A HYPERBOLIC TANGENT YIELD FUNCTION OF FLORIDA CITRUS

Lan Chen (*University of Florida*) and Charles B. Moss (*University of Florida*)

Presented at the American Agricultural Economics Association Meetings

in Portland, Oregon 2007

A HYPERBOLIC TANGENT YIELD FUNCTION OF FLORIDA CITRUS

Lan Chen (*University of Florida*) and Charles B. Moss (*University of Florida*)

Abstract: This study models Florida citrus production as a function of the age profile of a given tree stock. The age relationship is estimated using a modified hyperbolic tangent function and the parameters is solved by Spatial Process Models and Maximum Likelihood approach. The estimation is based on the production data of four citrus varieties in 25 regions of Florida from 1992 to 2005. The results show smooth "S"-shaped yield curves of Florida citrus. This analysis offers yield function of citrus as the first step for statistical modeling of the risks associated with citrus cancers aimed at pricing insurance rates.

Keywords: yield function, citrus, hyperbolic tangent, insurance, spatial autoregressive.

This study initiates the estimation of the effect of diseases such as Citrus Canker and Citrus Greening, and other natural calamities such as hurricanes by formulating an "S"-shaped yield function for citrus and incorporating the spatial dimension of such effects on citrus yields. Citrus Canker is a bacterial citrus disease that causes premature leaf and fruit drop. Remaining fruit can be unmarketable or much less valuable. Most commercial citrus varieties in Florida are susceptible to this disease, especially lime and grapefruit. Citrus canker is highly contagious and has several ways of transmission. It is mainly spread by human contact and wind driven rain, which possess a spatial nature of transmission. Citrus Greening (also known as huangblongbing or yellow dragon disease) is a bacterial disease that reduces citrus production. This disease is spread by the Asian citrus psyllid. Further, while most crops are susceptible to weather events, the effect that hurricanes had on the citrus industry in Florida in 2004. Specifically, Florida's citrus groves were directly affected by three hurricanes in 2004 (Charley, Francis, and Gene)

(See Figure 1). Of these storms Charley and Francis had the most severe implications for the citrus industry with the storm paths crossing over one Florida County.

Given the perennial nature of citrus production, modeling the effect of both disease and weather risks are somewhat more complex than for annual crops such as corn or soybeans. Unlike the yield functions for annual crops, the effect of disease and weather must incorporate the dynamic aspects of production. In the case of Citrus Canker and Citrus Greening, the disease infects the tree by reducing tree yield over time. In the case of hurricanes, the event usually has an immediate effect with dropped fruit from the weather event, as well as a long-term effect caused by the destruction of citrus trees. As an initial effort to address this time-dependent nature of citrus production we estimate a time dependent yield function where the yield per tree is initially small, but increases at an increasing rate, then increases at a decreasing rate before reaching a maximum yield per tree at around 20 years of age.

Methods and Procedures

Optimal Control Setup

We assume that the various decisions a grove manager makes follow the basic optimal control formulation. In this formulation, a manager determines the level of input usage that maximizes the expected value of profit through time. In the vernacular of optimal control, the fertilizer and other variable inputs are the control variables which are varied to control the level of state variables through time (Kamien and Schwartz 1991). In our formulation, producers optimize

$$\max_{u(t)} \int_0^T F(t, x(t), u(t))dt$$

$$s.t. \quad x'(t) = g(t, x(t), u(t)) \qquad (1)$$

$$x(t_0) = x_0$$

where $F(t, x(t), u(t))$ is the discounted profit function ($F(t, x(t), u(t)) = p(t)y(t, x(t), u(t)) - C(t, x(t), u(t), w(t))$ where $y(t, x(t), u(t))$ is the citrus yield, $p(t)$ is the output price, and $C(t, x(t), u(t), w(t))$ is the cost of production which is a function of input prices $w(t)$), $x(t)$ is the state function (the bearing area of the citrus tree), $u(t)$ is the control variable (such as fertilizer), and $g(t, x(t), u(t))$ is the equation of motion which depicts the growth in bearing area over time.

Critical to the optimal control problem is the idea that the level of the state variable cannot be instantaneously varied over time, but is only changed by variations in the control variable. Further complicating our formulation, however, is the fact that our control variable (the bearing area of each citrus tree) cannot be observed. Given this consideration, we start by focusing on the yield component of the optimal control model in Equation (1) $y(t, x(t), u(t))$. As a further simplification, we assume that the state variable (bearing area) is largely a function of tree age

$$x(t = v) = x_v = h(v) \qquad (2)$$

where v is the age of tree cohort. Replacing x with h in $y(t, x(t), u(t))$, the yield function y becomes a function of bearing age of trees as

$$y_v = y(v) = y(t = v, x(t = v), u(t = v)) \qquad (3)$$

Yield Function of Florida Citrus

Zanzig, Moss, and Schmitz (1998) recognized that perennial crops including citrus demonstrate two production characteristics that are critical to understanding the underlying economics: the first characteristic is the existence of a gestation period before the tree bears fruit; the second general characteristic is that the tree is not uniformly productive over its bearing years. Casual observation in the case of citrus suggests that: during the early bearing years, growth is relatively low, and changes from one year to the next are small; at some point, however, changes in growth increase but level off at some maximum yield; growth remains stable for a long period of time until at some age growth begins to decline (Zanzig, Moss and Schmitz 1998).

