Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Environmental Economics and Management 54 (2007) 245–261
JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT
www.elsevier.com/locate/jeem

The effectiveness of the US endangered species act: An econometric analysis using matching methods

Paul J. Ferraro[a,*], Craig McIntosh[b], Monica Ospina[a]

[a]Department of Economics, Andrew Young School of Policy Studies, Georgia State University, P.O. Box 3992, Atlanta, GA 30302-3992, USA
[b]School of International Relations and Pacific Studies, University of California-San Diego, San Diego, CA, USA

Received 5 May 2006
Available online 28 June 2007

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 matching 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.

Keywords: Endangered Species Act; Biodiversity; Program evaluation; Matching methods

1. Introduction

The Endangered Species Act (ESA) is the most important piece of biodiversity legislation in the United States, but its effectiveness is hotly debated [6,34,37]. 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 30 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 non-randomized nature of the ESA makes measuring its effect on species recovery difficult. As we will demonstrate, recent efforts have relied on a set of 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.

*Corresponding author. Fax: +1 404 413 0205.
E-mail address: pferraro@gsu.edu (P.J. Ferraro).

0095-0696/$ - see front matter
doi:10.1016/j.jeem.2007.01.002

As a starting point for our analysis, we study the ESA's species listing process. As in Metrick and Weitzman [24], 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, both of which use duration of listing to identify treatment effects [20,40]. In contrast, our use of matching estimators allows us to select counterfactual species that share charismatic, political and scientific characteristics with listed species. These methods offer a flexible way to control for the complex and shifting set of characteristics that determine selection.

Using different matching estimators, we compare species recovery between listed and unlisted species. 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 providing substantial flexibility in the choice of specification. 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 [4,16,27], we find that species that are listed with little or no funding experience worse outcomes than the comparison group.

Our estimators are non-experimental, and so will contain bias if the listing or funding processes are 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 negative effect of listing without funding is robust to the presence of unobserved heterogeneity, while the positive effect of listing with 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. This implies that using scarce conservation funding in the contentious process of listing a species may be less effective than using this funding to promote recovery directly. 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

We attempt to answer the question, "Are listed terrestrial and freshwater vertebrates better off than if they had not been listed?" Working against the ESA [4,43] are the Act'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. As evidence against the Act's effectiveness, critics cite the paucity of delisted species and of recovering species as defined by the US Fish and Wildlife Service (FWS) [13,21]. Furthermore, there is anecdotal [14,21,38], theoretical [16,27], and empirical [19,22,45] evidence that the Act encourages landowners to preemptively harm species and their habitat.

Working in favor of the ESA are the Act's strong regulatory powers (particularly Sections 7 and 9) and the recovery funds allocated by Congress annually. Evidence assembled in favor of the Act's effectiveness include arguments that the ESA has prevented extinctions [26,32,39], that the more years a species has been listed, the more likely the species would be declared to be recovering by the FWS [20,28,40,41], and that there is a positive correlation between government funding and reported FWS status [20,25].

All of the published evidence marshaled to date for and against the ESA has a common problem: the absence of a well-chosen counterfactual. For example, consider one of the more sophisticated analyses to date. Taylor et al. [40] 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 [26] and, given they are not constructed using transparent criteria, may be manipulated to achieve agency objectives.[1]

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. Similarly, Male and Bean [20] identify impacts of the ESA off of 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.[2]

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 (not 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 [5] and 2004 scores from NatureServe.

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 exclude species located outside of the 50 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 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 US 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. 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. In particular, we are interested in the average treatment effect on the treated (ATT), which is what a listed species' change in status would have been had it not been listed.[3] If listing is allocated randomly across species, we can estimate the counterfactual

[1]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.

[2]One unpublished study uses unlisted species to construct the counterfactual [17]. However, the paper has several limitations: the authors estimate changes over only a 4-year period, they include the outcome at the *end* of the period as an explanatory variable, and they compare listed species with *all* unlisted species, including species that are not endangered. These problems, combined with their small number of covariates, the dramatic differences on observable characteristics that affect outcomes between listed and unlisted species, and the highly parametric estimator (ordered probit), likely create bias.

[3]Because the intent of the ESA is precisely to list only those species that are endangered, the Average Treatment Effect, which would be the impact of listing the average species, is not of interest.

