Tetrahedron Vol. 37, pp. 509 to 521
Pergamon Press Ltd., 1981. Printed in Great Britain

0040-4020/81/0201-0509 $02.00/0

CAUTIONS REGARDING THE PHYSICAL INTERPRETATION OF STATISTICALLY BASED SEPARATION OF THE ORTHO SUBSTITUENT EFFECT INTO INDUCTIVE, MESOMERIC, AND STERIC COMPONENTS—II

ACID CATALYSED ESTERIFICATION OF BENZOIC ACIDS BY METHANOL AND CYCLOHEXANOL

H. MAGER
Academy of Sciences of the GDR, Institute of Plant Biochemistry, DDR-401 Halle/Weinberg, East Germany

(Received in Germany 27 September 1979)

Abstract—In order to attain a better insight into the composition of Taft E_s^o constants, the rate data of hydrion catalysed esterification of both *m,p*-substituted and *o*-substituted benzoic acids by cyclohexanol and methanol were submitted to a statistical analysis using inductive, mesomeric, and steric substituent constants and various dummy variables differently structured. Furthermore, a principal component analysis with subsequent identification of the first principal component *via* multiple regression analysis was applied. It has been demonstrated that in the reactions of *m,p*-substituted compounds some substituents capable of exerting strong mesomeric effects show peculiar characteristics deviating from the general trend. Since the same result was obtained in the correlations of ortho substituted benzoic acids this effect was taken into account using an appropriate dummy variable, which in all cases improved the multiple coefficient of determination. It is concluded that the esterification rates of the ortho substituted compounds depend essentially upon inductive and steric effects (taken away OMe, OEt and NO_2) as proposed by Taft. While generally the E_s^o values may be regarded as some measure of a steric effect, this is not true for the methoxy and ethoxy groups.

INTRODUCTION

In the foregoing paper[1] it was demonstrated that the interpretation of Taft E_s^o constants[2] in terms of predominating inductive, mesomeric, and steric contributions depends at least partly upon the sample size, although in nearly all cases about 80% of the variability of E_s^o could be accounted for by the van der Waals radius r_v. After consideration of the corresponding $\hat{t}$ values for testing the significance of the regression coefficients of the equation based on the largest sample size [eqn (7) in part I], one might conclude that E_s^o are composite constants which are 50% mesomeric in nature. This would be in accordance with some recent results of Charton.[3] It has been shown, however, that the regressors themselves are multicollinearly related, a fact from which arises some problems if an interpretation is intended.

The classical quantitative evaluation of steric effects is based on Taft's treatment[2] of acid catalysed hydrolysis and esterification. Since for the acidic hydrolysis or the related esterification reactions of m, p-substituted benzoic acids ρ values are relatively close to zero, in the range of -0.4 to $+0.5$, it was supposed[2] that the corresponding reactions of *o*-substituted benzoic acids/benzoates are mainly subject to steric effects whereas the basic hydrolysis should be governed by both the steric and the electronic properties of the substituents. In other words the fact that the rate data of *o*-substituted compounds show a greater variability than those of the *m,p*-substituted compounds was attributed to steric effects. The occurrence of small ρ values may have at least two reasons:

(i) the reaction under investigation does not obey the usual Hammett equation (1) in form of

$$\log k = \log k_0 + \rho\sigma \qquad (1)$$

or more generally written

$$y = b_0 + b_1\sigma$$

(ii) the reaction obeys the Hammett equation but the reactivity constants exhibit no great variability (the rate differences are small).

In the first case the question arises which factors other than polar determine the reactivity, while in the second case evidence should be provided that at least in a rough approximation an extension of this result to the reactions of corresponding *o*-substituted compounds is reasonable.

The E_s^o constants evaluated by Taft[2,4] are essentially based on the hydrion catalysed esterification of *o*-substituted benzoic acids with cyclohexanol and methanol, respectively. Therefore, the first point of interest is a more detailed investigation of the factors most probably affecting the reactivity differences in similar reactions of *m,p*-substituted benzoic acids. To this end, not only the Hammett equation (1) was used but multiple regressions were applied, too, e.g. the equation of Swain and Lupton[5]

$$y = b_0 + b_1\mathscr{F} + b_2\mathscr{R} \qquad (2)$$

where $\mathscr{F}$ is an inductive substituent constant and $\mathscr{R}$ a mesomeric substituent constant, etc. The σ constants required were taken from the tabulation of McDaniel

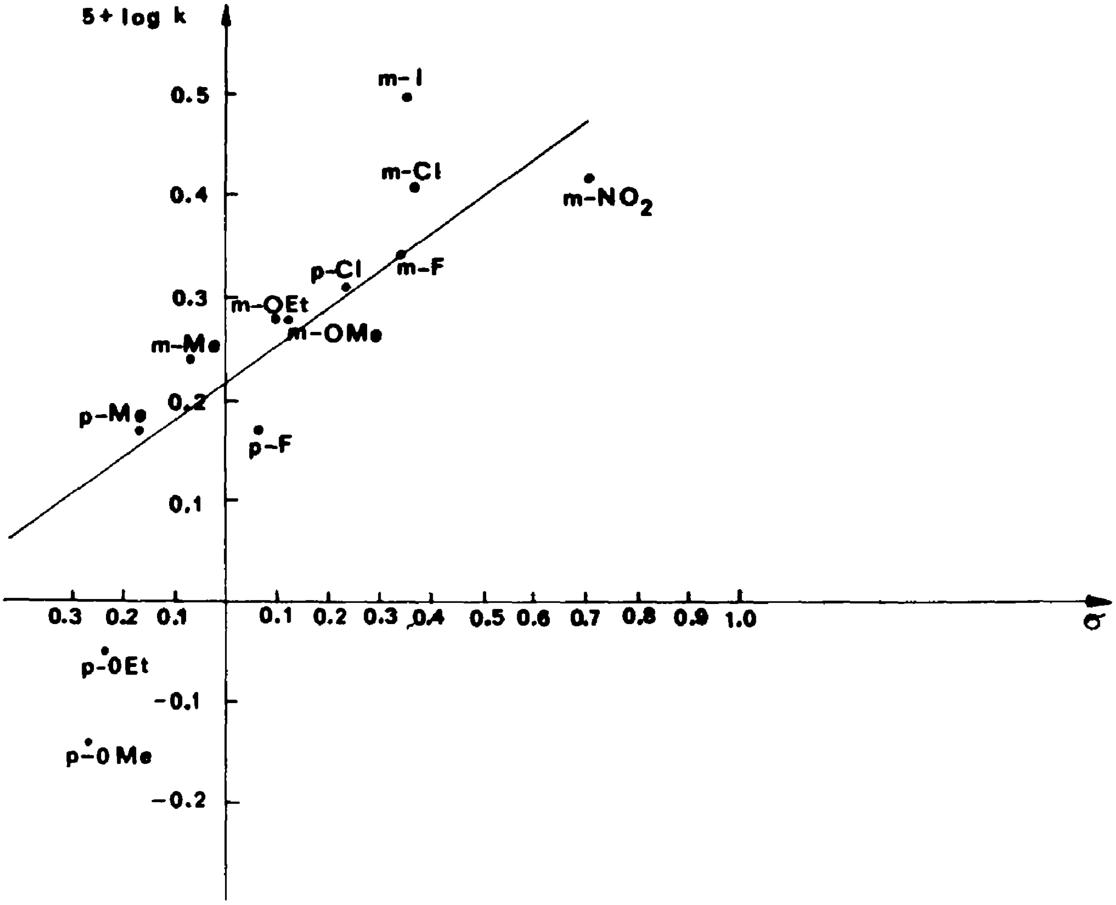

Fig. 1. Rate constants (5 + log k) for the esterification of *m*,*p*-substituted benzoic acids with cyclohexanol/HCl at 55°C vs σ. Straight line drawn in order to represent the assumed general trend, no least squares fit.

