A CRITIQUE OF CLOGG'S METHOD OF RATES ADJUSTMENT

M. RAFIQ SHAH

Vrije Universiteit, Brussels

IPD-Working Paper 1987-4

Acknowledgements:
I gratefully acknowledge the financial support of my research by the Belgian Administration for Overseas Development (ABOS-AGCD). I am grateful to Frans Willekens, Clifford C. Clogg, T.W. Pullum, G. Wunsch and S. Wijewickrema for their comments on an earlier version of this paper.

ORDERING INFORMATION

Extra copies of this working paper are available from :

Mr. E. Vanden Balck
Secretary/Librarian
Interuniversity Programme in Demography
c/o Centrum voor Sociologie
Vrije Universiteit Brussel
Pleinlaan 2
B-1050 Brussels
BELGIUM

at the cost of US $ 5.50 or BF 200. Please attach check or international money order.

ORDER FORM

Name ....
Address ....
....
....
....
....

wishes to receive ...... copies of Working Paper Nr..........
and includes payment of US $ .... or BF........

A CRITIQUE OF CLOGG'S METHOD OF RATES ADJUSTEMENT *

M. Rafiq Shah, Ph.D.

Interuniversity Programme in Demography
Vrije Universiteit Brussel

Abstract

Clogg (1978) has developed a method for the adjustement of demographic rates by purging the so called undesirable interactions from a saturated multiplicative model. The summary rates of the purged frequencies are known as the adjusted rates. This paper critically examines the proposed method of finding adjusted rates by purging two factor interactions, and relates it to the traditional method of direct standardization by defining the prevalence rates in terms of the parameters of the model. A comparison of the results reveals that unlike directly standardized rates, the adjusted rates carry with them the effects of the confounding factor. The case of finding adjusted rates based on purging both two and three factors interactions is also discussed. The examples quoted by Clogg (1978) are used as numerical illustrations.

* Prepared for the annual meeting of the Population Association Of America, to be held in Chicago during April 30 - May 2, 1987.

Introduction

A simple measure of summarizing the experience of a demographic process is the crude rate. This rate is, however, not suitable for the purpose of making a comparison of any given process due to the built-in defect of such a measure when seen as a weighted average; the relevant weights being the proportions of an associated variable or factor. It is precisely for this reason that the method of direct standardization was introduced as early as 1883 (Wolfenden, 1954), so that comparisons could be made on the basis of standardized rates.

According to the method of direct standardization, the weights proper to each population compared (for example, the proportions in various age groups of the population under comparison) are replaced by a common set of weights (the "standard age composition"). The expected summary rates caculated by applying the "standard" are known as directly standardized rates.

Even though these rates have the advantage of being easy to calculate, it is the selection of the "standard" that is found to be really problematic. In fact, one cannot find a "unique" standard for the calculation of directly standardized rates. Several methods have been suggested for finding summary indices either by using an endogenous standard or by using no standard at all.

In all these attempts, the use of statistical models have played an important role. In particular, the models that are commonly used for the analysis of contingency tables (e.g., multiplicative or log-linear models) have provided an alternative to the traditional techniques of standardization. Besides other advantages, each of these models provides unique estimates of the required summary indices without the intermediary use of a standard.

This paper deals with the method of adjusting summary rates, by postulating a saturated multiplicative model on the available data, as proposed by Clogg (1978). Since a saturated model fully describes the data (e.g. contains all possible main and interaction effects), the interaction that brings about most of the confounding effects on the summary measures (the undesirable interaction) is first identified. This interaction is then purged out of the data. The summary rates based on purged and rescaled data are known as adjusted rates. The method has been applied in labor force analysis (Clogg, 1979; Clogg and Shockey, 1985).

As the method of purging interaction according to Clogg's method and that of direct standardization both deal with the procedures of removing or controlling the confounding effects of the summary rates, this paper critically examines the properties of adjusted rates in comparison to directly standardized rates.

To begin with, the type of a contingency table that could be used for the calculation of prevalence rates and the rates of non-renewable events is described in terms of a general-purpose terminology. The model that characterizes various effects of the factors involved will follow immediately.

Terminology

Groups, populations or periods of observations in which a demographic process is required to be compared will be called the "state factor" or the "states". The demographic process in question will be called the "outcome factor" or the "outcome". The effects of the states will be assumed to depend on some background factor(s) influencing a demographic process under study. A background factor will be called a "confounding factor" if, besides its influence on the outcome, it is distributed differently in the various states. For example, in the problems of standardization and in most of the discussions in this paper, confounding factor, outcome and state factor represent age composition, employment status and groups or years of observations respectively.

Summary Rates in a Contingency Table and the Model

Assume that the outcome D is polytomous having K categories k ($k=1,...K$) and is classified by the confounding factor A, having R levels x ($x=1,...R$) and the state factor B having C categories i ($i=1,...C$). Denote the number of events in the i-th state, the k-th category of D and the x-th level of A by $E_{ik}(x)$. The presentation of such a table is found to be consistent with the description of prevalence rates and the rates of non-renewable events, some of which could be seen in numerical illustrations. A detailed account of the analysis of such tables may be found in Bishop et. al (1975 31-41), Goodman (1978, Chapter 4), Haberman (1978, Vol 1, Chapter 3) and Willekens (1983).

Note that events give rise to occurence /exposure rates and the presence in any specified category are used for the purpose of prevalence rates. As most of the discussion in this paper relates to the prevalence rates, events simply mean the presence of individuals in the specified categories of the factors involved. Accordingly, the proportion of events occuring in the i-th state that fall in the k-th class of D and the x-th level of A is the xi-specific rate of the k-th class of D:

$$r_{ik}(x) = E_{ik}(x)/\Sigma_k E_{ik}(x) = E_{ik}(x)/E_{i+}(x) \qquad (1)$$

The crude rate of the k-th class of D in the i-th state is.

$$r_{ik}(.) = \Sigma_x E_{ik}(x)/\Sigma_{kx} E_{ik}(x) = E_{ik}(+)/E_{i+}(+) \qquad (2)$$

The crude rate (2) can be expressed as a weighted average of the xi-specific rate $r_{ik}(x)$, the weights being the proportions of the confounding factor A:

$$r_{ik}(.) = \Sigma_x r_{ik}(x).V_i^A(x) \qquad (3)$$

where $V_i^A(x) = E_{i+}(x)/E_{i+}(+)$.

The model that describes the data fully in a contingency table is the saturated model:

$$E_{ik}(x) = w.w_x^A.w_i^B.w_k^D.w_{xi}^{AB}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}. \qquad (4)$$

The usual constraints on the parameters are:

$$\Pi_x w_x^A = \Pi_i w_i^B = \Pi_k w_k^D = \Pi_x w_{xi}^{AB} = \ldots = \Pi_x w_{xik}^{ABD} = \ldots = 1$$

In a saturated model, since there are as many independent parameters in the equation as there are types of different effects describing the data, the estimated value for $E_{ik}(x)$ is equal to the observed value. The parameters of the model are: the overall effect (w), main effects (w_x^A, w_i^B, w_k^D), first order interaction effects (w_{xi}^{AB}, w_{xk}^{AD}, w_{ik}^{BD}) and the second order interaction effect (w_{xik}^{ABD}).

Of interest here in the model (4) is the interpretation of the first order or two factor interaction effects. For instance, w_{xi}^{AB} representes state differences in the distribution of the confounding factor A, the parameter w_{xk}^{AD} measures the average effect of the confounding factor A on D, and w_{ik}^{BD} measures the state diffences in D.

Note that the additive form of the multiplicative model (4) is obtained by taking logarithms.

$$\log E_{ik}(x) = U + U_x^A + U_i^B + U_k^D + U_{xi}^{AB} + U_{xk}^{AD} + U_k^{BD} + U_{xik}^{ABD} \qquad (5)$$

with $U = \log(w)$, etc.

The usual restrictions of the parameters of the log-linear model (5) are:

$$\Sigma_x U_x^A = \Sigma_i U_i^B = \Sigma_k U_k^D = \Sigma_x U_{xi}^{AB} = \ldots = \Sigma_x U_{xik}^{ABD} = \ldots = 0$$

An immediate advantage of these models is that one can easily express the xi-specific rates $r_{ik}(x)$, the crude rates $r_{ik}(.)$ and the weights $V_i^A(x)$ in terms of the parameters of the model. For instance, the rate $r_{ik}(x)$, defined in (1) may be written as follows:

$$r_{ik}(x) = \frac{w.w_x^A.w_i^B.w_k^D.w_{xi}^{AB}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}}{\Sigma_k\, w.w_x^A.w_i^B.w_k^D.w_{xi}^{AB}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}}$$

$$= \frac{w_k^{D}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}}{\Sigma_k\, w_k^{D}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}}\,. \qquad (6)$$

The crude rate (2) is:

$$r_{ik}(\,) = \frac{w_k^{D}.\, w_{ik}^{BD}.\Sigma_x w_x^{A}.w_{xi}^{AB}.w_{xk}^{AD}.w_{xik}^{ABD}}{\Sigma_k\, w_k^{D}.\, w_{ik}^{BD}.\Sigma_x w_x^{A}.w_{xi}^{AB}.w_{xk}^{AD}.w_{xik}^{ABD}} \qquad (7)$$

and the weights $V_i^A(x)$ as in (3) are:

$$V_i^A(x) = \frac{w_x^{A}.w_{xi}^{AB}.\ \Sigma_k w_k^{D}.\ w_{ik}^{BD}.w_{xk}^{AD}.w_{xik}^{ABD}}{\Sigma_x w_x^{A}.w_{xi}^{AB}.\ \Sigma_k w_k^{D}.\ w_{ik}^{BD}.w_{xk}^{AD}.w_{xik}^{ABD}}\,. \qquad (8)$$

Clogg's Method of Adjustement

Due to the fact that the two factor interaction w_{xi}^{AB} in the multiplicative model (4) represents state (factor B) differences in the distribution of the confounding factor A (e.g. differences in the age compositions), Clogg (1978) proposed the adjustement of summary rates by purging this interaction from the saturated model (4) of counts. The summary rates based on purged frequencies are known as the "adjusted rates". As the adjusted rates do not possess the effects of the confounding factor, their differences are thought to show the "true" differences of the outcome in question. Clogg's method of adjustement based on purging and rescaling is explained in the following subsections.

1. Purging of Undesired Interaction

The interaction w_{xi}^{AB} which is thought to confound or obscure summary rates is removed from an assumed saturated model (4) of counts by the simple process of division. Consequently, frequencies expressed without this interaction, the so called "purged frequencies", are:

$$E_{ik}^{*}(x) = E_{ik}(x)/w_{xi}^{AB} = w.w_x^{A}.w_i^{B}.w_k^{D}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}. \quad (9)$$

Since the original frequencies change due to the process of division, they could be rescaled in order to sum to the observed total frequencies in each state.

2. Rescaling

As the sum $E_{++}^{*}(+)$ of the purged frequencies $E_{ik}^{*}(x)$ will not be equal to the corresponding sum $E_{++}(+)$ of the observed frequencies $E_{ik}(x)$, each $E_{ik}^{*}(x)$ must be rescaled in order for the sum of the purged frequencies to be equal to that of the observed frequencies. This is done by multiplying each $E_{ik}^{*}(x)$ by the ratio of the total observed frequency to the total purged frequency, i.e. by the ratio $E_{++}(+)/E_{++}^{*}(+)$. Denoting the rescaled frequencies so obtained by $E_{ik}^{**}(x)$, we have

$$E_{ik}^{**}(x) = w'.w_x^{A}.w_i^{B}.w_k^{D}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}. \quad (10)$$

where $w'=w.(E_{++}(+)/E_{++}^{*}(+))$.

