Appendix C—Description of the Wildlife-Viewing Valuation Model

In addition to pheasant populations, changes in the Conservation Reserve Program (CRP) may affect a variety of other wildlife populations, with resulting impacts on public participation in nonconsumptive wildlife-based outdoor recreation (wildlife viewing). To study this possible relationship, FWHAR data were used to model how the CRP influences the public's participation in wildlife viewing.

The nonconsumptive wildlife-associated recreation component of FHWAR data contains approximately 26,000 observations. This analysis uses the number of trips taken for the primary purpose of viewing wildlife, and focuses on the trips taken within the general vicinity of the individual's home, which roughly translates to all trips taken within approximately 100 miles of the individual's residence.

The size and extent of the FHWAR database are the primary features motivating its use. However, counterbalancing these positive features is the paucity of information on the recreational sites visited by respondents. As with the pheasant and water-oriented recreation models, site-specific information is very important, since landscape characteristics (that is, the extent of CRP) are likely to influence recreational behavior.

In contrast to the pheasant model, which used ancillary information (the Breeding Bird Survey) to designate visited sites, the available data could not be used to impute which (of several possible sites) the respondent visited. Thus, rather than select a particular site, the model examines how the aggregate trip-taking behavior of an individual is a function of the aggregated characteristics of all the sites available to that individual.

Briefly, the analysis involves the following steps:

1) Create "landscape characteristics" variables defined at a number of "semi-circular zones" around each respondent (**LC**), which are then aggregated into "circular-distance-band, aggregated" landscape characteristic variables (**Z**).

2) For each individual, extract visitation (**Q**), and socioeconomic (**X**), data from the 1991 FHWAR.[18]

3) Using the "distance to most frequently visited site" as a dependent variable, estimate a representative trip price (**P**).

4) Regress total number of trips against **X**, **P**, and **Z**.

5) Using coefficients from step 4, estimate predicted number of trips (and consumer surplus) under several scenarios.

The following sections provide greater detail on each of these steps.

Imputing Landscape Characteristics

Following the procedure used in the pheasant study, the ASH (Scott and Whittaker, 1996) technique is used to create a variety of landscape characteristics from National Resource Inventory (NRI) data. Several broad measures of land use as proxies for wildlife habitat (and potential populations) are created:

1) percentCRP. The percent of the land (in the sub-county region) that is in CRP.

2) percentCROP.

3) percentFOREST.

4) percentGRASS (rangeland and pasture).

5) RUC: Rural-Urban Continuum code (0 being most urban, 9 being most rural).

6) DIVERSITY: Landscape diversity, with higher values of DIVERSITY indicating a more variegated landscape (based on the interspersion of water bodies, forest land, grassland, and cropland).

[18]This requires knowledge of a key piece of information: the individual's residence. Since public-use releases of FHWAR do not contain this information (due to privacy concerns), analysis of the data necessitated using raw data at U.S. Census facilities.

For each FHWAR respondent, 19 "zones" are defined. These zones are equivalent to those 16 zones generated for the pheasant study, but with the inner zone divided into 4 components.[19]

To simplify the model (and avoid problems with missing data), the landscape characteristics (**LC**) of the 19 zones are aggregated into 5 "distance-band-aggregated" measures (**Z**). To account for the possibility that landscape heterogeneity may be important, we use a "constant elasticity of substitution" functional form to compute an aggregate measure.

Specifically, this measure is defined as:

(C.1) $$Z_k = \left(\sum_{j=1}^{J_k} (LC_{jk})^{1/\alpha} \right)^{\alpha}$$

where:

α is a parameter to be estimated,

J_k is the number of zones in the k-th *distance band* (that is, the 62-mile distance band four has 5 zones),

LC_{jk} is the value of the characteristic in the j zone of band k,

and

Z_k is the aggregated measure of the land characteristic of the k-th *band*.

Note that when

$\alpha = 0$: Maximum matters
$\alpha = 1$: Sum (or average) matters
$\alpha >> 1$: Variations in characteristics do not matter

[19]These four zones are defined as zone 1 being within 12 miles of the "own ZIP Code" centroid; and zones 2 through 4 between 12 and 25 miles of the centroid.

