Acta Genet Med Gemellol 36:297-312(1987)

Statistical Models for the Twinning Rate

J.O. Fellman[1], **A.W. Eriksson**[2]
[1] *Folkhälsan Institute of Genetics, Population Genetics Unit, Helsinki, Finland;* [2] *Institute of Human Genetics, Medical Faculty, Free University of Amsterdam, Netherlands*

Abstract. Linear regression models are used to explain the variations in the twinning rates. Data sets from different countries are analysed and maternal age, parity and marital status are the main regressors. The model building technique is also used in order to study the secular decline in the twinning rate. Linear regression technique makes it possible to compare the effect of different factors but the method requires sufficiently disaggregated data.

Key words: **Twinning rates, Linear regression models, Maternal age, Parity, Marital status**

INTRODUCTION

It is a well established fact that the human twinning rate varies to a great extent between different populations and within a given population during different time periods. These differences are mainly caused by the dizygotic (DZ) twinning rate. The monozygotic (MZ) twinning rate is rather constant for different populations and different time periods. In the literature these differences have been discussed and different explanations have been given. For a comprehensive discussion of the variations in the twinning rate see Eriksson [2] and the references given in that monography.

It is well known that the twinning rate and especially the DZ twinning rate depend on several factors. The most important are maternal age and birth order (parity). Other factors influencing the (DZ) twinning rate are the marital status and the social class of the mother. Especially, it has been observed that the twinning rate is higher in rural than in urban areas. Therefore, the secular decline in the twinning rate can to some extent be explained by an increasing urbanization.

All these factors have to be taken into account when the twinning rates for different populations are compared. Some attempt has been made to standardize the total twinning rate. In this way the effect of different maternal age distributions is reduced and the twinning rates for different populations can be compared [3].

This paper is intended to study the effect of the above mentioned factors by using statistical regression models.

MODEL BUILDING

In an earlier paper [6] linear regression models and logit models were built in order to study the effect of maternal age, parity and marital status on the twinning rate. The theory was applied on twinning data from Finland for the period 1953-1964. The linear regression model has the advantage that it is easy to explain the parameter estimates. On the other hand the linearity assumption may be too restrictive and the model assumptions cannot be used with full efficiency. These shortcomings are to a great extent eliminated if we use the logit model, which uses the distribution properties of the observed variables. The main result in our previous paper [6] was that both methods gave very similar results. Therefore, we now concentrate on the use of linear regression models.

In general, we start from a general model

$$T = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + e$$

where T is the (total, MZ or DZ) twinning rate per mil; $X_1, \dots, X_k$ are potential regressors, which may explain the twinning rate; $\beta_0, \dots, \beta_k$ are unknown parameters, which have to be estimated; and e is the error term.

In the set of regressors $X_1, \dots, X_k$ you may include variables that measure maternal age, parity, marital status of the mother, social class, time, etc.

Usually the estimation procedure starts from a large set of regressors. They are step by step included in the model. The inclusion demands that the corresponding parameter estimate differs significantly from zero. At some stage the process stops and a model with statistically significant coefficient estimates is obtained. This procedure can be forced to include regressors, which are of special interest although their coefficients are not significant.

If we intend to explain the twinning rate, then

$$\mathrm{Var}\,(e) = \frac{E(T)(1000 - E(T))}{n},$$

where n is the number of maternities in the cell in question and E(T) is the expected twinning rate (per mil). This indicates heteroscedasticity and weighted least squares (WLS) should be used. However, E(T) is not known and full efficiency cannot be obtained. Under such circumstances good alternative weights are the number of maternities (n).

We assume that maternal age obtains the values of the mid-points of the age classes. If parity as such is used as an explanatory variable we assume that the effect of parity is linear. In an earlier study [6] it was observed that this is not always acceptable and a modified parity variable was used. This linearity assumption can be totally avoided if we introduce the dummy variable technique. This technique means that we introduce variables that measure the difference between the effect of the different parity levels and the effect of the parity level one.

In the following we use data sets from different countries in order to investigate how the model building works. A major shortcoming is that the composition of the data sets from different countries differs markedly. Hence, we have great difficulties to compare the twinning rates from different countries. As a consequence of this we concentrate in this study on the model building technique and on how it can solve different problems.

RESULTS AND DISCUSSION

Finland

These data have been earlier studied in several papers [3,4,6]. The data consist of twinning data for Finland for the time period 1953-1964. The data are pooled over the whole time period an therefore secular changes cannot be studied. Maternal age, parity and the marital status of the mother have been registered. The complete set of data is published in Eriksson & Fellman [3].

In the first two studies the data were rather superficially analysed and only statistically significant effects of maternal age, marital status and parity were established. In a later paper by Fellman and Eriksson (1985) these data were analysed more throughly. In that paper we built linear regression models and logit models for the whole data set and separately for the data for married and unmarried mothers, respectively. The linear regression models and the logit models gave quite similar results. We observed a linear parity effect up to level 5. For higher levels the effect was rather constant. Therefore, we used in the models the modified birth order variable

$$P_5 = \begin{cases} \text{Parity if parity} \leqslant 5 \\ 5 \qquad \text{if parity} > 5 \end{cases}$$

Furthermore, we included the interactions between marital status and maternal age and parity, respectively. The detailed results can be seen in Fellman and Eriksson [6]. In a recent paper, Kostense et al [8] use logit models on Dutch data. They conclude that the parity effect is mainly a consequence of the strong correlation between maternal age and parity.

The analysis of our data has been continued. In order to obtain the best model within the most general set of linear regression models, we now include in the set of potential factors all types of interactions (age × marital status, age × parity, parity × marital status, age × parity × marital status).