To preserve the total of the observed frequencies in each state, one has to multiply the purged frequencies $E_{ik}^{*}(x)$ by the ratio $E_{i+}(+)/E_{i+}^{*}(+)$. The rescaled frequencies $E_{ik}^{\wedge}(x)$ so obtained are:

$$E_{ik}^{\wedge}(x) = E_{ik}^{*}(x)[E_{i+}(+)/E_{i+}^{*}(+)]$$

$$= w'.w_x^{A}.(w_i^{B})'.w_k^{D}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}, \quad (11)$$

where

$$(w_i^B)' = w/w'.[\Pi_{xk}(E_{ik}^{\wedge}(x)/E_{ik}^{*}(x))]^{1/(RK)}$$

$$= [E_{++}^{*}(+)/E_{++}(+)].[E_{i+}(+)/E_{i+}^{*}(+)].w_i^B.$$

3. Adjustement

The adjusted rate $r_{ik}^{s}()$ of the k-th class of D in the i-th state is obtained by replacing the observed frequencies $E_{ik}(x)$ by the purged frequencies $E_{ik}^{*}(x)$ or by the purged and rescaled frequencies $E_{ik}^{**}(x)$. Using the purged and rescaled frequencies $E_{ik}^{''}(x)$, the adjusted rate according to Clogg's method is:

$$r_{ik}^{s}() = \Sigma_x E_{ik}^{''}(x)/\Sigma_{xk} E_{ik}^{''}(x) \qquad (12)$$

$$= E_{ik}^{\wedge}(+)/E_{i+}^{''}(+) = E_{ik}^{*}(+)/E_{i+}^{*}(+)$$

Comments and Comparisons

The method of adjustement presented by Clogg (1978) displays a useful application of the saturated models in demographic analysis and is considered to be a breaktrough in the methodology of standardization. The method provides a basis for the replacement of the traditional method of components analysis (Kitagawa, 1955) in case of several interacting factors. Clogg's method is flexible and has the capacity of accomodating a number of factors and states simultaneously. Clogg (1982) has also made available a computer program (PURGE) which may be used for the comparison of several states classified by several factors where each factor could have many categories, and where purging of higher order interaction is felt necessary.

Keeping in view, however, an important criterion of a standardized rate, namely that a standardized rate should be independent of the compositions of the states under comparison, the adjusted rates according to Clogg's method are found to be lacking this property. The arguments in favour of illustrating the lack of such an important criterion could be presented as follows.

The lack of this criterion arises due to the process of purging undesirable interaction (w_{xi}^{AB}) only, from the saturated model (4). Note that purging of other interactions gives entirely different results (Shah, 1986)

According to the saturated model (4), the xi-specific rates $r_{ik}(x)$ are independent of the AB interaction as shown in (6). The observed rates do therefore not change during the process of purging the AB interaction. Since the $r_{ik}(x)$ do not change, one may compare the adjusted rate $r_{ik}^{s}(.)$ with the crude rate $r_{ik}(.)$ as follows.

A comparison of the crude rate (7) with the adjusted rate (12) expressed in terms of the parameters of the model (4) indicated that the difference between the two is entirely due to the absence of the AB interaction term in the adjusted rate.

Due to the fact that the $r_{ik}(x)$ do not change, the adjusted rate $r_{ik}^{s}(.)$ could be expressed as a weighted average of the xi-specific rates $r_{ik}(x)$ as follows.

$$r_{ik}^{s}(.) = \Sigma_x r_{ik}(x).V_i(x) \qquad (13)$$

where

$$V_i(x) = E_{i+}'(x)/E_{i+}'(+).$$

The weights $V_i(x)$ are the state factor specific proportions of the confounding factor, after having removed the AB interaction. The method thus produces a "purged confounding factor" only. According to (13), these weights depend on i and are not identical in all states under comparison. It follows that the adjusted rates based on Clogg's method include the effects of the confounding factor. The confounding factor is thus neither controlled nor eliminated.

Since the xi-specific rates $r_{ik}(x)$ are not affected by the method of adjustement as stated earlier, and the method of adjustement produces changes in the confounding factors, i.e. the weights $V_i^A(x)$, the adjusted rate being a summary measure could be compared with a directly standardized rate.

Given a set of standard weights $V^s(x)$ that is independent of the states under comparison, the directly standardized rate of the k-th class of D in the i-th state i is defined as

$$DSR_{ik} = \Sigma_x r_{ik}(x).V^s(x) \qquad (14)$$

where

$V^s(x) = E_{i+}{}^s(x)/E_{i+}{}^s(+)$, such that $\Sigma_x V^s(x) = 1$.

A comparison of (13) and (14) reveals that whereas $r_{ik}(x)$ is common in both $r_{ik}{}^s(.)$ and DSR_{ik}, the weights $V_i^A(x)$ and $V^s(x)$ are different. Common to these weights, however, is that both are independent of the AB interaction. Therefore, the difference between $r_{ik}{}^s(.)$ and DSR_{ik} cannot be attributed to the AB interaction, but to the difference in magnitude arising in $V_i^A(x)$ and $V^s(x)$.

It may be noted that while $V_i^A(x)$ depends on AB, BD and ABD interactions, $V^s(x)$ is independent of all types of interactions as it remains constant over all the states under comparison. This finding is by no means related to the proof that the direct method of standardization is better than the method based on a multiplicative model. It may be used merely for the identification of an important property of a standardized rate, namely that the rate is independent of the compositions of the states under comparison. This property is not fulfilled by the adjusted rate (after purging two factor interaction from a saturated model) as proposed by Clogg (1978).

We shall demonstrate through numerical illustrations that the bias arising due to the dependence of weights $V_i^A(x)$ on i could lead to different (misleading) inferences and conclusions. Out of several examples the ones quoted by Clogg (1978) are chosen to be presented here for ready reference and comparison with Clogg's results.

Numerical Results

1. Hypothetical Data

Table 1 shows the hypothetical data classified by confounding factor (composition), state factor (groups) and dichotomous outcome factor. The composition-specific rates of states 1 and 2 are the same but the crude rates differ because of compositional differences. On the other hand, the composition of states 1 and 3 is the same but the composition-specific rates are different. The purpose is to compare the states in order to identify differences in the prevalence rates of the outcome factor. A saturated log-linear model is fitted to the data of Table 1 using GLIM (Generalized Linear Interactive Modelling; Baker and Nelder, 1978). The parameters of the model are shown in Appendix 1. The computer listing and program are laid out in Appendix 2.

Table 1. Hypothetical data. frequencies $E_{ik}(x)$ and rates $r_{ik}(x)$ by state.

	$E_{ik}(x)$									$r_{ik}(x)$		
State factor	1			2			3			1	2	3
Outcome factor/ Confounding factor	1	2	Total	1	2	Total	1	2	Total			
1	25	25	50	50	50	100	25	25	50	.50	.50	.50
2	10	40	50	15	60	75	25	25	50	.20	.20	.50
3	8	92	100	2	23	25	50	50	100	.08	.08	.50
Total	43	157	200	67	133	200	100	100	200			
Crude rates										.215	.335	.500

For k=1 the rate $r_{i1}(x)= E_{i1}(x)/E_{i+}(x)$, e.g. in state 1, the rate at the 1st level of the confounding factor is $r_{11}(1)=25/50=0.5$.

Source: Clogg (1978), Table 1: 528.

The expected frequencies $E_{ik}^*(x)$ obtained after purging the AB interactions as proposed by Clogg (1978) are given in Table 2.a. Calculations based on these frequencies indicate that the xi-specific rates $r_{ik}^*(x)$ of the k-th category of D are equal to the observed rates $r_{ik}(x)$ as shown in Table 1. For instance, $r_{11}^*(1)$ based on Table 2.a is $r_{11}^*(1)=E_{11}^*(1)/E_{1+}^*(1)$ =29.656/59.312=0.50, and the same rate based on the data of Table 1 is $r_{11}(1)=E_{11}(1)/E_{1+}(1)$=25/50=0.50. These calculations correspond to the observations that the xi-specific rates of any category of D are independent of the AB interactions (see equation (6)).

Table 2.a. Frequencies $E_{ik}^*(x)$ purged of AB interactions

State factor	1		2		3	
Outcome factor/ Confounding factor	1	2	1	2	1	2
1	29.6560	29.6560	26.9251	26.9251	39.1236	39.1236
2	11.6009	46.4037	10.5411	42.1644	30.6373	30.6373
3	5.8140	66.8605	5.2770	60.6860	26.0417	26.0417
Total	47.0709		42.7432		95.8026	

$E_{ik}^*(x) = E_{ik}(x)/w_{xi}^{AB}$; e.g. $E_{11}^*(1) = 29.6560 = 25/\exp(-.1702)$

Source: Table 1 and Appendix 1.

Notice the variation in the distribution of the confounding factor over the states as shown in Table 2.b. Note that these proportions are used as weights for the calculation of adjusted rates by equation (13). Since these weights depend on the states under comparison, the adjusted rates obtained after purging AB interactions only do not satisfy the property that a standardized index be independent of the compositional distribution of the states under comparison. We shall see later that the absence of this property of the adjusted rate could give different (misleading) results. The proportions $V_i^A(x)$ vary according to the pattern of the xi-specific rates. The proportions are identical only in the states where the xi-specific rates are identical. For instance, since the rates $r_{ik}(x)$ in states 1 and 2 are identical, the weights $V_i^A(x)$ associated with states 1 and 2 are also identical.

Table 2.b. Frequencies $E_{i+}^*(x)$ and proportions $V_i(x)$ purged of AB interactions by state.

State factor	1		2		3	
Confounding factor	$E_{1+}^*(x)$	$V_1(x)$	$E_{2+}^*(x)$	$V_2(x)$	$E_{3+}^*(x)$	$V_3(x)$
1	59.3120	0.3122	53.8502	0.3121	78.2472	0.4084
2	58.0046	0.3053	52.7055	0.3055	61.2746	0.3198
3	72.6745	0.3825	65.9630	0.3824	52.0834	0.2718
Total	189.9911	1.0000	172.5187	1.0000	191.6052	1.0000

Source: Table 2.a.

Since the adjusted rates (13) are not really standardized in a conventional sense, as we have noted above, they cannot be used for decomposing the difference of rate and composition components as proposed by Kitagawa (1955). Except in situations where the xi-specific rates are identical, the estimates of "rate" and "composition" components of the difference of crude rates is biased. For example, the difference between the crude rate of state 1 and that of state 3 is 0.335-0.500=-0.165 or -16.50%. The corresponding difference in adjusted rates is 0.248-0.500=-0.252 or -25.20%. The difference between these quantities (8.7%, -16.50+25.20) is an estimate of the effect of the confounding factor. This estimate is, however, based on the assumption that the set of weights is common to both states (as is normally the case in direct standardization).

2. Empirical Data

The hypothetical data in Table 1 are constructed such that the xi-specific rates are chosen to be identical in the categories of the confounding factor in state 1 and state 2 whereas in state 3 these rates are identical. Such data conceal in part the drawbacks of Clogg's adjustement method. Therefore we consider another data set. Table 3 shows the U.S. civilian labour force data classified by age and year of reporting. The objective is to see if age composition has played an important role in the process of unemployment over the reported years.

Table 3. U.S. Civilian labor force classified by age and year of reporting ($E_{ik}(x)$ with i=year, k=employment status, x=age)

	1969			1971			1973		
Age	unempl.	empl.	total	unempl.	empl.	total	unempl.	empl.	total
14-19	668	5,122	5,790	986	4,949	5,935	907	5,247	6,154
20-34	784	18,581	19,365	1,522	18,011	19,533	1,223	19,103	20,326
35-49	413	19,155	19,568	751	17,672	18,423	547	16,112	16,659
50-64	310	14,250	14,560	534	13,607	14,141	399	12,421	12,820
65+	58	2,508	2,566	81	2,177	2,198	45	1,949	1,994
Total	2,233	59,616	61,849	3,874	56,356	60,230	3,121	54,832	57,953
Crude unemployment rate (percent)	3.61			6.43			5.39		

Source: Clogg, 1978, Table 5: 536 (Data from March Current Population Survey)

As before a saturated log-linear model is fitted on the data of Table 3 by using GLIM. Computer output and the parameters of the model are displayed in Appendix 4. The parameters required for the purpose of purging age-time interaction are shown in Appendix 5. Following our critical remarks on the uneven distribution of weights $V_i(x)$, when the adjusted rates $r_{ik}^{s}(.)$ are expressed as a weighted average of the xi-specific rates $r_{ik}(x)$, attention is focussed on the purged distribution of the state factor $E_{i+}{}^{*}(x)$.

Sets of purged distributions $E_{i+}{}^{*}(x)$ are obtained by dividing the observed distribution of the background factor $E_{i+}(x)$ by the appropriate interaction terms (see Appendix 5, Table 5). Note that the interaction term is common in both categories of the outcome factor D in a specified year and age group. Purged totals for each i and x, $E_{i+}{}^{*}(x)$, and purged weights $V_{i}(x)$ are shown in Table 4. The "purged weights" are not identical in all the years under study, implying that the background factor is still a confounding factor and the adjusted rates $r_{ik}^{s}(.)$ based on these weights carry with them confounding effects.

Table 4. U.S. Civilian labor force. Purged counts $E_{i+}{}^{*}(x)$ and proportions $V_{i}(x)$ by age and years.

