ESTIMATING IRRIGATED ACRES WITH MISSING DATA

Irfan Y. Tareen[1], Charles E. Rose[2] and Jimmy R. Bramblett[3]

AUTHORS:[1] Graduate Research Assistant, Department of Agricultural and Applied Economics, The University of Georgia, Athens, Ga. 30602. [2] Graduate Research Assistant, Department of Forest Biometrics, The University of Georgia, Athens, Ga. 30602. [3] Adjunct Faculty, Department of Agricultural and Applied Economics, University of Georgia, Athens, Ga. 30602
REFERENCE: Proceedings of the 2001 Georgia Water Resources Conference, held March 26-27, 2001, at the University of Georgia, Kathryn J. Hatcher, Editor, Institute of Ecology, The University of Georgia, Athens, Ga. 30602

Abstract. Irrigation for agriculture is a major component for water use planning in the state of Georgia. Despite its importance in planning little credible information is available on current water consumption patterns and future demand. The ongoing water rights battle between Alabama, Florida, and Georgia, where the other two states contend that Georgia is wrongfully taping into the waters that belong to them, has necessitated a better understanding of water needs over the next several years. Policy makers are often limited by information available to them which may be lacking due to missing or highly aggregated data. The main problem caused by missing data is that estimators of population characteristics are assumed to be biased unless evidence to the contrary is provided. Also, the mean square errors of such estimators are likely to be larger than those obtained from complete data. The challenge, therefore, lies in being able to estimate the missing values in an unbiased and consistent manner. These estimates may then be used to make forecasts. The first objective of the study was to derive a method to disaggregate the available information for the state of Georgia to a county and commodity level. The second objective is to use the disaggregated data to forecast irrigation water demand for the counties by commodity. This analysis focuses on seven commodity groups: corn, cotton, peanuts, rye, sorghum, soybeans, and tobacco for 1970 through 1998.

OBJECTIVES

Data is disaggregated to a county by crop irrigated acreage using the three methods outlined below. Such disaggregation makes it possible to account for a greater source of variability in modeling the data. Disaggregated data is used to forecast irrigated acreage for each county by commodity. An ARMA model is developed to better understand the historic agricultural irrigation patterns and to project future agricultural irrigation acreage. The application of this technique focuses on seven commodity groups in Georgia: corn, cotton, peanuts, rye, sorghum, soybeans and tobacco from 1970 through 1998.

DATA, METHOD AND MODEL

Two data sources were used for the analysis: The University of Georgia – Cooperative Extension Service (UGA-CES) and the National Agricultural Statistic Service (NASS) of the USDA. Irrigation data, published by UGA-CES came from the University of Georgia county extension agents. These data contain irrigated acreage by commodity for the state and irrigated acreage by county for all commodities combined and are reported in the annual publication the Georgia County Guide. All harvest data are from NASS. These data contain the commodity-harvested acreage by county. With few exceptions, these data are available for all commodities and counties from 1970 through 1998. Missing data are crop irrigated acreage by county. Aggregate acreage across all crops by county exists with only acreage by crop reported on a state level. Given these data sets, the problem is to develop a data set consisting of crop by county irrigated acreage. The state commodity and county total irrigated acres estimation for missing data assumes that irrigation acreage increases or decreases linearly between two time intervals. Three methods were used to disaggregate the data sets into a county irrigation acreage data sets by commodity. The first method (Method I) assumes that the amount of county irrigation for a given commodity is proportional to the amount of harvest for the county. The assumption is that the irrigation gradient for a given commodity across the state remains constant. This method ensures that the estimated state commodity total equals the known state commodity total. Therefore the state total equals the known state total. However, it does not require that the county total for the sum of the commodities equals the known county total. Table

Table 1. Example using Method I for disaggregating commodity state irrigated acres into county level commodity irrigated acres.

Commodity	CHA	SHA	CHA/SHA	State Irrigated acres (SIA)	County Irrigated acres (CIA)
Corn	2,000	40,000	0.05	10,800	540
Cotton	1,200	34,000	0.035	6,800	240
Peanuts	1,800	52,000	0.035	13,520	468
Rye	300	7,800	0.038	156	6
Sorghum	200	5,600	0.036	784	28
Soybeans	900	11,450	0.079	916	72
Tobacco	1,100	19,525	0.056	7,224	407
Other	1,800	47,625	0.038	20,003	756
Total	9,300	218,000		60,203	2522

Note: CHA is the county harvested acres and SHA is the state harvested acres.
SIA is the known state irrigated acres for a given commodity, CIA = (CHA/SHA)*SIA.

Table 2. Example using Method II for disaggregating commodity state irrigated acres into county level commodity irrigated acres.

Commodity	County irrigated acres (CIA)	Weight coefficient (WC)	Total county irrigated acres (TCIA)	Weighted county irrigated acres (WCIA)
Corn	540	0.215	3,200	688.0
Cotton	240	0.095	3,200	304.0
Peanuts	468	0.186	3,200	595.2
Rye	6	0.002	3,200	6.4
Sorghum	28	0.011	3,200	35.2
Soybeans	72	0.029	3,200	92.8
Tobacco	407	0.162	3,200	518.4
Other	756	0.300	3,200	960.0
Total	2517	1.0		3,200.0

Note: WC = (CIACommodity)/(CIATotal), WCIA= WC*TCIA.