Given these characteristics of perennial crop production, we assume that the form of the age-yield relationship follows an “S” shape. Because the hyperbolic tangent function provides an ideal “S”-shaped functional form for modeling perennial crop production, the yield function is approximated with a transformed hyperbolic tangent function proposed by Zanzig, Moss, and Schmitz (1998). The formal representation of average yield can be written as

$$y(\beta_{\max}, \beta_0, \beta_1, v) = \frac{\beta_{\max}}{2}(1 + \tanh(\beta_0 + \beta_1 v)) \tag{4}$$

where $\beta_{\max}$, β_0 , and β_1 are estimated parameters, $y(\beta_{\max}, \beta_0, \beta_1, v)$ is the yield of each citrus tree, and v is the tree age. The transformed hyperbolic tangent function yields an “S”-shaped function. The range of the hyperbolic tangent function is (-1, 1). Thus, in our formulation, the range of citrus yields is $y(v) \in (0, \beta_{\max})$. The remaining parameters shift the location and relative slope of the sigmoid shaped graph. In its original formulation, the hyperbolic tangent has an inflection point at $v = 0$. Hence, it increases at an increasing rate until $v = 0$ and then decreases at a decreasing rate.

Unfortunately we do not have tree-level yield data. Instead, we aggregate the yield of each citrus variety over each county based on this formulation as

$$\hat{y}_{it} = \sum_{v=a}^{z} y\left(\beta_{\max}, \beta_0, \beta_1, v\right) T_{itv} \tag{5}$$

where $\hat{y}_{it}$ is the estimated county level yield and T_{itv} is the number of each tree age v in county i at time t. In this formulation, $y(\beta_{\max}, \beta_0, \beta_1, v)$ is the expected yield for a particular age cohort.

Spatial Autoregressive Model

Given yield age relationship and spatial nature of several random events effecting citrus production such as freezes, hurricanes, and disease outbreaks (including both Citrus Canker and Greening), these parameters are estimated using Spatial Autoregressive Models using the maximum likelihood approach.

We will begin with the non-linear regression model with spatial autoregressive disturbances. We assume that spatial autocorrelation only exists in disturbance term but not in dependent variables. The spatial structure is written as:

$$\begin{aligned} y &= f(x, \beta) + \varepsilon \\ \varepsilon &= \rho W \varepsilon + \mu \end{aligned} \tag{6}$$

where y is the county-level average yield for a given citrus variety, $F(x, \beta)$ is the nonlinear model of citrus yields presented in Equation 5, β are the estimated parameters, x is a vector of exogenous factors including the tree age variable v and the numbers of trees in each age cohort is the share of trees in each county by age cohort T_{itv}, ρ is the spatial autoregressive coefficient, and μ is an identically and independently distributed error term ($\mu \sim N\left(0,\sigma^2\right)$) (Anselin 1988, Livanis et al. 2006). The W matrix is the spatial weight matrix that is determined by the specific location of counties. We define that if two counties are contiguous the corresponding cell of the two counties in W matrix sets into 1; otherwise, the cell sets into 0. In this formulation we usually assume that $\rho \in [0,1]$ with $\rho = 0$ representing the standard ordinary least squares model. It is mathematically, possible for ρ to be less than zeros, it raises some empirical questions. Further, following the intuition from Generalized Least Squares, we note that the true spatial formulation is always at least as efficient as ordinary least squares. However, Estimated Generalized Least Squares is not guaranteed to more efficient than ordinary least squares since estimating the heteroscedasticity process introduces some error. Hence, if $\rho > 0.10$ we are more confident that adjusting for spatial autocorrelation improves the efficiency of estimation.

