Estimation for County Level Cropland Cash Rental Rates

Emily Berg[1], Will Cecere[2], Malay Ghosh[3]

September 2, 2013

[1]Iowa State University, [2]Westat, [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

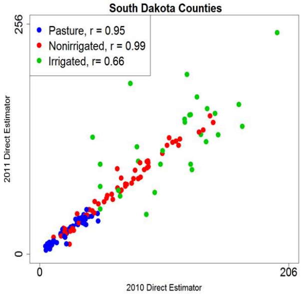
- 2008 Farm Bill requires annual cash rent survey
  - Land uses: irrigated cropland, nonirrigated cropland, pasture
  - Counties with at least 20,000 acres in cropland or pastureland

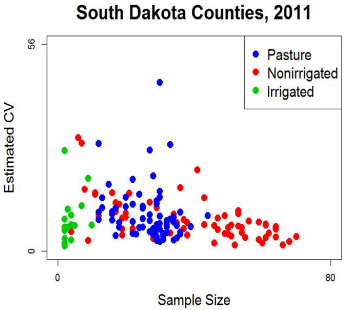
- NASS Annual Cash Rent Survey (2009-2012)
  - Stratified sample
  - $\approx$ 224,000 operations each year
  - Direct estimators weighted sums

$$\text{w} \;\propto \left[(\text{select. prob.})(\text{resp. prob.})\right]^{-1}$$

  - 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

Cash Rental Rates: Model Based Estimation

- 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

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

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

T $=$ Pitman-Morgan test statistic of H_o : Cor{(Avg,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

$$\widehat{y}_i = 0.5(\widehat{y}_{i,t} + \widehat{y}_{i,t-1})$$
$$= \theta_i + e_i, \theta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i$$
$$(u_i, e_i)' \sim [\mathbf{0}, \text{diag}(\sigma_u^2, \sigma_{ei,avg}^2)]$$

Difference

$$\widehat{d}_i = \text{ direct estimate of diff.}$$
$$= \Delta_i + \eta_i, \Delta_i = \boldsymbol{z}_i'\boldsymbol{\beta}_d + v_i$$
$$(v_i, \eta_i)' \sim [\mathbf{0}, \text{diag}(\sigma_v^2, \sigma_{\eta i,diff}^2)]$$

- Assume estimates of $\sigma_{ei,avg}^2$ and $\sigma_{\eta i,diff}^2$ are available.

Bivariate Model: Predictors

Average

$$\widehat{\theta}_i = \widehat{\gamma}_i \widehat{y}_i + (1 - \widehat{\gamma}_i)\boldsymbol{x}_i' \widehat{\boldsymbol{\beta}}$$

$$\widehat{\gamma}_i = \widehat{\sigma}_u^2 (\widehat{\sigma}_u^2 + \widehat{\sigma}_{ei,avg}^2)^{-1}$$

Difference

$$\widehat{\Delta}_i = \widehat{\lambda}_i \widehat{d}_i + (1 - \widehat{\lambda}_i)\boldsymbol{z}_i' \widehat{\boldsymbol{\beta}}_d$$

$$\widehat{\lambda}_i = \widehat{\sigma}_v^2 (\widehat{\sigma}_v^2 + \widehat{\sigma}_{\eta i,diff}^2)^{-1}$$

Time t Mean

$$\widehat{\theta}_{i,t} = \widehat{\theta}_i + 0.5\widehat{\Delta}_i$$

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

$$\widehat{MSE}_{i,w} = \widehat{MSE}(\widehat{\theta}_i) + 0.25\widehat{MSE}(\widehat{\Delta}_i)$$
$$\widehat{MSE}(\widehat{\theta}_i) = \widehat{\gamma}_i\widehat{\sigma}^2_{ei,avg} + \widehat{g}_{2i,a} + 2\widehat{g}_{3i,a}$$
$$\widehat{MSE}(\widehat{\Delta}_i) = \widehat{\lambda}_i\widehat{\sigma}^2_{\eta i,diff} + \widehat{g}_{2i,d} + 2\widehat{g}_{3i,d}$$

  - $\widehat{g}_{2i,a}$ and $\widehat{g}_{3i,a}$ for estimation of $\boldsymbol{\beta}$ and σ_u^2 (Prasad and Rao, 1990)
  - $\widehat{MSE}_{i,w}$ biased if average and difference correlated

