A FURTHER APPROACH TO THE ESTIMATION OF SWITCHING REGRESSIONS

Richard E. Quandt

ECONOMETRIC RESEARCH PROGRAM
Research Memorandum No. 122

March 1971

The research described in this paper was supported by NSF Grant GS 3264.

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

ABSTRACT

In recent years much attention has been focussed on the problem of discontinuous shifts in regression regimes at unknown points in the data series. The present paper approaches this problem by assuming that Nature chooses between regimes with probabilities λ and $1-\lambda$. This allows the formulation of the appropriate likelihood function which may be maximized with respect to the parameters in the regression equations and λ. The method is tested in some sampling experiments and in a realistic economic problem and is found satisfactory.

A FURTHER APPROACH TO THE ESTIMATION OF SWITCHING REGRESSIONS*

Richard E. Quandt
Department of Economics, Princeton University

I. INTRODUCTION

The standard problem of switching regimes in regression theory consists of (1) testing the null hypotheses that no switch in regimes took place against the alternative that the observations were generated by two (or possibly more) distinct regression equations and (2) estimating the two (or more) regimes that gave rise to the data. Given n observations and k independent variables, the null hypothesis is expressed by

$$Y = X\beta + U \qquad (1\text{-}1)$$

where Y is the nx1 vector of observations on the dependent variable, X the nxk matrix of observations on the independent variables, U the nx1 vector of unobservable error terms distributed as $N(0,\sigma^2 I)$ and β the kx1 vector of coefficients to be estimated. The alternative hypothesis is expressed by asserting that there exists some permutation of the rows of Y and X so that they may be partitioned

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} , \qquad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

---

*I am indebted to Gregory C. Chow, Ray C. Fair and John Tukey for useful comments and suggestions. Support was obtained from NSF Grant GS 3264.

and that

$$Y_1 = Y_1\beta_1 + U_1 \tag{1-2}$$

and

$$Y_2 = X_2\beta_2 + U_2 \tag{1-3}$$

where U_1 and U_2 are distributed as $N(0,\sigma_1^2 I)$ and $N(0,\sigma_2^2 I)$ respectively and where $(\beta_1,\sigma_1^2) \neq (\beta_2,\sigma_2^2)$.

Various special cases of this problem have been treated in the literature. Some of the most distinctive variants are as follows.

(1) If it is assumed that, under the alternative hypothesis, the subsets of observations corresponding to (1-2) and (1-3) respectively are identified on a priori grounds, the problem of testing the null-hypothesis is solved exactly by the Chow-test [2]. The corresponding estimation problem is solved simply by estimating (1-2) and (1-3) separately by least squares.

Other approaches to the problem deal with the more difficult circumstance in which it is not known which observations were generated by Regime 1 and which by Regime 2, if indeed there were two regimes at all.

(2) The relatively simplest of these variants assume that if there are two regimes, there occurred only a single switch between the t^{th} and $(t+1)^{th}$ observations (t being unknown), with the first t observations having been generated by Regime 1 and the last $n-t$ by Regime 2. At least three approaches have been suggested to cope with this variant: (a) Quandt ([9] and [10]) expresses the likelihood of the sample as a function of t and maximizes it with respect to the (continuously varying) variables β_1, σ_1^2, β_2, σ_2^2

and the (discrete) variable t. (b) Farley and Hinich [4] assume that all possible switching points are equally likely and derive a likelihood ratio test for the null hypothesis with power characteristics that appeared good in Monte Carlo experiments. (c) McGee and Carleton [8] examine the same problem from the point of view of the hierarchical clustering of adjacent observations successively included in successively recomputed regressions. At each stage left or right adjacent points continue to be included in existing clusters unless the Chow-test rules this out on the grounds that the new point could not have been generated by the same regime. The process continues until a set of optimal clusters is determined.

