Functional Form Model Specification: An Application to Hedonic Pricing

Jeff E. Brown and Don E. Ethridge

A combination of conceptual analysis and empirical analysis—partial regression and residuals analysis—was used to derive an appropriate functional form hedonic price model. These procedures are illustrated in the derivation of a functional form hedonic model for an automated, econometric daily cotton price reporting system for the Texas-Oklahoma cotton market. Following conceptualization to deduce the general shapes of relationships, the appropriate specific functional form was found by testing particular attribute transformations identified from partial regression analysis. Minimizing structural errors across attribute levels and estimation accuracy were used in determining when an appropriate functional form for both implicit and explicit prices was found.

Functional form (i.e., mathematical form or model structure) is critical in determining an accurate and consistent econometric model. Statistical procedures such as Box-Cox regression tend to rely on purely empirical results rather than drawing on knowledge about the actual physical, biological, or economic processes involved in specifying a relevant functional form. Empirical analysis alone cannot replace conceptual reasoning when estimating relationships of most economic phenomena.

Prior studies have examined model specification and functional form using a variety of methods and covering disparate areas. Some investigations have been in the development of production models (e.g., Carrasco-Tauber and Moffitt; Lichtenberg and Zilberman), probit and logit models (Ozuna *et al.*; Davidson and MacKinnon; Godfrey; Yatchew and Griliches; Pace and Gilley), hedonic models in housing (Can; Burgess and Harmon; Kang and Reichert) and hedonic models in cotton (Ethridge and Davis; Bowman; Ethridge *et al.*, 1992).

The functional form of hedonic models can be difficult to conceptualize. First stage hedonic models are different from any general pricing models where price is determined by general supply and demand factors. Hedonic models determine implicit prices of all recognized levels of quality characteristics embodied in a product on the basis of the utility or productivity of the characteristics. The conceptualization of the appropriate functional form is further complicated by the fact that the choice of the functional form cannot be determined *a priori*, given the lack of a theoretical basis (Can).

The cotton industry recognizes and values nine quality characteristics that influence the explicit price of cotton. The ability to conceptualize all relationships between and among attributes in the production or pricing process is difficult. Conceptual relationships can only provide a general shape of a price-quality relationship. A thorough empirical investigation is necessary to identify the specific market price-quality relationships and the appropriate functional form.

This paper presents the procedures used in the development of a functional form for the estimation of a daily hedonic price model for cotton. The objective of the research was to derive an efficient hedonic price model that accurately identifies implicit prices of attributes on a daily basis. The purpose of this paper is to describe the techniques and procedures used in that process and present the results. The techniques are derived from procedures advocated by Belsley et al., Godfrey, and Neter *et al.*, but extend beyond general concepts to their application in hedonic modeling. The approach is useful in other hedonic applications because functional form specification is a critical issue in all hedonic studies; if the functional form is not correct, derived implicit prices are not correct.

The first section of this paper describes the de-

The authors are Research Associate and Professor, Department of Agricultural Economics, Texas Tech University.

The authors acknowledge the contribution of Dale Shaw for providing the data for the analysis reported in this paper. They also thank Emmett Elam, Darren Hudson, Phillip Johnson, Sukant Misra, Dick Owens, Eduardo Segarra and an anonymous reviewer for their suggestions on the manuscript.

This research was supported by Cotton Incorporated and the Texas State Support Committee. Texas Tech University, College of Agricultural Sciences and Natural Resources Manuscript No. T-1-386.

velopment of a daily hedonic cotton market pricing model, the conceptual rationale for its structure, and a problem that suggested an improper functional form specification. The second section discusses the functional form tests, including the data and sample used in the application, the conceptual procedures, and use of partial regression analysis in finding an appropriate functional form. The remaining sections provide results of the partial regression analysis, the procedures used in examining various functional forms, and the appropriate functional form developed from the selection method of variable transformations which confirms both the conceptual hypothesis and the partial regression analysis.

Development of a Daily Hedonic Cotton Pricing Model

The market price for cotton may be viewed as a collection of implicit prices determined by the quantity and quality of cotton and recognized by the industry in the form of premiums and discounts (U.S. Dept. of Agriculture). General supply and demand forces determine the general level of prices for any given period of time (day, month, year, etc.). Supply and demand factors for individual attributes likewise affect the market valuation of attributes within any period of time. The mix of cotton quality attributes produced and their usefulness in the textile production process, individually and in combination with other attributes, determine the implicit prices of cotton quality attributes.

