ISSUES AND PROBLEMS IN AGRICULTURAL WATER DEMAND ESTIMATION FROM SECONDARY DATA SOURCES

Gary D. Lynne

Economists have a continued and justified interest in agricultural water demand estimation. Water is becoming a scarce resource throughout much of the United States, even in "water abundant" eastern states such as Florida [4, 15, 19]. As a result, new water institutions are evolving to address the water allocation problems [20, 22]. Estimates of values in alternative uses are important information in these water allocation decision processes.

Economists studying agricultural water demand generally have taken one of two approaches to estimation, referred to herein as the "analytic" and the "aggregate" approaches. In the analytic approach, attempts are made to determine the underlying soil-water-plant relations to estimate production functions for water.[1] Anderson et al. [1] provide an excellent review of some of this work. Other studies of special interest include [10, 11, 12]. Various quantitative methods and approaches have been used to represent the production relations, including econometrics, simulation, and mathematical programming. All of these efforts share two common problems: (1) use of the analytic approach depends on knowledge derived from basic experimentation by crop scientists, which is generally expensive, and (2) it is extremely difficult to identify all of the interactions with the water resource of factors in the experiments. As a result, the analytic models have limited usefulness in examining the long run impacts of changes in relative price ratios or changes in the institutional structure. The aggregate models appear necessary.

Ruttan [18] in his 1966 study suggested a more aggregate, econometric approach for understanding the value productivity over wider regions. The aggregate programming models have also been used [6, 23] in finding the range of values for water in major uses and areas of the country.[2]

The purpose of this article is to discuss problems inherent in some aggregate models, particularly the approach suggested by Ruttan. The discussion is based on the presumption that analytic models will *not* be able to provide all the answers to water allocation problems. Information derived from more aggregate models will be needed, especially to resolve substitution and/or interregional transfer issues.

CONCEPTUAL BASIS AND DATA NEEDS FOR AGGREGATE WATER DEMAND ESTIMATION

As background for discussion, some concepts related to water demand estimation from secondary sources (in particular, from the U.S. Agricultural Census, as developed in the Ruttan approach) are reviewed. The theory of derived demand is well documented [7, 13, 14] and is not discussed in its entirety here. Given some production relation

$$q = f(x_1, x_2, \ldots, x_n) \tag{1}$$

the first order conditions for a profit maximizing firm, operating under perfectly competitive conditions, can be used to derive several alternative forms of the demand for the factors, namely

$$r_i = r_i(r_1, r_2, .., r_{i-1}, r_{1+1}, ..., r_n, x_i, p) \tag{2}$$

$$r_i = r_i(r_1, r_2, .., r_{i-1}, r_{i+1}, ..., r_n, x_i, q) \tag{3}$$

$$r_i = r_i(r_1, r_2, .., r_{i-1}, r_{i+1}, ..., r_n\, x_i, C) \tag{4}$$

$$r_i = r_i(r_1, r_2, .., r_{i-1}, r_{i+1}, x_i, x_{i+2}, .., x_n, p) \tag{5}$$

$$r_i = r_i(x_1, x_2, .., x_i, x_n, p) \tag{6}$$

Gary D. Lynne is Assistant Professor, Food and Resource Economics Department, Institute of Food and Agricultural Sciences, University of Florida.

Florida Agricultural Experiment Station Journal Series No. 1384. This article benefited from the diligent efforts of *Journal* reviewers. Special thanks are also due to Roe Carriker and Max Langham for especially perceptive comments. Remaining errors belong to the author.

[1] This is the domain primarily of those economists attempting to work with other agricultural scientists in modeling and describing production functions for yield response to water.

[2] The literature contains a large number of examples of the "aggregate" mathematical programming models. These extensive discussions are time honored, well documented, and need not be discussed here.

where

r_i = factor prices
p = produce prices
x_i = factors of production
q = product
C = total costs of production.

Equations 2, 3, and 4 represent the constant marginal cost, constant output, and constant cost derived demand curves, respectively, for x_i [14]. Equation 5 represents the derived demand (a marginal value product) for factor x_i given that some of the factors are fixed (in this case, $x_{i+2}, \ldots, X_n$) [7]. Equation 6 is also a marginal value product ("short run" demand) relation, where all factors other than x_i are fixed. Any and all of these conceptual models can form the basis for agricultural water demand estimation; the appropriate model depends on the problem being addressed and on the expected use of the results. Generally, if one assumes the firms involved are profit maximizers having essentially unlimited resources and no imposed output constrains on their respective industries, equations 2, 5, and/or 6 are appropriate 2 for "long run" and 5 and 6 for "short run" analysis.