Given the K=5 distance-zones, and six characteristics (listed above), this yields 30 separate distance-band-aggregated landscape characteristic (**Z**) variables.[20]

Individual Data

Data on nonconsumptive wildlife-associated recreation were obtained from the 1991 FHWAR. For each surveyed individual, the number of visits to "non-distant" sites was extracted. Operationally, this involved several steps:

1) Trips to one's own State, and to all States for which the "most visited location" was within 100 miles of the resident's home, were summed to obtain total "non-distant" trips (for all "potential wildlife viewers").[21]

2) Using information on past participation, and on current plans, observations on individuals who were not likely to be "potential wildlife viewers" were dropped.[22]

3) Several socioeconomic variables were extracted for each individual, including male, caucasian,

[20]The five distance zones could be further aggregated into an overall measure by using an endogenous distance decay. Although this yields a more parsimoniously specified model, it also complicates estimation.

[21]This focus on trips to "non-distant" sites is necessitated by modeling concerns; such as the large number of "sites" one would have to include in order to account for far away trips. However, note that trips to these "non-distant" sites account for over 90 percent of nonconsumptive wildlife-oriented trips.

[22]The following table contains the percent of observations in four categories.

	0 Trips	>0 Trips
Excluded observation	18 percent	7 percent
Included observation	45 percent	30 percent

Ideally, the "excluded-observation / >0 trips" category should contain 0 percent (since individuals who took a trip should not be *a priori* excluded). Conversely, the "included-observation / 0 trips" category may contain a large percentage, since it is possible for potential participants to choose 0 trips in any given season. Overall, approximately 85 percent of all trips were accounted for by individuals retained in the sample.

rural residence, high school graduate, college graduate and household income. All but household income are dummy (0/1) variables. Household income is an approximation based on the center of seven broad ranges (in the $0 to >$75,000 interval).

4) Individual weights were also obtained for each observation. These are demographic weights, computed by the FHWAR survey designers, that are used when creating population level predictions.

Constructing an Imputed Price

For several reasons, it is desirable to include an explicit price information in the analysis. First, if explicit price information can be obtained (say, a "representative" price), then welfare estimates using consumer surplus may be readily computed. Second, including such price information should improve the model's performance.

The problem is, as with landscape characteristics, the paucity of knowledge about which sites were visited implies a lack of explicit price information; a problem that is exacerbated when individuals took zero trips. As a substitute, an imputed "representative" price can be used.

The imputed "representative" price is based on the FHWAR's distance to the most frequently visited site variable. This distance variable is converted into a travel cost, using average cost per mile information, and a simple time cost (based on a fraction of household income). This travel cost is then used as the dependent variable in a sample selection model. The use of this predicted price offers two advantages: as a control for potential simultaneity between "price" and "quantity of trips," and to impute a price for the (many) individuals who consumed zero trips.

To predict this price, a sample selection model is used:

$$\text{i) } \Pr ob\left[\text{Participant}; \gamma\right] = \Phi(\chi, \gamma)$$

(C.2) $$\text{ii) } E\left[\ln(\text{Price}); \beta_\chi, \beta_v | Q > 0\right] = \chi\beta_\chi + \frac{\phi(\chi, \overline{\gamma})}{\Phi(\chi, \gamma)}\beta_v$$

where:

χ = X and Z variables

$\overline{\gamma}$ = Predicted value of γ from step 1

ϕ = Normal pdf

Φ = Normal cdf

The first equation is a Probit on whether the individual took any trips, with Z and X used as regressors. The γ coefficients from this Probit estimator are used to compute a Mills ratio. This Mills ratio, along with the Z and X variables, are regressed against the log of observed price (using observations with non-zero trips) in a standard semi-log OLS. Lastly, the predicted values of γ, β_χ, and β_γ are used with equation C.2.ii to impute a price for all observations (including individuals who took zero trips).

The Demand Estimator

Using the **X** (socioeconomic), **P** (imputed price), and **Z** (aggregated landscape characteristics) variables, a "representative trip" demand curve is estimated.

