Modeling Conservation Program Impacts: Accounting for participation using bootstrapping

Daniel Hellerstein Economic Research Service, USDA-ERS*

Selected Paper for presentation at the American Agricultural Economics Association Annual Meeting, Long Beach, California, July 23-26, 2006.

*The views expressed here are those of the author, and may not be attributed to the Economic Research Service or the U.S. Department of Agriculture. This paper is prepared by a USDA employee as part of his official duties, hence is in the public domain and can be reproduced at will.

Modeling Conservation Program Impacts: Accounting for participation using

bootstrapping

Keywords: CRP, participation, EBI

Author: Daniel Hellerstein*

Abstract:

Many conservation programs, such as the Conservation Reserve Program (CRP), use indices to

select offers. When modeling how changes in the index weights effect program outcomes, one

must account for the attributes of available land, and which landowners chose to participate.

This paper introduces a methodology to account for changes in participation as index weights

change. Data on the actual CRP (all offers received) are combined with an artificial population

of available lands (based on National Resources Inventory data). Bootstrapping methods are

used to calibrate estimates of participation probability, and to account for errors-in-variables

when estimating how index scores effect this probability. Preliminary analysis suggests that

accounting for participation effects will effect estimated impacts of changing the CRP's index

weights.

* Economist, Economic Research Service, USDA. The views and opinions expressed in this

paper are solely those of the author, and do not necessarily reflect those of the Economic

Research Service or the USDA.

Modeling Conservation Program Impacts: Accounting for participation using bootstrapping

Introduction

Agricultural conservation programs, such as the Conservation Reserve Program (CRP), often have multiple objectives (such as reducing soil erosion, improving water quality, and enhancing wildlife habitat). Choosing what lands to enroll requires some method of ranking offers that will account for more than one objective. In practice, indices are used to score and rank offer. These indices, such as the CRP's Environmental Benefits Index (EBI), combine measures of several biophysical attributes of an offer (such as soil erodibility, and wildlife value of a proposed cover crop) with an index-weight

Given the difficulties of constructing weights that accurately represent social preferences, in practice, index-weights are the products of technical information on physical characteristics, input from stakeholders, and available economic insights. It is of interest to examine how program results would vary as these somewhat ad-hoc index-weights change? For example, are there combinations of index-weights that greatly improve the provision of one environmental benefit, with only small decreases in the others?

The impacts of a change in weights will be a function of the distribution of physical characteristics across the landscape – since each parcel of land provides a unique mix of environmental attributes. In addition, the impacts are a function of landowner willingness to offer their land.

This paper considers the latter problem: how to account for changes in landowner participation rates as index-weights change. I focus on the CRP. The overall goal is to estimate what the CRP would look like as the weights in its Environmental Benefits Index (EBI)¹ change.

EBI = N1 * E1 + N2 * E2 + + Nn * En

Where Ni is the relative **value** of the i'th factor (for a particular offer), that ranges between 0 and 1; Ei the **weight** for the i'th factor (it is the same for all offers). For example, the N3 factor (soil erosion) weight is 100. Note that Ni * Ei is referred to as the **score** for the i'th factor.

¹ An EBI score (for a particular offer) is a vector product:

Using estimated probability of participation to compute an expansion factor

Estimating the first order effect of a change in EBI weights is simple: as weights change, the scores granted to existing offers would change, leading to a re-ranking of existing offers, implying a that a different set of offers would be accepted (assuming, say, that only the top X% of offers are accepted).

However, this simple model suffers from at two sources of "participation" bias:

- a) As the EBI weights change, it is possible that different subsets of acres (out of the currently eligible acres) will be offered.²
- b) It is likely that the set of lands currently available are systematically different from the full set of lands that are eligible. In particular, land currently enrolled in the CRP can not be "reoffered", and this land may be different than land not currently enrolled. This can lead to biased projections of the impacts on future enrollments (given a change in weights).

To account for these concerns, I use a method based on augmenting current offers with offer specific *expansion factors*. The expansion factor is based on an estimated, offer-specific *probability of participation (PP)*.

The essential notion is that each observation in the CRP offer file is representative of a larger set of acreage that could be offered into the CRP. This implies that for a given offered acre, there are other observationally equivalent acres out there. Some of these observationally equivalent acres are already enrolled in the CRP, some were offered but were not accepted, while others belong to landowners who have decided not to offer them to the CRP during this signup.