Attribute pricing by the market relies on the accuracy and appropriateness of the quality information. The system of cotton quality information in the U.S., administered and maintained by the U.S. Department of Agriculture (USDA), measures a group of quality characteristics of cotton fibers on each bale of cotton produced. The quality attributes measured are trash (i.e., leaf, bark, grass content), color (i.e., reflectance and yellowness), fiber length, strength, micronaire (a measure of fiber fineness and maturity), and length uniformity of the fiber. This information is available to all participants trading the cotton. The reliability of the quality information is widely accepted. However, this quality information must be translated into price information for efficient market performance.

Prior studies using hedonic models to estimate implicit prices in cotton have found linear models to be effective when aggregated price-quality data were used (Ethridge and Davis; Hembree et al.) and non-linear models for less aggregated data (Ethridge and Mathews; Ethridge et al., 1977; Ethridge and Neeper; Ethridge et al., 1992; Bowman and Ethridge). These models provide understanding for a *daily* hedonic model for estimating implicit price relationships in cotton.

However, the daily model is also influenced by (1) conceptualization of relationships between various quality attributes and price and (2) expected interactions between/among attributes. Because each of the fiber quality attributes contributes to the production of a textile product—through manufacturing efficiency and/or quality dimensions of the product—each attribute is hypothesized to be non-linear in its relationship with price (makes a diminishing marginal contribution to the end product). The desirable attributes (length, strength) are hypothesized to contribute to price at a decreasing rate. The undesirable attributes (foreign matter, color) are hypothesized to decrease price at an increasing rate (a "mirror-image" of a desirable attribute). Interactions among attributes are based on understanding of textile manufacturing relationships. The conceptual rationale is described in greater detail in Ethridge *et al.* (1992).

Data for the daily model were obtained from two computerized electronic spot markets operating in Texas and Oklahoma. Accessing these marketing systems provided daily observations of individual sales/purchases on about one-third of all spot market sales and includes purchases from about 75% of buyers purchasing cotton in the region. Data collection and model estimation on a daily basis is performed by accompanying support software, referred to as the Daily Price Estimation System (DPES). Daily estimation was initiated in February, 1989 and continues to the present (see Hudson et al. for description of the DPES).

An initial model structure selected was a double-log (natural log) or multiplicative model. This model was specified with the lot sales price of cotton as a function of six quality attributes, a below grade binary variable, and a regional binary variable:

$$P_i = \beta_0 G1_i^{\beta_1 + \beta_2 BG_i} (G2_i + 1)^{\beta_3} STA_i^{\beta_4} e^{\beta_5 M_i + \beta_6 M_i^2} STR_i^{\beta_7} U_i^{\beta_8} e^{\beta_9 R_i} \epsilon_i \quad (1)$$

where P_i is the price of mixed lot i in cents per pound, $G1_i$ is an average of the first digit of the composite grade code, BG_i is a binary indicator variable denoting below grade cotton ($BG = 0$ if $G1 \leq 7$ and $BG = 1$ if $G1 > 7$), $G2_i$ is an average of the second digit of the composite grade code, STA_i is the average staple length in 32^{nds} inch, M_i

is the average micronaire index reading, STR_i is the average strength reading in grams per tex, U_i is the average length uniformity index, R_i is a binary indicator variable denoting region ($R = 0$ if market reporting region is West Texas and $R = 1$ if market reporting region is East Texas-Oklahoma), and ϵ_i is the stochastic error term.

This functional form is appealing for several reasons: (1) it is a simple mathematical configuration; (2) it allows a pattern of implicit prices to be curvilinear; (3) it allows interaction of all attributes in determining the implicit prices; (4) the coefficients are the price flexibilities of the attributes; and (5) no other standard model form tested (including Box-Cox transformation) provided better empirical results.

A means to test the model's accuracy was needed. The USDA' spot price quotations (DSCQ) are not valid as an indicator of "true" prices and implicit values of attributes because their accuracy has not been established. The DPES model estimates typically had R^2's of 0.85–0.95 for individual days' model runs, which seemed acceptable for cross-sectional data. Another test of accuracy was determining if the DPES contained any systematic error. To examine this, the estimated price (P_e) for each sale of cotton (i) was regressed against the actual price (P_a) for each sale. The model,

$$P_{e_i} = \alpha + \beta P_{a_i} + \epsilon_i \quad (2)$$

should have $\alpha = 0$ and $\beta = 1$ to confirm no systematic error. Any systematic error indicates a consistent over or under-estimation of prices over some range of prices—a bias in the model structure.