States	1969	1971	1973	1969	1971	1973
Confounding factor Age	Frequencies $E_{i+}{}^{*}(x)$			Proportions $V_{i}(x)$		
14-19	5934.2011	6325.9433	5632.9519	0.0926	0.1059	0.1000
20-34	21284.8980	19437.7550	18582.9200	0.3322	0.3254	0.3299
35-49	19857.9260	17926.4380	16869.8730	0.3099	0.3001	0.2995
50-64	14775.7260	13866.4440	12910.3730	0.2306	0.2322	0.2292
65+	2216.4637	2171.7222	2336.5362	0.0346	0.0364	0.0415
All ages	64069.2158	59728.3035	56332.6541	0.9999	1.0000	1.0001

Source: Table 3 and Appendix 5.b., e.g. the totals in each year are obtained by first dividing the number of employed and unemployed by the AB interaction of that year. For the required level of the confounding factor as shown in Appendix 5.b, see text.

The confounding effect (bias) may be estimated by the components method. For this purpose we apply first Kitagawa's method to "adjusted rates" and next we compare the inference based on "adjusted rates" with those based on directly standardized rates. Consider the unemployment rates $r_{i1}(x)$ based on the calculations from Table 3 and the weights $V_i(x)$ from Table 4. The adjusted rates for the years 1969, 1971 and 1973 are 0.0364, 0.0653 and 0.0525 respectively. The corresponding crude unemployment rates for the same years as noted in the bottom row of Table 3 are 0.0361, 0.0643 and 0.0539 respectively.

Before commenting on the inferences based on adjusted rates for the estimation of the compositional effect or the confounding bias, it is useful to recapitulate the main points about the components of the difference of two crude rates. According to Kitagawa's suggestion, the difference in crude rates of two populations (states) is composed of a "rate effect" and a "compositional effect". The part due to rates or rate effect is estimated from the difference of directly standardized rates (population composition is common in both populations). The compositional effect is then the difference of the crude rates minus the difference of the standardized rates.

Using adjusted rates instead of directly standardized rates we find that the difference of the crude rates of 1969 and 1971 is -0.0282 = 0.0361-0.0643. The corresponding difference in the adjusted rates is -0.0289 = 0.0364-0.0653. The difference between these quantities is -0.0007 = -0.0282+0.0289 or 0.07%.

Any inference based on this figure (0.07%) about the compositional effect cannot be correct, since the estimate of the "rate component" is based on "purged crude rates" (adjusted rates) and the weights $V_i(x)$ are not independent of the state factor unlike the weights commonly used in direct standardization. Moreover, any conclusion regarding the role of AB interaction in the increase of unemployment from 1969 to 1971 is uncalled for, as far as the components of the difference of the crude rates are concerned. This is due to the rate effect which is confounded by the differences of the AB interactions in 1969 and 1971, and to the presence of other interaction effects. Note That Clogg's (1978. p.537) inference about the role of age-time-period interaction is based on this figure (0.07%).

Due to reasons noted earlier and considering Kitagawa's procedure of decomposition in case of one factor as both logical and less complicated (compared to 2 or more factor cases), it is possible to estimate the "rate component" without bias by using directly standardized rates of the states (populations). Using the observed proportions of the populations in the years under study (Table 5) as standards, we calculated directly standardized rates for 1969, 1971 and 1973. Crude as well as standardized rates are shown in matrix M (Appendix 6). Since the standardized rates obtained by using these standards differ from those based on adjusted rates we used the average composition of 1969 and 1971 as a standard (Table 5). Directly standardized rates for the years 1969 and 1971, based on this standard are 0.0365 and 0.0638 respectively, giving a rate component of -0.0273 = 0.0365-0.0638. Since the difference of crude rates in 1969 and 1971 is -0.0282 the composition component is estimated as -0.0009 = -0.0282+0.0273 or -0.09%. The use of directly standardized rates of 1969 and 1971 results in a negative effect (-0.09%) of the population structure in contrast to the one (0.07%) based on "adjusted rates". The negative effect could be interpreted as a decrease in the prevalence of unemployment due to the compositional change that occured from 1969 to 1971. Note that due to problems of weights in the adjusted rates, the conclusion based on conventional directly standardized rates seems to be correct.

Our experiments with several other sets of data suggest that the adjusted rates based on purging two factor (AB) interactions gives results different from those based on directly standardized rates. We have, therefore, tried to purge out other interactions besides the AB interaction in order to solve the problem of estimating identical weights for all the states in question. First we tried to purge out the three factor (ABD) interaction from the saturated model, the results of which are discussed in the following section.

Table 5. Observed proportional distribution $V_i^A(x)$ of the population by age and years of observations and standard $V^s(x)$

	Observed proportions $V_i^A(x)$			Standard $V^s(x)$ [a]
		Years		
Age	1969	1971	1973	
14-19	.0936	.0985	.1062	.0961
20-34	.3131	.3243	.3507	.3187
35-49	.3164	.3059	.2857	.3112
50-64	.2354	.2348	.2212	.2351
65+	.0145	.0365	.0344	.0390
All ages	1.0000	1.0000	1.0000	1.0001

[a] Average age composition of 1969 and 1971.
Source. Table 3.

The Need of Purging Higher Order Interactions

The importance of purging higher order terms is felt necessary since purging two factor (AB) interactions yields biased results. Because of the non-hierarchical nature of the log-linear model of purged frequencies $E_{ik}^{*}(x)$, we also purged out the three factor interaction w_{xik}^{ABD}. Dividing $E_{ik}^{*}(x)$ by w_{xik}^{ABD} and denoting the purged frequencies so obtained by $E_{ik}^{**}(x)$, we have

$$E_{ik}^{**}(x) = w.w_x^A.w_1^B.w_k^D.w_{xk}^{AD}.w_{ik}^{BD} \qquad (15)$$

Using (15), the weights as shown in (13) without AB and ABD interaction terms are:

$$V_i^*(x) = \frac{w_x^A.\Sigma_k w_k^D.\ w_{ik}^{BD}.w_{xk}^{AD}}{\Sigma_x\ w_x^A.\Sigma_k w_k^D.\ w_{ik}^{BD}.w_{xk}^{AD}}. \qquad (16)$$

Note that since the xi-specific rates depend on ABD interactions, they will change (smooth out) unlike the ones obtained by purging AB interaction only. Using (15) and denoting the smoothed xi-specific rates purged of both AB and ABD interaction by $r_{ik}^*(x)$.

$$r_{ik}^*(x) = \frac{w_k^D\ w_{ik}^{BD}.w_{xk}^{AD}}{\Sigma_k w_k^D.\ w_{ik}^{BD}.w_{xk}^{AD}} = E_{ik}^{**}(x)/E_{i+}^{**}(x)\ . \qquad (17)$$

Using (17) and (16), the adjusted rates based on purging both AB and ABD interactions are:

$$r_{ik}^{s*}(.) = \Sigma_x r_{ik}^*(x).V_i^*(x) \qquad (18)$$

The adjusted rate defined by (18) is not comparable to a directly standardized rate (as we have been comparing in case of purging AB interactions only, i.e. when the rates $r_{ik}(x)$ did not change). An important point to note is that the weights used in the "smoothed" rates (18) still depend on factor x and state i. In other words, factor A is still a confounding factor.

The extent of such dependence depends on the question at hand. A series of exercises on hypothetical as well as emperical sets of data suggest that, whereas the weights $V_i^*(x)$ could practically be assumed to remain constant over the states under comparison, there are others where the difference in weights so obtained is considerably large. For instance, the results of Table 6 obtained after purging both AB and ABD interactions support the finding that the adjusted rates are not free from the effects of the confounding factor when these rates are considered as weighted averages and expressed in terms of the parameters of the proposed model.

Table 6. Frequencies $E_{i+}^{**}(x)$ and proportions $V_i^*(x)$ purged of both AB and ABD interactions.

State factor	1		2		3	
Confounding factor x	$E_{1+}^{**}(x)$	$V_1^*(x)$	$E_{2+}^{**}(x)$	$V_2^*(x)$	$E_{3+}^{**}(x)$	$V_3^*(x)$
1	67.31	0.33655	76.31	0.33655	84.26	0.42130
2	63.72	0.31860	63.72	0.31860	60.47	0.30235
3	68.97	0.34485	68.97	0.34485	55.27	0.27635
All categories	200.00	1.00000	200.00	1.00000	200.00	1.00000

Source: Table 2.a. and Appendix 1.

References

Baker, R.J. and J.A. Nelder (1978). The GLIM System Manual. Release 3. Oxford : The Numerical Algorithms Group, Royal Statistical Society, London.

Bishop, Y.M.M.; S.E. Fienberg and P.W. Holland (1975). Discrete Multivariate Analysis : Theory and Practice. M.I.T. Press.

Clogg, C.C. (1978). Adjustement of Rates Using Multiplicative Models. Demography, 15. 523-539.

---------- (1979). Measuring Unemployment. Demographic Indicators for the U.S. Labor Force. New York, Academic Press.

---------- (1982). Computer Program PURGE. Personal communication.

---------- and J.W. Shockey(1985). The Effect of Changing Demographic Position on Recent Trend in Underemployement. Demography 22. 395-414

Goodman, L.A. (1978). Analyzing Qualitative Categorical data. Log-linear Models and Latent Structures. Ed. J. Magidson, Cambridge Mass: Associates.

Haberman, S.J. (1978). Analysis of Qualitative Data. Vol. 1: Introductory Topics and Vol. 2: New Developments. New York, Academic Press.

Kitagawa, E.M. (1955). Components of a Difference between Two Rates. Journal of the Americal Statistical Association 50: 1165-1194.

Shah, M. Rafiq (1986). Standardization in Demographic Analysis. Ph. D. Thesis. Interuniversity Programme in Demography, Vrije Universiteit Brussels, Belgium.

Willekens, F. (1983). Specification and Calibration of Spatial Interaction Models: a Contingency Table Perspective and an Application to Intra-urban Migration in Rotterdam. Tijdschrift voor Economische Geografie 74: 239-252.

Wolfenden, H. H. (1954). Population Statistics and their Compilation. University of Chicago Press.

Appendix 1

Parameters of the saturated log-linear model for the data of Table 1.

$$\log E_{ik}(x) = U + U_x^{A} + U_i^{B} + U_k^{D} + U_{xi}^{AB} + U_{xk}^{AD} + U_k^{BD} + U_{xik}^{ABD}$$

Overall Mean: U = 3.233

Main Effects

U_1^A = .2174	U_1^B = -.0607	U_1^C = 4254
U_2^A = -.0272	U_2^B = -.1567	U_2^C = -.4254
U_3^A = -.1902	U_3^B = .2174	

Two Factor Interactions

AB			AD		BD	
-.1702	.6187	-.4485	.4254	-.4254	-.2127	.2127
-.1488	.3526	-.2038	-.0367	.0367	-.2127	.2127
.3190	-.9713	.6523	-.3887	.3887	.4254	-.4254

Three Factor Interactions (ABD)

.2126	-.2126	.2128	-.2128	-.4254	.4254
-.0183	.0183	-.0184	.0184	.0367	-.0367
-.1943	.1943	-.1944	.1944	.3887	-.3887

Source: Appendix 2.