Finding an appropriate data set is a major problem, however. Unfortunately, there are very few functioning markets for agricultural water; thus, it is virtually impossible to estimate the derived demand equations directly.[3] The researcher usually is forced to estimate the production functions for water, and to derive the demand curves of interest. The early work by Ruttan provides insight into the nature of the aggregate water production function.

Ruttan used one year of census data (1954) from irrigation counties within the major hydrologic regions of the United States [18]. Several alternative forms of the aggregate function were tested. Generally, the models had the form (Lynne's notation)

$$(7)\quad TVP = f(L, O, T, K, I, N)$$

where [18, pp. 38-39]

TVP = total value of farm products sold in the county in 1954
L = total number of family and hired farm workers
O = current operating expenses, measured by the summation of expenditures on purchased feed, fertilizer, and lime
T = value of machinery and equipment in a county for the western regions; number of tractors on farms for the eastern regions
K = value of livestock investment in a county for western regions; livestock was not considered in the eastern regions
I = irrigated cropland harvested and pasture irrigated in a county for the western regions; total irrigated land in a county for eastern regions
N = nonirrigated cropland harvested in the county.

Various combinations of these variables were included in the Ruttan test models [18, pp. 38-39]. The demand model used was a variation on equation 3.

CONSIDERATIONS IN THE AGGREGATE RUTTAN APPROACH

The Ruttan approach has been criticized and evaluated elsewhere (see especially Hoch [8]) and suggestions have been made for improvements in the method [2, 3]. The major criticism has been leveled at the manner in which multicollinearity problems were resolved by variable deletion,[4] which leads to specification error [8]. Brown and Beattie [3] suggest ridge regression as a means for resolving the nearly insurmountable multicollinearity problem, and isolate the instances when favorable results are most likely. Beattie [2] also suggests that appropriate delineation of the county observations into more homogeneous units may aid the estimation process. This suggestion may not be appropriate if homogeneity also means crop type, as discussed hereafter.

Two other problems not previously discussed in the literature are also associated with the aggregate Ruttan approach: (1) a large variation in the size of counties (in terms of land and the associated total farm sales value) may place an upward bias on the productivity coefficient for irrigated land, as well as other inputs; (2) the "price" of water may already be accounted for in an operating expense variable as generally used in the Ruttan type model. To include irrigated land and operating expenses as variables may be equivalent to including the

---

[3] The only "price" agricultural producers generally pay for water (especially in the eastern states) is the cost of getting water to specific crops and locations. If markets existed, this cost would be a component in determining the "price" for water. If there were data on the marginal factor costs of pumping and applying irrigation water, the researcher could estimate the demand curve for water by relating factor cost to quantities used. Such data are not readily available for the eastern U.S.

[4] Ruttan ultimately reduced the variables in equation 7 to simply I and O because of multicollinearity problems [18, pp. 19-26].

same independent variable in the equation twice. The results from direct consideration of these problems follow.

AN ALTERNATIVE MODEL AND EMPIRICAL RESULTS

A modified Ruttan approach guided the estimation process described in this section. The approach differs from the Ruttan approach in that time series (5 year intervals) as well as cross-section data are used. Also, the independent variables included in the production relation are defined differently. The general form of the model considered is:

$$(8)\quad TVP = h(OK, TR, CA, I, NI)$$

where

TVP = total value of farm products sold from commercial farms in a county for each of the census years, in constant 1910-1914 dollars
OK = intensity of operating capital use in the county, measured by operating capital per acre of all commercial farmland (1910-1914 dollars); all operating expenses are in OK, including expenditures for hired labor and an assumed return to operator labor
TR = intensity of machinery and equipment investment, measured by number of tractors per acre of all commercial farmland
CA = importance or intensity of livestock production in the county, measured by cattle per acre of all commercial farmland
I = total acres of irrigated land on commercial farms
NI = total acres of nonirrigated land in commercial farms.