Based on Equation 6, we derive the error term which is a function of production, spatial coefficient, independent variables and βs as:

$$\begin{aligned} y &= f(x,\beta) + (1-\rho W)^{-1}\mu \\ \mu &= (I-\rho W)(y - f(x,\beta)) \end{aligned} \tag{7}$$

The likelihood function for the specification in Equation 7 can be expressed as

$$\begin{aligned} L = &-\frac{N}{2}\ln(2\pi) + \ln|I-\rho W| - \frac{N}{2}\ln|\Omega| \\ &-\frac{1}{2}\left[(1-\rho W)(y - f(x,\beta))\right]' \Omega^{-1}\left[(1-\rho W)(y - f(x,\beta))\right] \end{aligned} \tag{8}$$

Maximization Equation 8 with respect to Ω, for a given ρ, yields

$$\hat{\Omega}_{MLE}(\beta,\rho) = \frac{(y - f(x,\beta))'(I - \rho W)'(I - \rho W)(y - f(x,\beta))}{N} \quad . \tag{9}$$

Substituting this result back into the log-likelihood function Equation 8 we get the concentrated likelihood function with respect to ρ :

$$L_C = -\frac{N}{2}\ln(2\pi) + \ln\left|I - \rho W\right| - \frac{N}{2}\ln\left|\hat{\Omega}_{MLE}(\beta,\rho)\right| - \frac{1}{2}N \tag{10}$$

Given that the eigenvalues of W matrix can be written as w_i , and that

$$\ln\left|I - \rho W\right| = \sum_{i=1}^{N}\ln(I - \rho w_i) \tag{11}$$

the final expression for the concentrated log-likelihood function is given by:

$$L_c = -\frac{N}{2}\ln(2\pi) + \sum_{i=1}^{N}\ln(I - \rho w_i) - \frac{N}{2}\ln\left|\frac{\left[(1-\rho W)(y - f(x,\beta))\right]'\left[(1-\rho W)(y - f(x,\beta))\right]}{N}\right| - \frac{N}{2} \tag{12}$$

Finally, we estimate this likelihood function for the same set of counties over several years so that the likelihood function becomes

$$L_c = -\frac{NR}{2}\ln(2\pi) + R\sum_{i=1}^{N}\ln(I - \rho w_i) - \frac{N}{2}\sum_{t=1}^{R}\ln\left|\frac{\left[(1-\rho W)(y_t - f(x_t,\beta))\right]'\left[(1-\rho W)(y_t - f(x_t,\beta))\right]}{N}\right| - \frac{RN}{2} \tag{13}$$

where R represents years, and N is number of observations (counties) for each year.

Data and Estimation

The Florida Agricultural Statistical Service conducted the state's complete citrus tree census survey as of January every two years since 1966 and the results of the census survey are presented in the Commercial Citrus Inventory and the Citrus Summary. We abstracted the number of trees by variety, county and year set from Commercial Citrus Inventory, and the total production by county and variety from Citrus Summary for this research. However, because Commercial Citrus Inventory is published every two years from 1966 while Citrus Summary is reported every year, data on number of trees in odd years had to be estimated to match with data on production. Assuming proportion relationship between the total bearing trees and that at each age is constant for two successive years, the quantities of trees in odd years can be estimated by Citrus Inventory of previous year and corresponding Citrus Summary. For example, for estimating tree number in 1993, first we mark trees as age from 0 and added total bearing tree number with the tree number at age 2 in 1992's Citrus Inventory, resulting in estimated total bearing tree number for 1993. Second, calculate ratio of total bearing trees which equals to total bearing tree number from 1993 citrus summary divided by estimated total bearing tree number resulted from last step. Third, the first 23 categories of bearing trees in modified 1992 citrus inventory were remarked as age from 1 to 23, and the rest after age 23 was aggregated as one group at age 24. The next step is multiplying ratio of bearing trees with number of bearing trees older than 2 years which resulted from step 3, resulting in estimated bearing tree number data for 1993 from age 3 to 24. Finally bearing tree number are transformed into percentage of trees dividing by total bearing tree number, as the same as production, for emphasizing the weight of trees at individual age among total bearing trees.

The gestation period before the tree bears fruit is set from tree age 0 to age 2, which means that citrus trees start to bear fruit from age 3. The range of yield-age profile is from age 3 to 24. We select four citrus varieties' data in 25 Florida counties from 1992 to

2005: Early and Midseason Oranges, Valencia Oranges, White Seedless Grapefruit, and Colored Seedless Grapefruit. All-round orange and all grapefruit category are skipped to avoid multicollinear problem. Since the share of seedy grapefruit is relatively small, it was eliminated from data set.

Spatial weight matrix W for Equations 11 and 13 are determined by the specific location of counties. In this study, we assume that Spatial Autocorrelation only exists between two contiguous counties because the closer the locations of two counties the more significant the spatial effect. Thus, the W matrix used in this study becomes

$$W = \begin{bmatrix} 0 & 1 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \cdots & 0 \end{bmatrix} \qquad (14)$$

where the cells of two counties are set to be 1 if the two counties are contiguous; otherwise the cells are set into 0.

The maximization of equation 13 generates estimated parameters βs and ρ. We also set constraints on the second-degree derivative of yield function with respective to bearing age v which is negative at tree age 24 and positive at 3. These settings make sure growth of average yield increases fast at early bearing years and but slow down after certain point.