Assuming that the logit models give quite similar results, we consider in this study only linear regression models. Regression models are built for the total set of Data (Model 1, Model 2, and Model 3) and for the legitimate data (Model L) and for the illegitimate data (Model I).

The best model (Model 1) for the total set of data contains both the original parity variable and the modified one. A model with entire use of only one of these variables is more easily interpreted. Therefore, Model 1 is compared with the best model containing the original parity variable (Model 2) and with the best model containing the modified parity variable (Model 3). The obtained models are as follows (the standard errors of the estimates are given in brackets):

Model 1

$$T = -5.33885 + \underset{(0.0385)}{0.63596}\ AGE + \underset{(0.0043)}{0.03988}\ AGE \times P_5 + \underset{(0.0150)}{0.08225}\ AGE \times MARSTAT \times PARITY$$

$$\bar{R}^2 = 0.9358$$

Model 2

$$T = -7.57536 + \underset{(0.0352)}{0.73754}\ AGE + \underset{(0.1120)}{0.87608}\ PARITY + \underset{(0.0162)}{0.08038}\ AGE \times MARSTAT \times PARITY$$

$$\bar{R}^2 = 0.9245$$

Model 3

$$T = -5.35921 + \underset{(0.0390)}{0.63013}\ AGE + \underset{(0.0044)}{0.04195}\ AGE \times P_5 + \underset{(0.0295)}{0.15475}\ AGE \times MARSTAT$$

$$\bar{R}^2 = 0.9341$$

Model L

$$T = -5.07230 + \underset{(0.0390)}{0.62033}\ AGE + \underset{(0.0043)}{0.04189}\ AGE \times P_5 \qquad \bar{R}^2 = 0.9643$$

Model I

$$T = -8.74226 + \underset{(0.1436)}{0.85832}\ AGE + \underset{(0.0248)}{0.07718}\ AGE \times PARITY \qquad \bar{R}^2 = 0.6919$$

Using the adjusted coefficient of determination, $\bar{R}^2$, as criterion we observe that none of Model 1, Model 2 and Model 3 is distinctly better than the other two. Following the ideas in the earlier study [6], the final comparison of the models is based on how well the expected pooled twinning rates for the different age groups correspond to the observed rates for legitimate data (Table 1) and illegitimate data (Table 2). In these tables we also compare the expected total twinning rates with the observed ones. It is to be noted that the expected total twinning rates contain the observed twin maternities for mothers in the age groups over 40 years and for mothers with unknown age. (These age groups were not included in the model building). We observe that all the models fit the observed legitimate data rather well. For illegitimate data the fit is not so good (Table 2, Fig. 1). There are two reasons for this. First, the illegitimate data set is markedly smaller than the data set for legitimate maternities. Therefore it is disturbed by greater random fluctuations. Furthermore, the legitimate data set dominates the total data set. Finally, we observe by comparing Model L and Model I that the parity effect is markedly stronger among illegitimate maternities than among legitimate maternities. Figure 2 shows the goodness of fit of the legitimate and the illegitimate model.

Table 1. Observed and estimated twinning rates in relation to maternal age among legitimate maternities in Finland, 1953-1964. The estimated total twinning rates include the observed number of twin maternities for mothers in the age groups above 40 years and for mothers with unknown age

Age Group	Twinning rates				
	Observed	Estimated according to			
		Model 1	Model 2	Model 3	Model L
-17	4.96	5.49	5.13	5.41	5.54
18-19	7.03	7.62	7.45	7.54	7.64
20-24	10.31	10.43	10.44	10.35	10.41
25-29	14.99	14.75	14.82	14.70	14.72
30-34	19.60	19.36	19.33	19.36	19.32
35-39	23.29	23.95	23.87	23.99	23.91
40-44	17.47				
45-	6.68				
Total	15.15	15.17	15.17	15.14	15.15

Table 2. Observed and estimated twinning rates in relation to maternal age among illegitimate maternities in Finland, 1953-1964. The estimated total twinning rates include the observed number of twin maternities for mothers in the age groups above 40 years and for mothers with unknown age

Age Group	Twinning rates				
	Observed	Estimated according to			
		Model 1	Model 2	Model 3	Model L
-17	6.18	6.83	6.43	7.88	6.25
18-19	9.74	9.19	8.97	10.39	9.11
20-24	12.28	12.30	12.28	13.44	12.68
25-29	17.57	17.25	17.41	17.96	18.09
30-34	26.50	22.46	22.77	22.52	23.73
35-39	26.93	27.42	27.83	26.99	29.21
40-44	10.03				
45-	0.00				
Total	15.56	15.01	15.04	15.79	15.55