and Brown,[6] $\mathscr{F}$ and $\mathscr{R}$ values are those of Swain and Lupton as given in their original paper,[5] and the van der Waals radii are the same as used in part I.[1]

Hydrion catalysed esterification of m, p-*substituted benzoic acids with cyclohexanol at* 55°C

Figure 1 represents 14 observations relating the rate constants (given in y = 5 + log k) to σ. These results show that only a moderate fit should be expected which is confirmed by means of regression analysis (see Table 1, set 1). Although set 1 is significant at the 5% level the quality of fit is outside the range usually accepted. Exclusion of substituents having the largest residuals leads to an improvement over set 1. However, since the residuals are not large enough to be considered as outliers on a statistical basis (cp Mager[7] for the problem of testing outliers in regressions), such a procedure remains unsatisfactorily because the arbitrary actions introduced in this way may disturb the symmetry of the system and give rise to a loss of information. Setting aside all *m*-substituents gives no significant and in both cases the weighting factors (estimated using the $\hat{t}$ values) differ from those obtained from the dissociation of benzoic acids. The contribution of the mesomeric effect is in the latter reaction more important regarding the *m*-position and is of lesser importance with respect to the *p*-position. From Fig. 1 it may be seen that essentially those substituents display serious deviations which have strong donor capacities (large $|\mathscr{R}|$ values: *p*-F, *p*-OMe, *p*-OEt). The addition of a so-called dummy variable† (x_4) to which for *p*-F, *p*-OEt, *p*-OMe the value + 1 was assigned and O otherwise, should facilitate a description of the whole data material using only one equation. Set 6 confirms this assumption and may be regarded as an indication that the significance of the resonance term in set 3 could be mainly due to the special characteristics of the three substituents mentioned above, decreasing the rate of esterification more than it could have been expected from the general trend. Such effects are to be taken into consideration in all cases where reactions are significant correlation and it was necessary, therefore, to fall back upon eqn (2). The sets 4, 5 and 3, respectively, indicate the cause responsible for the dispersion of observations about the straight line in set 1. Whereas from the position meta to the reaction center essentially the inductive effect predominates, in the *p*-position the resonance effect also proved

---

†Whereas the statistical significance of a dummy variable is to be checked using the standard t test, the physical meaning may be only judged comparing the different sets obtained and, if possible, using prior information concerning the substituents for which a value of 0 was assigned.

Table 1. Esterification of *m,p*-substituted benzoic acids with cyclohexanol/HCl at 55°C[a]

Set	Regressand[b, c]	Regression equation and parameters[d, e]
1	all substituents	$y = 0.175 + 0.554\,x_1$ $R^2 = 0.7587$ $\hat{F} = 37.72$ $F_{1;12;0.95} = 4.75$
2	p-OMe, p-OEt, m-I omitted	$y = 0.227 + 0.338\,x_1$ $R^2 = 0.8290$ $\hat{F} = 43.63$ $F_{1;9;0.95} = 5.12$
3	p-substituents	$y = 0.246 + 0.287\,x_2 + 0.920\,x_3$ $R^2 = 0.9271$ $\hat{F} = 19.08$ $F_{2;3;0.95} = 9.55$ $\hat{t}$: 3.19; 6.16 $t_{3;0.95} = 3.18$
4	m-substituents	$y = 0.227 + 0.221\,x_2 + 0.007\,x_3$ $R^2 = 0.6930$ $\hat{F} = 6.77$ $F_{2;6;0.95} = 5.14$ $\hat{t}$: 3.58; 0.06 $t_{6;0.95} = 2.45$
5	m-substituents, m-I omitted	$y = 0.224 + 0.200\,x_2 + 0.016\,x_3$ $R^2 = 0.8819$ $\hat{F} = 18.68$ $F_{2;5;0.95} = 5.79$ $\hat{t}$: 5.90; 0.27 $t_{5;0.95} = 2.57$
5a	"	$y = 0.219 + 0.202\,x_2$ $R^2 = 0.8802$ $\hat{F} = 44.08$ $F_{1;6;0.95} = 5.99$
6	all substituents	$y = 0.235 + 0.400\,x_1 - 0.181\,x_4$ $R^2 = 0.8886$ $\hat{F} = 43.88$ $F_{2;11;0.95} = 3.98$ $\hat{t}$: 5.19; 3.58 $t_{11;0.95} = 2.20$

[a] taken from R. J. Hartman, H. M. Hoogsteen and J. A Moede, *J. Amer. Chem. Soc.* **66**, 1714 (1944); [b] basis set: H, *m*-F, *p*-F, *m*-I, *m*-OMe, *p*-OMe, *m*-OEt, *p*-OEt, *m*-Me, *p*-Me, *m*-NO_2, *m*-Br, *m*-Cl, *p*-Cl; [c] $y = 5 + \log k$; [d] for the statistical notations cp part I[1]; [e] $x_1 = \sigma$, $x_2 = \mathscr{F}$, $x_3 = \mathscr{R}$, x_4 = dummy variable to which for *p*-F, *p*-OMe, *p*-OEt the value +1 and for *p*-NO_2 the value −1 was assigned.

investigated of which the measured overall rate constant is a composite quantity depending itself on the rate and equilibrium constants of several reaction steps on which the substituents exert different effects yielding small ρ values for the reaction as a whole ($\rho_1 < 0$, $\rho_2 > 0$, $\rho_1 + \rho_2 \approx 0$). If in such a case substituents having some "extremal" properties (strong (+)-M or (−)-M effect, etc.) are present then under certain circumstances a "tip over" should easily take place.

Hydrion catalysed esterification of m,p-*substituted benzoic acids with methanol at* 25°C

While in the foregoing case a poor but significant correlation with σ has been obtained, the rate constants for the esterification with methanol show such a disturbing scattering that the correlation proves non-significant (cp Fig. 2 and Table 2, set 7). The peculiar behaviour of strong electron donors is much enhanced and it seems that they act against the general trend. Exclusion of most of the *p*-substituents leads to a considerable improvement of the fit (set 8). Contrary to the esterification with cyclohexanol where electron acceptors are rate-accelerating in the present case *m*- and *p*-substituents seem to produce different mode of actions. In the position meta to the reaction center, substituents exerting small (−)-I and strong (+)-M effects accelerate the rate whereas in the *p*-position the inductive effect actually acts in the same direction, the mesomeric effect, however, has changed his sign (sets 9 and 10). Since the latter observation might be due to the peculiar behaviour of some substituents, especially of strong donors (cp Fig. 2) lowering the rate contrary to the general trend (they should exert an accelerating action), a dummy variable x_5 was introduced by virtue of the same arguments that were applied in the previous section (x_5: *p*-F, *p*-OMe, *p*-OEt, *p*-Br = +1; *p*-NO_2 = −1).

