# The Effectiveness of Listing under the U.S. Endangered Species Act

## An econometric analysis using matching methods

Paul J. Ferraro[1], Craig McIntosh[2] and Monica Ospina[3]

April 2006

**Abstract:**

Diametrically opposed views of the effectiveness of the United States Endangered Species Act (ESA) co-exist more than 30 years after the Act's creation. The evidence marshaled to date for and against the ESA suffers from a problem common in analyses of biodiversity protection measures: the absence of a well-chosen control group. We demonstrate how state-of-the-art statistical methods can be used to select such a control group and thereby estimate how species listed under the ESA would have fared had they not been listed. Our results show that listing a species under the ESA is, on average, detrimental to species recovery if not combined with substantial government funds. In contrast, listed species with such funding tend to improve. Our analysis offers not only new insights into a controversial debate, but also a methodology to guide conservation scientists in evaluating the effectiveness of society's responses to biodiversity loss.

[1] Assistant Professor, Department of Economics, Andrew Young School of Policy Studies, P.O. Box 3992, Georgia State University, Atlanta, GA 30302-3992; pferraro@gsu.edu
[2] Assistant Professor, School of International Relations and Pacific Studies, University of California-San Diego, San Diego, CA; ctmcintosh@ucsd.edu
[3] Graduate Research Assistant, Department of Economics, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA; prcmox@langate.gsu.e
**Acknowledgements.** The authors thank Andrew Metrick for providing us with the DEMES database, and Kiernan Suckling, Timothy Male, Bruce Stein, members of the Atlanta Audubon Society, and participants at the 2004 CAMP Resources Workshop, the 2005 Association of Environmental and Resource Economists Workshop and the 2005 Occasional California Workshop at UCSB's Bren School for providing useful comments.

**The Effectiveness of Listing under the U.S. Endangered Species Act: An econometric analysis using matching methods**

**Abstract**

Diametrically opposed views of the effectiveness of the United States Endangered Species Act (ESA) co-exist more than 30 years after the Act's creation. The evidence marshaled to date for and against the ESA suffers from a problem common in analyses of biodiversity protection measures: the absence of a well-chosen control group. We demonstrate how state-of-the-art statistical methods can be used to select such a control group and thereby estimate how species listed under the ESA would have fared had they not been listed. Our results show that listing a species under the ESA is, on average, detrimental to species recovery if not combined with substantial government funds. In contrast, listed species with such funding tend to improve. Our analysis offers not only new insights into a controversial debate, but also a methodology to guide conservation scientists in evaluating the effectiveness of society's responses to biodiversity loss.

## 1. Introduction

The Endangered Species Act (ESA) is the most important piece of biodiversity legislation in the United States, but its success at protecting listed species is hotly debated (Dalton and Marris 2005; Stokstad 2005). Supporters call the ESA the "crown jewel" of the nation's environmental legislation and an absolutely essential tool for protecting biodiversity. Opponents claim that the ESA imposes unreasonable costs on society while delivering few benefits. Such ambiguity exists more than thirty years after the Act's creation because the ESA was never designed to be evaluated, and a complex set of biological and political factors affects its implementation. The nonrandomized nature of the ESA makes measuring its effect on species recovery difficult. As we will demonstrate, recent efforts have relied on identifying assumptions that are likely to be flawed. This paper uses matching estimators and the most complete set of covariates employed to date in order to analyze the effect of the ESA on vertebrate species.

As a starting point for our analysis, we study the ESA's listing process. As in Metrick and Weitzman (1996), we find that non-scientific variables have strong effects on the probability of listing. By dividing listed species into cohorts, we then show that a 1982 reform to listing guidelines had the desired effect of concentrating the listing process on scientific considerations. These changes over time generate bias in the two most recent published analyses of the effects of the ESA, which use duration of listing to identify treatment effects (Male and Bean 2005; Taylor et al. 2005). In contrast, we incorporate into our matching estimators several variables that describe the environmental preferences of the politicians who represent a species' habitat. This approach allows us to select counterfactual species that share charismatic, political and scientific characteristics with listed species. Matching methods offer a flexible way to control for the complex and shifting set of characteristics that determine selection.

Using a variety of distance metrics, we calculate counterfactuals by averaging outcomes among the four unlisted species that are most similar to each ESA listed species  Using this set of matched species, we compare changes in endangerment status between 1993 and 2004.    This approach makes substantial improvements on efforts to identify the effects of the ESA.  First, the approach has an intuitive simplicity that makes the resulting estimates transparent.  Second, matching offers a semi-parametric way of comparing listed and unlisted species, thereby reducing specification bias.  Third, our expanded set of covariates incorporates variables on environmental voting records for politicians representing a species' habitat, and so our counterfactuals include the politics of the listing process in a novel way.

We measure impacts of two related treatments; the effects of listing under the ESA and the effects of expenditures to species-specific recovery plans.  The overall effect of listing under the ESA is insignificant, but point estimates are consistently negative.  In contrast, listing and funding together are strongly effective, indicating that focused and well-funded recovery efforts can work.  However, in line with theoretical predictions (Brown & Shogren 1998, Innes 1997; Polasky and Doremus 1998), we find that species that are listed with little or no funding experience substantially worse outcomes than the comparison group.

Our impact estimators are non-experimental, and so will contain bias if the listing process is determined by variables that are unobservable to us.  Because matching permits a great deal of flexibility in defining how the counterfactual is formed, we compare different ways of using the observable data and find the estimated impacts to be robust.  Our conclusions are also robust to alternative measures of species recovery.  As a final robustness check, we use Rosenbaum bounds to measure how strong an unobserved variable's effect on selection would have to be in order to undermine our conclusions.  The results indicate that the estimated positive effect of listing with funding is robust to the presence of unobserved heterogeneity, while the negative effect of listing without funding is more sensitive to such heterogeneity.

Understanding the efficacy of the ESA is a crucial step in efforts to protect North American biodiversity and to improve our understanding of species-specific environmental regulations. Our results indicate that success can be achieved when the ESA is combined with substantial species-specific spending, but listing in the absence of any funding appears to have adverse consequences for species recovery. The implication is that scarce biodiversity conservation funds would be better spent *in situ* rather than on the contentious listing process. We conclude by offering hypotheses for further testing and by describing experimental and quasi-experimental methods with which to test these hypotheses.

## 2. The controversy over the ESA

Enacted in 1973, the ESA has the simple but ambitious objective of reversing the decline of endangered species. Two important sections of the ESA are sections 7 and 9. Under section 7, federal agencies must insure that actions they take, fund, or authorize are not likely to jeopardize the continued existence of any listed species. Section 9 prohibits commerce in and the "taking" of endangered animal species. The prohibition on taking includes killing or capturing the animal, as well as harm caused by substantial habitat modification. The ESA provides no protection to species that are not officially listed under the ESA.

Both opponents and supporters are concerned that the ESA, as currently implemented, may be ineffective at halting or reversing the decline of endangered species. Constraints on its effectiveness arise from the ESA’s species-level rather than ecosystem-level focus, vague or contradictory legislative rulings, interest group pressures that warp listing decisions, and landowner actions that preemptively harm species and their habitat in order to avoid regulatory burdens (for an economist's perspective, see Brown and Shogren 1998; for a biologist's perspective, see Wilcove et al. 1993).

Although the ESA and its controversy have been around for decades, there are surprisingly few empirical studies of its effectiveness. A few studies have attempted to empirically verify widely cited anecdotes and theoretical modeling of preemptive habitat destruction (for anecdotes, see Mann and Plummer 1995; Hollingsworth 1998; and Stroup 1995; for theory, see Innes 1997; Polasky and Doremus 1998). Lueck and Michael (2003) and Zhang (2004), in the context of red-cockaded woodpeckers, and Margolis et al. (2004), in the context of pygmy owls, present data that are consistent with preemptive habitat manipulation. However, it is unclear if two cases can be extrapolated to nationwide trends.