Table 1
1993 scores for listed and unlisted species

1993 NS score	Definition	Unlisted species		Listed species	
		Number	%	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

simply by using the status of unlisted species because 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.[4] 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 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 [30,31]. Matching mimics random assignment through the *ex post* construction of a control group. 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, the non-treated outcomes are what the treated outcomes would have been had they not been treated. This ‘‘conditional independence 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.

In our analysis (Section 4.2), we use a variety of covariate and propensity score matching estimators. We assess the robustness of our results through ‘‘quality control’’ measures, including imposing common support, using calipers and forcing exact matching on important covariates. 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 knowledge and interest 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 scores from the League of Conservation Voters, or LCV). Further details on these covariates and their sources, as well as summary statistics and other covariates considered, can be found in Appendix A.

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]For more complete discussions of the ESA and its history, see [34,36,44].

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. In the late 1970s, many believed that listing decisions were often driven more by politics and preferences than by science [9]. In 1979, the ESA was amended to include a requirement that the FWS perform a formal review, which includes communication with ''experts in the field,'' to determine whether sufficient ''scientific and biological data'' exist to justify a listing proposal. 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 (biological) 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 Refs. [4,9]). It is an empirical question, however, whether this amendment removed the influence of preferences and politics given FWS employee preferences for mammals and birds[5] and given Congress's control over the FWS budget.

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 are designated by the years of listing: 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.

For the cohort of species that were listed under the ESA during each of the four periods, the relevant counterfactual is formed by similar species that were *not* listed during that period. The control species for the first cohort listed therefore consist of all other species. As we move through successive cohorts, we remove from the analysis those species that have already been listed. We explain the binary outcome listed/not listed using a linear probability model with heteroskedasticity-robust standard errors.

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 NatureServe 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 Appendix A 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 1983–1989 cohort, also become insignificant in the final cohort. We thus conclude that the FWS guidelines were successful in reducing the influence of non-scientific criteria on the listing process.[6]

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

[5] Metrick and Weitzman [24] 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. Dawson and Shogren [7] argue that this correlation could result from time invariant characteristics such as a well-developed scientific foundation or historical game use associated with these taxonomic groups. However, a survey of employee preferences in the FWS Office of Endangered Species also found that employees ranked mammals and birds above fish, amphibians and reptiles [4, p. 8].

[6] These changes in the selection process provide a shift in the rule that assigns species to the treatment and the control, and so it is tempting to try to use them as instruments. These changes, however, are likely to be endogenous both to the species up for consideration at a moment in time and to broader changes in environmental preferences, thus violating the exclusion restriction.

Table 2
Probability of being listed, by cohort

Cohort	# Obs.	Political	Scientific	Charisma
1967	430	362 (0.0065)	15.46 (0.0000)	14.65 (0.0000)
1968–1982	388	1.93 (0.1041)	9.92 (0.0000)	1.43 (0.2328)
1983–1989	351	1.49 (0.2042)	6.95 (0.0000)	2.84 (0.0379)
1989–1993	318	0.29 (0.8868)	4.28 (0.0022)	0.94 (0.4239)

F-statistics (*P*-values in parentheses).
Political variables: average LCV scores for House and Senate delegations; average number of pro-environment and pro-land use representatives (see Appendix A for details).
Scientific variables: 1993 NatureServe score; average # of citations; dummies for small genus and monotypic species.
Charisma variables: log of species length; dummies for birds and mammals.

construction of a counterfactual in the evaluation of the ESA must take into account selection on these observable characteristics. Matching methods control for selection on observables and allow one to easily confirm that the matched species share 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 quartile 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 4 years in which the FWS was required to conduct an annual accounting of ‘‘reasonably identifiable’’ expenditures associated with ESA (Appendix A).[7]

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 four matching estimators (based on work by [1,11,30]): (1) nearest-neighbor covariate matching estimator with an inverse variance weighting matrix to account for the difference in scale of the covariates; (2) nearest-neighbor covariate matching estimator with Mahalanobis weighting; (3) nearest-neighbor propensity score matching estimator; and (4) kernel (Gaussian) propensity score matching estimator.[8] The nearest-neighbor matching is with replacement and we resolve the mean-variance tradeoff in the match quality by using four nearest neighbors; the counterfactual outcome is the average among these four.[9] All matching is done across the full vector of control characteristics.

Based on recent work that demonstrates that bootstrapping standard errors is invalid with non-smooth nearest-neighbor estimators [1], we use Abadie and Imbens’ variance formula [1] for our nearest-neighbor estimators. We use the robust version of the formula to allow for heteroskedasticity (using four neighbors in the second-stage matching), which allows the treatment effect to be non-constant (i.e., outcome variance differs by treatment status and covariates). For the kernel-matching estimator, we bootstrap the standard errors (using 999 replications).