It can be seen that x_5 is capable to replace $\mathscr{R}$ and simultaneously the resulting fit is much improved when compared to set 10 (Table 2, set 11a). However, the variable x_5 will not reflect the extraordinary† resonance effects in a correct manner, since there is no significant correlation between all rate data of the total sample size and the linear combination of σ and x_5 (set 12). After setting aside all *p*-substituents the *m*-substituents obey the Hammett equation in a reasonable manner (set 13) yielding as expected a negative value for ρ, whereas the *p*-substituted compounds alone show no significant correlation with

---

†The terms "extraordinary" or "special" resonance are used throughout this paper with respect to σ.

Table 2. Esterification of *m,p*-substituted benzoic acids with methanol/HCl at 25°C[a]

Set	Regressand[b, c]	Regression equation and parameters[d]
7	all substituents	$y = 1.159 - 0.233\ x_1$ $R^2 = 0.2150 \quad \hat{F} = 3,83 \quad F_{1;14;0.95} = 4.60$
8	m-substituents and p-Me, p-Cl	$y = 1.287 - 0.536\ x_1$ $R^2 = 0.8420 \quad \hat{F} = 47.95 \quad F_{1;9;0.95} = 5.12$
9	m-substituents	$y = 1.288 - 0.341\ x_2 - 0.293\ x_3$ $R^2 = 0.9874 \quad \hat{F} = 235.94 \quad F_{2;6;0.95} = 5.14$ $\hat{t}$: 17.71; 8.47
10	p-substituents	$y = 1.311 - 0.354\ x_2 + 0.299\ x_3$ $R^2 = 0.8903 \quad \hat{F} = 20.29 \quad F_{2;5;0.95} = 5.79$ $\hat{t}$: 6.16; 2.94 $\quad t_{5;0.95} = 2.57$
11	p-substituents	$y = 1.271 - 0.390\ x_2 - 0.216\ x_3 - 0.175\ x_5{}^d$ $R^2 = 0.9462 \quad \hat{F} = 23.44 \quad F_{3;4;0.95} = 6.59$ $\hat{t}$: 8.06; 0.82; 2.04 $\quad t_{4;0.95} = 2.78$
11a	p-substituents	$y = 1.291 - 0.378\ x_2 - 0.108\ x_5$ $R^2 = 0.9372 \quad \hat{F} = 37.33 \quad F_{2;5;0.95} = 5.79$ $\hat{t}$: 8.46; 4.34 $\quad t_{5;0.95} = 2.57$
12	all substituents	$y = 1.206 - 0.384\ x_1 - 0.149\ x_5$ $R^2 = 0.3600 \quad \hat{F} = 3.66 \quad F_{2;13;0.95} = 3.80$
13	m-substituents	$y = 1.330 - 0.619\ x_1$ $R^2 = 0.9551 \quad \hat{F} = 148.98 \quad F_{1;7;0.95} = 5.59$
14	p-substituents	$y = 1.090 - 0.165\ x_1$ $R^2 = 0.1458 \quad \hat{F} = 1.02 \quad F_{1;6;0.95} = 5.99$
15	p-substituents	$y = 1.209 - 0.689\ x_1 - 0.314\ x_6{}^d$ $R^2 = 0.8980 \quad \hat{F} = 22.00 \quad F_{2;5;0.95} = 5.99$ $\hat{t}$: 6.49; 6.07 $\quad t_{5;0.95} = 2.57$
16	p-substituents	$y = 1.222 - 0.413\ x_1 - 0.301\ x_4$ $R^2 = 0.9116 \quad \hat{F} = 25.77 \quad F_{2;5;0.95} = 5.79$
17	p-substituents	$y = 1.282 - 0.308\ x_2 - 0.077\ x_3 - 0.179\ x_4$ $R^2 = 0.9808 \quad \hat{F} = 68.02 \quad F_{3;4;0.95} = 6.59$ $\hat{t}$: 10.67; 0.78; 4.34
17a	p-substituents	$y = 1.291 - 0.317\ x_2 - 0.151\ x_4$ $R^2 = 0.9779 \quad \hat{F} = 110.54 \quad F_{2;5;0.95} = 5.79$ $\hat{t}$: 12.62; 7.91 $\quad t_{5;0.95} = 2.57$
18	all substituents	$y = 1,262 - 0.583\ x_1 - 0.307\ x_6{}^d$ $R^2 = 0.7327 \quad \hat{F} = 17.82 \quad F_{2;13;0.95} = 3.80$ $\hat{t}$: 5.82; 5.02 $\quad t_{13;0.95} = 2.16$
19	p-NO_2 omitted	$y = 1,280 - 0.533\ x_1 - 0.377\ x_6$ $R^2 = 0.8726 \quad \hat{F} = 41.09 \quad F_{2;12;0.95} = 3.88$
20	p-NO_2 and m-OEt omitted	$y = 1.267 - 0.515\ x_1 - 0.362\ x_6$ $R^2 = 0.8918 \quad \hat{F} = 45.32 \quad F_{2;11;0.95} = 3.98$

Table 2. (*Contd.*)

Set	Regressand^{b,c}	Regression equation and parameters^{d}
21	all substituents	$y = 1.276 - 0.494\ x_1 - 0.367\ x_4$ $R^2 = 0.8749 \quad \hat{F} = 45.46 \quad F_{2;13;0.95} = 3.80$ $\hat{t}$: 8.45; 8.28 $\quad t_{13;0.95} = 2.16$
22	m-OEt omitted	$y = 1.263 - 0.477\ x_1 - 0.352\ x_4$ $R^2 = 0.8909 \quad \hat{F} = 49.01 \quad F_{2;12;0.95} = 3.88$

[a] rate constants taken from R. J. Hartman and A. G. Gassman, *J. Amer. Chem. Soc.* **62**, 1559 (1940); R. J. Hartman and A. M. Borders, *ibid.*, **59**, 2107 (1937); [b] basis set: see Table 1, footnote b and in addition *p*-NO_2, *p*-Br; [c] $y = 5 + \log k$; [d] x_5 = dummy variable with + 1 for *p*-F, *p*-OMe, *p*-OEt, *p*-Br and − 1 for *p*-NO_2, x_6 = dummy variable with + 1 for *p*-F, *p*-OMe, *p*-OEt and − 1 for *p*-NO_2, for other notations cp Table 1, footnotes d and e.

(set 14). Inclusion of the dummy variable x_6 (x_6: *p*-F, -OMe, *p*-OEt = + 1; *p*-NO_2 = − 1, set 15) or x_4 (set 16) gives a decisive improvement over set 14. From the orrelations the following conclusions may be drawn:

(i) the signs of the regression coefficients of x_6 and x_4, respectively, clearly indicate that the (+)-M effect of p-F, *p*-OMe, and *p*-OEt gives rise to a much more accentuated decrease of the reaction rates than it might have been expected on the basis of the general trend according to the σ values;

(ii) the particular mesomeric effect which the nitro group is able to exert should probably be of minor importance. These conclusions also rested on sets 17 and 17a which show that in the position para to the reaction center essentially the inductive effect acts, with the exception of *p*-F, *p*-OMe, and *p*-OEt. Therefore, the whole data material is not described in the best way using the linear combination σ and x_6 [cp sets 18, 19, 20; in set 18 the residual of *p*-NO_2 examined after computation of the standardized residuals (see Mager[17]) is significant at the 5% level] but if σ and x_4