Empirical Results and Discussion

The empirical results show that citrus has an “S”-shaped yield function and significant spatial coefficient ρ. The parameter estimates and related statistics for the yield function are presented in Tables 1.

White seedless oranges exhibit the largest maximum average yield per tree which is 4.47 boxes and Valencia the smallest, 3.11 boxes. β_0s are all negative which means that the minimum yield starts from value less than half of $\beta_{\max}$. Small positive β_1s reflect yield curve is increasing slowly. The spatial coefficients range from 0.48 to 0.64 and all are statistically different from zero at the 95% level, which verify that the spatial correlation has significant impacts on the estimation. Given that all the estimated parameters pass t-test except β_0 of white seedless oranges, the final estimated parameters are shown in Table 2.

The shapes of yield function for each variety are presented in Figure 2 through 5 which also show the estimated boxes of fruit per tree by USDA based on official end-of-season production estimates and the number of bearing trees indicated by the citrus tree inventory surveys. Although the estimation done by USDA does not have enough points to show the shape of yield, the trend and bounds of two curves match well. Several critical points coincide.

The estimated yield functions exhibit an "S" shaped curve which increases with a positive second derivative during the early bearing years and once reaching a certain age the growth rate decreases rapidly and the yield levels off. For example, Figure 2 shows that early and midseason oranges' estimated yield goes up rapidly with a concave curve until age 10. After that the curve becomes convex and then flat, ending with a maximum average yield of around 4.5 boxes. Other varieties have the similar characteristics of yield curve except white seedless grapefruit. The reason for white seedless grapes' non-"S" shape maybe involves that hurricane with citrus canker in 2004-2005 impact yields so much that characteristics of curve could not be captured from data.

Deviation of Yield

Although we assume that all the counties have the same yield function for the same variety, the deviations of counties differ much with each other. The average deviation for each county $\bar{\varepsilon}_i$ were calculated as

$$\hat{y}_{it} = \sum_{it=1}^{22} \hat{\delta} * T_{it}$$
$$\hat{y}_{it} - y_{it} = \varepsilon_{it} \qquad (15)$$
$$\bar{\varepsilon}_i = \frac{1}{T} \sum_{t=1}^{14} \varepsilon_{it}$$

Theoretically, $\bar{\varepsilon}_i$ s should converge to zero, but Figure 6 shows that some counties present deviations close to zero such as Hardee, Palm Beach and Seminole while others possess a large deviation like Hendry or one with an opposite direction like Hillsborough and Glades. The deviations of average yield range from Highland's 2337.147 thousand boxes to Hillsborough's negative 816.25 thousand boxes. The reason for this huge difference is various. One of the reasons may be the impact of tree density, because tree density is different among counties and impacts average yield per tree. Other reasons might be relative to the different locations of counties.

Summary and Conclusion

It is complicated to describe the production decisions for perennial crops due to several factors including the time-specific nature of yields. There is a proportional relationship between changes over time in the yields of perennial crops and the size of the trees in the case of citrus. This research provides a descriptive framework for analyzing the yield age relationship for Florida citrus. In particular, the focus of the work describes the response of the average yield of Florida citrus per tree to the age of trees. Further, spatial effect produced by factors from contiguous counties on yield is simulated under spatial autoregressive model and maximum likelihood estimation providing clear advantages than OLS.

The results indicate that yield curve depicts a clear "S" shape for Early and Midseason orange, Valencia Orange and Colored Seedless grapefruit. The possible explanation for the convex shape of White Seedless grapefruit may be involve even more significant influence produced by hurricane and citrus canker on its' yield. From the results, it is also indicated that spatial effects could not be ignored in the process of simulation. All estimated spatial coefficients are significant and larger than 0.4 within 1, which could include spatial effects of factors such as weather, temperature, catastrophic events, and human mobility. This would be useful for future studies that involve these factors.

This study provides the basic structure of citrus yield function that would be helpful for analyzing impacts of random events such as freezing, hurricanes and disease outbreak including citrus canker. Take citrus canker for example. Citrus canker mainly was not a severe problem that impacted citrus production until 2004. However, because infected citrus will not be eradicated any more, citrus canker can be added into model as a factor impacting betas in form like $(1+\tau Z)\beta$ where τ is canker coefficient and Z is index of citrus canker. Another factor impacting citrus yield is tree density. Brown noted that historic boxes per tree may be higher than future tree yields due to increasing tree densities, so that the projections based on tree yields may overstate future production. As for insurance program design, the next step is to simulate conditional probability density function of average yield τ . Mean of τ has been estimated by this research and its variance may be of interest for future study.

