Tree-based Estimation of Heterogeneous Dynamic Policy Effects

Steve Miller[a,*]

[a] *Department of Applied Economics, University of Minnesota, St. Paul, MN, USA*

October 6, 2017

Abstract

This paper develops and applies a method for estimating policy impacts that may vary both cross-sectionally and across time. The approach extends recent work using regression trees for causal inference to a setting of dynamic and staggered treatment. Potential benefits include not only improved targeting, but also setting realistic expectations for effect timing and evaluating whether policy implementation has improved. I demonstrate the method in an analysis of how individual quota systems impact the probability of natural resource collapse, finding substantial heterogeneity and nonlinear time dependencies not identified in prior work.

JEL Codes: C23; C22; Q22

[*] Corresponding author. Email: s-miller@umn.edu. Phone: 1-612-625-3212.

I am grateful for generous input from many others, including seminar participants at the University of Minnesota, Utah State University, and Iowa State University, as well as conference participants at AERE 2017.

1. Introduction

Understanding how the impacts of public policies vary is crucial to evaluating past efforts, setting expectations for the performance of current policies, and targeting future interventions. Variability in policy effects is pervasive: labor market interventions (Gerfin and Lechner 2002), changes to class size in schools (Bandiera et al. 2010), and transfer programs (Becker et al. 2013) all may impact different groups in different ways. Similarly, environmental policies may induce a range of emissions abatement across firms (Fowlie 2010), changes in ozone concentration across space (Auffhammer and Kellogg 2011), or modified consumption in response to informational programs (Shimshack et al. 2007; Allcott 2011). Ignoring this variability can lead to misleading conclusions; a policy may have no effect on average but strong impacts on specific people, firms, or locations. (Auffhammer et al. 2009).

One key challenge with many methods for studying heterogeneity in policy effects is that they require pre-specification of its form. Two common approaches are (1) to interact a treatment variable with one or more cross-sectional characteristics in a parametric model, or (2) to partition the data into groups and estimate a single average effect (parametrically or not) separately for each subset. Both routes require important assumptions about the way in which policy effects vary, codified either by choices of variables with which to interact a treatment variable or by groups of observations on which to estimate separate policy effects. Robustness of estimates can be probed by estimating multiple models under different assumptions, but the range of potential forms of heterogeneity is large.

Two recent contributions by Athey and Imbens (2016) and Wager and Athey (2017) offer a potential way forward, adapting tree-based methods from machine learning to uncover the form of heterogeneity while striving to maintain a causal interpretation of estimates. The logic is straightforward: start with the partitioning approach described above, derive a way to compare potential partitions, and modify automatic partitioning procedures (regression trees and random forests) to find the 'best' partition and corresponding set of estimates. The key insight from those studies is that partitions can be compared by the mean squared error between estimated and true policy effects, and crucially, that error can be estimated without knowledge of the true policy effect. The main benefit of the resulting method is that the researcher no longer has to pre-specify the form of heterogeneity, though she still must choose the set of variables with which the estimated treatment effect is allowed to vary.

A natural concern with this approach is that if tree-based algorithms are designed to find heterogeneity, they may overfit and yield spurious claims of variability. The methods in Athey and Imbens (2016) offer two safeguards. First, the criterion used to evaluate partitions is designed as an estimate of out-of-sample mean squared error of estimated policy effects. As a result, the criterion penalizes partitions with imprecise estimates, which favors fewer splits of the data. Second, the authors advocate 'honest' estimation in which one dataset is used to construct a partition, and a separate dataset is used to compute estimates given that partition. Together, both features reduce the potential for overfitting.

The principal goal of this paper is to extend the use of tree-based methods to study how policy effects vary not only cross-sectionally, but also through time. Effect strength may vary with policy duration for a variety of reasons, e.g., information diffusion, learning, repeated exposure, or the growth of a resource (Hamilton 1995; Lalive et al. 2008; Allcott and Rogers 2014; Costello et al. 2008). In addition, if policy exposure is staggered through time, individuals or firms subjected to the policy starting at a later date may experience different impacts if the effectiveness of the intervention has improved. Further, interactions among these dimensions of heterogeneity are possible: the way in which policy effects vary through time may depend on cross-sectional characteristics. As with cross-sectional heterogeneity, a model could be fit separately for each possible treatment duration and policy start year, or for particular time ranges identified by the researcher. The method proposed here automates that process, identifying temporal ranges over which treatment effects can be well approximated by an average.

The method I propose is composed of three key parts. First, identification is based on the Sequential Conditional Independence Assumption (SCIA) (Robins et al. 2000; Lechner and

Miquel 2010), which is a dynamic analog of the selection-on-observables assumption employed in Athey and Imbens (2016) and numerous econometric studies. Importantly, the tree-based methods play no role in identification. Second, policy effects are estimated using Inverse Propensity Score Weighting (IPSW) for each group in a given partition, which constructs a weighted difference between treated and control outcomes with weights inversely proportional to the probability that an entity experienced its entire observed sequence of policy exposure (Imbens 2015). The combination of SCIA and IPSW yields a weighting estimator appropriate for panel settings with dynamic selection into treatment (see, e.g., Azoulay et al. (2009)). Finally, modified versions of the tree-based procedures in Athey and Imbens (2016) and Wager and Athey (2017) are used to systematically and recursively divide the data into subsamples, with a single estimate generated for each subsample. The procedures select a partition for which estimated policy effects differ across subsamples but are similar and precisely estimated within subsamples. The key but straightforward modification here is that subsamples may be defined not only by one or more cross-sectional characteristics, but also by *potential* date of policy introduction and *potential* duration of policy exposure.

To illustrate the resulting approach, I revisit whether a common form of rights-based management helps avoid resource collapse in fisheries around the world. Costello et al. (2008) originally examined this question on a global scale using standard panel methods and propensity score adjustments, finding that individual quota (IQ) programs helped reduce the probability of collapse with effect sizes increasing with policy duration. Because the tree-based methods presented here rely on a similar selection-on-observables identification strategy, any differences in findings between the studies are plausibly attributable to the methods themselves.

Applying the tree-based methods to a newly compiled and updated global dataset paints a much richer picture of heterogeneity. Point estimates of the effect of IQs on the probability of collapse range from a reduction of 25% to an increase of 20%, with an overall mean of 8% reduction. The effectiveness of IQs varies with fishery characteristics, policy duration, and year of first implementation. Further, the way in which the effect size varies with duration depends upon fishery characteristics. In some cases, a significant impact of IQs on the probability of collapse takes several years to materialize, consistent with the intuition that natural resources take time to grow. Further, the largest decreases in the probability of collapse appear in fisheries with moderate catch levels just prior to the implementation of IQs. This finding is consistent with the intuition that fisheries experiencing extreme catch levels just before IQ implementation may see less biological benefit from IQs. Fisheries with high, stable levels of catch could already be well managed; conversely, fisheries that have already collapsed may take much longer to rebuild, and the first several years of IQ management may mandate extremely low catch or even closure. In summary, individual quota programs do appear to reduce the probability of fishery collapse, but in a manner that is far from uniform.

This paper complements a growing body of work on the use of machine learning methods for studying heterogeneity in program or policy effects. Applications of tree-based methods to study cross-sectional heterogeneity include Davis and Heller (2017), Handel and Kolstad (2017), and Bertrand et al. (2017). Asher et al. (2016) extend the idea of automated partitioning to generalized method of moments estimation in each subsample, but maintain a focus on cross-sectional heterogeneity. Chin et al. (2017) use trees to estimate counterfactual outcomes in a difference-in-differences framework rather than to directly estimate the treatment effect itself. A few studies examine effects at different intervals after program implementation, allowing for cross-sectional heterogeneity within each interval. Bertrand et al. (2017) look at short-run and long-run program effects, but estimate those effects in separate models. With many potential post-treatment effects to be estimated (e.g. one per potential treatment start year and duration), estimating a separate model per effect could become cumbersome or force the researcher to choose duration cutoffs in order to estimate a smaller set of models. Knaus et al. (2017) estimate cross-sectionally averaged effects for different durations, and hand-select a few durations for which to separately estimate cross-sectional heterogeneity, but stop short of automatically exploring potential duration splits or interactions between temporal and cross-sectional heterogeneity.[1]

The remainder of this paper is as follows. In the next section, I briefly review the causal tree method proposed by Athey and Imbens (2016) and an extension to random forests suggested by Wager and Athey (2017). In the third section, I formally adapt these proposals to estimating treatment effects which may also vary with policy start time and duration. The fourth section applies the resulting method to study the effectiveness of individual quota systems in fisheries, and the final section concludes.

2. Causal Trees and Forests

One way to understand how a policy's effects vary is to estimate its average impact on different subsamples. With an appropriate set of assumptions and a corresponding estimator, the resulting set of estimates can be interpreted as a step function approximation to the true policy effect. A key challenge, of course, is how to select a set of subsamples (locations of steps) so that the approximation of the policy effect function is a good one.

