Modeling Perennial Crop Supply: An Illustration from the Pecan Industry

Abdelmoneim H. Elnagheeb and Wojciech J. Florkowski*

Abstract

Two methodological approaches were applied to estimating the number of non-bearing trees in the absence of such data using data for the Southern USA pecan industry. The first approach distinguished between bearing and non-bearing phases of a tree life and directly estimated the number of non-bearing trees. The second focused on indirect estimating of the non-bearing tree number from changes in production. This approach relaxed the assumption of maintaining maximum yields for infinite period as used in earlier studies. Empirical applications used two data sets from the pecan industry. The comparison of empirical results suggested that the first method was more accurate than the alternative approach in predicting the number of newly planted trees over an extended period of time. Additional data collection will allow for further application of available methodology to the pecan industry.

Key words: non-bearing trees, pecans, tree crop, tree yields, volume produced.

Introduction

The literature on perennial crop supply has developed over the last two decades. The first attempt to model perennial crop supply was French's study of the Michigan and U.S.A. apple production. Later, French and Bressler developed a supply response model for lemons specifying planting and removal relationships. French and Matthews developed a structural multi-equation model for the supply response of asparagus. The equations were subsequently combined to give a reduced-form, single equation for output. Wickens and Greenfield modeled the investment and harvest decisions separately for the Brazilian coffee. However, the separate equations were combined to give a reduced-form, single-output equation which used a lagged-variable technique for estimating the model.

Some researchers advocate the structural approach for estimating separate equations for planting, removal, and harvesting (Akiyama and Trivedi; Bellman and Hartley; French, King, and Minami; and Hartley, Nerlove, and Peters). However, a common difficulty in estimating a structural model for perennial crops is data availability. For many perennial crops, data on the number (or area) of new plantings, tree age-distribution, and removals are not available. Because of data limitations the reduced-form, single-equation approach has dominated perennial crop-supply studies.

This paper utilizes and compares two alternative approaches that mitigate the paucity of data on new plantings. The two models, which are described below, were estimated using data from Georgia’s pecan industry. Pecans are an important

* A. H. Elnagheeb is a postdoctoral associate and W. J. Florkowski is an associate professor. Both are at the department of agricultural and applied economics, The University of Georgia College of Agricultural and Environmental Sciences, Griffin, GA.

Copyright 1993 Southern Agricultural Economics Association
source of income for many farmers in the Southeast. Georgia produces about 40 percent of the annual total pecan output in the U.S. Hence, it is important to understand how pecan supply responds to different policies. Previous pecan studies focused on pecan quality at the retail level (Williams, et al.), predicting pecan prices (Epperson and Allison), estimating pecan price flexibilities (Wells, et al.), evaluating the impact of pecan crop forecasts on pecan prices and value (Shafer) and differences in pecan prices by variety (Okunade and Cochran). This study expands the studies of the pecan industry by providing estimates of the pecan supply response.

The Model

Previous studies on perennial crop supply faced the problem of lack of data on new plantings and tree age-distribution. To overcome this problem, researchers modeled the unobserved variables as functions of observable variables (e.g. expected prices, costs, and output). These equations were substituted into other structural equations to obtain a single-equation, reduced-form model (Bateman; Behrman; French and Matthews; and Wickens and Greenfield).

The current study faces the similar data problem for new pecan plantings and the age distribution of trees. Two models were constructed. Estimates were compared and the performance of each model evaluated. In both models, the number of newly planted trees at time \( t \), \( I_t \), is assumed to be a linear function of lagged prices, \( P_{t-1} \), and costs, \( C_t \):

\[
I_t = \alpha_0 + \alpha_1 P_{t-1} + \alpha_2 C_t
\]

The first model uses a study by French and Matthews as its basis. The French and Matthews' (FM) model starts with the identity:

\[
B_t = B_{t-1} + aN_{t-k} + R_{t-1}
\]

where \( B_t \) is the total bearing acreage, \( N_t \) is new plantings in acres, \( R_{t-1} \) is bearing acreage removed at \( t-1 \), \( a \) is a constant with a value slightly less than one to account for plantings removed before reaching bearing age, and \( k \) is the number of years necessary for a tree to reach a bearing age.