References

Andrews, D. "A Note on the Unbiasedness of Feasible GLS, Quasi-Maximum Likelihood, Robust, Adaptive, and Spectral Estimators of the Linear Model." *Econometrica*. 54(3) (May, 1986): 687-698.

Anselin, L. *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic, 1988.

Zanzig, B. R., C. B. Moss, and A. Schmitz. "The 'S'-shaped Yield Curve of Florida Valencias." *Staff Paper Series* SP 98-8(1998).

Grigorios Livanis, Charles B. Moss, Vincent E. Breneman, and Richard F. Nehring. "Urban Sprawl and Farmland Prices" *American. Journal of Agricultural Econonomics* 88(4): 915–929. (2006)

Kamien, M.I. and Schwartz, N.L.. *Dynamic Optimization: The Calculations of Variations and Optimal Control in Economics and Management*. Second Edition. North-Holland , 1991.

United States Department of Agriculture, *Commercial Citrus Inventory*. Florida Department of Agricultural and Comsumer Service, Tallahassee, Fl, (1992-2005).

United States Department of Agriculture, *Citrus Summary*. Florida Agricultural Statistics Service, Tallahassee, Fl, (1992-2005).

Table 1—Parameter Estimates and Relative Statistics

Early& Midseason	estimation	Standard deviation	T-value
β_{max}	4.472	0.247	18.142
β_0	-1.778	0.653	-2.723
β_1	0.235	0.102	2.300
ρ	0.640	0.048	13.255
Valencia oranges			
β_{max}	3.111	0.133	23.468
β_0	-1.041	0.328	-3.177
β_1	0.177	0.049	3.618
ρ	0.576	0.052	11.140
White Seedless			
β_{max}	5.476	0.229	23.960
β_0	-0.463	0.370	-1.251
β_1	0.139	0.064	2.162
ρ	0.556	0.056	9.920
Colored Seedless			
β_{max}	4.961	0.216	22.997
β_0	-0.984	0.225	-4.370
β_1	0.175	0.038	4.582
ρ	0.482	0.060	8.064

Table 2—Final Estimation

	Early and midseason	Valencia	White seedless	Colored seedless
β_{max}	4.472	3.111	5.476	4.961
β_0	-1.778	-1.041	0.000	-0.984
β_1	0.235	0.177	0.139	0.175
ρ	0.640	0.576	0.556	0.482

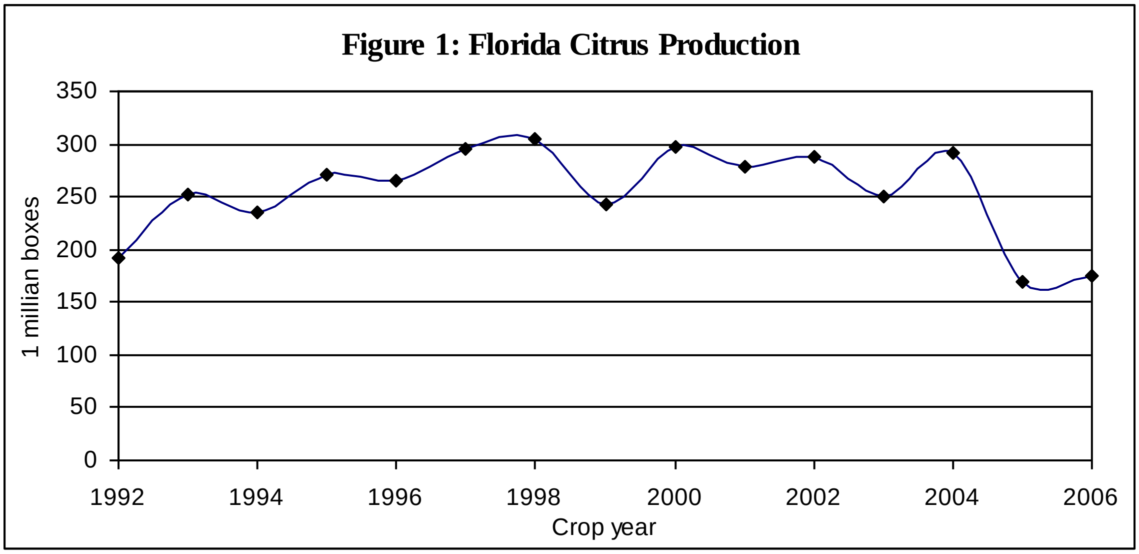
Figure 1: Florida Citrus Production
350
300
250
200
150
100
50
0
1 millian boxes
1992
1994
1996
1998
2000
2002
2004
2006
Crop year