A recent proposal by Athey and Imbens (2016) cleverly automates the selection of cross-sectional subsamples by adapting a regression tree algorithm. Regression trees are predictive models that make a single prediction for each subsample in a dataset, with the set of subsamples chosen via an algorithm that seeks to minimize squared prediction error.[2] The algorithm begins by estimating a quantity of interest and associated error for the entire sample. Next, the sample is split in two based on the values of a single covariate at a time, estimates are computed for each subsample, and the improvement in estimation error is noted. This calculation is repeated for many possible splits and the split yielding the largest improvement in estimation error is selected. Because the number of potential splits is large (and infinite if at

[1] An additional difference is that Knaus et al. (2017) also estimate heterogeneity using the least absolute shrinkage and selection operator (LASSO) rather than tree-based methods. Tree-based methods start with a single average effect and gradually increase heterogeneity ("top down"), while LASSO begins with a fully heterogeneous model and then selects a subset of interactions ("bottom up").

[2] Variants of the algorithm allow for non-uniform estimates/predictions within subsamples, while other objective functions are designed for classification tasks.

least one covariate is continuous), not all possible splits can be (practically) considered. The algorithm considers only threshold splits for continuous variables, with possible thresholds located at the values in the sample itself. Prior work has established that this restriction is immaterial (Fisher 1958); an analogous result limits the partitions that must be considered for categorical predictors (Breiman et al. 1984; Chou 1991). After the first split of the data, each subsample is then recursively split until further divisions give a reduction in prediction error that is sufficiently small. The resulting partition can be viewed as a tree with the final subsamples as leaves, giving the algorithm its name.

The key insight from Athey and Imbens (2016) is that this procedure can be adapted to estimate policy effects, even though the estimation error cannot be evaluated directly because true policy effects are not observed. Those authors illustrate that under a selection-on-observables assumption, the expected mean squared error (EMSE) of an unbiased policy effect estimator can be estimated, so that the regression tree procedure sketched above can be based on estimates of the policy effect alone. Specifically, if an unbiased estimator $\hat{\tau}(x)$ of a conditional average treatment effect $\tau(x)$ is available (conditioning on a vector of covariates x), Athey and Imbens (2016) show that minimizing EMSE is equivalent to maximizing the criterion

$$Q \equiv E[\tau^2(x)] - E[Var(\hat{\tau}(x))], \tag{1}$$

which can be estimated by

$$\hat{Q} \equiv \frac{1}{N_{tr}} \sum_i \hat{\tau}^2(x) - \left(\frac{1}{N_{tr}} + \frac{1}{N_{est}} \right) \sum_i \widehat{Var}(\hat{\tau}(x)), \tag{2}$$

where i indexes observations and N_{tr} and N_{est} are the number of observations in training and estimation datasets, respectively. The training dataset is used to select a partition, while the estimation dataset is used to estimate treatment effects given a fixed partition. This split-sample approach, which Athey and Imbens (2016) term “honest estimation”, partly addresses overfitting concerns in an adaptive procedure like the regression tree algorithm.

When evaluating potential splits in the regression tree procedure, the best split is defined as that which causes $\hat{Q}$ to increase by the largest amount.[3] There are two components to this criterion. The first term rewards identification of treatment effects which vary across subpopulations. The second term in the criterion penalizes variance of the treatment effect estimator within a subpopulation. This helps control too aggressively seeking heterogeneity through overfitting. Together, these components have intuitive appeal. The criterion will be highest when the treatment effect is well approximated by a step function: constant within but varying across subpopulations. The resulting modified algorithm maximizing this criterion is referred to as a causal tree.

A related paper by Wager and Athey (2017) addresses some limitations of the regression tree procedure by building many step-function approximations to the policy effect function: a forest of trees. While a single step-function approximation (tree) is highly variable, averaging over many different approximations can provide more stable results. The random forest algorithm (Breiman 2001) does exactly that, building each tree from a bootstrapped sample, considering a random subset of variables each time the data is split, and averaging over the set of tree estimates to generate a single forest estimate. Wager and Athey (2017) adapt the random forest procedure to use causal trees and establish key asymptotic properties of the resulting causal forest.

In what follows, I build on these ideas to examine heterogeneity in policy effects across time in a panel data setting. Both Athey and Imbens (2016) and Wager and Athey (2017) consider cross-sectional heterogeneity, but as shown below, these ideas can be readily extended to study how policy effects depend on both the time a policy started and how long it has been in place.

[3] The regression tree algorithm is greedy, seeking the largest local improvement to $\hat{Q}$. It does not guarantee that the final partition will yield the global maximum of $\hat{Q}$.

3. Estimating policy effects which vary both cross-sectionally and temporally

In this section, I detail how causal trees can be adapted to estimate policy effects that may vary not only with cross-sectional characteristics, but also with both calendar time and policy duration. The proposed method retains the broad structure of a causal forest: a set of causal trees is constructed from bootstrapped data samples, and the conditional average policy effect is estimated as the average of relevant estimates across the trees. However, three main modifications are needed to deal with the dynamic nature of treatment. The first two modifications borrow from existing work on dynamic treatment regimes. First, the assumptions necessary to identify a conditional average policy effect must be changed to reflect a policy adoption (selection) process that occurs over and may vary with time. Second, the estimation method used on any subsample must be updated to reflect the dynamic nature of both treatment and the corresponding effects. Finally, the handling of control observations during the splitting procedure requires elaboration. In a dynamic treatment setting, a control observation can potentially aid in estimation of a counterfactual for treated units first subjected to a policy at different times. As such, when splitting the data on either policy duration or policy start time, it may be useful to have a control observation appear in both subsamples. These modifications are all relatively straightforward, and I address them below in turn after setting up the estimation objective.

Estimation target

The goal of estimation is to approximate the effect of a policy on an outcome of interest, where that effect may vary with the characteristics of an affected entity, the time the policy went into effect ("start time"), and how long an entity has been subject to the policy ("duration"). The true policy effect will be approximated with a step function, where breaks in the step function may occur with cross sectional characteristics, start time, or duration. Each step corresponds to a constant average policy effect estimated on a subsample.

To formalize the problem, suppose we observe outcomes Y_{it}, characteristics X_{it}, and a binary treatment status T_{it} for entity i at time t. The treatment indicator T_{it} is equal to 1 if entity i is treated at time t and is 0 otherwise. For brevity, let the history of any variable be denoted by its boldface version, e.g., $\mathbf{X}_{it} = \{X_{i1}, ..., X_{it}\}$. To simplify analysis, in what follows I focus on cases in which treatment is permanent – once a policy goes into effect, it remains in effect. This allows representation of a sequence of treatments by a duration $D_{it} = \sum_{t'=1}^{t} T_{it'}$, which simplifies both notation and interpretation of results. Further, to simplify explanation I refer to units of time as years.

With this notation the estimation target can be defined. In the language of potential outcomes popularized by Rubin (1974), each step is a conditional average treatment effect (CATE), with conditioning ensuring an observation falls within a particular step. Denoting a *potential* start year by s and a *potential* policy duration by d, the CATE is defined by

$$\tau(\mathcal{X}, \mathcal{D}, \mathcal{S}) \equiv E[Y_{it}(d,s) - Y_{it}(0,s) | \mathbf{X_{is-1}} \in \mathcal{X}, d \in \mathcal{D}, s \in \mathcal{S}, t = s + d - 1], \tag{3}$$

where $Y_{it}(d,s)$ is the potential outcome for entity i in year t if subjected to d years of treatment beginning in year s. The CATE is simply the average difference in outcomes for an entity in year t if it had been subjected to d years of a policy starting in year s compared to no treatment in that same timeframe. The sets $\mathcal{X}$, $\mathcal{D}$, and $\mathcal{S}$ define which pre-treatment characteristics, potential treatments, and potential treatment start years the CATE applies to. Note that the characteristics $\mathbf{X_{is-1}}$ used for conditioning are restricted to those appearing prior to the start of the potential treatment.[4]

Identifying assumption

Since is impossible to observe i in year t under two different treatment histories, further assumptions are needed to estimate (3). Because of the dynamic, non-random assignment to

[4] Characteristics occurring after the initial exposure to treatment could be used for conditioning if we were interested in, e.g. comparing the effect of d years of treatment to that of $d-1$ years of treatment. However, the focus here is on the total effect of d years of treatment as compared to a counterfactual of no treatment.

treatment, a suitable choice is the Sequential Conditional Independence Assumption (SCIA) (Robins et al. 2000; Lechner and Miquel 2010). Like conditional independence in a static context, SCIA essentially states that after conditioning on enough information, whether or not an entity is treated in the current period is as good as random. The conditioning information includes current and past exogenous characteristics, and past endogenous variables, including any prior treatment exposure. Letting X_{it}^{EX} and X_{it}^{EN} denote the exogenous and potentially endogenous subsets of covariates, respectively, the SCIA states:

$$Y_{it}(d, s) \perp\!\!\!\perp T_{it} | \mathbf{T}_{it-1}, \mathbf{X}_{it}^{EX}, \mathbf{X}_{it-1}^{EN}. \tag{4}$$

Importantly, the conditioning set includes past treatment and lagged endogenous variables (possibly including lagged outcomes), since both may affect current treatment and outcomes and thus confound estimation.

