WORKING PAPERS

A POLYNOMIAL APPROXIMATION FOR SWITCHING REGRESSIONS WITH APPLICATION TO MARKET STRUCTURE-PERFORMANCE STUDIES

Donald T. Sant

WORKING PAPER NO. 9

February 1978

BUREAU OF ECONOMICS
FEDERAL TRADE COMMISSION
WASHINGTON, DC 20580

Abstract: A Polynomial Approximation for Switching Regressions with application to Market Structure-Performance Studies

This paper describes an estimation method for the switching regression model when the choice of regression regime is determined by a critical value of an exogenous variable. The advantage of the method over standard regression techniques is that standard errors can be calculated for the estimated critical value. The technique is applied to a model of price-cost margins and industry concentration ratios.

A Polynomial Approximation for Switching Regressions with application to Market Structure-Performance Studies

Donald T. Sant

1. Introduction

In many empirical studies in which a regression approach is appropriate, it is not realistic to assume that all observations are generated by the same regression equation. Economic examples include questions of "structural change" and models of disequilibrium such as Fair and Jaffe (1972). More generally, all switching regression models are of this form. The estimation difficulties for these types of models, stem from the fact that standard regression theory can only be used if prior knowledge is available which can classify observations to their respective generating equation. Without this prior information, it is necessary to make additional assumptions in order to obtain estimates with desirable properties.

Most assumptions describe some mechanism for determining which regression regime will generate an observation. Goldfeld and Quandt (1972, 1973), specify a probability model for determining the regression regime. In their formulation, nature selects the regime which generates the ith observation according to the probability p_i where p_i may be a function of some exogeneous variables. In Quandt (1958), the choice of regression regime is determined by a critical value of an exogeneous variable. If the exogeneous variable is smaller than some critical value,one regime generates the observation and if the exogenous variable is larger than this critical value a second regime generates the observation. In either

formulation additional parameters have been specified which can be used to classify observations to their respective generating equation. The substantive difference between these two approaches is the amount of stochastic variation allowed into the model. The first approach includes the second approach if we can specify and estimate a probability model which allows for the degenerate case where the probabilities are identically zero or one. But if our probability model of regression regimes does not allow for this restriction, the two approaches are not nested except possibly in the limit as some parameters converge to plus or minus infinity. Since this is the case for the logit and probit probability models, it is still necessary to consider both approaches separately. This paper will start with the nonstochastic formulation of selecting the regression regime and will describe an approach for estimating the critical value of the exogeneous variable.

The particular model considered in this paper is one in which the mean of a random variable can be represented by a step function of some exogenous variable but where the discontinuity point is unknown. The suggested approach is to approximate this model by a model in which the mean is represented by a continuous grafted polynomial. The advantage of this approach is that estimates of precision can be made about where the discontinuity occurs. Further, specification checks can be made to see if the step function is an appropriate representation of the data.

In section 2, the statistical model is presented along with some estimation theory. Section 3 presents some results on artificial data and section 4 presents an economic illustration using price-cost margins and concentration ratios.

2. The Statistical Model

The regression model of interest is

$$(1a)\quad y_t = \mu_t + \varepsilon_t \qquad x_t \leq \alpha$$

$$(1b)\quad y_t = \mu_t + \gamma + \varepsilon_t \qquad \alpha \leq x_t$$

where μ_t is an estimable function of exogenous variables, which may or may not include x_t, ε_t is a random error term with mean 0 and variance σ^2, and γ is the step increment which occurs at the discontinuity point α. The problem is to obtain estimates of all of the parameters, including α.

