Estimation for County Level Cropland Cash Rental Rates

Emily Berg\textsuperscript{1}, Will Cecere\textsuperscript{2}, Malay Ghosh\textsuperscript{3}

September 2, 2013

\textsuperscript{1}Iowa State University, \textsuperscript{2}Westat, \textsuperscript{3}University of Florida

The United States Department of Agriculture National Agricultural Statistics Service (NASS) funded this research.
Outline

- Cropland cash rental rates
  - NASS annual Cash Rent Survey
  - Objectives for model based estimation
  - Auxiliary information
- General bivariate model for two time means
  - Decomposition as average (level) and difference (change)
  - Add predictor of average to one half predictor of difference
- Specific procedures for Cash Rent Survey
  - Estimation of differences
  - Estimation of sampling variances
  - Definitions of covariates
- Results for 2011
- Simulation
- Discussion
Cash Rental Rates

- Cropland or pastureland rented in $/acre
  - Useful to producers, policy-makers, researchers
  - Farm Service Agency uses cash rental rates for guidance in calculating rates for the Conservation Reserve Program

- 2008 Farm Bill requires annual cash rent survey
  - Land uses: irrigated cropland, nonirrigated cropland, pasture
  - Counties with at least 20,000 acres in cropland or pastureland

- NASS Annual Cash Rent Survey (2009-2012)
  - Stratified sample
  - ≈ 224,000 operations each year
  - Direct estimators weighted sums
    \[ w \propto [(\text{select. prob.})(\text{resp. prob.)}]^{-1} \]
  - Jackknife variance estimator
Cash Rental Rates: Cash Rent Survey

Direct estimators for 2011 and 2010 correlated

Wide range in county sample sizes and estimated CVs
Objectives
- Efficient estimates of average cash rental rates at the county level
  - Irrigated, nonirrigated, permanent pasture
  - Counties with $\geq 20,000$ acres of cropland or pastureland
- Mean squared error estimators
- Computational simplicity

Data
- Survey data for 2 years $t - 1, t$ (eg. 2010, 2011)
  - Focus on South Dakota (SD) and Florida (FL) for this presentation
- Auxiliary data
Cash Rental Rates: Auxiliary Data

- 2007 Census of Agriculture
  - Total value of agricultural production in a county
- NASS published county yields from 2005-2009
  - Yield indexes: irrigated, nonirrigated, total, hay
- National Commodity Crop Productivity Indexes (NCCPI)
  (Developed by the Natural Resources Conservation Service)
  - Three indexes: corn, cotton, wheat
    - Reflect the productivity of the soil for growing nonirrigated crops in different climates

- Covariates constant across two consecutive years
  - Information about level
  - Only provide information about change to the extent that level and change are correlated
Bivariate Model: Relationships between Level and Change

- Average cash rent/acre for time $t$ is a sum of the average and half of the difference:

$$\theta_{i,t} = \text{true avg. cash rent/acre, county } i \text{ time } t$$

$$\theta_{i,t} = \theta_i + 0.5\Delta_i$$

$$\theta_i = 0.5(\theta_{i,t-1} + \theta_{i,t}), \Delta_i = \theta_{i,t} - \theta_{i,t-1}$$

- We can construct a predictor of a time $t$ mean by adding a predictor of an average to half of a predictor of a difference.
Bivariate Model: Relationships between Level and Change

- $\hat{y}_{i,t} = \text{direct est. of } \theta_{i,t}$
- $V\{\hat{y}_{i,t}\} = V\{\hat{y}_{i,t-1}\}$ implies $\hat{y}_{i,t} - \hat{y}_{i,t-1}$ and $0.5(\hat{y}_{i,t} + \hat{y}_{i,t-1})$ are uncorrelated

<table>
<thead>
<tr>
<th>State</th>
<th>Use</th>
<th>$C{n_{11}, n_{10}}$</th>
<th>$C{(\text{Avg}, \text{Diff})}$</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>Nir</td>
<td>0.94</td>
<td>0.37</td>
<td>3.18</td>
</tr>
<tr>
<td>FL</td>
<td>Nir</td>
<td>0.96</td>
<td>-0.14</td>
<td>-0.83</td>
</tr>
<tr>
<td>SD</td>
<td>Pas</td>
<td>0.79</td>
<td>0.17</td>
<td>1.34</td>
</tr>
<tr>
<td>FL</td>
<td>Pas</td>
<td>0.90</td>
<td>0.08</td>
<td>0.63</td>
</tr>
<tr>
<td>SD</td>
<td>Irr</td>
<td>0.86</td>
<td>0.41</td>
<td>2.33</td>
</tr>
<tr>
<td>FL</td>
<td>Irr</td>
<td>0.85</td>
<td>0.24</td>
<td>1.35</td>
</tr>
</tbody>
</table>