(3) The most complex variants hypothesize that the system may switch numerous times, either to successively newer regimes or back and forth between two particular regimes. In these cases the likelihood approaches corresponding to the previous set of variants tend to become impractical. For example, the condensed likelihood function, as a function of t only, which was suggested in [8], must be evaluated in the second set of variants (due to the discreteness in the partitioning of Y and X) a number of times that is of the order of n. In the present variant this likelihood function would have to be evaluated a number of times that is of the order of 2^n. Some of the approaches that have been suggested for this variant are as follows. (a) Brown and Durbin [1] define recursive residuals u_r $(r = k+1,\ldots,n)$ where u_r is obtained from employing the r^{th} row of Y and X and an estimate of β based on the first $r-1$ rows of Y and X. A generalization of the Helmert transformation is used by

them to transform these u_r into a set of variables w_r, distributed as $N(0,\sigma^2 I)$. They then propose testing the cumulative sums of the w_r for significant departures from zero. (b) Fair and Jaffee [3] assume, as may be frequently justified in an economic context, that there exists on a priori grounds an extraneous variable z on which observations $z_i (i=1,\ldots,n)$ are available and which may be used to classify the x and y-observations between the two regimes; accordingly, if $z_i \leqq z_o$ (z_o assumed to be a known number) the i^{th} value of y is assumed to have been generated by Regime 1 and conversely for $z_i > z_o$. (c) Goldfeld and Quandt [7] generalize the approach of Fair and Jaffee by not requiring knowledge of the cut-off value z_o. Maximum likelihood estimates for the regression parameters as well as for z_o can then be operationally obtained as follows. Let D be a diagonal matrix of order n with diagonal elements $d(z_i)$ where $d(z_i) = 0$ if $z_i \leqq z_o$ and $d(z_i) = 1$ otherwise. The observations can then be thought to have been generated by

$$Y = (I-D)X\beta_1 + DX\beta_2 + W \qquad (1\text{-}4)$$

where W is the vector of unobservable (and heteroscedastic) error terms $(I-D)U_1 + DU_2$, and where β_1,β_2 as well as the elements of D must be estimated. This basically intractable combinatorial problem is rendered computationally feasible by replacing the unit step functions $d(z_i)$ in D by the continuous approximation

$$d(z_i) = \int_{-\infty}^{z_i} \frac{1}{\sqrt{2\pi}\sigma} \exp \{-\frac{1}{2}(\frac{\xi - z_o}{\sigma})^2\} d\xi \qquad (1\text{-}5)$$

This introduces two new (unknown) parameters z_o and σ but makes it now possible to write down the likelihood function for (1-4) without the disturbing combinatorial aspects that are otherwise inherent in this formulation.[1] The logarithmic likelihood function clearly is

$$L = \text{constant} - \frac{1}{2}\log|\Omega| - \frac{1}{2}\{[Y-(I-D)X\beta_1-DX\beta_2]'\Omega^{-1}[Y-(I-D)X\beta_1-DX\beta_2]\}$$

where

$$\Omega = (I-D)^2\sigma_1^2 + D^2\sigma_2^2 .$$

The maximum likelihood estimate for z_o gives the estimate for the desired cut-off value and σ has been interpreted as measuring the "mushiness" of discrimination between the two regimes.[2]

2. AN ALTERNATIVE MODEL

In the model to be analyzed in the remainder of this paper it is assumed that there is an unknown probability λ that Nature will choose Regime 1 for generating observations and a probability $1-\lambda$ that

---

[1]A simpler approach, not unlike that in [9] will also work if there is only a single extraneous variable z. This more complicated approach seems unavoidable if there are several extraneous variables $z_1,\ldots,z_p$ and if Nature is thought to make the choice between regimes on the basis of some function of the z's, say $\Sigma\gamma_i z_i$, where the γ_i are also unknown.

