Ecosystem Service of Recreation for GEP

Matthew Braaksma Lifeng Ren Spencer Wood
Ryan McWay

September 20, 2024

Table of contents

1 Motivation

- We want to estimate the monetary value of recreation as an ecosystem service.
- We would like to do this credibly at a global scale to produce a country-year panel.
- This is a subcomponent of the larger GEP value for each nation and likely a sizeable value.

2 Components of Recreation GEP

- Nature-based recreation can be sub-divided by the type of recreation (park, camping, beach, lake activites, eco-tourism, etc.).
- The key distinction for nature-based recreation is that these are activities where the main attraction (or attribute deriving utility) comers from nature.
- For our purposes, we can start by narrowing our examination to camping or hiking recreation in protected areas. But in general, we can modify the following approach to consider a wider definition of nature-based recreation.
  - Consider what are the $i \in I$ that we wish to estimate and what we can't estimate.

3 Previous Estimates of Recreation Value

1. SEEA Accounts (Vallecilloa et al. 2019, Ecological Modelling)
   - ESTIMAP-Recreation (EU Only)

    - 17% of land suitable for daily recreation
    - 500 million inhabitants (62% close to suitable land)
    - Value: EUR 50 Billion

2. GEP of Ecotourism in Qinghai (Ougang et al., 2020)
    - Contingent Valuation survey data of travel expenditures (P, Q) for 'eco-tourism'
    - They conduct the questionnaire at 3 parks of Qinghai and get the following key information
        - Visits Staying Time (days)
        - Travel Expenditure (X $/person)
        - Salary of Visitor (Y $/person/days)
        - Where does the visitor come from (Zone Z)
    - Based on this information, they can get for all the zones they are interested in. And they get for each zone Z, what is the frequency/probability that people will visits 3 parks in Qinghai.
    - Then they reweight by zones for all respondents' visiting rates and zonal population and sum up all costs (Travel Expenditure and Time expenditure) to equal to the GEP of ecotourism.
    - 462 respondents; 2 parks, 1 lake
    - Value: Yuan 3 Billion in 2000, Yuan 21.6 Billion in 2015

4 GEP Formula

Deterministics (accounting) formula:

- Panel of monetary value over year t and country c over the set of recreation activities I
- Requires aggreagtion to value by year and country to be consistent with system of national accounts (SNA) as presented with SEEA accounts.

$$GEP_{c,t} = \lambda \cdot (\sum_{i \in I} P_{i,c,t} \cdot Q_{i,c,t})$$

- λ: Nature's contribution
- P: Price $\rightarrow$ Travel cost for park visitors
- Q: Quantity $\rightarrow$ Number of park visitors

While λ may vary over time and space, there is a theoretical and empirical decision to assert it is constant. Empirically, difficult to measure a change over time. Theoretically, if the value of recreation comes from contributions of land, labor, and capital – is the relative contribution of land compared to labor and capital changing (meaningfully) over time? Does it vary over country (componsition of the country)? Probably linked to structural transformation (development of a country).

5 Estimating Recreation GEP

5.1 Method 1: Extrapolation from GDP Measures

5.1.1 General Model

$$GEP_{c,y} = \lambda \cdot \theta \times X_{c,y} \times Z_{c,y}$$

- X: Share of GDP from tourism
- Z: Annutal GDP
- λ: Nature's contribution adjustment
- θ: % of tourism that is nature based

5.1.2 Data Requirements

1. Share GDP from tourism related to nature
    - Tourism Satellite Accounts (TSA) for scenic and sight seeing travel cost
    - BEA
    - ORR
2. National GDP measures
    - From the world bank

5.1.3 Calibration

- Test if this aligns with SEEA measures or UNWTO measures

5.2 Method 2: Outdoor Recreation Extrapolation

- USA-ITA breaks down to GDP based on the recreation activities for both national level and state level
- This data not available for the other countries (can extrapolate to other countries at global level) H- owever, it is hard to tease out the labor portion for calculating GDP_recreation.

What we calculate now is an upper bound assuming

$$GEP_{rec} \leq GDP_{rec}$$

5.2.1 General Model

Estimate the relationship (elasticity) between GDP recreation and size of the recreation areas. Then if we know the recreation area size we can use this as a proxy for GEP and translate this into monetary value through GDP. The logic is that larger recreation areas will generate more GDP.