To clarify, lacking good information on where people went on their "wildlife-associated" trips, rather than selecting a "visited site" (using ancillary information) and estimating a RUM model (viz., the pheasant model), the wildlife-viewing model focuses on the total number of wildlife-associated trips within a few hours' drive of an individual's residence. Hence, the use of the aggregated landscape characteristic (**Z**) variables to estimate total trips is best interpreted as arising from a reduced-form model of the site-selection problem solved by an individual recreator. That is, the reduced-form model combines trips to all sites into a "total number of trips," and uses aggregated landscape characteristics to explain the total number of trips taken. Thus, the determinants of a set of chosen trips (to unknown-to-the-analyst sites) are "represented" by these aggregated characteristics.

To control for the prevalence of zero trips, a double-hurdle Poisson estimator is used to estimate the representative trip demand curve.

(C.3) $$\Pr ob(q; q > 0 | \lambda, \gamma) = \left(\frac{e^{-\lambda}\lambda^q}{q!}\right)(1 - e^{-\gamma})$$

$$\Pr ob(q = 0 | \lambda, \gamma) = e^{-\lambda} + \left((1 - e^{-\lambda})e^{-\gamma}\right)$$

where:

λ is the quantity parameter: $\lambda = \exp(R\beta)$

γ is the participation parameter: $\gamma = \exp(S\tau)$

β and τ are parameters to estimate

R are factors that influence the number of trips; including **P** (price information), **X** (socioeconomic factors), and **Z** (aggregated landscape characteristics)

S are factors that influence participation (a subset of **X** is used)

Note that, along with the imputed price term (**P**), the use of distance zone variables allows cost differences to affect demand.

Summarizing the Model

The model to be estimated is:

$Z = Z(LC, \alpha)$

$P = P(X_2, Z; \gamma)$

$q = F(X_1, X_2, Z, P, W; \beta_1, \beta_2, \beta_p, \beta_z)$

where:

q = Number of trips.

X_1 = Individual specific variables that influence probability of participation; typically socio-economic variables.

X_2 = Individual specific variables that influence quantity of trips; typically socioeconomic variables. Note that X_1 and X_2 may contain the same variables.

Z = Aggregated landscape characteristic variables, for z different variables and k=1..K bands. These will be a function of the 19 **LC** variables and **α** (the "CES" aggregation parameter).

P = The imputed price of a trip. Based on a sample selection model with the observed "distance to favorite site" as the dependent variable, and X_2 and Z as the independent variables.[23]

Z() = The "distance-band" landscape aggregation variables (equation i).

P() = The "sample selection" imputed price model (equation ii).

F() = The hurdle Poisson model (equation C.3).

W = Population weight correction factor.[24]

and

$\beta_1, \beta_2, \beta_p, \beta_z \alpha, \gamma,$ = Parameters to be estimated.

Although simultaneous estimation of the above would be optimal, operational difficulties dictate a multi-stage model, to wit:

---

[23]The observed price term is computed as the sum of an out-of-pocket cost and a time cost:
P= [0.3 * DIST] + [WAGE * 0.33 (DIST/50)]
Where:
DIS: Distance to site (in miles)
0.3: Approximate per mile cost of using a car
WAGE: Imputed wage rate = Household income divided by 2040.
DIST/50: Time required to travel DISTANCE
0.33: Fraction of travel time that is "onerous." The assumption is that recreational travel is not as unpleasant as work, hence should not be valued at the wage rate (Shaw, 1992).
Note that the following is assumed:
1) The WAGE rate assumes that the trip taker is the sole wage earner in the household; and freely chooses to work 2,040 hours. 2) Out-of-pocket costs (0.3 * DIST) assume a group size of one (no cost sharing, and no variation in fuel economy, depreciation rates, etc. 3) An average speed of 50 miles per hour.

[24]When using the FHWAR weights to scale up to the population, the desired equivalence between "observed" and "predicted" (using the baseline data) number of trips need not hold. There are several ways of addressing this inconsistency; including ex-post calibration, weighted estimation, or inclusion of the weight as a correction factor. Though all of these are problematic, the use of the weight as a correction factor involves the fewest ad hoc assumptions.

1) Using a grid search, select a candidate value of α.

a) For each candidate value of α, the $\mathbf{Z_k}$ variables are generated.

b) Given **Z**, **P** is imputed.

c) Given **Z** and **P**, estimate F(·).

d) Record the log-likelihood from c.

2) Reiterate step 1 for different values of α.

3) Given a set of coefficient vectors (one vector for each value of α) choose the one with the best log-likelihood. The β coefficients associated with this best log-likelihood are the estimated parameters of the model.