There is no conceptual difficulty in using least-squares or the likelihood approach (Quandt, 1958) to estimate the parameters of equation (1), however, an algorithm other than the exhaustive search procedure of Quandt (1958) is not immediately obvious. The major difficulty is that this search procedure must be expanded beyond finding the minimum error sum of squares if estimates of standard errors are desired. For example, to obtain a 95% confidence interval for α, one must find the values of α^*, (Scheffe 1959), which solve

$$(2)\quad (T-k)\frac{S_\omega - S_\Omega}{S_\Omega} = F_{.05;1,T-k}$$

where

T= number of observations
k= number of parameters
$F_{.05;1,T-k}$= upper 5% point of the F-distribution with 1 and T-k degrees of freedom

S_Ω= residual sum of squares which has been minimized with respect to all parameters (including α) and

S_ω= residual sum of squares which has been minimized with respect to all parameters except $\alpha = \alpha^*$.

Since this will become a very complex problem when more complicated confidence ellipsoids are desired, an alternative approach seems desirable. The alternative suggested here is to choose a functional form which is easier to estimate and which will approximate equation (1) reasonably well. A functional form which satisfies this criterion is a segmented polynomial, (Gallant and Fuller, 1973). In a small neighborhood of the point $x = \alpha$, we specify a polynomial in x, which will join the two segments of equation (1). To reduce estimation problems, the polynomials can be constrained to be continuous and differentiable in x. Another consideration of this approach is that it seems to be more consistent with the assumptions of economic theory where continuous functions are most often used to describe behavior. But it is still true that if equation (1) is the correct description of reality, we will be making a specification error if we estimate a different set of equations.

To describe the approximation, consider the model

(3a) $y_t = \mu_t + \varepsilon_t \qquad x_t \leq \alpha_1$

(3b) $y_t = \mu_t + g(x_t) + \varepsilon_t \qquad \alpha_1 \leq x_t \leq \alpha_2$

(3c) $y_t = \mu_t + \gamma + \varepsilon_t \qquad \alpha_2 \leq x_t.$

We have added the joining equation $g(x_t)$, and if we want the function y_t to be continuous in x_t, $g(x_t)$ must satisfy

(4) $g(\alpha_1) = 0$

(5) $g(\alpha_2) = \gamma$

and for the function y_t to be differentiable in x_t, $g(x_t)$ must satisfy

(6) $g'(\alpha_1) = 0$

(7) $g'(\alpha_2) = 0.$

If the function g is to be a polynomial the smallest degree polynomial which will satisfy these constraints is one of degree three.

(8) $g(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3.$

Looking at equation (3), one can easily see that the specification error is only made in the range $\alpha_1 \leq x_t \leq \alpha_2$. If the parameters of equation (8) can be chosen which satisfy the constraints (4)-(7) and make the difference $\alpha_2 - \alpha_1$ small, equation (3) should be a very good approximation to equation (1). Imposing the constraints on equation (8) one obtains

(9) $\beta_0 = \frac{\beta_3}{2} \alpha_1^2 (\alpha_1 - 3\alpha_2)$

(10) $\beta_1 = 3\beta_3 \alpha_1 \alpha_2$

(11) $\beta_2 = \frac{3}{2} \beta_3 (\alpha_1 + \alpha_2)$

(12) $\gamma = \frac{-\beta_3}{2} (\alpha_2 - \alpha_1)^3.$

One can now reparameterize equation (3) into a form that is convenient to use standard optimization routines to find the least squares estimates. Letting $\beta_3 = \beta$ and defining the function $I(z)$ to be

(13a) $I(z) = 0 \qquad z<0$

(13b) $I(z) = 1 \qquad z \geq 0$

equation (3) can be written

$$(14)\quad y_t = \mu_t + g(x_t) + \frac{3}{2}\beta((\alpha_2 - x_t) - (\alpha_1 - x_t))(\alpha_1 - x_t)^2 I(\alpha_1 - x_t) + \beta(\alpha_1 - x_t)^3 I(\alpha_1 - x_t) + \frac{3}{2}\beta((\alpha_1 - x_t) - (\alpha_2 - x_t))(\alpha_2 - x_t)^2 I(x_t - \alpha_2) + \beta(\alpha_2 - x_t)^3 I(x_t - \alpha_2) + \varepsilon_t$$

which is a once continuously differentiable function of β, α_1, α_2, and x_t. It is only once continuously differentiable since

(15) $z^k I(z)$

is only continuously differentiable for $k \geq 2$. The function (14) is not twice differentiable at the points $x=\alpha_1$ and $x=\alpha_2$.