Run the following state level regression

$$GDP_{s,y} = \alpha + \beta X_{s,y} + \varepsilon$$

- GDP: State s by year y recreational GDP
- X: % of recreational area divided by total size of the state
- β: elasticity between size of recreation area and GDP recreation share

With this estimate we can use the β to determine GEP for other nations using the same approach

$$GEP_{c,y} = \lambda \times (\beta \cdot X_{c,y})$$

- X: % of protected area divided by total size of the country
- λ: Nature's contribution adjustment

A big limitation of this is that we are assuming that the relationship of GDP to GEP for recreation in the US is representative of other countries – especially those in lower income brackets. This is unlikely to be true.

5.2.2 Data Requirement

1. Use USA Outdoor Recreation GDP at state level
    - Conventional recreation, boating/fishing, RVing, Snow activities
2. Measure accessibility to state parks
    - PAD-US-AR

5.3 Method 3: Balmford et al. 2015 Measure

5.3.1 Estiamting Quantity Q

For each country c, we can use the number of estimated visits annually for each protected area using the Balmford et al. 2015 measure (~ 8 Billion visits annually).

$$Q_c = \sum_{i \in I} Visits_{ic}$$

The main issue with this quantity it is likely can over-estimate of visitations. It models predictions of visitation by:

- PA size
- Local population size within 100KM of PA
- Remoteness by travel time of city of >50,000 using global friction surface
- Natural attractiveness using a 1-5 rating scale by 3 conservation scientists
- National income

In their materials, they use a generalized linear model but (1) do not show the estimation equation and (2) do not show validation of the model or out of sample prediction.

5.3.2 Estimating Price P

Try to estimate the average expenditure for a recreation visit by dividing the GDP per capita attributable to recreation by the number of park visitors annually. You could measure the expenditure on recreation through UNWTO share of GDP or as international tourism reciept (a bit more broadly defined).

Two issue:

- GDP measure will be for all recreation (includes tourism, etc.). So this will over-value the price
- It is difficult to find data on only nature-based recreation GDP

$$P_c = \frac{GDP_{pc} fromRec_j}{Q_c}$$

5.4 Method 4: Ryan's Imagination

5.4.1 Estimating Quantity Q

- Main issue: We do not observe park vistors in systematic way across global and across time.

Inspired by the Balmford et al., 2015 method. Ideally improved. What I am aiming to do is measure a propensity to visit using attributes of the park that are attractive and the distance of the park. Then multiplying this probability to visit by the known population from the origin to obtain the number of visitors a park gets each year.

Start with the gridded population data. The idea is to estimate a likelihood function to determine the probability that a representive agent in any grid cell will visit each park. We can estimate this primarly using distance to parks. In this setting, we assume that likelihood to visit is primarly determined by distance. But we add in other attributes that make a park attractive (just as is done in Balmford).

The measure of distance can be determined by:

- Euclidean distance using raster algebra
- Travel time calculated by Open Street Maps

We can place this framework within a gravity model.

$$V_{ij} = G\frac{M_i^{\beta_1} M_j^{\beta_2}}{D_{ij}exp(Border)} \cdot \eta$$

- G: Base rate of gravity. In our context this is the base rate of visitiation between grid cell i and park j. This would be the frictionless rate of visitation (perfectly accessible park)
- M_i and M_j: These are the attration properties of grid cell i to park j. For grid cells, this may be that they are urban/rural, income, or other attributes that make them have increased desire to visit a park. For parks, this could be attributes about the park like features, scenic value, size of the park, designation as national, state or local.
- D: the distance between grid cell i and park j
- $Border$: This is dummy variable that applifies the fricition to distance if the visitor has to cross a national border (administrative or logistical fricitions)
- η: A multipilicative error term for all other attributes that pull the bodies together and push the apart. Otherwise, this is a deterministic model.

We can log-linearize this model:

$$log(V_{ij}) = \beta_0 log(G) + \beta_1 log(M_i) + \beta_2 log(M_j) + \beta_3 log(D_{ij}) + \beta_4 Border + \eta$$

We will need to match this model to real data. So to take this to the data we will likely want to use a Poisson Pseudo Maximum Likelihood (PPML) method that accounts for heteroskedastic errors (Silva and Tenreyro, 2006). We will need some data on the number of visits from i to j. So this is some park level information on annual visits generally from across the world. This is difficult to do for all parks globally. But suppose you can do it for some subset of parks that are representative of all parks (externally valid). Then you can estimate the above model and use the β coefficents as a calibration method to extrapolate the number of visitations V for each park given attributes that are known globally about M_i and M_j and our estimated distance between the two locations.

