REGIONAL ACREAGE RESPONSE BY QUARTER FOR FRESH TOMATOES: AN EXAMPLE OF THE USE OF MIXED ESTIMATION

Michael D. Hammig

The study reported was motivated by a USDA study to develop complete quarterly models of supply and demand for a selected set of fresh salad vegetables. The acreage planted component enters recursively into both the acreage harvested and yield relations used in many of these models. Consequently, predictions of acreage planted are instrumental in predicting total supply and resulting market equilibrium solutions.

In modeling acreage planted over relevant seasons within four regions, various sources of information can be brought to bear. Obviously, data series on past plantings, costs, and prices provide the foundation of statistical estimation of an acreage response model. However, additional information from previous studies, economic theory, and subjective judgment on the part of the researcher also can be incorporated into the model through the use of the mixed estimation technique developed by Theil and Goldberger [8].

The purpose of this article is to appraise the usefulness of mixed estimation, and to provide examples of relations estimated by pure, mixed, and restricted least squares regression as sources of comparison of the techniques. Though the mixed estimation technique is not new, its use has been limited and its performance in comparison with that of other techniques is not widely known. The U.S. fresh market tomato model offers an appropriate means for examining such a comparison.

A relatively large number of acreage relations are required, and the specifications of all acreage equations can be identical. Fresh tomatoes are produced in the Southeast during all four quarters.[1] In the South and Southwest they are produced in the spring, summer, and fall. In the "All Other" region, production is significant only during the summer months. Thus, a model representing total U.S. production, disaggregated by season and by region, requires 11 separate acreage planted equations.

MODEL SPECIFICATION

The model specification adopted for this study is:

$$AP^r_{q,t} = AP^r_{q,t}\,(P^{e,r}_{q,t}, R^r_{q,t}, CP_{q,t}, AP^r_{q,t-1}, u^r_{q,t})$$

where

$AP^r_{q,t}$ = acreage of fresh tomatoes planted in region r in quarter q of year t

$P^{e,r}_{q,t}$ = expected price for fresh tomatoes, in \$/cwt, in region r in quarter q of year t

$R^r_{q,t}$ = farmer's subjective risk associated with fresh tomatoes in region r in quarter q of year t

$CP_{q,t}$ = costs of production in quarter q of year t

$AP^r_{q,t-1}$ = acreage of fresh tomatoes planted in region r in quarter q of year t−1

$u^r_{q,t}$ = disturbance term associated with acreage planted of fresh tomatoes in region r in quarter q of year t.

The expected price variable is defined as a three-year geometrically declining weighted average of past observed prices:

$$P^{e,r}_{q,t} = \sum_{i=1}^{3} w^i P^{e,r}_{q,t-i}$$

where w = .54369.[2] This expectation is motivated, in part, by Nerlove's [5] well-known "Adaptive expectations hypothesis" of geometrically declining weighted differences

Michael D. Hammig is Agricultural Economist, Fruits, Vegetables, and Sweeteners, National Economics Division, ESCS, USDA.

[1]Quarters are defined as: Winter - January, February, March; Spring - April, May, June; Summer - July, August, September; Fall - October, November, December. Regions are defined as: South - Texas, Oklahoma, Arkansas, Louisiana; Southeast - Kentucky, Virginia, Tennessee, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi; Southwest - California, Nevada, Utah, Colorado, New Mexico, Arizona.

[2]This value is obtained by solving the expression $w^3 + w^2 + w = 1$.

between actual and expected outcomes on price. Adherence to the method by which Nerlove's method becomes estimable implies that the weights are determined in the estimation process. To readily incorporate an expectations variable into the mixed estimation process, the weights are specified *a priori* as geometrically declining over a three-year period and summing to one. In this way, the price expectation is predetermined in each time period prior to estimation.