It follows from Malinvaud (1970, p. 331) that the least squares estimates of equation (14) are consistent and, under suitable conditions on the parameter space, asymptotically normally distributed. To use standard theorems to find the limiting distribution, we need derivatives of the first three orders which restricts the relevant domain of the parameter space of α_1 and α_2 to any closed and bounded region which does

not include any of the values of the exogenous variables x_t. Given this parameter space, it then follows that the parameter values of α_1, α_2, β, and θ (the parameters of the function μ_t) which minimize

$$(16) \quad \sum_{t=1}^{T} (y_t - f_t)^2$$

where f_t is the right hand side (excluding ε_t) of equation (14), are consistent and asymptotically normally distributed with covariance matrix given by

$$(17) \quad \sigma^2 \sum_{t=1}^{T} \nabla f_t \, \nabla f_t'$$

where ∇f_t is the gradient of the function f_t.

3. Some Nonrandom Experiments

In order to describe the adequacy of the approximation and to obtain methods for estimation, some nonrandom experiments were conducted. Three forms of experiments were conducted. Experiment 1 could be described exactly by equation (1) with μ_t and ε_t identically equal to zero and experiment 2 set $\mu_t = 0$ but included a nonzero error term. Experiments 3 and 4 used data that could be best described by a Bernoulli random variable. Experiment 5 used data described by a linear equation without an error term.

3.1 Experiments 1 and 2.

In experiment 1a and 1b, equation (1) was a perfect representation of the data. In both cases μ_t and ε_t were identically

equal to zero. In experiment 1a, $\gamma = .3$, $\alpha = .51$, and in experiment 1b, $\gamma = -.3$, $\alpha = .51$. In both cases there were 25 observations where $X_t \leq .5$ and 25 observations where $X_t \geq .52$, and all values of x_t where spaced .02 apart. As is evident from table I, the function (14) provides a very good representation of the data. The error sum of squares is essentially zero, and to 8 significant digits, γ is estimated, using equation (12), to be .3. The parameter estimates, judged by their standard errors, are very precise and one learns everything one could learn about α from the data. There are no observations between .50 and .52 and (to 5 significant digits) since $\hat{\alpha}_1 = .50$, $\hat{\alpha}_2 = .52$, one could conclude from this evidence that $\alpha \varepsilon (.50, .52)$. Experiment 1b conveys no additional evidence about goodness of fit, except it numerically exhibits the symmetry of the function. If $\alpha_1 < \alpha_2$, positive values of β imply a negative value of γ, and negative values of β imply a positive value of γ (evident from equation (12)). However, the algorithm used for minimization (Davidon-Fletcher-Powell, see Goldfeld and Quandt, 1972) did not impose the requirement that $\alpha_1 < \alpha_2$. And depending upon the initial values, the minimizing values were sometimes such that $\hat{\alpha}_1 > \hat{\alpha}_2$, but in these cases, the sign of β also changed so as to give precisely accurate results.

In experiment 2, the values of y were either -.3 or .02 if $x_t \leq .50$ and were either .3 or -.02 if $x_t \geq .52$. The values

of y_t were all .02 or -.02 when $x_t \varepsilon (.38, .62)$ and were either .3 or -.3 when $X_t \notin (.38, .62)$. Judged by the point estimates, the procedure has determined the value of α to a reasonable tolerance, but our standard errors of the estimates are fairly large. But considering the data, this should be expected. The value of the dependent variable is overlapping, and one should not expect the classification of observations to the appropriate regression regime to be very precise. Judged by these experiments, the procedure is a good method for estimating equations of type (1).

3.2 Experiments 3 and 4.

In experiments 3 and 4, there where 50 observations, where $X_t = .02t$, and $y_t = 0$ if t is odd and $y_t = .3$ if t is even. This data was chosen as representative of the alternative hypothesis that $\gamma = 0$ and would be best represented by the equation

(18) $y_t = \mu + \varepsilon_t$

where here $\mu = .15$.