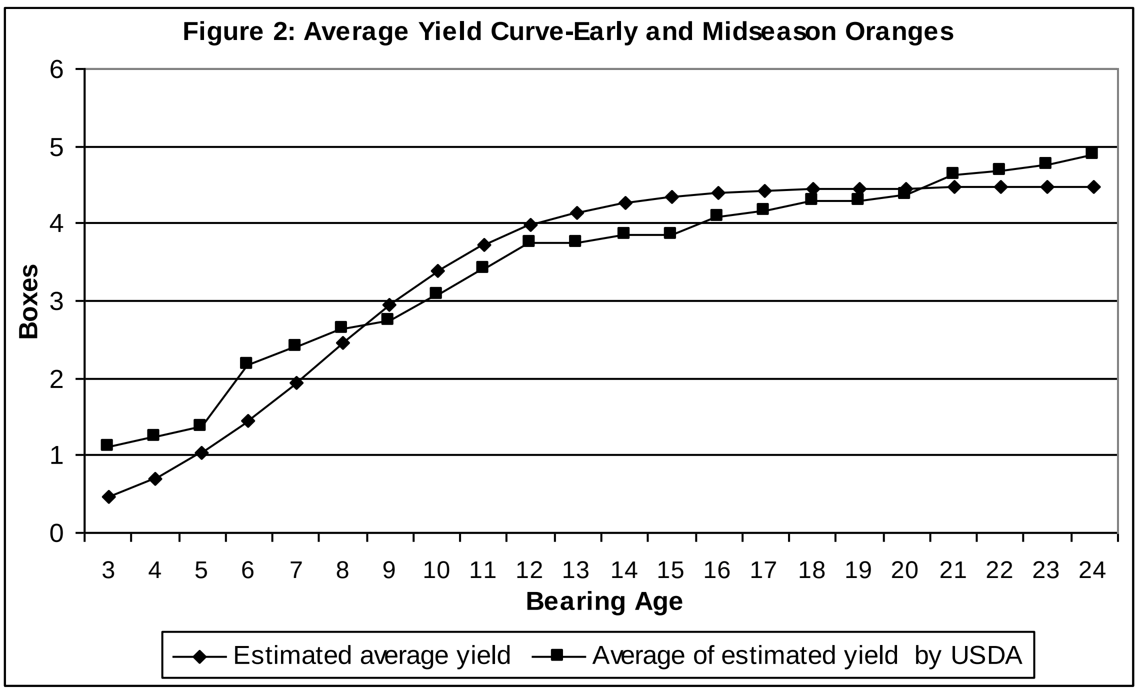
Figure 2: Average Yield Curve-Early and Midseason Oranges
6
5
4
3
2
1
0
Boxes
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Bearing Age
Estimated average yield
Average of estimated yield by USDA

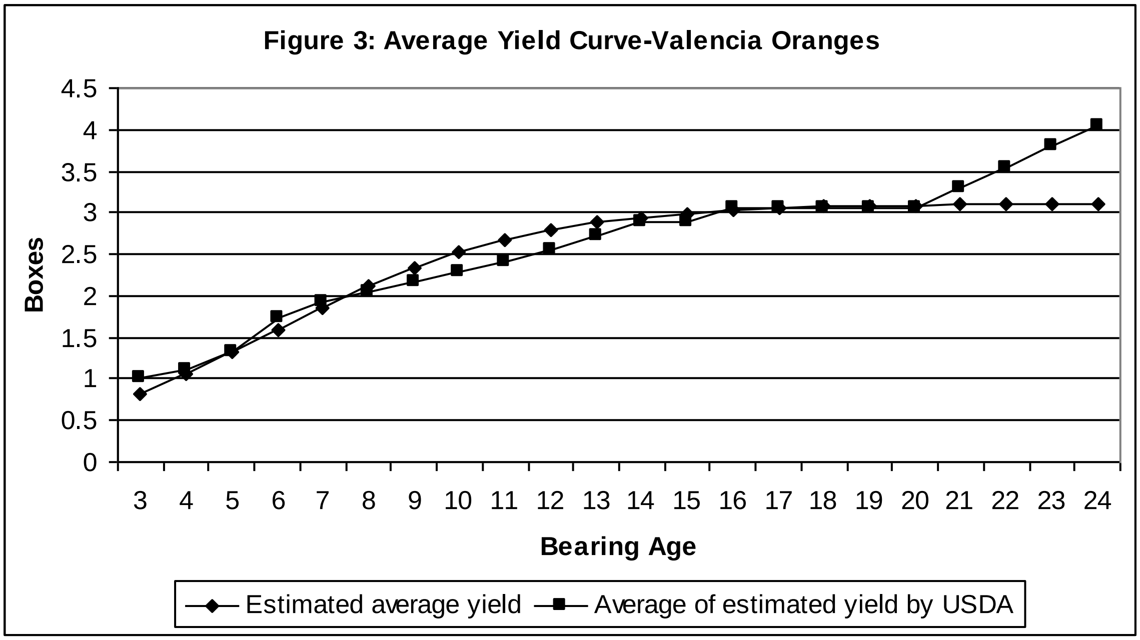
Figure 3: Average Yield Curve-Valencia Oranges
Boxes
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Bearing Age
Estimated average yield
Average of estimated yield by USDA