Estimation

Under the SCIA, the average effect of a treatment sequence compared to the absence of treatment in any subsample can be estimated several ways, including using inverse propensity score weighting (IPSW). The intuition behind IPSW is identical to that for weighting estimators in cross-sectional settings: selection bias overweights observations selected into a particular treatment sequence, and the IPSW estimator attempts to recover a sample that reflects the underlying population by inverting that weighting. The key difference in a dynamic setting is that the weights correspond to the inverse probability that an entity experienced its entire sequence of treatment exposure.

The probability of an entity experiencing its actual sequence of treatment is the product of conditional probabilities of observed treatment in each year. Formally,

$$p(\mathbf{T}_{it}) = \prod_{t'=1}^{t} P(T_{it'} | \mathbf{T}_{it'-1}, \mathbf{X}_{t'}^{EX}, \mathbf{X}_{t'-1}^{EN}). \tag{5}$$

The aforementioned focus on permanent treatment regimes, in which an entity subjected to treatment will be treated in all following years, allows for simplification of these probabilities. Specifically, the probability of remaining in treatment is one, and the probabilities of interest can be rewritten as functions of duration and start year:

$$p(d,s,t) = \\ \begin{cases} p(T_{is}=1|\mathbf{T}_{is-1}=\mathbf{0},\mathbf{X}^{EX}_{is},\mathbf{X}^{EN}_{is-1}) \prod_{t'=1}^{s-1} p(T_{it'}=0|\mathbf{T}_{it'-1}=\mathbf{0},\mathbf{X}^{EX}_{it'},\mathbf{X}^{EN}_{it'-1}) & \text{if } d>0 \\ \prod_{t'=1}^{t} p(T_{it'}=0|\mathbf{T}_{it'-1}=\mathbf{0},\mathbf{X}^{EX}_{it'},\mathbf{X}^{EN}_{it'-1}) & \text{if } d=0 \end{cases} . \tag{6}$$

Intuitively, the probability of nonzero treatment duration is the probability of beginning treatment in the specified start year and not before, while the probability of zero duration is simply the probability of never getting treated in any year.

Each conditional probability can be estimated as the probability of starting treatment in a single year conditioned on observed information, and the overall probability of a treatment sequence can be constructed from those estimates. For example, in the empirical application, I estimate the conditional probabilities using a pooled logit model with year effects, though more flexible estimation approaches are possible.

As with cross-sectional analyses involving estimated probability of treatment, both overlap and balance assessments should be conducted before probabilities are used to construct weights. While overlap could be assessed for the probability of entire treatment sequences, since selection happens on an annual basis, it may be more practical to assess overlap for the conditional probability of transitioning into treatment. That conditional probability is the model that underlies the treatment sequence probabilities and can be assessed using standard methods. Further, as in cross-sectional IPSW estimation, it may be prudent to use trimming to cap the influence of observations with extremely low estimated probability of observed treatment sequence.

Because the conditioning set for the SCIA and treatment probabilities will differ across observations from different points in time, some care must be taken in estimating a treatment effect using a subsample with observations from multiple years. One option is to estimate $\tau(\mathcal{X}, \mathcal{D}, \mathcal{S})$ by first estimating $\tau(\mathcal{X}, \{d\}, \{s\})$ for each $(d, s) \in \mathcal{D} \times \mathcal{S}$, then averaging across those estimates. Choosing a specific duration d and start date s implies a specific observation year t, which in turn ensures that the conditioning set used is coherent across treated and control observations.

With the preceding groundwork in place, the IPSW estimator can be defined. The average treatment effect within a sub-population is estimated as follows:

$$\hat{\tau}(\mathcal{X}, \mathcal{D}, \mathcal{S}) = \frac{1}{|\mathcal{D} \times \mathcal{S}|} \sum_{d \in \mathcal{D}} \sum_{s \in \mathcal{S}} \hat{\tau}(\mathcal{X}, d, s), \tag{7}$$

where

$$\hat{\tau}(\mathcal{X}, d, s) = \sum_{\substack{s=t-d+1, \\ i,t: X_{is-1} \in \mathcal{X}, \\ D_{it}=d}} w_{it}^{d}(d, s) y_{it} - \sum_{\substack{s=t-d+1, \\ i,t: X_{is-1} \in \mathcal{X}, \\ D_{it}=0}} w_{it}^{0}(d, s) y_{it}, \tag{8}$$

$$w_{it}^{a}(d, s) = \frac{\frac{1(D_{it}=a)}{\hat{p}(a,s,t)}}{\sum_{\substack{s=t-d+1, \\ i,t: X_{is-1} \in \mathcal{X}, \\ D_{it}=a}} \frac{1(D_{it}=a)}{\hat{p}(a,s,t)}} \text{ for } a \in \{0, d\}. \tag{9}$$

The estimated treatment effect for a set of potential start years and durations is the average across estimated treatment effects for each pair of a potential duration and potential start date. Each duration- and start year-specific estimated effect is estimated via IPSW, which is a weighted difference in mean outcomes between entities experiencing the specified duration of treatment beginning in the specified year and those experiencing no treatment. Weights are equal to the inverse estimated probability that the entity received its observed sequence of treatment, with weights normalized to one separately for the treated and control groups. An estimate of the variance of this estimator, which is required to calculate the partitioning criterion (2), can be computed based on the linear representation (8). Doing so requires an estimate of $Var(y_{it}|\mathbf{X_{it}})$, which can be calculated by matching observation i to another having the same treatment duration and the smallest difference in covariates from observation i. See Imbens (2015) for details.

With the estimator defined, the splitting procedure defined in Athey and Imbens (2016) can be applied to splits on s or d with one final modification. Because s and d represent *potential* start year and *potential* duration, they do not exist in observed data. Assigning d and s is simple for treated observations: set $d = D_{it}$ and $s = t - d + 1$. In contrast, since control observations have $D_{it} = 0$, it is not obvious which values of d and s should be assigned. One option, which I adopt, is to replicate control observations for any values of s and d for which the control observation could plausibly aid in counterfactual estimation. For example, and untreated observation in 2017 could act as a control for $d = 1$ and $s = 2017$, or as a control for $d = 2$ and $s = 2016$. This approach is similar to matching with replacement, where the same control can contribute to multiple counterfactual estimates. Mechanically, this is accomplished as a data pre-processing step before the tree procedure starts.[5]

Inference

Some guidance on inference for causal trees and forests is provided in Athey and Imbens (2016) and Wager and Athey (2017). However, if a forest rather than a single tree is used, the inference approach suggested in Athey and Imbens (2016) (which applies to individual leaves in a single tree) is not directly applicable. Second, the asymptotic results for causal forests in Wager and Athey (2017) are derived under the assumption of randomized treatment, which does not hold for many policies. Third, the other regularity assumptions underlying the Wager and Athey (2017) may not hold in all datasets and, given the recency of their contribution, the finite sample performance of the asymptotic approximations are not yet well known.

For these reasons, I construct bootstrap confidence intervals for the conditional average treatment effects estimated using the forest procedure outlined above. The bootstrap wraps the entire forest-building procedure, including both estimates of treatment probabilities as

[5] To save on computation, each untreated observation could be replicated for a random subset of the (d, s) pairs for which it could serve as a potential control

well as the construction of all trees in the forest. Further, to account for correlation across time given the assumed panel structure of the data, bootstrapping resamples at the fishery level, so that if a fishery is selected for a bootstrap sample, the fishery's entire history is included (Kapetanios 2008). Altogether, the bootstrap procedure then builds a set of forests, and the quantiles of the estimates produced by the set of forests can be used to build confidence intervals. Bootstrapping the entire forest-building procedure is also employed in Asher et al. (2016); similarly, Knaus et al. (2017) bootstrap their LASSO-based procedure for identifying heterogeneity.

4. Application: Individual Transferable Quota Markets in Fisheries

Beginning in the mid 1970s, a number of commercial fisheries around the world have shifted to management based on individual quotas (IQs)[6] that allocate the rights to catch fish to specific entities (individuals or groups). Using the methods described above, I examine the effect of IQs on a catch outcome, with the primary goal of understanding what new insights these methods might provide as compared to more conventional approaches.