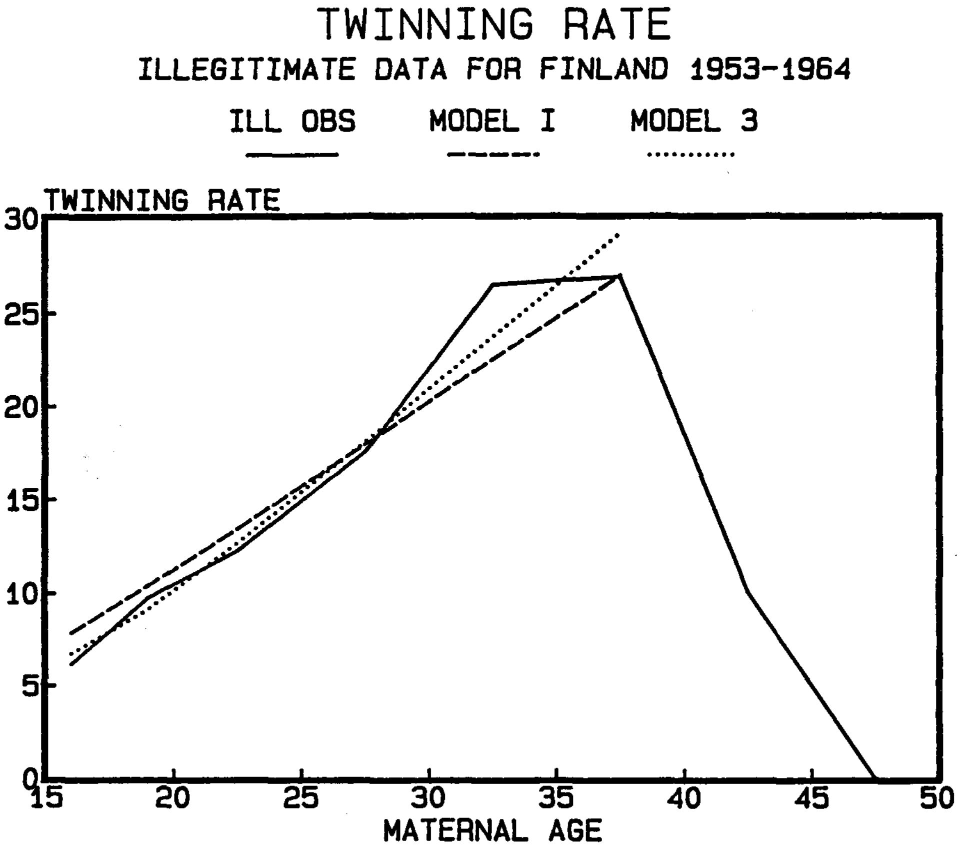

Fig. 1. Age-specific twinning rates among illegitimate maternities in Finland, 1953-1964. The observed data (ILL OBS) are compared with the values calculated according to Model 3 (MODEL 3) and the illegitimate model (MODEL I).

Denmark

The twinning data from Denmark consist of yearly data for the period 1973-1984. The total number of births and twin births are classified according to:
- maternal age
- marital status of the mother
- sex composition of the twin pairs
- year.

Maternal age is given in one-year classes but the sex compositions of the twin pairs are classified only for five-year maternal age groups and therefore only five-year age classes are used. Information about parity was not available. The number of MZ and DZ twinning maternities was estimated according to Weinberg's law. For the MZ, DZ and total twinning rates regression models were built. Maternal age, marital status and time were used as basic regressors. In order to study interactions between these variates we included in the analysis also the regressors:

https://doi.org/10.1017/S000156600000605X Published online by Cambridge University Press

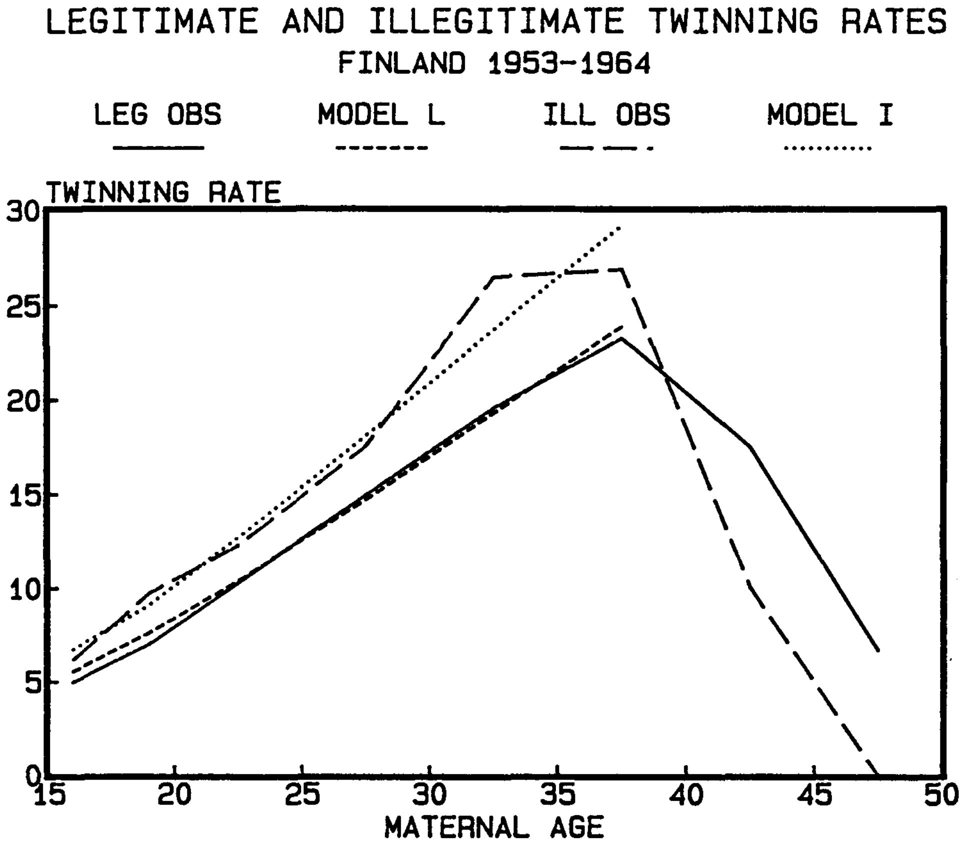

Fig. 2. Age-specific twinning rates among legitimate and illegitimate maternities in Finland, 1953-1954. The observed data (LEG OBS and ILL OBS) are compared with the data calculated according to the legitimate model (MODEL L) and the illegitimate model (MODEL I), respectively.

