
Benjamin M. Gramig1, Christopher A. Wolf2, and Frank Lupi1,2
1 Department of Agricultural Economics, Michigan State University, 2 Department of Fisheries and Wildlife, Michigan State University; send paper requests to gramigbe@msu.edu

Research Problem
- Veterinary epidemiologists routinely analyze data linking livestock serological tests to farm management and biosecurity practices to identify "risk factors" for infection.
- The key results from statistical analysis of such data are reported as odds-ratios interpreted as the "relative risk" of infection associated with an individual practice.
- When herd-level data are analyzed this way, information available to a manager is limited to the change in the likelihood of the herd being infected, which may be one or more animals is infected. Although this may be helpful for managers with contagious diseases, for other diseases managers might want to know how practices affect prevalence rates.
- For diseases that are widespread, such as Bovine Leukosis Virus (BLV), managers may not find eradication cost effective, but instead may want to know how control practices reduce prevalence rates and economic losses.
- We provide an example of statistical analysis capable of predicting herd-level disease prevalence for BLV and use this in an analysis of adoption of BLV control practices.

Stage I: Estimate fractional response: Disease prevalence = fn (BLV controls, etc.)
- Fractional logit was used to directly estimate the logistic outcome "disease prevalence" per farm which falls in the unit interval (0,1).
- Fractional logit was compared to standard logit of disease presence/absence per farm which doesn't use all information (ie ignores prevalence rates). The box on far right provides more information on fractional logit.

Stage II: Estimate BLV Control Practice Adoption: Adoption = fn (costs avoided, etc.)
- Construct cost avoided variable for each farm by combining (1) Economic loss per % point of BLV per cow per year from veterinary literature, and (2) Predicted Stage I marginal effect of BLV reducing practices.
- Resulting variable = losses avoided by adopting control practice
- Adoption is estimated using standard logit to determine the influence of potential economic damages on adoption

Table 1. Weighted NAHMS Dairy 1996 Sample Statistics for BLV

<table>
<thead>
<tr>
<th>Variable</th>
<th>BLV (hard size)</th>
<th>p-value from rank sum test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herd Roll Avg Milk Production (gal/yr)</td>
<td>Positive</td>
<td>17342.02 17560.00</td>
</tr>
<tr>
<td>Negative</td>
<td>18023.01 18000.00</td>
<td>0.5513</td>
</tr>
<tr>
<td>Entire sample</td>
<td>17429.26 17800.00</td>
<td>0.5513</td>
</tr>
<tr>
<td>Cost per COW</td>
<td>Positive</td>
<td>7.0300 8.0000</td>
</tr>
<tr>
<td>Negative</td>
<td>7.0406 7.0830</td>
<td>0.1893</td>
</tr>
<tr>
<td>Entire sample</td>
<td>6.0705 8.0000</td>
<td>0.1893</td>
</tr>
<tr>
<td>BLV prevalence</td>
<td>Positive</td>
<td>0.0489 0.0030</td>
</tr>
<tr>
<td>Negative</td>
<td>0 0</td>
<td>0.000000</td>
</tr>
<tr>
<td>Entire sample</td>
<td>0.0489 0.0030</td>
<td>0.4039707 0.4</td>
</tr>
</tbody>
</table>

Note: Asterisk denotes level of statistical significance: *10%, **5%, ***<1%.

Data
- USDA-APHIS maintains the National Animal Health Monitoring System (NAHMS) which contains extensive farm-level behavioral information from a complex random stratified sample that includes:
  - Animal inventory and operational characteristics
  - Health management and biosecurity practices
  - Feed/flock feeding and animal management
  - Morbidity, mortality, and culling
  - Some cost of production information

Data for Bovine Leukosis Virus (BLV) from the 1996 NAHMS dairy survey are used in this study. Sample statistics are found in Table 1.

Table 2. Estimation results of fractional logit (Rogit) for BLV prevalence rate in a herd and binary logit for presence/absence of BLV in a herd

| Variable | Rogit: E(y|x) | Deprate (BLV: prevalence [marginal effect] | logit: E(y|x) = p = 0.1 (BLV infection rate [marginal effect]) |
|----------|---------------|------------------------------------------|------------------------------------------|
| Location region | 0.000000 | 0.000000 | 0.000000 |
| Total Cows (thousands) | 0.000000 | 0.000000 | 0.000000 |
| Rolling Avg Milk Prod (thousands gal/yr) | 0.000000 | 0.000000 | 0.000000 |
| Safe dehorning practice | 0.000000 | 0.000000 | 0.000000 |
| Clean Injections to Cows | 0.000000 | 0.000000 | 0.000000 |
| Clean Injections to Heifers | 0.000000 | 0.000000 | 0.000000 |
| Herd | 0.000000 | 0.000000 | 0.000000 |

Notes: n=993. Only statistically significant variables reported. Asterisk denotes level of statistical significance: *10%, **5%, ***<1%

Figure 1. Predicted Probability of Adoption and Estimated Economic Damages

- Figure 1 plots predicted probability of adoption against estimated economic damages from not adopting a "safe" dehorning practice using Stage II estimation results.

Conclusions
- Fractional logit regression can provide valuable prevalence information about diseases where it may not be economically optimal to eradicate infection.
- NAHMS provides a wealth of data on disease prevalence rates that are well suited to analysis using the fractional logit method.

References
- For our paper containing full details, contact gramigbe@msu.edu

Fractional Logit Method
We illustrate fractional logit estimation (Papke and Wooldridge, 1996, Wooldridge 2002) for analysis of dependent variables with fractional or propositional values in the unit interval, i.e., y ∈ [0, 1].

Examples of fractional response variables encountered by agricultural economists include:
- Livestock herd disease prevalence rates
- Proportion of farmers using in production of a particular crop
- Proportion of land in a city, county, or other jurisdiction in a specific use (ag, commercial, urban, residential, forest, etc.)

Fractional logit overcomes several problems encountered when other methods are used with fractional response variables:
- OLS cannot guarantee predictions fall within unit interval
- Log-ODds regression requires arbitrary adjustment for all methods used for fractional data
- MLK not robust to distributional failure
- Nonlinear LS requires Var(x|y) [x2] for relative efficiency which is unlikely for fractional y

Fractional logit estimation improves upon previously used statistical methods because it only requires that the conditional mean be specified correctly to obtain consistent parameter estimates and it allows for direct estimation of fractional response variables.

Fractional logit is a quasi-MLE method with conditional mean assumption

E(y|x) = exp(x′β)/(1+exp(x′β)) = y | x

The quasi-log-likelihood for observation i is exactly the same as for the logistic regression model

(1) | y i | log(1+exp(x′β)) - (1) | y i | log(1+exp(x′β)) |

where (x′β) is the logistic CDF and y ∈ [0,1] (which differs from fractional logit when limits for values of y ∈ [0,1]). The mechanism of obtaining the parameter estimates is identical to the binary response case, but a fully robust variance estimator should be described. This bivariate CDF variance estimation required for generally robust inference

Var(x|y) = (x′β)[1-2y(x′β)]

This procedure can be programmed directly or implemented using a statistical package that does not have the capability of doing fractional variables as binary. When a bivariate normal distribution is used with the logit link function is specified in estimating generalized linear models.

For our paper containing full details, contact gramigbe@msu.edu