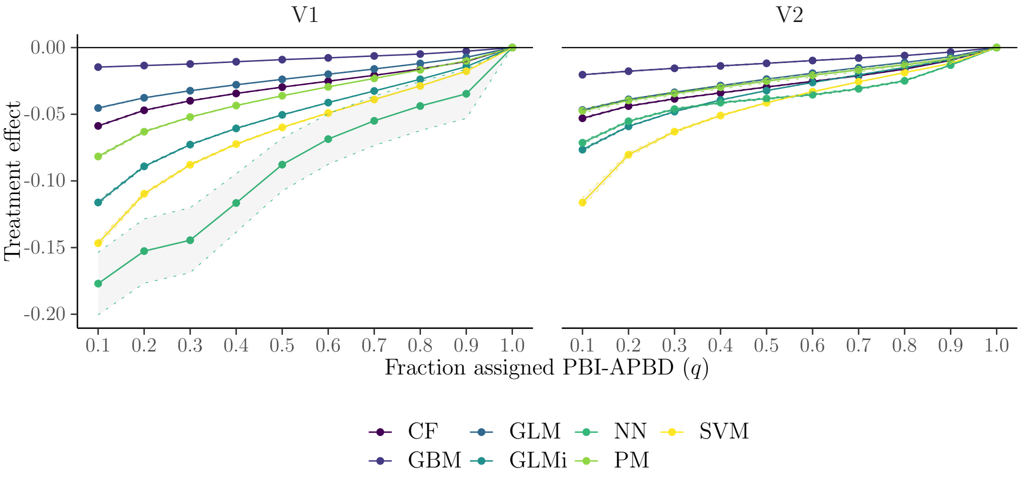
Figure 2: Covariate balance between households assigned to PBI-APBD and PBI-APBN.



Note: Standardised mean differences between the "treated" (PBI-APBD) and the "controls" (PBI-APBN) are reported in the unweighted untrimmed sample, unweighted trimmed sample, and weighted trimmed sample (using inverse probability of treatment weights for the ATE). HoH = head of household. HH = household. See Appendix A for further details on the included covariates.

there is evidence of heterogeneity in treatment effects. An evaluation that focuses solely on overall ATEs would conclude that there are limited impacts of assigning APBD over APBN, and the evident variation in impacts would be missed. The sorted GATEs tell us that households in the higher quintiles (predominantly Q4 and Q5) have a lower risk of suffering catastrophic health expenditures from receiving APBD over APBN, while those in the lower quintiles (Q1 and Q2) are at a higher risk. In summary, our heterogeneity analysis provides justification for

Figure 3: Targeting Operator Characteristic (TOC) curve



Note: TOC curve plots the cumulative estimated ATEs on catastrophic expenditure from assigning PBI-APBD (compared to PBI-APBN) to increasing fractions of the population, ranked according to the estimated CATEs from threshold-based candidate estimators of the policy rule $\hat{\tau}_j(X_i)$. Separate TOC curves are displayed for CATE estimators that consider $V1$ and $V2$.

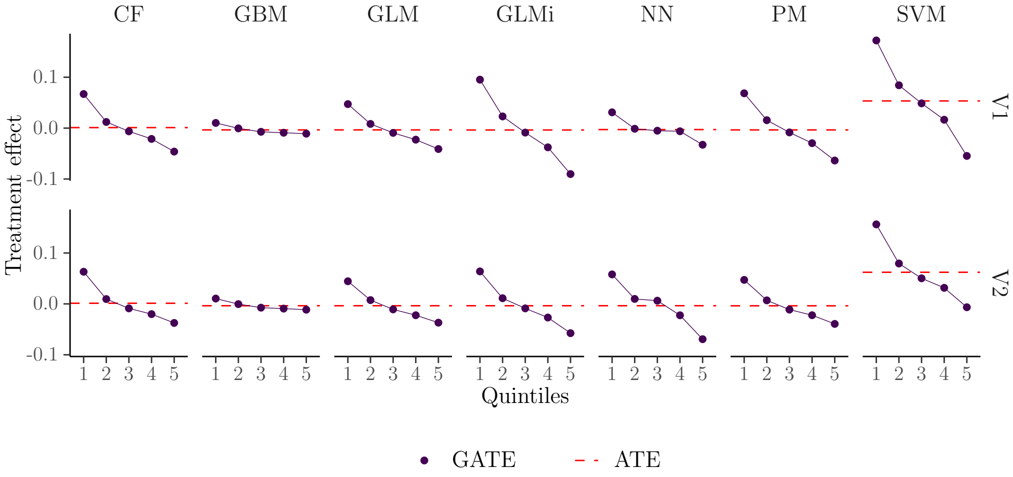
learning CATE-based policy rules.

6.4 Evaluating the estimated policy rules

Figure 5 presents point estimates and 95% confidence intervals of the counterfactual mean outcomes under the super learner estimates of the optimal policy rules and the static rules, in comparison to the mean outcome under the actual policy assignment D_i of 0.127 (95% CI +/- 0.003). Using our primary performance measure, the AIPTW estimator of the mean value of the learned policy, we find that the unconstrained rule has the best performance, with an estimated mean outcome of 0.115 (95% CI +/- 0.006). The constrained rule performs slightly worse with an estimated mean outcome of 0.125 (95% CI +/- 0.005), which is still better than the actual assignment. Assigning all households to APBD generates a mean outcome of 0.123 (95% CI +/- 0.007), and assigning all to APBN gives 0.127 (95% CI +/- 0.004). The TMLE estimator supports our primary findings, with very similar point estimates and confidence intervals.

Table B.2 presents estimates of the counterfactual mean outcomes for the candidate estimators included in the super learner. We find that the majority of our algorithms, including simple linear models, generalised boosted models, causal forests, regression splines and policy trees (of all depths), perform particularly well with estimates in the range of 0.114-0.118. Overall, candidates rules that are based on simpler or restricted functional forms (for example, linear models and and shallow policy trees) outperform those based on more complex models (for example,

Figure 4: Estimated sorted GATEs from threshold-based candidate estimators of the CATE

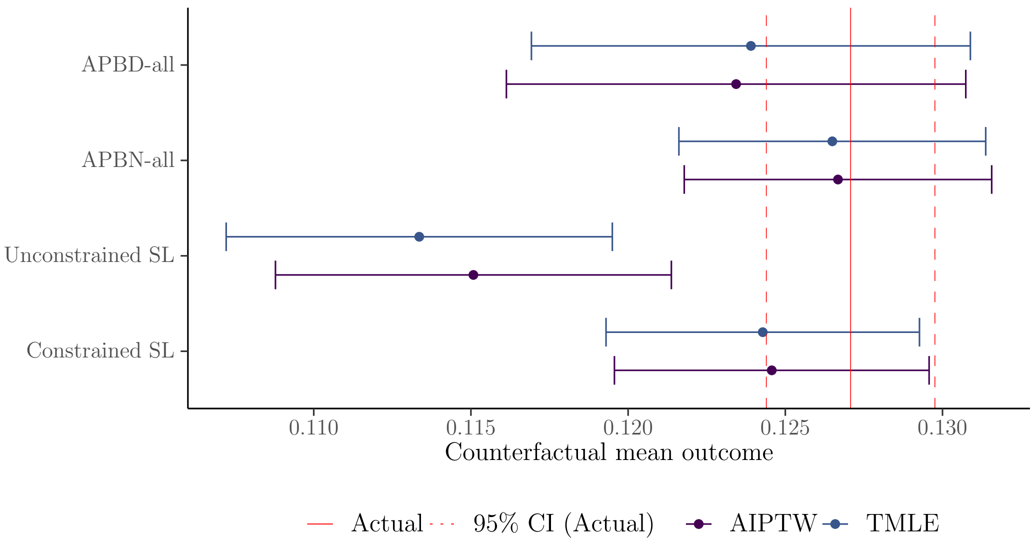


Note: Sorted GATEs are the estimated ATEs for each quintile of the population, ranked (in descending order) by predicted CATEs. The bottom quintile (Q1) denotes the least affected population subgroup (i.e. households with the smallest CATEs), and the top quintile (Q5) denotes the most affected subgroup (i.e. households with the largest CATEs). The red dashed line is the candidate-specific ATE estimate. Separate plots are displayed for CATE estimators that consider $V1$ and $V2$.

neural networks and support vector machines). Figure B.3 presents the average weighted contributions of the candidate estimators across the sample-specific super learner estimates of the policy rules. Causal forests (fitted on $V1$) and generalised linear models (fitted on $V2$) receive the largest weightings in the respective unconstrained and constrained rules.