– time × age
– time × marital status
– age × marital status
– age × time × marital status.

The results were rather surprising. Using WLS and the stepwise estimation procedure only maternal age was accepted as a statistically significant regressor. No statistical significant time effect or marital status effect were discernible. We obtained the following models:

$$T_{TOT} = -2.78580 + \underset{(0.0257)}{0.47975}\ AGE \qquad \bar{R}^2 = 0.7456$$

$$T_{DZ} = -5.37828 + \underset{(0.0289)}{0.41366}\ AGE \qquad \bar{R}^2 = 0.6312$$

$$T_{MZ} = +2.59248 + \underset{(0.0257)}{0.06609}\ AGE \qquad \bar{R}^2 = 0.0450$$

It is remarkable that in the model for the MZ twinning rate, the parameter estimate differs statistically significantly from zero (t = 2.573; 118 DF; $P < 0.02$). However, the low $\bar{R}^2$ value indicates that the association between the MZ twinning rate and maternal age is very slight. Fig. 3 shows the goodness of fit. A more remarkable result is that there seems to be no association between twinning rates and marital status. In several studies such an association has been observed [3,4,6]. This contradiction to earlier findings may have two explanations. First, our analysis does not take parity into account. The married mothers may have on average higher parity and this may to some extent compensate higher age-specific twinning rates among unmarried mothers.

According to our opinion another explanation is more credible. We think that the classification, marital status, has lost its original meaning. In Denmark today especially younger couples live as married couples with no formal wedding. However, they are classified as unmarried. This tendency increases continuously. This can be indirectly observed in our series. During the period 1973-1984 the proportion of maternities of unmarried mothers has increased from 17.2% to 42.1% (Table 3 and Fig. 4).

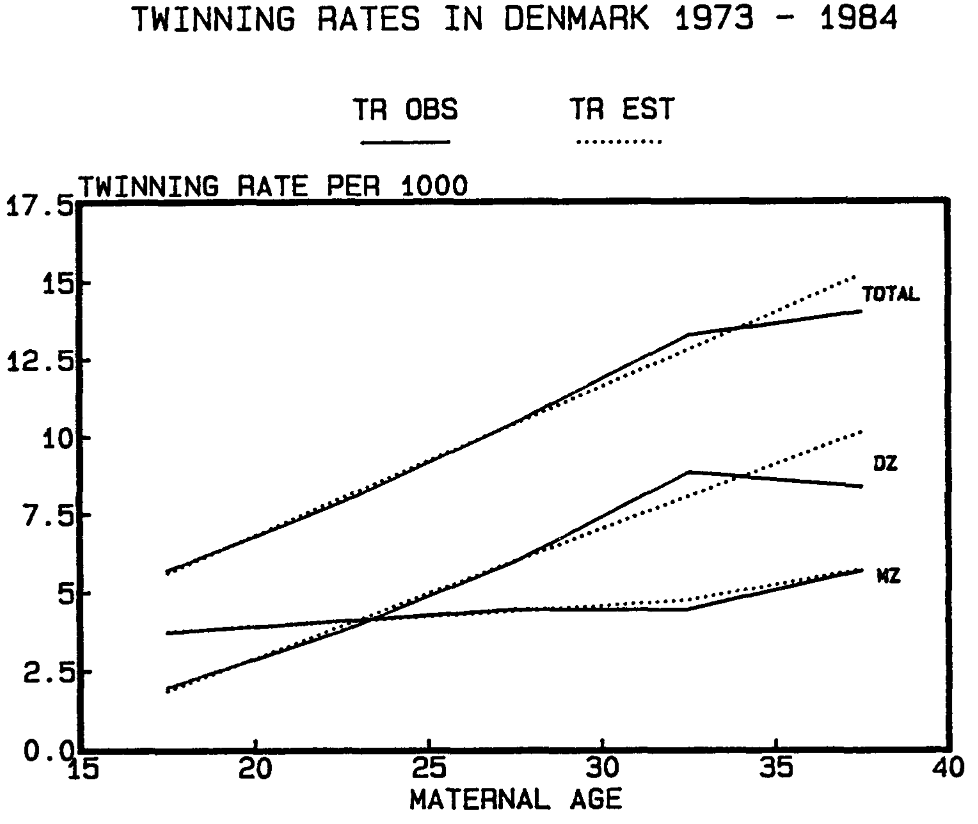

Fig. 3. Age-specific twinning rates in Denmark, 1973-1984. The rates for MZ twins, DZ twins and all twins are compared with the values calculated according to the obtained models.

Table 3. Summary statistics for the twinning rates in Denmark, 1973-1984.