To compute the *probability of participation*, we assume that a landowners decision to offer an acre to the CRP is influenced by an acre's EBI scores, along with profitability and other concerns. By modeling the probability of making an offer as a function of the EBI score (hence as functions of EBI weights), we can estimate a new probability (of making an offer) should the

² If landowner transaction costs (such as the time and effort of submitting an offer to the USDA) are non-negligible, some landowners may not bother making an offer. In particular, landowners who judge, based on their estimate of their EBI score, that they have a small chance of being accepted; and won't submit an offer even if they would like to be in the program.

EBI weights change. Using this new probability (in conjunction with a probability at the status quo), it is straightforward to derive a new *expansion factor*.

The expansion factor measures how many acres (across the entire nation) are represented by an acre in an actual offer. As this expansion factor changes (due to changes in the underlying probability of participation), so will estimates of what lands are offered to the CRP. In particular, simulations that predict just what the CRP will look like (as index-weights change) will use the expansion factor, along with changes in an offer's EBI score, when "choosing lands" to be part of the CRP.

This simulation strategy, which augments data on existing offers with an estimated expansion factor, can be contrasted with micro-level approaches that carefully model participation using detailed survey instruments on the general population (for example, Lambert et al). Such a micro-approach has the appeal of clarity – given that one starts with a representative sample, prediction is a relatively straightforward exercise. However, the leveraging of available data is the primary advantage of this simulation approach. This leveraging is especially useful if ones goal is to detail the correlation of environmental impacts (as index-weights change) – since the offers provide a rich census of what the actual possible tradeoffs are.

The basic probability of participation model

The probability of participation model uses an estimated *probability of participation* to account for changes in the probability that a landowner will offer his land to the CRP, changes that may be due to changes in the EBI weight vector. This is a several step process that combines data from several sources.

1) LTB: The USDA Farm Services Agency (FSA) likelihood-to-bid (LTB) model is used. The LTB model is based on NRI data. It determines which NRI points represent acres are eligible for the CRP, predicts EBI factor scores for these NRI points (given an EBI weight vector), and predicts whether the land will be offered into the CRP or not.

The LTB model provides a simulated "universe" of data on US agricultural lands. In particular, the LTB can provide estimates of the acres eligible for the CRP (eligible-acres³) in each Major Land Resource Area (MLRA)⁴.

- 2) **OFFER**: The complete set of offers made to the CRP's 26^h signup form a "basis" from which we compute the total acreage in a MLRA offered into the CRP (offered-acres). Each offer contains locational information as well as information on EBI scores
- 3) **CONTRACT**: The complete set of currently active contracts contains the same information, per observation, as the OFFER file.

The model uses the LTB and OFFER data to compute MLRA specific offer rates (Or_{mlra}):

(1) $OR_{\rm m} = (acres offered in this MLRA) / (eligible acres in this MLRA).$

The participation probability can then be modeled by regressing the *offer rate* on several explanatory variables.

(2) OR =
$$f(X, \beta)$$

where

X is a vector of independent variables including an offer's EBI score, measures of land productivity, and average farmer characteristics (such as county-wide median age): β is a vector of coefficients to be estimated

The results of this regression can be used to generate an expansion factor is a function of indexweights (given an alternative vector of EBI weights):

1. For each observation in the offer file, predict

(3)
$$\overline{OR}_{i0} = f(X_{i0}, \beta)$$
 and $\overline{OR}_{i1} = f(X_{i1}, \beta)$;

³ Eligible acres are defined as land that meets crop history and other criteria for enrollment in the CRP, and that are not currently enrolled in the CRP.

⁴ More precisely, eligible acres can be estimated for each of the approximately 300 "MLRAS-within-state" areas. This choice of aggregation is based on the level at which the Natiional Resource Conservation Service deems NRI data to be "statistically reliable".

the "old" and "new" predicted offer rates.. These predictions use each offer's attributes (such as its EBI factor scores) and the estimated values of β . \overline{OR}_{i1} , uses the alternative EBI weight vector to compute the EBI scores for each offer, while \overline{OR}_{i0} is based on the weight vector in place when the actual offers were made.

4) Compute offer specific expansion factors using:

(4) XPi =
$$\overline{OR}_{i,1} / \overline{OR}_{i,0}$$

Thus, if the predicted offer rate increases from 25% to 50%, then the **expansion factor** will be 2.0.