ESA opponents also point to the paucity of delisted species as evidence of the ESA's failure (e.g., Mann and Plummer; Gordon et al. 1997). Of the twenty-four delisted U.S. species, sixteen were delisted because they went extinct or because the original data were in error.[1] However, as noted by Doremus and Pagel (2001), delisting requires ongoing protective measures, which under current law are often effectively provided only by the ESA itself. Consequently, lengthy residence on the list should be expected for the majority of species.

Opponents, however, are also able to point to the paucity of listed species exhibiting recovery (improved status). According to the Fish and Wildlife Service's (FWS) 1992 biannual report on the ESA, less than 10% of all listed species were known to be improving. The observation that listed species continue to decline may seem to imply that the ESA is failing, but

[1] From http://ecos.fws.gov/tess_public/TESSWebpageDelisted?listings=0, last accessed 3/5/05. Mann and Plummer assert that only the delisting of the American alligator and Aleutian Canada goose (a subspecies of the widespread Canadian Goose) can be attributed to the ESA. Other cases, such as the brown pelican and peregrine falcon that benefited from the ban of the pesticide DDT, recovered from actions taken outside the context of the ESA.

the observation lacks consideration of the counterfactual: would listed species be even worse off had they not been listed? As we will demonstrate in section 3, listed species tend to be more endangered, on average, than other species. Thus the fact that the average listed species has seen no improvement in its status or does not seem to be doing any better than the average unlisted species does not necessarily imply the ESA is ineffective. The average unlisted species does not represent what the average listed species would look like in the absence of the ESA.

Like their opponents, supporters of the ESA marshal evidence to support their view that the ESA is effective despite its limitations. For example, ESA proponents claim that the ESA may not have led to improvements in the populations of many species, but it has prevented hundreds of extinctions (NRC 1995, Schwartz 1999). However, determining if a species is in fact extinct is difficult. Ascribing the absence of the extinction to the ESA is even more difficult. Suckling et al. (2004) note that of the one hundred eight U.S. species known to have become extinct in the first 21 years following the creation of the ESA, the majority (85) were unlisted species. However, the pool of unlisted species is much larger than the pool of listed species and thus absolute numbers of extinctions do not reveal much.

Recently, more sophisticated analyses of the ESA's effect have been undertaken, but we argue that these analyses are based on erroneous identifying assumptions. Building on a simpler analysis by Rachlinski (1997), Taylor et al. (2005) use logistic regression to examine the effect of the number of years listed on the likelihood that a species has been declared to be recovering or declining by the FWS in the 1990s. They find that the longer a species was listed, the more likely it was to be improving and the less likely it was to be declining. Although this analysis makes great improvements over previous arguments for and against the act, it still suffers from two important shortcomings. The first, and less serious, is the use of FWS measures of species status as the outcome variable. The FWS measures of species status have been criticized as too subjective (NRC) and, given they are not constructed using transparent criteria, may be manipulated to achieve agency objectives.[2]

A more important shortcoming is the use of a flawed counterfactual. The sample consists of only listed species, and impact is identified by comparing outcomes across the duration of listing. For this counterfactual to be valid, the listing process must be the same over time. However, there is substantial evidence that the listing process has in fact changed over time (see section 4.1). Evaluating the direction of the bias is not straightforward. Many of the first listed species may have been close to extinction and thus perhaps hard to recover (e.g., California Condor). The coefficient on "time listed" would thus be biased down. However, many of the first species listed may have also been simply charismatic species that, while certainly threatened, were relatively easier to recover (e.g., birds of prey, which could be recovered without listing through the banning of DDT). The coefficient on "time listed" would thus be biased up. Male and Bean (2005) similarly identify impacts of the ESA off of duration of listing.

The only paper to use unlisted species to construct the counterfactual is a working paper by Kerkvliet and Langpap (2002). However, this paper has several potential problems: they estimate changes over only a four-year period, they include the outcome at the *end* of the period as an explanatory variable, and they compare listed species with <u>all</u> unlisted species (including unlisted species that were not endangered). We show in section 4.1 that listed and unlisted species are substantially different in characteristics that affect both selection into the ESA and recovery outcomes. Thus, the highly parametric estimator that they use (ordered probit) is likely

[2] Since at least 1994, the FWS has attempted to measure the correlation between the length of time that a species has been listed and its status measure in an attempt to show that the ESA is successful.

to be biased. In contrast, we use a semi-parametric matching estimator with an expanded set of controls and a sample restricted to only endangered species.

## 3. Data and Methodology

Our outcome variable is "change in endangerment status from 1993 to 2004." We use national endangerment scores from NatureServe, which is the most comprehensive measure of species endangerment for the set of listed and unlisted vertebrates. Based on the Natural Heritage Methodology, NatureServe's system assigns an endangerment score to each species on a scale of 0 (extinct) to 5 (least endangered). Each of the scores has a well-defined meaning and a serious effort is made to apply the scores consistently. We obtain 1993 scores from the DEMES database (29) and 2004 scores from NatureServe.

For reasons outlined in the Appendix, we limit our study to native endangered terrestrial and freshwater vertebrates that have full species status and are present in one or more of the 50 states. We include only endangered species (i.e., with 1993 scores of less than 4) and exclude any species that had scores indicating they were extinct or potentially extinct in 1993. Our sample consists of 135 listed species and 295 unlisted species.

The goal of program evaluation is to construct a proper counterfactual. Let $L_i$ denote the listing decision for the $i$th species, where $L_i = 1$ when the species is listed and $L_i = 0$ when it is not listed. Let $S_{1i}$ denote the $i$th species' endangerment status when it is listed and $S_{0i}$ denote status when the species is not listed. Higher values of $S$ are preferred (they imply more secure populations). Thus the expected change in status for a species that is listed (i.e., the average treatment effect on the treated, or ATT) is: ATT = $E[S_{1i} - S_{0i} \mid L_i = 1]$.

If the ESA is effective, ATT > 0. Of course, estimating ATT is not straightforward because we cannot observe the counterfactual, $E[S_{0i} \mid L_i = 1]$, which is what the species' status would have been had it not been listed. If listing is allocated randomly across species, then we can estimate the counterfactual simply by using the status of unlisted species because: $E[S_{0i} \mid L_i = 1] = E[S_{0i} \mid L_i = 0]$ (i.e., the expected status in the absence of the ESA is identical for listed and unlisted species).

However, decisions to list species under the ESA are determined by observable characteristics of the species and their circumstances (see section 4.1). Thus listed and unlisted species, on average, differ in characteristics that may also affect status changes after listing (i.e., propensity to recover or decline). In other words, in a model that regresses changes in a species' status on its ESA listing and other covariates likely to affect recovery, the error term will be correlated with one or more of the covariates, leading to a biased estimate of the ESA's effect.

In the presence of such potential bias, the methods of matching provide one way to assess the effect of listing under the ESA. Matching works by, *ex post*, identifying a comparison group that is "very similar" to the treatment group with only one key difference: the comparison group did not participate in the program of interest. Matching mimics random assignment through the *ex post* construction of a control group.