Year	Illeg. births/1000	Maternal mean age (yr)			Total twinning rate								
					MZ			DZ			TOTAL		
		Legit.	Illeg.	Total	Legit.	Illeg.	Total	Legit.	Illeg.	Total	Legit.	Illeg.	Total
1973	17.2	26.86	23.55	26.30	4.30	2.94	4.06	6.05	4.30	5.68	10.35	7.24	9.75
1974	18.9	26.93	23.60	26.31	4.13	4.51	4.20	5.77	5.11	5.65	9.90	9.62	9.84
1975	21.8	27.01	23.59	26.26	4.41	5.27	4.60	5.27	3.85	4.96	9.69	9.12	9.56
1976	24.1	27.29	23.85	26.46	3.61	4.18	3.75	6.24	4.63	5.85	9.85	8.80	9.60
1977	26.0	27.37	23.93	26.48	4.20	4.71	4.33	5.74	4.02	5.29	9.94	8.72	9.62
1978	28.0	27.50	24.29	26.60	5.11	4.66	4.98	5.43	4.19	5.08	10.54	8.85	10.07
1979	30.8	27.69	24.50	26.71	4.90	3.26	4.39	6.20	5.97	6.13	11.10	9.22	10.52
1980	33.3	27.81	24.74	26.79	4.77	4.19	4.58	6.10	4.78	5.57	10.87	8.97	10.24
1981	35.9	27.94	25.05	26.90	4.25	4.36	4.29	5.94	4.89	5.56	10.19	9.25	9.86
1982	38.5	28.20	25.31	27.08	4.60	4.65	4.62	6.88	3.40	5.54	11.49	8.05	10.16
1983	40.7	28.37	25.61	27.23	3.90	3.97	3.93	6.46	6.07	6.30	10.36	10.04	10.23
1984	42.1	28.59	25.80	27.40	4.98	4.09	4.61	6.91	5.86	6.47	11.89	9.96	11.07

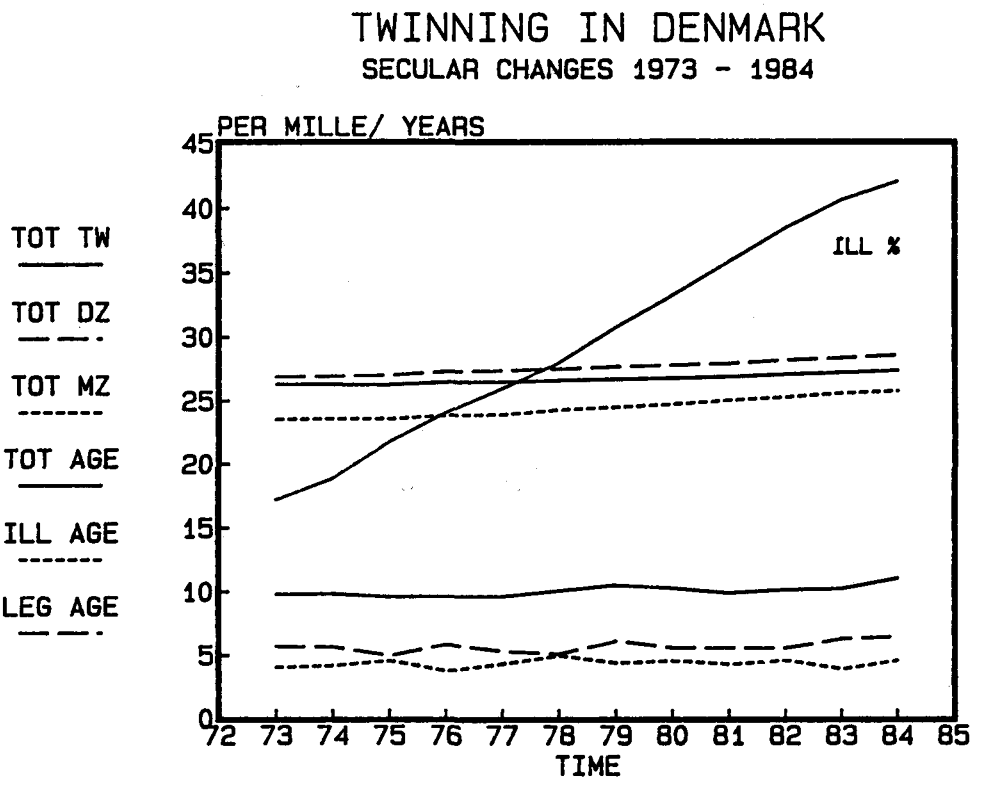

Fig. 4. Summary twin statistics for Denmark, 1973-1984 (see Table 3).

Italy

James [7] gives twinning data for Italy. He gives the rate of opposite-sex twins per 1000 legitimate maternities for the years 1957-1969. The series are grouped according to maternal age and parity. He observes a secular decline and he states that this is not solely due to secular changes in birth order or maternal age. He observes a general tendency that there is a secular decline in every parity and maternal age group. He also observes that this decline varies from group to group. He cannot obtain an association between the decline and social class. His problem will be analysed by our estimation techniques. However, he does not give the total number of births. Hence, the estimation of the parameters in the models must be based on ordinary least squares (OLS). Given that the model is correct, OLS gives parameter estimates, which are unbiased but inefficient.

In our study we build three different models. In the first one we assume that the decline is due to a decline in the age-specific twinning rates and the parity effects are constant. We consider the following potential regressors:

– maternal age
– time × maternal age, where time = year -1956
– dummy variables $D_1,...,D_6$, which correspond to the different parity levels.

In the second model we assume that the decline is due solely to secular changes in the parity effect. We allow the model to contain the regressors:
- maternal age
- the dummies $D_1,...,D_6$
- the variates $D_i \times$ time (i = 1,...,6).

In the third model we allow declines both in the parity effect and in the maternal age effect. We consider the following potential regressors:
- maternal age
- time
- parity dummies $D_1,...,D_6$
- age × time
- $D_i \times$ time (i = 1,...,6)
- $D_i \times$ age (i = 1,...,6)
- $D_i \times$ age × time (i = 1,...,6)

We obtain the best model:

$$T = -0.57024 - \underset{(0.0101)}{0.08040}\ \text{TIME} +$$

$$+ [\underset{(0.0075)}{0.24592} - \underset{(0.0037)}{0.09855}\ D_1 - \underset{(0.0037)}{0.08856}\ D_2 - \underset{(0.0038)}{0.05925}\ D_3 - \underset{(0.0038)}{0.03098}\ D_4]\ \text{AGE}$$

$$\bar{R}^2 = 0.9122$$

If we compare the three attempts and use the adjusted $\bar{R}^2$ as criterion, the second one is slightly better than the first and the third one. In Fig. 5 we see how well the third model fits the declining effect of parity and age. Fig. 5 is a modified version of the figures given by James in his paper [7].