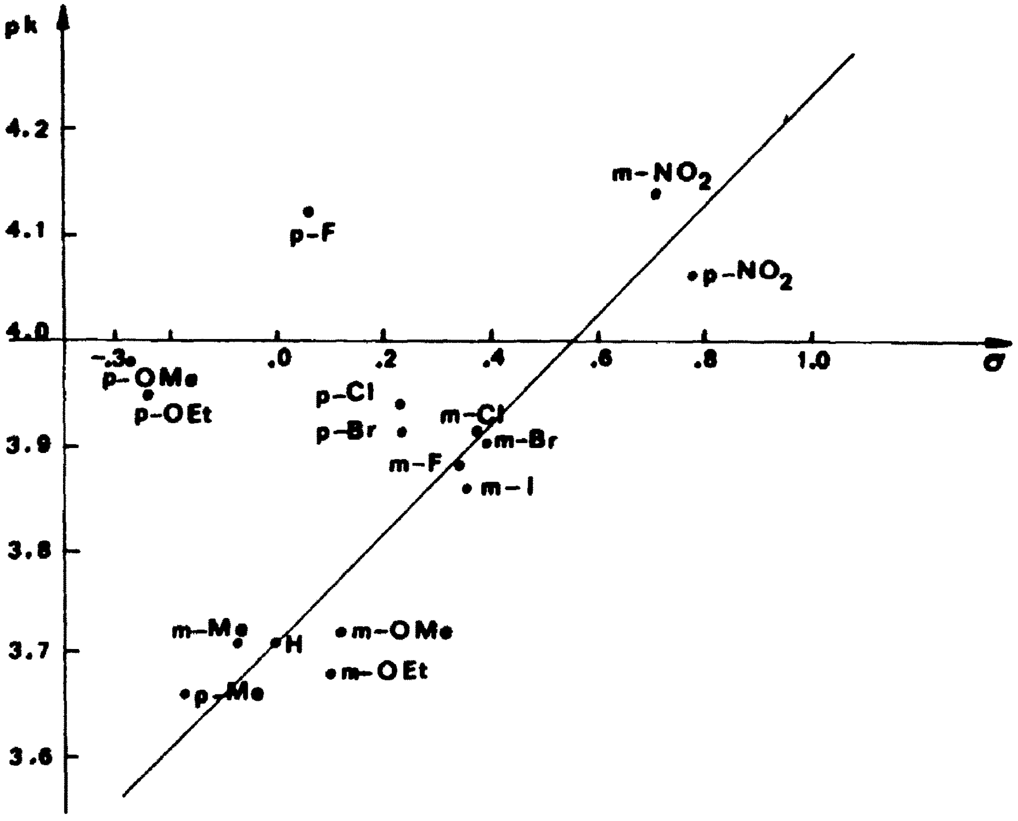

Fig. 2. Hammett plot for the reaction of *m*,*p*-substituted benzoic acids with methanol/HCl at 25°C representing the general trend.

were included (sets 20 and 21). In conclusion, we have found that the esterification reactions of *m,p*-substituted benzoic acids with cyclohexanol and with methanol, respectively, can be described with satisfactory precision by applying a Hammett type relationship where the resonance effects connected with such groups as *p*-F, *p*-OMe, and *p*-OEt in a particular manner are accounted for by use of a dummy variable. These special resonance affectuations are much more accentuated in the esterification reaction with methanol, leading in this case to an overwhelming retardation. It should be noted that a principal component analysis (see below) based on the rate data of: (i) the esterification with cyclohexanol at 55°C and 65°C and (ii) the esterification with methanol at 25°C and 40°C, and subsequent identification of the first principal component (eigenvalues: $l_1 = 61{,}93$; $l_2 = 0{,}74$; $l_3, l_4 \cong 0$) is in keeping with the previously cited results. The large eigenvalue of the first eigen vector indicates that an investigation of the esterification reactions at other temperatures should not reveal any new aspects.

Esterification of ortho substituted benzoic acids with cyclohexanol at 55°C

Recently, it was demonstrated[1] that the separation of inductive, mesomeric, and steric contributions to E_s^0 depends at least partly upon the sample size. This renders more difficult the investigation of many reactions of *o*-substituted compounds and is for example very conspicuous, if the correlation between E_s^0 and σ_0^* is regarded (Table 3). Assuming that σ_0^* actually represents a pure polar effect, then with set 26 one could conclude that E_s^0 is also purely polar in nature. Furthermore, the statistical analysis of the reactions of *o*-substituted compounds generally implies that the steric parameter used as one of the regressors (in most cases the van der Waals radius r_v and other measures deduced from it) actually may be regarded as a measure of the steric contribution to the so-called ortho effect. It is in the nature of matter that such an assumption could not have been verified up to now. Additionally, the variability of the van der Waals radii depending upon the method of determination is a well established fact.[8,9] Correlating the reactivity data of *o*-substituted benzoic acids we used the same values for r_v as in parts I.[1] However, it should be emphasized from the beginning that the use of the van der Waals radius of oxygen (1·52) for OMe and OEt is surely not entirely justified (cp Förster and Vögtle[9] and the literature cited therein), although we can find many papers in which the van der Waals radii for heteroatoms have been extended to the corresponding alkylated heteroatoms (cp Charton,[10] Fujita and Nishioka[11]). One of the many indications that such an extension should be somewhat questionable results from the comparison among the dipole moments of corresponding benzene and mesitylene derivatives.[12] Especially in the case of strong donors and acceptors the ranking of substituents according to their "size" may be very different. So there is no correlation between the n values which have been regarded as a measure of the size of the substituents[9] and r_v (Fig. 3 and Table 4, set 28). Only after elimination of the nitro and methoxy groups does the correlation become acceptable (sets 29 and 30).

In this connection the relationship between r_v and the linear combination of $\mathscr{R}$ and n is of interest (set 31), finally leading to a significant correlation between r_v and $\mathscr{R}$ (set 31a). From the correlations given above and by inspection of Figs. 3 and 4, evidence is accumulated that a reasonable discrimination among inductive, mesomeric, and steric effects is generally not possible if only substituents such as H, Me, Cl, Br, and I are included.

If the experimentalist seeks to distinguish between rival models, many benefits can be derived from an examination of linear relationships (multicollinearities) among independent variables and their influence on the estimated regression coefficients. Such multicollinearities often are present if the objective is a

Table 3. The dependence of the correlation between E_s^0 and σ_0^* upon the sample size

Set	Regressand[a]	Regression equation and parameters
23	all substituents	$E_s^0 = 0.335 - 1.276\,\sigma_o^*$ $R^2 = 0.7248 \quad \hat{F} = 18.43$
24	Me dropped	$E_s^0 = 0.422 - 1{,}434\,\sigma_o^*$ $R^2 = 0.8753 \quad \hat{F} = 42.11$
25	Me and F dropped	$E_s^0 = 0.369 - 1.494\,\sigma_o^*$ $R^2 = 0.9586 \quad \hat{F} = 115.81$
26	Me, F, and I dropped	$E_s^0 = 0.409 - 1.459\,\sigma_o^*$ $R^2 = 0.9919 \quad \hat{F} = 489.21$
27	Me, F, I, and NO_2 dropped	$E_s^0 = 0.401 - 1{,}455\,\sigma_o^*$ $R^2 = 0.9928 \quad \hat{F} = 411.24$

[a] basis set: OMe, OEt, F, Cl, Br, Me, I, NO_2, H.