The risk associated with price expectations is represented by the square root of weighted squared deviations between actual and expected prices over the preceding three-year period, as observed in relation to the current expectation.

$$R^r_{q,t} = \frac{\sum_{i=1}^{3} [w^i(P^r_{q,t-1} - P^{e,r}_{q,t-i})^2]^{1/2}}{P^{e,r}_{q,t}}$$

where $P^r_{q,t}$ is the actual price in region r in quarter q of year t and other variables are as defined before.

Variable costs of production are represented by an index of prices paid by farmers for items used in production. Fixed factors, as well as psychological influences that induce the producer to continue production of the same crop over time, are represented by the preceding year's level of acreage planted.

Definition of major substitutes in production for fresh tomatoes is very difficult. Many crops compete for land suitable for vegetable production; however, no manageable subset can be defined, even on a regional basis, that can be conveniently incorporated into an econometric analysis. For this reason the effect of substitute alternatives is disregarded.

A PRIORI INPUT

Considerable *a priori* information is available on the anticipated signs and magnitudes of certain parameters. Results contrary to this prior information can lead to respecification of the relation and another round of estimation. To avoid such an experimental search for an acceptable specification and to take advantage of present knowledge presumed relevant to the proposed relation, *a priori* information was incorporated directly into the relation and the mixed estimation technique was applied. Prior information in the form of linear probabilistic constraints was developed for the coefficients on expected price, risk, and costs of production. The prior information was drawn from results of previous studies and subjective judgments of the researcher.

For example, studies by Nerlove and Addison [6] and Hammig [2] show the elasticity for fresh tomato acreage with respect to expected price, on an annual aggregate basis, to be .16 and .20, respectively. Economic theory dictates that this elasticity has a positive sign. Relations disaggregated by region and by quarter can be expected to reflect elasticities that are greater than those obtained on an annual basis. Because the decreased level of aggregation more nearly approximates the true economic conditions facing individual producers, effects of "averaging out" due to aggregation are diminished. Jesse and Machado [3] have shown this to be true for the California tomato crop. For these reasons the coefficient on expected price is constrained to reflect an elasticity of:

$$\epsilon_{AP,pe} = .5 \pm .5 \text{ with .95 probability.}$$ [3]

The point estimate is considerably higher than the previously derived elasticities. However, the variance around this estimate permits results in the range of those previously obtained, as well as elasticities that are greater. In essence, the constraint is that, with 95 percent certainty, acreage response to expected price will be inelastic.

Prior information about the elasticity of acreage with respect to risk also is available. Lin [4] obtained a value of −.06 for the elasticity of planted acreage of Kansas wheat with respect to risk, where risk was defined as the three-year moving standard deviation of past observations on price. A similar risk measure was used by Traill [9] in a national study of onion production. In a model for the total annual crop he obtained a value of −.02, whereas in a similar model for the late summer crop the risk elasticity was −.01. Hammig [2] obtained elasticities for risk, for a set of seven fresh vegetables on an annual national basis, ranging from −.004 to −.03. The effect of risk on tomatoes is assumed comparable to its effect as evaluated for other crops. The production process for tomatoes does not differ

[3]Note that, because expected price is included as a divisor in the risk term, the elasticity of acreage planted with respect to expected price involves a linear combination of regression coefficient. Redefine $R_{q,t}$ as $\frac{SD}{pe}$ and consider the relation

$$AP = \beta o + \beta_1 pe + b_2 \frac{SD}{pe}.$$

The elasticity of AP with respect to p^e is

$$\epsilon_{AP,pe} = \beta_1 \frac{p^e}{AP} - \beta_2 \frac{SD}{p^e \cdot AP}.$$

It is to this linear combination that the constraint is applied.

enough from production processes of other vegetable crops to warrant a unique influence from the risk factor.

The coefficient on risk is constrained to reflect an elasticity of

$$\epsilon_{AP,R} = -.05 \pm .05 \text{ with .95 probability.}$$

If tomato producers are assumed to be risk averse, the upper limit of the risk elasticity will be zero. The point estimate may appear somewhat high in relation to previous results; however, in a regional study, producer response to risk should be greater than that in more aggregated models.