This test was conducted for a random sample of days and sales within days. Results (Ethridge *et al.*, 1992) were that $\alpha = -0.27$ (significantly different from 0 at the .05 level) and $\beta = 0.997$ (significantly different from 1 at the .05 level); $R^2 = 0.88$, $F = 1189.3$ (significant at the .0001 level). Thus, evidence of a systematic error suggested an incorrect mathematical form of the hedonic model[1] and posed the problem in development of the DPES model to find a correct functional form. This was the purpose of the research effort reported here.

Functional form specification was further complicated by a change in the grading system in August, 1993 which introduced a modified set of attributes into cotton marketing. The 1993 grading system separated the grade code designations of trash and color into more objective measures of classer's trash (*CTR*), bark levels (*LB* = level 1 and *HB* = level 2), grass levels, reflectance (*Rd*), and yellowness ($+b$). However, the pre-1993 mathematical form still provided a conceptual basis to determine the appropriate functional form for the 1992 marketing year. This matter is addressed later in relation to functional form analysis procedures.

Tests for Functional Form

Conceptual analysis provides the basis for general model form (e.g., relationships increasing at a decreasing rate, etc.) but the other important issue with respect to implicit prices is specific model formulation (e.g., how sharply the curves bend) is an empirical question. The procedures described below draw from the procedures described by Neter *et al.* and Belsley *et al.* The empirical procedures described are based on analysis of residuals performed on alternate functional forms, each developed by considering both partial regression analysis and conceptual analysis (already discussed). Residual analysis conducted in this manner insures the absence of structural error and a functional form model that can accurately represent the pricing structure.

The remainder of this section is organized as follows. The data used in the residuals analysis and partial regressions are discussed. The method of using partial regression analysis to determine the most accurate implicit price relationships (functional form) is then described.

The Data

The data used were daily lot cotton sales from the 1990/91 and 1991/92 cotton marketing years collected in the course of running the DPES. The data set contained over 400 days of sales and was extensive in the range of cotton quality and daily market conditions. The data set included information on quality characteristics of the cotton for the grading system that became effective in 1993 as well as the pre-1993 grading system data. Consequently, the data provided the means to examine functional form for the pricing structure that was approaching rather than being confined solely to a historical relationship that would be obsolete for a current market.

Because the total volume of the data set was so large (400 daily sets of data), making the compu-

[1] Although the systematic error was about half the magnitude of the error associated with USDA's DSCQ, a significant systematic error or any magnitude makes the results unreliable.

tation demands of using data from all days so great, a sample of days was drawn. A random sample of days was selected from the two year period and compared to the entire data set to ensure a true representation of the population. The number of observations in the sample, 5,564 lot sales, represented 10.37% of the entire data set, which contained 53,647 lot sales. Means comparisons indicated the sample was a true representation of the population.

Partial Regression Analysis

Partial regression analysis was used to identify price-quality (attribute) relationships when all other attributes recognized in trading are present. This portion of the analysis helped accomplish two objectives: (1) determine the price-quality relationships that yield the implicit prices and (2) correct for the systematic error in model estimates discussed earlier. Partial regression analysis provides evidence on the empirical nature of the dependent-independent variable relationships. The partial regression analysis regressed both price and attributes against the other attributes in the model and the residuals were retained for each. These residuals were then regressed against each other to reveal (1) the nature of the regression relationship for the attributes under consideration and (2) the marginal importance of that attribute in reducing the residual variance (Neter *et al.*; Belsley *et al.*). The partial regression analysis procedure is presented below, followed by an abbreviated illustration to clarify its use. The regressions were performed using ordinary least squares and residuals calculated for each attribute for each of the sample days as follows:

$$\hat{P}_i(Z_k) = f(Z_{ik}), \tag{3}$$

$$e_i(P|Z_k) = P_i - \hat{P}_i(Z_k), \tag{4}$$

$$\hat{Z}_{ij}(Z_k) = f(Z_{ik}), \tag{5}$$

$$e_i(Z_j|Z_k) = \hat{Z}_{ij}(Z_k), \tag{6}$$

where the caret over a variable indicates the estimated value, P_i is price of cotton for all i mixed lot sales, Z_k is all attributes in the model not under consideration (i.e., excluding attribute Z_j), e_i is the residual term, and Z_j is the attribute under consideration (i.e., excluded from Z_k). The sample days were then combined and the respective residuals regressed against each other using