The study of IQs is a useful test case for two reasons. First, a high-profile benchmark study (Costello et al. 2008) used a mix of propensity score and fixed effects methods to examine whether IQs reduce the probability of fishery collapse, with collapse defined as catch falling below 10% of its historical maximum (Worm et al. 2006). Their results indicate that IQ adoption reduces probability of collapse on average, but their specifications allow for only limited heterogeneity in policy effects: a linear trend with policy duration. Because the tree-based methods used here rely on a similar identification strategy, a comparison of results should offer illustration of what new insights tree-based methods might provide.[7] Second, policy effects are likely heterogeneous: prior work has identified dependency on, e.g.,

[6] The analysis here does not distinguish between tradable and non-tradable forms of harvest rights.

[7] Regression trees and related techniques have been applied before to examine factors associated with improved fisheries outcomes, though not using causal trees or forests. Gutiérrez et al. (2011) use random forests to predict outcomes for co-managed fisheries. Melnychuk et al. (2016) use random forests to examine factors associated with desired outcomes for fisheries under individual quota management. In both cases, the authors attach a causal interpretation to estimates, but do not directly tackle the issue of causality.

social factors (Gutiérrez et al. 2011), scientific robustness of quotas set (Mora et al. 2009), biological characteristics of targeted species (Pinkerton and Edwards 2009), and country development (Ban et al. 2009).[8] More recent work also suggests that quota stringency itself (and hence outcomes) depends upon the strength of property rights in a fishery (Costello and Grainger forthcoming). Figure 1 illustrates the potential for heterogeneous effects, depicting trends in collapse among fisheries not under IQ management, and two groups under IQ management: those with low catch prior to policy implementation and those with high catch. Differences between the IQ groups and the non-IQ group vary through time, and there are clear differences in both levels and trends across the two IQ groups as well.

In this setting, i indexes a fishery, t a year, and D_{it} the number of years that fishery i was subjected to IQ management as of year t. The outcome of interest, Y_{it}, is a binary variable taking value one if and only if the catch c_{it} for fishery i in year t is less than 10% of the maximum catch observed in fishery i in all years $t' < t$, i.e., $Y_{it} \equiv \mathbb{1}(c_{it} \leq 0.1 \cdot c_{it}^{max})$ with $c_{it}^{max} \equiv \max_{t':t'<t} c_{it'}$.[9] The covariates X_{is} considered are one year lagged relative catch $(c_{is-1}/c_{is-1}^{max})$, one year lagged relative catch trend $((c_{is-1} - c_{is-2})/c_{s-1}^{max})$, and the Von-Bertalanffy growth rate K_i of the species. Catch variables are expressed relative to the historical max for consistency with the definition of collapse. In addition to estimating a model allowing for policy effects to vary with any of these variables, duration, and start year, I also present results allowing for heterogeneity only in particular dimensions.

The probability of transitioning to IQ management is estimated as a pooled logit model using the set of observations that are not already under IQ management in a given year.

[8] For a broader review of potential limitations of IQs see Copes et al. (1986).

[9] There are clear drawbacks to the use of catch-based measures of collapse, including mathematical reasons (Wilberg and Miller 2007) and differences from management-based measures of collapse (de Mutsert et al. 2008; Branch et al. 2011). Still, 1) catch data are available for far more fisheries than biomass estimates, 2) the purpose of the application is to illustrate what new methods offer, motivating use of the same outcome measure as prior work, and 3) some of the mathematical concerns with the chosen collapse metric (Wilberg and Miller 2007) are mitigated by the estimator used here. Those concerns state that even without a trend in true rates of collapse, the random nature of catch statistics make it more likely for a fishery to experience an extremely low catch year as time passes. However, since the estimator outlined above first compares IQ and non-IQ fisheries in a given year (see (8)), trends in collapse metrics are of less concern.

Explanatory variables include the lagged catch and lagged catch trend variables defined above, their squares, and year effects. The logic behind such a model is that recent outcomes are likely to influence adoption of a new management policy.

I apply this method to a global dataset of fisheries catch, IQ management status, and species characteristics ranging from 1950-2014. For consistency with Costello et al. (2008), a fishery is defined by a species and Large Marine Ecosystem (LME) pair. Catch records (tons per fishery per year) come from the Sea Around Us project, IQ status (yes/no) is derived from a database complied by the Environmental Defense Fund, and Von Bertalanffy growth rate (year^{-1}), where available, is collected from FishBase. Where a species is subjected to multiple forms of management within a single LME, the earliest date of IQ implementation for a species in an LME is used to determine IQ status for the fishery. Records are matched across data sources using species, location, and year. The dataset is constructed with the intent of being an updated and expanded version of the original data used in Costello et al. (2008). See the Appendix for more details on dataset assembly.

Prior to estimation, the dataset is cleaned and filtered in two primary ways. First, attention is restricted to fisheries occurring in LMEs which contain at least one fishery under IQ management during the study period, which should improve similarity between IQ fisheries and control fisheries. Second, the data are filtered to complete records containing all outcomes and explanatory variables.

The resulting dataset has 181,659 observations from 5,612 fisheries, with 337 fisheries managed via an IQ system at some point during the study window. For models including species growth rate, data availability limits the sample to 111,850 observations from 3,287 fisheries, 261 of which enter into IQ management during the study window. Many of the observations without available growth rates correspond to shellfish fisheries for which Von Bertalanffy growth is not applicable.

During the course of estimation, after propensity scores are estimated, records are eliminated in which estimated propensity scores are numerically equivalent to zero or one in order to satisfy overlap requirements. In addition, trimming is used to eliminate records with probability of observed treatment sequence below 0.001.

4.1. Results

4.1.1. Standard methods and heterogeneity

As a baseline for the tree-based approach, I first provide results using conventional estimation. Doing so has two main benefits. First, using a logit model identical to a primary model used in Costello et al. (2008) establishes that any differences in results obtained from the new methods presented here are not likely due to differences between the original and updated data. Second, examining results from alternate logit specifications that encode different types of treatment effect heterogeneity highlights the challenge of correctly specifying the form of heterogeneity.

One of the primary results from Costello et al. (2008) suggests that while the probability of fishery collapse increases through time, each additional year under IQ management can provide a counteracting and potentially completely offsetting effect on the probability of collapse. While those authors investigate many models, a basic specification they explore is a logit model with a constant and three regressors: whether a fishery is ever an IQ, a year effect, and how many years the fishery has been under IQ management. The last regressor is simply an interaction between a dummy variable indicating that a fishery is under IQ management and a continuous year variable, thereby capturing treatment effect heterogeneity.

I estimate the same model using an analog to their dataset produced by subsetting the updated data described earlier. Specifically, I filter to include only data up to 2003 and, for the purposes of this comparison, I consider only fisheries established as IQs in 2003 or before to be IQs. The marginal effects from estimating the logit model described above on this dataset are presented in the first column of Table 1. The key findings are as follows: while the probability of collapse increases around half a percent a year, an additional year of IQ management reduces the probability of collapse by around the same amount. These results are consistent with the main results presented Costello et al. (2008), suggesting the datasets are sufficiently similar up to 2003 that substantive differences in results from the tree-based methods are likely due to the methods themselves.

One concern with the logit specification just used is that the way in which the policy effect varies may be misspecified. To provide just one example, the effectiveness of IQ management may depend upon the (relative) catch in a fishery in the year just prior to IQ implementation. If a fishery is already collapsed or nearly so, management under IQs may optimally impose very low harvest levels in order to rebuild the stock, while a healthy fishery may enjoy continued high landings. The former case may show up as collapse in the dataset, while the latter will not. To explore this possibility, I estimate a second logit model which includes a dummy variable for whether a fishery is under IQ management in a given year and the interaction of that indicator with the relative catch just prior to IQ introduction.

The results of this exercise are presented in the second column of Table 1. While the probability of collapse increases through time at roughly the same rate as in the earlier logit specification, the effect of an additional year of IQ management is now estimated to compound that effect. Further, IQ management results in larger reductions in the probability of collapse for fisheries with higher relative catch just prior to IQ introduction. These findings highlight the both the challenges inherent in and consequences of choosing a specification for treatment effect heterogeneity.

4.2. Causal forest methods

To investigate the potential for the modified honest forest procedure outlined above to aid in identifying heterogeneous effects of IQs, the method was applied to two samples. First, the procedure was run without biological information, allowing for the use of more data. That procedure allowed splitting on four dimensions: pre-treatment relative catch, pre-treatment relative catch trend, policy duration, and policy start year. Overall, substantial heterogeneity in estimated treatment effects was found (Figure 2). The second set of results add the option to split on Von-Bertalanffy growth, using only observations for which that information is also available. Before presenting more detailed results, I briefly summarize two diagnostics for propensity score estimation.

4.2.1. Assessing the treatment propensity model

Balance plots for pre-treatment catch and catch trends are presented in Figure 3. The associated standardized difference in means after reweighting are 0.03 and 0.10, respectively, which satisfy the criteria for balance in Austin and Stuart (2015). In addition, Figure 4 presents distributions of estimated probability of transition into treatment among observations that do and do not change management. The two distributions display clear overlap.