Very few specific empirically derived values for elasticities of acreage with respect to production costs are available. However, economic theory suggests the range that these values should take. Neoclassical theory of supply is predicated on the assumption that producers are motivated by a desire to maximize profits. Because profit implies a tradeoff between product and factor prices, producer response to changes in these prices should be of like magnitude, but opposite direction. For this reason the coefficient on costs of production is restricted *a priori* to reflect an elasticity of

$$\epsilon_{AP,CP} = -.5 \pm .5 \text{ with .95 probability.}$$

In a manner directly analogous to the constraint on the expected price elasticity, this constraint implies that acreage response to changes in production costs will be in the inelastic range.

RESULTS

All equations were assumed linear and the 11 acreage relations were estimated by pure, restricted, and mixed least squares. Pure least squares represents the case in which prior knowledge is assumed to be completely uncertain. Restricted least squares represents the opposite extreme; the prior knowledge is as-

TABLE 1. ESTIMATION RESULTS FOR ACREAGE PLANTED OF FRESH MARKET TOMATOES, BY REGION AND BY QUARTER

Method[(a)]	Dependent Variable	Independent Variables: Constant	$pe^{,r}_{q,t}$	$R^{r}_{q,t}$	$CP_{q,t}$	$AP^{r}_{q,t}$	R^2	$\chi^{2(b)}_3$	$\Theta p^{(c)}$	$\epsilon_{AP,pe}^{(d)}$	$\epsilon_{AP,R}^{(d)}$	$\epsilon_{AP,CP}^{(d)}$
		Southeast										
1. OLS	$AP^{SE}_{W,t}$	17170.0	-209.59	-5295.8	-8.37	.1174	.22			-.203	-.030	-.067
		(3.26)	(.23)	(.29)	(.09)	(.48)						
RLS	$AP^{SE}_{S,t}$	12268.2	667.7	-9815.1	-62.9	.1752	.05			.5	-.05	-.5
		(4.07)				(.91)						
ME	$AP^{SE}_{W,t}$	15509.0	409.32	-10366.0	-67.75	.2107	.20	.78	.40	.284	-.053	-.538
		(3.23)	(1.82)	(2.22)	(2.96)	(1.00)						
2. OLS	$AP^{SE}_{SP,t}$	23485.0	749.86	23288.0	-89.81	.1137	.29			.396	.082	-.422
		(3.94)	(1.42)	(.86)	(1.62)	(.56)						
RLS	$AP^{SE}_{SP,t}$	21309.5	1313.8	-14125.1	-108.0	.1761	.04			.5	-.05	-.5
		(3.92)				(.88)						
ME	$AP^{SE}_{SP,t}$	22655.0	903.99	-11951.0	-105.61	.2745	.23	1.99	.33	.336	-.042	-.496
		(4.00)	(2.74)	(1.75)	(3.10)	(1.73)						
3. OLS	$AP^{SE}_{SU,t}$	7514.1	498.81	-21323.0	-35.42	.6327	.76			.181	-.080	-.247
		(3.46)	(2.29)	(2.06)	(1.58)	(4.81)						
RLS	$AP^{SE}_{SU,t}$	-2085.4	1051.9	-13382.2	-71.7	1.1181	.45			.5	-.05	-.5
		(.40)				(3.74)						
ME	$AP^{SE}_{SU,t}$	7898.9	640.81	-15363.0	-50.92	.6222	.76	1.21	.21	.278	-.057	-.355
		(3.74)	(3.70)	(2.75)	(2.92)	(4.75)						
4. OLS	$AP^{SE}_{F,t}$	6506.7	950.08	6654.1	-79.26	.1982	.55			1.084	.046	-.861
		(2.96)	(2.55)	(.52)	(2.32)	(1.02)						
RLS	$AP^{SE}_{F,t}$	3602.7	503.4	-7245.4	-46.1	.6895	.34			.5	-.05	-.5
		(1.38)				(2.94)						
ME	$AP^{SE}_{F,t}$	6340.8	623.12	-5179.6	-50.58	.3431	.48	4.18	.35	.645	-.036	.-549
		(3.43)	(4.00)	(1.51)	(3.49)	(2.00)						
		Southwest										
5. OLS	$AP^{SW}_{SP,t}$	2527.2	173.05	-581.3	-21.73	.2207	.09			.805	-.013	-.785
		(2.10)	(.93)	(.07)	(.80)	(.90)						
RLS	$AP^{SW}_{SP,t}$	2896.8	116.3	-2155.1	-13.8	.1329	.02			.5	-.05	-.5
		(3.32)				(.54)						
ME	$AP^{SW}_{SP,t}$	2692.7	112.06	-2054.7	-12.78	.1709	.08	.37	.44	.485	-.048	-.461
		(2.74)	(2.83)	(1.94)	(2.24)	(.79)						