- 5) For each observation, the <u>effective acres</u>, EA, is computed as:
 - (5) $Ea_i = actual \ acres_i * XPi$.

Where actual_acres_i is the actual acreage of observation i.

6) All the offers are sorted by EBI scores, and the "''best" offers are entered into the simulated CRP. Note that each offer's effective acres, rather then actual acres, is used when adding lands to the simulated CRP.

A bootstrapping estimator

Prediction of the "offer rate" for each observation in the offer file is odd – after all, if an offer is received, its "rate" is 1.0! However, if one considers that each offer is representative of other lands, lands that are already enrolled and unenrolled lands that were not offered (but are observationally equivalent), then an "observation specific offer rate" does make sense.

As detailed above, one can use actual offers and information on the "eligible acres" data (from the LTB dataset) to compute regional (say, county-wide) offer rates. One can regress these regional offer rates against regional (county-average) measures of independent variables (such as the EBI factor scores). The notion is that this regional data will be

representative of actual individuals, so that the coefficients from this regression can then be applied to individual observations from the offer file.

However, aggregation bias is likely to be present, especially if non-linear functions (such as probits) are estimated. In this case, it would be convenient to introduce other variables to help control for aggregation bias; variables such as standard deviations, ranges, and other such measures of the dispersion of the independent variables.

Unfortunately, there is no obvious way to use coefficients on such dispersion variables in the prediction phase. That is, for actual observations from the offer file, there is no "dispersion" information -- the attribute measures are exact.

To control for this bias, a bootstrapping estimator, I use simulated draws from an underlying population of landowners. Basicallly, a simulated dataset of "bootstrapped" observations is generated, and used in a probit estimator. The idea is to convert errors in the independent variable (aggregation bias) into errors in the dependent variable (inexact measures of outcome); a conversion that should reduce bias.

The bootstrapping estimator has several steps:

- 1. For each MLRA (m), compute several non-parametric probability density functions (PDF_m) defined over a multi-variate vector of attributes (Z). This PDF_m(Z) reports what fraction of acres (in MLRA m) are in the *cohort* that possesses attribute values of Z. The details of these probability distribution functions are discussed below.
- 2. Separate PDFs for the OFFER data (PDF_0_m), the CONTRACT data ((PDF_C_m) and the eligible acres (LTB) data (PDF_E_m) are generated. Note that a separate version of each of these functions is defined for each of the 300 "MLRA-within-state" regions. The Z- attributes over which these functions are defined are the values of the 6 EBI factors.
- 3. For each MLRA, draw (with replacement) j=1...J different *bootstrap observations* from the eligible acres contained in the LTB file. The notion is to draw a representative sample of the types of eligible land present in each MLRA.

4. For each bootstrap observation (j) from an MLRA (m), use its attributes (Zj) to lookup two cohort probabilities—

$$COHO_{j} = PDF_{j}O_{m}(Z_{j})$$
 and $COHO_{j} = PDF_{j}E_{m}(Z_{j})$ (for the offer and eligible cohorts respectively).

5. Randomly assign a dependent variable value of 0 (no-offer) or 1 (offer) to each bootstrap observation. The probability of a 1 will be:

$$P_1_i = OR_m * (COHO_O_i / COHO_E_j)$$

where $OR_{\rm m}$ is the actual MLRA-wide offer rate.

The idea is to adjust the MLRA-specific offer rate, accounting for observations from cohorts that are over (or under) represented in the "population" of offered acres (relative to the population of eligible acres).

For example,

- if PDF_O_m(Zj) predicts that 5% of offered acres are in COHO_O_i
- if PDF_E_m(Zj) predicts that 4% of eligible acres are in COHO_E_j
- Then, for offers in this "cohort" < the overall MLRA probability ($OR_{\rm m}$) will be increased by 25% (multiplied by 1.25).
- 6. Estimate the β coefficient vector, using a probit model applied to all these bootstrap observations (and the randomly assigned dependent variables accomplished in step 5)

Steps 3 to 6 are repeated R times (R=100) times to form a R row matrix (B) of coefficient vectors. The average of B vectors would be the estimate of β , with a coefficient covariance matrix also derived from the covariance of the columns of B.

Notes that the chance of a bootstrap observation (that is drawn from the LTB file) being an offer (having a dependent variable of 1) will increase as OR_m (the overall offer rate for the MLRA) increases. It also increases as COHO_O_j, the (the size of the cohort of offered acres that "look like this bootstrap observation")"increases relative COHO_E_j (the size of the cohort of eligible acres "that looks like this bootstrap observation").