$T = \text{Pitman-Morgan test statistic of } H_o : \text{Cor}\{(\text{Avg}, \text{Diff})\} = 0.$

- A working assumption of constant variance for two time points appears reasonable for Cash Rent Survey data.
Bivariate Model: Decomposition as Level and Change

- Univariate models for average and difference (Fay and Herriot, 1979)

**Average**

\[
\hat{y}_i = 0.5(\hat{y}_{i,t} + \hat{y}_{i,t-1})
\]

\[
= \theta_i + e_i, \quad \theta_i = x_i'\beta + u_i
\]

\[
(u_i, e_i)' \sim [0, \text{diag}(\sigma^2_u, \sigma^2_{ei,avg})]
\]

**Difference**

\[
\hat{d}_i = \text{direct estimate of diff.}
\]

\[
= \Delta_i + \eta_i, \quad \Delta_i = z_i'\beta_d + v_i
\]

\[
(v_i, \eta_i)' \sim [0, \text{diag}(\sigma^2_v, \sigma^2_{\eta_i,diff})]
\]

- Assume estimates of \(\sigma^2_{ei,avg}\) and \(\sigma^2_{\eta_i,diff}\) are available.
Bivariate Model: Predictors

<table>
<thead>
<tr>
<th>Average</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\theta}_i = \hat{\gamma}_i \hat{y}_i + (1 - \hat{\gamma}_i) x'_i \hat{\beta} )</td>
<td>( \hat{\Delta}_i = \hat{\lambda}_i d_i + (1 - \hat{\lambda}_i) z'_i \hat{\beta}_d )</td>
</tr>
<tr>
<td>( \hat{\gamma}_i = \hat{\sigma}<em>u^2 (\hat{\sigma}<em>u^2 + \hat{\sigma}</em>{e</em>{i,avg}}^2)^{-1} )</td>
<td>( \hat{\lambda}_i = \hat{\sigma}<em>v^2 (\hat{\sigma}<em>v^2 + \hat{\sigma}</em>{\eta</em>{i,diff}}^2)^{-1} )</td>
</tr>
</tbody>
</table>

**Time t Mean**

\( \hat{\theta}_{i,t} = \hat{\theta}_i + 0.5 \hat{\Delta}_i \)

- Estimated generalized least squares estimators of model parameters
  - Modification of Wang, Fuller, and Qu (2008)
  - Positive estimator of \( \sigma_u^2 \) and \( \sigma_v^2 \)

**Working assumption:** \( C\{u_i, v_i\} = 0 \) and \( C\{e_i, \eta_i\} = 0 \)

- Estimator of optimal linear predictor if errors in average and difference are uncorrelated
- Unbiased but possibly inefficient if working model not true
Bivariate Model: MSE Estimators

- Working assumption that average and difference uncorrelated
  
  \[ \hat{MSE}_{i,w} = \hat{MSE}(\hat{\theta}_i) + 0.25\hat{MSE}(\hat{\Delta}_i) \]
  
  \[ \hat{MSE}(\hat{\theta}_i) = \hat{\gamma}_i\hat{\sigma}_{avg}^2 + \hat{g}_{2i,a} + 2\hat{g}_{3i,a} \]
  
  \[ \hat{MSE}(\hat{\Delta}_i) = \hat{\lambda}_i\hat{\sigma}_{diff}^2 + \hat{g}_{2i,d} + 2\hat{g}_{3i,d} \]

- \( \hat{g}_{2i,a} \) and \( \hat{g}_{3i,a} \) for estimation of \( \beta \) and \( \sigma_u^2 \) (Prasad and Rao, 1990)