Table 4 presents results from the classification analysis of the learned subgroups under the super learner estimates of the optimal policy rule for selected variables in X. We compare the characteristics of households under the optimal policy to those under the actual assignment. Compared to households actually assigned to APBD, those that are assigned to APBD under the unconstrained rule are more likely to be urban-based, where health care facilities are even more accessible. They are slightly less likely to reside in the regions of Maluku-Papua and Sumatera, where the current APBD allocation is concentrated, but in Sulawesi, Bali, and Java. The opposite is true for APBN, in that compared to the actual assignment, the unconstrained rule assigns APBN to less urban households, particularly in the regions of Sumatera and Maluku-Papua, where health care is less accessible. There are also some important socioeconomic differences that strengthen this urban-rural distinction. Compared to the actual assignment, households assigned to APBD under the optimal policy rule have more educated members, their homes are more likely to have basic amenities and assets, and they are also more likely to access technology (for example, internet and cellphones). Households assigned to APBN under the rule are slightly worse off in terms of the same features. With regard to demography, APBD-

Figure 5: Estimated counterfactual mean outcomes



Note: Point estimates and 95% confidence intervals (CI) are reported for the AIPTW and TMLE estimators. APBD-all (APBN-all) is the static rule that assigns APBD (APBN) to everyone. Unconstrained (constrained) SL is the super learner estimate of the unconstrained (constrained) optimal policy rule. The red solid line denotes the mean outcome under the actual policy assignment D_i.

assigned households under the rule are smaller, with heads of households that are older (aged 45 and over) and more likely to be employed in the secondary and tertiary sectors, compared to the current assignment. On the contrary, APBN-assigned households under the rule are larger, with heads of households that are younger (aged below 45) and working in the primary sector. Under the constrained rule, we find similar differences between households that are assigned to the two schemes compared to the actual assignment. However, certain differences in characteristics are more pronounced. For example, households that are assigned to APBD under the optimal policy rule are even more likely to be educated and have access to technology, while the opposite is true for households that are assigned to APBN.

Figures B.4-B.9 plot the learned policy trees that are candidates in the super learner estimate of the unconstrained optimal policy. The splitting covariates support our findings from the subgroup analyses that households assigned to APBD under the optimal policy rule are socioeconomically richer than those assigned to APBN, with better household amenities (for example, protected drinking water, and cooking and toilet facilities) and assets (for example, vehicles and cellphones). The shallower trees (of depth 1) only split on the availability of household amenities, while the deeper trees (of depths 1 and 2) introduce characteristics associated with household members (for example, the household size and the number of members in work or with education).

7 Discussion

In this paper, we addressed the assignment of Indonesia's two subsidised health insurance schemes to eligible households, considering the perspective of the state government. We learned optimal policy rules that assign households to either treatment (PBI-APBD) or control (PBI-APBN) to reduce the expected probability of households incurring catastrophic health. Using a standard 10% threshold of household income for catastrophic expenditure, we were able to gauge the impact on households' enrolment to APBD versus APBN on their financial outcomes. Although average treatment effects appear small, the heterogeneous treatment effects indicate that utility gains could be maximised with CATE-based policy rules, estimated using the super learner.

Our unconstrained optimal policy rule, which targets specific covariates, achieves a lower expected risk of catastrophic expenditure than both the actual assignment and static rules. Applying a 10% budget constraint, our constrained policy outperforms the actual assignment and the static rule that assigns APBN to all. Although the utility gains under the estimated optimal policy rules appear modest in percentage terms, the reduction in the total number of uninsured households facing catastrophic health expenditure compared to the actual assignment strategy could be substantial at the population level. For example, in a population of approximately 245 million with 22.5% uninsured (=55 million) (as reported by Mahendradhata et al. (2017)), the relative reduction in the number of uninsured facing catastrophic health expenditure is approximately 660,000 for the unconstrained rule and 110,000 for the constrained rule.

One of the main differentiating features between the assignment strategy under the estimated rules and the actual assignment is geography, particularly the urban-rural distinction, which is known to be linked to the availability of health services, and a key determinant of health spending (Agustina et al., 2019; Johar et al., 2018; Sambodo et al., 2021). An optimal policy rule that shifts a small proportion of urban enrollees from APBN to APBD could improve outcomes, assuming that this strategy is chosen over one that improves the availability of health services in less-developed regions. Socioeconomic differences are also present within the assignment strategy. Households that are assigned to APBD under the optimal policy rule are more likely to be better off than those assigned to APBN, in terms of characteristics that can be easily verified by decision makers (for example, whether the household has electricity, toilet facilities or a vehicle). The wider goal of JKN is to achieve full population coverage, which would involve assigning PBI to eligible uninsured households based on their observed covariates. Differences in characteristics between households that are assigned to APBN and APBD under the optimal policy rule, which can be identified from our descriptive statistics and the visual decision trees, could be used to match uninsured households to the appropriate PBI scheme.

Our study has some limitations. We assume no unobserved confounding, given our rich set of included controls, and rely on a doubly robust adjustment for observed confounding to estimate our target causal parameter. We previously discussed our justification for restricting the sample to households that are enrolled into either PBI scheme, and excluding uninsured households. By controlling for head-of household characteristics, household assets and health care service accessibility, we try to minimise unobserved confounding that may explain the process of selection into APBD or APBN, which can create bias in our estimates of the treatment effects, and consequently, the policy rules. We also make explicit the fact that we trim observations with extreme propensity scores, which are becoming more common in larger datasets. We acknowledge that trimming methods are sensitive to the pre-defined cut-off points, and can result in a substantial sample size reduction (Li et al., 2019). We also apply a 10% budget constraint to reflect the national government healthcare funding allocation to local governments, which assists in modelling realistic constraints but may not fully capture additional funding sources for APBD. For a representative policy rule, further consultation would be necessary to understand precise budgetary restrictions on PBI allocations. Furthermore, our choice to define catastrophic expenditure using a 10% threshold of household income – while standard (Wagstaff et al., 2018) – may not fully capture financial strain across income levels or regions. Although we did not perform a sensitivity analysis with alternative thresholds (e.g. 5%, 15%), future studies could explore this approach to assess the robustness of policy implications. Lastly, in our super learner library of candidate estimators, we include rules that are based on the CATE function (threshold- and tree-based rules) and static rules. In theory, other candidate estimators that do not rely on estimating the CATE function could be included.

Potential criticisms of relying on ensembling tools for policy learning are that their underlying methods are based on a black-box, meaning that the exact contribution of each covariate to prediction is unclear. The use of machine learning in policy decisions has raised some concerns about their potential ethical and equity implications (Kube et al., 2019). Hence, the learned rule may need to be constrained to belong to a policy class that imposes restrictions on the functional form, for example, for added interpretability (Kitagawa and Tetenov, 2018; Zhou et al., 2022). In general, the requirement for human interpretability to justify policy decisions could result in a sub-optimal rule in terms of the objective function. If the expected loss in utility between an optimal, uninterpretable rule and a nearly-optimal, interpretable rule is small, decision makers may choose to opt for the latter. Learning optimal policies using the super learner offers a structured solution to addressing the trade-off between optimality and interpretability, as the candidate library can include a diverse set of estimators according to the decision maker's preferences. In our assignment problem, we find that a policy that belongs to a class of interpretable and transparent rules (for example, simple linear models and shallow decision trees) may achieve similar or even better outcomes in finite samples than the black-box ensemble. Further, the comparable performance of candidate estimators that consider $V1$ and $V2$ also implies that a policy rule that targets only the most important predictors of treatment effect heterogeneity, which we identify using regularisation on the CATE model, can also produce an interpretable solution. Since our policy rule is based on decision criteria that were curated by policymakers, and we can, to some extent, explain the construction of the rule, we hope that the potential impacts of algorithmic bias on our findings are reduced (Panch et al., 2019).

Learning policy rules can be time and resource-intensive. A well-constructed rule relies on gathering sufficient data and having an appropriately skilled workforce to conduct the analyses. Sustaining utility gains would require regularly updating the rule using new information that reflects any changes in covariate distributions, particularly if these changes are in response to the rule itself. However, this is unlikely in our empirical problem given the relatively comparable characteristics between treated and control populations.

Our findings highlight the potential of combining ex-ante and ex-post targeting measures to improve the delivery of social programmes. Evaluating the progress of implemented policies, through an assessment of impact heterogeneity, could strengthen policy corrections and guide future decisions, by potentially identifying a superior assignment strategy that achieves larger expected benefits. We acknowledge that policymaking at the national level is incredibly complex. Health care may be one of several objectives of the state, so aligning targeting criteria across the social welfare function is important. Improving equity is a policy priority in most health care systems, including Indonesia (Johar et al., 2018). We do not directly encode equity constraints into our objective function, although in theory this could be possible with some algorithmic modifications. For example, an equity constraint could assign centrally-funded insurance to all households that receive social security, regardless of their observed covariates. We do, however, include proxies for socioeconomic status in our pre-specified targeting criteria that could indirectly address equity considerations. Since our learned policy trees define an assignment strategy that is primarily based on socioeconomic disparities, policymakers could use these to guide fairness decisions.

8 Main conclusions

This study demonstrates the potential of data-driven methods to optimise the assignment of Indonesia's subsidised health insurance schemes. The proposed optimal policy rules can significantly reduce catastrophic health expenditure, with substantial benefits at the population level. By targeting specific geographic and socioeconomic groups, the rules could enhance equity in health insurance access. While real-world implementation is beyond the scope of this study, the findings provide a basis for future discussions with policymakers on adopting data-driven approaches to improve health insurance allocation and outcomes.

Table 4: Classification analysis of selected variables in X, for households assigned to PBI-APBN and PBI-APBD under the learned policy rules, estimated using the unconstrained and resource-constrained super learners.