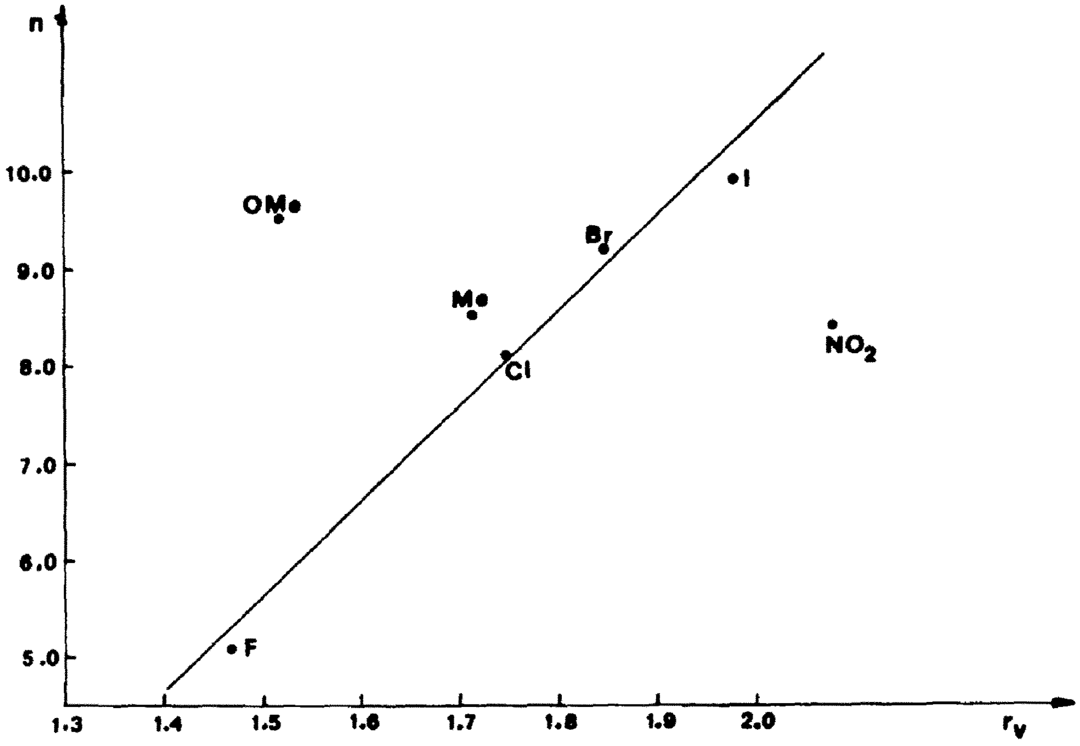

Fig. 3. Relationship between the n values and the van der Waals radii.

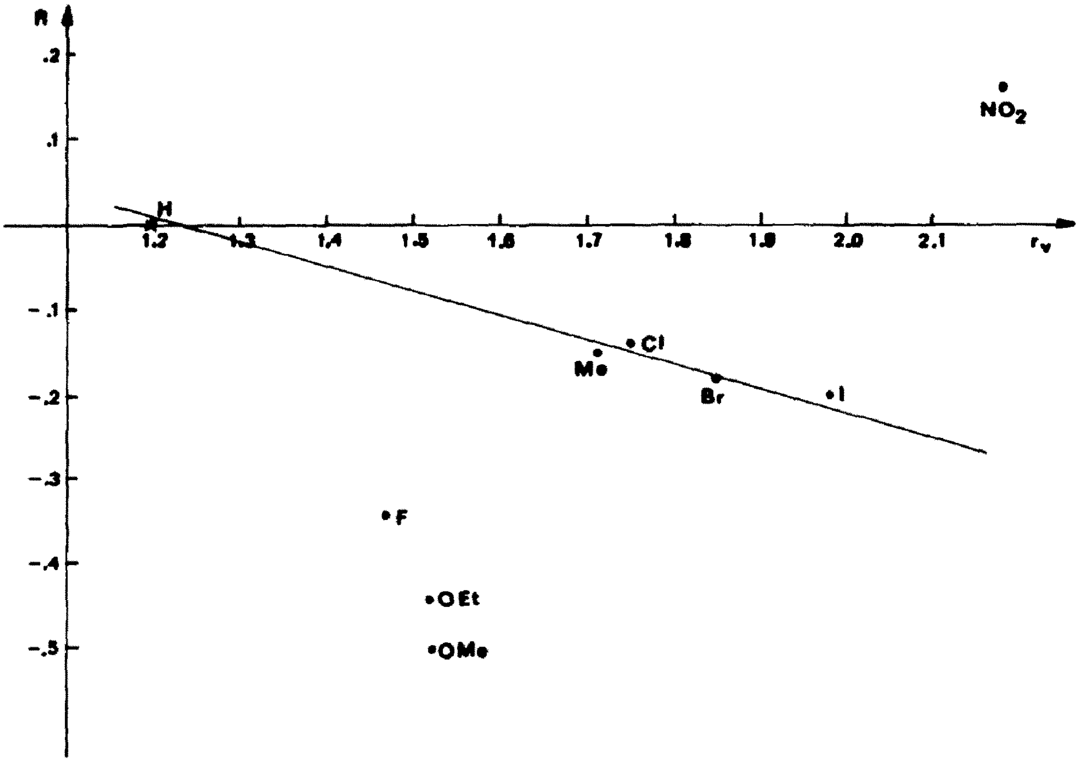

Fig. 4. Plot of the resonance constant $\mathscr{R}(x_R)$ vs r_v.

Table 4. Some simple relationships between the steric measures n and r_v

Set	Regressand[a, b]	Regression equation and parameters
28	all substituents	$n = 4.045 + 2.397\ r_v$ $R^2 = 0.1405\quad \hat{F} = 0.82\quad F_{1;5;0.95} = 6.61$
29	NO_2 omitted	$n = -1.560 + 5.801\ r_v$ $R^2 = 0.4186\quad \hat{F} = 2.88\quad F_{1;4;0.95} = 7.71$
30	NO_2 and OMe omitted	$n = -8.439 + 9.469\ r_v$ $R^2 = 0.9398\quad \hat{F} = 46.86\quad F_{1;3;0.95} = 10.13$
31	all substituents	$r_v = 1.476 + 1.070\ x_3 + 0.061\ n$ $R^2 = 0.8942\quad \hat{F} = 16.90\quad F_{2;4;0.95} = 6.94$ $\hat{t}$: 5.34; 2.41 $t_{4;0.95} = 2.78$
31a	all substituents	$r_v = 1.984 + 1.060\ x_3$ $R^2 = 0.7404\quad \hat{F} = 14.26\quad F_{1;5;0.95} = 6.61$

[a] *n* values taken from H. Förster and F. Vögtle, *Angew. Chem.* **89**, 443 (1977); [b] basis set: F, I, OMe, Me, NO_2, Br, Cl.

statistical examination of the ortho effect. In the same way as in the case of independent variables subject to measurement error where the error involved in x_j can influence the magnitude and even the sign of the regression coefficient $b_{j'}$ $(j' \neq j)$[7,13,14] it is also possible that in the case of regressors having negligible measurement errors certain values of the regressor x_j exert an important influence on both the size and the significance of the regression coefficient $b_{j'}$. In order to get a corresponding measure, especially which values