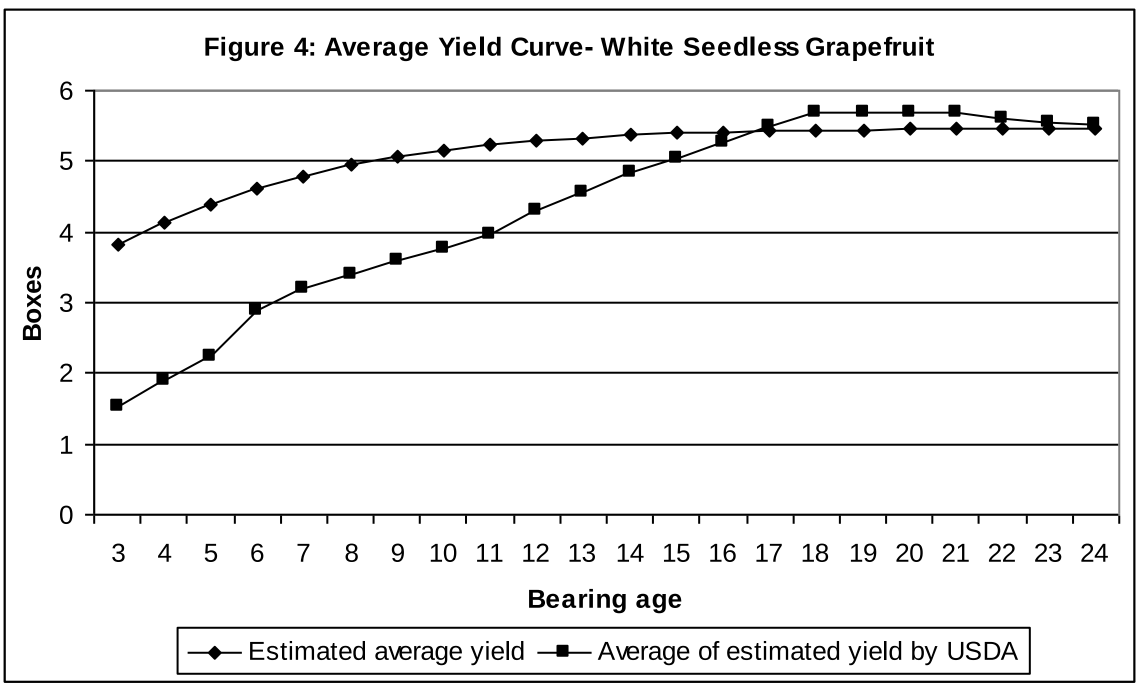
Figure 4: Average Yield Curve- White Seedless Grapefruit
Boxes
6
5
4
3
2
1
0
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Bearing age
Estimated average yield
Average of estimated yield by USDA

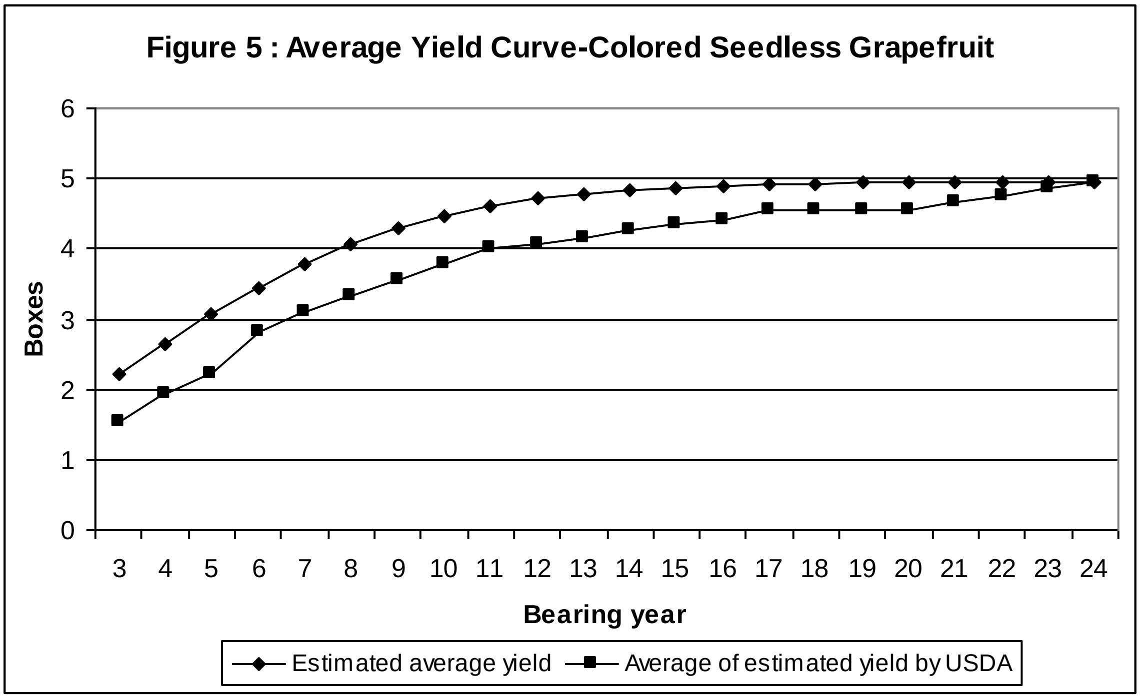
Figure 5 : Average Yield Curve-Colored Seedless Grapefruit
Boxes
6
5
4
3
2
1
0
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Bearing year
Estimated average yield
Average of estimated yield by USDA