4.2.2. Treatment effect estimates

The results of estimation excluding biological parameters produce compelling results. First, beginning by allowing splitting only on policy duration, results are broadly consistent with those from Costello et al. (2008): over the course of 30 years of policy implementation, the probability of collapse declines by around 0.005 per year. However, even without allowing for heterogeneity in other dimensions, the method already reveals substantial nonlinearities. Much of the decline in the probability of collapse with continued IQ implementation appears to accrue between years 5-10 and 20-28, with no significant affect in years 1-4, and a relatively flat trend between years 10-20. This suggests that expectations of immediate benefits may need to be tempered, and that reductions in the probability of collapse may not be steady.

Relaxing assumptions to allow policy effects to also depend on catch levels just prior to the policy change reveals a richer picture. Figure 6 depicts partial effects of policy duration at three different levels of pre-treatment relative catch (10%, 50%, and 90%, top to bottom). When catch is already low prior to IQ implementation (top panel), the policy change has no significant effect on the probability of collapse regardless of policy duration. With pre-treatment catch at more moderate levels, the effects of duration are more similar to those uncovered in the first model allowing heterogeneity only based on duration. Finally, when pre-treatment catch is already high, the policy yields significant reductions in the probability of collapse for more than half of the possible policy durations, but point estimates exhibit a much weaker time dependency. In short, duration matters for policy effects, but in a way that depends upon the history of the resource. This interaction highlights the benefit of using tree-based methods to examine temporal and cross-sectional heterogeneity simultaneously.

Allowing for heterogeneity in further dimensions yields intuitive results. Figure 7 uses partial dependence plots to examine the effect of the start year of IQ implementation (at 5 year intervals) as a function of the pre-implementation relative catch trends ranging from -0.4 to 0 to 0.4 (top to bottom). In all cases, pre-IQ relative catch is fixed at 0.5 and duration is 15 years. Comparison of the panels reveals that IQs tend to be more beneficial when catch was changing prior to implementation, whereas effects are smaller in magnitude and less precisely estimated when catch was stable prior to implementation. In addition, policies implemented since 2000 appear to have a less beneficial and weaker effect. Including species growth rates reveals little evidence of heterogeneity (Figure 8), though what differences there are in point estimates suggest stronger reductions in the probability of collapse for slower growing species. One plausible explanation is that faster growing species are less susceptible to collapse overall, so that management interventions will have less effect.

On a final note, a caveat on interpretation of results is in order. In some fisheries the introduction of individual quotas represents the first introduction of any kind of binding harvest quota. For those cases any improved outcomes reflect the composite effect of both changes, which could differ from the effect of introducing individual-level rights alone. Nevertheless, the preceding analysis highlights important heterogeneity in how the probability of collapse under IQ management differs from that without it.

5. Conclusion

Public policies are likely to impact people (or firms, or countries, etc) in different ways, and those effects may vary depending on when a policy first took effect and for how long it has been in place. Learning more about that heterogeneity has many potential benefits. At the least, we can set expectations as to when policy effects might start to arise for a new policy and who stands to benefit the most from it. Knowledge of heterogeneity in policy effects might also help better target a particular type of policy to settings in which it is likely to have the most beneficial effect.

This paper has shown how to adapt a recently proposed suite of methods – Causal Trees (Athey and Imbens 2016) and Causal Forests (Wager and Athey 2017) – to examine how policy effects vary not only cross-sectionally, but also with both policy duration and calendar time. The estimation approach uses a modified regression tree procedure to identify subpopulations in which the treatment effect is relatively constant, and makes use of inverse propensity score weighting to construct estimates in each subsample. The key benefit of this method is that it requires weaker assumptions about how policy effects vary. The researcher need only specify which variables might influence policy effects, but not the functional form of that influence. This allows data to potentially reveal a richer pattern of heterogeneity than what may be hypothesized.

Applying these methods to examine the effect of a market-based natural resource management policy has revealed considerable heterogeneity across fisheries and across time. Individual Quota programs substantively reduce the probability of fishery collapse for some fisheries, while having no or even detrimental estimated effects for others. Moreover, the strongest effects take years to materialize and do not always follow a linear pattern. These findings suggest that application of these methods to other policy questions could reveal heterogeneity that has been previously ignored or assumed away.

6. Appendix

Dataset construction

The list of fisheries subjected to IQ management was based on the database of catch share programs provided by the Environmental Defense Fund. Since that database provides a single start year for an entire management program, which may involve multiple species, start dates for each species in each LME were subsequently compiled from published government sources or academic research papers. In some cases, a single species may be subjected to several forms of management in a single LME. Some LMEs span national jurisdictions, and even within a jurisdiction and LME, different groups of vessels (e.g. those using different fishing gear) may be subjected to different management. The start date used for IQs in such cases was the earliest year in which the species was subjected to IQ management in that LME. A more refined definition of a fishery could be used to avoid this issue, but defining a fishery as an LME x species pair allows for a clean comparison with Costello et al. (2008).

Catch data from the Sea Around Us project was downloaded per LME, and records were restricted to reported landings data (exluding estimates of unreported catch and fish which were discarded prior to returning to port).

As explained in the main text, species name and LME were used to match management information with catch records. Records from the management database without an exact (species x LME) match in the catch records were manually reviewed. Where possible, exact matches were supplemented with manual matches, correcting differences in spelling or the use of non-standard species synonyms in one source or the other. Catch records were matched to trait data from FishBase based on species names alone. If no exact match was found, several additional steps were used to increase coverage. First, the species name from the catch data was matched against genus names in FishBase. In some cases, catch data was actually reported at the genus level, so that the 'species' name actually represented a genus. Genus traits were computed as means and modes of species-level traits for species in that genus. Supplementary steps included attempting matches based on synonyms for species names used in the catch data, as well as manually mapping taxonomic family names used in catch data to a representative genus for matching to the genus-level traits just described.

References

Allcott, H., 2011. Social norms and energy conservation. Journal of Public Economics 95 (9), 1082–1095.

Allcott, H., Rogers, T., 2014. The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. The American Economic Review 104 (10), 3003–3037.

Asher, S., Nekipelov, D., Novosad, P., Ryan, S. P., December 2016. Classification trees for heterogeneous moment-based models. Working Paper 22976, National Bureau of Economic Research.
URL http://www.nber.org/papers/w22976

Athey, S., Imbens, G., 2016. Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences 113 (27), 7353–7360.

Auffhammer, M., Bento, A. M., Lowe, S. E., 2009. Measuring the effects of the clean air act amendments on ambient pm10 concentrations: The critical importance of a spatially disaggregated analysis. Journal of Environmental Economics and Management 58 (1), 15–26.

Auffhammer, M., Kellogg, R., 2011. Clearing the air? the effects of gasoline content regulation on air quality. The American Economic Review, 2687–2722.

Austin, P. C., Stuart, E. A., 2015. Moving towards best practice when using inverse probability of treatment weighting (iptw) using the propensity score to estimate causal treatment effects in observational studies. Statistics in medicine 34 (28), 3661–3679.

Azoulay, P., Ding, W., Stuart, T., 2009. The impact of academic patenting on the rate, quality and direction of (public) research output. The Journal of Industrial Economics 57 (4), 637–676.

Ban, N. C., Caldwell, I. R., Green, T. L., Morgan, S. K., O'Donnell, K., Selgrath, J. C., 2009. Diverse fisheries require diverse solutions. Science 323 (5912), 338–339.

Bandiera, O., Larcinese, V., Rasul, I., 2010. Heterogeneous class size effects: New evidence from a panel of university students. The Economic Journal 120 (549), 1365–1398.

Becker, S. O., Egger, P. H., Von Ehrlich, M., 2013. Absorptive capacity and the growth and investment effects of regional transfers: A regression discontinuity design with heterogeneous treatment effects. American Economic Journal: Economic Policy 5 (4), 29–77.

Bertrand, M., Crépon, B., Marguerie, A., Premand, P., 2017. Contemporaneous and post-program impacts of a public works program: Evidence from côte d'ivoire. Working paper, World Bank Group.
URL `http://documents.worldbank.org/curated/en/361281506439891614/Contemporaneous-and-p`

Branch, T. A., Jensen, O. P., Ricard, D., Ye, Y., Hilborn, R., 2011. Contrasting global trends in marine fishery status obtained from catches and from stock assessments. Conservation Biology 25 (4), 777–786.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A., 1984. Classification and regression trees. CRC press.

Chin, S., Kahn, M. E., Moon, H. R., April 2017. Estimating the gains from new rail transit investment: A machine learning tree approach. Working Paper 23326, National Bureau of Economic Research.
URL `http://www.nber.org/papers/w23326`

Chou, P. A., 1991. Optimal partitioning for classification and regression trees. IEEE Transactions on Pattern Analysis & Machine Intelligence (4), 340–354.