- \( \hat{MSE}_{i,w} \) biased if average and difference correlated

- Relax assumption that sampling errors in avg. and diff. uncorrelated

  \[ \hat{MSE}_{i,t} = \hat{g}_{1it,cor} + \hat{g}_{2i,a} + 2\hat{g}_{3i,a} + 0.25(\hat{g}_{2i,d} + 2\hat{g}_{3i,d}) \]
  
  \[ \hat{g}_{1it,cor} = (1 - \hat{\gamma}_{i,avg})^2\hat{\sigma}_u^2 + 0.25(1 - \hat{\gamma}_{i,diff})^2\hat{\sigma}_v^2 \]
  
  \[ + (1, 0.5)\text{diag}(\hat{\gamma}_{i,avg}, \hat{\gamma}_{i,diff})\hat{\Sigma}_{e\eta,i}\text{diag}(\hat{\gamma}_{i,avg}, \hat{\gamma}_{i,diff})(1, 0.5)' \]

  \[ \hat{\Sigma}_{e\eta,i} = \hat{\text{Cov}}\{(e_i, \eta_i)'\} \]
Specific Procedures for Cash Rent Survey

1. Estimation of differences
2. Estimation of sampling variances
3. Definitions of covariates
Unit Level Cash Rental Rates

- Responses for units in both survey years are correlated
- Variance of difference decreases as acres increase
- Outliers relative to normality
  - Eg. FL irrigated, unit-level differences can exceed $3 \times \text{SD}$ of differences
Convex combination of $\hat{y}_{i,t} - \hat{y}_{i,t-1}$ and a weighted average of the difference for respondents in both time points

$$\hat{d}_i = \alpha_i \hat{y}_{di} + (1 - \alpha_i) \bar{y}_{di}$$

$$\hat{y}_{di} = \hat{y}_{i,t} - \hat{y}_{i,t-1},$$

$$\bar{y}_{di} = \left( \sum_{j=1}^{n_{it,t-1}} d_{ij} a_{ij} \right) \left( \sum_{j=1}^{n_{it,t-1}} a_{ij} \right)^{-1}$$

- $j = \text{unit in county } i$, $n_{it,t-1} = \text{number in both years}$
- $\alpha_i$ optimal if sample sizes and variances for two time points equal
- $d_{ij} = \text{unit-level difference after modification to outliers}$
- $a_{ij} = \text{average acres rented}$
Cash Rent Specifics: Estimation of Sampling Variances

**Jackknife variances**

- Undefined for sample size < 2
- Large variances for small sample sizes
  - Avg. 2010 sample sizes between 2 and 60 (roughly)
- Correlated with direct estimators of means
  - Correlation between jackknife s.d. and $\hat{y}_{i,2011}$ between 0.20 and 0.85

**Implication**

- Predictor undefined for sample size of 1
- Large prediction variance
- Biased GLS estimators
Cash Rent Specifics: Estimation of Sampling Variances

- Hierarchical model for sampling variances
  - \( n_{it} \) = sample size for county \( i \) and year \( t \), \( n_{it} > 1 \)
  - \( n_{it}^{-1} (\sigma^{2}_{ei,t}, S^{2}_{it}) \) = (true sampling variance, jackknife estimator)

\[
\frac{(n_{it} - 1)S^{2}_{it}}{\sigma^{2}_{ei,t}} \mid \sigma^{2}_{ei,t} \sim \chi^{2}_{(n_{it}-1)}, \quad \frac{1}{\sigma^{2}_{ei,t}} \overset{d}{=} \frac{1}{\sigma^{2}_{0it} \nu} X, \quad X \sim \chi^{2}_{\nu},
\]

\[
E[S^{2}_{it}] = \sigma^{2}_{0it} = (\mu^{2}_{it}) \alpha
\]

\[
(\mu_{it-1}, \mu_{it}) = (x'_{i} \beta - 0.5z_{i}\Delta_{i}, x'_{i} \beta + 0.5z_{i}\Delta_{i})
\]

- Method of moments estimators: \( \hat{\alpha}, \hat{\nu} \)
- Estimator of sampling variance for county \( i \) and year \( t \)

\[
\hat{V}_{ei,t} = n_{it}^{-1} E[\sigma^{2}_{ei,t} \mid \hat{\nu}, \hat{\alpha}, S^{2}_{it}, \hat{\mu}_{it}] = \frac{d^{*}_{it}}{d^{*}_{it} - 2} \sigma^{2*}_{it},
\]

\[
d^{*}_{it} = \hat{\nu} + n_{it}, \quad \sigma^{2*}_{it} = \frac{\hat{\nu}}{\hat{\nu} + n_{it}} \hat{\alpha} \hat{\mu}^{2}_{it} + \frac{n_{it}}{\hat{\nu} + n_{it}} S^{2}_{it}
\]
Correlations between $\hat{y}_i$ and Auxiliary Variables for SD