In our covariate-matching estimators, we use a post-matching bias-correction procedure that asymptotically removes the conditional bias term in finite samples [2]. For all propensity score estimators, we enforce a common support. Balancing tests were also conducted for the propensity score estimators.[10] In the first two treatments (listing, high funding), balance was achieved on all covariates. In the third treatment (low funding),

[7]The expenditure data clearly separate species into a high funding and a low-funding group, and so impact estimates are not sensitive to local changes in the dollar value or in the percentile used to define the cutoff.

[8]With the exception of the kernel matching (Stata v.9; [18]), matching was done in Ref. [33].

[9]This is a standard approach in nearest-neighbor matching [1,23]. We varied the number of neighbors from one to fifteen and the ATT estimate changes very little. Results available through JEEM’s online archive of supplementary material, which can be accessed at http://www.aere.org/journal/index.html.

[10]Available through JEEM’s online archive (see footnote 7).

Table 3
Treatment effect estimates

	Average treatment effect on the treated		
	Listing	Listing and high funding (compared to unlisted)	Listing and low funding (compared to unlisted)
Nearest-neighbor covariate (inverse variance)	−0.0191 (0.839)	0.4537*** (0.001)	−0.2128** (0.027)
Nearest-neighbor covariate (Mahalanobis)	−0.0189 (0.823)	0.4091*** (0.001)	−0.1806** (0.047)
Nearest-neighbor propensity score	−0.1074 (0.161)	0.3355 (0.130)	−0.2307*** (0.004)
Kernel (Gaussian) propensity score	−0.1419* (0.100)	0.4295** (0.050)	−0.2188** (0.050)
# Observations	430	329	396
# Listed species	135	34	101
# Species off common support	18	15	5

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

balance was achieved on all covariates except that there is a slightly greater percentage of amphibians among the unlisted species (6.7% difference; $p = 0.09$).

Table 3 presents the treatment effect estimates. The estimated treatment effect of listing alone is small, negative and not statistically different from zero in three of the four specifications (weakly significant using kernel matching). Among listed species, however, we see sharply differentiated effects across funding status. Listing a species with high funding increases its NatureServe score by almost a full half point, and the effect is significantly different from zero in most specifications (not significant using nearest-neighbor propensity score matching).

As a robustness check on the funding result, we re-do the analysis using the Mahalanobis estimator, reinterpreting the set of species receiving ‘‘substantial funding’’ as the top third of fund recipients. This cohort receives 97% of all funds allocated in our sample. The effect of listing, combined with this notion of substantial funding, remains positive (0.2690) and significantly different from zero ($p = 0.006$). In contrast, species that are listed with little or no funding show a decline that, while roughly half as strong in magnitude as the improvement in well-funded species, is strongly significant.[11]

5. Robustness checks

5.1. Additional constraints on selecting the counterfactual

Table 4 presents treatment effect estimates using covariate and propensity score matching in combination with additional constraints placed on the selection of counterfactuals. As an additional form of quality control, we implement caliper matching in the context of our bias-adjusted, nearest-neighbor Mahalanobis matching estimator [35]. In the first two rows of Table 4, we use calipers of 3.5 and 3 standard deviations, meaning that any control unit outside the range of this caliper in the space of the distance metric is dropped

[11]A natural baseline regression would be to use ordinary least squares regression. However, OLS is problematic because it conflates the selection and outcome equations, and thus assumes that the determinants and coefficients are the same for these two processes. OLS 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 mis-specified and thus inconsistent. OLS also uses control units for estimating the counterfactual regardless of whether they are on a common support with treated units (OLS also has difficulty with extreme imbalances in the covariate densities on the common support). 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 without funding has a significant negative effect, and listing with funding is positively significant only at the 90% level.