Experiment 3 assumed that $\mu_t = 0$ and tried to fit the approximate step function to this data. The representation of the observed data was as good as equation (18) since all observations are in the range where $x_t \geq \hat{\alpha}_2$. In effect it just used γ to estimate μ. The estimated $\hat{\gamma}$ is .15 which would be the estimate of μ using these observations. One indication of the inappropriateness of the specification however is that both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ were less than the actual values of x_t. Another

Table I

Parameter Estimates For The Controlled Experiments

Experiment	β	α_1	α_2	constant	Error Sum of Squares
1a	-74945.072 (4.523)	.500 ($.285 \times 10^{-6}$)	.520 ($.285 \times 10^{-6}$)		4.53×10^{-16}
1b	74938.415 (4.688)	.500 ($.295 \times 10^{-6}$)	.520 ($.295 \times 10^{-6}$)		6.11×10^{-16}
2	-22583.469 (1.014×10^{6})	.507 (.200)	.536 (.200)	-.146 (.033)	1.249
3	-9.959	-1.224	-.091		1.125
4	-10.001 (1418.007)	.414 (2.859)	.526 (2.848)	.146 (.035)	1.124
5	-.018 (.0026)	-2.567 (.226)	3.589 (.228)	-.776 (.075)	5.55×10^{-6}

indication is that equation (17) becomes a singular matrix with these estimates and this data set.

Experiment 4 uses the same data set but includes a general mean μ in the estimation formulation. This also eliminates the singular matrix of equation (17), since at the minimizing values we get some variability in the function f_t in the range of the observed X_t. Here also we would be inclined to make correct inferences. The standard errors of each of the estimates are large relative to their value, which would lead one to suspect something wrong with the approximating function (or in the specification of the model).

3.3 Experiment 5.

Experiment 5 used data generated from the linear equation

(19) $y_t = .5x_t \qquad t=1,50$

where $x_t = .02t$. Even though we were able to fit equation (19) fairly closely, there is evidence which would lead to the conclusion that our model was not specified correctly. The standard errors are not large relative to the coefficient estimates, but the difference, $\hat{\alpha}_2 - \hat{\alpha}_1$, is much larger than is consistent with the model of equation (1) and the observed data. It is the middle part of our function (3) which is being used to fit the data and as a result, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ yield no evidence (since there isn't any) about the range of α.

The interval $(\hat{\alpha}_1, \hat{\alpha}_2)$ encompasses all of the observed data, so we have learned nothing more about α from the estimation process and from the data.

The evidence from these nonrandom experiments suggests that equation (3) is a useful approach to the estimation of equation (1). The model yields precise estimates when specified correctly, and it provides signals for specification error when appropriate. The next section uses this model on data about price-cost margins and market concentration.

4. An Economic example

A much studied question is the impact of industry structure on performance. The general hypothesis is (Bain 1968) that high seller concentration within industries should be associated with substantial excesses of selling price over long-run average costs and low seller concentration should be associated with no excess at all. After a survey of the empirical work however Weiss (1974) concludes that the question still remains whether there really is a critical level of concentration which can be used to distinguish between high and low levels of seller concentration.

The hypothesis of substantial profits has been tested using specifications of the form of equation (1), but where α or the critical concentration level was given by assumption instead of estimated. Since α is not determined by

theory, assuming α is known when it should be estimated, will give an all too optimistic sense of the precision of our estimates. This section will use the technique of section II to analyze the question of a critical concentration level and provide estimates of precision for that level.

The data are those used by Kwoka (1977) and include the four-digit SIC industry aggregates taken from the 1972 Census of Manufactures and another source described later. The profit measure is the price-cost margin defined as

$$Y = \frac{\text{Value Added - Payroll}}{\text{Value of Shipments}}$$

Independent variables other than concentration are:

(1) A measure of the capital-output ratio given by

z_1 = gross book value of fixed assets divided by value of shipments.