We adopt the method of nearest-neighbor matching, developed in Rosenbaum and Rubin (1983, 1984). This method has been used extensively in the labor (e.g., Dehejia and Wahba, 1999, 2002; Heckman *et al.*, 1997, 1998a, 1998b) and development economics literature (e.g., Deininger et al. 2004; Duflo 2001; Jalan and Ravallion 2003), as well as in a few published studies in environmental economics (Frederiksson and Millimet 2004; List et al. 2003). If the researcher can select observable characteristics so that any two species with the same value for these characteristics will display homogenous responses to the treatment, then the treatment effect can be measured without bias. Measuring the average treatment effect on the treated without bias requires that, given a vector of covariates, $Z$: $E[S_{0i} \mid L_i = 1, Z] = E[S_{0i} \mid L_i = 0, Z] = E[S_{0i} \mid Z]$. Called the "conditional independence assumption" (CIA), this assumption implies that given $Z$, the non-treated outcomes are what the treated outcomes would have been had they not been treated. This assumption requires that selection into treatment occurs only on observable characteristics. Hence an unbiased estimator requires that we have included all of the determinants of the political selection problem. Arguably one can satisfy this requirement in the case of the ESA because the species themselves exert no idiosyncratic influence, and so the problem is only one of *eligibility* and not one of self-selection.

Even given the CIA, however, matching on $Z$ can be difficult when one has more than a few covariates. Rosenbaum and Rubin (1983) proved that that matching on the probability of participating conditional on the vector of covariates ("propensity score") can achieve consistent estimates of the treatment effect in the same way as matching on all covariates. The propensity score can be estimated by a simple logit or probit model with the treatment as the dependent variable. The PSM method constructs probabilistically equivalent groups on observed characteristics. If matching is successful, the outcome of the matched, non-experimental control group is equivalent to the outcome of an experimental control group and thus serves as an appropriate counterfactual. Abadie and Imbens (2004) develop an alternative, bias-corrected matching estimator that uses a vector norm in the covariate space as the relevant distance metric (the norm is defined using the inverse variance weighting matrix to account for the difference in scale of the covariates). We use both matching methods in the analysis.

Based on narrative and empirical evidence presented in the next section, we match listed and unlisted species using covariates on taxonomy and size, 1993 NatureServe endangerment status, the amount of scientific interest and knowledge in a species (as proxied by scientific publications), and the historical environmental preferences of the citizens and legislators from the states in which the species are found (as proxied by more than two decades of Senate and Congressional voting records). Further details on these covariates and their sources, as well as other covariates considered, can be found in the Appendix.

One potential concern is that the listing process has been a perfect policy instrument that has listed all endangered species and thus there is no control group for very endangered, listed species. Table 1 shows how the distribution of 1993 NatureServe scores differs across treatment status. Although the listed species within the sample are more likely to be endangered than are the unlisted, it is clear by inspection that there is a common support (i.e., overlap at all levels).

## 4. Analysis

### 4.1. The Listing Process

The nature of the selection process is a central issue in any analysis of a non-randomized intervention. The ESA was an extension of the 1969 Endangered Species Conservation Act, which itself was an extension of the 1966 Endangered Species Preservation Act. The latter was the first comprehensive federal attempt to address species extinction. The 1966 Act directed the Secretary of Interior to determine which vertebrates were threatened with extinction. Thus some species listed under the 1973 ESA were identified as needing protection in 1967. However, the 1966 Act did not result in substantial actions to address species loss. Moreover, although the 1966 Act directed the Secretary to seek scientific advice, it did not require the use of particular criteria in making the listing decision. The 1973 ESA offered protection to plants as well as animals, authorized protection of species that had not yet reached the very brink of extinction, added prohibitions on private actions harmful to listed species, and subjected all federal agencies to an obligation not to jeopardize the continued existence of listed species (Doremus)[3].

In the late 1970s, a debate about ESA listing practices developed. Many believed that listing decisions were often driven more by politics and preferences than by science (Doremus). There was an official policy to favor “higher” animals in the following order: mammal, bird, fish, reptile, and amphibian (Metrick and Weitzman). The 1979 ESA amendments thus included an explicit requirement that the FWS perform a formal review to determine whether sufficient "scientific and biological data" existed to justify a listing proposal. Congress directed the FWS to communicate with "experts in the field" as part of the review process.

In the 1982 amendments, Congress required that listing decisions be made "solely" on the basis of the best available scientific and commercial information. Thus listing decisions were to be made based on scientific criteria without reference to the taxonomic preferences of FWS staff and the public, without pressure from politicians, and without consideration of economic costs (see discussion in Brown and Shogren, and Doremus). In the ESA listing context, science is synonymous with biology. Doremus notes that, “Congress repeatedly equated the two” and that “a listing petition need only present ‘biological information.’”

In response to the 1982 amendments, the FWS created formal guidelines to guide its expenditure decisions. The FWS created a “priority system” index, which took into account the degree of threat to the species, the size of the genus to which it belonged (i.e., its contribution to diversity), and its potential for recovery. Note that this index was used to guide expenditure

[3] For more complete discussions of the ESA and its history, see Yaffee (1982), Shogren (1999) and SELS (2001)

decisions (Brown and Shogren), not listing decisions, but it is likely that the first two components were important factors in listing decisions.

Although Congress made clear that listing decisions should be based on biological considerations alone, a survey of employee preferences in the FWS Office of Endangered Species found that employees ranked mammals and birds above fish, amphibians and reptiles (Brown 1990, cited by Brown and Shogren: 8). Likewise, Metrick and Weitzman's study (1996) of listed vertebrate found larger animals from higher forms of life (mammal, birds) were more likely to be listed. They note that larger species and higher life forms are likely to be "charismatic," and thus enjoy stronger political support than others (see Dawson and Shogren 2001 for alternative interpretations). Given that the senior staff in the Department of the Interior is appointed and that Congress controls the Department's budget, FWS staff may be unable to avoid political influence.

To confirm the narrative evidence that the determinants of the selection process are varied and have changed over time, we conduct an empirical analysis. Because the date of ESA listing is not observed in the control species, analysis of changes in the listing determinants over time is not easily performed through interactions. Instead, we divide the enlistees under the ESA into four roughly equally-sized cohorts. The cohorts, designated by the years of listing, are 1967, 1968-1982, and the post-FWS guideline cohorts of 1983-1988 and 1989-1993. If the guidelines were successful in de-politicizing the process, we should see political and charismatic characteristics becoming less important over time.

We run linear probability regressions with robust standard errors on each cohort. Our outcome variable is defined as 0 for all controls, and as 1 for the species that are listed in that cohort. Species that have already been listed are eliminated from subsequent cohorts, decreasing the numbers of observations for each regression.  We define a set of *political* explanatory variables that includes the average League of Conservation Voter scores for House and Senate representation (up to the end of each cohort), and the average annual number of pro-environment and pro-land-use congressional representatives.  The vector of *scientific* controls comprises the average number of journal citations in the years up to the end of the cohort, the 1993 Nature Serve score, and dummies for being from a very small genus and for being a monotype.  We also define a set of '*charisma'* variables that are the dummies for being a mammal or bird, and the log length of the species (see data appendix for more details on all of the variables used).

In Table 2, we report F-statistics of the hypothesis that the political, scientific and charisma controls are each jointly equal to zero.  While the scientific variables remain strongly significant in every cohort, the influence of the political variables diminishes over time and is entirely absent in the selection process by 1983.  The charisma variables, while seeing a brief resurgence in the 1983-89 cohort, also become insignificant in the final cohort.  We thus conclude that the FWS guidelines were successful, as their introduction coincides perfectly with the point at which politics ceases to influence listings under the ESA, and the process became less charisma-driven in the latest cohort.  While it is tempting to use the political changes as instruments, or to identify a difference-in-differences off of them, identification would be undermined by the endogeneity of changes in political preferences over time.

The fact that the listing process is based on observable characteristics is important because these same characteristics are also likely to affect the ability of a species to recover.  For example, larger animals and higher taxonomic classes tend to require larger habitat areas and reproduce more slowly.  They are also more likely to be seen as "worthy" of sacrifices by the general population to secure their recovery.  Thus the construction of a counterfactual in the evaluation of the ESA must take into account selection on these observable characteristics. Further, because some key covariates are not balanced (scientific publications), we suggest that matching is a natural way to proceed in forming impact estimates, because it is straightforward to ensure that the matched sample shares a common support with the treated species.