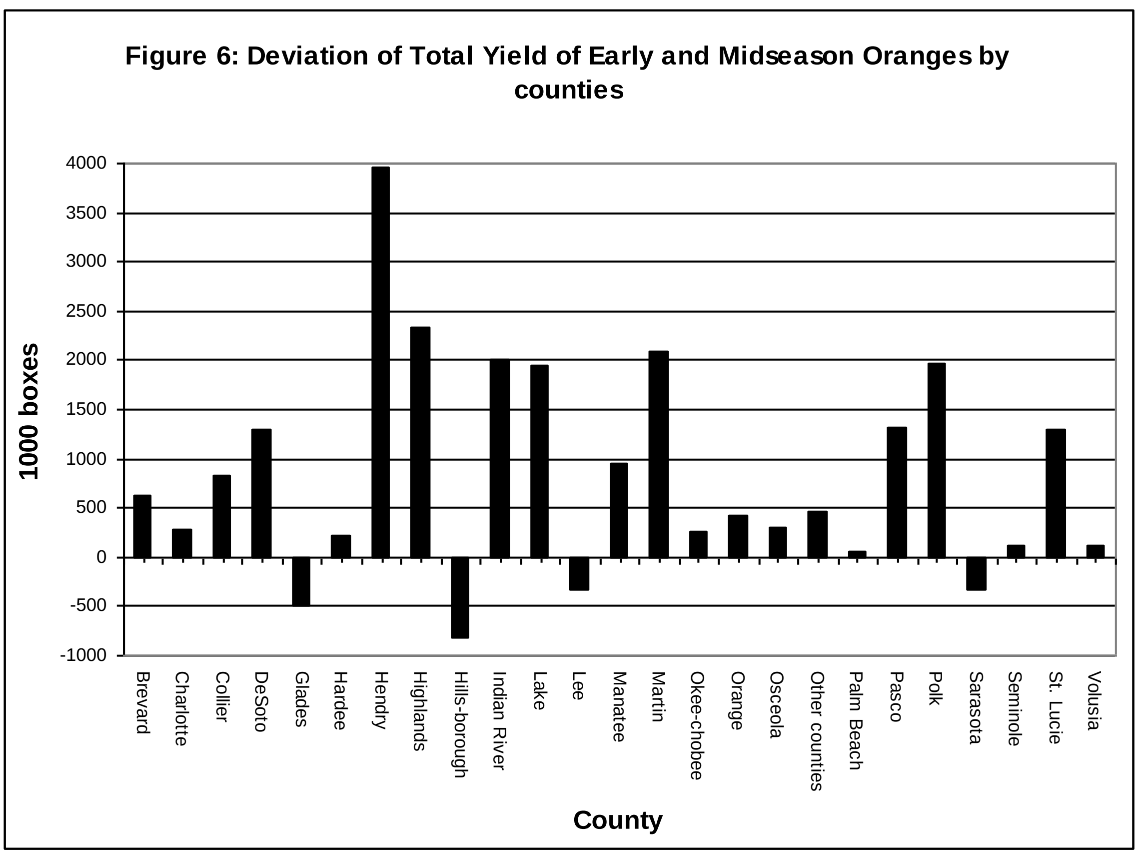
Figure 6: Deviation of Total Yield of Early and Midseason Oranges by counties
1000 boxes
4000
3500
3000
2500
2000
1500
1000
500
0
-500
-1000
Brevard
Charlotte
Collier
DeSoto
Glades
Hardee
Hendry
Highlands
Hills-borough
Indian River
Lake
Lee
Manatee
Martin
Okee-chobee
Orange
Osceola
Other counties
Palm Beach
Pasco
Polk
Sarasota
Seminole
St. Lucie
Volusia
County