(2) A measure to account for the local and regional nature of some markets given by

z_2 = Collins and Prestons (1968, 1969) geographical dispersion index defined as the sum of the absolute values of the differences between the percent of an industry's value added and all manufacturing value added for all four Census regions of the country.

(3) A measure of industry growth to account for short-run phenomena defined as

$$z_3 = \frac{\text{Value of shipments in 1972 - Value of shipments in 1967.}}{\text{Value of shipments in 1972}}$$

(4) A dummy variable z_4 equalling one for consumer goods industries.

(5) A constant.

The usual measure of industry concentration is the four firm concentration ratio (x_t) which is also used here in the first formulation. So the equation to be estimated is

$$(20)\quad y_t = q(x_t) + \delta_0 + \delta_1 z_{1t} + \delta_2 z_{2t} + \delta_3 z_{3t} + \delta_4 z_{4t} + \varepsilon_t$$

where $q(x_t)$ represents the segmented polynomial approximation to the step function. Estimates of equation (20) are given in table II.

Table II

Parameter Estimates Using Four Firm Concentration Ratio

$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}$	$\hat{\delta}_0$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$
.460	.470	-88106.902	.225	.081	-.035	.044	.039
(.956)	(.979)	(5.55×10^5)	(.021)	(.017)	(.021)	(.012)	(.861)

Error Sum of Squares 1.994

These results are consistent with those of previous investigators. The values of $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\beta}$ imply that for industries where the concentration ratio is above 47 percent, the price-cost margin is 4.4 percentage points higher than in those industries where the concentration ratio is below 46 percent. However, when the "break point" is also estimated, our precision of these results is not very good. In fact, we can not make any definitive statement about the critical level of concentration. The estimated variance of $\hat{\gamma}$ is .076 which permits us to be precise about the price-cost differential, but the precise point where monopoly profits become prevalent is not obvious. From the experimental results, this probably results from the "fuzzy" region as in experiment 2 in the middle range of concentration. Scanning the

error sum of squares for different values of the parameters confirmed this suspicion. Holding the difference $(\hat{\alpha}_1 - \hat{\alpha}_2)$ constant at .01 and allowing $\hat{\alpha}_2$ to vary between .5 and .68 yielded a change in the error sum of squares of only 2 percent.

A possible reason for these results is that no allowance is made for the distribution of output between the four leading firms. The use of the four firm concentration ratio constrains the impact of all four firms to be the same. This constraint precludes tests of hypotheses regarding dominant firm models and collusive dealings models. A priori reasoning would suggest that the performance of an industry with a four firm concentration ratio of .60 but where one firm had 50 percent of the market would be different from the performance of an industry where each of the four firms provided 15 percent of the market. This reasoning is not contained in the formulation using the four firm concentration measure.

The next formulation will disaggregate the measure of concentration by firm and enter the share of the market of the four largest firms individually. The individual firm share data is from Economic Information Systems, Inc. A description of this data and further discussion of all data used in this study can be found in Kwoka (1977). The equation to be estimated using the individual shares is

$$(21)\quad y_t = q_1(s_{1t}) + q_2(s_{2t}) + q_3(s_{3t}) + q_4(s_{4t}) + \delta_0 + \delta_1 z_{1t} + \delta_2 z_{2t} + \delta_3 z_{3t} + \delta_4 z_{4t} + \varepsilon_t$$

where s_{it} is the share of the output in the t^{th} industry produced by the i^{th} largest firm. These parameter estimates are given in table III.