[2]An alternative interpretation, suggested by John Tukey, is that the extraneous variable values may decide Nature's choice of regimes only with some uncertainty. The assumption of the cumulative normal integral as the suitable approximation in (1-5) then represents the assumption that Nature's choice of regimes as a function of z is normally distributed. If one assumed, for example, that Nature's choice of regimes has, say, the Cauchy distribution, the appropriate approximation would be $d(z_i) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(z_i-z_o)$, the cumulative Cauchy integral.

it will choose Regime 2. Accordingly it addresses itself to the same type of problem as the third set of variants described in Sec. 1. Assuming that the error terms in the two regimes are normally and independently distributed, the conditional density of the i^{th} value of the dependent variable, y_i, conditional on the values of the k independent variables $x_{1i}, x_{2i}, \ldots, x_{ki}$ is

$$h(y_i | x_{1i}, \ldots, x_{2i}) = \frac{\lambda}{\sqrt{2\pi}\sigma_1} \exp \{- \frac{1}{2\sigma_1^2} (y_i - \sum_{j=1}^{k} \beta_{1j} x_{ji})^2\} + \frac{1-\lambda}{\sqrt{2\pi}\sigma_2} \exp \{- \frac{1}{2\sigma_2^2} (y_i - \sum_{j=1}^{k} \beta_{2j} x_{ji})^2\} \qquad (2\text{-}1)$$

where β_{1j} and β_{2j} denote the j^{th} regression coefficient in the two regimes and x_{ji} is the i^{th} observation on the j^{th} independent variable. The logarithmic likelihood function is obtained by summing the logarithms of (2-1) over i, yielding

$$L = \sum_{i=1}^{n} \log [\frac{\lambda}{\sqrt{2\pi}\sigma_1} \exp \{- \frac{1}{2\sigma_1^2}(y_i - \sum_{j=1}^{k} \beta_{1j} x_{ji})^2\} + \frac{1-\lambda}{\sqrt{2\pi}\sigma_2} \exp \{- \frac{1}{2\sigma_2^2}(y_i - \sum_{j=1}^{k} \beta_{2j} x_{ji})^2\}] \qquad (2\text{-}2)$$

This function is to be maximized with respect to β_{1j}, β_{2j} $(j=1,\ldots,k)$, $\sigma_1, \sigma_2 \geqq 0$ and $0 \leqq \lambda \leqq 1$.

This involves a nonlinear maximization problem and in order to explore the feasibility of obtaining estimates by maximizing (2-2), the sampling experiments performed for the purpose of examining the properties of the method using the function $d(z_i)$ and reported in [7] were repeated using the present method. These experiments employed the following equations for Regimes 1 and 2 for generating observations:

$$y_i = 1.0 + 1.0x_i + u_{1i} \tag{2-3}$$

and

$$y_i = 0.5 + 1.5x_i + u_{2i} \tag{2-4}$$

The u_{1i} and u_{2i} were normally and independently distributed with variances σ_1^2 and σ_2^2 and the x_i were uniformly distributed and identical in repeated samples. The remaining characteristics of the experiments are summarized in Table 1.

Table 1

Characteristics of Sampling Experiments

Case	Number of Observations	Range of x-values	σ_1^2	σ_2^2	λ
1	60	10 to 20	2.0	2.5	.5
2	120	10 to 20	2.0	2.5	.5
3	60	10 to 20	2.0	2.5	.75
4	60	10 to 20	2.0	25.0	.5
5	60	0 to 40	2.0	2.5	.5

The experiments were not intended to be exhaustive in any sense but were designed merely to test the computability of estimates under slightly varying circumstances. Each experiment was replicated 30 times. The likelihood function (2-2) was maximized by employing the modified quadratic hill-climbing algorithm ([5], [6]). The initial approximation to the location of the maximum used to initiate the iterative computation were the same in each replication as that employed by Goldfeld and Quandt [7], with the initial value of λ set at .5, in order to facilitate comparison of the two sets of results. Satisfactory convergence of the maximization algorithm failed to take place in 22 per cent of all replications.