The FM model formulates desired new plantings as a function of expected long-run profits and a disturbance term. Such formulation reduces the problem of limited data availability on new plantings and removals. A function accounting for removals included a short-run profit expectations measure, a measure of acreage exceeding the age at which productivity typically begins a significant decline, and a disturbance term. New plantings and removal functions replaced explicit measures in identity (2). Because nonlinear forms for the new planting and removal functions will yield an equation difficult to estimate empirically, the FM model assumes linear functional forms. French and Willett expressed all variables, including \( B_t \) and \( B_{t-1} \), in the log form as an approximation for an unknown nonlinear relationship.

The approach proposed in this study is similar to the FM's approach. Pecan tree life encompasses bearing and non-bearing phases. A study by Gorman et al. shows pecan trees as bearing at the age of six years. Thus, the following relationship was formulated:

\[
NB_t = \sum_{i=0}^{5} I_{t-i} + e_t
\]

where \( NB_t \) is the number of non-bearing trees at time \( t \), \( I_t \) is the number of trees planted at time \( t \), and \( e_t \) is an independently identically distributed error term with zero mean and a constant variance \( \sigma^2 \). The error term, \( e_t \), can be construed as the error in measurement. Equation (3) states that at time \( t \), the group of non-bearing trees consists of trees from zero to five years old. It was assumed that no removal occurs in the non-bearing group. The removal could be incorporated into equation (3) by multiplying the \( I_{t-i} \) by \( a \), where \( a \) is slightly less than one, as used in the FM's model. However, \( a \) will be unidentifiable when the model derived below is estimated. In the case of pecans, the removal of trees occurs seldom and therefore \( a \) should be very close to one.
Substituting equation (1) into equation (3) and rewriting, we obtain:

\[ NB_t = 6\alpha_0 + \alpha_1 P_t^* + \alpha_2 C_t^* + e_t \]

where,

\[ P_t^* = \sum_{i=0}^{s} P_{t-1-i} \]  
\[ C_t^* = \sum_{i=0}^{s} C_{t-i} \]

Equation (4) states that the number of non-bearing trees is influenced by past pecan prices and costs. Equation (4) can be estimated by ordinary least squares or generalized least squares, if the error terms are not homoscedastic, yielding consistent and unbiased estimates of the unknown parameters.

The second approach benefits from the models developed by Bateman; Behrman; Baritelle and Price; and Wickens and Greenfield. In those studies, authors measured output as the summation of the product of different bearing age tree groups (or acres) and the yield per bearing tree (acre)

\[ Q_i = \sum_{i=0}^{n} y_i l_{t-i} \]

where \( Q_i^* \) is the total potential output, \( y_i \) is the yield per tree (or per acre) for trees in age group \( i \), and \( l_{t-i} \) is the number of trees (or acres) in age group \( i \) at time \( t \). Baritelle and Price used \( I_{i,t} \) as the number of trees, while Bateman used \( I_{i,t} \) to denote the number of bearing acres. More specifically, Bateman’s model (p. 389) assumes that \( n \) equals infinity and yield is represented by a two-step process as described by a yield function:

\[ Q_t = b_1(\sum_{i=0}^{x-1} I_{t-i}) + b_2(\sum_{i=x}^{m} I_{t-i}) \]

where \( Q_t^* \) is as defined above, \( b_1 \) the output per acre attained after the first increase in yield, \( b_2 \) the output per acre after reaching the plateau, \( k \) is the age at which trees begin to bear, and \( s \) is the year in which the second distinct increase in yield occurs. The assumption made by Bateman is that plants maintain peak yields for an infinite length of time.