Appendix 2

GLIM program for fitting the saturated model on the data of Table 1 and the estimated parameters under the usual constraints.

```
$UNITS  18
$DATA Y
$FACTOR  A 3  B 3  D 2
$CALC  A=%GL(3,6) :  B=%GL(3,2) : D=%GL(2,1)
$DINPUT 1
$LOOK  Y
$YVAR  Y
$ERROR  P
$CALC   A2=%EQ(A,1)-%EQ(A,2)
      : A3=%EQ(A,1)-%EQ(A,3)
      : B2=%EQ(B,1)-%EQ(B,2)
      : B3=%EQ(B,1)-%EQ(B,3)
      : D2=%EQ(D,1)-%EQ(D,2)
$CALC   P1=A2*B2 :  P2=A2*B3  :  P3=A3*B2 :  P4=A3*B3
      : P5=A2*D2 :  P6=A3*D2  :  P7=B2*D2 :  P8=B3*D2
$CALC  Q1=A2*B2*D2 :  Q2=A2*B3*D2 :  Q3=A3*B2*D2 :  Q4=A3*B3*D2
$FIT  A2+A3+B2+B3+D2+P1+P2+P3+P4+P5+P6+P7+P8+Q1+Q2+Q3+Q4
$DISPLAY  MERT
$STOP

     25   25   50   50   25   25
     10   40   15   60   25   25
      8   92    2   23   50   50

GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

 CYCLE  DEVIANCE      DF
   3     .4394E-11     0

Y-VARIATE  Y
ERROR POISSON  LINK LOG

LINEAR PREDICTOR
%GM A2 A3 B2 B3 D2 P1 P2 P3 P4 P5 P6 P7 P8 Q1 Q2 Q3 Q4

       ESTIMATE        S.E.        PARAMETER
   1    3.233        .6105E-01     %GM
   2    .2724E-01    .7998E-01     A2
   3    .1902        .1019         A3
   4    .1567        .9950E-01     B2
   5   -.2174        .7469E-01     B3
   6   -.4254        .6105E-01     D2
   7    .3526        .1220         P1
   8   -.2038        .1023         P2
   9   -.9713        .1779         P3
  10    .6523        .1155         P4
  11    .3669E-01    .7998E-01     P5
  12    .3887        .1019         P6
  13    .2127        .9950E-01     P7
  14   -.4254        .7469E-01     P8
  15   -.1835E-01    .1220         Q1
  16    .3669E-01    .1023         Q2
  17   -.1944        .1779         Q3
  18    .3887        .1155         Q4
 SCALE PARAMETER TAKEN AS     1.000

 UNIT  OBSERVED    FITTED     RESIDUAL
   1      25        25.00      .1990E-11
   2      25        25.00      .1918E-11
   3      50        50.00      .2110E-11
   4      50        50.00      .2713E-11
   5      25        25.00      .1421E-11
   6      25        25.00      .1705E-11
   7      10        10.00      .4494E-13
   8      40        40.00      .1887E-11
   9      15        15.00      .9907E-12
  10      60        60.00      .2202E-11
  11      25        25.00      .1137E-11
  12      25        25.00      .1279E-11
  13       8        8.000     -.3135E-11
  14      92        92.00      .3271E-11
  15       2        2.000     -.4994E-07
  16      23        23.00      .8860E-12
  17      50        50.00      .1306E-11
  18      50        50.00      .1708E-11
```

Appendix 3

GLIM program and the parameters of the model after purging the two factor (AB) interaction.

```
$UNITS  18
$DATA   Y
$FACTOR  A 3  B 3  D 2
$CALC  A=%GL(3,6)  :   B=%GL(3,2)  :   D=%GL(2,1)
$DINPUT 1
$LOOK  Y
$YVAR  Y
$ERROR  P
$CALC    A2=%EQ(A,1)-%EQ(A,2)
       : A3=%EQ(A,1)-%EQ(A,3)
       : B2=%EQ(B,1)-%EQ(B,2)
       : B3=%EQ(B,1)-%EQ(B,3)
       : D2=%EQ(D,1)-%EQ(D,2)
       : P5=A2*D2  :  P6=A3*D2  :  P7=B2*D2  :  P8=B3*D2
$CALC  Q1=A2*B2*D2  :  Q2=A2*B3*D2  :  Q3=A3*B2*D2  :  Q4=A3*B3*D2
$FIT  A2+A3+B2+B3+D2+P5+P6+P7+P8+Q1+Q2+Q3+Q4
$DISPLAY MAR
$STOP

   29.6560   29.6560   26.9251   26.9251   39.1236   39.1236
   11.6009   46.4037   10.5411   42.1644   30.6373   30.6373
    5.8140   66.8605    5.2770   60.6860   26.0417   26.0417

 GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

          SCALED
  CYCLE   DEVIANCE        DF
     2     .3784E-04       4

 Y-VARIATE  Y
 ERROR POISSON  LINK LOG

 LINEAR PREDICTOR
 %GM A2 A3 B2 B3 D2 P5 P6 P7 P8 Q1 Q2 Q3 Q4

        ESTIMATE          S.E.          PARAMETER
   1     3.232           .5313E-01      %GM
   2     .2732E-01       .7168E-01      A2
   3     .1902           .7915E-01      A3
   4     .1568           .7541E-01      B2
   5    -.2172           .6602E-01      B3
   6    -.4253           .5116E-01      D2
   7     .3678E-01       .6854E-01      P5
   8     .3885           .7391E-01      P6
   9     .2126           .7101E-01      P7
  10    -.4253           .6638E-01      P8
  11    -.1858E-01       .9021E-01      Q1
  12     .3678E-01       .9110E-01      Q2
  13    -.1941           .9116E-01      Q3
  14     .3885           .9687E-01      Q4
 SCALE PARAMETER TAKEN AS        1.000

 UNIT   OBSERVED       FITTED        RESIDUAL
   1         30        29.65          .1575E-02
   2         30        29.65          .1575E-02
   3         27        26.92          .4664E-03
   4         27        26.92          .4664E-03
   5         39        39.13         -.1758E-02
   6         39        39.13         -.1758E-02
   7         12        11.61         -.1740E-02
   8         46        46.41         -.8696E-03
   9         11        10.54          .1480E-02
  10         42        42.16          .7411E-03
  11         31        30.64          .2008E-03
  12         31        30.64          .2008E-03
  13          6        5.817         -.1107E-02
  14         67        66.86         -.3244E-03
  15          5        5.284         -.3155E-02
  16         61        60.69         -.9286E-03
  17         26        26.03          .1937E-02
  18         26        26.03          .1937E-02
```

Appendix 4

Fitting a saturated log-linear model on the data of civilian labor force classified by age and years of reporting, i.e. 1969, 1971 and 1973.

```
$UNITS  30
$DATA Y
$FACTOR  A 5  B 3  D 2
$CALC    A=%GL(5,6) :   B=%GL(3,2) :   D=%GL(2,1)
$DINPUT 1
$LOOK Y
$YVAR Y
$ERROR P
$CALC    A2=%EQ(A,1)-%EQ(A,2)
        :A3=%EQ(A,1)-%EQ(A,3)
        :A4=%EQ(A,1)-%EQ(A,4)
        :A5=%EQ(A,1)-%EQ(A,5)
        :B2=%EQ(B,1)-%EQ(B,2)
        :B3=%EQ(B,1)-%EQ(B,3)
        :D2=%EQ(D,1)-%EQ(D,2)
$CALC   P1=A2*B2 :  P2=A2*B3 :  P3=A3*B2 :  P4=A3*B3 :  P5=A4*B2
      : P6=A4*B3 :  P7=A5*B2 :  P8=A5*B3
$CALC   P9=A2*D2 :  P10=A3*D2 :  P11=A4*D2 :  P12=A5*D2 :  P13=B2*D2
       : P14=B3*D2
$CALC   Q1=A2*B2*D2 :  Q2=A2*B3*D2 :  Q3=A3*B2*D2 :  Q4=A3*B3*D2
      : Q5=A4*B2*D2 :  Q6=A4*B3*D2 :  Q7=A5*B2*D2 :  Q8=A5*B3*D2
$FIT  A2+A3+A4+A5+B2+B3+D2+P1+P2+P3+P4+P5+P6+P7+P8+P9+P10+P11+P12
     +P13+P14+Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8
$DISPLAY  MERT
$STOP

   668    5122     986    4949     907    5247
   784   18581    1522   18011    1223   19103
   413   19155     751   17672     547   16112
   310   14250     534   13607     399   12421
    58    2508      81    2117      45    1949

 GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON
 CYCLE  DEVIANCE       DF
   2    .1475E-08      0

Y-VARIATE  Y
ERROR POISSON  LINK LOG

LINEAR PREDICTOR
%GM A2 A3 A4 A5 B2 B3 D2 P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12 P13 P14 Q1 Q
Q3 Q4 Q5 Q6 Q7 Q8

       ESTIMATE       S.E.        PARAMETER
   1    7.551       .9061E-02    %GM
   2   -.8805       .1143E-01    A2
   3   -.4953       .1334E-01    A3
   4   -.2019       .1458E-01    A4
   5    1.664       .3112E-01    A5
   6   -.1271       .1193E-01    B2
   7    .3760E-01   .1342E-01    B3
   8   -1.516       .9061E-02    D2
   9    .4874E-02   .1505E-01    P1
  10    .8968E-01   .1649E-01    P2
  11    .2732E-01   .1754E-01    P3
  12   -.1263E-01   .1925E-01    P4
  13    .1958E-01   .1924E-01    P5
  14   -.7002E-02   .2097E-01    P6
  15    .1203E-01   .4086E-01    P7
  16   -.1585       .4694E-01    P8
  17   -.1182       .1143E-01    P9
  18    .2140       .1334E-01    P10
  19    .2350       .1458E-01    P11
  20    .2840       .3112E-01    P12
  21   -.1413       .1193E-01    P13
  22   -.6380E-02   .1342E-01    P14
  23    .2069E-01   .1505E-01    Q1
  24    .1685E-01   .1649E-01    Q2
  25    .9189E-02   .1754E-01    Q3
  26    .3186E-01   .1925E-01    Q4
  27   -.9619E-02   .1924E-01    Q5
  28    .2521E-01   .2097E-01    Q6
  29    .2677E-01   .4086E-01    Q7
  30   -.9083E-01   .4694E-01    Q8
  SCALE PARAMETER TAKEN AS    1.000
```

Appendix 5

Table 5.a. Log-linear parameters of "age-time interaction" U_{xi}^{AB}

Age	Years (States) 1969	1971	1973
14-19	-.0246	-.0638	.0884
20-34	-.0946	.0049	.0897
35-49	-.0147	.0273	-.0126
50-64	-.0126	.0196	-.0070
65+	.1465	.0120	-.1585

Source: Appendix 4.

Table 5.b. "Age-time interaction" parameters w_{xi}^{AB} of the multiplicative model.

Age	Years (States) 1969	1971	1973
14-19	.9757	.9382	1.0924
20-34	.9097	1.0049	1.0938
35-49	.9854	1.0277	.9875
50-64	.9875	1.0198	.9930
65+	1.1578	1.0121	.8534

Source: Table 5.a.

Appendix 6

Matrix M - crude and standardized rates [a]

Standard used $V^s(x)$	States 1969	1971	1973
1969	.0361	.0633	.0513
1971	.0368	.0643	.0522
1973	.0380	.0663	.0539

[a] Obtained by using the age-compositions as a standard for the prevalence rates in each year.
Source. Table 3.

IPD-WORKING PAPERS 1983

PRICE PER COPY	
BF 230 US$ 6.00	1983-1. R. Lesthaeghe : "A Century of Demographic and Cultural Change in Western Europe : An Exploration of Underlying Dimensions".
BF 300 US$ 8.00	1983-2. R. Lesthaeghe, C. Vanderhoeft, S. Becker, M. Kibet : "Individual and Contextual Effects of Education on Proximate Determinants and on Life Time Fertility in Kenya".
BF 230 US$ 6.00	1983-3. F. Eelens : "Impact of Breast-feeding on Infant and Child Mortality with varying Incidence of Malaria - Evidence from the Kenya Fertility Survey 1977-78".
BF 230 US$ 6.00	1983-4. S. Becker, A. Chowdhury, S. Huffman : "Determinants of Natural Fertility in Matlab, Bangladesh".
BF 350 US$ 9.00	1983-5. C. Vanderhoeft : "A Unified Approach to Models for Analysis of Zero-One Data with Applications to Intermediate Fertility Variables".
BF 300 US $ 8.00	1983-6. S. Wijewickrema, R. Bulté : "Migration Impact on Population Growth in Belgium : A Multiregional Investigation with Detailed Projection Results for the Period 1976-2001".
BF 230 US$ 6.00	1983-7. S. Wijewickrema et Alii : "Marital Status Trends in Belgium, 1961-77 : Application of Multi-State Analysis".
BF 230 US$ 6.00	1983-8. H.J. Page : "Fertility and Family : New Currents and emerging Emphases in Research and Policy" (Overview of the UN Expert Group Meeting on Fertility and Family (New Delhi, 1983) preparatory to the International Conference on Population).
BF 230 US$ 6.00	1983-9. R. Lesthaeghe, C. Vanderhoeft, S. Becker, M. Kibet : "Individual and Contextual Effects of Female Education on the Kenyan Marital Fertility Transition" (abbreviated non-technical version of Working Paper 1983-2).

IPD-WORKING PAPERS 1984

PRICE PER COPY	
BF 230 US$ 6.00	1984-1. F. Rajulton : "An Age Dependent Semi-Markov Model of Marital Status in Belgium : An Application of Littman's Algorithm to Period Data, 1970".
BF 400 US$10.00	1984-2. R. Lesthaeghe : "Fertility and its Proximate Determinants in sub-Saharan Africa : The Record of the 1960's & 70's".
BF 300 US$ 8.00	1984-3. C. Vanderhoeft : "Life Tables and Relational Models : A Unified Approach, and Applications in Demography".
BF 230 US$ 6.00	1984-4. R. Lesthaeghe : "Demographic Renewal and Problems of Social Organization".
BF 230 US$ 6.00	1984-5. S. Becker : "A Response Bias in the reporting of Month of Birth in Pregnancy History Surveys".
BF 230 US$ 6.00	1984-6. R. Lesthaeghe : "On the Adaptation of sub-Saharan Systems of Reproduction".