	Actual assignment		**Unconstrained rule**				**Constrained rule**			
	APBN (n=39,367)	APBD (n=22,375)	APBN (n=23,598)		APBD (n=38,144)		APBN (n=58,168)		APBD (n=3,574)	
	Mean	Mean	Mean	Diff	Mean	Diff	Mean	Diff	Mean	Diff
Head of household characteristics										
Male	0.814	0.850	0.828	0.013	0.827	-0.023	0.827	0.012	0.834	-0.016
Age 0-24	0.013	0.018	0.015	0.002	0.015	-0.003	0.015	0.002	0.013	-0.006
Age 25-44	0.307	0.391	0.340	0.033	0.336	-0.055	0.338	0.030	0.341	-0.051
Age 45-64	0.508	0.461	0.487	-0.021	0.494	0.033	0.492	-0.017	0.487	0.025
Age 65+	0.154	0.114	0.140	-0.014	0.139	0.025	0.139	-0.015	0.144	0.030
Married	0.770	0.805	0.781	0.011	0.783	-0.021	0.782	0.012	0.791	-0.014
In employment	0.855	0.892	0.867	0.012	0.869	-0.023	0.869	0.014	0.862	-0.030
Employment: primary sector	0.437	0.524	0.469	0.032	0.469	-0.055	0.469	0.032	0.457	-0.067
Employment: secondary sector	0.063	0.047	0.058	-0.005	0.057	0.010	0.057	-0.006	0.056	0.010
Employment: tertiary sector	0.354	0.321	0.340	-0.014	0.344	0.022	0.342	-0.012	0.348	0.027
Literate: Latin/Arabic letters	0.914	0.907	0.913	-0.001	0.910	0.004	0.911	-0.003	0.911	0.004
Compulsory education	0.684	0.683	0.683	-0.001	0.684	0.001	0.684	-0.001	0.687	0.004
Had a cellphone in previous 3 months	0.585	0.610	0.594	0.009	0.593	-0.016	0.594	0.009	0.587	-0.023
Used internet in previous 3 months	0.117	0.106	0.116	-0.001	0.112	0.005	0.114	-0.003	0.107	0.001
Household characteristics										
Number of members	3.538	3.693	3.590	0.053	3.596	-0.097	3.593	0.055	3.618	-0.075
Number of productive members (aged 15-64)	2.366	2.340	2.358	-0.008	2.356	0.015	2.356	-0.010	2.365	0.025
Number of children in school	0.394	0.352	0.376	-0.018	0.381	0.029	0.378	-0.016	0.388	0.037
Number of members in employment	1.684	1.695	1.689	0.004	1.688	-0.007	1.688	0.004	1.693	-0.002
Number of members with compulsory education	2.211	2.136	2.179	-0.032	2.187	0.051	2.182	-0.029	2.218	0.082
Number of rooms	5.851	5.796	5.836	-0.014	5.828	0.031	5.830	-0.021	5.845	0.049
Location: urban	0.372	0.315	0.349	-0.023	0.352	0.038	0.351	-0.021	0.355	0.041
Household asset ownership										
Electricity	0.958	0.908	0.942	-0.016	0.938	0.031	0.940	-0.018	0.935	0.027
Purchases drinking water	0.596	0.559	0.586	-0.010	0.581	0.022	0.583	-0.014	0.580	0.021
Private/shared toilet	0.794	0.790	0.793	-0.001	0.792	0.003	0.792	-0.002	0.804	0.015
Refrigerator	0.407	0.422	0.408	0.001	0.415	-0.007	0.413	0.006	0.407	-0.015
Gas canister >5.5kg	0.062	0.053	0.059	-0.003	0.058	0.005	0.058	-0.003	0.059	0.006
Car/motorcycle	0.638	0.660	0.648	0.009	0.646	-0.015	0.646	0.007	0.652	-0.008
Gold/jewellery >10g	0.131	0.139	0.135	0.005	0.133	-0.006	0.134	0.003	0.131	-0.007
District characteristics										
Easy access: secondary care	0.793	0.681	0.755	-0.038	0.750	0.069	0.752	-0.040	0.750	0.069
Easy access: community care	0.979	0.924	0.960	-0.019	0.958	0.034	0.959	-0.020	0.957	0.034
Easy access: primary care	0.887	0.815	0.866	-0.021	0.858	0.043	0.861	-0.026	0.857	0.042
Easy access: maternity care	0.867	0.781	0.839	-0.028	0.834	0.052	0.836	-0.031	0.828	0.047
Region										
Sumatera	0.311	0.360	0.328	0.017	0.329	-0.030	0.329	0.018	0.317	-0.042
Jakarta	0.003	0.000	0.002	-0.001	0.002	0.002	0.002	-0.001	0.003	0.003
Java	0.312	0.270	0.296	-0.015	0.297	0.027	0.296	-0.016	0.306	0.036
Bali,NTB,NTT	0.069	0.030	0.055	-0.014	0.054	0.024	0.054	-0.015	0.064	0.034
Kalimantan	0.072	0.088	0.079	0.007	0.077	-0.011	0.078	0.006	0.073	-0.015
Sulawesi	0.165	0.075	0.135	-0.030	0.131	0.056	0.133	-0.032	0.129	0.054
Maluku-Papua	0.069	0.177	0.104	0.036	0.110	-0.067	0.108	0.039	0.108	-0.069

Note: Sample means are reported for selected variables included in X for households in the trimmed 70% data partition that are assigned to PBI-APBD and PBI-APBN under the actual assignment, and the optimal unconstrained and constrained policy rules (estimated using the super learner). The absolute differences ("Diff") in covariate means for households that are assigned to APBN and APBD under the estimated optimal policy rules compared to the actual assignment are also reported. Coloured cells denote the p-values from Welch's two-sample t-test on whether these differences in covariate means are significant. Two-sample t-test is conducted at a significance level of 0.05. If p-value on t-statistic <0.05: cell colour is green; white otherwise.

References

Agustina, R., Dartanto, T., Sitompul, R., Susiloretni, K. A., Suparmi, Achadi, E. L., Taher, A., Wirawan, F., Sungkar, S., Sudarmono, P., Shankar, A. H., Thabrany, H., and Indonesian Health Systems Group (2019). Universal health coverage in Indonesia: concept, progress, and challenges. *Lancet*, 393(10166):75–102.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., Purnamasari, R., and Wai-Poi, M. (2016). Self-targeting: Evidence from a field experiment in indonesia. *J. Polit. Econ.*, 124(2):371–427.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., and Tobias, J. (2012). Targeting the poor: evidence from a field experiment in Indonesia. *Am. Econ. Rev.*, 102(4):1206–1240.

Amram, M., Dunn, J., and Zhuo, Y. D. (2022). Optimal policy trees. *Mach. Learn.*, 111(7):2741–2768.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *Ann. Stat.*, 47(2):1148–1178.

Athey, S. and Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89(1):133–161.

Badan Pusat Statistik (2017). Survei sosial ekonomi nasional (SUSENAS) 2017.

Badan Pusat Statistik (2018). Village potential statistics (PODES) 2018.

Bahamyirou, A., Schnitzer, M. E., Kennedy, E. H., Blais, L., and Yang, Y. (2022). Doubly robust adaptive LASSO for effect modifier discovery. *Int. J. Biostat.*, 18(2):307–327.

Bernal, N., Carpio, M. A., and Klein, T. J. (2017). The effects of access to health insurance: evidence from a regression discontinuity design in Peru. *J. Public Econ.*, 154:122–136.

Bertsimas, D., Dunn, J., and Mundru, N. (2019). Optimal prescriptive trees. *INFORMS Journal on Optimization*, 1(2):164–183.

Bhattacharya, D. and Dupas, P. (2012). Inferring welfare maximizing treatment assignment under budget constraints. *J. Econom.*, 167(1):168–196.

Bookwalter, J. T., Fuller, B. S., and Dalenberg, D. R. (2006). Do household heads speak for the household? A research note. *Soc. Indic. Res.*, 79(3):405–419.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018a). Double/debiased machine learning for treatment and structural parameters. *Econom. J.*, 21(1):C1–C68.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2018b). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. Technical Report 24678, National Bureau of Economic Research.

Coyle, J. R. (2017). *Computational Considerations for Targeted Learning.* PhD thesis. University of California, Berkeley.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1):187–199.

Dartanto, T., Halimatussadiah, A., Rezki, J. F., Nurhasana, R., Siregar, C. H., Bintara, H., Usman, Pramono, W., Sholihah, N. K., Yuan, E. Z. W., and Soeharno, R. (2020). Why do informal sector workers not pay the premium regularly? Evidence from the national health insurance system in Indonesia. *Appl. Health Econ. Health Policy*, 18(1):81–96.

Dehejia, R. H. (2005). Program evaluation as a decision problem. *J. Econom.*, 125(1):141–173.

Dudoit, S. and van der Laan, M. J. (2005). Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Stat. Methodol.*, 2(2):131–154.

Dutta, A., Ward, K., Setiawan, E., and Prabhakaran, S. (2020). Fiscal space for health in Indonesia: public sector opportunities and constraints in achieving the goals of Indonesia's mid-term development plan (RPJMN) 2020–2024. Technical report, Jakarta: Kementerian PPN/Bappenas.