The basic results are displayed in Tables 2, 3 and 4. Table 2 contains the mean biases, Table 3 the mean square errors, and Table 4 the ratio of the mean values of the asymptotic variance estimates for the successful replications in each experiment to the mean square error. Thus, denoting by $\hat{\theta}_i$ the estimate for some coefficient in the i^{th} replication of an experiment, by θ the true value and by n the number of successful replications, Table 2 contains values of

$(1/n) \sum_{i=1}^{n} (\hat{\theta}_i - \theta)$, Table 3 contains values of $(1/n) \sum_{i=1}^{n} (\hat{\theta}_i - \theta)^2$,

and Table 4 contains values of $\sum_{i=1}^{n} \hat{\sigma}_{ii} / \sum_{i=1}^{n} (\hat{\theta}_i - \theta_i)^2$, where

$\hat{\sigma}_{ii}$ is the appropriate diagonal element of the negative inverse of the matrix of second partial derivatives of (2-2).[3] For the sake of comparability, Tables 1 and 2 report the earlier results of Goldfeld and Quandt in columns A and the results of the present experiments under column B.

[3]As is well-known, the asymptotic covariance matrix of the maximum likelihood estimators is consistently estimated by the negative inverse of the matrix of second partial derivatives if the maximum likelihood estimates are jointly sufficient.

Inspection of Tables 2 and 3 reveals that the two methods are of comparable quality in the experiments performed. Out of the 20 possible coefficient-case pairings in which comparisons are possible the present method exhibits smaller mean biases in 10 and smaller mean square errors in 7 instances. The procedure proposed in this paper has a distinct advantage over the procedure employed in [7] in terms of the ratios of the mean asymptotic variance to the mean square error. This ratio converges to unity as the sample size is increased indefinitely. For the procedure reported in [7], convergence seemed to occur for samples of size 120 but not for samples of size 60; in the present method nearly all the ratios are close to 1 (but note the value for λ in Case 5).[4] The relatively good performance of this method is all the more impressive in that, unlike the methods of Fair and Jaffee [3] or Goldfeld and Quandt [7], it does not require prior knowledge of an extraneous variable z. It is by the same token inferior to those methods in that it does not allow individual observations to be identified with particular regimes but computes only the probability that one or the other regime was operative during the sample period.

---

[4]These ratios are easily affected by outliers in the estimates for the asymptotic variances. Outliers may occur relatively more often than would be expected since second derivatives were computed by numerical differencing.

Table 2

Mean Biases

Coefficient	Case 1 A	1 B	2 A	2 B	3 A	3 B	4 A	4 B	5 A	5 B
a_1	-.4972	-.3216	.0070	.0098	-.1304	-.1411	-.5616	.0139	-.1464	-.1837
b_1	.0287	.0194	.0016	-.0049	.0078	.0011	.0325	.0395	.0003	.0061
a_2	.6313	.4622	.2674	.2525	.2624	-1.1361	.9600	3.1806	.2381	.4340
b_2	-.0457	-.0259	-.0245	-.0122	-.1028	.0642	-.0503	-.1416	-.0128	-.0014
λ	-	.0386	-	-.0046	-	.0514	-	.0411	-	.0092

Table 3

Mean Square Errors

Coefficient	Case 1 A	1 B	2 A	2 B	3 A	3 B	4 A	4 B	5 A	5 B
a_1	2.0310	2.9119	.7009	1.0121	1.3154	1.0745	1.7264	3.9031	.2798	.4185
b_1	.0074	.0108	.0035	.0040	.0050	.0040	.0063	.0136	.0004	.0005
a_2	1.7747	2.1271	1.7629	1.6834	6.7845	14.4618	14.1377	32.4337	.4257	.4080
b_2	.0076	.0073	.0066	.0057	.0288	.0566	.0496	.1674	.0008	.0004
λ	-	.0021	-	.0003	-	.0042	-	.0117	-	.0004

Table 4

Mean Asymptotic Variance/Mean Square Error

Coefficient	Case 1	2	3	4	5
a_1	1.0221	1.4141	1.3675	1.3674	1.0508
b_1	1.0348	1.4212	1.4781	1.6648	1.1801
a_2	1.5468	.9211	1.1646	.9304	1.1634
b_2	1.8365	1.1368	1.1136	1.1626	2.4603
λ	2.1354	.7694	.9238	.7492	13.2621