Australia

In a recent paper Doherty and Lancaster [1] study the secular changes of the twinning rate in Australia during 1907-1982. They give the total, the age-specific, and the standardized twinning rate.

The data for the year 1907 are not complete and were excluded. Using the remaining data we give an alternative method of standardization. Assume that the age-specific twinning rate is a linear function of maternal age up to age 40 years. Only a very few maternities involve mothers over 40 years of age, so these can be excluded. The disturbing effect of these two approximations is very slight.

If mean maternal age is μ_A, then the total twinning rate (TTR) satisfies the approximate equation

$$\text{TTR} \simeq \alpha + \beta\mu_A$$

where α and β are the parameters in the model

$$\text{TR} = \alpha + \beta\ \text{AGE}$$

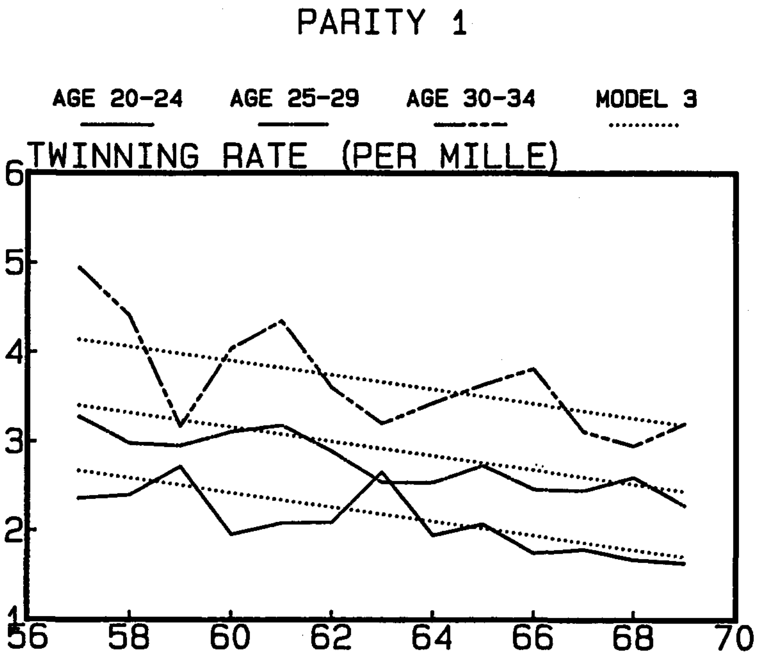

FIGURE a

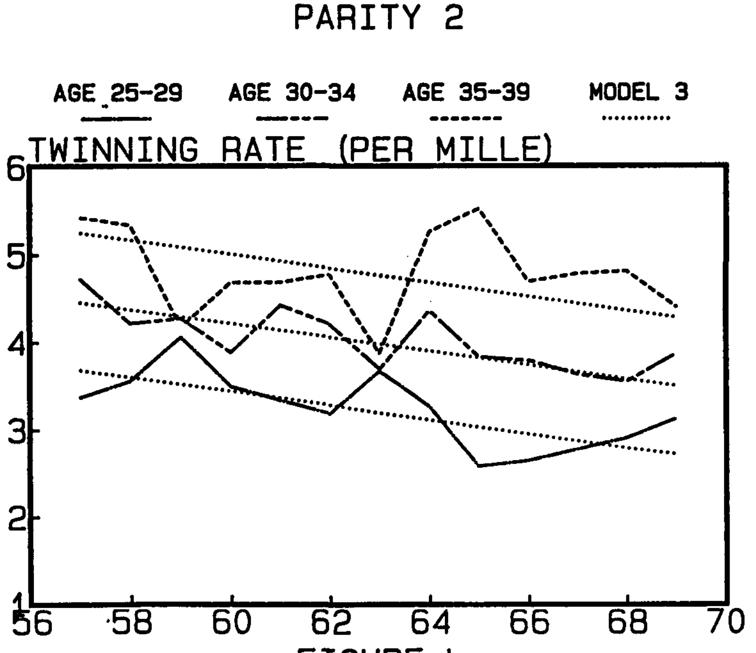

FIGURE b

https://doi.org/10.1017/S000156600000605X Published online by Cambridge University Press