Copes, P., et al., 1986. A critical review of the individual quota as a device in fisheries management. Land economics 62 (3), 278–291.

Costello, C., Gaines, S. D., Lynham, J., 2008. Can catch shares prevent fisheries collapse? Science 321 (5896), 1678–1681.

Costello, C., Grainger, C., forthcoming. Property rights, regulatory capture, and exploitation of natural resources. Journal of the Association of Environmental and Resource Economists.

Davis, J. M., Heller, S. B., 2017. Using causal forests to predict treatment heterogeneity: An application to summer jobs. American Economic Review: Papers and Proceedings 107 (5), 546–550.

de Mutsert, K., Cowan, J. H., Essington, T. E., Hilborn, R., 2008. Reanalyses of gulf of mexico fisheries data: landings can be misleading in assessments of fisheries and fisheries ecosystems. Proceedings of the National Academy of Sciences 105 (7), 2740–2744.

Fisher, W. D., 1958. On grouping for maximum homogeneity. Journal of the American statistical Association 53 (284), 789–798.

Fowlie, M., 2010. Emissions trading, electricity restructing, and investment in pollution abatement. The American Economic Review, 837–869.

Gerfin, M., Lechner, M., 2002. A microeconometric evaluation of the active labour market policy in switzerland. The Economic Journal 112 (482), 854–893.

Gutiérrez, N. L., Hilborn, R., Defeo, O., 2011. Leadership, social capital and incentives promote successful fisheries. Nature 470 (7334), 386–389.

Hamilton, J. T., 1995. Pollution as news: media and stock market reactions to the toxics release inventory data. Journal of Environmental Economics and Management 28 (1), 98–113.

Handel, B., Kolstad, J., 2017. Wearable technologies and health behaviors: New data and new methods to understand population health. American Economic Review 107 (5), 481–85.

Imbens, G. W., 2015. Matching methods in practice: Three examples. Journal of Human Resources 50 (2), 373–419.

Kapetanios, G., 2008. A bootstrap procedure for panel data sets with many cross-sectional units. The Econometrics Journal 11 (2), 377–395.

Knaus, M., Lechner, M., Strittmatter, A., 2017. Heterogeneous employment effects of job search programmes: A machine learning approach. Working Paper 23326, IZA Institute of Labor Economics.

Lalive, R., Van Ours, J. C., Zweimüller, J., 2008. The impact of active labour market programmes on the duration of unemployment in switzerland. The Economic Journal 118 (525), 235–257.

Lechner, M., Miquel, R., 2010. Identification of the effects of dynamic treatments by sequential conditional independence assumptions. Empirical Economics 39 (1), 111–137.

Melnychuk, M. C., Essington, T. E., Branch, T. A., Heppell, S. S., Jensen, O. P., Link, J. S., Martell, S. J., Parma, A. M., Smith, A. D., 2016. Which design elements of individual quota fisheries help to achieve management objectives? Fish and fisheries 17 (1), 126–142.

Mora, C., Myers, R. A., Coll, M., Libralato, S., Pitcher, T. J., Sumaila, R. U., Zeller, D., Watson, R., Gaston, K. J., Worm, B., 06 2009. Management effectiveness of the world's marine fisheries. PLOS Biology 7 (6), 1–11.

Pinkerton, E., Edwards, D. N., 2009. The elephant in the room: the hidden costs of leasing individual transferable fishing quotas. Marine Policy 33 (4), 707–713.

Robins, J. M., Hernán, M. Á., Brumback, B., 2000. Marginal structural models and causal inference in epidemiology. Epidemiology, 550–560.

Rubin, D. B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational Psychology 66 (5), 688.

Shimshack, J. P., Ward, M. B., Beatty, T. K., 2007. Mercury advisories: information, education, and fish consumption. Journal of Environmental Economics and Management 53 (2), 158–179.

Wager, S., Athey, S., 2017. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association.

Wilberg, M. J., Miller, T. J., 2007. Comment on "impacts of biodiversity loss on ocean ecosystem services". Science 316 (5829), 1285–1285.

Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., Jackson, J. B., Lotze, H. K., Micheli, F., Palumbi, S. R., et al., 2006. Impacts of biodiversity loss on ocean ecosystem services. science 314 (5800), 787–790.

Tables

	(1)	(2)
Ever IQ	-0.076^{***}	-0.064^{***}
	(0.003)	(0.004)
Year	0.005^{***}	0.005^{***}
	(0.000)	(0.000)
Is IQ x Years in IQ	-0.006^{***}	0.009^{***}
	(0.001)	(0.001)
Is IQ x Pre-IQ catch		-0.407^{***}
		(0.032)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Logit estimation of collapse using (1) model in Costello et al. (2008) and (2) adding interaction between treatment indicator and relative catch in the year prior to IQ implementation.

Figures

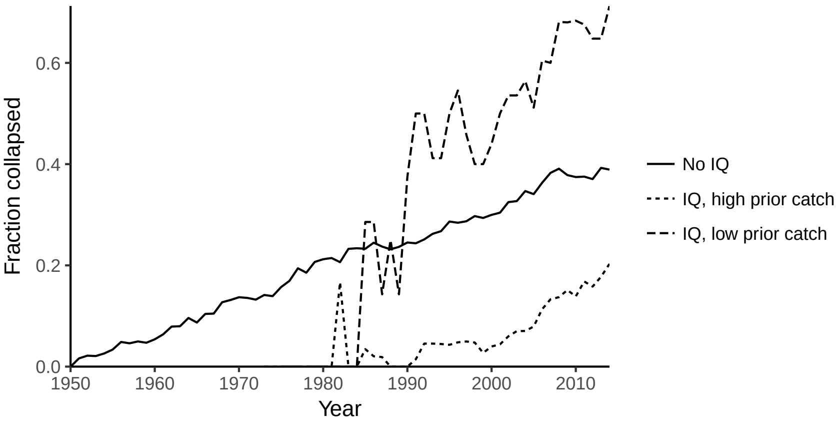

Figure 1: Fraction of fisheries collapsed in a given year in three groups: those not under IQ management (solid), those under IQ management with low catch ($\leq$ 20% of historical max) just prior to IQ implementation (dotted, and those under IQ management with high catch ($>$ 20% of historical max) just prior to IQ implementation (dashed).

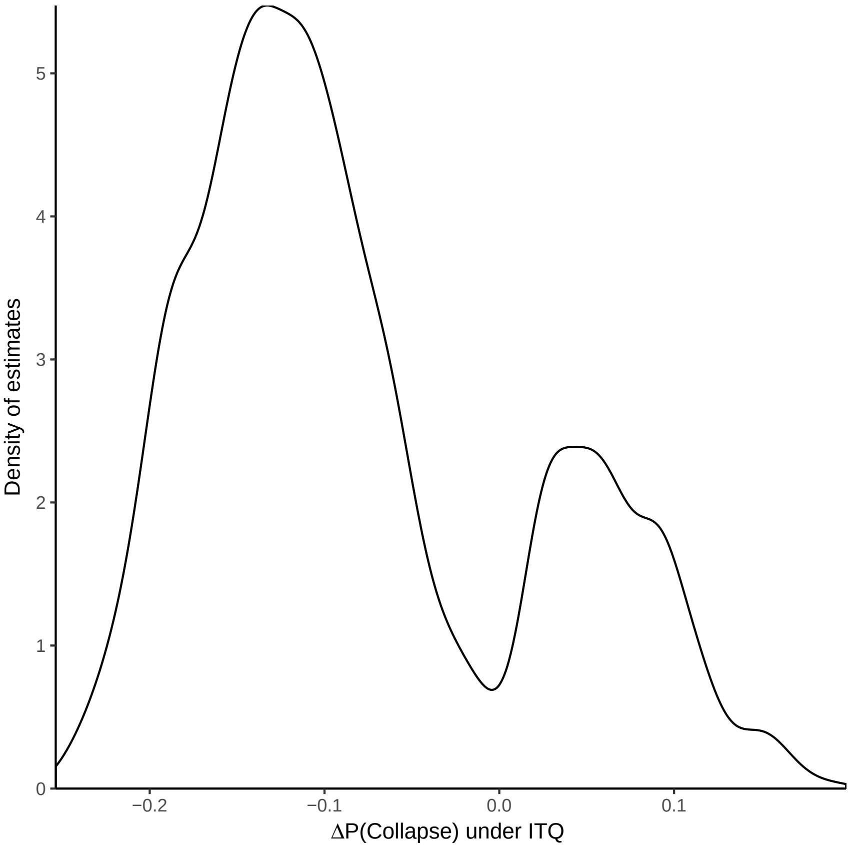

Figure 2: Density of estimated treatment effects in model allowing splits on pre-IQ catch and catch trends, potential duration, and potential start year. Negative numbers represent reduction in the probability of fishery collapse.