At the end of this exercise, we will have the number of estimated visits from each park to each grid cell. This allows us to generate:

1. The total number of visitations to each park.
2. The total visitations from each grid cell.

Ideally, we can validate our results for each with some aggregated data on park vistations from tourism and park services for parks and from cities. This should add to the credibility of our estimates for the quantity of visitors.

In the end, for GEP we will just need to aggregate up the number of visits by *parks* to get the Q in the forumla for each country. We should use park value (rather than grid values) because that is where the ecosystem service is being derived from. So we don't actually care where the visitor comes from. What we care about is where they go. That is where the production value is being produced (and quinsendentally consumed as this is a non-tradeable good).

A Suggestion by Steve: Do a zonal calculation. Only consider people within a certian proximity of each park rather than the whole world. This relies on the assumption that after a certain buffer, travel to a PA is prohibitive (either in expense or legally).

5.4.2 Estimating Price P

- Main issues: (1) We do not observe prices for park visitors, and (2) most recreation areas are often free access or subsidized so we have a non-market good to consider.

The underlying principle to estimate the price for recreation is to measure a Willingness to Pay (WTP) through revealed preference of travel cost. The underlying assumption in this setting is that the primary determinate of visiting a park is distance. So the further the distance, the more valuable the park must be to the visitor. Note that this can be confirmed from the gravity model in the previous section using β_3. This should be the largest coefficient relative to observable attributes in M_i and M_j. If it is not, then distance is not the main driver of revealed preferences.

A nice attribute of travel cost is that we can associate this cost to known costs of travel in monetary terms. So, persume (naively) that everyone drives or flies to get to their destination.

Then we can use gasoline prices as a measure of cost and multiply that by the distance traveled. You can imagine how this could be refined. In this way, we avoid the infeasible necessity to use a survey method for visitors to report costs. This can be a way to validate our cost measures if there are parks/studies with this information. Such an exercise would be designed to change the travel costs to better reflected survey reports to total cost. So you could imagine the following simple regression to go beyond a basic gas prices × distance measure.

$$C_{ij} = \alpha_0 + \alpha_1 G_i \times D_{ij} + \alpha_2 X_i + \varepsilon$$

- C: Cost paid by visitor i to go to j
- G: Gas prices paid by visitor i
- D: Distance for visitor i to j
- X: Any other costs that are reported

From this regression, what can use from the subset of survey data is determine the influence of gas and distance α_1 relative to the importance of all other costs α_2. We can also confirm in this way that gas costs from distance is the main travel cost. We know the distance and gas globally, but not the other information. So we can try to measure the porportional importance of other costs to total cost. Say something as simple as $\frac{D}{C}$ as a probability weight for portion of cost attributable to distance. Flip this an we can re-weight all of our travel costs with an inverse probability weight. Namely, $w = \frac{1}{\frac{D}{C}} = \frac{C}{D}$.

From here forward I am laying out the groundwork for backing out the WTP measure in monetary terms. In case it gets lost in my explanation, the core aspect we care about is likelihood to visit determined by importance of income net of travel costs. This should give us the price for any given park as determined by distance.

Let us start with a discrete choice model. In this model, visitor n is selecting which park j they would like to visit. In any time t they can only select one park j from the set of parks J. We observe ex-post where a person has choosen to visit. So we will develop a behavioral model to describe why they made that choice using observable attributes about the individual visitor and the selection of parks avaliable to them.

Let us start with a random utility model (REM). The visitor has observed attributes s_n. We will say that these are heterogenous preferences across the population. So if we observe differences in income or being a foreign national we can distinguish visitation by differences in tastes. The visitory selects over alternative parks j which have observable attributes $x_{nj} \forall j \in J$. These are things like scenic quality, size of the park, amenities, travel cost to get there, etc. Each visitor recieves some representative utlity for each park they *could* visit:

$$V_{nj} = V(x_{nj}, s_n) = \beta_n' x_{nj} \forall j$$

β_n' allows for variation in taste and this is described by an i.i.d. extreme value distribution $\beta_n \sim f_n(\beta|\Theta)$. The β captures the s_n attributes varying across vistors. This is their difference