El-Sayed, A. M., Vail, D., and Kruk, M. E. (2018). Ineffective insurance in lower and middle income countries is an obstacle to universal health coverage. *J. Glob. Health*, 8(2):020402.

Erlangga, D., Suhrcke, M., Ali, S., and Bloor, K. (2019). The impact of public health insurance on health care utilisation, financial protection and health status in low- and middle-income countries: A systematic review. *PLoS One*, 14(11):e0225237.

Fink, G., Robyn, P. J., Sié, A., and Sauerborn, R. (2013). Does health insurance improve health?: Evidence from a randomized community-based insurance rollout in rural Burkina Faso. *J. Health Econ.*, 32(6):1043–1056.

Finkelstein, A. and Notowidigdo, M. J. (2019). Take-up and targeting: Experimental evidence from SNAP. *Q. J. Econ.*, 134(3):1505–1556.

Grogger, J., Arnold, T., León, A. S., and Ome, A. (2015). Heterogeneity in the effect of public health insurance on catastrophic out-of-pocket health expenditures: the case of mexico. *Health Policy Plan.*, 30(5):593–599.

Hahn, R. P., Murray, J. S., and Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Anal.*, 15(3):965–1056.

Hanna, R. and Olken, B. A. (2018). Universal basic incomes versus targeted transfers: Anti-poverty programs in developing countries. *J. Econ. Perspect.*, 32(4):201–226.

Harimurti, P., Pambudi, E., Pigazzini, A., and Tandon, A. (2013). The nuts and bolts of Jamkesmas: Indonesia's government-financed health coverage program for the poor and near-poor. Technical report, World Bank.

Hatamyar, J. and Kreif, N. (2023). Policy learning with rare outcomes. *arXiv [econ.EM]*.

Johar, M., Soewondo, P., Pujisubekti, R., Satrio, H. K., and Adji, A. (2018). Inequality in access to health care, health insurance and the role of supply factors. *Soc. Sci. Med.*, 213:134–145.

Kennedy, E. H. (2023). Towards optimal doubly robust estimation of heterogeneous causal effects. *Electron. J. Stat.*, 17(2).

Kitagawa, T. and Tetenov, A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616.

Kruse, I., Pradhan, M., and Sparrow, R. (2012). Marginal benefit incidence of public health spending: evidence from Indonesian sub-national data. *J. Health Econ.*, 31(1):147–157.

Kube, A., Das, S., and Fowler, P. J. (2019). Allocating interventions based on predicted outcomes: a case study on homelessness services. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33 of *33*, pages 622–629.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl. Acad. Sci. U. S. A.*, 116(10):4156–4165.

Li, F., Thomas, L. E., and Li, F. (2019). Addressing extreme propensity scores via the overlap weights. *Am. J. Epidemiol.*, 188(1):250–257.

Limwattananon, S., Tangcharoensathien, V., Tisayaticom, K., Boonyapaisarncharoen, T., and Prakongsai, P. (2012). Why has the universal coverage scheme in thailand achieved a pro-poor public subsidy for health care? *BMC Public Health*, 12 Suppl 1(S1):S6.

Luedtke, A. R. and van der Laan, M. J. (2016a). Optimal individualized treatments in resource-limited settings. *Int. J. Biostat.*, 12(1):283–303.

Luedtke, A. R. and van der Laan, M. J. (2016b). Super-learning of an optimal dynamic treatment rule. *Int. J. Biostat.*, 12(1):305–332.

Mahendradhata, Y., Trisnantoro, L., Listyadewi, S., Soewondo, P., Marthias, T., Harimurti, P., and Prawira, J. (2017). The Republic of Indonesia health system review. Technical report, WHO Regional Office for South-East Asia.

Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246.

Mardiansjah, F. H., Rahayu, P., and Rukmana, D. (2021). New patterns of urbanization in Indonesia: emergence of non-statutory towns and new extended urban regions. *Environment and Urbanization ASIA*, 12(1):11–26.

Maulana, N., Soewondo, P., Adani, N., Limasalle, P., and Pattnaik, A. (2022). How Jaminan Kesehatan Nasional (JKN) coverage influences out-of-pocket (OOP) payments by vulnerable populations in Indonesia. *PLOS Global Public Health*, 2(7).

Montoya, L., Skeem, J., van der Laan, M. J., and Petersen, M. (2021). Performance and application of estimators for the value of an optimal dynamic treatment rule. https://arxiv.org/abs/2101.12333. Accessed: 2023-2-5.

Montoya, L. M., van der Laan, M. J., Luedtke, A. R., Skeem, J. L., Coyle, J. R., and Petersen, M. L. (2023). The optimal dynamic treatment rule superlearner: considerations, performance, and application to criminal justice interventions. *Int. J. Biostat.*, 19(1):217–238.

National Team for the Acceleration of Poverty Reduction (Indonesia) (2015). The road to national health insurance (JKN). Technical report.

Newey, W. K. and Robins, J. R. (2018). Cross-fitting and fast remainder rates for semiparametric estimation. https://arxiv.org/abs/1801.09138. Accessed: 2023-2-5.

Nie, X. and Wager, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319.

Panch, T., Mattie, H., and Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. *J. Glob. Health*, 9(2):010318.

Phillips, R. V., van der Laan, M. J., Lee, H., and Gruber, S. (2022). Practical considerations for specifying a super learner. https://arxiv.org/abs/2204.06139. Accessed: 2023-2-5.

Polley, E. C. and van der Laan, M. J. (2010). Super learner in prediction. Technical report, University of California, Berkeley.

Prabhakaran, S., Dutta, A., Fagan, T., and Ginivan, M. (2019). Financial sustainability of Indonesia's Jaminan Kesehatan Nasional: Performance, prospects, and policy options. Technical report, Washington, DC: Palladium, Health Policy Plus, and Jakarta, Indonesia: National Team for the Acceleration of Poverty Reduction (TNP2K).

Pratiwi, A. B., Setiyaningsih, H., Kok, M. O., Hoekstra, T., Mukti, A. G., and Pisani, E. (2021). Is Indonesia achieving universal health coverage? Secondary analysis of national data on insurance coverage, health spending and service availability. *BMJ Open*, 11(10):e050565.

Sambodo, N. P., Van Doorslaer, E., Pradhan, M., and Sparrow, R. (2021). Does geographic spending variation exacerbate healthcare benefit inequality? A benefit incidence analysis for indonesia. *Health Policy Plan.*, 36(7):1129–1139.

Shalit, U., Johansson, F. D., and Sontag, D. (2017). Estimating individual treatment effect: generalization bounds and algorithms. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3076–3085. PMLR.

Sparrow, R., Budiyati, S., Yumna, A., Warda, N., Suryahadi, A., and Bedi, A. S. (2017). Sub-national health care financing reforms in Indonesia. *Health Policy Plan.*, 32(1):91–101.

Stürmer, T., Webster-Clark, M., Lund, J. L., Wyss, R., Ellis, A. R., Lunt, M., Rothman, K. J., and Glynn, R. J. (2021). Propensity score weighting and trimming strategies for reducing variance and bias of treatment effect estimates: a simulation study. *Am. J. Epidemiol.*, 190(8):1659–1670.

Sverdrup, E., Kanodia, A., Zhou, Z., Athey, S., and Wager, S. (2020). policytree: Policy learning via doubly robust empirical welfare maximization over trees. *J. Open Source Softw.*, 5(50):2232.

van der Laan, M. J. and Dudoit, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: finite sample oracle inequalities and examples. U.C. Berkeley Division of Biostatistics Working Paper Series.

van der Laan, M. J., Dudoit, S., and van der Vaart, A. W. (2006). The cross-validated adaptive epsilon-net estimator. *Statist. Decisions*, 24(3).

van der Laan, M. J. and Luedtke, A. R. (2014). Targeted learning of an optimal dynamic treatment, and statistical inference for its mean outcome. U.C. Berkeley Division of Biostatistics Working Paper Series.

van der Laan, M. J. and Luedtke, A. R. (2015). Targeted learning of the mean outcome under an optimal dynamic treatment rule. *J Causal Inference*, 3(1):61–95.

van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.*, 6(1).

van der Laan, M. J. and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer Series in Statistics. Springer Publishing.

van der Laan, M. J. and Rose, S. (2018). *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Series in Statistics. Springer Publishing.

van der Vaart, A. W., Dudoit, S., and van der Laan, M. J. (2006). Oracle inequalities for multi-fold cross validation. *Statist. Decisions*, 24(3):351–371.

Vilcu, I., Probst, L., Dorjsuren, B., and Mathauer, I. (2016). Subsidized health insurance coverage of people in the informal sector and vulnerable population groups: trends in institutional design in Asia. *Int. J. Equity Health*, 15(1):165.

Wagstaff, A., Flores, G., Hsu, J., Smitz, M.-F., Chepynoga, K., Buisman, L. R., van Wilgenburg, K., and Eozenou, P. (2018). Progress on catastrophic health spending in 133 countries: a retrospective observational study. *Lancet Glob Health*, 6(2):e169–e179.