Table 4
Treatment effect estimates with additional constraints on counterfactual selection

	Listing	Listing and high funding (compared to unlisted)	Listing and low funding (compared to unlisted)
Nearest-neighbor covariate, caliper 3.5[a]	−0.0752	0.3373***	−0.1763**
	(0.339)	(0.002)	(0.052)
Nearest-neighbor covariate, caliper 3	−0.0772	0.2530***	−0.1896**
	(0.319)	(0.010)	(0.034)
Nearest-neighbor covariate, exact taxonomy	0.0045	0.5188***	−0.1573*
	(0.960)	(0.000)	(0.097)
Nearest-neighbor covariate, exact 1993 score	−0.0299	0.3087***	−0.1839**
	(0.728)	(0.004)	(0.048)
Nearest-neighbor covariate, exact 1993 score and taxonomy	−0.0222	0.5265***	−0.2667***
	(0.799)	(0.000)	(0.002)
Nearest-neighbor propensity score, exact taxonomy	−0.0671	0.1645	−0.1849*
	(0.479)	(0.384)	(0.060)
Nearest-neighbor propensity score, exact 1993 score	−0.0833	0.3355*	−0.2500***
	(0.282)	(0.090)	(0.001)
Nearest-neighbor propensity score, exact 1993 score and taxonomy	−0.1079	0.1250	−0.2787***
	(0.140)	(0.458)	(0.001)

P-values in parentheses, *** = 99% confidence, ** = 95%, * = 90%.
[a]Number of treated units dropped because no controls were found in the caliper: *Caliper* 3.5: 4 (listing), 3 (high funding), and 1 (low funding); *Caliper* 3.0: 10 (listing), 6 (high funding), and 9 (low funding).

when the counterfactual mean is calculated.[12] Listing alone continues to have negative but insignificant treatment effects. Listing combined with funding continues to have positive treatment effects; they are lower but still significantly different from zero. Listing with little or no funding continues to have negative and significant treatment effects.

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 using both covariate matching (Mahalanobis) and propensity score matching (nearest-neighbor). We also restrict the match to be exact on the 1993 NatureServe baseline score, and then on both the 1993 score and taxonomy.[13] None of our qualitative conclusions change.

To address the concern that our control species may simply not be as endangered as listed species despite having, on average, the same 1993 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 unlisted species.

Constructing our counterfactual from this reduced set does not change our conclusions. Treatment effect estimates are presented in Table 5. Most estimators indicate that listing has zero or a weakly negative effect on recovery, listing with funding has a substantial and significant positive effect on recovery (not significant using nearest-neighbor propensity score matching), and listing without funding has a substantial and significant negative effect.

[12]Reducing the caliper size below 3 standard deviations drops too many observations in our high-funding treatment to be useful. Note, however, that reducing the caliper successively to 2.5, 2, 1.5 and 1 increases our treatment effect estimates in absolute value in all treatments (and they are significantly different from zero). Thus, if anything, smaller calipers strengthen our qualitative conclusions.

[13]For some treated species, an exact match did not exist (particularly when trying to match on both taxonomy and the 1993 score). In Table 4, we present results from the analysis in which treated units without an exact match are dropped. Keeping them in the analysis does not change our qualitative conclusions.

Table 5
Treatment effect estimates with controls restricted to candidate species

	Listing	Listing and high funding (compared to unlisted)	Listing and low funding (compared to unlisted)
Nearest-neighbor covariate (inverse variance)	−0.0664	1.0043***	−0.2314***
	(0.471)	(0.011)	(0.005)
Nearest-neighbor covariate (mahalanobis)	−0.0852	0.4686***	−0.2652***
	(0.314)	(0.005)	(0.004)
Nearest-neighbor propensity score	−0.1466**	0.3214	−0.1399**
	(0.042)	(0.205)	(0.067)
Kernel (Gaussian) propensity score	−0.1602**	0.4157***	−0.1815**
	(0.050)	(0.001)	(0.050)
# Observations	278	177	244
# Listed species	135	34	101
# Species off common support	38	27	24

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

5.2. *Alternative outcome variable*

One concern with the analysis is the accuracy of our outcome variable: 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 A). 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 artifacts of the outcome variable we use.

Treatment effect estimates using this outcome variables are presented in Table 6. 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. Our estimate of the effect of listing alone becomes more negative and significantly different from zero in all estimators. Our estimate of the effect of listing without funding remains substantially negative and significantly different from zero in all estimators. Unlike the estimated treatment

Table 6
Treatment effect estimates with alternative outcome variable (short-term trend)

	Listing	Listing and high funding (compared to unlisted)	Listing and low funding (compared to unlisted)
Nearest-neighbor covariate (inverse variance)	−0.1855**	−0.1734	−0.2302**
	(0.042)	(0.417)	(0.019)
Nearest-neighbor covariate (mahalanobis)	−0.1569*	0.0435	−0.2360**
	(0.088)	(0.796)	(0.017)
Nearest-neighbor propensity score	−0.3723***	−0.4375*	−0.2932**
	(0.004)	(0.094)	(0.023)
Kernel (Gaussian) propensity score	−0.3033**	−0.3561	−0.2531*
	(0.050)	(0.250)	(0.100)
# Observations	269	193	241
# Listed species	104	28	76
# Species off common support	15	12	7

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

effects of listing and listing without funding, however, our estimate of the effect of high funding is unstable and varies greatly by the matching method and the number of neighbors used.