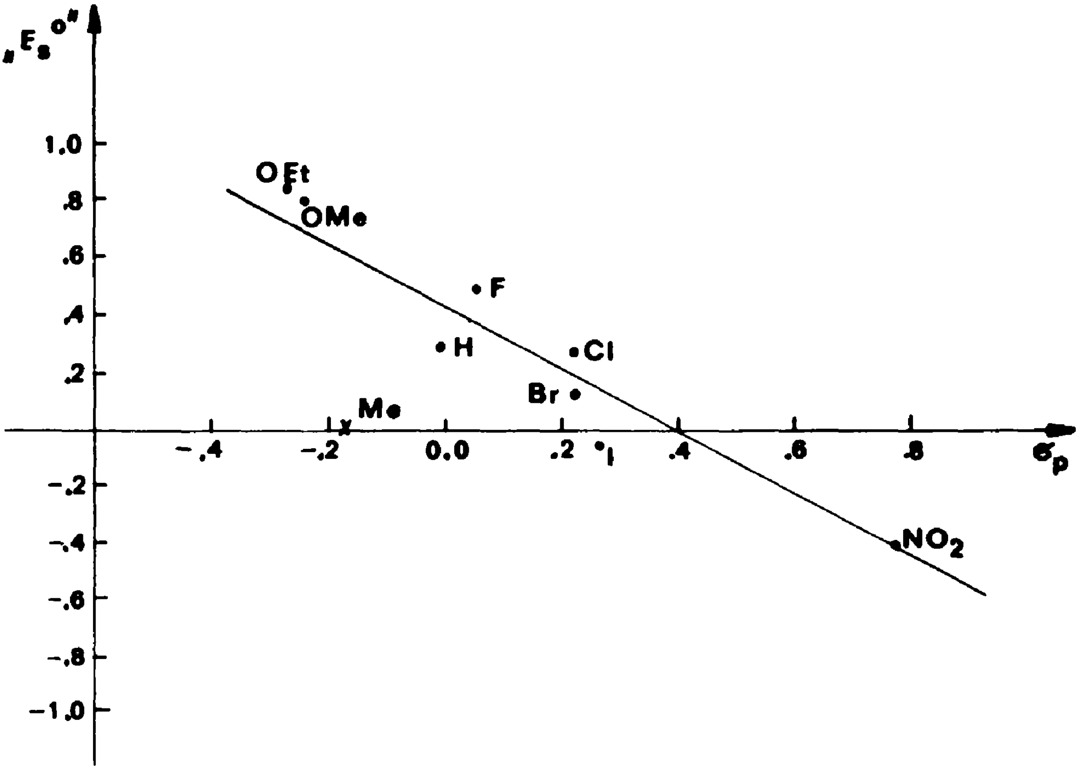

Fig. 5. "E_s^o" values estimated using the esterification of *o*-substituted benzoic acids with cyclohexanol/HCl at 55°C as the basis reaction vs σ_p.

Table 5. Esterification of o-substituted benzoic acids with cyclohexanol/HCl at 55°C[a]

Set	Regressand[b, c]	Regression equation and parameters[d]
32	all substituents	$y = 0.238 - 1.016\ x_1$ $R^2 = 0.6742 \quad \hat{F} = 14.49 \quad F_{1;7;0.95} = 5.59$
32a	Me omitted	$y = 0.331 - 1.220\ x_1$ $R^2 = 0.9312 \quad \hat{F} = 81.17 \quad F_{1;6;0.95} = 5.99$
33	all substituents	$y = 1.163 + 0.188\ x_2 - 1.277\ x_3 - 0.814\ r_v$ $R^2 = 0.9451 \quad \hat{F} = 28.71 \quad F_{3;5;0.95} = 5.41$ $\hat{t}$: 1.16; 5.59; 3.70 $\quad t_{5;0.95} = 2.57$
34	Me omitted	$y = 0.975 - 0.010\ x_2 - 1.281\ x_3 - 0.631\ r_v$ $R^2 = 0.9496 \quad \hat{F} = 25.12 \quad F_{3;4;0.95} = 6.59$ $\hat{t}$: 0.03; 5.38; 1.92 $\quad t_{4;0.95} = 2.78$
34a	Me omitted	$y = 0.985 - 1.280\ x_3 - 0.640\ r_v$ $R^2 = 0.9496 \quad \hat{F} = 47.09 \quad F_{2;5;0.95} = 5.79$ $\hat{t}$: 6.03; 4.32 $\quad t_{5;0.95} = 2.57$
35	H and OEt omitted	$y = 2.704 + 0.529\ x_2 + 0.102\ x_3 + 0.127\ n - 2.250\ r_v$ $R^2 = 0.9862 \quad \hat{F} = 35.62 \quad F_{4;2;0.95} = 19.25$ $\hat{t}$: 3.04; 0.16; 2.46; 3.38 $\quad t_{2;0.95} = 4.30$
35a	H and OEt omitted	$y = 2.574 + 0.510\ x_2 + 0.120\ n - 2.148\ r_v$ $R^2 = 0.9860 \quad \hat{F} = 70.32 \quad F_{3;3;0.95} = 9.28$ $\hat{t}$: 4.76; 5.47; 13.31 $\quad t_{3;0.95} = 3.18$
36	all substituents	$y = 1.169 + 0.376\ x_2 - 0.244\ x_3 - 0.783\ r_v + 0.430\ x_7{}^{e}$ $R^2 = 0.9848 \quad \hat{F} = 64.67 \quad F_{4;4;0.95} = 6.39$ $\hat{t}$: 3.36; 0.70; 6.02; 3.23 $\quad t_{4;0.95} = 2.78$
36a	all substituents	$y = 1.200 + 0.417\ x_2 - 0.790\ r_v + 0.516\ x_7{}^{e}$ $R^2 = 0.9829 \quad \hat{F} = 95.79 \quad F_{3;5;0.95} = 5.41$ $\hat{t}$: 4.62; 6.43; 10.54 $\quad t_{5;0.95} = 2.57$

[a] taken from R. J. Hartman, H. M. Hoogsteen and J. A. Moede, *J. Amer. Chem. Soc.* **66**, 1714 (1944); [b] basis set: H, F, I, OMe, OEt, Me, NO_2, Br, Cl; [c] $y = 5 + \log k$; [d] for the notations used cp footnotes d and e of Table 1; [e] x_7 = dummy variable with +1 for OMe, OEt and, −1 for NO_2.

x_{ij} (which substituents) cause such effects, the computation of the partial derivatives

$$\frac{\partial b_k}{\partial x_{ij}} = c_{kj}e_i - \left(\sum_{l=1}^{q} c_{kl}x_{il} \right) b_j \qquad (3)$$

can have advantages.[7] The differential quotients of eqn (3) reflect the sensitivity of the ith value of the regressor x_j with respect to the regression coefficient b_k, where

$$\{c_{kj}\} = (X'X)^{-1}$$

and

$$e_i = y_i - \hat{y}_i = y_i - \sum_{j=1}^{q} b_j x_{ij} - b_o.$$

These sensitivities may be surprisingly different, both in value and ranking, from the sensitivities to change in the dependent variable. They will in particular give details of those observations most liable to cause estimation error, which is especially important where it is hoped to interpret and use the estimated regression equation directly.