Continued--

TABLE 1. Estimation Results for Acreage Planted of Fresh Market Tomatoes, by Region and by Quarter--Continued

Method[(a)]	Dependent Variable	Constant	Independent Variables $pe^{r}_{q,t}$	$R^{r}_{q,t}$	$CP_{q,t}$	$Apr_{q,t}$	R^2	$\chi^{2\,(b)}_{3}$	θ_p[(c)]	$\varepsilon_{AP,pe}$[(d)]	$\varepsilon_{AP,R}$[(d)]	$\varepsilon_{AP,CP}$[(d)]
				Southwest								
6. OLS	$AP^{SW}_{SU,t}$	9191.1	287.99	-9512.6	-23.16	.4254	.20			.154	-.051	-.185
		(2.89)	(.81)	(.64)	(.71)	(1.80)						
RLS	$AP^{SW}_{SU,t}$	8929.9	773.1	-9376.6	-62.5	.4119	.12			.5	-.05	-.5
		(2.20)				(1.55)						
ME	$AP^{SW}_{SU,t}$	9348.4	540.38	-9869.5	-45.86	.4188	.18	.92	.32	.332	-.053	-.367
		(2.95)	(2.55)	(2.23)	(2.48)	(1.94)						
7. OLS	$AP^{SW}_{F,t}$	21220.0	-537.28	-25404.0	-6.35	.1838	.75			-.549	-.125	-.051
		(2.47)	(1.16)	(1.08)	(.18)	(.58)						
RLS	$AP^{SW}_{F,t}$	4113.0	633.9	-10164.3	-62.0	.6538	.55			.5	-.05	-.5
		(1.78)				(4.58)						
ME	$AP^{SW}_{F,t}$	8274.5	188.26	-12584.0	-42.20	.6799	.69	6.28	.37	.087	-.062	-.340
		(1.33)	(.78)	(2.56)	(2.28)	(2.91)						
				South								
8. OLS	$AP^{SO}_{SP,t}$	750.2	-115.09	-2907.0	20.57	.7799	.95			-.126	-.018	.179
		(.23)	(.24)	(.15)	(.29)	(7.19)						
RLS	$AP^{SO}_{SP,t}$	3640.9	585.9	-8015.8	-57.5	.6441	.91			.5	-.05	-.5
		(3.85)				(13.27)						
ME	$AP^{SO}_{SP,t}$	3253.8	340.32	-8728.6	-51.76	.8294	.95	1.51	.37	.265	-.054	-.450
		(1.44)	(2.35)	(2.28)	(2.25)	(14.98)						
9. OLS	$AP^{SO}_{SU,t}$	1015.4	-8.48	-1542.1	-.76	.8524	.63			-.067	-.038	-.022
		(1.33)	(.06)	(.12)	(.04)	(2.67)						
RLS	$AP^{SO}_{SU,t}$	801.2	162.3	-2033.4	-17.2	.8139	.63			.5	-.05	-.5
		(1.21)				(5.40)						
ME	$AP^{SO}_{SU,t}$	1262.1	108.11	-2072.0	-15.89	.8517	.62	.91	.42	.316	-.051	-.463
		(1.86)	(2.07)	(2.05)	(2.39)	(5.02)						
10. OLS	$AP^{SO}_{F,t}$	1050.9	123.49	2749.2	-18.58	.4304	.72			1.312	.185	-1.435
		(1.84)	(2.25)	(1.11)	(2.73)	(2.59)						
RLS	$AP^{SO}_{F,t}$	136.9	60.3	-742.9	-6.5	.8633	.54			.5	-.05	-.5
		(.39)				(4.46)						
ME	$AP^{SO}_{F,t}$	1018.9	53.39	-612.9	-9.60	.6131	.63	7.10	.42	.446	-.041	-.741
		(2.67)	(2.62)	(1.68)	(3.93)	(4.59)						
				All Other								
11. OLS	$AP^{AO}_{SU,t}$	8458.6	392.29	2793.1	-19.72	.6757	.98			.109	-.05	-.074
		(2.08)	(1.05)	(.26)	(1.85)	(3.69)						
RLS	$AP^{AO}_{SU,t}$	-42499.8	2072.0	-30737.1	-132.4	2.2348	.95			.5	-.05	-.5
		(10.22)				(18.28)						
ME	$AP^{AO}_{SU,t}$	9743.4	514.41	-6528.9	-21.67	.6277	.98	8.54[(e)]	.09	.126	-.011	-.082
		(2.58)	(1.50)	(.74)	(2.17)	(3.78)						