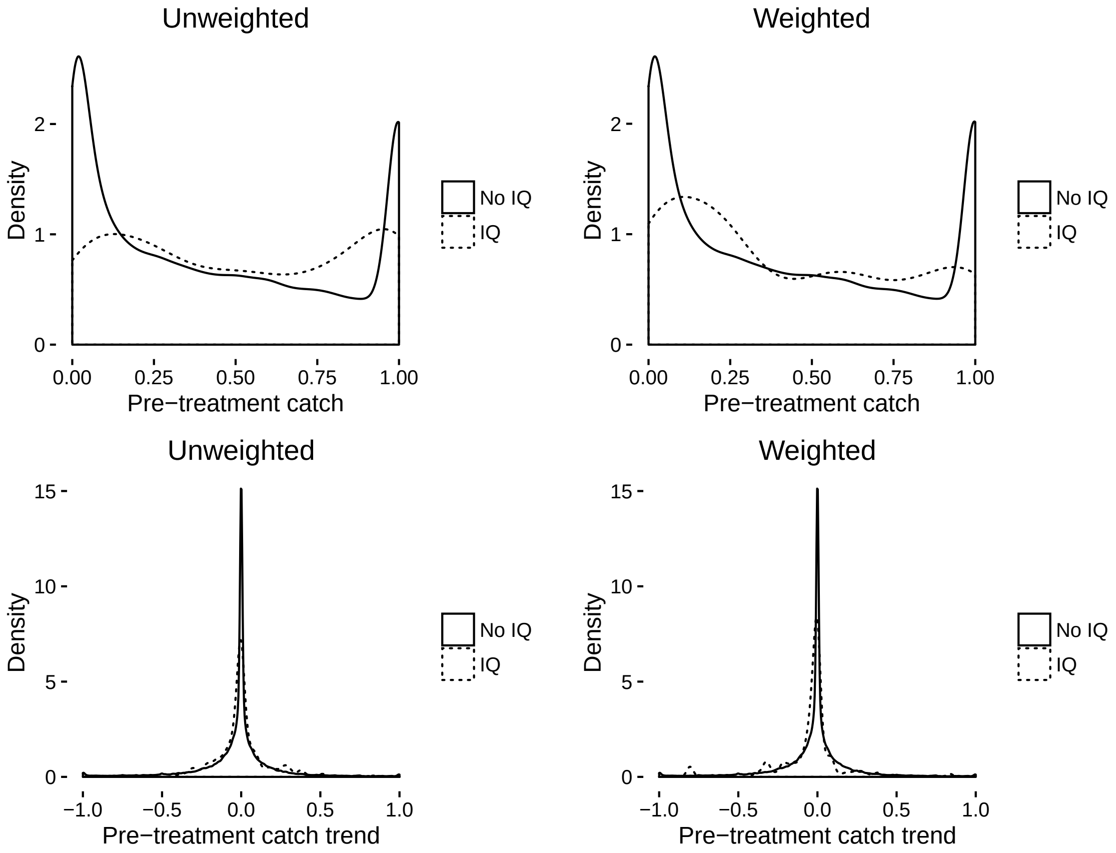

Figure 3: Covariate balance across treatment status before (left column) and after (right column) reweighting by stabilized inverse probability of transition from non-IQ to IQ management. Solid lines represent fisheries that do not transition to IQ management (control group); dashed lines correspond to fisheries that do transition to IQ management (treatment group).

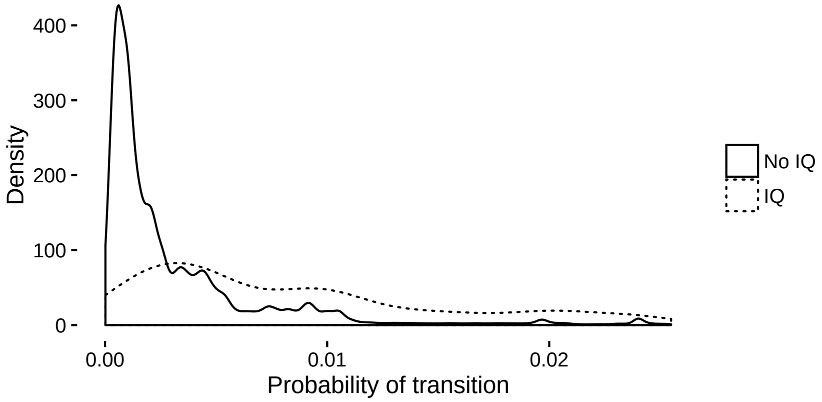

Figure 4: Densities of probability of transition from non-IQ to IQ management among fisheries that do (dashed line) and do not (solid line) transition.

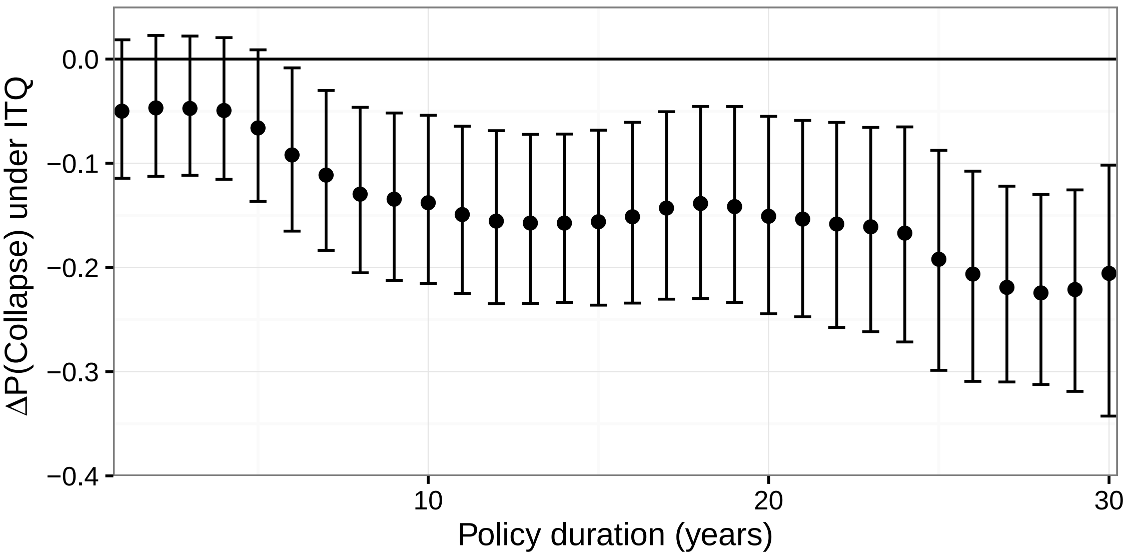

Figure 5: Change in probability of collapse due to IQs as a function of the number of years IQ management has been in effect. Points represent mean estimates and error bars coincide with 95% confidence intervals, with all quantities computed via panel bootstrapping at the fishery level.