The OK, TR, and CA "intensity" variables are chosen to isolate the effects of I and NI on TVP, and are not measures of the "factors of production," in the usual sense.[5] Rather, these variables serve to represent the technology set in the study area. Thus, quantification of this model facilitates estimates of the marginal value of irrigated and nonirrigated land as specified in equation 5, over a "short run" situation.

A 13-county hydrologic region in South Florida was selected, encompassing one of the five water management districts in the state.[6] This area is very heterogeneous across counties with respect to crop type. Major agricultural crops are citrus, vegetables, sugar cane, and livestock. Nearly all counties have some pasture. A large share of the pasture is irrigated, as are nearly all vegetables, all of the sugar cane, and much of the citrus.

Ordinary least squares estimation procedures were used to generate the results in Table 1, where county size is not explicitly considered. Time (T) was included to remove trend effects. The other variables were as described in equation 8. The finding that the coefficient on I is significant at the 0.05 level (two-

TABLE 1. TOTAL VALUE PRODUCTIVITY (AGRICULTURAL) FUNCTION, ESTIMATES FOR SOUTH FLORIDA AREA (13 COUNTIES), 1949-69.

$$(9)\quad \ln TVP^{a} = \underset{(1.725)^{b}}{1.9097} + \underset{(6.556)}{0.6953}\ln OK + \underset{(3.277)}{0.3785}\ln TR + \underset{(0.689)}{0.0822}\ln CA + \underset{(2.371)}{0.1150}\ln I + \underset{(8.645)}{0.9410}\ln NI + \underset{(8.820)}{0.0902}\ln T$$

$R^2 = 0.88$ $\quad F = 73.64$, 6 and 58 degrees of freedom.

[a] Variables defined as:
TVP = deflated (1910-1914 dollars), value of farm sales in county on commercial farms
OK = deflated operating capital per acre of all farmland, including expenditures for hired labor and an assumed return to operator labor (1910-1914 dollars)
TR = number of tractors per acre of all farmland
CA = number of cattle per acre of all farmland
I = total acres irrigated land in the county
NI = total acres non-irrigated land in the county
T = time, linear sequence, 1 = 1949, ..., 5 = 1969.

[b] The t-statistics are in the parentheses.

---

[5] Ideally, as argued in the seminal work by Ruttan, the model would be formulated so as to facilitate estimates of the resource substitution possibilities. This may not be possible for all resources because of multicollinearity problems. The fomulation in equation 8 facilitates estimates of the marginal value product (MVP) of I and NI, given various intensities of labor and capital (variable and fixed). Also, the substitution relation between I and NI can be derived. The multicollinearity was reduced considerably by use of equation 8 as the guiding model.

[6] Data were derived from the U.S. Agricultural Census for Broward, Collier, Dade, Glades, Hendry, Highlands, Lee, Martin, Okeechobee, Orange, Osceola, Palm Beach, and Saint Lucie Counties, Florida, for the 1949, 1954, 1959, 1964, and 1969 census years. Price indexes were obtained from [21].

tailed t-test), as are the OK, TR, NI, and T variables, suggests a useful equation for estimating the marginal value of I. Multicollinearity does not appear to be severe in the equation. The highest simple correlation coefficient is 0.88 between OK and TR. All the rest are less than 0.40. No significant linear combinations were found in further testing.

The problems suggested earlier now can be highlighted. The size of counties included in the sample varied greatly, with a coefficient of variation of 1.13 for TVP, 1.17 for I, 0.65 for NI, and 0.56 for total farmland. Thus, a portion of the variation in TVP is due entirely to size of the county. As a result, the coefficients of the independent variables were expected to be biased upward. This hypothesis was tested by resorting to a covariance model.

A covariance model is based on the presumption that each cross-sectional unit during each period of time has a unique intercept [9], such as could be caused, in part, by the county size influence. Thus 0-1 "dummy" variables were incorporated to represent the differences among counties over time. More directly, one of the 13 counties was chosen as the base county against which to test each of the other counties for significant differences. The census year 1949 was chosen as the base year for the qualitative time variables. Ordinary least squares regression procedures were then used, retaining the variables specified in equation 8. The results are shown in Table 2.