Note: Figures in parentheses are absolute t-statistics.

(a) OLS refers to ordinary least squares; RLS to restricted least squares; and ME to mixed ordinary least squares.

(b) The critical value of χ^2_3 at the .95 level of significance is 7.81.

(c) θ_p refers to the share of information used to form estimates that is attributable to the prior information.

(d) Elasticities are calculated at the mean level of data, 1954-1977.

(e) The probability of greater noncentral χ^2_3, with noncentrality parameter ½, for this equation is .088.

sumed to be known with complete certainty (see Brook and Wallace [1]). Mixed estimation incorporates uncertain prior knowledge; but the degree of uncertainty is regulated by the subjective variances associated with the prior point estimates on the constrained coefficients. Results are given in Table 1. In general, the pattern of results is varied. Explanatory power, in terms of R^2, ranges from .02 for the Southwest spring crop estimated by RLS to .98 for the All Other summer crop estimated by OLS. In most cases, if R^2 is low the amount of variability in the dependent variable also is low.

The manner in which the incorporation of prior information affects regression results can be seen in Table 1. Various degrees of multicollinearity, a persistent problem in many economic statistical analyses, are present in the set of equations presented. The unique effect of a given variable may be difficult to distinguish from the effect of another with which it is highly correlated. Observation of the simple correlations among explanatory variables shows that in virtually all cases the expected price and cost variables are very highly correlated. The average correlation between these variables for the 11 equations is .90. Other explanatory variables have high correlations in certain instances. The correlation is as high as .96 for expected price and lagged acreage, .90 for risk and lagged acreage, and .74 for costs and lagged acreage. Correlations between risk and cost of production do not contribute enough to multicollinearity problems to warrant concern.

Among the pure OLS estimates, four coefficients on expected price have a negative sign, four risk coefficients are positive, and one cost coefficient is positive. As a result of the imposition of constraints on these coefficients through mixed estimation, the estimated effects of the associated explanatory variables are more appealing intuitively. In the mixed OLS results all coefficients have the proper sign.

The set of pure OLS estimates contains many statistically insignificant coefficients, also probably because of multicollinearity. As a result of the application of mixed estimation, a generalized least squares procedure, the standard errors of parameter estimates are reduced and, hence, t-statistics are increased. Within the set of mixed coefficients, only three have t-statistics less than one. The increase in precision among the estimates is substantial in many instances.