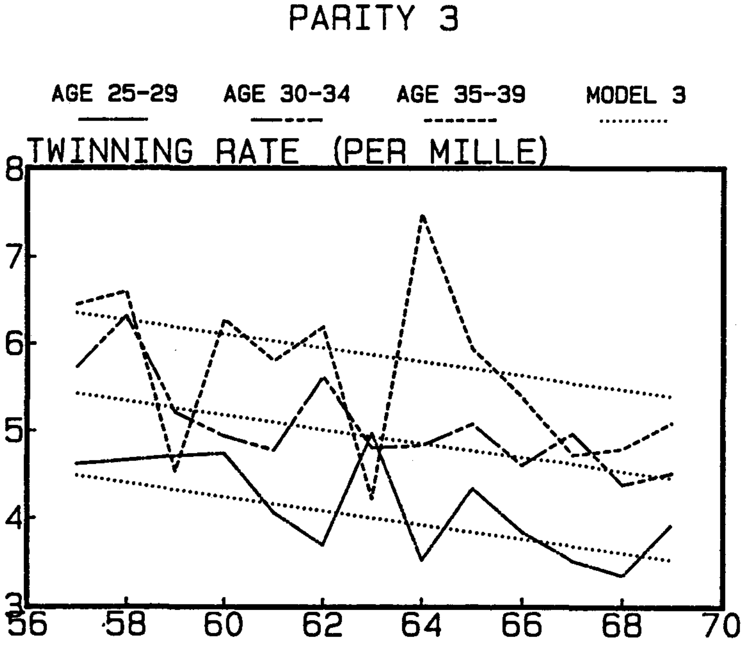

FIGURE c

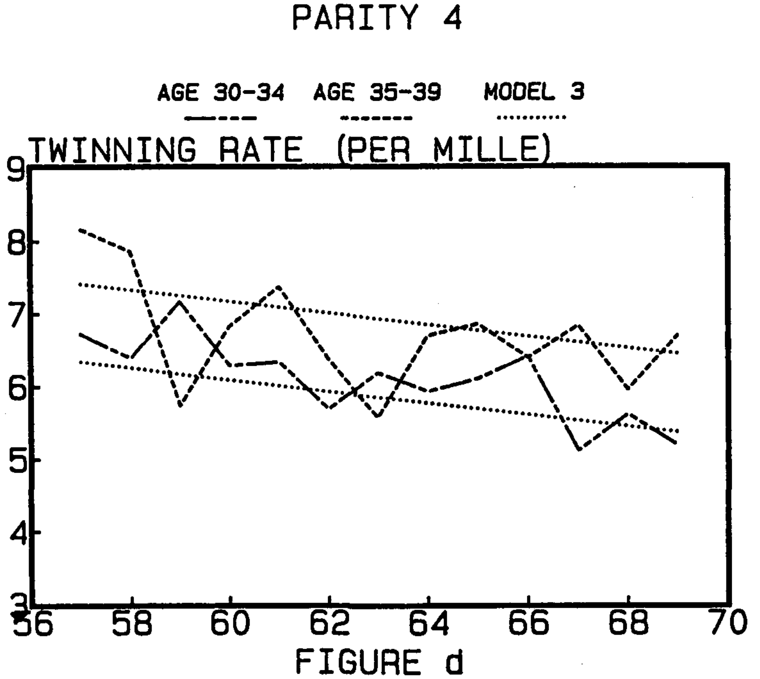

FIGURE d

Fig. 5. The age- and parity-specific twinning rates in Italy, given by James (1975), compared with the values calculated according to our third model.

where TR is the age-specific twinning rate. If we estimate α and β for every five-year time period and we use a common maternal age, μ_s, then the formula

$$STR \simeq \alpha + \beta\mu_s$$

gives standardized twinning rates for the different time periods. In Fig. 6 we compare our standardization (μ_s = 27.5 years) with the observed series and with the standardized series given by Doherty and Lancaster [1].

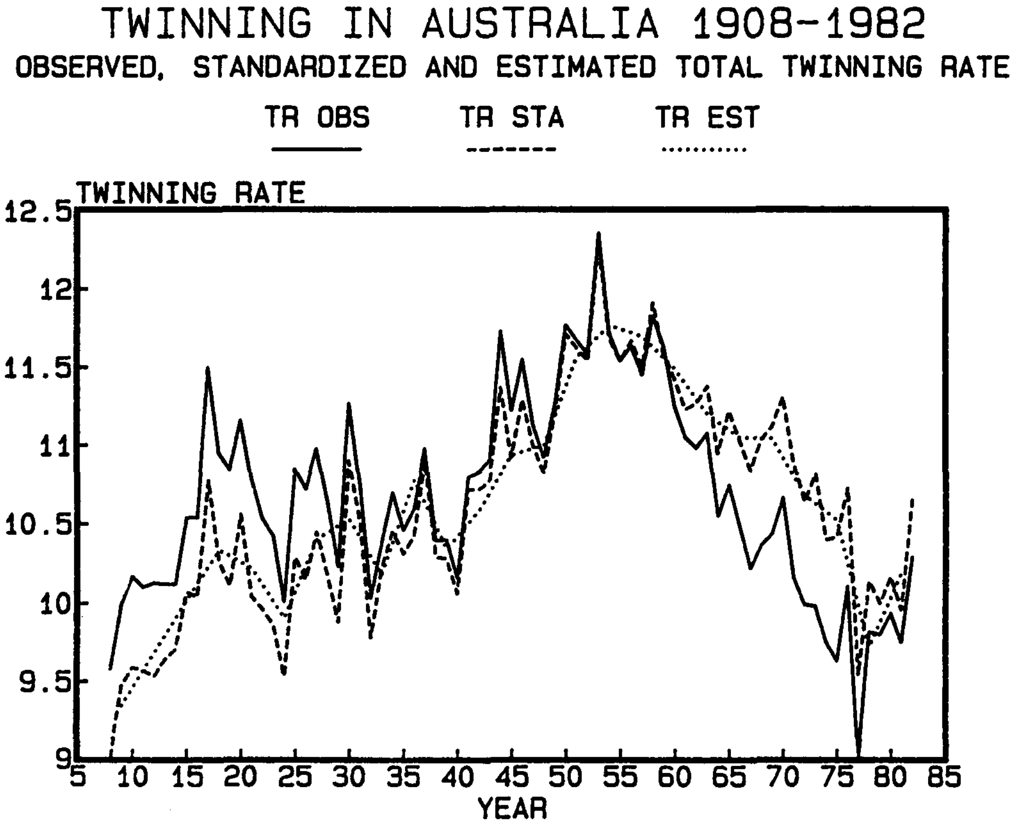

Fig. 6. The secular change in the twinning rate in Australia, 1908-1982. Observed rates (TR OBS) are compared with the rates standardized by Doherty and Lancaster (1986) (TR STA) and by the authors (TR EST).