The regression coefficients did decline for all the variables except CA (Table 2). Thus, the hypothesis is not rejected.[7] The county "dummy" variables (D1 - D12, many of which are significant in equation 10 remove at least some of the size effect, which is implicit in equation 9. The existence of some significant county "dummies" gives indication that the 13-county area is heterogeneous, in part because of the size influence. The t-statistic for I in equation 10 is lower than in equation 9. However, the coefficient for I in equation 10 should give a more accurate estimate of the demand price for irrigated land (and thus for water), to the extent the county dummies removed the county size effect.

The coefficient on I in equation 10 still may not generate a very useful measure of value, however. To some extent, both OK and I measure the same influence. To illustrate, assume I was proportional to the total water applied (W), or I = kW. Including I and the total cost (C) of applying water (C = rW where r is the average cose of applying W, and C is contained in OK) is equivalent to having both kW and rW as independent variables. Thus, part of the influence of the irrigation water is being measured in OK. Unfortunately, C cannot be determined from the census data and separated out of the OK variable.

The county dummies reduce the heterogeneity in the data, but also reduce the significance levels on all of the independent variables, except for CA. The results for the I variable, of special interest here, are particularly revealing. The regression coefficient on I is no longer significant at usually acceptable levels. The reason for this reduction in significance is important in light of the suggestion by Beattie to choose more homogeneous study areas. For this data set, the use of county dummies (which reduce heterogeneity in size) also reduces heterogeneity in crop type as certain crops dominate within each county for this area, whereas nearly all counties have some pasture.[8] In addition, the dummies serve to reduce differences between counties having similar crop mixes. The result is a rejection of the hypothesis that irrigated land has a nonzero marginal value. Therefore, choosing a homogeneous study area, especially where

TABLE 2. COVARIANCE ANALYSIS OF TOTAL VALUE PRODUCT RESPONSE OF AGRICULTURAL PRODUCTION, ESTIMATES FOR SOUTH FLORIDA AREA (13 COUNTIES) 1949-1969.

$$\begin{aligned}
(10)\quad \ln TVP^{a} = {} & \underset{(1.263)^{b}}{3.2918} + \underset{(3.394)}{0.4742}\ln OK + \underset{(0.036)}{0.0066}\ln TR + \underset{(0.700)}{0.1338}\ln CA + \underset{(1.088)}{0.0829}\ln I \\
& + \underset{(3.343)}{0.6127}\ln NI + \underset{(0.438)}{0.1800}\ln D1 - \underset{(-1.434)}{0.4864}\ln D2 + \underset{(1.421)}{0.6078}\ln D3 - \underset{(-2.1441)}{0.7636}\ln D_4 \\
& - \underset{(-0.084)}{0.0246}\ln D5 - \underset{(-0.381)}{0.1002}\ln D6 - \underset{(-0.861)}{0.2256}\ln D7 - \underset{(-2.129)}{0.5419}\ln D8 - \underset{(-2.396)}{0.8779}\ln D9 \\
& + \underset{(2.530)}{0.7190}\ln D10 - \underset{(-2.601)}{0.9519}\ln D11 + \underset{(2.553)}{0.6679}\ln D12 + \underset{(1.253)}{0.2412}\ln T2 + \underset{(2.820)}{0.6704}\ln T_3 \\
& + \underset{(3.617)}{0.8716}\ln T4 + \underset{(2.478)}{0.7418}\ln T5
\end{aligned}$$

$R^2 = 0.94$ $\qquad$ F = 35.32, 21 and 43 degrees of freedom

[a]Variables defined as:

TVP, OK, TR, CA, I, NI = as defined in Table 1.

DI = "dummy" variables for counties i = 1 through i = 12; value of 3 = 2.718 ... or 1.

tj = time "dummy", j = 2 for 1954 through j = 5 for 1969 (ag. census years); value of 3 = 2.718 ... or 1.

[b]The t-statistics are in the parentheses.

---

[7] An explanation for the rise in the coefficient on CA is given hereafter.

[8] This statement is supported by the increase in the coefficient on CA. High valued crops are concentrated in some counties and lower valued livestock enterprises (and associated crops) dominate in others. Yet nearly *all* counties had some cattle. Thus, a portion of the variation in TVP was due to variation in value of sales *among crops*. The qualitative "dummies," then, removed some of this crop effect, increasing the influence of the CA variable on TVP.

homogeneity is taken to mean similar crop types, could ensure a low level of significance on the I variable. This statement is explained more fully hereafter. At minimum, the researcher may wish to retain the intracrop heterogeneity in the data set.