These results also are not very precise and would not refute very many theories of oligopoly behavior. All of the qualitative findings agree with Kwoka (1977) where the least-squares search procedure of Quandt (1958) was used, but the statistical findings are much less significant when one accounts for the estimated break point. Two observations can be made to help explain this lack of significance. The residual variance is large relative to γ, making it difficult to estimate precisely (without a fairly large sample) the point where monopoly profits emerge or stop increasing. The second observation is that there may be no nice relationship between profits and concentration. Even if the results were estimated with greater precision, the results would be consistent with several theories of oligopoly development. Some industries could be behaving like a dominant firm model, other industries could be guilty of collusive agreements, but the above formulation, with one equation explaining all behavioral relations, ignores the mechanism producing concentrated industries and will not easily distinguish among different theories. Since the reason for concentration is also important, greater

Table III

Parameter Estimates Using Individual Shares

	β	δ_1	δ_2
s_1	-103.12 (407.68)	.216 (.068)	.311 (.072)
s_2	-123.57 (39.433)	.090 (.036)	.185 (.046)
s_3	5.82 (3737.866)	.147 (.083)	4.892 (3007.955)
s_4	38.976 (12294.27)	-.635 (204.088)	.034 (.051)

$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$	$\hat{\delta}_0$
.072 (.018)	-.035 (.013)	.036 (.019)	.035 (.011)	.608 (3507)

Error Sum of Squares 1.808

use of structural equation models will be most useful in explaining the correlation between profitability and concentration.

5. Conclusion

This paper has suggested a different approach to the modelling and the estimation of relationships between variables in situations where the form of the relationship is determined by a critical value of an additional exogenous variable. I have focused on the case where there is only one critical value and where only the constant term in a linear model is hypothesized to vary. However, it is obvious that this approach can be generalized to allow for different slope coefficients by interacting the constrained polynomial equation (8) with additional slope parameters. The procedure for allowing more than two regression regimes is also straight forward conceptually, although it will be more difficult to implement. The advantage of this technique over the use of linear splines (Poirier, 1975), is that it is not necessary to specify a priori the join points (the critical values) since they are estimated jointly with the coefficients. The advantage of this technique over the search procedure of Quandt (1958) is that we have a simple procedure for estimating standard errors of the estimated critical values. The stochastic structure is different from the one specified in Goldfeld and Quandt (1972, 1973), so the choice here must be determined on theoretical grounds. It seems plausible in describing some relationships to specify that the regression regime is chosen in a deterministic

manner which would mean the technique described in this paper would be appropriate. However, in cases where the choice of regression regime is determined by a stochastic process, the techniques described by Goldfeld and Quandt (1972, 1973) would be more appropriate.

Bibliography

Bain, J.S., (1968), Industrial Organization, second edition, John Wiley and Sons, New York.

Collins, N.R. and L.E. Preston, (1969), "Price-Cost Margins and Industry Structure," Review of Economics and Statistics, 51, 271-286.

Collins, N.R. and L.E. Preston, (1968), Concentration and Price-Cost Margins, University of California Press, Berkeley.

Fair, R.C. and D.M. Jaffee, (1972), "Methods of Estimation for Markets in Disequilibrium," Econometrica, 40, 497-514.

Gallant, A.R. and W.A. Fuller, (1973), "Fitting Segmented Polynomials Regression Models Whose Join Points Have to be Estimated," Journal of the American Statistical Association, 68, 144-147.

Goldfeld, S.M. and R.E. Quandt, (1972), Nonlinear Methods in Econometrics, North-Holland Publishing Co., Amsterdam.

Goldfeld, S.M. and R.E. Quandt, (1973), "A Markov Model for Switching Regressions," Journal of Econometrics, 1, 3-16.

Kwoka, J.E. (1977), "The Effect of Market Shares and Share Distribution on Industry Performance," Federal Trade Commission Working Paper No. 2, Washington, D.C. March 1977.

Malinvaud, E. (1970), Statistical Methods of Econometrics, North-Holland Publishing Co., Amsterdam.

Poirier, D.J. (1975), "On the Use of Bilinear Splines in Economics," Journal of Econometrics, 3, 23-34.

Quandt, R.E.7 (1958), "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," Journal of the American Statistical Association, 53, 873-880.

Scheffe; H., (1959), The Analysis of Variance, John Wiley, New York.

Weiss, L., (1974) "The Concentration Profits Relationship and Antitrust," in Goldschmid, H.H, et. al., Industrial Concentration: The New Learning, Little Brown, Boston.