Wagstaff, A. and Lindelow, M. (2008). Can insurance increase financial risk? The curious case of health insurance in China. *J. Health Econ.*, 27(4):990–1005.

Wagstaff, A., O'Donnell, O., van Doorslaer, E., and Lindelow, M. (2007). *Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and their Implementation.* World Bank Publications.

Wagstaff, A. and van Doorslaer, E. (2003). Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993-1998. *Health Econ.*, 12(11):921–934.

World Health Organization (2010). Health systems financing: The path to universal coverage. Technical report, World Health Organization.

Xu, K., Evans, D. B., Kawabata, K., Zeramdini, R., Klavus, J., and Murray, C. J. L. (2003). Household catastrophic health expenditure: a multicountry analysis. *Lancet*, 362(9378):111–117.

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., and Wager, S. (2021). Evaluating treatment prioritization rules via rank-weighted average treatment effects. https://arxiv.org/abs/2111.07966. Accessed: 2023-2-5.

Zhou, Z., Athey, S., and Wager, S. (2022). Offline multi-action policy learning: generalization and optimization. *Oper. Res.*, 71(1):148–183.

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.*, 101(476):1418–1429.

A List of all covariates used for confounder adjustment

Head of household-level (binary)

- Male
- Marital status: married
- Age 0-24
- Age 25-44
- Age 45-64
- Age 65+
- Has a national identity number
- Literate: Latin/Arabic letters
- Educated (at the compulsory level)
- Travelled domestically for tourism in 2016
- Had a cellphone in previous 3 months
- Used internet in previous 3 months
- Employment status: in employment
- Employment sector: primary
- Employment sector: secondary
- Employment sector: tertiary

Household-level (count)

- Educated (at the compulsory level)
- Employment status: in employment
- Productive members (aged 15-64)
- Children at school
- Size (members)
- Size (families)
- Rooms

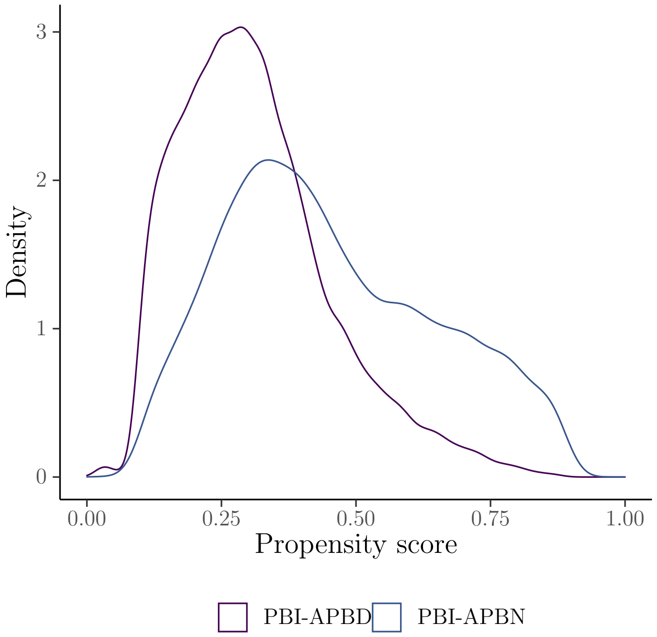
Household-level (binary)

- Location: urban area
- Home occupancy status: owner
- Has a second residence
- Roof material: concrete/tile
- Wall material: concrete
- Floor material: marble/granite/ceramic/parquet/vinyl/carpet

Toilet facility: private/shared
Protected drinking water source
Purchases drinking water
Electricity
Cooking fuel: gas/electric
Sewage disposal: septic tank/sewage system
Experienced a natural diaster in previous year
Natural tourism in residential area
Savings account
Goods ownership: gas (over 5.5kg)
Goods ownership: refrigerator
Goods ownership: air conditioning
Goods ownership: radiator
Goods ownership: landline
Goods ownership: computer
Goods ownership: gold (over 10g)
Goods ownership: boat
Goods ownership: car/motorcycle
Goods ownership: television
Goods ownership: land
District-level (binary)
Easy access to primary health care
Easy access to community health care
Easy access to maternal health care
Easy access to secondary (hospital) health care
Easy access to pharmacy
Regional-level (binary)
Region: Sumatera
Region: Jakarta
Region: Jawa
Region: Bali, NTB, NTT
Region: Kalimantan
Region: Sulawesi
Region: Maluka-Papua

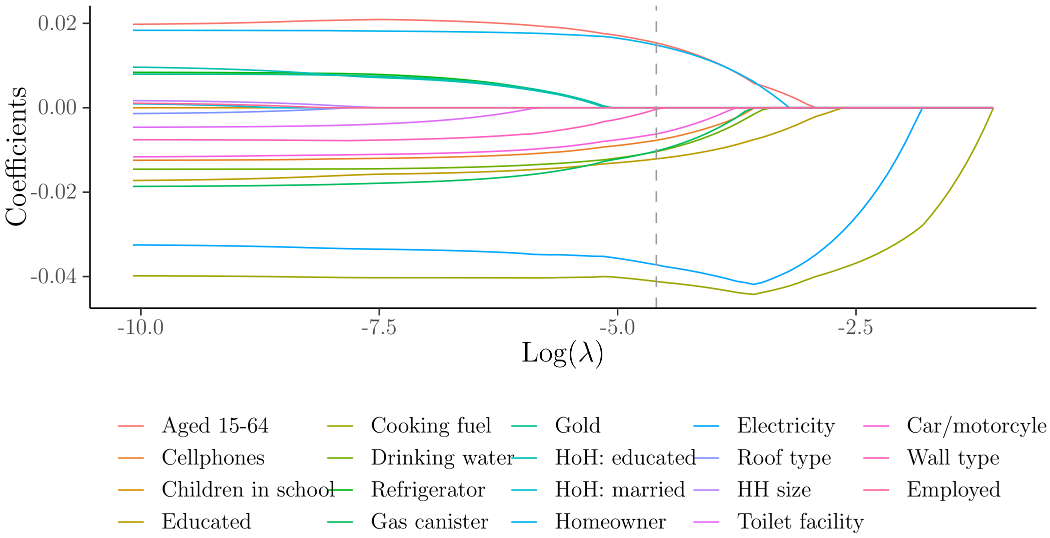
B Additional figures/tables

Figure B.1: Overlap plot



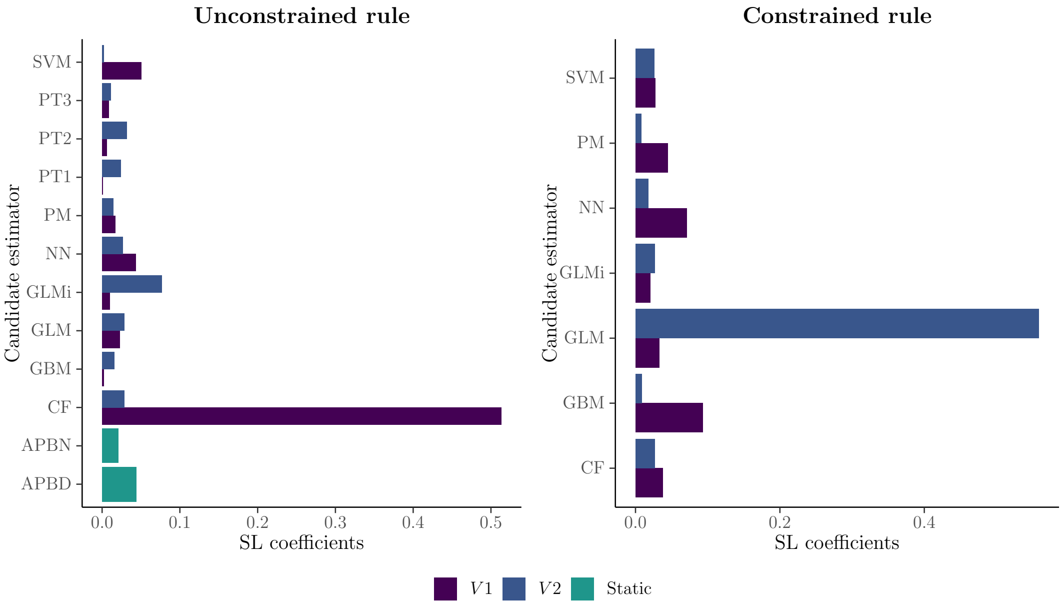
Note: Density plot showing the distribution of propensity scores for households enrolled into PBI-APBD and PBI-APBN in the trimmed 70% data partition.

Figure B.2: Variable trace plot of adaptive LASSO fit



Note: Plot shows the coefficients on $V1$, as a function of the log(λ) values used in the cross-validated adaptive LASSO model. The grey dashed line represents the value of log(λ) that minimises the cross-validated mean squared error. See Table 2 for a detailed description of the covariates in $V1$. HoH = head of household.

Figure B.3: Weighted contributions of candidate estimators to the super learner



Note: Average contribution of each candidate estimator in the super learner across cross-fitted samples are displayed. Static rules are not included in the candidate library for the constrained rule.