Because the quantity harvested may differ from that produced, Bateman allowed for the effects of economic and climatic factors, by incorporating producer’s price \( (P_t) \) and climatic variables, \( W_t \) (e.g., rainfall and humidity) into equation (7):

\[ Q_t = b_1(\sum_{i=0}^{x-1} I_{t-i}) + b_2(\sum_{i=x}^{m} I_{t-i}) + cW_t + dP_t \]

where \( Q_t \) is the amount harvested at time \( t \). A problem with equation (8) is the infinite horizon which presents estimation difficulties. To overcome this problem, Bateman took the first difference of equation (8) to obtain:

\[ \Delta Q_t = b_1(\sum_{i=t-k}^{x-1} I_{t-i}) + (b_2-b_1)I_{t-3} + c\Delta W_t + d\Delta P_t \]

where \( \Delta \) indicates a change in the respective variable from one period to the next. Since data on newly planted acreage were not available, Bateman expressed \( I_t \) as a function of prices and costs and substituted that equation into equation (9). In this study, \( I_t \) represents the number of trees rather than the number of acres.

In this paper, equation (8) is modified to:

\[ Q_t = \gamma_0 + \gamma_1 Q_t^* + \sum_{i=1}^{m} \beta_i P_{t-i} + \beta_{m+1} C_t + \mu W_t + \phi Q_{t-1} + u_t \]
where $Q_t$, $P_t$, $C_t$, and $W_t$ are as defined before; $Q_t^*$ is the potential total output as given by equation (6); $u_i$ is independently and identically distributed error term with zero mean and constant variance, $\sigma^2_u$. The differences between equations (8) and (10) include the addition of the term $\Phi Q_t$ into equation (10) and the scaling of $Q_t^*$ term with the parameter $\gamma$, which was set equal to one in equation (8). The lagged dependent variable in equation (10) was added to capture the alternate bearing effect which characterizes many perennial crops.

\[
\Delta Q_t = \gamma_1 \Delta Q_t^* + \sum_{i=1}^{m} \beta_i \Delta P_{t-i} + \beta_{m+1} \Delta C_t + \mu' \Delta W_t + \Phi \Delta Q_{t-1} + \Delta u_t
\]

where

\[
\begin{align*}
\alpha_0^* &= \gamma_1 \alpha_0 (y_6 + \sum_{i=7}^{13} \mu_i) \\
\alpha_i^* &= \gamma_1 \alpha_i, & i=1,2 \\
\hat{p}_i &= y_6 P_{t-7} + \sum_{i=7}^{13} \mu_i P_{t-1-i} \\
\hat{C}_t &= y_6 C_{t-7} + \sum_{i=7}^{13} \mu_i C_{t-1-i}
\end{align*}
\]

Data on yield, $y$, were obtained from Gorman, et al. and were used to calculate the transformations in equations (14).

**Estimation Procedures**

Equations (4) and (13) represent the reduced forms of the two models to be estimated. For the model in equation (4), data on non-bearing pecan trees are available every five years. That is, the index on $NB$ in equation (5) is $i+i$ where $i = 0, 5, 10, 15, \ldots$. Because the error term, $e_i$, was assumed to be independently and identically distributed (iid), equation (4) can be estimated with ordinary least squares (OLS). However, initial results indicated the presence of serial correlation and the model was re-estimated using Prais-Winsten’s method to correct for serial correlation.

For ease of exposition, the model in equation (13) is rewritten as:

\[
M_t = \alpha_0^* + X_t' \pi + \Phi M_{t-1} + \varepsilon_t
\]
where $M_t = Q_t - Q_{t-1}$, $\varepsilon_t = u_t - u_{t-1}$, and $X_t$ contains all other explanatory variables in equation (13) and $\pi$ is a conformably defined vector of parameters. Applying OLS to equation (15) encounters two problems. First, the explanatory variables contain a lagged dependent variable. The OLS estimates are consistent but may suffer from small sample bias (Johnston). The second problem is the serial correlation of the error terms which will result in inconsistent OLS estimates (Johnston). Although the $u_t$'s were assumed iid, the error terms, $\varepsilon_t$'s, are not:

\begin{align*}
E[\varepsilon_t | \varepsilon_{t \neq s}] &= -\sigma^2 \\
&= \begin{cases} 
2\sigma^2 & s = 0 \\
-\sigma^2 & s = \pm 1 \\
0 & |s| \geq 2
\end{cases}
\end{align*}