3. A CONCRETE EXAMPLE

In their analysis of markets in disequilibrium Fair and Jaffee examine the market for housing starts with the following demand and supply functions:

$$y_t = \alpha_o + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{3t} + u_{1t} \tag{3-1}$$

and

$$y_t = \beta_o + \beta_1 x_{1t} + \beta_4 x_{4t} + \beta_5 x_{5t} + \beta_6 x_{6t} + u_{2t} \tag{3-2}$$

where y_t is the observed number of housing starts in month t, x_{1t} is a time trend, x_{2t} a measure of the stock of houses, x_{3t} the mortgage rate lagged two periods, x_{4t} a moving average of private deposit flows in savings and loan associations and mutual savings banks lagged one period, x_{5t} a moving average of borrowings by savings and loan associations from the Federal Home Loan Bank lagged two periods and $x_{6t} = x_{3t+1}$. They estimated their model in several ways, two of the principal ones being based on the outside variable $z_t = x_{6t+1} - x_{6t}$, i.e., the change in the mortgage rate. They classified the observations into two groups depending on the sign of z_t: according to the elementary economic theory of markets in disequilibrium $z_t > 0$ implies that there is an excess demand and hence the observed value of y_t lies on the supply function and conversely for $z_t < 0$. Their two methods differ with respect to the treatment of observations corresponding to $z_t = 0$.

Their model was reestimated employing the λ-method developed in this paper. A slight additional complication is due to the fact that the error terms u_{1t} and u_{2t} must be assumed to be serially correlated. Assuming first order Markov processes $u_{1t} = \rho_1 u_{1t-1} + \varepsilon_{1t}$ and $u_{2t} = \rho_2 u_{2t-1} + \varepsilon_{2t}$ for the error generation, equations (3-1) and (3-2) can be transformed to contain only the independent errors ε_{1t} and ε_{2t} and the likelihood function corresponding to (2-2) can then be written down. Maximizing it yields results that are displayed in Table 5 together with those of Fair and Jaffee. The parenthesized figures are the absolute values of the estimates divided by the square root of the estimates of the asymptotic variances. It may be noted that the results are broadly comparable to those of Fair and Jaffee with the notable difference that the responsiveness of demand is less for each independent variable than had been found by Fair and Jaffee. In a rough way the λ-method can also be said to be in between the two methods of Fair and Jaffee; for example the estimates of α_3, β_1, β_5 are close to their Model I; the estimates of ρ_1 and ρ_2 are close to their Method II. It is finally interesting to observe that the estimate of λ is .181. The probability λ was associated in the formulation with the demand regime. Of the 127 observations used by Fair and Jaffee 41 corresponded to price increases and thus represent points on the supply function, 19 corresponded to price declines and thus represent points on the demand function, and the remaining 67 corresponded to a zero price change and can be thought to lie on both functions. The number of demand points as a fraction of all points is 19/127 = .149; the number of demand points as a fraction of all points not lying on both functions is 19/60 = .316. Our estimate of λ is between these two limits and seems accordingly quite reasonable.

4. CONCLUDING REMARKS

The main results of this paper exhibit (a) that the λ-method proposed for estimating different regimes when the switching point is unknown is a computationally feasible procedure and (b) that it yields results that appeared sensible both in a limited set of sampling experiments and in a realistic economic problem. The one notable disadvantage of the method is that it does not allow individual observations to be identified with particular regimes.

There are at least three possible extensions of the procedure, of which the first two are thoroughly straightforward.

(1) The number of regimes does not have to be limited to two. If it is assumed that the number of regimes is h, with probabilities of being selected by Nature $\lambda_1, \lambda_2, \ldots \lambda_h$, $\sum_{r=1}^{h} \lambda_r = 1$, and that the conditional density of y given the x's for the r^{th} regime is $f_r(y|x_1,\ldots,x_k)$, then the conditional density corresponding to (2-1) is

$$h(y_i|x_{1i},\ldots,x_{ki}) = \sum_{r=1}^{h} \lambda_r f_r(y_i|x_{1i},\ldots,x_{ki})$$