### 4.2. Impact Estimates

We consider three treatments: (a) being listed under the ESA between 1973 and 1993; (b) being listed and receiving "substantial" federal and state funds for recovery between 1989 and 1993; and (c) being listed but not receiving "substantial" federal and state funds for recovery between 1989 and 1993. For the latter two treatment effects, we define the set of species that received "substantial" funds as the top 25% of fund recipients. Collectively, these species received 95% of all funds allocated in our sample from 1989 to 1993, a period which corresponds to the first four years in which the FWS was required to conduct an annual accounting of "reasonably identifiable" expenditures associated with ESA (Appendix).

#### MATCHING ESTIMATORS

We estimate the average treatment effect on the treated (ATT) by comparing the change in endangerment scores from 1993 to 2004 between listed and unlisted species using nearest-neighbor matching (with replacement) across the full vector of control characteristics. We resolve the mean-variance tradeoff in the match quality by using four nearest neighbors; the counterfactual outcome is the average among these four.[4]

[4] This is a standard approach in nearest-neighbor matching; see Abadie & Imbens (2004) or McIntosh (2004).

Abadie and Imbens show that matching estimators in finite samples without exact matching (and two or more continuous covariates) include a conditional bias term. Thus we use their post-matching bias-correction procedure that removes the conditional bias asymptotically. Most analyses in the 1990s and early 2000s used simple bootstrapping methods to estimate the standard errors of the treatment effect. Abadie and Imbens point out, however, that simple bootstrapping methods ignore the variance that arises from multiple matches to the same unit. We use Abadie and Imbens' variance formula for the nearest-neighbor ATT estimator. They also provide a robust estimator that allows for heteroskedasticity, which allows the treatment effect to be nonconstant (i.e., outcome variance differs by treatment status and covariates). This modification is implemented by using a second matching procedure that matches treated units to treated units and control units to control units.

Using the Mahalanobis distance metric, which weights distances across different covariates by the inverse of their variances, we first match unconditionally to the nearest neighbors. Due to concerns raised by ecologists that matching a bird to a reptile, for example, cannot form a credible counterfactual, we then restrict the match to be exact on taxonomy, and choose the nearest neighbors within that group. Finally we use propensity scores to calculate the distance weighting metric so that 'nearest' neighbors are selected relative to their propensity to be listed under the ESA.

Table 3 presents the treatment effect estimates. The estimated treatment effect of listing alone is small, negative and not statistically different from zero in all specifications. Among listed species, however, we see sharply differentiated effects across funding status. Listing a species with funding increases its Nature Serve ranking by almost a full half point, and the effect is strongly significant. As a robustness check on this result, we also estimate a treatment effect for the top third of funding recipients, which receive 97% of all funds allocated in our sample (instead of the top quarter). The effect is smaller (0.2690), yet significantly different from zero (p=0.006). In contrast, species that are listed but do not receive funding show a decline that, while roughly half as strong in magnitude as the improvement in listed funded species, is strongly significant.[5]

### PEARSON CHI-SQUARED TESTS

One potential criticism of the analysis thus far is the use of the NatureServe scores as continuous variables. Although these scores represent underlying continuous variables (numbers of occurrences or individuals), one could argue that a more conservative approach would be to treat the scores as ordinal categories and consider changes in a species' status from 1993 to 2004 as represented by three ordinal values: more endangered, unchanged, and less endangered.

To measure the effect of listing and funding on this ordinal outcome, we conduct nearest-neighbor propensity-score matching using our vector of controls, and then use a simple Pearson's chi-square test to detect differences in the distribution of outcomes in the treatment and the counterfactual. Table 4 reports the outcomes given each treatment (Listing, and Listing and High Funding) as well as counterfactual outcomes in the matched controls.

We find no significant difference between treatment and control outcome distributions when the treatment is listing alone (upper panel). A more powerful effect is seen in the lower

---

[5]A natural baseline regression would be to use OLS. Ordinary Least Squares regression is problematic because it conflates the selection and outcome equations, and thus assumes that the determinants and coefficients be the same for these two processes. It also assumes the data generating process is linear in the parameters. Under any non-linearity in the joint selection and outcome equations, these equations are misspecified and thus inconsistent. OLS also uses control units for estimating the counterfactual regardless of whether they are on a common support with treated units. However, if one runs robust OLS regressions using our vector of controls and including the treatment terms described above as binary dummies, one will see all treatment effects shifted in the negative direction. Hence listing alone has a significant negative effect, listing with funding is positively significant only at the 90% level, and listing without funding has a t-statistic of -3.25. The full output is available upon request.

panel of Table 4, where we conduct the same exercise for the high funding treatment. The distributions imply that, compared to unlisted species, well-funded species were more likely to become less endangered and less likely to become more endangered. To conserve space, we do not present a table for the distributions of listed species without substantial funding and their matched counterfactuals, but we note the difference between the distributions is significant (Chi(2) = 6.89, p = 0.032) and implies listed, low-funded species are more likely to become more endangered and less likely to become less endangered. Conducting the matching exercises with our ordinal outcome variable coded as integers yields the qualitative results of Table 3.

To generate more intuitive estimates of the treatment effects in terms of the three ordinal outcomes, we conduct nearest-neighbor, propensity-score matching using our vector of controls, and then estimate marginal effects from a weighted, ordered probit using the matched listed and unlisted species (weighted by the number of times a control is matched with treated units).[6] With such a model, we estimate that listing increases the probability that a species becomes more endangered by 1.4% and decreases the probability that a species becomes less endangered by 0.7%. Listing with high funding decreases the probability that a species becomes more endangered by 9.6% and increases the probability that a species becomes less endangered by 8.4%. In contrast, listing <u>without</u> substantial federal funding increases the probability that a species becomes more endangered by 11.5% and decreases the likelihood that a species becomes less endangered by 7.4%.

## 5. Robustness Checks

### 5.1. Alternative Outcome Variable

---

[6] In addition to imposing parametric assumptions on the data, this approach generates standard errors that lack a variance adjustment for repeated matches (Abadie and Imbens). Marginal effects, however, remain consistent.

One concern with our analysis is the accuracy of the outcome variable we use: changes in NatureServe scores between 1993 and 2004. Changes in these scores may reflect declines or improvements in a species' population, but they also may reflect new information and hence measure scientific interest in a species. We therefore reanalyzed our data using NatureServe's "short-term trend" variable, which explicitly avoids attributing changes in status to new information (see Appendix).

Observations on the trend variable exist for 77% of our listed species but only 56% of our unlisted species. Thus one should treat any analysis using this outcome variable with caution, but we present an analysis because it provides evidence that our results are not likely to be artefacts of the outcome variable we use. Our estimate of the effect of listing becomes more negative (-0.1569) and significantly different from zero (p=0.067). Our estimate of the effect of listing without substantial funding remains substantially negative (-0.2360) and significant (p=0.012) (similar qualitative results come from propensity score matching). Our estimated positive effect of listing with funding remains positive (0.0435), but is no longer significantly different from zero (p= 0.810). Unlike the estimated treatment effects of listing and listing without funding, our estimate of the effect of substantial funding is highly unstable and varies greatly by the matching method used. For example, propensity score matching with a single nearest neighbor generates a treatment effect of 0.3125, but using four nearest-neighbors it yields -0.0859 (neither estimate is significantly different from zero). Using trend data, the ESA's impact is less favorable because many of the unlisted species with missing trend data do poorly in terms of changes in their NatureServe scores.

We are therefore able to confirm the negative effect of listing without funding using an alternative outcome variable, but our sample size does not allow us to confirm the positive effect of listing combined with funding using the same outcome variable.