Plotting "E_s^o" values from the esterification with cyclohexanol at 55°C against σ_p displays that only the methyl group shows a marked deviation (Fig. 5). The methyl group is of essential significance in all investigations concerning the ortho effect and should in no case be omitted from the analyses. The same assertion holds for the hydrogen substituent: the special role which it plays following already from the consideration of the so-called "weighted squared distances" computed for multiple regressions concerning the reactions of o-substituted compounds.[15] The correlation of all data points with $\mathcal{F}$, $\mathcal{R}$, and r_v supplies as expected a subset, which is significant in $\mathcal{R}$ and r_v (cp set 33). This result is not altered by dropping the methyl group (sets 34 and 34a). By inspection of the partial derivatives according to eqn (3) the effects of the elimination of Me can be roughly predicted. (Table 6). According to these values the elimination of Me from the correlation should yield a change of the regression coefficients for the inductive and the steric terms with b_2 remaining approximately constant. This forecasting is confirmed by the equation obtained (set 34). The computed $\partial b_k/\partial x_{ij}$ reveal the interesting fact that the van der Waals radii of the methoxy and the ethoxy groups exert an important influence on b_2, the regression coefficient for the resonance term $\mathcal{R}$. This might be due to at least two different reasons: (i) the van der Waals radii of OMe and OEt are inappropriate and therefore, the steric information has partly to be governed by the resonance term, or, more probably; (ii) the rate constants of these groups do not represent a pure steric effect but some kind of resonance action or a mixture of both.

A very interesting phenomenon resulted from the correlation with $\mathcal{F}$, $\mathcal{R}$, n, and r_v, where $\mathcal{R}$ lost its significance (set 35, cp also sets 31 and 31a). An interpretation of the subset obtained in this way is not possible in a simple manner, but there is some indication that ranking the substituents according to their "sizes" may not lead in all cases to the same result, in other words, the overall steric effect is composite in nature (similar to the resonance effect) and exhibits some variation depending upon the reaction or physical property investigated. The linear combination of n and r_v can replace the resonance constant $\mathcal{R}$, but it is clear by inspection of the signs of the regression coefficients that it does not represent a uniform effect.

From the statistically based examination of the rate data for the esterifications of *m,p*-substituted benzoic acids, the conclusion was drawn that such substituents as *p*-F, *p*-OMe, and *p*-OEt exert special resonance effects which were taken into account using the dummy variable x_4. For the different structurations of the dummy variable possible from the physico-chemical viewpoint, multiple regression equations with $\mathcal{F}$, $\mathcal{R}$, and r_v have been computed (backward elimination procedure, ordinary least squares). The best subsets were obtained using the dummy variable x_7 (structuration: OMe, OEt $= +1$: $NO_2 = -1$) as it can be seen from sets 36 and 36a. The comparison between sets 36a and 3 shows that the inductive effect is in the same order in both cases (being very particular about this, only a comparison of the standardized regression coefficients[16] should be made).

Esterification of ortho substituted benzoic acids with methanol at 25°C

The corresponding correlations of the rate constants for the esterification of *o*-substituted benzoic acids with methanol/HCl gave essentially results comparable with those obtained previously. In the present case also the correlation with σ_p proved significant at the 5% level (Table 7, set 37) and the goodness of fit is considerably improved by dropping the methyl group (set 38). Correlation of the reactivity data with $\mathcal{F}$, $\mathcal{R}$, and r_v would give rise to the assumption that resonance and steric effects are predominant (set 39) and after deleting the methoxy, ethoxy, and hydrogen substituents only r_v remains significant (sets 40 and 40a), the latter result is not changed after resumption of H. The differential quotients of eqn (3) display essentially the same structure as in the esterification reaction with cyclohexanol. Introduction of the dummy variable x_7 yields a nonsignificant regression coefficient for $\mathcal{R}$ (set 41) and thus to a subset consisting of $\mathcal{F}$, r_v, and x_7 (41a).

Table 6. Selected partial derivatives according to eqn (3) of the regression coefficients computed for set 34[a]

$\frac{\partial b_1}{\partial x_{i2}}$	$\frac{\partial b_1}{\partial x_{i3}}$	$\frac{\partial b_2}{\partial x_{i3}}$	$\frac{\partial b_3}{\partial x_{i2}}$	$\frac{\partial b_3}{\partial x_{i3}}$	Group
-2.611887	-1.680011	-1.355786	5.127215	3.301430	H
-1.760319	-0.811362	-2.231075	5.567689	2.864598	F
-2.887278	-1.626154	-2.471044	7.523198	4.323512	I
-2.551133	-1.776460	-3.170020	6.386158	4.402083	OMe
-2.630332	-1.888681	-3.051481	6.345332	4.512070	OEt
-3.929002	-2.407842	-2.278171	7.767551	4.738165	Me
-2.091566	-1.461279	-1.762193	6.542253	4.452802	NO_2
-2.427678	-1.499503	-2.369463	6.681384	4.153458	Br
-2.247398	-1.601202	-2.306018	6.114679	4.269556	Cl

[a] The partial derivatives not included can be regarded as being nearly constant in this context (equally weighted).

Table 7. Esterification of *o*-substituted benzoic acids with methanol/HCl at 25°C[a]

Set	Regressand[b, c]	Regression equation and parameters[d]
37	all substituents	$y = 1.186 - 1.850\ x_1$ $R^2 = 0.8131\quad \hat{F} = 30.45\quad F_{1;7;0.95} = 5.59$
38	Me omitted	$y = 1.308 - 2.119\ x_1$ $R^2 = 0.9776\quad \hat{F} = 261.45\quad F_{1;6;0.95} = 5.99$
39	all substituents	$y = 2.733 - 0.073\ x_2 - 1.938\ x_3 - 1.230\ r_v$ $R^2 = 0.9663\quad \hat{F} = 47.77\quad F_{3;5;0.95} = 5.41$ $\hat{t}$: 0.35; 6.52; 4.30 $\quad t_{5;0.95} = 2.57$
40	H, OMe, OEt omitted	$y = 2.802 - 0.045\ x_2 - 1.358\ x_3 - 1.259\ r_v$ $R^2 = 0.9755\quad \hat{F} = 26.55\quad F_{3;2;0.95} = 19.16$ $\hat{t}$: 0.26; 2.11; 2.75 $\quad t_{2;0.95} = 4.30$
40a	H, OMe, OEt omitted	$y = 4.432 - 2.064\ r_v$ $R^2 = 0.9206\quad \hat{F} = 46.35\quad F_{1;4;0.95} = 7.71$
41	all substituents	$y = 2.744 + 0.193\ x_2 - 0.483\ x_3 - 1.187\ r_v + 0.604\ x_7$[e] $R^2 = 0.9945\quad \hat{F} = 180.89\quad F_{4;4;0.95} = 6.39$ $\hat{t}$: 1.73; 1.40; 9.15; 4.55 $\quad t_{4;0.95} = 2.78$
41a	all substituents	$y = 2.804 + 0.275\ x_2 - 1.202\ r_v + 0.775\ x_7$ $R^2 = 0.9918\quad \hat{F} = 202.13\quad F_{3;5;0.95} = 5.41$ $\hat{t}$: 2.66: 8.53; 13.82 $\quad t_{5;0.95} = 2.57$

[a] rate constants taken from R. J. Hartman and A. G. Gassman, *J. Amer. Chem. Soc.* **62**, 1559 (1940); R. J. Hartman and A. M. Borders, *ibid.*, **59**, 2107 (1937); [b] basis set: cp Table 5, footnote b; [c] $y = 5 + \log k$, [d] for the notations used cp footnotes d and e of table 1; [e] cp Table 5 footnote e

Principal component analysis

The principal component analysis is suitable for analysing dependence structures if one attempts to decry those hidden factors which have generated the dependencies. That is, the observable variables are represented as functions of a smaller number of latent factor variables, which can be used to generate the covariances among the variables from which we started. If that form is simple, and if the latent variables are few in number, a more economical description of the dependence structure can be obtained. By means of this method the data material can be concentrated and to this end the rate constants of the following esterification reactions were used: (i) esterification by cyclohexanol/HCl at 55 and 65°C; and (ii) esterification by methanol/HCl at 25 and 40°C. We denote the rate constants in the following way:

$5 + \log k$ (esterification by
cyclohexanol/HCl at 55°C) $= y_1$,
$5 + \log k$ (esterification by
cyclohexanol/HCl at 65°C) $= y_2$,

etc., and form the "covariance" matrix $Y'Y$ with

$$Y = (y_1 y_2 y_3 y_4),$$

where y_1 to y_4 are the vectors of the rate constants. The determinantel equation to be solved for each eigen value l_k then becomes

$$\det (Y'Y - l_k I) = 0$$
$$[I = \text{unit matrix of dimension } (4, 4).]$$

After having computed the eigenvalues the eigen vectors v_k were extracted from

$$(Y'Y - l_k I)v_k = 0.$$

We obtained the following eigenvalues: $l_1 = 40{\cdot}30$; $l_2 = 1{\cdot}14$; l_3, l_4 close to zero. Thus about 97% of the total "variance" of the rate constants in all reactions investigated can be accounted for by the first principal component z_1 with

$$z_1 = Yv_1$$

(for an introduction to principal component analysis cp Morrison[17]). The correlation of z_1 versus the corresponding physicochemical parameters yields entirely similar results as obtained using the primary data:

1. Correlation vs $\mathcal{F}$, $\mathcal{R}$, r_v

$$z_1 = 4.575 - 0.009\mathcal{F} - 3.029\mathcal{R} - 1.935r_v$$
$$R^2 = 0.9520, \quad \hat{F} = 33.09, \quad F_{3;5;0.95} = 5.41$$
$$\hat{t}(b_1) = 0.02, \quad \hat{t}(b_2) = 5.52, \quad \hat{t}(b_3) = 3.66$$

backward elimination

$$z_1 = 4.585 - 3.027\mathcal{R} - 1.944r_v$$
$$R^2 = 0.9520, \quad \hat{F} = 59.55, \quad F_{2;6,0.95} = 5.14$$
$$\hat{t}(b_2) = 6.10, \quad \hat{t}(b_3) = 5.59, \quad t_{6;0.95} = 2.45.$$

2. Correlation with $\mathcal{F}$, $\mathcal{R}$, r_v, x_7

$$z_1 = 4.592 + 0.484\mathcal{F} - 0.324\mathcal{R} - 1.853r_v + 1.124x_7,$$
$$R^2 = 0.9933, \quad \hat{F} = 148{,}38 > F_{4,4;0.95} = 6.39,$$
$$\hat{t}(b_1) = 2.54, \quad \hat{t}(b_2) = 0.55, \quad \hat{t}(b_3) = 8.37, \quad \hat{t}(b_4) = 4.97$$

backward elimination

$$z_1 = 4.632 + 0.538\mathcal{F} - 1.863r_v + 1.239x_7, \tag{4}$$
$$R^2 = 0.9928, \quad \hat{F} = 229.90 > F_{3;5;0.95} = 5.41,$$
$$\hat{t}(b_1) = 3.58, \quad \hat{t}(b_3) = 9.11, \quad \hat{t}(b_4) = 15.22, \quad t_{5,0.95} = 2.57.$$

From the equations presented the following conclusions can be drawn.

1. In both the reactions of *m,p*-substituted and *o*-substituted benzoic acids strong donor and/or acceptor substituents may exert special resonance effects which can be accounted for by use of various dummy variables.
2. The esterification rates of *o*-substituted benzoic acids depend upon inductive, mesomeric, and steric effects.
3. The appearance of the resonance term is mainly a result of peculiar resonance contributions performed by such substituents as methoxy, ethoxy, and nitro. If this fact is taken into account using an appropriate structured dummy variable, the resonance term can be dropped from the regression equations.
4. Taking away the special resonance effects connected with the above mentioned groups, the rate of esterification of *o*-substituted benzoic acids is essentially influenced by inductive and steric effects, where the influence of the former is of minor importance.

Table 8. Comparison among some measures of the steric substituent effect

Substituent	n^a	$r_v{}^b$	E_s^{oc}	$E_s^o(\text{corr.})^d$
H	-	1.20	-	0.79
F	5.1	1.47	0.49	0.42
OEt	-	1.52	0.90	0.29
OMe	9.5	1.52	0.99	0.34
Me	8.5	1.715	0.00	0.00
Cl	8.1	1.75	0.18	-0.17
Br	9.2	1.85	0.00	-0.42
I	9.9	1.98	-0.20	-0.74
NO_2	7.9	2.18	-0.75	-0.87

[a] taken from reference 9; [b] taken from reference 1; [c] taken from reference 2; [d] this paper.

5. The E_s^0 values of the ethoxy and the methoxy groups as given by Taft are actually composite in nature and should not be regarded as a measure of the steric effect.

These conclusions are somewhat contrary to those drawn previously.[1] However, the most important conclusion, namely that E_s^0 should generally represent a predominant steric effect, has not been affected and could be maintained. Since it has been demonstrated that the correlations of the esterification rates exhibit no uniform pattern, these reactions and the corresponding hydrolyses should not further be used for the estimation of steric substituent constants. Basing on eqn (4) modified steric substituent "constants" may be derived which are "corrected" for inductive effects and particular mesomeric contributions (Table 8).

Finally, it is our objective to emphasize vigorously that the values displayed in Table 8 are essentially based on the supposition that all regressors used in the correlations truly measure what they purport to measure, a supposition which is by no means self-evident (cp the discussion concerning the van der Waals radii).

REFERENCES

[1] Part I. H. Mager, P. P. Mager and A. Barth, *Tetrahedron* **35**, 1953 (1979).

[2] R. W. Taft, Separation of Polar, Steric, and Resonance Effects in Reactivity. In: M. S. Newman (Ed.), Steric Effects in Organic Chemistry, p. 556, Wiley, New York (1956).

[3] M. Charton, personal communication.

[4] R. W. Taft, *J. Amer. Chem. Soc.* **74**, 3120 (1952).

[5] C. G. Swain and E. C. Lupton, *Ibid.* **90**, 4328 (1968).

[6] D. H. McDaniel and H. C. Brown, *J. Org. Chem.* **23**, 420 (1958).

[7] H. Mager, Moderne Regressionsanalyse in Chemie und Biologie. Textbook, in preparation.

[8] H. A. Stuart, Molekülstruktur, p. 81. Springer, Berlin 1967.

[9] H. Förster and F. Vögtle, *Angew. Chem.* **89**, 443 (1977).

[10] M. Charton, *J. Amer. Chem. Soc.* **91**, 615 (1969).

[11] T. Fujita and T. Nishioka, The Analysis of the Ortho Effect. In: R. W. Taft (Ed.), *Progr. Phys. Org. Chem.* **12**, 49 (1976).

[12] H. G. O. Becker, Einführung in die Elektronentheorie organisch-chemischer Verbindungen, p. 69. Deutscher Verlag der Wissenschaften, Berlin 1974.

[13] H. Mager, Dissertation, Akademie der Wissenschaften der DDR, 1977.

[14] H. Mager, P. P. Mager and A. Barth, *Sci. Pharm.* **47**, 187 (1979).

[15] H. Mager, unpublished results (available from the author on request).

[16] H. Mager and A. Barth, *Pharmazie* **34**, 557 (1979).