IPD-WORKING PAPERS 1985

PRICE PER COPY	
BF 330 US$ 8.50	1985-1. R. Lesthaeghe and F. Eelens : "Social Organization and Reproductive Regimes : Lessons from sub-Saharan Africa and historical Western Europe".
BF 300 US$ 8.00	1985-2. P. Willems and S. Wijewickrema : "The Evolution of Nuptiality in Belgium from 1954 to 1981".
BF 400 US$10.00	1985-3. F. Eelens and L. Donné : "The Proximate Determinants of Fertility in Sub-Saharan Africa: A Factbook based on the Results of the World Fertility Survey".
BF 130 US$ 3.50	1985-4. Ronald C. Schoenmaeckers : "Current Fertility Behaviour in Africa. Results from a Birth Interval Analysis of WFS Data".
BF 400 US$10.00	1985-5. Camille Vanderhoeft : Stratified Proportional Hazards Models. A GLIM oriented approach, with special reference to the problem of competing risks.
BF 250 US$ 6.50	1985-6. Stanley Wijewickrema : "Childlessness in Belgium and Flanders".
BF 250 US$ 6.50	1985-7. R. Lesthaeghe : "Value Orientations, Economic Growth and Demographic Trends - Toward a Confrontation?
BF 250 US$ 6.50	1985-8. Fernando Rajulton and Stan Wijewickrema : "An Enquiry into the Two Basic Assumptions of Multistate Increment-Decrement Life Tables".

IPD-WORKING PAPERS 1986

PRICE PER COPY	
BF 250 US$ 6.50	1986-1. F. Rajulton : "Marriageability and divorceability : a simulation of the unobservables through the conditional gaussian diffusion process.
BF 180 US$ 4.50	1986-2. H.J. Page : "Child-bearing versus Child-rearing : Co-residence of Mothers and Children in Sub-Saharan Africa".
BF 450 US$11.50	1986-3. R. Lesthaeghe, G. Kaufmann, D. Meekers : "The Nuptiality Regimes in Sub-Saharan Africa".
BF 230 US$ 6.00	1986-4. R. Lesthaeghe, G. Kaufmann, D. Meekers : "The Nuptiality Regimes in Sub-Saharan Africa : Data Files NUPFILE 1 and NUPFILE 2".
BF 200 US$ 5.00	1986-5. Hwa Young LEE and Fernando RAJULTON : "A Semi-Markovian Multistate Analysis of Sexual Union Formation in Flanders, Belgium".
BF 450 US$11.50	1986-6. Zeng Yi : "Fertility and Household Dynamics in China".
BF 300 US$ 8.00	1986-7. R. Lesthaeghe & D. Meekers : "Value Changes and the Dimensions of Familism in the EEC".

IPD-WORKING PAPERS 1987

PRICE PER COPY	
BF 250 US$ 6.50	1987-1. C. Vanderhoeft : "Hidden heterogeneity in composite link models : further developments".
BF 250 US$ 6.50	1987-2. M. Rafiq Shah : "The use of a piecewise constant proportional hazards model in issues related to standardisation : a review and some results".
BF 200 US$ 5.50	1987-3. R. Lesthaeghe : "The breast-feeding hypothesis and regional differences in marital fertility and infant mortality in the Low Countries during the 19th Century. - Comments on a debate".
BF 200 US$ 5.50	1987-4. M. Rafiq Shah : "A critique of Clogg's method of rates adjustment".

A CRITIQUE OF CLOGG'S METHOD OF RATES ADJUSTEMENT *

M. Rafiq Shah, Ph.D.

Interuniversity Programme in Demography
Vrije Universiteit Brussel

Abstract

Clogg (1978) has developed a method for the adjustement of demographic rates by purging the so called undesirable interactions from a saturated multiplicative model. The summary rates of the purged frequencies are known as the adjusted rates. This paper critically examines the proposed method of finding adjusted rates by purging two factor interactions, and relates it to the traditional method of direct standardization by defining the prevalence rates in terms of the parameters of the model. A comparison of the results reveals that unlike directly standardized rates, the adjusted rates carry with them the effects of the confounding factor. The case of finding adjusted rates based on purging both two and three factors interactions is also discussed. The examples quoted by Clogg (1978) are used as numerical illustrations.

* Prepared for the annual meeting of the Population Association Of America, to be held in Chicago during April 30 - May 2, 1987.

Introduction

A simple measure of summarizing the experience of a demographic process is the crude rate. This rate is, however, not suitable for the purpose of making a comparison of any given process due to the built-in defect of such a measure when seen as a weighted average; the relevant weights being the proportions of an associated variable or factor. It is precisely for this reason that the method of direct standardization was introduced as early as 1883 (Wolfenden, 1954), so that comparisons could be made on the basis of standardized rates.

According to the method of direct standardization, the weights proper to each population compared (for example, the proportions in various age groups of the population under comparison) are replaced by a common set of weights (the "standard age composition"). The expected summary rates caculated by applying the "standard" are known as directly standardized rates.

Even though these rates have the advantage of being easy to calculate, it is the selection of the "standard" that is found to be really problematic. In fact, one cannot find a "unique" standard for the calculation of directly standardized rates. Several methods have been suggested for finding summary indices either by using an endogenous standard or by using no standard at all.

In all these attempts, the use of statistical models have played an important role. In particular, the models that are commonly used for the analysis of contingency tables (e.g., multiplicative or log-linear models) have provided an alternative to the traditional techniques of standardization. Besides other advantages, each of these models provides unique estimates of the required summary indices without the intermediary use of a standard.

This paper deals with the method of adjusting summary rates, by postulating a saturated multiplicative model on the available data, as proposed by Clogg (1978). Since a saturated model fully describes the data (e.g. contains all possible main and interaction effects), the interaction that brings about most of the confounding effects on the summary measures (the undesirable interaction) is first identified. This interaction is then purged out of the data. The summary rates based on purged and rescaled data are known as adjusted rates. The method has been applied in labor force analysis (Clogg, 1979; Clogg and Shockey, 1985).

As the method of purging interaction according to Clogg's method and that of direct standardization both deal with the procedures of removing or controlling the confounding effects of the summary rates, this paper critically examines the properties of adjusted rates in comparison to directly standardized rates.

To begin with, the type of a contingency table that could be used for the calculation of prevalence rates and the rates of non-renewable events is described in terms of a general-purpose terminology. The model that characterizes various effects of the factors involved will follow immediately.

Terminology

Groups, populations or periods of observations in which a demographic process is required to be compared will be called the "state factor" or the "states". The demographic process in question will be called the "outcome factor" or the "outcome". The effects of the states will be assumed to depend on some background factor(s) influencing a demographic process under study. A background factor will be called a "confounding factor" if, besides its influence on the outcome, it is distributed differently in the various states. For example, in the problems of standardization and in most of the discussions in this paper, confounding factor, outcome and state factor represent age composition, employment status and groups or years of observations respectively.

Summary Rates in a Contingency Table and the Model

Assume that the outcome D is polytomous having K categories k (k=1,...K) and is classified by the confounding factor A, having R levels x (x=1,...R) and the state factor B having C categories i (i=1,...C). Denote the number of events in the i-th state, the k-th category of D and the x-th level of A by $E_{ik}(x)$. The presentation of such a table is found to be consistent with the description of prevalence rates and the rates of non-renewable events, some of which could be seen in numerical illustrations. A detailed account of the analysis of such tables may be found in Bishop et. al (1975: 31-41), Goodman (1978, Chapter 4), Haberman (1978,Vol 1, Chapter 3) and Willekens (1983).

Note that events give rise to occurence /exposure rates and the presence in any specified category are used for the purpose of prevalence rates. As most of the discussion in this paper relates to the prevalence rates, events simply mean the presence of individuals in the specified categories of the factors involved. Accordingly, the proportion of events occuring in the i-th state that fall in the k-th class of D and the x-th level of A is the xi-specific rate of the k-th class of D:

$$r_{ik}(x) = E_{ik}(x)/\Sigma_k E_{ik}(x) = E_{ik}(x)/E_{i+}(x) \tag{1}$$

The crude rate of the k-th class of D in the i-th state is:

$$r_{ik}(.) = \Sigma_x E_{ik}(x)/\Sigma_{kx} E_{ik}(x) = E_{ik}(+)/E_{i+}(+) \tag{2}$$

The crude rate (2) can be expressed as a weighted average of the xi-specific rate $r_{ik}(x)$, the weights being the proportions of the confounding factor A:

$$r_{ik}(.) = \Sigma_x r_{ik}(x).V_i^A(x) \tag{3}$$

where $V_i^A(x) = E_{i+}(x)/E_{i+}(+)$.

The model that describes the data fully in a contingency table is the saturated model:

$$E_{ik}(x) = w.w_x^A.w_i^B.w_k^D.w_{xi}^{AB}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}. \tag{4}$$

The usual constraints on the parameters are:

$$\Pi_x w_x^{A} = \Pi_i w_i^{B} = \Pi_k w_k^{D} = \Pi_x w_{xi}^{AB} = \ldots = \Pi_x w_{xik}^{ABD} = \ldots = 1$$

In a saturated model, since there are as many independent parameters in the equation as there are types of different effects describing the data, the estimated value for $E_{ik}(x)$ is equal to the observed value. The parameters of the model are: the overall effect (w), main effects (w_x^{A}, w_i^{B}, w_k^{D}), first order interaction effects (w_{xi}^{AB}, w_{xk}^{AD}, w_{ik}^{BD}) and the second order interaction effect (w_{xik}^{ABD}).

Of interest here in the model (4) is the interpretation of the first order or two factor interaction effects. For instance, w_{xi}^{AB} representes state differences in the distribution of the confounding factor A, the parameter w_{xk}^{AD} measures the average effect of the confounding factor A on D, and w_{ik}^{BD} measures the state diffences in D.

Note that the additive form of the multiplicative model (4) is obtained by taking logarithms:

$$\log E_{ik}(x) = U + U_x^{A} + U_i^{B} + U_k^{D} + U_{xi}^{AB} + U_{xk}^{AD} + U_k^{BD} + U_{xik}^{ABD} \qquad (5)$$

with $U = \log(w)$, etc.

The usual restrictions of the parameters of the log-linear model (5) are:

$$\Sigma_x U_x^{A} = \Sigma_i U_i^{B} = \Sigma_k U_k^{D} = \Sigma_x U_{xi}^{AB} = \ldots = \Sigma_x U_{xik}^{ABD} = \ldots = 0$$

An immediate advantage of these models is that one can easily express the xi-specific rates $r_{ik}(x)$, the crude rates $r_{ik}(.)$ and the weights $V_i^{A}(x)$ in terms of the parameters of the model. For instance, the rate $r_{ik}(x)$, defined in (1) may be written as follows:

$$r_{ik}(x) = \frac{w.w_x^{A}.w_i^{B}.w_k^{D}.w_{xi}^{AB}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}}{\Sigma_k\, w.w_x^{A}.w_i^{B}.w_k^{D}.w_{xi}^{AB}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}}$$

$$= \frac{w_k^{D}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}}{\Sigma_k\, w_k^{D}.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}}\,. \tag{6}$$

The crude rate (2) is:

$$r_{ik}(.) = \frac{w_k^{D}.\ w_{ik}^{BD}.\Sigma_x w_x^{A}.w_{xi}^{AB}.w_{xk}^{AD}.w_{xik}^{ABD}}{\Sigma_k\, w_k^{D}.\ w_{ik}^{BD}.\Sigma_x w_x^{A}.w_{xi}^{AB}.w_{xk}^{AD}.w_{xik}^{ABD}} \tag{7}$$

and the weights $V_i^A(x)$ as in (3) are:

$$V_i^A(x) = \frac{w_x^{A}.w_{xi}^{AB}.\ \Sigma_k w_k^{D}.\ w_{ik}^{BD}.w_{xk}^{AD}.w_{xik}^{ABD}}{\Sigma_x w_x^{A}.w_{xi}^{AB}.\ \Sigma_k w_k^{D}.\ w_{ik}^{BD}.w_{xk}^{AD}.w_{xik}^{ABD}}\,. \tag{8}$$

<u>Clogg's Method of Adjustement</u>

Due to the fact that the two factor interaction w_{xi}^{AB} in the multiplicative model (4) represents state (factor B) differences in the distribution of the confounding factor A (e.g. differences in the age compositions), Clogg (1978) proposed the adjustement of summary rates by purging this interaction from the saturated model (4) of counts. The summary rates based on purged frequencies are known as the "adjusted rates". As the adjusted rates do not possess the effects of the confounding factor, their differences are thought to show the "true" differences of the outcome in question. Clogg's method of adjustement based on purging and rescaling is explained in the following subsections.