### 5.2 Alternative Constraints on Selecting the Counterfactual

To address the concern that our control species may simply not be as endangered as listed species despite having, on average, the same NatureServe endangerment score, we further restrict the set of species from which the counterfactual is constructed. In 1980, the FWS began maintaining a list of "candidate species" that contained all species "being considered by the Secretary for listing as an endangered or threatened species but not yet the subject of a proposed rule" (50 CFR 424.02). We therefore restrict our set of unlisted species to those on the 1993 candidate list (categories 1 and 2, 1994 Federal Register). This restriction reduces our set of potential matches by 152 species.

Constructing our counterfactual from this reduced set does not change our conclusions. We continue to find listing has an insignificant effect on recovery, listing with funding has a substantial and significant ($p<0.01$) positive effect on recovery, and listing without funding has a substantial and significant ($p<0.01$) negative effect In fact, the treatment effect estimates for high funding are larger using only candidate species as controls (0.469 and 1.602 for the Mahalanobis matching), and the estimates for listing without funding are more negative (-0.265 and -0.203 for the Mahalanobis matching), with little change in the standard errors.

### 5.3. Sensitivity tests: Rosenbaum Bounds

Given the natural suspicion that some degree of selection bias might remain even after careful matching, we use Rosenbaum bounds (Rosenbaum 2002, Diprete & Gangl 1994) to determine how strongly an unmeasured confounding variable must affect selection into the treatments in order to undermine our conclusions. If the probability of agent *j* selecting into the treatment is $\pi_j$, the odds are then $\frac{\pi_j}{1-\pi_j}$. The log odds can be modeled as a generalized function of a vector of controls $x_j$ and a linear unobserved term, so $\log(\frac{\pi_j}{1-\pi_j}) = \kappa(x_j) + \gamma u_j$, where $u_j$ is an unobserved covariate scaled so that $0 \le u_j \le 1$.

Take a set of paired observations where one of each pair was treated and one was not, and identical observable covariates within pairs. In a randomized experiment or in a study free of bias, $\gamma = 0$. Thus under the null hypothesis of no treatment effect, the probability that the treated outcome is higher equals 0.5. In general, the possibility that $u_j$ is correlated with the outcome means that the mean difference between treated and control units may contain bias.

The odds ratio between unit *j* which receives the treatment and the matched control outcome *k* is: $\frac{\pi_j(1-\pi_k)}{\pi_k(1-\pi_j)} = \exp\{\gamma(u_j - u_k)\}$. Because of the bounds on $u_j$, a given value of $\gamma$ constrains the degree to which the difference between selection probabilities can be a result of hidden bias. Defining $\Gamma = e^{\gamma}$, setting $\gamma = 0$ and $\Gamma = 1$ implies that no hidden bias exists, and hence is equivalent to the usual regression assumptions. Increasing values of $\Gamma$ imply an increasingly important role for unobservables in the selection decision. The differences in outcomes between the treatment and control are calculated and ranked. A Wilcoxon’s signed rank statistic is then used to compare the sums of the ranks of the pairs in which the treatment was higher than the control (calculated using Stata code ‘rbounds’; Gangl 2004).

The intuitive interpretation of the statistic for different levels of $\Gamma$ is that matched species may differ in their odds of being listed by a factor of $\Gamma$ as a result of hidden bias. The higher the level of $\Gamma$ to which the difference remains significantly different from zero, the stronger the relationship is between treatment and differences in recovery.

The first column of Table 5 indicates that, using the sign rank test, the effect of listing is negative and just significant at the 90% level, but only in the complete absence of any hidden bias. The second column examines the sensitivity of our estimated positive treatment effect of the combination of funding and listing. Here, the effect remains significantly positive even in the presence of substantial unobserved bias. The results imply that if an unobserved covariate caused the odds ratio of listing to differ between listed and unlisted cases by a factor of as much as 5, the 90% confidence interval would still exclude zero. The third column reports the sensitivity of the treatment effect of being listed without funding. Here, if the odds of a declining species being listed for unobserved reasons were two times as high, the effect seen in these data would not remain significant at the 90% level. Note that the assumed unobserved covariate is a strong confounder: one that not only affects selection but also determines whether the recovery is better for the treatment or the control units for matched units in our sample.

Thus the positive estimated effects of listing with funding are robust to the presence of unobserved bias, while the negative effects of listing without funding are less so. However, if selection bias were to explain our estimates, the bias for listing and funding would have to work in opposite directions. To consider this possibility, we briefly highlight key aspects of the two decision processes underlying our treatments.

The listing process for the ESA prior to the 1982 FWS guidelines was ill-defined, but the guidelines restricted the allowable determinants of listing. Doremus (23) states, “Congress…

expressly restricted the scope of listing decisions, requiring that they be made 'solely' on the basis of the best available scientific and commercial information. This change was made to 'prevent non-biological considerations from affecting' listing decisions. The primary 'non-biological' considerations at issue were those included in the administration's economic impact analyses, that is, the economic costs of protecting species." Thus, the listing process followed rules that were not explicit but which sought to disallow all "non-scientific" determinants.

Funding decisions from 1989-1993, on the other hand, follow a set of guidelines that are explicit and include economic costs. Congress requires that expenditures on species should vary within an 18-point priority system (6) formed by considering "degree of threat," "recovery potential," "taxonomy," and "conflict with development." Given recovery potential is included in this metric, we should be concerned that some bias may exist across this selection criterion.

However, empirical analysis shows that one of the strongest determinants of funding decisions is whether the species is in conflict with development or other human activity (24). Species that are in conflict have substantially higher spending allotted to them. The conflict variable is more influential than other determinants that, on paper, are assigned greater weight in determining priority scores for funding decisions. *Ceteris paribus*, species facing conflict seem less likely to rebound. Hence the selection bias present in spending decisions may in fact be negative, which would strengthen our finding of a positive effect from substantial funding. It is thus difficult to tell any simple bias story that would generate the patterns in these data.

## 6. Discussion and Conclusion

The decision to list or fund a species is contentious, involving complex tradeoffs of scientific, political and financial concerns. Further, the selection process is affected by observable characteristics of species and the relative importance of these characteristics has changed over time. After controlling for selection bias through several different means, we find no evidence that listed species fare any better than their counterfactual unlisted species on average. In fact, listed species that receive little or no federal and state funding do worse on average than their counterfactual unlisted species. We do, however, find evidence that the <u>combination</u> of listing and funding for recovery efforts can be effective in assisting recovery.

One interpretation for these results is that the ESA is not effective, and only money works. However, because we do not observe any unlisted species that receive high funding, this conclusion cannot be clearly drawn from the evidence here. Rather, we find that the ESA works when it is backed up with money, and not otherwise. Why could this be the case? The most plausible explanation for the negative effect of listing alone is that, because the ESA imposes perverse incentives on private landowners, it causes them to undertake pre-emptive actions to eliminate listed species from their land (the so-called 'shoot, shovel, and shut up' response).[7] The penalties for doing so, if landowners are caught, are substantial, and so it may be the case that the species-specific funding creates a sufficient level of perceived monitoring to overcome these perverse incentives. Seen in this light, it is only when the ESA is enforced that it has any chance of being effective, and provided sufficient money to create this monitoring and enforcement effect, the ESA works.

Given the significance of expenditures for species recovery, identifying the channels through which spending achieves its impact becomes a central policy question. Unfortunately the DEMES expenditure data are aggregated by agency and not by use, and only in one year (1993) are land expenditures by the FWS split out as a separate category. However, analysis of these available data can help guide future research. Using the control variables and ordered

[7] In the early 1990s, nearly 80% of all listed species occur partially or entirely on private lands (FWS 1996).

probit model of section 4 (outcomes: improve, same, worsen), we observe that Forest Service spending has the strongest positive effect, followed by the Bureau of Land Management and the Fish and Wildlife Service. State spending appears to have no effect, despite the fact that states spend 38% of the money in the sample. Indeed, once we split out spending by the three 'effective' agencies, cumulative spending by all other agencies put together has no effect on outcomes whatsoever.[8] These results, when combined with our treatment effect estimates, suggest that more detailed analysis of the channels through which species-specific funds are spent will be a fruitful path of inquiry for future ESA research.