Table B.1: Results from heterogeneity test using difference-in-means estimator

	V1				V2			
	Est	SE	Unadj *p*-val	Adj *p*-val	Est	SE	Unadj *p*-val	Adj *p*-val
CF								
Q2-Q1	-0.055	0.000	0.000	0.000	-0.054	0.000	0.000	0.000
Q3-Q1	-0.073	0.000	0.000	0.000	-0.072	0.000	0.000	0.000
Q4-Q1	-0.088	0.000	0.000	0.000	-0.083	0.000	0.000	0.000
Q5-Q1	-0.113	0.000	0.000	0.000	-0.101	0.000	0.000	0.000
GBM								
Q2-Q1	-0.011	0.000	0.000	0.000	-0.011	0.000	0.000	0.000
Q3-Q1	-0.017	0.000	0.000	0.000	-0.018	0.000	0.000	0.000
Q4-Q1	-0.019	0.000	0.000	0.000	-0.020	0.000	0.000	0.000
Q5-Q1	-0.021	0.000	0.000	0.000	-0.022	0.000	0.000	0.000
GLM								
Q2-Q1	-0.039	0.000	0.000	0.000	-0.037	0.000	0.000	0.000
Q3-Q1	-0.057	0.000	0.000	0.000	-0.055	0.000	0.000	0.000
Q4-Q1	-0.070	0.000	0.000	0.000	-0.067	0.000	0.000	0.000
Q5-Q1	-0.088	0.000	0.000	0.000	-0.081	0.000	0.000	0.000
GLMi								
Q2-Q1	-0.072	0.000	0.000	0.000	-0.053	0.000	0.000	0.000
Q3-Q1	-0.104	0.000	0.000	0.000	-0.073	0.000	0.000	0.000
Q4-Q1	-0.133	0.000	0.000	0.000	-0.091	0.000	0.000	0.000
Q5-Q1	-0.185	0.000	0.000	0.000	-0.121	0.000	0.000	0.000
SVM								
Q2-Q1	-0.088	0.001	0.000	0.000	-0.077	0.001	0.000	0.000
Q3-Q1	-0.124	0.001	0.000	0.000	-0.106	0.001	0.000	0.000
Q4-Q1	-0.156	0.001	0.000	0.000	-0.125	0.001	0.000	0.000
Q5-Q1	-0.227	0.001	0.000	0.000	-0.163	0.001	0.000	0.000
NN								
Q2-Q1	-0.032	0.001	0.000	0.000	-0.048	0.000	0.000	0.000
Q3-Q1	-0.036	0.001	0.000	0.000	-0.052	0.000	0.000	0.000
Q4-Q1	-0.037	0.001	0.000	0.000	-0.080	0.000	0.000	0.000
Q5-Q1	-0.064	0.001	0.000	0.000	-0.127	0.000	0.000	0.000
PM								
Q2-Q1	-0.053	0.000	0.000	0.000	-0.040	0.000	0.000	0.000
Q3-Q1	-0.077	0.000	0.000	0.000	-0.058	0.000	0.000	0.000
Q4-Q1	-0.098	0.000	0.000	0.000	-0.069	0.000	0.000	0.000
Q5-Q1	-0.132	0.000	0.000	0.000	-0.086	0.000	0.000	0.000

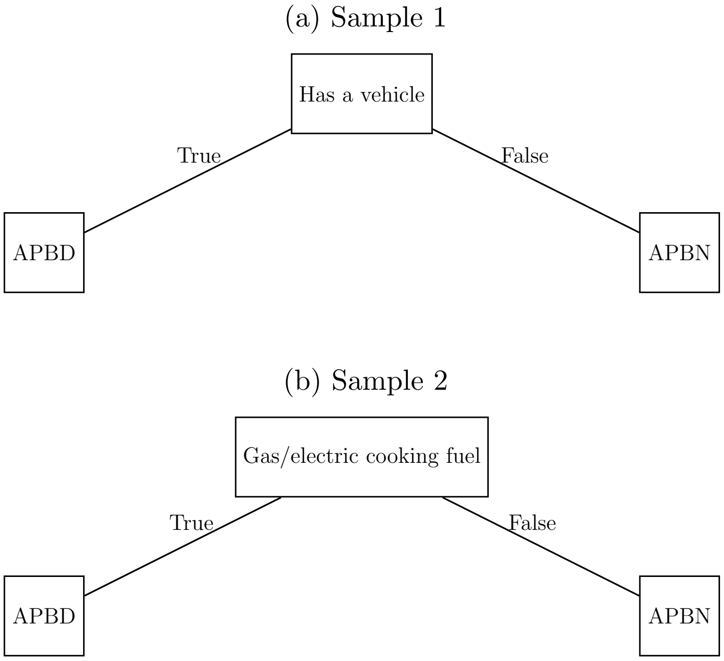
Note: Table reports estimates and standard errors of the differences in sorted GATEs (for quintiles of predicted CATEs) between the lowest quintile (Q1) and higher quintiles (Q2-Q5). Unadj *p*-val does not correct for multiple hypothesis testing. Adj *p*-val uses the Romano-Wolf procedure to correct for multiple hypothesis testing.

Table B.2: Counterfactual mean outcomes for candidate estimators included in the super learner

	AIPTW		**TMLE**	
	Est	SE	Est	SE
Static rules				
APBD-all	0.123	0.004	0.124	0.003
APBN-all	0.127	0.002	0.126	0.002
Threshold-based rules				
GLM-$V1$	0.114	0.003	0.113	0.003
GLMi-$V1$	0.117	0.003	0.115	0.003
PM-$V1$	0.116	0.003	0.114	0.003
NN-$V1$	0.122	0.003	0.120	0.003
SVM-$V1$	0.120	0.003	0.122	0.003
GBM-$V1$	0.114	0.003	0.112	0.003
CF-$V1$	0.118	0.003	0.116	0.003
GLM-$V2$	0.114	0.003	0.112	0.003
GLMi-$V2$	0.114	0.003	0.113	0.003
PM-$V2$	0.114	0.003	0.113	0.003
NN-$V2$	0.119	0.003	0.117	0.003
SVM-$V2$	0.126	0.003	0.127	0.003
GBM-$V2$	0.115	0.003	0.112	0.003
CF-$V2$	0.117	0.003	0.116	0.003
Tree-based rules				
PT1-$V1$	0.116	0.003	0.114	0.003
PT2-$V1$	0.116	0.003	0.115	0.003
PT3-$V1$	0.117	0.003	0.115	0.003
PT1-$V2$	0.116	0.003	0.114	0.003
PT2-$V2$	0.116	0.003	0.114	0.003
PT3-$V2$	0.116	0.003	0.114	0.003

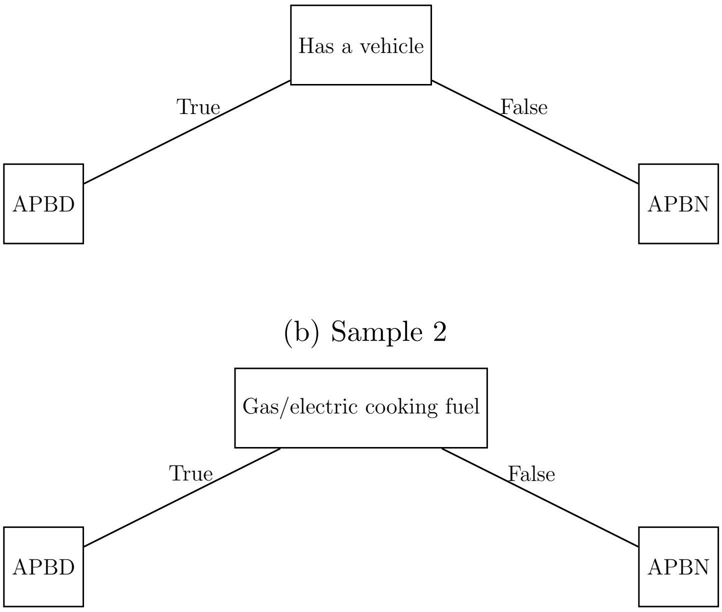
Note: Point estimates (Est) and standard errors (SE) are reported for the AIPTW and TMLE estimators.