Further constraints on the selection of the counterfactuals (calipers, exact matching) yields the same qualitative conclusions.[14] 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.

5.3. Sensitivity tests: Rosenbaum bounds

Given the natural suspicion that some degree of selection bias might remain even after careful matching, we use Rosenbaum bounds to determine how strongly an unmeasured confounding variable must affect selection into the treatments in order to undermine our conclusions [8,29]. Although there are other sensitivity tests available (e.g. [15]), we use Rosenbaum bounds because they are relatively free of parametric assumptions and because the test provides a single, easily interpretable measure of the way in which unobservables enter.

If the probability of agent j selecting into the treatment is π_j, the odds are then $\pi_j/(1-\pi_j)$. The log odds can be modeled as a generalized function of a vector of controls x_j and a linear unobserved term, so $\log(\pi_j/(1-\pi_j)) = \kappa(x_j) + \gamma u_j$, where u_j is an unobserved covariate scaled so that $0 \leqslant u_j \leqslant 1$. 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. 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: $\pi_j(1-\pi_k)/\pi_k(1-\pi_j) = \exp\{\gamma(u_j - u_k)\}$. Because of the bounds on u_j, a given value of γ 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 Γ imply an increasingly important role for unobservables in the selection decision. The differences in outcomes between the treatment and control are calculated and ranked. We contrast outcomes using matched species from the kernel propensity score-matching estimator from Table 3. 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 [12].

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

Table 7 presents the results from the Rosenbaum bounds analysis. Because listing has an insignificant impact even under the null of no unobserved bias, we perform robustness checks only on listing with and without funding. The first column indicates that, using the sign rank test, the estimated positive treatment effect of being listed with substantial funding is robust to only moderate levels of unobserved heterogeneity. If an unobserved covariate caused the odds ratio of listing to differ between listed and unlisted cases by a factor of 2.5, the 90% confidence interval would include zero. The second column indicates that the estimated negative treatment effect of being listed without funding remains significantly negative 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.

We conclude that the negative estimated effects of listing without funding are robust to the presence of unobserved bias, though the positive effects of listing with 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

[14] Available through JEEM's online archive (see footnote 7).

Table 7
Rosenbaum critical P-values for treatment effects

Γ	Test of the null of zero effect for	
	Listing and high funding (compared to unlisted)	Listing and low funding (compared to unlisted)
1	0.0050	0.0000
1.5	0.0323	0.0000
2	0.0838	0.0000
2.5	0.1498	0.0002
3	0.2216	0.0016
3.5	0.2935	0.0063
4	0.3624	0.0170
4.5	0.4266	0.0361
5	0.4856	0.0649

Table reports P-value for Wilcoxon sign-rank test of significance under hidden bias.

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 [9] 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 to 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 formed by considering ''degree of threat,'' ''recovery potential,'' ''taxonomy,'' and ''conflict with development'' [4]. Given recovery potential is included in this metric, we should be concerned that some bias may exist across this selection criterion.

However, others have shown that funding is not correlated with recovery potential [25]. Moreover, one of the strongest determinants of funding decisions is whether the species is deemed by the FWS to be in conflict with human activities [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 combination 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? One 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).[15] The potential penalties for pre-emptive actions 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 the credible potential of enforcement that renders the ESA effective.

Given the significance of expenditures for species recovery, identifying the channels through which spending achieves its impact becomes a central policy question. Unfortunately the expenditure data are aggregated by agency and not by use, and only in 1 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 an ordered probit model (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.[16] 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 the act of listing itself that matters, but rather high levels of expenditures for recovery combined with listing. Simply listing a species in the absence of such expenditures appears to lead 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. 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 answered by pairing species *ex ante* based on 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 [10]. 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.