In terms of policy, the results of our empirical analyses indicate that the rancorous debate over listing more species under the ESA may be missing the point. Our analysis suggests that it is not listing that is effective, but rather high levels of expenditures for recovery combined with listing. Simply listing a species in the absence of such expenditures actually leads to a decline.

We recognize that these claims are controversial among conservationists and biologists. One might reasonably suspect that, despite the large set of covariates used in this analysis, some form of selection bias remains in our analyses (although in order to completely explain our results, the bias would need to have opposite signs for the listing and funding decisions). Indeed, such doubts can never be fully resolved using non-experimental data, which leads to the question of how we might design testing strategies for the impact of the ESA that are more robust.

Randomization of listing under the ESA is unlikely to ever be politically or legally feasible. Funding decisions, on the other hand, are much more amenable to experimental methods because such funds will necessarily be scarce, and so some rationing rule must exist. Instead of funding only the best or the worst candidates, important research questions could be

[8] These other agencies are States (38%), the Army Corps of Engineers (3%), the Federal Highway Administration (2%), the National Parks Service (2%), land acquisition by the FWS (1.5%), and an assortment of other federal agencies, which collectively spend 15% of the money in the sample.

answered by pairing up species *ex ante* in terms of their predicted recovery probabilities, and then randomly choosing one of each pair to receive funding.

Ultimately, designing effective endangered species policy requires policymakers to develop interventions and collect data with the intention of evaluating the intervention's effectiveness. The absence of such efforts is a widespread problem in the field of biodiversity protection (Ferraro and Pattanayak 2006). Thus the current debate about the ESA should be less about dramatic changes to the ESA regulations and more about dramatic changes in the way in which the ESA is implemented and evaluated in the field.

**References**

Abadie A., D. Drukker, J.L. Herr, G.W. Imbens. 2001. "Implementing Matching Estimators for Average Treatment Effects in Stata." The Stata Journal 1:1-18. (updated 10 June 2004) http://ksghome.harvard.edu/~.aabadie.academic.ksg/software.html

Abadie, A. and G. Imbens. 2004. "Large sample properties of Matching Estimators for average treatment effects." Working paper.

Ando, A.W. 1999. "Waiting to Be Protected under the Endangered Species Act: The Political Economy of Regulatory Delay." Journal of Law and Economics 42(1): 29-60.

Brown, G.M. and J.F. Shogren. 1998. "Economics of the Endangered Species Act." Journal of Economic Perspectives 12:3-20.

Cash, D.W., A. Metrick, S. Shapiro, T. Schatzki, and M. Weitzman. 1997. "Database on the Economics and Management of Endangered Species Codebook (DEMES)." Prepared for presentation at the Social Order and Endangered Species Act Conference at the University of Wyoming.

Dalton, R. with E. Marris. (2005) Congress attacked over species bill. *Nature* **438**, 140-141.

Dawson, D., and J.F. Shogren. 2001. An Update on Priorities and Expenditures under the Endangered Species Act 77(4): 527-532.

Dehejia, R.H. and S. Wahba. 1999. "Causal Effects in Non-Experimental Studies: re-evaluating the evaluation of training programs." Journal of the American Statistical Association 94(448): 1053-1062.

Dehejia, R.H. and S. Wahba. 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies." The Review of Economics and Statistics 84(1): 151-161.

Deininger, K., H. Hoogeveen, and B. H. Kinsey. 2004. "Economic Benefits and Costs of Land Redistribution in Zimbabwe in the Early 1980s." World Development 32 (10): 1697-1709

Diprete, T.A. and M. Gangl. 1994. "Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments." Sociological Methodology 34 (forthcoming).

Doremus, H. 1997. "Listing Decisions under the Endangered Species Act: Why better science isn't always better policy." Washington University Law Quarterly 75.

Doremus, H. and J. E. Pagel. 2001. "Why Listing may be forever: Perspectives on Delisting under the U.S. Endangered Species Act." Conservation Biology 15: 1258–1268.

Duflo, E. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia:  Evidence from an Unusual Policy Experiment." The American Economic Review 91(4): 795-813.

Frederiksson, P.G., and D. Millimet. 2004. "Comparative Politics and Environmental Taxation." Journal of Environmental Economics and Management 48(1):705-722.

Gordon, R.E. Jr., J.L. Lacy, and J.R Streeter. 1997. "Conservation under the Endangered Species Act." Environment International 23: 359-400.

Heckman, J., H. Ichimura, and P. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. " Review of Economic Studies 64(4): 605-54.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998a. "Characterizing Selection Bias Using Experimental Data." Econometrica 66(5): 1017-1098.

Heckman, J., H. Ichimura, and P. Todd. 1998b. "Matching As an Econometric Evaluation Estimator." Review of Economic Studies 65:261-294.

Hollingsworth, D. 1998. "Why the US Regulatory Endangered Species Model is a Disaster for Small Property Owners and Hurts Species: Analysis and Case Studies." Paper presented at The Fraser Institute conference, Protecting Endangered Species: Alternatives to Legislation. Vancouver, British Columbia.

Innes, R. 1997. "Takings, Compensation and Equal Treatment of Owners of Developed Property." Journal of Law and Economics 40(2): 403-32.

Jalan, J. and M. Ravallion. 2003. "Does piped water reduce diarrhea for children in rural India?" Journal of Econometrics 111(1): 153-173.

Kerkvliet, J.R. and C. Langpap. 2002. "Success or Failure: Measuring the Effectiveness of the Endangered Species Act." Working paper.

List, J., D.L. Millimet, W.W. McHone and P.G. Frederiksson. 2003. "Effects of Environmental Regulations on Manufacturing Plant Births: Evidence from a Propensity Score Matching Estimator." The Review of Economics and Statistics 85(4): 944-952.

Lueck, D. and J.A. Michael. 2003. “Preemptive Habitat Destruction under the Endangered Species Act.” Journal of Law & Economics 46(1): 27-60.

Male, T., and M. Bean. 2005. “Measuring Progress in US Endangered Species Conservation.” *Ecology Letters*, Vol 8.

Mann, C. and M. Plummer. 1995. “Noah's choice: the future of endangered species.” A. Knopf, New York.

Margolis, M., D.E. Osgood, and J.A. List. 2004. “Measuring the Preemption of Regulatory Takings in the US Endangered Species Act: Evidence from a Natural Experiment.” Working paper.

McIntosh, C. 2004. “The Use of Two Control Groups in Quasi-Experimental Program Evaluation.” Working paper.

Metrick, A. and M. Weitzman. 1996. “Patterns of behavior in endangered species preservation.” Land Economics 72:1-16.

National Research Council (NRC). 1995. Science and the Endangered Species Act. National Academy Press, Washington, D.C.

Polasky, S. and H. Doremus. 1998. “When the truth hurts: Endangered species policy on private land with incomplete information.” Journal of Environmental Economics and Management 35(1):22-47.

Rachlinski, J.J. 1997. “Noah by the numbers: An empirical evaluation of the Endangered Species Act.” Cornell Law Review 82: 356-89.

Rosenbaum, P. 2002. “Observational Studies”, Second edition. Springer Series in Statistics, Springer-Verlag, New York.

Rosenbaum, P. and D. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” Biometrika 70(1): 41-55.

Rosenbaum, P. and D. Rubin. 1984. “Reducing bias in observational studies using subclassification on the propensity score.” Journal of the American Statistical Association 79: 516-524.