$$e_i(P|Z_k) = \phi_0 + \phi_1\,(e_i(Z_j|Z_k)) + \phi_2(e_i(Z_j|Z_k))^2, \tag{7}$$

where the Φ's are estimated regression coefficients, which indicate the relationship between price and the attribute, providing evidence of the appropriate transformation that attribute might have when placed in the model. The quadratic specification was used as a means of testing for a non-linear relationship.

To illustrate the partial regression analysis, a two independent variable (attribute) example is provided. Consider the following fully specified model:

$$P_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \epsilon_i, \tag{8}$$

where P_i is the price of cotton for all i mixed-lot sales, α's are the regression parameters, X_{1i} and X_{2i} are the only two attributes that conceptually explain price, and ϵ_i is the residual term.

In partial regression analysis, one attribute at a time is under consideration, (X_{1i} in this example). In the analysis X_{2i} is excluded from the model and the first regression performed and residuals obtained:

$$P_i = \beta_0 + \beta_1 X_{1i} + \lambda_i, \tag{9}$$

where β's are the estimated regression coefficients and λ_i is the error term. These residuals, λ_i, contain any random error in price and the effect of the excluded attribute on price [i.e., $\alpha_2 X_{2i} + \epsilon_i$ from equation 8 (Mosteller and Tukey)]. The next regression identifies any relationship between the attributes:

$$X_{2i} = \theta_0 + \theta_1 X_{1i} + \mu_i, \tag{10}$$

where θ's are estimated regression coefficients and μ_i is the error term. These residuals, μ_i, contain any random error in the excluded attribute and the effect of the excluded attribute not linearly associated with the included attributes. If there is no correlation between these attributes, then θ_1 is not significant and μ_i approximately equals X_{2i}. By contrast, if θ_1 is significant, X_{1i} explains X_{2i} and μ_i is only random error.

Finally, the residual regressions are performed to determine the effect of each attribute by regressing the two series of residuals obtained above (λ and μ) against each other:

$$\lambda_i = \gamma_0 + \gamma_1\mu_i + \gamma_2\mu_i^2, \tag{11}$$

where the γ's are estimated regression coefficients, which indicate the relationship between price and the excluded attribute, providing evidence of the appropriate transformation that attribute might have when placed in the model. This equation determines the shape and significance of the non-random error as related to price and the excluded attribute, given that there is some non-random error. If only random error exists, the attribute does not decrease any error variance and thus should not be included in the model. Used in this manner, partial regression can provide an indication of the direction and rate of change in the slope for each attribute and if variables should be included in the model.

Partial Regression Analysis Results

Partial regression analysis was performed on the attributes as if there were no preconceived knowledge of how the attributes should appear in the model. The partial regression results of the data, including the *t*-value probabilities for each statistic are shown in Table 1.

The analysis indicated that all attributes introduced in this model except length uniformity were significant in explaining the variation in the price of cotton at the 0.05 level of significance. An explanation of the classer's trash attribute (*CTR*) is given as an example in lieu of analyzing each attribute's results. *CTR* should be included in the model in a curvilinear fashion which decreases at an increasing rate. The coefficients were all significant with a positive intercept and a negative slope and slope change, which indicates *CTR*'s price-quality relationship is decreasing at low trash levels and decreasing more rapidly at higher trash levels. This implies that the price-CTR relationship should be concave to the origin, ceterus paribus. The other attributes' results can be interpreted similarly (Table 1).

