Using Regression-based Sensitivity Analysis in Exploratory Modeling of Complex Spatial Systems: An Example of Simulating the Impact of Agricultural Water Withdrawals on Fish Habitat

Glenn O’Neil (Institute of Water Research – Michigan State University)
Arika Ligmann-Zielinska, Ph.D. (Department of Geography – Michigan State University)
What issues arise in employing regression techniques to evaluate the sensitivity of a complex, agent-based spatial model to its individual parameters?
Overview:

- Conceptual agent-based model of fish habitat impacts from agricultural water withdrawals.

- Model sensitivity analysis through OLS regression.

- Issues with count data

- OLS alternative: negative binomial with hurdle
Conceptual Model

- Farmlands (CLUs) (land owner agents)
- Agricultural Demand Index
- Likelihood of CRP enrollment
- Land use change decisions – withdrawal installations
- Reductions in baseflow\(^1\)
- Decline in fish sustainability\(^2\)

---

Study Area

Branch County, Michigan

- Mainly Ag (65%)
- Well draining soils (85% B soils, 10% A)
- CLU data available
Fish Habitat Data

M-DNR:

- Tolerable baseflow reductions

Sensitive Fish Sustainability
Available GW Depletion (GPM)

- 107 - 243
- 244 - 515
- 516 - 1,887
- 1,888 - 3,140
- 3,141 - 10,507
Model Output / Dependent Variable

- Change in fish habitat sustainability over time (*Years To Stop*)
  - Reduction in baseflow
  - Change in stream fish habitat classification

40 model runs

Years until sensitive fish no longer supported at more than 75% of streams

Count

- 1
- 2 to 5
- 6 to 10
- 11 to 14
- 15 to 18
- 19 to 23
- > 23
Model Parameter Categories / Regression Independent Vars.

- **Crops**
  - area %
  - prices
  - price variability

- **CRP enrollment**
  - starting enrollment
  - probability of re-enrollment
  - contract length

- **Land cover change probabilities**
  - Given revenues of $X$, probability that a producer would convert Y to Z.

- **Decision thresholds**
  - revenue level above which producers will consider increased irrigation, below which they will consider CRP
Model Sensitivity Analysis

• Ran the model over 1,400 times with randomly selected parameter values

• Employed OLS regression
  - DV: Years until 75% of streams no longer support sensitive fish
  - IVs: model parameters

• Expectations

<table>
<thead>
<tr>
<th></th>
<th>Revenue threshold to move land into production</th>
<th>Revenue threshold to move land into conservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting corn prices</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Starting soy prices</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Corn area %</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Soy area %</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Crop price variability</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Soy price variability</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Corn yield per acre</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Soy yield per acre</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Ratio of market increase to CRP decrease: -
Starting % enrolled in CRP: +
CRP contract length: +
CRP renewal probability: +
Probability of conversion to pasture: +
Probability of conversion to forest: +
Probability of conversion to wetland: +
Model Sensitivity Analysis

• Identified best models through an exhaustive approach
  - 17 model parameters
  - max of 7 independent variables at a time
  - 41,226 regressions
  - sorted by R2, F-statistic, % of significant terms

• Is this rummaging?
  - Not trying to explore or discover variable relationships
  - The model is programmed to have relationships
  - Trying to identify weights of individual variables
Model Sensitivity Analysis

- Best OLS model

\[
\ln(\text{Years to stop}) = 2.23 - (0.173* \text{corn price}) - (0.141* \text{corn price variability}) - (0.009* \text{corn yield}) - (0.010* \text{soy yield}) + (0.002 * \text{land production revenue threshold})
\]

- \( R^2 \) 0.35
- \( F \)-statistic prob. < 0.001
- Sig. ind. vars all

- Standardized coefficients

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn price</td>
<td>-0.444</td>
</tr>
<tr>
<td>Soy yield</td>
<td>-0.328</td>
</tr>
<tr>
<td>Corn yield</td>
<td>-0.303</td>
</tr>
<tr>
<td>Land production revenue threshold</td>
<td>0.268</td>
</tr>
<tr>
<td>Corn price variability</td>
<td>-0.230</td>
</tr>
</tbody>
</table>
Model Sensitivity Analysis

Further inspection showed a poor fit

Why?

DV is a count, OLS won’t work
Model Sensitivity Analysis

- Employed a negative binomial regression, with a hurdle component as an alternative\(^3\),\(^4\).

- Useful for over-dispersed, skewed data with large zero counts.

- Function `hurdle()` from the `pscl` R package.

Hurdle component models zero values separately, essentially as a logit model.

Remainder modeled with negative binomial regression.

---

3. cran.r-project.org/web/packages/pscl/vignettes/countreg.pdf
Model Sensitivity Analysis

- Best negative binomial hurdle model (sorted by AIC)
  
  \[
  \text{Years to stop} = 7.72 - (0.731 \times \text{corn price}) - (0.657 \times \text{corn price variability}) \\
  - (0.038 \times \text{corn yield}) - (0.047 \times \text{soy yield}) \\
  + (0.009 \times \text{land production revenue threshold}) - \ln(\theta)
  \]

  \[
  \text{Years to stop} = 6.76 - (0.352 \times \text{corn price}) - (0.019 \times \text{corn yield}) - (0.020 \times \text{soy yield}) - \ln(\theta)
  \]

- Significant independent variables: All but \(\theta\)
- AIC: 74741

- Difficult to standardize coefficients
Model Sensitivity Analysis

- Hurdle coefficient standardization options
  - z-score ratios

<table>
<thead>
<tr>
<th>Model</th>
<th>Corn price</th>
<th>Soy yield</th>
<th>Corn yield</th>
<th>Corn price variability</th>
<th>Land production revenue threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>count model</td>
<td>-0.349</td>
<td>-0.298</td>
<td>-0.260</td>
<td>-0.229</td>
<td>0.137</td>
</tr>
<tr>
<td>zero model</td>
<td>-0.386</td>
<td>-0.309</td>
<td>-0.305</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- hierarchical partitioning
  - Murray and Connor 2009
  - R package *hier.part*
Model Sensitivity Analysis

Further inspection showed hurdle was still a poor fit

Why?

Still struggling with skewness.

Transformation of DV makes it no longer a count.
Regression can be utilized to estimate parameter weights in complex spatial model.

Issues arise when the dependent variable is count data
- Poisson and negative binomial regression are viable alternatives for over-dispersed data
- hurdle models for large zero counts

Dependent variable skewness is significant challenge
- normally distributed continuous data is preferable
- not always feasible for agent-based models based on steps

The example fish habitat sustainability model was most sensitive to market-based parameters (corn price, price variability, production revenue thresholds).
**Special Thanks:**

Dr. Jon Bartholic (Institute of Water Research - Michigan State University)  
Brad Love (Branch County Conservation District)  
Robert Pigg (Michigan Department of Agriculture)

**References:**