The chi-square statistic given in Table 1 is a test of the hypothesis that prior notions about the coefficients are not contradicted by the data under analysis, i.e., a test of compatibility between sample and *a priori* information [7]. For 10 of the 11 equations estimated, the chi-square test indicates compatibility at the .95 level of significance. Compatibility by this test implies unbiasedness of the stochastic restrictions implied by the prior information. In the one case in which unbiasedness is rejected, equation 11, a noncentral chi-square test of strong mean square error (MSE) superiority of mixed estimates indicates that the mixed estimates are superior in the strong MSE sense.[4] The researcher must decide whether to accept bias in order to achieve increased precision. In the case of equation 11, the mixed result has added appeal because of the change in sign on the risk coefficient between the pure and mixed results.

Review of the set of elasticities obtained shows that in most cases the pure and mixed results are similar. For the equations in which all signs are acceptable in pure estimation, all elasticities are within two standard deviations, as implied by the subjective variances, of the prior point estimates. In most cases, even if one or more signs in the equation are implausible, the elasticities given by coefficients with plausible signs are within two standard deviations of their expected values. The elasticities given by the mixed results all conform to prior expectations.

SUMMARY

Eleven equations are estimated to explain quarterly and regional acreage planted of fresh tomatoes. All equations are estimated by pure, restricted, and mixed least squares regression. Of these 11, only three of the pure OLS equations conform to the expected pattern of signs and magnitudes of coefficients. In contrast, mixed estimation results in plausible signs and magnitudes of parameter estimates in all cases. It appears reasonable, on the basis of these results, to consider mixed estimation a feasible alternative to pure regression in the estimation of relations if prior information is available that can be formulated confidently as linear probabilistic constraints on regression coefficients.

[4] For an estimator $\hat{\theta}$ to be strong MSE superior to another estimator, $\hat{\theta}^*$, the MSE of any linear combination of the elements of $\hat{\theta}$ must be less than the MSE of the corresponding linear combination of $\hat{\theta}^*$. See Yancey, Judge, and Bock [10].

REFERENCES

[1] Brook, R. and T. D. Wallace. "A Note on Extraneous Information in Regression," *Journal of Econometrics*, Volume 1, 1973, pp. 315-316.

[2] Hammig, M. D. "Supply Response and Simulation of Supply and Demand for the U. S. Fresh Vegetable Industry," unpublished Ph D dissertation, Washington State University, 1978.

[3] Jesse, E. V. and M. J. Machado. *Trends in Production and Marketing of California Fresh Market Tomatoes*, Division of Agricultural Sciences Bulletin No. 1871, University of California, 1975.

[4] Lin, W. "Measuring Aggregate Supply Response Under Instability," *American Journal of Agricultural Economics*, Volume 59, 1977, pp. 903-907.

[5] Nerlove, M. "Estimates of the Elasticities of Supply of Selected Agricultural Commodities," *Journal of Farm Economics*, Volume 38, 1956, pp. 496-509.

[6] Nerlove, M. and W. Addison. "Statistical Estimation of Long-Run Elasticities of Supply and Demand," *Journal of Farm Economics,* Volume 40, 1958, pp. 861-880.
[7] Theil, H. "On the Use of Incomplete Prior Information in Regression Analysis," *Journal of the American Statistical Association,* Volume 58, 1963, pp. 401-414.
[8] Theil, H. and A. S. Goldberger. "On Pure and Mixed Statistical Estimation in Economics," *International Economics Review,* Volume 2, 1961, pp. 65-78.
[9] Traill, W. B. "Risk Variables in Econometric Supply Response Models," paper presented at American Agricultural Economics Association annual meetings, State College, Pennsylvania, 1976.
[10] Yancey, T. A., G. G. Judge, and M. E. Bock. "A Mean Square Error Test When Stochastic Restrictions are Used in Regression," *Communications in Statistics,* Volume 3, 1974, pp. 755-768.