Table 1. Partial Regression Statistics for Sample[1]

Dependent Variable (Z_j)	Independent Variables: Intercept	Residual	$(\text{Residual})^2$
Classer's Trash	0.1016	−21.1491	−24.5470
(*CTR*)	(0.0056)	(0.0001)	(0.0001)
Reflectance	0.0390	0.5658	−0.0149
(*Rd*)	(0.2408)	(0.0001)	(0.0007)
Yellowness	0.0729	−1.0229	−0.3024
(+*b*)	(0.0346)	(0.0001)	(0.0001)
Staple	0.0582	0.7561	−0.0911
(*STA*)	(0.1084)	(0.0001)	(0.0017)
Strength	0.0895	0.1395	−0.0443
(*STR*)	(0.0206)	(0.0001)	(0.0001)
Uniformity	0.0493	−0.0775	−0.1166
(*U*)	(0.1310)	(0.1163)	(0.0001)
Micronaire	0.3095	3.0408	−3.3214
(*M*)	(0.0001)	(0.0001)	(0.0001)
Level 1 Bark	0.0256	−1.4159	−0.3824
(*LB*)	(0.4872)	(0.0001)	(0.1938)
Level 2 Bark	0.0109	−5.3846	−1.8944
(*HB*)	(0.7298)	(0.0001)	(0.0441)

[1]Using Equation 7 for each attribute and regressing the residual value, $e_i(P|Z_k)$, against residual values, $e_i(Z_j|Z_k)$ and residual values squared, $(e_i(Z_j|Z_k))^2$. Numbers in parenthesis below estimated coefficients are the t-value probabilities.

Functional Form Analysis Procedures

The analysis was performed for both pre-1993 and 1993 grading systems (Brown), but only the 1993 grading system model is illustrated here. The basic double-log model structure of Ethridge *et al.* (1992), along with the partial regression analysis results, provided a starting point for this analysis. Ordinary least squares (OLS) regression was run for each sample day. The residuals were obtained from those models. All sample days' observations were then combined to analyze the functional form across the sample. Residuals were analyzed to determine any inconsistencies, inefficiencies, or inaccuracies in each functional form.

The error terms of each day's sales included in the random sample were regressed against each attribute in the residual analysis. This test revealed any problems in the functional fit of the model. The following evaluations were made: (1) whether the error terms were linear across attribute levels for all attributes in their transformed state; (2) whether the error terms exhibited equal variance across attribute quantities, i.e., homoscedasticity; and (3) whether consistent/non-random patterns of the error terms existed across attribute levels. This analysis provided insight into the cause of the inaccuracy for any particular functional form.

Any significant deviations from the expected patterns of the error terms provided a basis for determining a correct functional form. A selection method of other functional forms or a combination of transformations of selected attributes were tested to correct the previously identified errors. These transformations were made in conjunction with the results of the partial regression analysis and the residual tests of the previously tested functional form. All functional forms were tested in the same manner to determine an appropriate functional form for estimating the daily market price of cotton. Once any pattern in residuals across each attribute was eliminated, an appropriate functional form model was designated. This model contained no systematic error, thus providing reliable implicit price estimates.

Residuals Test Results from Functional Form Search

As noted, the double-log functional form model was estimated for each of the days in the sample and the residuals obtained for each observation in the sample. The initial functional form was:

$$P = b_0 CTR_i^{b_1} Rd_i^{b_2} (+b)_i^{b_3} STA_i^{b_4} STR_i^{b_5} U_i^{b_6} e^{b_7 M_i + b_8 M_i^2 + b_9 R_i + b_{10} LB_i + b_{11} HB_i} \epsilon_i, \quad (12)$$

where Rd_i is an average of the reflectance percentage in mixed lot i, $+b_i$ is an average of the yellowness readings (Hunter's $+b$ calculation), LB_i is the percentage of level 1 bark, HB_i is percentage of level 2 bark, P_i and the other attributes are the same as defined for equation 1. The observations, consisting of residuals from each daily model were combined to perform the residuals analysis on the sample.

The regression statistics for residuals from model 12 regressed against each attribute within the model should be zero if that attribute transformation and/or functional form is correct. If the residual statistics are not zero and are significant, the transformation is incorrect because the residuals are not linear across attribute levels. The residual regression line indicates the direction in which the attribute transformation is not allowing a change in that attribute to accurately track the change in price, *ceteris paribus*. There is no significant relationship between the attribute and the residuals if the transformation accurately tracks changes in price from changes in the attribute, holding all other attributes constant. The residual analysis results for model 12 are shown in Table 2.