1. Purging of Undesired Interaction

The interaction w_{xi}^{AB} which is thought to confound or obscure summary rates is removed from an assumed saturated model (4) of counts by the simple process of division. Consequently, frequencies expressed without this interaction, the so called "purged frequencies", are:

$$E_{ik}^*(x) = E_{ik}(x)/w_{xi}^{AB} = w.w_x^A.w_i^B.w_k^D.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}. \quad (9)$$

Since the original frequencies change due to the process of division, they could be rescaled in order to sum to the observed total frequencies in each state.

2. Rescaling

As the sum $E_{++}^*(+)$ of the purged frequencies $E_{ik}^*(x)$ will not be equal to the corresponding sum $E_{++}(+)$ of the observed frequencies $E_{ik}(x)$, each $E_{ik}^*(x)$ must be rescaled in order for the sum of the purged frequencies to be equal to that of the observed frequencies. This is done by multiplying each $E_{ik}^*(x)$ by the ratio of the total observed frequency to the total purged frequency, i.e. by the ratio $E_{++}(+)/E_{++}^*(+)$. Denoting the rescaled frequencies so obtained by $E_{ik}^{**}(x)$, we have

$$E_{ik}^{**}(x) = w'.w_x^A.w_i^B.w_k^D.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}. \quad (10)$$

where $w'=w.(E_{++}(+)/E_{++}^*(+))$.

To preserve the total of the observed frequencies in each state, one has to multiply the purged frequencies $E_{ik}^*(x)$ by the ratio $E_{i+}(+)/E_{i+}^*(+)$. The rescaled frequencies $E_{ik}^{\hat{}}(x)$ so obtained are:

$$E_{ik}^{\hat{}}(x) = E_{ik}^*(x)[E_{i+}(+)/E_{i+}^*(+)]$$

$$= w'.w_x^A.(w_i^B)'.w_k^D.w_{xk}^{AD}.w_{ik}^{BD}.w_{xik}^{ABD}, \quad (11)$$

where

$$(w_i^B)' = w/w'.[\Pi_{xk}(E_{ik}^{''}(x)/E_{ik}^{*}(x))]^{1/(RK)}$$

$$= [E_{++}^{*}(+)/E_{++}(+)].[E_{i+}(+)/E_{i+}^{*}(+)].w_i^B.$$

3. Adjustement

The adjusted rate $r_{ik}^s(.)$ of the k-th class of D in the i-th state is obtained by replacing the observed frequencies $E_{ik}(x)$ by the purged frequencies $E_{ik}^{*}(x)$ or by the purged and rescaled frequencies $E_{ik}^{**}(x)$. Using the purged and rescaled frequencies $E_{ik}^{''}(x)$, the adjusted rate according to Clogg's method is:

$$r_{ik}^s(.) = \Sigma_x E_{ik}^{''}(x)/\Sigma_{xk} E_{ik}^{''}(x) \qquad (12)$$

$$= E_{ik}^{''}(+)/E_{i+}^{''}(+) = E_{ik}^{*}(+)/E_{i+}^{*}(+)$$

Comments and Comparisons

The method of adjustement presented by Clogg (1978) displays a useful application of the saturated models in demographic analysis and is considered to be a breaktrough in the methodology of standardization. The method provides a basis for the replacement of the traditional method of components analysis (Kitagawa, 1955) in case of several interacting factors. Clogg's method is flexible and has the capacity of accomodating a number of factors and states simultaneously. Clogg (1982) has also made available a computer program (PURGE) which may be used for the comparison of several states classified by several factors where each factor could have many categories, and where purging of higher order interaction is felt necessary.

Keeping in view, however, an important criterion of a standardized rate, namely that a standardized rate should be independent of the compositions of the states under comparison, the adjusted rates according to Clogg's method are found to be lacking this property. The arguments in favour of illustrating the lack of such an important criterion could be presented as follows.

The lack of this criterion arises due to the process of purging undesirable interaction (w_{xi}^{AB}) only, from the saturated model (4). Note that purging of other interactions gives entirely different results (Shah, 1986).

According to the saturated model (4), the xi-specific rates $r_{ik}(x)$ are independent of the AB interaction as shown in (6). The observed rates do therefore not change during the process of purging the AB interaction. Since the $r_{ik}(x)$ do not change, one may compare the adjusted rate $r_{ik}^{s}(.)$ with the crude rate $r_{ik}(.)$ as follows.

A comparison of the crude rate (7) with the adjusted rate (12) expressed in terms of the parameters of the model (4) indicated that the difference between the two is entirely due to the absence of the AB interaction term in the adjusted rate.

Due to the fact that the $r_{ik}(x)$ do not change, the adjusted rate $r_{ik}^{s}(.)$ could be expressed as a weighted average of the xi-specific rates $r_{ik}(x)$ as follows.

$$r_{ik}^{s}(.) = \Sigma_x r_{ik}(x).V_i(x) \qquad (13)$$

where

$$V_i(x) = E_{i+}^{*}(x)/E_{i+}^{*}(+).$$

The weights $V_i(x)$ are the state factor specific proportions of the confounding factor, after having removed the AB interaction. The method thus produces a "purged confounding factor" only. According to (13), these weights depend on i and are not identical in all states under comparison. It follows that the adjusted rates based on Clogg's method include the effects of the confounding factor. The confounding factor is thus neither controlled nor eliminated.

Since the xi-specific rates $r_{ik}(x)$ are not affected by the method of adjustement as stated earlier, and the method of adjustement produces changes in the confounding factors, i.e. the weights $V_i^A(x)$, the adjusted rate being a summary measure could be compared with a directly standardized rate.

Given a set of standard weights $V^s(x)$ that is independent of the states under comparison, the directly standardized rate of the k-th class of D in the i-th state i is defined as

$$DSR_{ik} = \Sigma_x r_{ik}(x).V^s(x) \qquad (14)$$

where

$$V^s(x) = E_{i+}{}^s(x)/E_{i+}{}^s(+), \text{ such that } \Sigma_x V^s(x) = 1.$$

A comparison of (13) and (14) reveals that whereas $r_{ik}(x)$ is common in both $r_{ik}{}^s(.)$ and DSR_{ik}, the weights $V_i^A(x)$ and $V^s(x)$ are different. Common to these weights, however, is that both are independent of the AB interaction. Therefore, the difference between $r_{ik}{}^s(.)$ and DSR_{ik} cannot be attributed to the AB interaction, but to the difference in magnitude arising in $V_i^A(x)$ and $V^s(x)$.

It may be noted that while $V_i^A(x)$ depends on AB, BD and ABD interactions, $V^s(x)$ is independent of all types of interactions as it remains constant over all the states under comparison. This finding is by no means related to the proof that the direct method of standardization is better than the method based on a multiplicative model. It may be used merely for the identification of an important property of a standardized rate, namely that the rate is independent of the compositions of the states under comparison. This property is not fulfilled by the adjusted rate (after purging two factor interaction from a saturated model) as proposed by Clogg (1978).

We shall demonstrate through numerical illustrations that the bias arising due to the dependence of weights $V_i^A(x)$ on i could lead to different (misleading) inferences and conclusions. Out of several examples the ones quoted by Clogg (1978) are chosen to be presented here for ready reference and comparison with Clogg's results.

Numerical Results

1. Hypothetical Data

Table 1 shows the hypothetical data classified by confounding factor (composition), state factor (groups) and dichotomous outcome factor. The composition-specific rates of states 1 and 2 are the same but the crude rates differ because of compositional differences. On the other hand, the composition of states 1 and 3 is the same but the composition-specific rates are different. The purpose is to compare the states in order to identify differences in the prevalence rates of the outcome factor. A saturated log-linear model is fitted to the data of Table 1 using GLIM (Generalized Linear Interactive Modelling; Baker and Nelder, 1978). The parameters of the model are shown in Appendix 1. The computer listing and program are laid out in Appendix 2.

Table 1. Hypothetical data: frequencies $E_{ik}(x)$ and rates $r_{ik}(x)$ by state.

	$E_{ik}(x)$									$r_{ik}(x)$		
State factor	1			2			3			1	2	3
Outcome factor/ Confounding factor	1	2	Total	1	2	Total	1	2	Total			
1	25	25	50	50	50	100	25	25	50	.50	.50	.50
2	10	40	50	15	60	75	25	25	50	.20	.20	.50
3	8	92	100	2	23	25	50	50	100	.08	.08	.50
Total	43	157	200	67	133	200	100	100	200			
Crude rates										.215	.335	.500

For k=1 the rate $r_{i1}(x)= E_{i1}(x)/E_{i+}(x)$, e.g. in state 1, the rate at the 1st level of the confounding factor is $r_{11}(1)$=25/50=0.5.
Source: Clogg (1978), Table 1: 528.

The expected frequencies $E_{ik}^{*}(x)$ obtained after purging the AB interactions as proposed by Clogg (1978) are given in Table 2.a. Calculations based on these frequencies indicate that the xi-specific rates $r_{ik}^{*}(x)$ of the k-th category of D are equal to the observed rates $r_{ik}(x)$ as shown in Table 1. For instance, $r_{11}^{*}(1)$ based on Table 2.a is $r_{11}^{*}(1)=E_{11}^{*}(1)/E_{1+}^{*}(1)$ =29.656/59.312=0.50, and the same rate based on the data of Table 1 is $r_{11}(1)=E_{11}(1)/E_{1+}(1)$=25/50=0.50. These calculations correspond to the observations that the xi-specific rates of any category of D are independent of the AB interactions (see equation (6)).

Table 2.a. Frequencies $E_{ik}^{*}(x)$ purged of AB interactions.

State factor	1		2		3	
Outcome factor/ Confounding factor	1	2	1	2	1	2
1	29.6560	29.6560	26.9251	26.9251	39.1236	39.1236
2	11.6009	46.4037	10.5411	42.1644	30.6373	30.6373
3	5.8140	66.8605	5.2770	60.6860	26.0417	26.0417
Total	47.0709		42.7432		95.8026	

$E_{ik}^{*}(x) = E_{ik}(x)/w_{xi}^{AB}$; e.g. $E_{11}^{*}(1) = 29.6560 = 25/\exp(-.1702)$
Source: Table 1 and Appendix 1.

Notice the variation in the distribution of the confounding factor over the states as shown in Table 2.b. Note that these proportions are used as weights for the calculation of adjusted rates by equation (13). Since these weights depend on the states under comparison, the adjusted rates obtained after purging AB interactions only do not satisfy the property that a standardized index be independent of the compositional distribution of the states under comparison. We shall see later that the absence of this property of the adjusted rate could give different (misleading) results. The proportions $V_i^A(x)$ vary according to the pattern of the xi-specific rates. The proportions are identical only in the states where the xi-specific rates are identical. For instance, since the rates $r_{ik}(x)$ in states 1 and 2 are identical, the weights $V_i^A(x)$ associated with states 1 and 2 are also identical.

Table 2.b. Frequencies $E_{i+}^*(x)$ and proportions $V_i(x)$ purged of AB interactions by state.

State factor	1		2		3	
Confounding factor	$E_{1+}^*(x)$	$V_1(x)$	$E_{2+}^*(x)$	$V_2(x)$	$E_{3+}^*(x)$	$V_3(x)$
1	59.3120	0.3122	53.8502	0.3121	78.2472	0.4084
2	58.0046	0.3053	52.7055	0.3055	61.2746	0.3198
3	72.6745	0.3825	65.9630	0.3824	52.0834	0.2718
Total	189.9911	1.0000	172.5187	1.0000	191.6052	1.0000

Source: Table 2.a.

Since the adjusted rates (13) are not really standardized in a conventional sense, as we have noted above, they cannot be used for decomposing the difference of rate and composition components as proposed by Kitagawa (1955). Except in situations where the xi-specific rates are identical, the estimates of "rate" and "composition" components of the difference of crude rates is biased. For example, the difference between the crude rate of state 1 and that of state 3 is 0.335-0.500=-0.165 or -16.50%. The corresponding difference in adjusted rates is 0.248-0.500=-0.252 or -25.20%. The difference between these quantities (8.7%; -16.50+25.20) is an estimate of the effect of the confounding factor. This estimate is, however, based on the assumption that the set of weights is common to both states (as is normally the case in direct standardization).