An insignificant coefficient on an I variable (as normally included in such aggregate modeling efforts) could result for several reasons, each of which should be contemplated by the researcher. The following statements are offered for consideration and as testable hypotheses. Agricultural producers would be expected to apply irrigation water (for each crop) to the point where the marginal value of water (MVP) is equal to the marginal (factor) cost (MFC) of application. Basic agricultural scientists (and their extension counterparts) in Florida, however, generally recommend that water be applied on the basis of "optimum growing conditions" for maximum yield (see e.g. [16, 17]), where $r_i = 0$ for equations 2-6. This is expected to be true in other areas of the United States as well. In addition, because of the common property aspects of the water resource, one would attempt to capture as much as possible (where equation 5 and/or 6 could actually be negative).[9] As a result, it can be hypothesized the MVP estimates may be close to zero (or negative) because (1) many agricultural producers probably follow the advise of the physical scientists and/or (2) the theoretical result for the expected outcome in a common property resource case actually occurs. Thus, if **each county observation reflected the zero value** of the marginal value product of I, it is unlikely that one would obtain a significant coefficient for the I variable with the aggregate approach — assuming, of course, that firms were operating on the same production function.

The same problem (insignificant coefficient on I) could arise in such aggregate data if all producers tended to use the same irrigation technology and to follow similar operating procedures. More directly, if producers across counties face a similar MFC function and tend toward the same maximum profit level of water use, a regression analysis with these data points will give an insignificant coefficient for I. Thus, the researcher may wish to consider a grouping of the data to incorporate heterogeneity in irrigation technology and operating procedures — again assuming a common production function across counties.

As is now apparent, the sources of variation in TVP across counties include (1) the county size influence, (2) the intercrop influence (among vegetables, citrus, sugar cane, and pasture, for this case), and (3) within crop influence. The analyst concerned with the MVP must decide which influence is to be isolated. Generally, the county size effects must be removed from the data. This step could be accomplished by grouping similar size counties within a hydrologic region. The intercrop and intracrop differences, if present, should be retained.

CONCLUSIONS

This article is a discussion of some of the problems inherent in using secondary census data at the county level in agricultural water demand estimation. Major conclusions are:

1. An adjustment for county size to a more homogeneous data set is generally desirable.
2. Using a regression coefficient on irrigated land to estimate the marginal value of water when total operating cost is also included as an independent variable may lead to "double counting."
3. Grouping counties to form more homogeneous data sets in terms of crop type will not allow estimation of the intercrop effect on the marginal value of water, which may also be of interest. It is also possible that there are no intracrop variations in a homogeneous study area because of the influence of recommendations to produce maximum yields and/or the "common property" influence.
4. Covariance analysis can be used to test for homogeneity among a group of counties. Thus, this type of analysis should become a corollary to the Beattie recommendation for homogeneous study areas.

Improvements in secondary data will be necessary to improve significantly the accuracy of water demand estimates from the aggregate approaches. Further research using the "aggregate" models should continue only in conjunction with further efforts to improve the data set. Because of the inadequacy of current secondary data sources, more researchers may have to resort to the development of "analytic" models, which start with the basic underlying soil, water, and plant relations.

---

[9] The "common property problem," as outlined for the fishery in the early work by Gordon [5], appears apropos to this case as well. It is expected that irrigators may also act to equate the average value of the product from irrigation water to the marginal cost of application, especially in cases where water supplies are short. This influence may have been small during 1949-1969 in Florida, however, because of the relative abundance of water in that period.

REFERENCES

[1] Anderson, Raymond L., Dan Yaron, and Robert Young. *Models Designed to Efficiently Allocate Irrigation Water Use Based on Crop Response to Soil Moisture Stress,* Technical Report 8, Environmental Research Center, Colorado State University, May 1977, 82 pp.

[2] Beattie, Bruce R. "Marginal Value Productivity of Water in Irrigated Agriculture: A Modification of the Ruttan Approach," *Western Agricultural Economics Association 44th Annual Meeting,* Squaw Valley, California, 1971, pp. 109-112.