Figure B.4: Depth 1 policy trees (fitted on $V1$)


(a) Sample 1

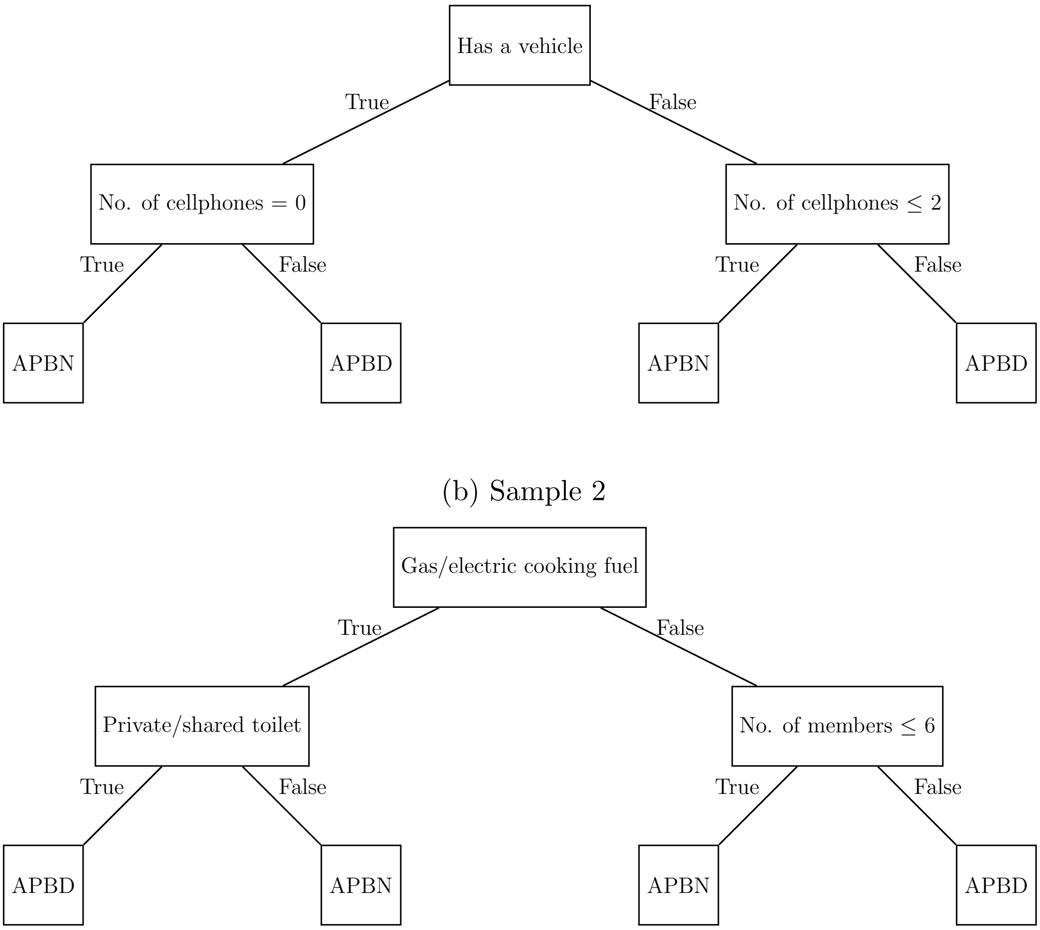
(b) Sample 2

Figure B.5: Depth 1 policy trees (fitted on $V2$)


(a) Sample 1

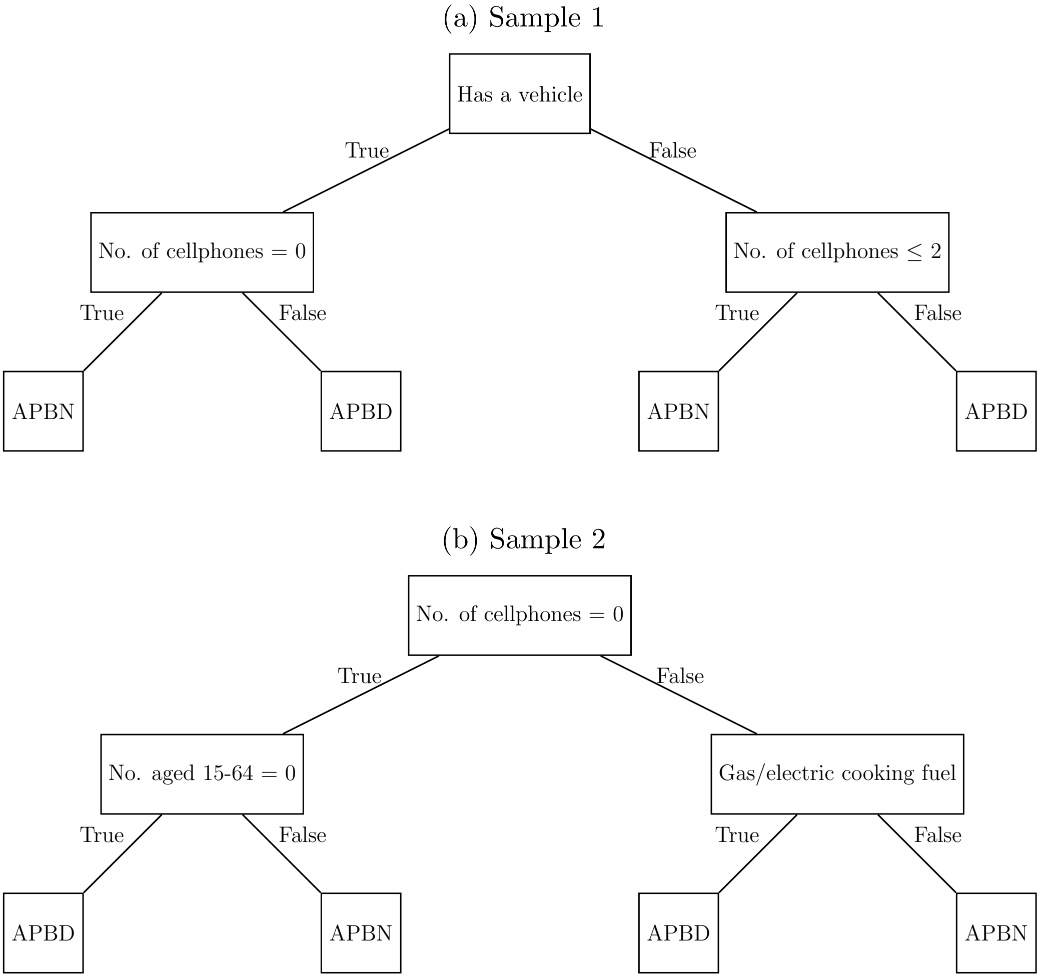
(b) Sample 2

Figure B.6: Depth 2 policy trees (fitted on $V1$)


(a) Sample 1

(b) Sample 2

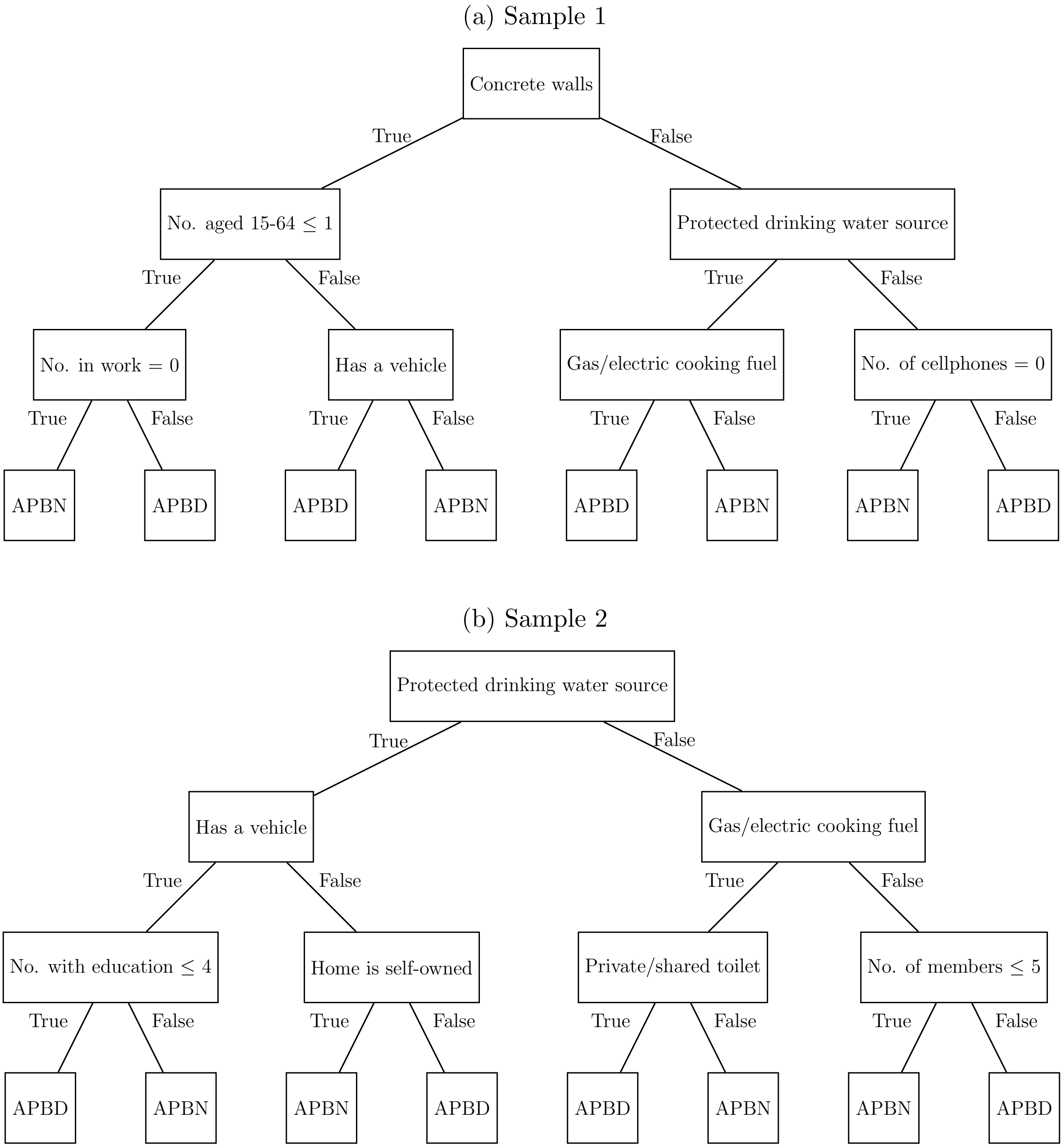
Figure B.7: Depth 2 policy trees (fitted on $V2$)