2. Empirical Data

The hypothetical data in Table 1 are constructed such that the xi-specific rates are chosen to be identical in the categories of the confounding factor in state 1 and state 2 whereas in state 3 these rates are identical. Such data conceal in part the drawbacks of Clogg's adjustement method. Therefore we consider another data set. Table 3 shows the U.S. civilian labour force data classified by age and year of reporting. The objective is to see if age composition has played an important role in the process of unemployment over the reported years.

Table 3. U.S. Civilian labor force classified by age and year of reporting ($E_{ik}(x)$ with i=year, k=employment status, x=age)

	1969			1971			1973		
Age	unempl.	empl.	total	unempl.	empl.	total	unempl.	empl.	total
14-19	668	5,122	5,790	986	4,949	5,935	907	5,247	6,154
20-34	784	18,581	19,365	1,522	18,011	19,533	1,223	19,103	20,326
35-49	413	19,155	19,568	751	17,672	18,423	547	16,112	16,659
50-64	310	14,250	14,560	534	13,607	14,141	399	12,421	12,820
65+	58	2,508	2,566	81	2,177	2,198	45	1,949	1,994
Total	2,233	59,616	61,849	3,874	56,356	60,230	3,121	54,832	57,953
Crude unemployment rate (percent)	3.61			6.43			5.39		

Source: Clogg, 1978, Table 5: 536 (Data from March Current Population Survey)

As before a saturated log-linear model is fitted on the data of Table 3 by using GLIM. Computer output and the parameters of the model are displayed in Appendix 4. The parameters required for the purpose of purging age-time interaction are shown in Appendix 5. Following our critical remarks on the uneven distribution of weights $V_1(x)$, when the adjusted rates $r_{ik}^{\$}(.)$ are expressed as a weighted average of the xi-specific rates $r_{ik}(x)$, attention is focussed on the purged distribution of the state factor $E_{i+}^{*}(x)$.

Sets of purged distributions $E_{i+}^{*}(x)$ are obtained by dividing the observed distribution of the background factor $E_{i+}(x)$ by the appropriate interaction terms (see Appendix 5, Table 5). Note that the interaction term is common in both categories of the outcome factor D in a specified year and age group. Purged totals for each i and x, $E_{i+}^{*}(x)$, and purged weights $V_i(x)$ are shown in Table 4. The "purged weights" are not identical in all the years under study, implying that the background factor is still a confounding factor and the adjusted rates $r_{ik}^{\$}(.)$ based on these weights carry with them confounding effects.

Table 4. U.S. Civilian labor force. Purged counts $E_{i+}^{*}(x)$ and proportions $V_i(x)$ by age and years.

States	1969	1971	1973	1969	1971	1973
Confounding factor Age	Frequencies $E_{i+}^{*}(x)$			Proportions $V_i(x)$		
14-19	5934.2011	6325.9433	5632.9519	0.0926	0.1059	0.1000
20-34	21284.8980	19437.7550	18582.9200	0.3322	0.3254	0.3299
35-49	19857.9260	17926.4380	16869.8730	0.3099	0.3001	0.2995
50-64	14775.7260	13866.4440	12910.3730	0.2306	0.2322	0.2292
65+	2216.4637	2171.7222	2336.5362	0.0346	0.0364	0.0415
All ages	64069.2158	59728.3035	56332.6541	0.9999	1.0000	1.0001

Source: Table 3 and Appendix 5.b., e.g. the totals in each year are obtained by first dividing the number of employed and unemployed by the AB interaction of that year. For the required level of the confounding factor as shown in Appendix 5.b, see text.

The confounding effect (bias) may be estimated by the components method. For this purpose we apply first Kitagawa's method to "adjusted rates" and next we compare the inference based on "adjusted rates" with those based on directly standardized rates. Consider the unemployment rates $r_{i1}(x)$ based on the calculations from Table 3 and the weights $V_i(x)$ from Table 4. The adjusted rates for the years 1969, 1971 and 1973 are 0.0364, 0.0653 and 0.0525 respectively. The corresponding crude unemployment rates for the same years as noted in the bottom row of Table 3 are 0.0361, 0.0643 and 0.0539 respectively.

Before commenting on the inferences based on adjusted rates for the estimation of the compositional effect or the confounding bias, it is useful to recapitulate the main points about the components of the difference of two crude rates. According to Kitagawa's suggestion, the difference in crude rates of two populations (states) is composed of a "rate effect" and a "compositional effect". The part due to rates or rate effect is estimated from the difference of directly standardized rates (population composition is common in both populations). The compositional effect is then the difference of the crude rates minus the difference of the standardized rates.

Using adjusted rates instead of directly standardized rates we find that the difference of the crude rates of 1969 and 1971 is -0.0282 = 0.0361-0.0643. The corresponding difference in the adjusted rates is -0.0289 = 0.0364-0.0653. The difference between these quantities is -0.0007 = -0.0282+0.0289 or 0.07%.

Any inference based on this figure (0.07%) about the compositional effect cannot be correct, since the estimate of the "rate component" is based on "purged crude rates" (adjusted rates) and the weights $V_i(x)$ are not independent of the state factor unlike the weights commonly used in direct standardization. Moreover, any conclusion regarding the role of AB interaction in the increase of unemployment from 1969 to 1971 is uncalled for, as far as the components of the difference of the crude rates are concerned. This is due to the rate effect which is confounded by the differences of the AB interactions in 1969 and 1971, and to the presence of other interaction effects. Note That Clogg's (1978: p.537) inference about the role of age-time-period interaction is based on this figure (0.07%).

Due to reasons noted earlier and considering Kitagawa's procedure of decomposition in case of one factor as both logical and less complicated (compared to 2 or more factor cases), it is possible to estimate the "rate component" without bias by using directly standardized rates of the states (populations). Using the observed proportions of the populations in the years under study (Table 5) as standards, we calculated directly standardized rates for 1969, 1971 and 1973. Crude as well as standardized rates are shown in matrix M (Appendix 6). Since the standardized rates obtained by using these standards differ from those based on adjusted rates we used the average composition of 1969 and 1971 as a standard (Table 5). Directly standardized rates for the years 1969 and 1971, based on this standard are 0.0365 and 0.0638 respectively, giving a rate component of -0.0273 = 0.0365-0.0638. Since the difference of crude rates in 1969 and 1971 is -0.0282 the composition component is estimated as -0.0009 = -0.0282+0.0273 or -0.09%. The use of directly standardized rates of 1969 and 1971 results in a negative effect (-0.09%) of the population structure in contrast to the one (0.07%) based on "adjusted rates". The negative effect could be interpreted as a decrease in the prevalence of unemployment due to the compositional change that occured from 1969 to 1971. Note that due to problems of weights in the adjusted rates, the conclusion based on conventional directly standardized rates seems to be correct.

Our experiments with several other sets of data suggest that the adjusted rates based on purging two factor (AB) interactions gives results different from those based on directly standardized rates. We have, therefore, tried to purge out other interactions besides the AB interaction in order to solve the problem of estimating identical weights for all the states in question. First we tried to purge out the three factor (ABD) interaction from the saturated model, the results of which are discussed in the following section.

Table 5. Observed proportional distribution $V_i^A(x)$ of the population by age and years of observations and standard $V^s(x)$

	Observed proportions $V_i^A(x)$			Standard $V^s(x)$ [a]
		Years		
Age	1969	1971	1973	
14-19	.0936	.0985	.1062	.0961
20-34	.3131	.3243	.3507	.3187
35-49	.3164	.3059	.2857	.3112
50-64	.2354	.2348	.2212	.2351
65+	.0145	.0365	.0344	.0390
All ages	1.0000	1.0000	1.0000	1.0001

[a] Average age composition of 1969 and 1971.
Source: Table 3.

The Need of Purging Higher Order Interactions

The importance of purging higher order terms is felt necessary since purging two factor (AB) interactions yields biased results. Because of the non-hierarchical nature of the log-linear model of purged frequencies $E_{ik}^*(x)$, we also purged out the three factor interaction w_{xik}^{ABD}. Dividing $E_{ik}^*(x)$ by w_{xik}^{ABD} and denoting the purged frequencies so obtained by $E_{ik}^{**}(x)$, we have

$$E_{ik}^{**}(x) = w.w_x^A.w_i^B.w_k^D.w_{xk}^{AD}.w_{ik}^{BD} \qquad (15)$$

Using (15), the weights as shown in (13) without AB and ABD interaction terms are:

$$V_i^*(x) = \frac{w_x^A.\Sigma_k w_k^D.\ w_{ik}^{BD}.w_{xk}^{AD}}{\Sigma_x\ w_x^A.\Sigma_k w_k^D.\ w_{ik}^{BD}.w_{xk}^{AD}}. \qquad (16)$$

Note that since the xi-specific rates depend on ABD interactions, they will change (smooth out) unlike the ones obtained by purging AB interaction only. Using (15) and denoting the smoothed xi-specific rates purged of both AB and ABD interaction by $r_{ik}^*(x)$.

$$r_{ik}^*(x) = \frac{w_k^D.\ w_{ik}^{BD}.w_{xk}^{AD}}{\Sigma_k w_k^D.\ w_{ik}^{BD}.w_{xk}^{AD}} = E_{ik}^{**}(x)/E_{i+}^{**}(x)\ . \qquad (17)$$

Using (17) and (16), the adjusted rates based on purging both AB and ABD interactions are:

$$r_{ik}^{s*}(.) = \Sigma_x r_{ik}^*(x).V_i^*(x) \qquad (18)$$

The adjusted rate defined by (18) is not comparable to a directly standardized rate (as we have been comparing in case of purging AB interactions only, i.e. when the rates $r_{ik}(x)$ did not change). An important point to note is that the weights used in the "smoothed" rates (18) still depend on factor x and state i. In other words, factor A is still a confounding factor.

The extent of such dependence depends on the question at hand. A series of exercises on hypothetical as well as emperical sets of data suggest that, whereas the weights $V_i^*(x)$ could practically be assumed to remain constant over the states under comparison, there are others where the difference in weights so obtained is considerably large. For instance, the results of Table 6 obtained after purging both AB and ABD interactions support the finding that the adjusted rates are not free from the effects of the confounding factor when these rates are considered as weighted averages and expressed in terms of the parameters of the proposed model.

Table 6. Frequencies $E_{i+}^{**}(x)$ and proportions $V_i^*(x)$ purged of both AB and ABD interactions.

State factor	1		2		3	
Confounding factor x	$E_{1+}^{**}(x)$	$V_1^*(x)$	$E_{2+}^{**}(x)$	$V_2^*(x)$	$E_{3+}^{**}(x)$	$V_3^*(x)$
1	67.31	0.33655	76.31	0.33655	84.26	0.42130
2	63.72	0.31860	63.72	0.31860	60.47	0.30235
3	68.97	0.34485	68.97	0.34485	55.27	0.27635
All categories	200.00	1.00000	200.00	1.00000	200.00	1.00000

Source: Table 2.a. and Appendix 1.

References

Baker, R.J. and J.A. Nelder (1978). The GLIM System Manual. Release 3. Oxford : The Numerical Algorithms Group, Royal Statistical Society, London.

Bishop, Y.M.M.; S.E. Fienberg and P.W. Holland (1975). Discrete Multivariate Analysis : Theory and Practice. M.I.T. Press.

Clogg, C.C. (1978). Adjustement of Rates Using Multiplicative Models. Demography, 15: 523-539.

---------- (1979). Measuring Unemployment: Demographic Indicators for the U.S. Labor Force. New York, Academic Press.

---------- (1982). Computer Program PURGE. Personal communication.

---------- and J.W. Shockey(1985). The Effect of Changing Demographic Position on Recent Trend in Underemployement. Demography 22: 395-414

Goodman, L.A. (1978). Analyzing Qualitative Categorical data: Log-linear Models and Latent Structures. Ed. J. Magidson, Cambridge Mass: Associates.

Haberman, S.J. (1978). Analysis of Qualitative Data. Vol. 1: Introductory Topics and Vol. 2: New Developments. New York, Academic Press.

Kitagawa, E.M. (1955). Components of a Difference between Two Rates. Journal of the Americal Statistical Association 50: 1165-1194.

Shah, M. Rafiq (1986). Standardization in Demographic Analysis. Ph. D. Thesis. Interuniversity Programme in Demography, Vrije Universiteit Brussels, Belgium.

Willekens, F. (1983). Specification and Calibration of Spatial Interaction Models: a Contingency Table Perspective and an Application to Intra-urban Migration in Rotterdam. Tijdschrift voor Economische Geografie 74: 239-252.

Wolfenden, H. H. (1954). Population Statistics and their Compilation. University of Chicago Press.