Table 2. Regression Statistics for Residual Terms from Model 12 Regressed Against Variables Within the Model[1]

Variable[2]	Intercept	Variable	(Variable)^2
Classer's Trash	−0.0971	−0.2492	−0.1488
(*LCTR*)	(0.0001)	(0.0001)	(0.0001)
Reflectance	−0.3790	0.1800	−0.0214
(*LRd*)	(0.8501)	(0.8501)	(0.8501)
Yellowness	−1.1324	0.9892	−0.2153
[*L*(+*b*)]	(0.0001)	(0.0001)	(0.0001)
Staple	−12.7675	7.3738	−1.0646
(*LSTA*)	(0.0232)	(0.0229)	(0.0226)
Strength	1.3724	−0.8495	0.1314
(*LSTR*)	(0.0979)	(0.0982)	(0.0987)
Uniformity	−28.2059	13.0023	−1.4983
(*LU*)	(0.3637)	(0.3597)	(0.3558)
Micronaire	0.0000	0.0000	0.0000
(*M*)	(1.0000)	(1.0000)	(1.0000)
Level 1 bark	0.0005	−0.0037	0.0034
(*LB*)	(0.8220)	(0.7276)	(0.7250)
Level 2 Bark	0.0009	−0.1402	0.1753
(*HB*)	(0.3316)	(0.0001)	(0.0001)

[1]Example of residuals regressed against the log of classer's trash.

$e_i = -0.0971 - 0.2492(LCTR) - 0.1488(LCTR^2)$

[2]The "L" preceding the attribute abbreviations indicates a natural log transformed variable.

The partial regression results for *STA* indicated a price-*STA* relationship increasing at a decreasing rate, as allowed by the double-log model. However, the residuals analysis indicated that a sharper bend than allowed by the double-log model should exist because a structural error remained. A similar situation was found for *HB* except in an inverse relationship. The log-linear, slightly convex relationship did not allow for enough bend in the price-*HB* relationship.

Evaluation of Model 12

Structural errors existed in four of the nine attributes [*CTR*, (+*b*), *STA*, and *HB*]. The results of the partial regression analysis also suggested that the log transformation of the *CTR* and +*b* attributes were incorrect. The partial regression analysis indicated *CTR* should form a concave price-*CTR* relationship (Table 1), where the log transformation allows only a convex price-*CTR* relationship. The existence of structural error for *CTR* is indicated by the significance of the statistics (Table 2). A similar structural error existed for (+*b*) because of the convex relationship forced by the log transformation, which was also indicated to be concave by the partial regression analysis.

Attribute Transformations

A number of transformations to *CTR*, (+*b*), *STA*, and *HB* were tested to determine an appropriate functional form model. This process allowed their respective price-quality relationships to behave as prior results had indicated. These tests were conducted in the same manner as the testing of model 12. Those functional forms that were tested included transformations on the structurally incorrect attributes. In total, four different functional form models were evaluated before all structural errors were eliminated. The model search for an appropriate functional form consisted of various transformations which are not discussed here (see

Brown for a detailed presentation); however, the correct transformations are discussed below.

The *CTR* attribute was transformed by squaring it to allow for a concave price-*CTR* curve. The partial regression results had indicated that yellowness $(+b)$ should have a concave downward sloping price-$(+b)$ relationship. However, squaring $(+b)$ to force a concave shape over-corrected for the error; an appropriate transformation was found to be log-linear.

STA was transformed in a quadratic form, and forced to have a positive *STA* coefficient and a negative STA^2 coefficient. The residual regression analysis for *HB* indicated its price-*HB* curve should have a sharper bend than the log-linear form allowed. The sharper bend was induced by placing *HB* in the model in a quadratic form, forcing the *HB* coefficient to be negative and the HB^2 coefficient to be positive. The other attributes in the model were unchanged since no structural error was identified and they did not conflict with either the partial regression analysis or prior conceptual knowledge.

These transformations were made over the testing of three functional forms. The first model addressed residual errors in *CTR* and *HB* to determine if CTR^2 and the quadratic transformation of *HB* were correct. The second model dealt with the residual errors in $(+b)$ and *STA*, which found that *STA* in quadratic form was correct, but $(+b)^2$ was not correct. The third functional form tested was as follows:

$$P_i = b_0 e^{b_1 CTR_i^2} Rd_i^{b_2} e^{b_3 (+b)_i + b_4 STA_i + b_5 STA_i^2} STR_i^{b_6} U_i^{b_7} e^{b_8 M_i + b_9 M_i^2 + b_{10} R_i + b_{11} LB_i + b_{12} HB_i + b_{13} HB_i^2} \epsilon_i. \tag{13}$$