<table>
<thead>
<tr>
<th>Land Use</th>
<th>Total Value of Production</th>
<th>Yield Total</th>
<th>Yield Hay</th>
<th>NCCPI Corn</th>
<th>NCCPI Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonirrigated</td>
<td>0.66</td>
<td>0.93</td>
<td>0.88</td>
<td>0.89</td>
<td>0.24</td>
</tr>
<tr>
<td>Pasture</td>
<td>0.68</td>
<td>0.86</td>
<td>0.92</td>
<td>0.85</td>
<td>0.35</td>
</tr>
<tr>
<td>Irrigated</td>
<td>0.32</td>
<td>0.66</td>
<td>0.60</td>
<td>0.64</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Challenges

- Availability and nature of relationships vary by state and land use
- Colinearity among covariates
  - Eg. Cor(Hay Yield, Total Yield) for SD is 0.89
- Negative predicted values in the model for the average are possible.
Cash Rent Specifics: Definitions of Covariates

Univariate Covariate - Index of Productivity: $x_i^*$

- Linear combination of auxiliary variables
- Scaled to be strictly positive and have similar mean and variance as cash rental rate
- Little loss of information

For nonirrigated cropland for SD, $R^2$ of model with five covariates is 0.89, and $R^2$ of model with univariate covariate index is 0.83.
Cash Rent Specifics: Definitions of Covariates

**Average vs. Index of Productivity**

SD Nonirrigated

- Segmentated regression for average

\[ x_i = (1, x_i^*, x_i^*, 1, x_i^*, 2) \]

\[ x_{i,1} = (x_i^* - x_{(m/3)})I[x_i^* > x_{(m/3)}] \]

\[ x_{i,2} = (x_i^* - x_{(2m/3)})I[x_i^* > x_{(2m/3)}] \]

- Coefficient \( \beta = (\beta_0, \beta_1, \beta_2, \beta_3) \)
- Restrict estimates to ensure positive predicted value
  - \( \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 < 0, \rightarrow \text{set } \hat{\beta}_3 = 0 \)
  - \( \hat{\beta}_1 + \hat{\beta}_2 < 0, \rightarrow \hat{\beta}_2 = 0 \)

- Covariate for difference

\[ z_i = (1, x_i^*) \]
Results for 2011 Cash Rent Survey

- Separate models for different states and land uses
- Benchmark to state level estimates (Ghosh and Steorts, 2012)

- Compare to predictor based on univariate area-level model for one time point

<table>
<thead>
<tr>
<th></th>
<th>Nonirrigated</th>
<th>Pasture</th>
<th>Irrigated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\theta}_{biv}$</td>
<td>$\hat{\theta}_{uni}$</td>
<td>$\hat{\theta}_{biv}$</td>
</tr>
<tr>
<td>SD</td>
<td>0.62</td>
<td>0.89</td>
<td>0.60</td>
</tr>
<tr>
<td>FL</td>
<td>0.86</td>
<td>0.95</td>
<td>0.70</td>
</tr>
</tbody>
</table>

- Medians of ratios of estimated MSEs of bivariate (biv) and univariate (uni) predictors to estimated variances of direct estimators.
Medians of estimated CVs (percent) for bivariate predictors

<table>
<thead>
<tr>
<th>State</th>
<th>Nonirrigated</th>
<th>Irrigated</th>
<th>Pasture</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>5.11</td>
<td>6.27</td>
<td>11.26</td>
</tr>
<tr>
<td>FL</td>
<td>11.40</td>
<td>17.80</td>
<td>20.28</td>
</tr>
</tbody>
</table>
Results for 2011 Cash Rent Survey

- Standardized residuals for SD nonirrigated cropland

\[ r_{i,avg} = \frac{\hat{y}_{i,avg} - \mathbf{x}'_{i}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}_{u}^2 + \hat{\sigma}_{ei,avg}^2}}, \quad r_{i,diff} = \frac{\hat{d}_{i} - \mathbf{z}'_{i}\hat{\boldsymbol{\beta}}_{d}}{\sqrt{\hat{\sigma}_{v}^2 + \hat{\sigma}_{\eta_{i,diff}}^2}} \]
Simulation