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

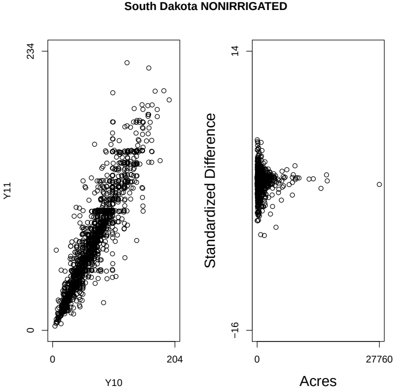
$$\widehat{MSE}_{i,t} = \widehat{g}_{1it,cor} + \widehat{g}_{2i,a} + 2\widehat{g}_{3i,a} + 0.25(\widehat{g}_{2i,d} + 2\widehat{g}_{3i,d})$$
$$\widehat{g}_{1it,cor} = (1-\widehat{\gamma}_{i,avg})^2\widehat{\sigma}^2_u + 0.25(1-\widehat{\gamma}_{i,diff})^2\widehat{\sigma}^2_v$$
$$+ (1, 0.5)\text{diag}(\widehat{\gamma}_{i,avg},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

Specific Procedures for Cash Rent Survey

1. Estimation of differences
2. Estimation of sampling variances
3. Definitions of covariates

Cash Rent Specifics: Estimation of Differences

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 × SD of differences

Cash Rent Specifics: Estimation of Differences

- Convex combination of $\widehat{y}_{i,t} - \widehat{y}_{i,t-1}$ and a weighted average of the difference for respondents in both time points

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

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

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

- $j =$unit in county i, $n_{it,t-1} =$ number in both years
- α_i optimal if sample sizes and variances for two time points equal
- $d_{ij} =$ unit-level difference after modification to outliers
- $a_{ij} =$ 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 $\widehat{y}_{i,2011}$ 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)

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

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

$$(\mu_{it-1}, \mu_{it}) = (\boldsymbol{x}_i'\boldsymbol{\beta} - 0.5\boldsymbol{z}_i\Delta_i, \boldsymbol{x}_i'\boldsymbol{\beta} + 0.5\boldsymbol{z}_i\Delta_i)$$

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

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

$$d^*_{it} = \widehat{\nu} + n_{it}, \quad \sigma^{2*}_{it} = \frac{\widehat{\nu}}{\widehat{\nu} + n_{it}}\widehat{\alpha}\widehat{\mu}^2_{it} + \frac{n_{it}}{\widehat{\nu} + n_{it}} S^2_{it}$$

Cash Rent Specifics: Definitions of Covariates

- Correlations between $\widehat{y_i}$ and Auxiliary Variables for SD

Land Use	Total Value of Production	Yield		NCCPI	
		Total	Hay	Corn	Wheat
Nonirrigated	0.66	0.93	0.88	0.89	0.24
Pasture	0.68	0.86	0.92	0.85	0.35
Irrigated	0.32	0.66	0.60	0.64	0.16

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.

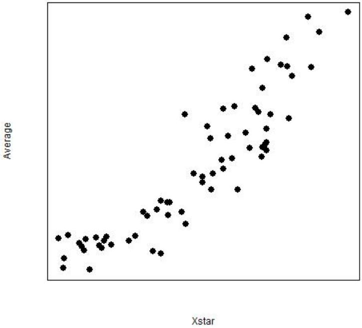
Cash Rent Specifics: Definitions of Covariates

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

Avgerage vs. Index of Productivity
SD Nonirrigated



- Nonlinear relationship

- Segmented regression for average

$$\boldsymbol{x}_i = (1, x_i^*, x_{i,1}^*, x_{i,2}^*)$$
$$x_{i,1}^* = (x_i^* - x_{(m/3)}^*) I[x_i^* > x_{(m/3)}^*]$$
$$x_{i,2}^* = (x_i^* - x_{(2m/3)}^*) I[x_i^* > x_{(2m/3)}^*]$$

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

- Covariate for difference

$$\boldsymbol{z}_i = (1, x_i^*)$$

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

	Nonirrigated		Pasture		Irrigated	
	$\widehat{\theta}_{biv}$	$\widehat{\theta}_{uni}$	$\widehat{\theta}_{biv}$	$\widehat{\theta}_{uni}$	$\widehat{\theta}_{biv}$	$\widehat{\theta}_{uni}$
SD	0.62	0.89	0.60	0.82	0.58	0.85
FL	0.86	0.95	0.70	0.83	0.61	0.86

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

Results for 2011 Cash Rent Survey

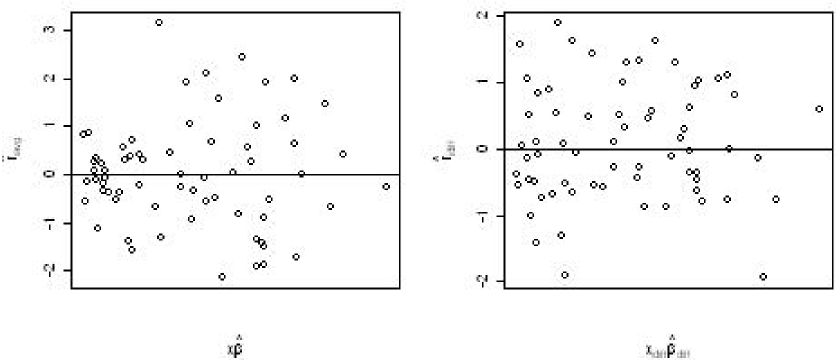
- Medians of estimated CVs (percent) for bivariate predictors