Table 1 presents the Z variables, and the estimated coefficients β , from estimating the above model using data from the CRP's 26^{th} signup. Note that the some Z values are available at the "offer level", while others are derived from aggregate (MLRA-within-state, or county-wide) measures. While most variables did not have significant impacts, the overall EBI score did, as did the contract rate (the fraction of lands, in a region, currently enrolled in the CRP).

Creating a non-parametric probability distribution

As discussed above, to increase the accuracy of several components of the model, *cohorts of a region* are used. A cohort is a subpopulation, of a region, described by a vector of attributes – each acre in the cohort will have the same value of this vector. In a sense, to further the goal of using small and homogeneous aggregates, we are defining regions both over physical space (MLRAs) and "attribute space" (Z).

To do this, probability distribution functions (PDF) are defined for each MLRA; and for each type of data (offered acres, contracted acres, and eligible acres). These PDFs report the probability of observing an acre with a given vector of attributes.

One approach is to define PDFs using a multivariate normal distribution for each MLRA, with a mean and variance computed using observed data within the MLRA. However, although this is straightforward, it imposes a single-peaked structure on what might be a variously peaked distribution.

Instead, use of a non-parametric PDFs allows greater flexibility. In the simple case, of a single attribute, a histogram based algorithm could be used – with the attributes divided into a finite set of classes, and a probability computed for each class. However, when there are many attributes a histogram method becomes difficult to implement (to avoid a crippling number of empty cells requires unobtainable amounts of data). Therefore, we adopt a distance based metric.

Our non-parametric, empirical PDF is defined as follows. The PDF will return a probability-measure within a particular MLRA, for a point (P) with attributes Z.

- 1. Extract R: all observations in the MLRA
- 2. Compute a pythagorean distance from P to each point in R, where the distance is in attribute (Z) space. Actually, to avoid scaling problems, the distance is in normalized Z space the Z values of P and of each point in R, are normalized by the mean and standard deviation (of each element of Z), computed across R.
- 3. Invert these distances
- 4. Take the average of these inverses.

This average is a measure of the relative probability. Hence, if P is close to the bulk of the points in R: the average distance will be small, the average inverse will be large, and the probability will be large. The use of an inverse helps control for multi-peaked distributions – points falling near a peak (and far from another peak) will be assigned higher probabilities than points falling in between two peaks.

Reiterating, each MLRA and each set of acres (offered, contracted, and eligible) has its own, unique, PDF.

Note that this "probability" measure is relative. Not only is it relative to the particular MLRA, it is not meant to be taken as a true density function (there is no attempt to force the implied distribution function to integrate to unity). However, since these measures are used in ratios (equation 5 above), absolute accuracy is not required.⁵

Some results

In our analysis of the effects of changing the EBI (Cattaneo et al), we created multiple simulations. Each simulation is based on a different set of EBI weights, and yields a different set of "lands accepted into the CRP". In each simulation, we computed the average value (across all accepted lands) of factor scores; and then computed elasticies of factor scores with respect to index-weights.

Tables 2a and 2b compare the with, and without, probability of participation effects models. No striking differences are revealed. However, the without model has somewhat smaller elasticity values, suggesting that changes in the index-weights have a lessened impact on the benefits when participation effects are ignored. For example, reductions in expected erosion differ the most, with a 10 percent increase in the soil erosion weight leading to a 2.8 percent in the without model, compared to a versus 3.6 percent increase in the with model. This is not surprising, since weight changes can induce different bids to be submitted that favor concerns with higher weights.

⁵ It may be more appropriate to call these functions "similarity" functions.

Conclusions

In this paper we highlight how participation effects of changes in index-weights can be modeled, and then used to simulate the CRP. A simulation model, designed to fully leverage the data contained in a full census of actual offers to the CRP, was devised to account for changes in participation probability. Preliminary empirical work suggest that accounting for such change can have some impact on predicted results of a change in EBI weights.

While accounting for several sources of bias, the model presented above was fairly simple. A number of issues were not discussed. These include:

• How does land currently in the CRP differ from land not in the CRP? Using cohort weights, it is relatively straightforward to adjust for observable differences. More troublesome is question of whether can one assume that currently enrolled acres are more likely to be offered into the CRP (assuming that the CRP were to be reset to zero) then observationally equivalent acres that are not currently enrolled. If so, using the offer file (containing offers from a single, 2 million acre, signup) to compute offer-rates will lead to biased predictions. In particular, offer-rates will be underpredicted for lands that tend to have high contract rates. That is, the predicted offer rates for land most likely to be accepted into the CRP will be too low.