Evaluation of Model 13

The residual regression results of model 13, shown in Table 3, indicated all the structural errors across attributes had been corrected. The residual analysis confirmed the absence of patterns in error terms with the attribute transformation, indicating that the model accurately represented implicit attribute prices. Overall, model accuracy for explicit prices was tested using the Ethridge *et al.* (1992) test to insure that the original problem of systematic error had been corrected. A random sample of 325 of the observed population of sales was taken, prices for these sales were estimated using the daily parameters for equation 13 from the 1993 crop and regressed against the actual sale prices as shown in Equation 2. Results were

Table 3. Regression Statistics for Residual Terms from Model 13 Regressed Against Variables Within the Model

Variable	Intercept	Variable	(Variable)^2
Classer's Trash	0.0013	−0.0005	0.0001
(*CTR*)	(0.9272)	(0.9264)	(0.9273)
Reflectance	0.7809	−0.3711	0.0441
(*LRd*)	(0.6852)	(0.6852)	(0.6851)
Yellowness	0.0001	−0.0001	0.0000
$(+b)$	(0.9727)	(0.9249)	(0.9208)
Staple	0.1410	−0.0101	0.0002
(*STA*)	(0.7523)	(0.7163)	(0.6812)
Strength	1.4118	−0.8741	0.1352
(*LSTR*)	(0.0763)	(0.0765)	(0.0769)
Uniformity	−13.5840	6.2926	−0.7286
(*LU*)	(0.6489)	(0.6446)	(0.6403)
Micronarie	0.0000	0.0000	0.0000
(*M*)	(1.0000)	(1.0000)	(1.0000)
Level 1 Bark	−0.0015	0.0112	−0.0105
(*LB*)	(0.4974)	(0.2686)	(0.2574)
Level 2 Bark	0.0000	0.0000	0.0000
(*HB*)	(0.9999)	(0.9991)	(0.9991)

$$Pe = 0.89 + .98(Pa) \tag{14}$$

where the t-values were 1.26 for α and 1.40 for β. Tests of null hypotheses of $\alpha = 0$ and $\beta = 1$ were not rejected at the .001 significance level. This equation also had a coefficient of determination of .94 and F-value significantly different from 0 at the .01 significance level. The model show in Equation 13 was therefore judged to be the most appropriate functional form hedonic model for the daily cotton market in Texas and Oklahoma for the new (1993 and beyond) grading system.

Summary and Conclusions

The primary purpose of this paper was to illustrate a set of procedures, described generally by Belsley *et al.*, Godfrey, and Neter *et al.*, for determining an appropriate functional form hedonic model. A hedonic model for determining daily *implicit* prices for attributes uses both conceptual rationale and empirical procedures for arriving at an improved functional form. Some maintain that hedonic theory provides no guidance on functional form. However, as illustrated here, prior knowledge or expectations about some aspects of economic behavior (production relationships—marginal productivities of attributes—in this instance) may suggest the *general* shape of relationships. Empirical analysis such as that described in this paper may be needed to arrive at a more exact functional form.

The use of partial regression analysis provides initial information about the empirical price-quality relationships found in the data. Other statistical methods of determining the appropriate model consider primarily the accuracy of the estimation and disregard the consistency of the implicit price structures. Partial regression analysis considers both of these criteria. Partial regression analysis identifies those attributes which will reduce the variance (i.e., increase the accuracy) in the estimation error, while providing evidence into the attribute relationships to price (i.e., price structure).

A selection method of searching for an appropriate functional form as indicated by the partial regression analysis was used. In the case for the cotton model presented, the fourth model evaluated revealed an appropriate functional form. This does not say that other transformations could not provide the same results, but only that this method was effective. That is, this approach can be useful for identifying *an* appropriate functional form, but it cannot claim to identify the singular most efficient functional form.

There may be other limitations to this approach. The OLS regression technique may be restricting. Experimenting with non-linear regression techniques would appear warranted, but was beyond the scope of this study. This study does illustrate the usefulness of prior conceptual knowledge, which in conjunction with empirical (partial regression and residual) analysis provides objectivity to the discovery process required to identify appropriate functional form models.

References

Belsley, D.A., E. Kuh, and R.E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* New York: John Wiley and Sons, Inc, 1980.

Bowman, K.R. "A Multi-Stage Hedonic Market Model of Cotton Characteristics With Separable Supply and Demand." Unpublished Ph.D. Dissertation, Texas Tech University, 1989.