For estimation purposes, the model in equation (15) is rewritten as:

\begin{equation}
W_t = \alpha_0^* + X_t^t\pi + \phi M_{t-1} + W_{t-1}
\end{equation}

where $W_t = Q_t - u_t$. Repeated substitution for $W_{t-1}$ into equation (17) allows us to write equation (17) as:

\begin{equation}
W_t = \alpha_0^* \sum_{i=1}^{t} 1 + \sum_{i=1}^{t} X_t^t\pi + \phi \sum_{i=0}^{t-1} M_i + W_0
\end{equation}

Substituting for $W_t = Q_t - u_t$ and $\sum_{i=0}^{t-1} M_i = \sum_{i=0}^{t-1} (Q_t - Q_{t-1}) = Q_{t-1} - Q_{t-1}$, leads to:

\begin{equation}
Q_t = \alpha_0^* t + \sum_{i=1}^{t} X_t^t\pi + \phi Q_{t-1} + W_0 + \phi Q_{t-1} + u_t
\end{equation}

Define:

\begin{equation}
Z_t = \sum_{i=1}^{t} X_t \\
\lambda = W_0 + \phi Q_{t-1} = Q_t - u_t + \phi Q_{t-1}
\end{equation}

Now, the model in equation (19) can be written as:

\begin{equation}
Q_t = \lambda + \alpha_0^* t + Z_t^t\pi + \phi Q_{t-1} + u_t
\end{equation}

In equation (21), $\lambda = Q_t - u_t + \phi Q_{t-1}$ is assumed to be a constant. OLS applied to equation (21) will now result in consistent estimates of parameters (Johnston).

Data and Empirical Results

The empirical specification of the model in equation (4) is given by:

\begin{equation}
NB_t = 6\alpha_0 + \alpha_1 P_i^* + \alpha_2 C_i^* + \varepsilon_t
\end{equation}

where $NB_i$ is the number of non-bearing pecan trees in 1,000 and $P_i^*$ and $C_i^*$ are as defined in equations (5).

Using equations (19) - (21), the empirical model of equation (13) is given by:

\begin{equation}
Q_t = \lambda + \alpha_0^* t + \sum_{i=1}^{t} \hat{P}_i + \alpha_2 \sum_{i=1}^{t} \hat{C}_i + \sum_{i=1}^{m} \beta_1 \sum_{i=1}^{m} \Delta P_{i-j} + \beta_2 \sum_{i=1}^{m} \Delta C_i + \phi Q_{t-1} + u_t
\end{equation}
where $Q_t$ is harvested pecans in 1,000 pounds, $P_t$ is the average annual price of pecan in cents/pound; $C_t$ is the index of input prices paid by farmers; $\alpha_0$, $\alpha_1$, $\alpha_2$, $\hat{P}_t$, and $\hat{C}_t$ are as defined in equations (14). Weather variables (annual rainfall and temperature) were included in the initial model but dropped from the final specification because their coefficients were statistically insignificant.

A series of Georgia Agricultural Facts provided data on pecan prices and quantities and the number of non-bearing trees. Agricultural Statistics supplied the series regarding the index numbers of input prices. National Agricultural Census provided data on bearing and non-bearing pecan trees which were reported every five years. The data for the first model (equation 22) cover the period 1924 to 1987 with a five year gap, while data for the second model (equation 23) cover the period 1962-1987.

The model in equation (22) was estimated by the Prais-Winsten’s method because initial results suggested first order autocorrelation as indicated by the Durbin-Watson’s statistic. The results are presented in table 1. The coefficients on the price and cost variables were both significantly different from zero as indicated by the $t$-statistics. The signs were as expected. The number of non-bearing trees increased with an increase in last year’s pecan prices, while it decreased with an increase in the index of input prices paid by farmers.