Note that the CRATE variable of Table 1 was included to partially control for this effect – a high CRATE capturing the proclivity of landowners in that region for enrolling into the CRP. We also experimented with a two-stage sample selection model; with the first stage estimating an enrollment rate (prior to a signup), and the second the probability of offering land during a signup. However, this model didn't reveal a correlation between stages, although empirical evidence (from large signups in the mid 1990's) suggest that most CRP landowners will re-enroll their land.

• How will practices change as weights change. Landowners can effect their EBI scores by planting different cover crops. The current methodology does not allow for this – it

assumes that offers are representative of a fixed pool, a pool whose attributes do not change.

• The CRP is not formed from a single signup. Thus, if infra marginal effects are of interest (say, if the CRP were to be started from scratch), the dynamics of enrollment over multiple signups may be important, since landowners have several chances to offer their land.⁶

Lastly, the statistical properties of this methodology are unknown. Future work will use controlled simulations, using artificial universes with based on a generated (hence known) set of underlying data.

⁶ Note that in other work, we simulate a "full CRP", via a repeated simulation model, that reduces eligible acres over the course of several signups.

References:

Lambert, Dayton, Patrick Sullivan, Roger Claassen, and Linda Foreman, "Conservation-Compatible Practices and Programs: Who Participates". Economic Research Report (Economic Research Service/USDA.) no. 14. 2006.

Cattaneo, Andrea, Daniel Hellerstein, Cynthia Nickerson, and Chritina Myers, "Balancing the Multiple Objectives of Conservation Programs". Economic Research Report (Economic Research Service/USDA.), in press. 2006.

TABLE 1: Coefficient and estimated coefficients of the probit model

Var	Estimated beta	stderr	tstat
Constant	-2.97337	0.937	-3.17
SRR	0.000834442	0.0021	0.38
EBI_TOTA	0.00155435	0.000203	2.69
CPA	-0.091982	0.071	-1.29
AVGAGE	0.00628953	0.017	0.36
MEAN_BID	-0.000779105	0.001	-0.44
HIGH_COS	-0.0630138	0.128	-0.49
RETURNBY	-0.604321	0.478	-1.26
CRATE	0.853022	0.127	6.68

Where:

SRR	Offer's soil rental rate (one of the EBI factors).
EBI	Offer's EBI score.
CPA	0/1 dummy if the county is in a Conservation Priority Area
HIGH_COST	0/1 dummy: 1 if the county is in a high-cost region
AVGAGE	average age of proprietor in county
MEAN_BID	County average of mean minimum bid acceptable to farmer (estimated from
	values generated by the LTB model)
RETURNBY	county level measures of total net cash returns divided by total cropland
CRATE	county level contract rate (fraction of eligible lands currently enrolled in the
	CRP)

Table 2a Simulated impacts of changing EBI weights – with participation effects

Dependent Variable	Independent variables				
	Wildlife Weight	Water quality weight	Erosion reduction weight	Enduring benefits weight	Air quality weight
Wildlife impacts	0.133	-0.015	-0.126	0.002	0.003
Water quality impacts	0.034	0.240	-0.022	-0.010	0.002
Erosion reduction impacts	-0.104	-0.039	0.362	-0.045	-0.025
Enduring benefits impacts	0.049	-0.118	-0.262	<u>0.324</u>	-0.017
Air quality impacts	-0.010	-0.068	-0.124	-0.016	<u>0.040</u>

Elasticities computed across 1000 simulations

Table 2b Simulated impacts of changing EBI weights – without participation effects

Dependent Variable	Independent variables				
	Wildlife Weight	Water quality weight	Erosion reduction weight	Enduring benefits weight	Air quality weight
Wildlife impacs	0.106	-0.016	-0.102	0.002	0.004
Water quality impacts	0.022	<u>0.204</u>	-0.019	0.002	0.010
Erosion reduction impacts	-0.091	-0.034	0.282	-0.013	-0.011
Enduring benefits impacts	0.015	-0.116	-0.177	0.260	0.002
Air quality impacts	-0.008	-0.055	-0.105	-0.011	0.033

Elasticities computed across 1000 simulations