Since it might be expected that recreational behavior may vary across the country, this model was applied separately to the five sub-national regions: the West, Northern Plains, Southern Plains, North, and the South.[25]

Constructing Alternative Scenarios

Total recreational trips under different allocations of the CRP are estimated under the three scenarios discussed in the body of the paper: 1992 CRP (signups 1 to 11), a "no-CRP" scenario, and a "15th EBI" scenario.

[25]The five regions consist of:
1) West: CA, WA, OR, MT, ID, WY, NV, UT, CO, AZ, NM.
2) Northern Plains: SD, ND, NE, KS.
3) Southern Plains: OK, TX.
4) North: MN, WI, MI, IA, MO, IL, IN, OH, ME, VT, NH, CT, RI, MA, NY, PA, MD, NJ, DE, DC.
5) South: AR, LA, MS, AL, GA, SC, FL, KY, TN, WV, VA, NC.

Appendix table 5 lists the percentCRP (and percentCROP) "perceived" by the FHWAR sample under these scenarios. Since the FHWAR sample is not uniformly distributed geographically, these percentages will differ from the actual landscape distribution in the regions.[26]

Across scenarios, the **LC** variables for each observation will be different. The impacts of these changes are examined by recomputing the predicted number of trips, using equation C.1 to recompute the aggregated landscape characteristics (**Z**), equation C.2 to recompute **X** and **P**, and the estimated coefficients from the model (equation C.3) to generate new predictions of trip demand.

Some Results

Screening information (on past wildlife-associated recreation) was used to classify approximately two-thirds of the sample as being potentially interested in "nonconsumptive wildlife-associated" recreation; the remaining one-third of the sample was classified as uninterested and was not included in the estimation. About one-half of the potentially interested individuals (one-third of the sample) actually took at least one trip (appendix table 6 gives further details). Note that the average reported trip value is based on a contingent valuation question asked of everyone who took a nonconsumptive, wildlife-oriented trip (Waddington and others, 1993).

The canonical estimator for this model, as described above, is based on a double-hurdle Poisson model and an imputed price. Given the large number of

[26]These "perceived" values are derived in the following manner. First, for each respondent, compute an average percentCRP (and average percentCROP) in the 19 "zones" (in the approximately 100-mile band surrounding his/her residence). Second, average these "100-mile-band" averages.

Appendix table 5—Perceived percentCRP and percentCROP

	1991 CRP (34 million NRI acres)		15th EBI		No CRP	
Region	percentCRP	percentCROP	percentCRP	percentCROP	percentCRP	percentCROP
West	1.2	9.9	1.1	9.9	0	11.2
Northern Plains	4.5	52.3	4.3	52.4	0	56.9
Southern Plains	0.9	19.2	1.4	18.6	0	20.0
North	1.1	28.9	1.6	27.8	0	29.9
South	0.9	14.6	1.3	13.6	0	15.4

Source: USDA, ERS.

Appendix table 6—Regional summary of participation in nonconsumptive wildlife-associated recreation

	Number of observations	Number retained	Number of participants	Average number of trips	Average distance	Average reported trip value
West	5,561	3,391	1,624	9.33 (16.7)	22 (33)	30
Northern Plains	2,075	1,679	659	11 (23)	13 (18)	25
Southern Plains	992	785	270	8.9 (19)	25 (44)	31
North	9,827	7,878	3,122	13 (23)	14 (30)	32
South	6,451	4,699	1,547	10 (19)	15 (19)	31

Source: USDA, ERS.

variables, appendix table 7 lists some of the more important variables. We also list the "sum" of the β_z coefficients for each landscape characteristic, which can be interpreted as the effect given a uniform change in landscape characteristics.

Note that the coefficients are best interpreted as the percent change given a unit change in the variable. The probability variables range from 0 to 100; the RUC (rural-urban continuum code) ranges from 0 to 9, and the diversity variable ranges from 1 to 4.

These coefficients are somewhat difficult to interpret, as they show no strong pattern. PercentCRP seems to be more often positive then negative, with the exception of the Southern Plains.

As a measure of model quality, the correlation between the weighted observed and weighted predicted number of trips (based on the original scenario) can be used in lieu of an R-square statistic.