Table 5

Results for the Fair and Jaffee Model

	Fair and Jaffee Method I	Fair and Jaffee Method II	λ-Method
α_0	193.16 (3.10)	328.43 (6.06)	190.09 (4.23)
α_1	6.78 (2.01)	3.94 (1.69)	2.25 (3.26)
α_2	- .055 (1.93)	- .032 (1.63)	- .016 (2.73)
α_3	- .241 (2.27)	- .471 (5.73)	- .195 (2.50)
β_0	-40.84 (1.29)	-75.87 (1.74)	-4.58 (.14)
β_1	- .236 (3.12)	- .332 (2.71)	- .247 (2.66)
β_4	.048 (6.20)	.047 (4.32)	.057 (6.39)
β_5	.033 (2.76)	.012 (.62)	.033 (3.79)
β_6	.116 (2.69)	.190 (2.74)	.143 (2.72)
σ_1^2	76.38 -	65.45 -	27.53 (2.04)
σ_2^2	57.61 -	47.06 -	34.45 (4.85)
ρ_1	.731 -	.499 -	.538 (5.76)
ρ_2	.574 -	.697 -	.698 (9.71)
λ	- -	- -	.181 (2.73)

from which a likelihood function corresponding to (2-2) can be immediately derived.

(2) The method of Goldfeld and Quandt [7] may be combined with the present method if one assumes that the probability with which Nature selects regimes depends on some extraneous variable z. Thus, defining $d(z_i)$ as in (1-5), the conditional density (2-1) becomes

$$h(y_i|x_{1i},\dots,x_{ki}) = \frac{d(z_i)}{\sqrt{2\pi}\sigma_1}\exp\{-\frac{1}{2\sigma_1^2}(y_i - \sum_{j=1}^{k}\beta_{1j}x_{ji})^2\} +$$

$$\frac{1-d(z_i)}{\sqrt{2\pi}\sigma_2}\exp\{-\frac{1}{2\sigma_2^2}(y_i - \sum_{j=1}^{k}\beta_{2j}x_{ji})^2\} .$$

The corresponding likelihood function is then maximized with respect to the β's, σ_1, σ_2 as well as the two additional parameters appearing in $d(z_i)$.

(3) It may be plausible to argue that if the estimated value of λ is very close to either 0 or 1, there is in fact only one regression regime in operation.[5] This suggests that a likelihood ratio test is feasible for testing the null hypothesis that only one regime is in operation (λ=0 or 1) against the alternative that there are two regimes (λ unrestricted between 0 and 1).

---

[5]This may, of course, only indicate that the numbers of observations coming from the two regimes is very unbalanced.

REFERENCES

[1] Brown, R. L. and J. Durbin, "Methods of Investigating Whether a Regression Relationship is Constant Over Time," Paper presented at the Rueopean Statistical Meeting, Amsterdam, 1968.

[2] Chow, G., "Tests of the Equality Between Two Sets of Coefficients in Two Linear Regressions," Econometrica, 28 (1960), 561-605.

[3] Fair, R. C., and D. M. Jaffee, "Methods of Estimation for Markets in Disequilibrium," Econometrica, forthcoming.

[4] Farley, J. U. and M. J. Hinich, "A Test for A Shifting Slope Coefficient in a Linear Model," Journal of the American Statistical Association, 65 (1970), 1320-1329.

[5] Goldfeld, S. M., R. E. Quandt and H. F. Trotter, "Maximization by Improved Quadratic Hill-Climbing and Other Methods," Econometric Research Program, Research Memorandum No. 95, Princeton University, April 1968, pp. 1-20.

[6] Goldfeld, S. M., R. E. Quandt and H. F. Trotter, "Maximization by Quadratic Hill-Climbing," Econometrica, 34 (1966), 541-555.

[7] Goldfeld, S. M. and R. E. Quandt, Nonlinear Econometric Methods, North-Holland Press, forthcoming.

[8] McGee, V. E. and W. T. Carleton, "Piecewise Regression," Journal of the American Statistical Association, 65 (1970), 1109-1124.

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

[10] Quandt, R. E., "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes," Journal of the American Statistical Association, 55 (1960), 324-330.