• Population means

$$(\theta_{i1}, \theta_{i2})' \sim \mathcal{N} \left( (\mu_{i1}, \mu_{i2})', \Sigma_{uu} \right), \quad \Sigma_{uu} = \sigma_u^2 \begin{pmatrix} 1 & \rho_u \\ \rho_u & 1 \end{pmatrix}$$

• Unit-level data

$$(y_{ij1}, y_{ij2})' \sim \mathcal{N} \left( (\theta_{i1}, \theta_{i2})', \Sigma_{ee} \right), \quad j = 1, \ldots, n_{i12}$$

$$y_{ij1} \sim \mathcal{N}(\theta_{i1}, \sigma_{ei1}^2), \quad j = n_{i12} + 1, \ldots, n_{i12} + n_{i11}$$

$$y_{ij2} \sim \mathcal{N}(\theta_{i2}, \sigma_{ei2}^2), \quad j = n_{i12} + n_{i11} + 1, \ldots, n_{i22}$$

$$\Sigma_{eii} = \text{diag}(\sigma_{ei1}, \sigma_{ei2}) \begin{pmatrix} 1 & \rho_e \\ \rho_e & 1 \end{pmatrix} \text{diag}(\sigma_{ei1}, \sigma_{ei2}).$$

• Parameter values based on SD nonirrigated cropland

• Model 1: $\sigma_{ei1} = \sigma_{ei2}, \quad n_{i11} = n_{i22}$

• Model 2: $\sigma_{ei1} \neq \sigma_{ei2}, \quad n_{i11} \neq n_{i22}$
Simulation

- **Four predictors**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Definition</th>
<th>Covariates</th>
<th>Bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{y}_{i,2}$</td>
<td>sample mean for $t = 2$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>$\hat{\theta}_{i,ad}$</td>
<td>$\hat{y}_i + 0.5\hat{d}_i$</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>$\hat{\theta}_{i,2}$</td>
<td>area-level model for $t = 2$</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>$\hat{\theta}_{i,biv}$</td>
<td>proposed predictor</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

$\hat{\theta}_{i,ad}$ illustrates gain (if any) due to estimating change as a weighted average of the difference between the direct estimators and the weighted average of the differences for paired observations.
Simulation

- Model 1 - equal sampling error variances for two time points
- Model 2 - unequal sampling error variances for two time points

\[
\begin{align*}
\hat{y}_{i2} & \quad \hat{\theta}_{i,ad} & \quad \hat{\theta}_{i2} \\
\hline
\text{Model 1} & 1.17 & 1.05 & 1.12 \\
\text{Model 2} & 1.28 & 1.14 & 1.14
\end{align*}
\]

- Medians of ratios of Monte Carlo (MC) MSEs of alternative predictors to MC MSEs of \( \hat{\theta}_{i,biv} \)
Simulation

- MC properties of MSE estimators
  - $\widehat{MSE}_{i,w}$: working model that sampling variances for two time points equal
  - $\widehat{MSE}_{i,2}$: allows unequal sampling variances in leading term

<table>
<thead>
<tr>
<th>MSE Estimator</th>
<th>Model 1 (Equal var.)</th>
<th>Model 2 (Unequal var.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rel. Bias</td>
<td>Cov. of 95% CI</td>
</tr>
<tr>
<td>$\widehat{MSE}_{i,w}$</td>
<td>-1.13</td>
<td>0.95</td>
</tr>
<tr>
<td>$\widehat{MSE}_{i,2}$</td>
<td>-0.22</td>
<td>0.95</td>
</tr>
</tbody>
</table>

- Medians of MC relative biases of MSE estimators
- MC coverages of normal theory confidence intervals (CI) with nominal coverage 0.95
Discussion: Summary

- Combine predictors based on separate univariate area-level models for averages and differences
- Incorporates data from previous year and covariates
- Computationally simple
- Incorporates survey weights
- Does not rely on normality of unit-level data
Discussion: Possible Improvements

- Explore data for additional sources of structure (e.g., spatial relationships, more than two time points, correlations between land uses)

- Integrate estimators of differences and sampling variances into the statistical model

- Formalize covariate selection

- Account for estimation of sampling variances, adjustments to outliers, and benchmarking in MSE estimation

- Directly consider nonsampling errors (e.g., nonresponse, definitions of pasture, effects of arms-length transactions)
The National Agricultural Statistics Service supported this research.
The authors thank the Center for Survey Statistics and Methodology and Iowa State University and Wayne A. Fuller.

References