Appendix 1

Parameters of the saturated log-linear model for the data of Table 1.

$$\log E_{ik}(x) = U + U_x^{A} + U_i^{B} + U_k^{D} + U_{xi}^{AB} + U_{xk}^{AD} + U_k^{BD} + U_{xik}^{ABD}$$

Overall Mean: U = 3.233

Main Effects

U_1^A = .2174	U_1^B = -.0607	U_1^C = .4254
U_2^A = -.0272	U_2^B = -.1567	U_2^C = -.4254
U_3^A = -.1902	U_3^B = .2174	

Two Factor Interactions

AB			AD		BD	
-.1702	.6187	-.4485	.4254	-.4254	-.2127	.2127
-.1488	.3526	-.2038	-.0367	.0367	-.2127	.2127
.3190	-.9713	.6523	-.3887	.3887	.4254	-.4254

Three Factor Interactions (ABD)

.2126	-.2126	.2128	-.2128	-.4254	.4254
-.0183	.0183	-.0184	.0184	.0367	-.0367
-.1943	.1943	-.1944	.1944	.3887	-.3887

Source: Appendix 2.

Appendix 2

GLIM program for fitting the saturated model on the data of Table 1 and the estimated parameters under the usual constraints.

```
$UNITS  18
$DATA Y
$FACTOR   A 3   B 3   D 2
$CALC   A=%GL(3,6)  :   B=%GL(3,2)  :  D=%GL(2,1)
$DINPUT 1
$LOOK   Y
$YVAR   Y
$ERROR   P
$CALC      A2=%EQ(A,1)-%EQ(A,2)
        :  A3=%EQ(A,1)-%EQ(A,3)
        :  B2=%EQ(B,1)-%EQ(B,2)
        :  B3=%EQ(B,1)-%EQ(B,3)
        :  D2=%EQ(D,1)-%EQ(D,2)
$CALC      P1=A2*B2  :   P2=A2*B3   :   P3=A3*B2  :   P4=A3*B3
         : P5=A2*D2  :   P6=A3*D2   :   P7=B2*D2  :  P8=B3*D2
$CALC   Q1=A2*B2*D2  :   Q2=A2*B3*D2  :   Q3=A3*B2*D2  :   Q4=A3*B3*D2
$FIT   A2+A3+B2+B3+D2+P1+P2+P3+P4+P5+P6+P7+P8+Q1+Q2+Q3+Q4
$DISPLAY   MERT
$STOP

     25    25    50    50    25    25
     10    40    15    60    25    25
      8    92     2    23    50    50

GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

 CYCLE  DEVIANCE       DF
   3     .4394E-11      0

Y-VARIATE   Y
ERROR POISSON   LINK LOG

LINEAR PREDICTOR
%GM A2 A3 B2 B3 D2 P1 P2 P3 P4 P5 P6 P7 P8 Q1 Q2 Q3 Q4

          ESTIMATE        S.E.         PARAMETER
     1     3.233        .6105E-01     %GM
     2     .2724E-01    .7998E-01     A2
     3     .1902        .1019         A3
     4     .1567        .9950E-01     B2
     5   -.2174         .7469E-01     B3
     6   -.4254         .6105E-01     D2
     7     .3526        .1220         P1
     8   -.2038         .1023         P2
     9   -.9713         .1779         P3
    10     .6523        .1155         P4
    11     .3669E-01    .7998E-01     P5
    12     .3887        .1019         P6
    13     .2127        .9950E-01     P7
    14   -.4254         .7469E-01     P8
    15   -.1835E-01     .1220         Q1
    16     .3669E-01    .1023         Q2
    17   -.1944         .1779         Q3
    18     .3887        .1155         Q4
    SCALE PARAMETER TAKEN AS      1.000

    UNIT   OBSERVED      FITTED        RESIDUAL
     1          25        25.00          .1990E-11
     2          25        25.00          .1918E-11
     3          50        50.00          .2110E-11
     4          50        50.00          .2713E-11
     5          25        25.00          .1421E-11
     6          25        25.00          .1705E-11
     7          10        10.00          .4494E-13
     8          40        40.00          .1887E-11
     9          15        15.00          .9907E-12
    10          60        60.00          .2202E-11
    11          25        25.00          .1137E-11
    12          25        25.00          .1279E-11
    13           8        8.000        -.3135E-11
    14          92        92.00          .3271E-11
    15           2        2.000        -.4994E-07
    16          23        23.00          .8860E-12
    17          50        50.00          .1306E-11
    18          50        50.00          .1708E-11
```

Appendix 3

GLIM program and the parameters of the model after purging the two factor (AB) interaction.

```
$UNITS  18
$DATA   Y
$FACTOR   A 3   B 3   D 2
$CALC   A=%GL(3,6)  :   B=%GL(3,2)  :   D=%GL(2,1)
$DINPUT 1
$LOOK   Y
$YVAR  Y
$ERROR   P
$CALC    A2=%EQ(A,1)-%EQ(A,2)
       : A3=%EQ(A,1)-%EQ(A,3)
       : B2=%EQ(B,1)-%EQ(B,2)
       : B3=%EQ(B,1)-%EQ(B,3)
       : D2=%EQ(D,1)-%EQ(D,2)
       : P5=A2*D2 : P6=A3*D2 : P7=B2*D2 : P8=B3*D2
$CALC   Q1=A2*B2*D2 : Q2=A2*B3*D2 : Q3=A3*B2*D2 : Q4=A3*B3*D2
$FIT  A2+A3+B2+B3+D2+P5+P6+P7+P8+Q1+Q2+Q3+Q4
$DISPLAY MAR
$STOP

   29.6560   29.6560   26.9251   26.9251   39.1236   39.1236
   11.6009   46.4037   10.5411   42.1644   30.6373   30.6373
    5.8140   66.8605    5.2770   60.6860   26.0417   26.0417

 GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

            SCALED
  CYCLE     DEVIANCE       DF
    2       .3784E-04       4

 Y-VARIATE   Y
 ERROR POISSON   LINK LOG

 LINEAR PREDICTOR
 %GM A2 A3 B2 B3 D2 P5 P6 P7 P8 Q1 Q2 Q3 Q4

        ESTIMATE        S.E.          PARAMETER
    1    3.232        .5313E-01       %GM
    2    .2732E-01    .7168E-01       A2
    3    .1902        .7915E-01       A3
    4    .1568        .7541E-01       B2
    5   -.2172        .6602E-01       B3
    6   -.4253        .5116E-01       D2
    7    .3678E-01    .6854E-01       P5
    8    .3885        .7391E-01       P6
    9    .2126        .7101E-01       P7
   10   -.4253        .6638E-01       P8
   11   -.1858E-01    .9021E-01       Q1
   12    .3678E-01    .9110E-01       Q2
   13   -.1941        .9116E-01       Q3
   14    .3885        .9687E-01       Q4
  SCALE PARAMETER TAKEN AS            1.000

  UNIT   OBSERVED       FITTED        RESIDUAL
    1        30          29.65          .1575E-02
    2        30          29.65          .1575E-02
    3        27          26.92          .4664E-03
    4        27          26.92          .4664E-03
    5        39          39.13         -.1758E-02
    6        39          39.13         -.1758E-02
    7        12          11.61         -.1740E-02
    8        46          46.41         -.8696E-03
    9        11          10.54          .1480E-02
   10        42          42.16          .7411E-03
   11        31          30.64          .2008E-03
   12        31          30.64          .2008E-03
   13         6           5.817        -.1107E-02
   14        67          66.86         -.3244E-03
   15         5           5.284        -.3155E-02
   16        61          60.69         -.9286E-03
   17        26          26.03          .1937E-02
   18        26          26.03          .1937E-02
```

Appendix 4

Fitting a saturated log-linear model on the data of civilian labor force classified by age and years of reporting, i.e. 1969, 1971 and 1973.

```
$UNITS  30
$DATA Y
$FACTOR  A 5  B 3  D 2
$CALC   A=%GL(5,6) :  B=%GL(3,2) :  D=%GL(2,1)
$DINPUT 1
$LOOK Y
$YVAR Y
$ERROR P
$CALC   A2=%EQ(A,1)-%EQ(A,2)
       :A3=%EQ(A,1)-%EQ(A,3)
       :A4=%EQ(A,1)-%EQ(A,4)
       :A5=%EQ(A,1)-%EQ(A,5)
       :B2=%EQ(B,1)-%EQ(B,2)
       :B3=%EQ(B,1)-%EQ(B,3)
       :D2=%EQ(D,1)-%EQ(D,2)
$CALC  P1=A2*B2 : P2=A2*B3 : P3=A3*B2 : P4=A3*B3 : P5=A4*B2
     : P6=A4*B3 : P7=A5*B2 : P8=A5*B3
$CALC  P9=A2*D2 : P10=A3*D2 : P11=A4*D2 : P12=A5*D2 : P13=B2*D2
     : P14=B3*D2
$CALC  Q1=A2*B2*D2 : Q2=A2*B3*D2 : Q3=A3*B2*D2 : Q4=A3*B3*D2
     : Q5=A4*B2*D2 : Q6=A4*B3*D2 : Q7=A5*B2*D2 : Q8=A5*B3*D2
$FIT  A2+A3+A4+A5+B2+B3+D2+P1+P2+P3+P4+P5+P6+P7+P8+P9+P10+P11+P12
     +P13+P14+Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8
$DISPLAY  MERT
$STOP

    668    5122     986    4949     907    5247
    784   18581    1522   18011    1223   19103
    413   19155     751   17672     547   16112
    310   14250     534   13607     399   12421
     58    2508      81    2117      45    1949

 GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

 CYCLE  DEVIANCE       DF
   2     .1475E-08      0

Y-VARIATE  Y
ERROR POISSON  LINK LOG

LINEAR PREDICTOR
%GM A2 A3 A4 A5 B2 B3 D2 P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12 P13 P14 Q1 Q
Q3 Q4 Q5 Q6 Q7 Q8

       ESTIMATE         S.E.        PARAMETER
   1    7.551        .9061E-02     %GM
   2   -.8805        .1143E-01     A2
   3   -.4953        .1334E-01     A3
   4   -.2019        .1458E-01     A4
   5    1.664        .3112E-01     A5
   6   -.1271        .1193E-01     B2
   7    .3760E-01    .1342E-01     B3
   8   -1.516        .9061E-02     D2
   9    .4874E-02    .1505E-01     P1
  10    .8968E-01    .1649E-01     P2
  11    .2732E-01    .1754E-01     P3
  12   -.1263E-01    .1925E-01     P4
  13    .1958E-01    .1924E-01     P5
  14   -.7002E-02    .2097E-01     P6
  15    .1203E-01    .4086E-01     P7
  16   -.1585        .4694E-01     P8
  17   -.1182        .1143E-01     P9
  18    .2140        .1334E-01     P10
  19    .2350        .1458E-01     P11
  20    .2840        .3112E-01     P12
  21   -.1413        .1193E-01     P13
  22   -.6380E-02    .1342E-01     P14
  23    .2069E-01    .1505E-01     Q1
  24    .1685E-01    .1649E-01     Q2
  25    .9189E-02    .1754E-01     Q3
  26    .3186E-01    .1925E-01     Q4
  27   -.9619E-02    .1924E-01     Q5
  28    .2521E-01    .2097E-01     Q6
  29    .2677E-01    .4086E-01     Q7
  30   -.9083E-01    .4694E-01     Q8
 SCALE PARAMETER TAKEN AS    1.000
```

Appendix 5

Table 5.a. Log-linear parameters of "age-time interaction" U_{xi}^{AB}

Age	Years (States) 1969	1971	1973
14-19	-.0246	-.0638	.0884
20-34	-.0946	.0049	.0897
35-49	-.0147	.0273	-.0126
50-64	-.0126	.0196	-.0070
65+	.1465	.0120	-.1585

Source: Appendix 4.

Table 5.b. "Age-time interaction" parameters w_{xi}^{AB} of the multiplicative model.

Age	Years (States) 1969	1971	1973
14-19	.9757	.9382	1.0924
20-34	.9097	1.0049	1.0938
35-49	.9854	1.0277	.9875
50-64	.9875	1.0198	.9930
65+	1.1578	1.0121	.8534

Source: Table 5.a.

Appendix 6

Matrix M - crude and standardized rates [a]

		States	
Standard used $V^s(x)$	1969	1971	1973
1969	.0361	.0633	.0513
1971	.0368	.0643	.0522
1973	.0380	.0663	.0539

[a] Obtained by using the age-compositions as a standard for the prevalence rates in each year.
Source: Table 3.