[3] Brown, William G. and Bruce R. Beattie. "Improving Estimates of Economic Parameters by Use of Ridge Regression with Production Function Applications," *American Journal of Agricultural Economics,* Volume 57, February 1975, pp. 21-32.

(4) Economic Research Service. *Our Land and Water Resources. Current and Prospective Supplies and Uses,* Miscellaneous Publication 1290, USDA, May 1974.

[5] Gordon, H. Scott. "The Economic Theory of a Common-Property Resource: The Fishery," *Journal of Political Economy,* Volume 65, 1954, pp. 124-142.

[6] Heady, Earl O., Howard C. Madsen, Kenneth J. Nicol, and Stanely H. Hargrove. "National and Interregional Models of Water Demand, Land Use, and Agricultural Policies," *Water Resources Research,* Volume 9, August 1973, pp. 777-791.

[7] Heady, Earl O. and Luther Tweeten. *Resource Demand and Structure of the Agricultural Industry.* Ames: Iowa State University Press, 1963.

[8] Hoch, Irving. "Book Review: The Economic Demand for Irrigated Acreage: New Methodology and Some Preliminary Projections, 1954-1980," by Vernon W. Ruttan, *Natural Resources Journal,* Volume 7, 1967, pp. 466-471.

[9] Hoch, Irving. "Estimation of Production Function Parameters Combining Time-Series and Cross-Sectional Data," *Econometrica,* Volume 30, January 1962, pp. 34-53.

[10] Mapp, Harry P. Jr., Vernon R. Eidman, John F. Stone, and James M. Davidson. *Simulating Soil Water and Atmospheric Stress-Crop Yield Relationships for Economic Analysis,* Technical Bulletin T-140, Agricultural Experiment Station, Oklahoma State University, February 1975, 63 pp.

[11] Miller, Stanley O., Larry L. Boersma, and Emery N. Castle: *Irrigation Water Values in the Williamette Valley: A Study of Alternative Valuation Methods,* Technical Bulletin 85, Agricultural Experiment Station, Corvallis, Oregon, September 1965.

[12] Minhas, B. S., K. S. Parikh, and T. N. Srinivasen. "Toward the Structure of a Production Function for Wheat Yields with Dated Inputs of Irrigation Water," *Water Resources Research,* Volume 10, June 1974, pp. 383-393.

[13] Mosak, J. "Interrelations of Production, Price, and Derived Demand," *Journal of Political Economy,* Volume 46, December 1938, pp. 761-787.

[14] Mundlak, Y. "Elasticities of Substitution and the Theory of Derived Demand," *Review of Economic Studies,* Volume 2, 1968, pp. 225-236.

[15] Parker, G. "Water and Water Problems in the Southwest Florida Water Management District and Some Possible Solution," *Water Resources Bulletin,* Volume 11, February 1975, pp. 1-20.

[16] Reitz, H. J. *Water Requirements for Citrus,* WRC-4, Institute of Food and Agricultural Sciences, University of Florida, 1977.

[17] Rogers, J. S. and D. S. Harrison. *Irrigation Water Requirements for Agronomic Crops in Florida,* WRC-5, Institute of Food and Agricultural Sciences, University of Florida, 1977.

[18] Ruttan, Vernon W. *The Economic Demand for Irrigated Acreage; New Methodology and Some Preliminary Projections, 1954-1980.* Baltimore: Johns Hopkins Press, 1965.

[19] Storch, W. "Water Management During Drought Events," *In Depth Report 2.* West Palm Beach, Florida: Central and Southern Florida Flood Control District, April 1974.

[20] Trelease, F. J. "The Model Water Code, the Wise Administrator and the Goddam Bureaucrat," *Natural Resources Journal,* Volume 14, April 1974, pp. 207-229.

[21] U. S. Crop Reporting Board. *Agricultural Prices, 1969 Annual Summary* PR 1-3(70), Washington, D.C., June 1970.

[22] Wadley, J. B. "A Summary Guide to Florida's Water Rights," Extension Circular 412, Institute of Food and Agricultural Sciences, University of Florida, 1976.

[23] Wolman, N. *The Value of Water in Alternative Uses.* Albuquerque, New Mexico: The University of New Mexico Press, 1962, 426 pp.