### Table 1. Parameter Estimates of the First Model (equation 22)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Coefficient estimate</th>
<th>$t$-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant X 6</td>
<td>$6\alpha_0$</td>
<td>150.940</td>
<td>0.670</td>
</tr>
<tr>
<td>$P_t^*$</td>
<td>$\alpha_1$</td>
<td>2.9655$^c$</td>
<td>3.203</td>
</tr>
<tr>
<td>$C_t^*$</td>
<td>$\alpha_2$</td>
<td>-0.17032$^c$</td>
<td>-2.162</td>
</tr>
<tr>
<td>Rho$^a$</td>
<td>$\rho$</td>
<td>0.82154$^c$</td>
<td>4.992</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson$^b$</td>
<td></td>
<td>2.375</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Rho is the autocorrelation coefficient. Prais-Winsten’s method was used to estimate the parameters.

$^b$ D-W statistic is for the transformed residuals.

$^c$ denotes significance at the 5 percent level.
The second model (equation 23) was estimated by OLS and results were corrected for heteroscedasticity (White). The value of the Durbin's h-statistic indicated no serial correlation could be detected (table 2). The value of \( m \) (lag length - see equation 23) was empirically determined. The value for \( m \) that gave the best statistical results was 3. The high \( R^2 \) indicated a good explanatory power of the model.

The \( \alpha \) coefficients, together, reflect the indirect impact of past plantings on the quantity supplied. This impact was significant as indicated by the corresponding \( t \) statistics. The signs of the \( \alpha \) coefficients were as expected. An increase in the lagged pecan price positively influenced current plantings and hence, had an indirect positive influence on the supply of pecan. On the other hand, an increase in the index of input prices had a negative effect on new plantings and consequently, an indirect negative effect on pecan supply. The \( \beta \) coefficients represent the direct effects of pecan prices and the index of input prices on pecan supply. The interpretation of these coefficients should be according to the original model of equation (10). Increases in lagged pecan prices had positive effects on pecan supply. However, the greatest effect of current price on supply occurred after two years (\( \beta_2 \) is the

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Coefficient estimate</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>( \lambda )</td>
<td>106020.000&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8.869</td>
</tr>
<tr>
<td>( t )</td>
<td>( \alpha_0 )</td>
<td>-7346.200&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-1.873</td>
</tr>
<tr>
<td>( P_0 )</td>
<td>( \alpha_1 )</td>
<td>32.995&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.015</td>
</tr>
<tr>
<td>( C_0 )</td>
<td>( \alpha_2 )</td>
<td>-2.775&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-1.803</td>
</tr>
<tr>
<td>( P_{t-1} )</td>
<td>( \beta_1 )</td>
<td>338.34</td>
<td>1.165</td>
</tr>
<tr>
<td>( P_{t-2} )</td>
<td>( \beta_2 )</td>
<td>1624.600&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5.720</td>
</tr>
<tr>
<td>( P_{t-3} )</td>
<td>( \beta_3 )</td>
<td>319.830</td>
<td>0.954</td>
</tr>
<tr>
<td>( C_t )</td>
<td>( \beta_4 )</td>
<td>-9.448</td>
<td>-0.231</td>
</tr>
<tr>
<td>( Q_{t-1} )</td>
<td>( \phi )</td>
<td>-0.680&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-6.578</td>
</tr>
</tbody>
</table>

\( R^2 = 0.86 \)

Durbin's h-statistic = 0.261

Note: Results are corrected for heteroscedasticity (White).

<sup>a</sup> For variable definitions see text and equations (14) and (23).

<sup>b</sup> denotes significance at 10% level.

<sup>c</sup> denotes significance at 5% level.
statistically significant coefficient). The direct effect of the index of input prices on pecan supply was negative as expected. The coefficient of the lagged dependent variable was negative and significantly different from zero. The result suggests the presence of the alternate bearing phenomenon in pecan production. An increase in the supply of pecans last year lowered supplies in the current year.

**New Plantings Prediction**

In order to predict new plantings, the estimates of the structural parameters of equation (1), the $\alpha$ parameters, must be recovered. Plugging these estimates back into equation (1) and using values of the observed price and cost variables, allows the prediction of new plantings.