The Aggregated Data

This last example shows that the aggregated data, which are standard in the yearbooks of the Central Statistical Offices in different countries, are rather worthless for this kind of model building. Our data consist of yearly data of the total twinning rate in different provinces in Finland for the period 1974-1983. Furthermore, we have information about mean maternal age, mean parity, the percentage of illegitimate maternities and the urbanization level for these provinces and for these years. These factors can be expected to be potential regressors. In all, we have 120 twinning rate values. There are statistically significant differences in the twinning rate for different provinces. Our attempt to build a model in order to measure age, parity, urbanization and marital status effects failed. The only significant effect is the urbanization level (U). The obtained model is

$$T = 11.94414 - \underset{(0.8379)}{2.0346}\, U \qquad \bar{R}^2 = 0.0391$$

We observe that the coefficient is significant and has the logical sign. However, we also observe that the model is a very poor one ($\bar{R}^2 = 0.0391$).

Fig. 7 shows the regional distribution of the twinning rate, the regional distribution of the urbanization level and the mean residuals between the observed and the expected twinning rate. For some of the provinces there is a good agreement between the model and the observed twinning rate but for some provinces the discrepancies are marked.

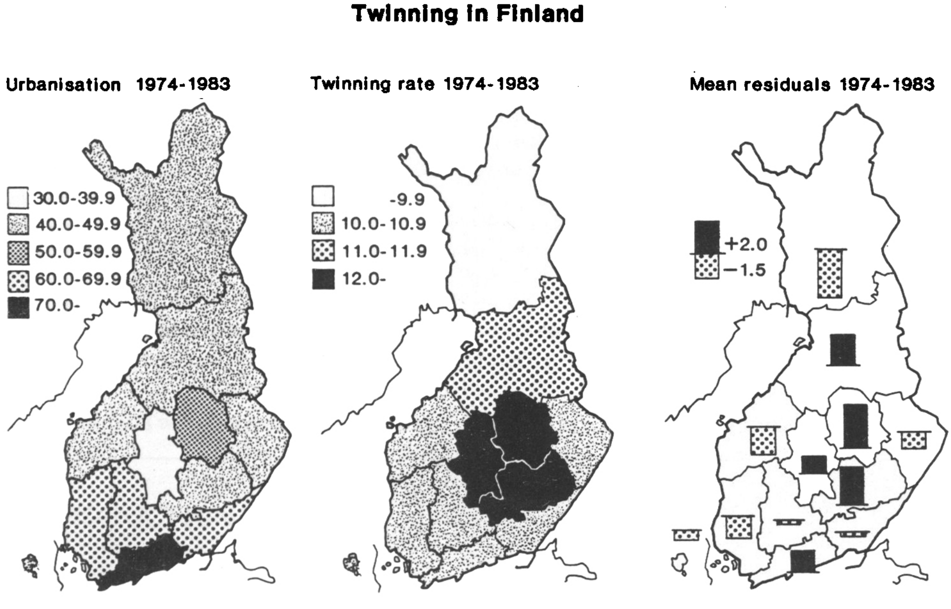

Fig. 7. The regional distribution of the urbanization level, of the twinning rate and of the mean residuals between the observed and estimated twinning rates.

CONCLUSIONS

The composition of the data from the different countries differs to a great extent. Therefore, it is difficult to make more exact comparisons between the different countries. Usually the only common regressor is maternal age. However, this study shows that the

https://doi.org/10.1017/S000156600000605X Published online by Cambridge University Press

proposed method gives promising results. The most important advantage of this model building technique is that it takes simultaneously in account all the factors and measures their effects. Therefore, it is a valuable method when the effects of different factors have to be compared. However, the method requires sufficiently disaggregated data.

Acknowledgements. The HP3000 computer at the Swedish School of Economics and Business Administration and the BMDP statistical software were used for the statistical analysis. For the maps in Fig. 7 valuable help was given by Kerstin Busck-Nielsen, MBA.

REFERENCES

1. Doherty JDH, Lancaster PAL (1986): The secular trend of twinning in Australia, 1853-1982. Acta Genet Med Gemellol 35: 61-76.
2. Eriksson AW (1973): Human twinning in and around the Åland Islands. Commentationes Biologicae 64: 1-154.
3. Eriksson AW, Fellman J (1967): Twinning and legitimacy. Hereditas 57: 395-402.
4. Eriksson AW, Fellman J (1967): Twinning in relation to the marital status of the mother. Acta Genet Statist Med 17: 385-398.
5. Eriksson AW, Fellman J (1973): Differences in the twinning trends between Finns and Swedes. Am J Hum Genet 25: 141-151.
6. Fellman J, Eriksson AW (1985): Factors affecting the birth prevalence of twins. Paper read at the Nordic Symposium on Biometrical Problems in Population Biology, Asker, Norway, Jan. 14-15.
7. James WH (1975): The secular decline in dizygotic twinning rates in Italy. Acta Genet Med Gemellol 24: 9-14.
8. Kostense PJ, Bressers M, Tas RFJ, Eriksson AW (1986): Maternal age and parity as determinants of human twinning rate: An analysis of Dutch population data. Paper read at the Fifth International Congress on Twin Studies, Amsterdam 15-19 Sept. 1986.

Correspondence: Dr. Johan Fellman, Swedish School of Economics and Business Administration, Arkadiagatan 22, 00100 Helsinki, Finland.