(a) Sample 1

(b) Sample 2

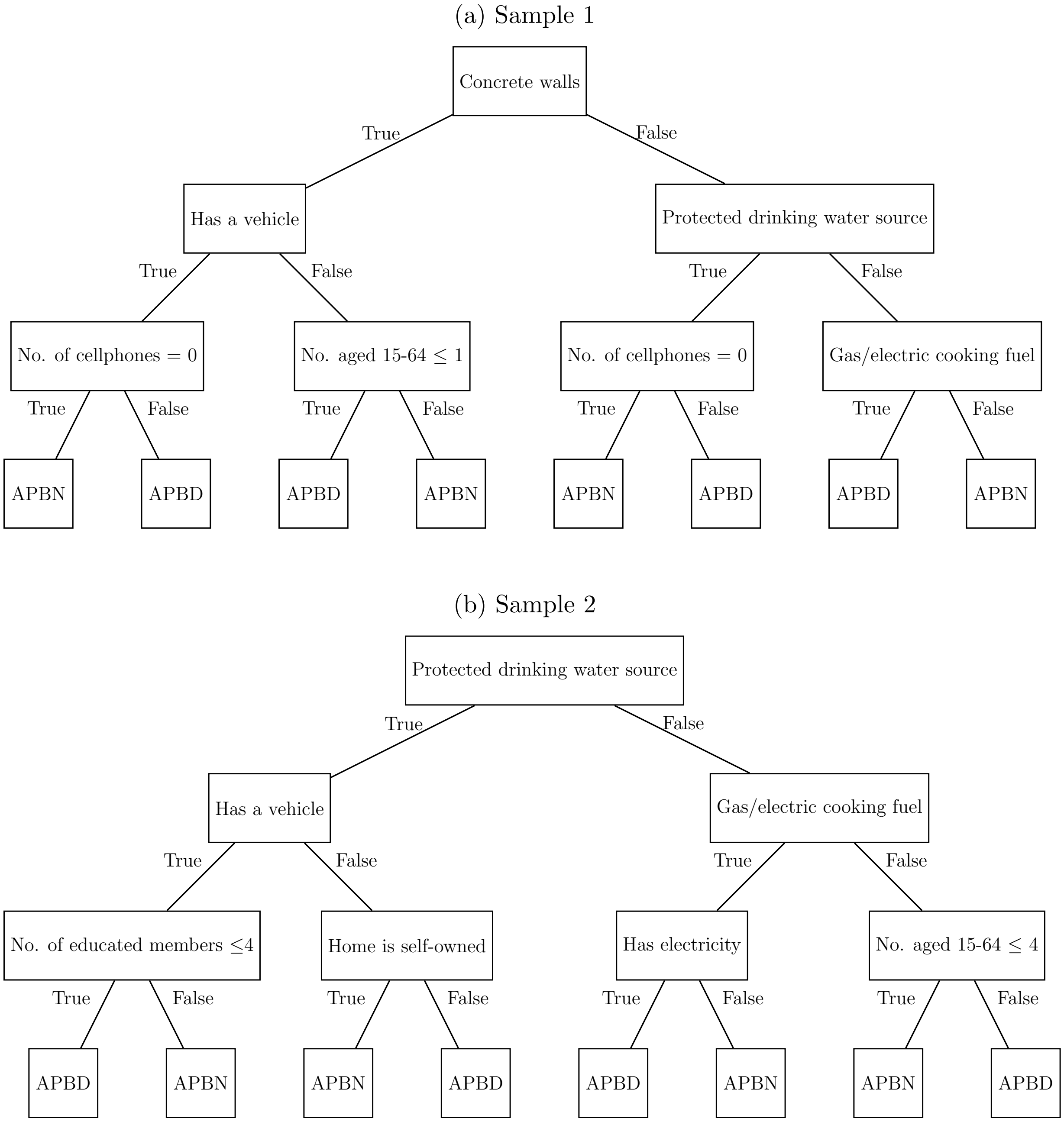
Figure B.8: Depth 3 policy trees (fitted on $V1$)



(a) Sample 1

(b) Sample 2

Figure B.9: Depth 3 policy trees (fitted on $V2$)



(a) Sample 1

(b) Sample 2

C Optimal policy learning with resource constraints

Luedtke and van der Laan (2016a) model the problem of estimating optimal resource constrained policy rules by imposing a constraint κ on the maximum proportion of units that can be treated, and defining a set of solutions that satisfy κ. The optimal policy rule is the optimal solution among the set of solutions that respect the constraint.

The formal theorem starts by defining S_P as the survival function of the CATE function $\hat{\tau}(V_i)$, i.e. the probability that $\hat{\tau}(V_i)$ is greater than some varying threshold T: $T \mapsto P(\tau(V_i) > T$.

Then, let

$$\begin{aligned} \eta &:= \inf\{T : S_P(T) \leq \kappa\} \\ T_P &:= \max\{\eta_P, 0\}, \end{aligned}$$

where η identifies the largest threshold value for which the survival probability is less than κ.

The optimal policy rule can be defined as follows:

$$\hat{\pi}^* := \begin{cases} \frac{\kappa - S_{P(T_P)}}{P(\hat{\tau}(V_i) = T_P)}, & \text{if } \hat{\tau}(V_i) = T_P \text{ and } T_P > 0 \\ \mathrm{I}(\hat{\tau}(V_i) > T_P), & \text{otherwise.} \end{cases}$$

D Classical Derivations

D.1 Equivalence of the Optimization Objectives

In this section, we provide classical derivations for the interested reader. In the following, we emphasize the equivalence of the full optimization objective in Equation (1) and the double-robust-based optimization objective in Equation (8). In particular, for $d = \{0, 1\}$, note that

$$\begin{aligned}
E[\hat{\Gamma}_i(d) \mid X_i] &= E[\hat{\mu}_d(X_i) + \frac{I(D_i = d)}{\hat{e}_d(X_i)}(Y_i - \hat{\mu}_d(X_i)) \mid X_i] \\
&= E[\hat{\mu}_d(X_i) \mid X_i] + E[\frac{I(D_i = d)}{\hat{e}_d(X_i)}(Y_i - \hat{\mu}_d(X_i)) \mid X_i] \\
&= \hat{\mu}_d(X_i) + \frac{1}{\hat{e}_d(X_i)}\hat{e}_d(X_i)(E[Y_i(d) \mid X_i] - \hat{\mu}_d(X_i)) \\
&= E[Y_i(d) \mid X_i].
\end{aligned}$$

D.2 Equivalence of the Double Robust Scores

We emphasize the equivalence of the double robust score in Equation (7) and (9):

$$\begin{aligned}
\hat{\Gamma}_i^{cf} &= \hat{\tau}^{cf}(V_i) + \frac{D_i - \hat{e}(X_i)}{\hat{e}(X_i)(1 - \hat{e}(X_i))}[Y_i - \hat{m}(X_i) - (D_i - \hat{e}(X_i))\hat{\tau}^{cf}(V_i)] \\
&= (\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)) + \frac{D_i - \hat{e}(X_i)}{\hat{e}(X_i)(1 - \hat{e}(X_i))}[Y_i - \hat{m}(X_i) - (D_i - \hat{e}(X_i))(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))] \\
&= (\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)) + \frac{D_i - \hat{e}(X_i)}{\hat{e}(X_i)(1 - \hat{e}(X_i))}[Y_i - \hat{\mu}_0(X_i) - \hat{e}(X_i)(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)) \\
&- (D_i - \hat{e}(X_i))(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))] \\
&= (\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)) + \frac{D_i - \hat{e}(X_i)}{\hat{e}(X_i)(1 - \hat{e}(X_i))}[Y_i - \hat{\mu}_0(X_i) - (\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))(D_i - \hat{e}(X_i) + \hat{e}(X_i))] \\
&= (\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)) + \frac{D_i - \hat{e}(X_i)}{\hat{e}(X_i)(1 - \hat{e}(X_i))}[Y_i - \hat{\mu}_0(X_i) - D_i(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))] \\
&= (\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)) + \frac{D_i - \hat{e}(X_i)}{\hat{e}(X_i)(1 - \hat{e}(X_i))}[Y_i - \hat{\mu}_D(X_i)] \\
&= \hat{\Gamma}_i.
\end{aligned}$$

We note that the second to last line follows since

1. if $D_i = 1$,

$$\begin{aligned}
&Y_i - \hat{\mu}_0(X_i) - D_i(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)) \\
&\quad = Y_i - \hat{\mu}_1(X_i).
\end{aligned}$$

2. if $D_i = 0$,

$$\begin{aligned} &Y_i - \hat{\mu}_0(X_i) - D_i(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)) \\ &\quad = Y_i - \hat{\mu}_0(X_i). \end{aligned}$$