Schwartz, M.F. 1999. “Choosing an appropriate scale of reserves for conservation.” Annual Review of Ecology and Systematics 30: 83-108.

Shogren,J.F. (ed.) 1999. “Private Property and the Endangered Species Act: Saving Habitats, Protecting Homes.” University of Texas Press. Austin, TX.

Stanford Environmental Law Society (SELS). 2001. The Endangered Species Act A Stanford Environmental Law Society Handbook. Stanford University Press, Stanford, CA.

Stokstad, E. (2005) What’s wrong with the Endangered Species Act? *Science* **309**, 2150-2152.

Stroup, R. 1995. “The Endangered Species Act: Making Innocent Species the Enemy.” PERC Policy Series, PS-3. Bozeman, MT: PERC.

Suckling, K., R. Slack, and B. Nowicki. 2004. “Extinction and the Endangered Species Act.” Center for Biological Diversity. Working paper.

Taylor, M., K. Suckling, and J. Rachlinski. 2005. “The Effectiveness of the Endangered Species Act: A Quantitative Analysis.” BioScience 55(4): 360-367.

U.S. Fish and Wildlife Service (FWS). 1992. Report to Congress: Endangered and Threatened Species Recovery Program 17-20. Washington, D.C: U.S. Department of the Interior.

U.S. Fish and Wildlife Service (FWS). 1996. Report to Congress on the Recovery Program for Threatened and Endangered Species. Washington, D.C: U.S. Department of the Interior.

Wilcove D.S., M.McMillan, K.G. Winston. 1993. “What exactly is an endangered species? An analysis of the US Endangered Species List: 1985-1990.” Conservation Biology 7: 87-93.

Yaffee, S.L. 1982. “Prohibitive Policy: Implementing the Federal Endangered Species Act”. MIT Studies in American Politics and Public Policy. The MIT Press, Boston, MA.

Zhang, D. 2004. “Endangered Species and Timber Harvesting: The Case of Red-Cockaded Woodpeckers.” Economic Inquiry 42(1):150-165.

## Table 1. 1993 Scores for Listed and Unlisted Species

| 1993 NS Score | Definition | Unlisted Species. Number | % | Listed Species. Number | % |
|---|---|---|---|---|---|
| **1** | Critically imperiled in range | 35 | 11.9 | 77 | 57.0 |
| **1.5** | | 10 | 3.4 | 0 | 0.0 |
| **2** | Imperiled in range | 63 | 21.4 | 40 | 29.6 |
| **2.5** | | 10 | 3.4 | 3 | 2.2 |
| **3** | Vulnerable in range | 162 | 54.9 | 14 | 10.4 |
| **3.5** | | 15 | 5.1 | 1 | 0.7 |
| **Total** | | 295 | 100 | 135 | 100 |

## Table 2. Probability of being listed, by Cohort

| Cohort: | # obs: | Political | Scientific | Charisma |
|---|---|---|---|---|
| 1967 | 430 | 3.62 | 29.79 | 14.65 |
| | | (0.0065) | (0.0000) | (0.0000) |
| 1968-82 | 388 | 1.93 | 18.65 | 1.43 |
| | | (0.1041) | (0.0000) | (0.2328) |
| 1983-1989 | 351 | 1.49 | 13.62 | 2.84 |
| | | (0.2042) | (0.0000) | (0.0379) |
| 1989-1993 | 318 | 0.29 | 8.17 | 0.94 |
| | | (0.8868) | (0.0004) | (0.4239) |

F-statistics (p-values in parentheses)
Political variables: LCV averages for house and senate,
and average of dummies for above 25/below 75.
Scientific variables: 1993 NS score and average # of citations.
Charisma variables: dummies for birds & mammals, log(length).

## Table 3. Treatment effect estimates.

**Average Treatment Effect on the Treated:**

| Distance metric: | Listing | Listing and High Funding (compared to unlisted) | Listing and Low Funding (compared to unlisted) |
|---|---|---|---|
| **Mahalanobis.** | -0.0189 | 0.4091*** | -0.1806** |
| | (.796) | (.000) | (.035) |
| **Exact match on Taxonomy, then Mahalanobis.** | 0.0045 | 0.5188*** | -0.1572* |
| | (.952) | (.000) | (.068) |
| **Propensity.** | -0.0639 | 0.5147** | -0.2166** |
| | (.489) | (.031) | (.034) |
| # treated (listed) species: | 135 | 34 | 101 |

P-values in parentheses, ***=99% confidence, **=95%, *=90%

## Table 4. Pearson Chi-squared Tests.

**Comparison of treated outcomes to matched counterfactuals.**

| **Treatment: Listing.** Species Outcome: | Unlisted Number | Unlisted % | Listed Number | Listed % | Total |
|---|---|---|---|---|---|
| More endangered | 34 | 25.2 | 24 | 17.8 | 58 |
| Unchanged | 85 | 63.0 | 100 | 74.1 | 185 |
| Less Endangered | 16 | 11.9 | 11 | 8.1 | 27 |
| Total | 135 | 100 | 135 | 100 | 270 |

Pearson chi(2) = 3.8663
p = 0.145

| **Treatment: High Funding.** Species Outcome: | Unlisted Number | Unlisted % | High Funding & Listed Number | High Funding & Listed % | Total |
|---|---|---|---|---|---|
| More endangered | 9 | 6.7 | 2 | 1.5 | 11 |
| Unchanged | 24 | 17.8 | 24 | 17.8 | 48 |
| Less Endangered | 1 | 0.7 | 8 | 5.9 | 9 |
| Total | 34 | 100 | 34 | 100 | 68 |

Pearson chi(2) = 9.8990
p = 0.007

## Table 5. Rosenbaum critical p-values for treatment effects.

**Test of the null of zero effect for:**

| $\Gamma$ | **Listing** | **Listing and High Funding (compared to unlisted)** | **Listing and Low Funding (compared to unlisted)** |
|---|---|---|---|
| **1** | 0.0886 | 0.0000 | 0.0001 |
| **1.5** | 0.7010 | 0.0003 | 0.0177 |
| **2** | 0.9691 | 0.0019 | 0.1465 |
| **2.5** | 0.9983 | 0.0061 | 0.4006 |
| **3** | 0.9999 | 0.0138 | 0.6551 |
| **3.5** | 1.0000 | 0.0247 | 0.8294 |
| **4** | 1.0000 | 0.0387 | 0.9243 |
| **4.5** | 1.0000 | 0.0550 | 0.9690 |
| **5** | 1.0000 | 0.0732 | 0.9880 |
| N (treated) | 135 | 34 | 101 |

Table reports p-values for Wilcoxon sign-rank test of significance under hidden bias.

**Appendix - Data**

Sample. We limit our study to endangered U.S. terrestrial and freshwater vertebrates that have full species status. We exclude listed species located outside of the 50 United States because (a) the FWS has little or no control over the protection of these species and (b) creating a counterfactual for these species is more difficult given data limitations in these foreign locations. We exclude plants because data on their endangerment status over time are not available for a large set of species. We exclude marine mammals because many are listed and managed by the National Marine Fisheries Service, not the FWS. We exclude subspecies for three reasons: (a) an exhaustive list of subspecies in the United States does not exist; (b) our biological database (NatureServe), which is the only exhaustive database of all U.S. vertebrate full species, only tracks "selected" unlisted sub-species; and (c) the concept of a "subspecies" is controversial in the biological community. In the FWS literature, full species are supposed to be afforded higher priority than subspecies. We also exclude exotic species because their protection is not a conservation objective, and we drop species that were listed after 1993 because we wish to construct a counterfactual using only species unlisted through 2004.

Outcome. The outcome variable is "change in endangerment status from 1993 to 2004." We choose to begin the period in 1993 for practical reasons. There are no available pre-1993 objective measures of endangerment status for a large set of listed and unlisted species of different taxonomic classes.