Bowman, K.R., and D.E. Ethridge. "Characteristic Supplies and Demands in a Hedonic Framework: U.S. Market for Cotton Fiber Attributes." *American Journal of Agricultural Economics* 74(1992):991–1002.

Brown, J.E. "The Optimal Hedonic Model Structure for Daily Cotton Market Price Reporting." M.S. Thesis, Texas Tech University, 1993.

Burgess, J.F. Jr., and O.R. Harmon. "Specification Tests in Hedonic Models." *Journal of Real Estate Finance and Economics* 4(Dec. 1991):375.

Can, A. "Specification and Estimation of Hedonic Housing Price Models." *Regional Science and Urban Economics* 22(Sept. 1992):453.

Carrasco-Tauber, C., and L.J. Moffitt. "Damage Control Econometrics: Functional Specification and Pesticide Productivity." *American Journal of Agricultural Economics* 74(Feb. 1992):158–62.

Davidson, R., and J.G. MacKinnon. "Convenient Specification Tests for Logit and Probit Models." *Journal of Economics* 25(July 1984):241–62.

Ethridge, D.E. and B. Davis. "Hedonic Price Estimation for Commodities: An Application to Cotton." *Western Journal of Agricultural Economics* 7(1982):156–63.

Ethridge, D.E., C. Engels, and J.E. Brown. "An Econometric Approach for Estimating Daily Market Prices." *1992 Beltwide Cotton Conference Proceedings,* Cotton Economics and Marketing Conference, National Cotton Council: 190–4.

Ethridge, D.E. and K.H. Mathews. "Reliability of Spot Cotton Quotations for Price Discovery in the West Texas Cotton Market." Agricultural Economics Dept., Texas Tech University College of Agr. Sciences Pub No. T-1-212. Aug. 1983.

Ethridge, D.E. and J.T. Neeper. "Producer Returns From Cotton Strength and Uniformity: An Hedonic Price Approach." *Southern Journal of Agricultural Economics* 19(1987):91–7.

Ethridge, D.E., D.L. Shaw, and J.E. Ross. "An Evaluation of the Impact of Instrument Test Line Values on Cotton Marketing and Use: Progress Report." CED Working Paper, Economic Research Service, USDA, Aug. 1977.

Godfrey, L.G. *Misspecification Tests in Econometics: The Lagrange Multiplier Principle and Other Approaches.* Cambridge: Cambridge University Press, 1988.

Hembree, J.F., D.E. Ethridge, and J.T. Neeper. "Market Values of Fiber Properties in Southeastern Textile Mills." *Textile Research Journal* 56(Feb. 1986):140–41.

Hudson, D., D. Ethridge, and J. Brown. "Producer Prices in Cotton Markets: Evaluation of Reported Price Information Accuracy." *Agribusiness: An International Journal* (in press).

Kang, H.B., and A.K. Reichert. "An Evaluation of Alternative Estimation Techniques and Functional Forms in Developing Statistical Appraisal Models." *Journal of Real Estate Resources* 87(Fall 1987):1.

Lichtenberg, E. and D. Zilberman. "The Econometrics of Damage Control: Why Specification Matters." *American Journal of Agricultural Economics* 68(1986):261–73.

Mosteller, F. and J.W. Tukey. *Data Analysis and Regression* Reading, Mass: Addison-Wesley, 1977.

Neter, J.A., W. Wasserman, and M.H. Kutner. *Applied Linear Regression Models.* 2nd ed. Homewood, IL.: Richard D. Irwin, Inc., 1989.

Ozuna, T. Jr., K.Y. Jang, and J.R. Stoll. "Testing for Misspecification in the Referendum Contingent Valuation Approach." *American Journal of Agricultural Economics* 75(May 1993):332–8.

Pace, R.K., and O.W. Gilley. "Estimation Employing A Priori Information Within Mass Appraisal and Hedonic Pricing Models." *Journal of Real Estate Finance and Economics* 3(Mar. 1990):55.

U.S. Department of Agriculture, "Daily Spot Cotton Quotations." Market News Branch, Cotton Division, Agricultural Marketing Service, Memphis, TN. Daily issues.

Yatchew A. and Z. Griliches. "Specification Errors in Probit Models." *Review of Economics and Statistics* 67(Feb. 1985):134–9.