Acknowledgments

The authors thank Andrew Metrick for providing us with the DEMES database, and Jason Delaney and Harini Kannan for assistance with data assembly. We also thank two anonymous referees, Timothy Male, Chuck Mason, Daniel Phaneuf, Bruce Stein, Kiernan Suckling, members of the Atlanta Audubon Society, and participants at the 2004 CAMP Resources Workshop, the 2005 Association of Environmental and Resource Economists Workshop, the 2005 Occasional California Workshop at UCSB's Bren School, and the 2006 Triangle Resource and Environmental Economics Seminar Series for providing useful comments.

[15] In the early 1990s, nearly 80% of all listed species occurred partially or entirely on private lands [42]. However, this explanation seems unlikely in the case of fish, which make up a substantial portion of our sample.

[16] 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.

Appendix A

A.1. 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.

A.1.1. 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).

A.2. 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.

A.2.1. Funding

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. In most years since 1989, fewer than 10% of the listed species received 90% of the available funds. As with the listing decision, funding decisions are non-random. 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 Cash et al. [5] for the years 1989–1993. The top 25% of funding recipients in our sample received, on average, \$10,950,640 from 1989 to 1993. The bottom 75% of listed species (our "low funding" treatment) received \$195,394, on average. The top 33% of funding recipients received, on average, \$8,290,187, whereas the bottom 67% received \$120,313. The average unlisted (control) species received only \$293 (only four of the 295 species received funds; an average of \$21,638).

Our analysis, which treats funding as a binary treatment variable, does not require the absolute amount of money reported by the FWS to be precise (there may indeed be expenses that are not "reasonably identifiable"). It requires only that the rank order by species is accurate and there is no species-specific bias in reporting absolute expenditures. We observe the sum of federal and state expenditures from a relatively rich variety of sources over a 5-year period, and use this total to define our "high-funding" treatment. Given the assumption of homogeneous treatment effects, the marginal impact of funding can properly be thought of as the impact of moving from the mean of the low-funding treatment (\$195,394) to the mean of the high-funding treatment (\$10,950,640), and the magnitude of this impact is the difference in differences of these two categories relative to the counterfactual; e.g., $0.41-(-0.18)=0.59$, for the nearest-neighbor Mahalanobis covariate estimator. Alternatively, the marginal effect of listing a species and increasing expenditures from the mean expenditure among unlisted species ($293) to the mean expenditure for the top funded species is 0.41. Because we do not observe funding data for every year during our sample and nor do we observe the universe of possible funding sources, total expenditures on species recovery are higher than what we observe. Hence our marginal effects should not be thought of as the impact of shifting total expenditures over the entire interval of the study from all sources, but rather the impact of shifting the federal and state expenditures over a 5-year period which lies at the beginning of the study period. To the extent that non-governmental expenditures (and government expenditures over time) co-move with governmental expenditures in our sample, these marginal effects are correct. Because private and foundation expenditures on recovery are not variables under the direct control of policymakers, we believe that the impact of government expenditure on recovery probabilities is the most policy-relevant.

A.3. Covariates

For summary statistics of the covariates, see Table A1.

A.3.1. 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 affect listing and recovery [24]. We obtain these measures from NatureServe’s Explorer database.

A.4. Endangerment status

A species’ level of endangerment affects 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 Metrick and Weitzman [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

Table A1
Sample summary statistics for listed and unlisted species

Variable	Listed species		Unlisted species	
	Mean (SD)	(Min, Max)	Mean (SD)	(Min, Max)
Mammal	0.08 (0.27)	0, 1	0.09 (0.29)	0, 1
Bird	0.27 (0.45)	0, 1	0.09 (0.28)	0, 1
Reptile	0.10 (0.30)	0, 1	0.09 (0.28)	0, 1
Amphibian	0.06 (0.24)	0, 1	0.18 (0.38)	0, 1
Fish	0.49 (0.50)	0, 1	0.55 (0.50)	0, 1
Small genus	0.22 (0.42)	0, 1	0.14 (0.35)	0, 1
Monotype genus	0.11 (0.32)	0, 1	0.06 (0.23)	0, 1
Length	2.76 (1.09)	1.1, 6.13	2.55 (0.82)	0.69, 5.06
1993 Score	1.56 (0.71)	1, 3.5	2.51 (0.74)	1, 3.5
Pro-env house	7.96 (18.20)	0.04, 111.2	6.38 (9.29)	0, 58.37
Pro-land house	11.29 (18.85)	0, 114.08	9.94 (10.06)	0, 86.96
House LCV score	42.44 (13.80)	22.96, 67.42	39.85 (12.37)	12.62, 80.83
Senate LCV score	40.76 (14.87)	14.37, 71.42	41.07 (15.82)	14.37, 90.12
Citations	0.69 (1.65)	0, 12.92	0.21 (0.42)	0, 4.75