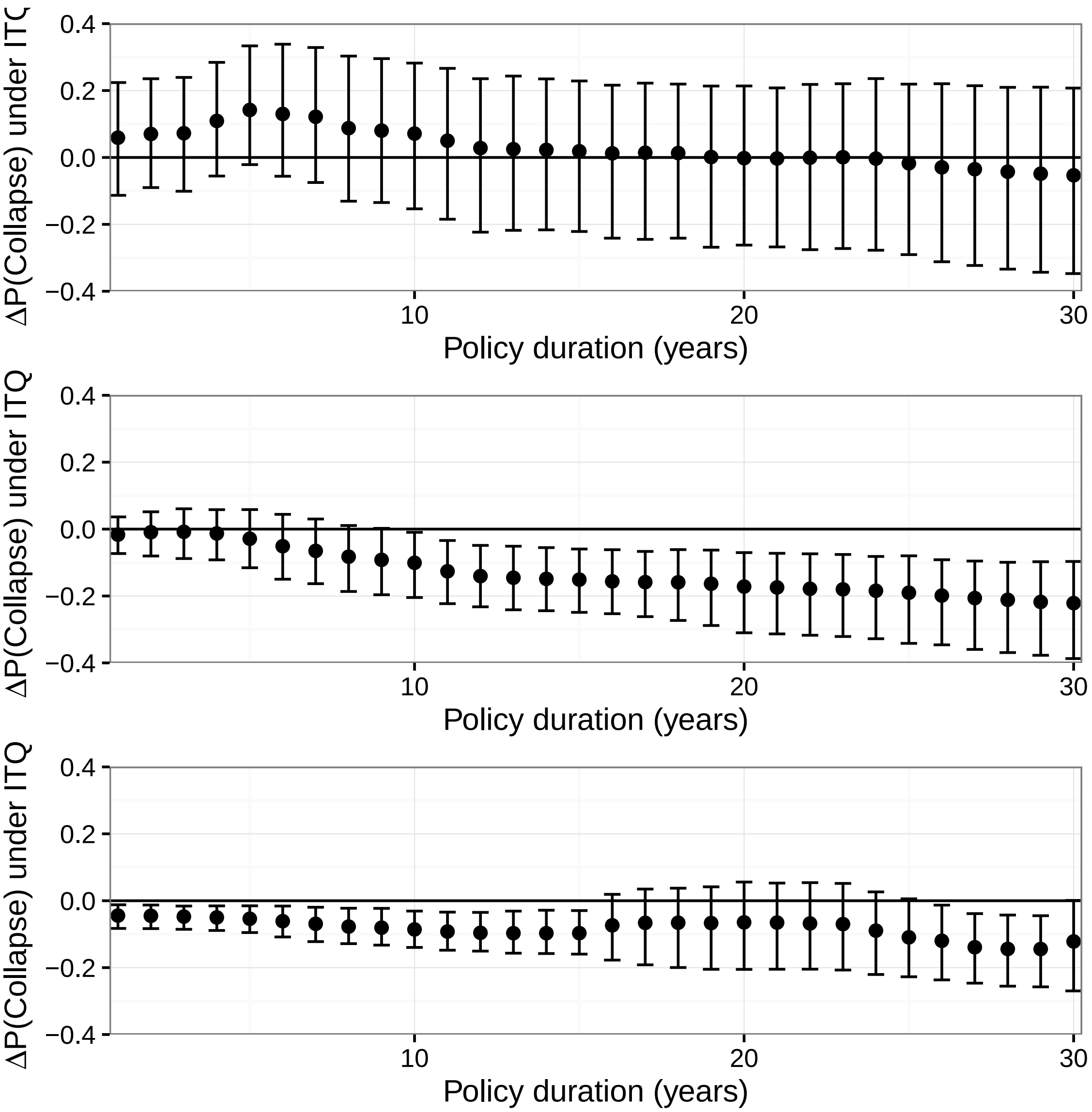

Figure 6: Change in probability of collapse due to IQs as a function of the number of years IQ management has been in effect. Panels, top to bottom, represent partial effects with pre-treatment relative catch at 10%, 50%, and 90% of the historical maximum catch in that fishery. Points represent mean estimates and error bars coincide with 95% confidence intervals, with all quantities computed via panel bootstrapping at the fishery level.

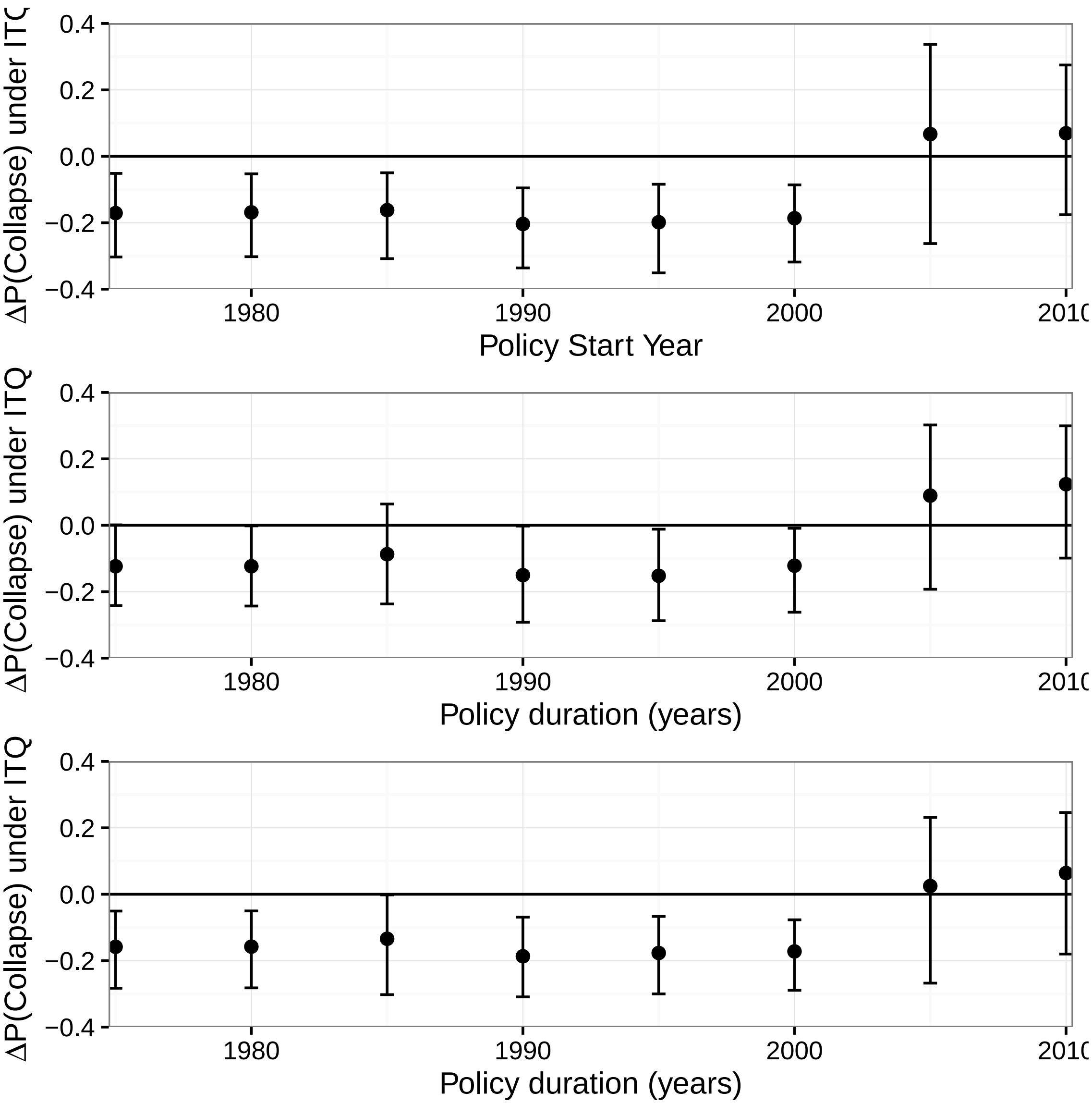

Figure 7: Change in probability of collapse due to IQs as a function of the first year of implementation (grouped by 5 year intervals). Panels, top to bottom, represent partial effects with pre-treatment relative catch trends at -40%, 0%, and 40% of the historical maximum catch in that fishery. Points represent mean estimates and error bars coincide with 95% confidence intervals, with all quantities computed via panel bootstrapping at the fishery level.

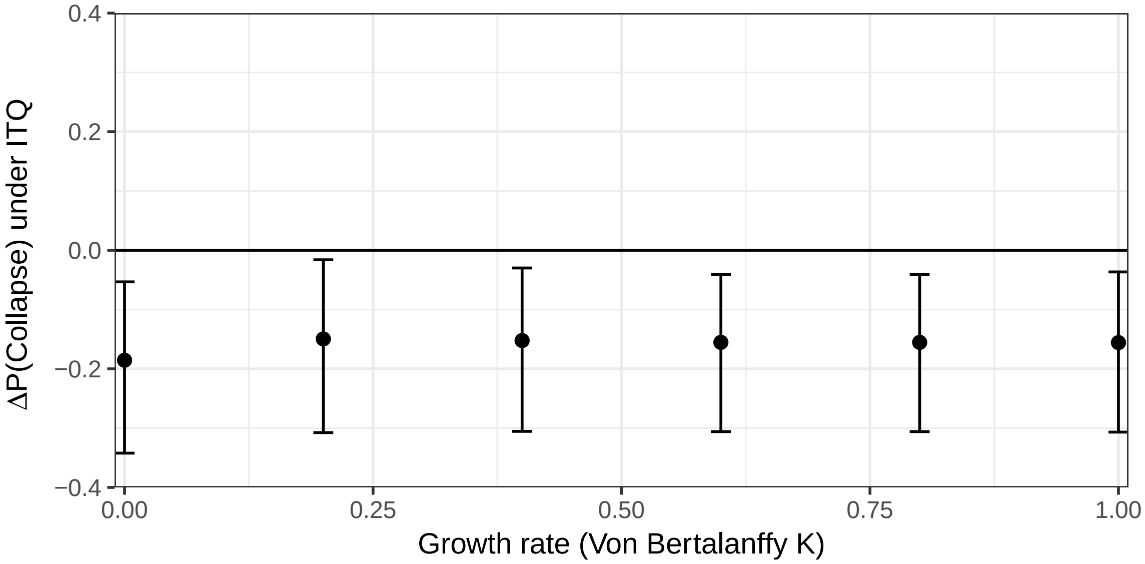

Figure 8: Change in probability of collapse due to IQs as a function of Von-Bertalanffy species growth. Partial effects computed for IQ start year of 1990, policy duration of 15 years, and pre-treatment relative catch and catch trends of 0.5 and 0.4, respectively. Points represent mean estimates and error bars coincide with 95% confidence intervals, with all quantities computed via panel bootstrapping at the fishery level.