State	Nonirrigated	Irrigated	Pasture
SD	5.11	6.27	11.26
FL	11.40	17.80	20.28

Results for 2011 Cash Rent Survey

- Standardized residuals for SD nonirrigated cropland

$$r_{i,avg} = \frac{\widehat{y}_{i,avg} - \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}}}{\sqrt{\widehat{\sigma}_u^2 + \widehat{\sigma}_{ei,avg}^2}}, \quad r_{i,diff} = \frac{\widehat{d}_i - \boldsymbol{z}_i' \widehat{\boldsymbol{\beta}}_d}{\sqrt{\widehat{\sigma}_v^2 + \widehat{\sigma}_{\eta i,diff}^2}}$$



Simulation

- Population means

$$(\theta_{i1}, \theta_{i2})' \sim \mathsf{N}\left[(\mu_{i1}, \mu_{i2})', \mathbf{\Sigma}_{uu}\right], \quad \mathbf{\Sigma}_{uu} = \sigma_u^2 \begin{pmatrix} 1 & \rho_u \\ \rho_u & 1 \end{pmatrix}$$

- Unit-level data

$$\begin{aligned}
(y_{ij1}, y_{ij2})' &\sim \mathsf{N}\left[(\theta_{i1}, \theta_{i2})', \mathbf{\Sigma}_{ee}\right], \quad j = 1, \ldots, n_{i12} \\
y_{ij1} &\sim \mathsf{N}(\theta_{i1}, \sigma_{ei}^2), j = n_{i12} + 1, \ldots, n_{i12} + n_{i11} \\
y_{ij2} &\sim \mathsf{N}(\theta_{i2}, \sigma_{ei}^2), j = n_{i12} + n_{i11} + 1, \ldots, n_{i22} \\
\mathbf{\Sigma}_{eei} &= \mathsf{diag}(\sigma_{ei,1}, \sigma_{ei,2}) \begin{pmatrix} 1 & \rho_e \\ \rho_e & 1 \end{pmatrix} \mathsf{diag}(\sigma_{ei,1}, \sigma_{ei,2}).
\end{aligned}$$

- Parameter values based on SD nonirrigated cropland
- Model 1: $\sigma_{ei,1} = \sigma_{ei,2}$, $n_{i11} = n_{i22}$
- Model 2: $\sigma_{ei,1} \neq \sigma_{ei,2}$, $n_{i11} \neq n_{i22}$

Simulation

- Four predictors

Predictor	Definition	Covariates	Bivariate
$\widehat{y}_{i,2}$	sample mean for $t = 2$	No	No
$\widehat{\theta}_{i,ad}$	$\widehat{y}_i + 0.5\widehat{d}_i$	No	Yes
$\widehat{\theta}_{i,2}$	area-level model for $t = 2$	Yes	No
$\widehat{\theta}_{i,biv}$	proposed predictor	Yes	Yes

- $\widehat{\theta}_{i,ad}$ 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

	$\widehat{y}_{i2}$	$\widehat{\theta}_{i,ad}$	$\widehat{\theta}_{i2}$
Model 1	1.17	1.05	1.12
Model 2	1.28	1.14	1.14

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

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

MSE	Model 1 (Equal var.)		Model 2 (Unequal var.)	
Estimator	Rel. Bias	Cov. of 95% CI	Rel. Bias	Cov. of 95% CI
$\widehat{MSE}_{i,w}$	-1.13	0.95	-8.30	0.94
$\widehat{MSE}_{i,2}$	-0.22	0.95	-2.97	0.94

- 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 (eg., 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 (eg., nonresponse, definitions of pasture, effects of arms-length transactiions)

Thank You

- 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

  - Fay, R. and Herriot, R. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *JASA* 74, 341-53.
  - Ghosh, M. and Steorts, R. (2012). Two-stage benchmarking as applied to small area estimation. Submitted.
  - Prasad, N.G.N., and Rao, J.N.K. (1990). The estimation of mean squared errors of small area estimators, *JASA*, 85, 163-171.
  - Wang, J., Fuller, W.A. and Qu, Y. (2008). Small area estimation under a restriction. *Survey Methodology*, 34, 1, 29-36.