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 1000 individuals. A score of 2 implies that the species is ‘‘imperiled’’ in its range, having between 6 and 20 occurrences, or fewer than 3000 individuals. A score of 3 implies the species is ‘‘vulnerable’’ in range, have fewer than 100 occurrences, or fewer than 10,000 individuals. A score of 4 implies the species is ‘‘apparently secure’’ throughout range (but possibly rare in parts of its range). A score of 5 implies the species is ‘‘demonstrably secure’’ throughout range (however, it may be rare in certain areas). 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 exist but lack persuasive documentation, or species that have not been observed in some time but have the potential to exist, we assign a score of 0.5. As noted by Metrick and Weitzman [24], the NatureServe system is similar to the ‘‘degree of threat’’ measure in the FWS's priority scoring system. However, unlike NatureServe, the FWS has not published specific standards 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.

A.4.1. 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 inception [9]. 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 [9]. 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 [9]. 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.

A.4.2. 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 [3], 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 [3] 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 (*House LCV score*; *Senate LCV score*). 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 (*Pro-Env House*, *Pro-Land House*) can matter in listing and recovery outcomes. Using

Ando's [3] 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 FWS surveys) on Federal and State land ownership by state (from http://www.nrcm.org/documents/publiclandownership.pdf), and on the comprehensiveness of state Endangered Species laws (from http://www.defenders.org/pubs/sesa01.html). However, we find that these variables are not important in any of our selection models and we exclude them in the analysis (citizen preferences are only important when variables for politician preferences are excluded).

References

[1] A. Abadie, G. Imbens, Large sample properties of matching estimators for average treatment effects, Econometrica 74 (2006) 235–267.
[2] A. Abadie, G. Imbens, On the failure of the bootstrap for matching estimators, Working Paper, Harvard University, 2006b.
[3] A.W. Ando, Waiting to be protected under the Endangered Species Act: the political economy of regulatory delay, J. Law Econ. 42 (1999) 29–60.
[4] G.M. Brown, J.F. Shogren, Economics of the Endangered Species Act, J. Econ. Perspect. 12 (1998) 3–20.
[5] D.W. Cash, A. Metrick, S. Shapiro, T. Schatzki, M. Weitzman, Database on the economics and management of endangered species codebook (DEMES). Presented at the Social Order and Endangered Species Act Conference, University of Wyoming, 1997.
[6] R. Dalton, Congress attacked over species bill, Nature 438 (2005) 140–141.
[7] D. Dawson, J.F. Shogren, An update on priorities and expenditures under the Endangered Species Act, Land Econ. 77 (2001) 527–532.
[8] T.A. Diprete, M. Gangl, Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments, Sociol. Methodol. 34 (2004) 271–310.
[9] H. Doremus, Listing decisions under the Endangered Species Act: why better science isn't always better policy, Wash. Univ. Law Quart. 75 (1997) 1029–1153.
[10] P.J. Ferraro, S.K. Pattanayak, Money for nothing? A call for empirical evaluation of biodiversity conservation investments, PLoS Biol. 4 (2006) e105.
[11] M. Frolich, Finite-sample properties of propensity score matching and weighting estimators, Rev. Econ. Statist. 86 (2004) 77–90.
[12] M. Gangl, RBOUNDS: Stata module to perform Rosenbaum sensitivity analysis for average treatment effects on the treated, Statistical Software Components S438301, Boston College Department of Economics, 2004.
[13] R.E. Gordon, J.K. Lacy, J.R. Streeter, Conservation under the Endangered Species Act, Environ. Int. 23 (1997) 359–419.
[14] D. Hollingsworth, Why the US regulatory endangered species model is a disaster for small property owners and hurts species: analysis and case studies, Presented at The Fraser Institute Conference, Protecting Endangered Species: Alternatives to Legislation, Vancouver, British Columbia, 1998.
[15] A. Ichino, F. Mealli, T. Nannicini, From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? IZA Discussion Paper, 2006
[16] R. Innes, Takings, compensation, and equal treatment for owners of developed and undeveloped property, J. Law Econ. 40 (1997) 403–432.
[17] C. Langpap, J.R. Kerkvliet, Success or failure: measuring the effectiveness of the Endangered Species Act, Working Paper, 2002.
[18] E. Leuven, B. Sianesi, PSMATCH2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing, Statistical Software Components S432001, Boston College Department of Economics, 2003 (revised 28 December 2006).
[19] D. Lueck, J.A. Michael, Preemptive habitat destruction under the Endangered Species Act, J. Law Econ. 46 (2003) 27–60.
[20] T.D. Male, M.J. Bean, Measuring progress in US endangered species conservation, Ecol. Lett. 8 (2005) 986–992.
[21] C.C. Mann, M.L. Plummer, Noah's Choice: The Future of Endangered Species, Knopf, New York, 1995.
[22] M. Margolis, D.E. Osgood, J.A. List, Measuring the preemption of regulatory takings in the US Endangered Species Act: evidence from a natural experiment, Working paper, 2005.
[23] C. McIntosh, Estimating treatment effects from spatial policy experiments: an application to Ugandan microfinance, Rev. Econ. Statist. (2007), http://irps.ucsd.edu/assets/003/5272.pdf.
[24] A. Metrick, M. Weitzman, Patterns of behavior in endangered species preservation, Land Econ. 72 (1996) 1–16.
[25] J.K. Miller, J.M. Scott, C.R. Miller, L.P. Waits, The Endangered Species Act: dollars and sense?, BioScience 52 (2002) 163–168.
[26] National Research Council (US), Science and the Endangered Species Act, National Academy Press, Washington, DC, 1995.
[27] S. Polasky, H. Doremus, When the truth hurts: endangered species policy on private land with incomplete information, J. Environ. Econ. Manage. 35 (1998) 22–47.
[28] J.J. Rachlinski, Noah by the numbers: an empirical evaluation of the Endangered Species Act, Cornell L. Rev. 82 (1997) 356–389.
[29] P.R. Rosenbaum, Observational Studies, Springer, New York, 2002.