Ideally, the coefficient on the imputed price could be used to generate consumer surplus values. Unfortunately, the imputed price coefficients are often positive (or negative but very small in magnitude), which yields impossible (or implausible) consumer surplus values. It would appear that the distribution of quality sites obscures the price relationship.[27]

[27]If the distribution of site quality varies over the population (with some individuals living close to better sites, while others must travel long distances to attain better sites), then the imputed price should be correlated with number of trips. That is, better quality sites nearby should yield more trips to closer sites; hence a negative sign on the imputed price coefficient. On the other hand, if the shape of the distribution of site quality is similar across the population (say, increasing with respect to distance), but with some individuals having better all-around choices (say, the slope of the distance/quality relationship varies across individuals), then high prices may be associated with high number of trips. That is, individuals who can pay a high price for a "fabulous" site may take more trips than individuals who choose a closer "mediocre" site instead of a farther out "slightly better than mediocre" site.

Goodness of fit: Observed and predicted trips

	Correlation: individual trips	Correlation: weighted trips
West	0.18	0.40*
Northern Plains	0.31	0.15
Southern Plains	0.43	0.41
North	0.24	0.18
South	0.19	0.17

*When a large outlier was not removed from the West, the weighted correlation was 0.81.

However, since the "imputed price" does allow extra information (the "distance to last site" data) to be incorporated, we will retain the results with the understanding that the "imputed price" is to be interpreted loosely.[28]

Instead of directly computing consumer surplus, we use a benefit's transfer value. In particular, the "average per day" value of wildlife watching is used as a proxy for per-trip value. Although several sources for such a value exist, the "self-reported" value from the FHWAR is most appropriate for this exercise. The regional averages of these values are used to report the "consumer surplus" of wildlife-viewing trips under the three scenarios.

[28]A number of other specifications were attempted, including models without imputed price terms, and models that used the simple Poisson model. The results from these models were qualitatively similar to the double-hurdle, imputed "price" model.

Appendix table 7—Some coefficients from the double-hurdle model with imputed price (t-stats in parentheses)

	West	Northern Plains	Southern Plains	North	South
Aggregation parameter α	0.47	2.33	1.5	4.7	0.47
Some probability stage coefficients					
INCOME	-2.13e-7	-1.58e6	4.14e-6	-1.02	1.22e-6
	(-.19)	(-0.82)	(1.5)	(-1.3)	(1.0)
OWN_CRP	-0.019	-0.052	0.13	0.042	-0.047
	(-12.1)	(-2.3)	(3.4)	(4.1)	(-1.66)
Some quantity stage coefficients					
Income	-9.34e-6	5.13e-6	6.1e-6	-9.5	-5.0e-6
	(-10.7)	(8.4)	(4.4)	(-15.5)	(-6.5)
CRP0	1 0.352	0.023	-1.17	0.011	0.89
	(12.2)	(0.61)	(-5.8)	(1.17)	(17.4)
CRP02	-0.234	-0.002	0.30	0.00053	-0.65
	(-9.9)	(-0.44)	(5.7)	(7.7)	(-15.4)
CRP03	0.0077	0.0010	-0.13	-0.00066	0.15
	(0.86)	(0.40)	(-4.7)	(-9.0)	(8.3)
CRP04	0.061	0.0013	-0.004	-4.31	0.016
	(12.8)	(2.1)	(-0.42)	(-0.5)	(1.9)
CRP05	-0.028	-8.06e-5	0.037	1.6e-5	-0.003
	(-8.3)	(-0.4)	(-5.60)	(10.0)	(-0.056)
Price	0.029	-0.0093	-0.002	0.133	0.027
	(4.8)	(-0.90)	(-0.34)	(11.9)	(3.14)
Summation of landscape characteristic coefficients					
Σ percentCRP	0.16	0.02	-0.96	0.01 0.	40
Σ percentCROP	0.57	.10	0.02	-0.01	-0.001
Σ percentForest	0.15	.27	-0.04	0.003	0.008
Σ percentGrass	0.005	.03	-0.01	-0.03	0.02
Σ percentDiversity	-0.31	-2.58	.37	-0.02	1.6
Σ RUC	0.11	0.05831	0.29	0.06	-0.003
Log likelihood	17671	10733	3239	59802	20316

Source: USDA, ERS.