The first model (equation 22) gives direct estimates of the $\alpha$ parameters ($\alpha_i$ is obtained by dividing the intercept estimate by 6). However, the second model (equation 23) gives estimates of the transformations of the $\alpha$ parameters as described in equations (14). From equations (14) one could calculate the estimates of the $\alpha$ parameters if the value of $\gamma_i$ is known. In the following analyses $\gamma_i$ was assumed to be equal to one. That is, it was assumed that all potential output would be harvested, an assumption which seems to be reasonable in case of pecans.

The only report on new pecan plantings was for the year 1929 (Jones et al.). A survey of over two million trees of improved varieties reported that about 109,000 (or 5.4 percent of the total number of pecan trees) were planted in 1929. However, the estimate of the total number of trees reported by Jones et al. was 49 percent higher than the National Agricultural Census estimate for the same year.

Using the $\alpha$ estimates from the first model (equation 22), the predicted new planting was 77,816 trees in 1929. This estimate is equal to about 71 percent of the estimate reported in Jones et al. The second model (equation 23) predicted planting of 280,230 trees or about 257 percent of the estimate reported by Jones et al. However, if Jones et al. overestimated new plantings by the same percentage as for the total number of trees, then the adjusted estimate of new trees is 73,154. After this adjustment, the first model predicted about 106 percent and the second model about 383 percent of Jones et al.’s estimate of new plantings in 1929. Moreover, while the Jones et al. estimate was for the improved pecan varieties only, the estimates in this study are for both improved varieties and seedlings. This may explain why both models predicted larger new plantings than reported by Jones et al. after adjustment particularly because at that time the common knowledge suggested planting of several cultivars including seedlings for improved pollination.

Because non-negativity constraints were not placed on either model, negative prediction of new plantings were possible due to the linear functional form of the planting equation (equation 1). However, negative predictions of new plantings were only among the results obtained using the second model. All predictions (for the period 1924 to 1987) from the first model were positive.

The above results suggest that the first model performed better in terms of predicting the number of new plantings. The first model is simpler and more direct than the second model. Hence, results indicate that simple approaches to model pecan new planting may outperform the complex ones, especially if detailed data are not available.

**Concluding Comments**

This study focused on improving methods of estimating the number of non-bearing trees, modifying and adopting the models to specific data limitations, and tree crop characteristic. The modeling of the tree crop planting response such as tree nuts is difficult due to gaps in data on the number of trees, new additions to existing orchards and removal of trees. The first of the proposed methods accounted directly for new plantings and the non-bearing and bearing growing stage. The second approach considered the dynamic yield-tree age relationship allowing for yield increases as a tree matures.
The application of both approaches to the pecan industry provided empirical illustration. The majority of the parameter estimates were statistically significant. The first approach more accurately estimated the pecan planting response in case of 1929 data than the model based on the second approach.

From practical standpoint, the first approach contains the basic information readily available to a grower as it includes past prices and a measure of production costs. This approach, less complex, predicted a positive number of new plantings confirmed by the actual observation of the developments in the pecan industry over time. Therefore, the first approach may be used in generating information for making investment decisions. The second approach incorporates more information by combining the harvesting and investment decisions and by incorporating the dynamic yield-tree age relationship. However, predictions of new plantings obtained using the second approach were occasionally negative.

The models offer a methodological framework for estimating the planting response of the pecan industry. The empirical applications of the models could be expanded as more data become available. Progress in data reporting for pecan industry has been slow but steady as illustrated by initiating pecan storage reports in the past and including Arizona and California as individual areas in pecan price and production reports. The importance of data needs and progress in understanding pecan tree physiological mechanism will balance efforts in developing methodological framework enhancing the ability to provide increasingly accurate information needed for the practical decision-making.

References

*Agricultural Statistics*, USDA, various issues.


Nerlove, M. "Distributed Lags and Estimation of Long-Run supply and Demand Elasticities: Theoretical Considerations." J. Farm Econ. 40:301-310.