[30] P.R. Rosenbaum, D.B. Rubin, The central role of the propensity score in observational studies for causal effects, Biometrika 70 (1983) 41–55.
[31] P.R. Rosenbaum, D.B. Rubin, Reducing bias in observational studies using subclassification on the propensity score, J. Amer. Statist. Assoc. 79 (1984) 516–524.
[32] M.W. Schwartz, Choosing the appropriate scale of reserves for conservation, Annu. Rev. Ecol. Syst. 30 (1999) 83–108.
[33] J. Sekhon, Matching: Algorithms and software for multivariate and propensity score matching with balance optimization via genetic search, Working Paper and Software, 2006.
[34] J.F. Shogren, Private property and the Endangered Species Act: saving habitats, protecting homes, University of Texas Press, Austin, TX, 1998.
[35] J. Smith, P. Todd, Does matching overcome LaLonde's critique of nonexperimental estimators?, J. Econometrics 125 (2005) 305–353.
[36] Stanford Environmental Law Society, The Endangered Species Act (Stanford Environmental Law Society Handbook), Stanford University Press, Stanford, CA, 2001.
[37] E. Stokstad, What's wrong with the Endangered Species Act?, Science 309 (2005) 2150–2152.
[38] R. Stroup, The Endangered Species Act: making innocent species the enemy, PERC Policy Series, 1995.
[39] K. Suckling, R. Slack, B. Nowicki, Extinction and the Endangered Species Act, Center for Biological Diversity Working Paper, 2004.
[40] M. Taylor, K. Suckling, J. Rachlinski, The effectiveness of the Endangered Species Act: a quantitative analysis, BioScience 55 (2005) 360–367.
[41] US Fish and Wildlife Service (FWS), Report to Congress: Endangered and Threatened Species Recovery Program, US Department of the Interior, Washington, DC, 1992, pp. 17–20.
[42] US Fish and Wildlife Service (FWS), Report to Congress on the Recovery Program for Threatened and Endangered Species, US Department of the Interior, Washington, DC, 1996.
[43] D.S. Wilcove, M. McMillan, K.G. Winston, What exactly is an endangered species? An analysis of the US endangered species list: 1985–1990, Conserv. Biol. 7 (1993) 87–93.
[44] S.L. Yaffee, Prohibitive Policy: Implementing the Federal Endangered Species Act, The MIT Press, Boston, MA, 1982.
[45] D. Zhang, Endangered species and timber harvesting: the case of red-cockaded woodpeckers, Econ. Inquiry 42 (2004) 150–165.