1 illustrates the Method I computations for hypothetical data.

Method I is biased for county total irrigated acreage, i.e. for this example the county total irrigated acreage is 678 acres less than the known county total irrigated acreage. The 678 acres is the difference between the Method II, constrained to equal the known value, and Method I total county irrigated acreage.

Method II uses a weighted average to allocate the irrigated acres to a commodity within a county. The initial computations are identical to Method I. Therefore the county irrigated acres obtained using Method I (Table 1) are used as the initial values for the Method II example. The commodity county irrigated acres are then summed and the proportionality for each county is calculated. This weighted proportionality is used to compute the commodity acreage within a county. Table 2 illustrates Method II for a hypothetical county.

Method II ensures that the county irrigation acreage equals the known county irrigation total. Therefore the state total (sum of the county totals) equals the known state total. Method III is the average of Methods I and II for irrigated acres by county and commodity. This method is correct for the state total irrigated acreage but is biased for the state total of a given commodity and county total irrigated acres. However, the bias magnitude is less than the respective bias of the other methods.

A generalized statistical model was developed for the disaggregated data to forecast the county

Table 3. The transformed covariance parameter (REML) estimates of Methods I, II, and III

Cov. Parameter	Subject	Method I	Method II	Method III
	County			
AR (1)		0.72714569	0.87318865	0.82051933
Residual		5.48145580	44.2214136	12.1331265
Model Fit Statistics				
AIC		-1907.65	-2583.89	-2110.59
SBC		-1912.55	-2588.78	-2115.48
-2ResLL		3811.31	5163.78	4217.17

Table 4. The fitted mixed model results of fixed effects for Method I, II, and III

	Source	NDF	DDF	Type III F	Pr > F
Method I	County	47	623	36.59	0.0001
Method II	County	47	879	46.06	0.0001
	Time	1	256	22.76	0.0001
Method III	County	47	986	151.92	0.0001
	Time	1	624	145.86	0.0001

commodity irrigated acreage. The model assumes a linear trend with county specific intercept, slope, and a county by time interaction. The model assumes both random and serially correlated errors. Serial correlation is a function of the distance (d) between time j and k, (f (d_{jk})). The generalized statistical model assuming a linear trend is:

$$Y_{it} = (\alpha_0 + \alpha_i) + \beta_{0t} + \beta_{it} + \varepsilon_{it} + \delta_{it} \qquad (1)$$

where Y_{it} is the i^{th} county estimated irrigated acres at time t, α_0 is the intercept, α_i is the additive intercept effect for the i^{th} county, β_{ot} is the slope coefficient, β_{it} is the i^{th} county and t^{th} time interaction effect. ε_{it} are i.i.d. $\sim N(0, \sigma^2)$, $\delta_{it} \sim N(0, \sigma^2 [f(d_{jk})])$, with $i = 1, \ldots, n$, $t = 1, \ldots, 22$ and $cov(\varepsilon_{it}, \delta_{it}) = 0$. The data are modeled using mixed models in statistical software SAS®.

RESULTS AND DISCUSSION

Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC) (Akaike, 1978) were used to determine which function modeled the covariance and provided the best fit. Forecasts were obtained using the best fitting model. State forecasts of irrigated acreage for the three methods were compared to the known state total for commodity-irrigated acres for a selected year. The associated standard error and 95% forecast intervals were contrasted among the three disaggregation methods. The following table summarize the finding of the prediction model.

All methods have a high lag one correlation. The model fits best using the data from Method I as illustrated by the AIC, SBC and the –2(log-likelihood) values. The interaction term is not significant for either of the methods. Table 4 presents the fixed effects tests and associated statistics.

Method I is sufficiently modeled as an intercept model with an AR(1) covariance structure. The other two methods are a function of both time and county. All these results are highly statistically significant as seen by the p-values (0.0001) for all parameters.

Finally, all the methods are tested via in-sample validation. The irrigated acreage was predicted for the years 1996, 1997 and 1998 and compared against the *known* state totals. Method I outperforms the other two methods on account of more precise estimated values and appreciably smaller standard errors. As expected, the standard errors get larger for all three methods as forecast is extrapolated further in the future. As a word of caution to the practitioner, the choice of method of disaggregation depends upon the focus of analysis, i.e. the desired accuracy on either the state or county level.

LITERATURE CITED

Akaike, H. 1978. "Time series analysis and control through parametric models, *Applied Time Series Analysis*, D.F. Findley (ed.), Academic Press, New York.

Harrison, K. and A. Tyson. 1999. Irrigation survey for Georgia. p.421-424. In. K. J. Hatcher (ed.) *Proceedings of the 1999 Georgia Water Resources Conference.* Univ. of Georgia Institute of Ecology, Athens, Ga.

United States Department of Agriculture. *Georgia Agricultural Facts – Several Editions.* 1970 - 1990. USDA, National Agricultural Statistical Service. Athens, Georgia.

The University of Georgia Cooperative Extension Service. 1970 - 1998. The Georgia County Guide. Several editions.