Treatments. As is common in analyses of the ESA, we do not distinguish between species listed as "endangered" (in danger of extinction throughout all of a significant portion of its range) or as "threatened" (likely to become endangered in the foreseeable future). The words "endangered" and "threatened" are not precise scientific terms with generally accepted biological meaning, and in practice, both categories are afforded the same protection under the ESA. The motivation for considering listing and funding jointly arises from the fact that many listed species get no more than their names in the Federal Register and a nominal amount of money for recovery efforts (32). For example, although almost 90% of species received funds in 2003, the median funding level of species that received funds was $20,100. Less than 2% of listed species received half of all the funds allocated, and less than 8% received 90% of the funds.

<u>Covariates</u>.

*Taxonomy and Size*: We include: taxonomic class (bird, amphibian, mammal, fish, and reptile), which captures human affinity for species that are more closely related to humans and important biological characteristics such as reproductive capacity and habitat requirements; length, which, along with taxonomic class, is a proxy for "charismatic megafauna" as well as capturing important biological characteristics such as metabolism and habitat requirements; and taxonomic distinctiveness (monotypic or from a small genus with 2-5 species), which captures the species value-added to biodiversity. These covariates plausibly have an effect on both listing and recovery (24). We obtain these measures from NatureServe's Explorer database.

*Endangerment Status.* A species level of endangerment affects both its probability of listing as well as its probability of recovering. To measure endangerment status, we use national endangerment scores from NatureServe, which tracks all native vertebrates in the United States (also used by (24)). Based on the Natural Heritage Methodology, NatureServe's system assigns an endangerment score to each species on a scale of 0 (extinct) to 5 (least endangered). The NatureServe scoring system is the most comprehensive measure of species endangerment for the set of listed and unlisted vertebrates. Each of the scores has a well-defined meaning and a serious effort is made to apply the scores consistently. A score of 1 implies that the species is “critically imperiled” in its range, having fewer than 6 occurrences in the world, or fewer than 1,000 individuals. A score of 2 implies that the species is “imperiled” in its range, having between 6 and 20 occurrences, or fewer than 3,000 individuals. A score of 3 implies the species is “vulnerable” in range, have fewer than 100 occurrences, or fewer than 10,000 individuals. When a species falls between two scores, we give it an average value (e.g., 2/3 implies 2.5). For species that are reported to still exist but lacking persuasive documentation, or species that have not been observed in some time but have the potential to still exist, we assign a score of 0.5. As noted by others (24), the NatureServe system is similar to the “degree of threat” measure in the FWS’s priority scoring system. However, unlike the NatureServe system, no specific standards have been published by the FWS to explain why different species are assigned different degrees of threat. We include only species that were “endangered” in 1993: those with scores between 1 and 3.5. We remove species with scores of 0 or 0.5 because it would be difficult for such species to show any change between 1993 and 2004. Not all vertebrates received scores in 1993 and thus this variable is the limiting variable in our data set.

One potential concern is that the listing process has been a perfect policy instrument that has listed all endangered species and thus there is no control group for very endangered, listed species. Table 4 in Supporting Information shows how the distribution of 1993 NatureServe scores differs across treatment status. Although the listed species within the sample are more likely to be endangered than are the unlisted, it is clear by inspection that there is a common support (i.e., overlap at all levels).

*Science*: The FWS claims to base listing decisions on available scientific evidence. If the science does not warrant listing (either because the science indicates the species is not imperiled or because sufficient data are lacking), the FWS will not propose a species for listing in the Federal Register. Relying on scientific information has been, to different degrees, an important part of the ESA from its birth (23). Scientific information can affect listing decisions through its direct effect on the FWS, but also indirectly through conversations that the FWS has with scientists who might be interested in seeing “their” species listed (23). Obviously scientific information will also influence the success of a species recovery: the more well understood the species, the more likely the species can be successfully recovered.

We use the annual number of journal articles as a measure of scientific influence on the listing and recovery processes. The number of such articles is not a perfect measure of scientific (and scientist) influence on the processes of listing and recovery. However, although the FWS is allowed to consult unpublished reports and first-hand observations, it tends to be reluctant to do so (23). For every species in our database, we used BIOSIS Previews to record the annual number of citations to that species from 1969 to 1993. In our analysis, we use the average annual citations to a species as a measure of scientific influence.

*Politics*: Interference in the listing and recovery process by federal legislators is commonly assumed, although with the exception of one published article on the listing process (33), data for such interference are lacking. Pro-environment politicians may be more active in seeking, or less active in preventing, the listing of species in their states. Pro-environment politicians may also reflect pro-environmental preferences of their constituents which may make the FWS more inclined to list a species from the state (because there will be less resistance) or less inclined to list the species (because the citizens and politicians have or will take action themselves and the species is less likely to need federal protection). Opposite effects would stem from pressure by pro-land-use politicians and their constituents. Clearly, political influence can affect not only listing, but species recovery as well.

To measure the environmental preferences of federal legislators, we follow Ando (33) and use League of Conservation Voter (LCV) scores for every House and Senate delegation back to 1971, the first year the League published their scorecard. These data were derived from on-line and hard copy content from the League. We construct two measures of environmental preferences. (A) For both the Senate and House delegations, we estimate the average annual LCV score between 1971 and 1993. When a species is found in more than one state, we take an average of the annual scores across states (we do not have precise data on the proportions of a species' habitat in each state, and thus weigh each state equally). (B) Every Senate delegation has two members, but the number of members of each House delegation varies by the size of the state. Presumably the total numbers of pro-environment and pro-land-use House representatives can matter in listing and recovery outcomes. Using Ando's score cutoffs for designating a representative as "pro-environment" (score>75) or "pro-land-use" (score<25), we estimate the average annual number of pro-environment and pro-land-use congressional representatives that have influence over a given species' habitat.

We also collected data on citizen environmental preferences (proxied by the number of citizens of each state that recreationally observe wildlife in a non-hunting context according to the FWS), on Federal and State land ownership by state (from http://www.nwi.org/Maps/LandChart.html), and on the comprehensive of state Endangered Species laws (from http://www.defenders.org/pubs/ sesa01.html). However, we find that these variables were not important in any of our selection models and we exclude them in the analysis.

*Funding*: As with the listing decision, funding decisions are nonrandom. For example, species with the highest spending include many "charismatic" species, such as the red-cockaded woodpecker and the bald eagle. We use FWS annual accounting of "reasonably identifiable"

ESA expenditures compiled in (29). Our analysis, which treats funding as a binary treatment variable, does not require the absolute amount of money reported by the FWS to be accurate (there may indeed be expenses that are not “reasonably identifiable”). It requires only that the rank order by species is accurate (i.e., there is a strong positive correlation between the identified expenditures in the FWS reports and the actual expenditures).

*Short-term Trend*: Changes in the NatureServe scores between 1993 and 2004 may reflect declines or improvements in a species’ population, but they also may reflect new information (i.e., new populations of a species are discovered and thus the species status improves; one species is separated into two species and thus the status of the species maintaining the original name declines). NatureServe’s “short-term trend” variable explicitly avoids attributing changes in status to new information and attempts to capture “the observed, estimated, inferred, suspected, or projected short-term trend in population size, extent of occurrence, area of occupancy, number of occurrences (EOs), and/or viability/ecological integrity of occurrences (whichever most significantly affects the Heritage Conservation Status Rank) within the specified geographic level” (quoted from the metadata file accompanying the data file from NatureServe). This variable consists of four categories of declining trend (>70%, 50-70%, 30-50%, 10-30%), one category for stable (unchanged or within a 10% fluctuation), one category for increasing trend (>10%), and one category for unknown. We code declining as -1, stable as 0, and improving as 1 (we exclude species with unknown trend).