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.
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 ADJUSTMENT *

M. Rafiq Shah, Ph.D.

Interuniversity Programme in Demography
Vrije Universiteit Brussel

Abstract

Clogg (1978) has developed a method for the adjustment 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 calculated 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 occurrence/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 occurring in the i-th state that fall in the k-th class of D and the x-th level of A is the $x_i$-specific rate of the k-th class of D:

$$r_{ik}(x) = \frac{E_{ik}(x)}{\sum_{k} E_{ik}(x)} = \frac{E_{ik}(x)}{E_{ik}(x)}$$

(1)

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

$$r_{ik}(.) = \frac{\sum_{x} E_{ik}(x)}{\sum_{x} E_{ik}(x)} = \frac{E_{ik}(.)}{E_{ik}(.)}$$

(2)

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

$$r_{ik}(.) = \sum_{x} r_{ik}(x) \cdot V_{i}^A(x)$$

(3)

where $V_{i}^A(x) = E_{ik}(x)/E_{ik}(.)$.

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

$$E_{ik}(x) = w_{i}w_{k}^{A}w_{i}^{B}w_{k}^{D}w_{i}^{AB}w_{i}^{AD}w_{i}^{BD}w_{i}^{ABD}.$$  

(4)
The usual constraints on the parameters are:

\[ \Pi W_k = \Pi \Pi = \Pi = \Pi \Pi = \ldots = \Pi \Pi = \ldots = 1 \]

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 \( E_k(x) \) is equal to the observed value. The parameters of the model are: the overall effect (\( w \)), main effects (\( w_i^A \), \( w_i^B \), \( w_k^D \)), first order interaction effects (\( w_{x_i}^{AB} \), \( w_{x_k}^{AD} \), \( w_{ik}^{BD} \)) and the second order interaction effect (\( w_{x_k}^{ABD} \)).

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

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

\[
\log E_k(x) = U + U + U + U + U + U + U + U + U + U + U
\]

(5)

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

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

\[
\sum U_i^A = \sum U_i^B = \sum U_i^D = \sum U_i^{AB} = \ldots = \sum U_i^{ABD} = \ldots = 0
\]

An immediate advantage of these models is that one can easily express the \( x_i \)-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:

\[
\Gamma_{ik}(x) = \frac{w_i^A w_k^B w_{x_i}^D w_{x_k}^{AB} w_{x_i}^{AD} w_{ik}^{BD} w_{x_k}^{ABD}}{\sum w_i^A w_k^B w_{x_i}^D w_{x_k}^{AB} w_{x_i}^{AD} w_{ik}^{BD} w_{x_k}^{ABD}}
\]
The crude rate (2) is:

$$r_{ik}(x) = \frac{W_{ik}^D \cdot W_{ik}^B \cdot \sum_x W_{ixk}^A \cdot W_{ixk}^B \cdot W_{ixk}^A \cdot W_{ixk}^B}{\sum_k W_{ik}^D \cdot W_{ik}^B \cdot \sum_x W_{ixk}^A \cdot W_{ixk}^B \cdot W_{ixk}^A \cdot W_{ixk}^B}$$

(7)

and the weights $V_1^A(x)$ as in (3) are:

$$V_1^A(x) = \frac{W_{ix1}^A \cdot W_{ix1}^B \cdot \sum_k W_{ik}^D \cdot W_{ik}^B \cdot \sum_x W_{ixk}^A \cdot W_{ixk}^B \cdot W_{ixk}^A \cdot W_{ixk}^B}{\sum_k W_{ik}^D \cdot W_{ik}^B \cdot \sum_x W_{ixk}^A \cdot W_{ixk}^B \cdot \sum_x W_{ixk}^A \cdot W_{ixk}^B \cdot W_{ixk}^A \cdot W_{ixk}^B}$$

(8)

**Clogg's Method of Adjustment**

Due to the fact that the two factor interaction $w_{ix1}^{AB}$ 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 adjustment 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 adjustment based on purging and rescaling is explained in the following subsections.
1. Purging of Undesired Interaction

The interaction $w_{XAB}$ 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:

$$E_{ik}^*(x) = E_{ik}(x) / w_{XAB} = w_{X}^A w_{i}^B w_{D}^D w_{xk}^{AD} w_{ik}^{AD} w_{xik}^{ABD}.$$ (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

$$E_{ik}^{**}(x) = w'.w_{X}^A w_{i}^B w_{D}^D w_{xk}^{AD} w_{ik}^{AD} w_{xik}^{ABD}.$$ (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 $E_{++}(+)/E_{++}^*(+)$. The rescaled frequencies $E_{ik}^*(x)$ so obtained are:

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

$$= w'.w_{X}^A(w_{i}^B).w_{D}^D w_{xk}^{AD} w_{ik}^{AD} w_{xik}^{ABD}.$$ (11)
where
\[
(w_{i_k}^P)' = w/w' \prod_{x \in x_k} (E_{i_k}^{-}(x)/E_{i_k}^{*}(x))^{1/(R_k)} \\
= [E_{i_+}^{-}(+)/E_{i_+}^{*}(+)] [E_{i_k}^{-}(+)/E_{i_k}^{*}(+)] w_{i_k}^P.
\]

3. Adjustment

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

\[
r_{i_k}^S(\cdot) = \frac{\sum_x E_{i_k}^{-}(x)}{\sum_x E_{i_k}^{-}(x)} = E_{i_k}^{-}(+)/E_{i_+}^{*}(+) = E_{i_k}^{*}(+)/E_{i_+}^{*}(+)
\]

Comments and Comparisons

The method of adjustment presented by Clogg (1978) displays a useful application of the saturated models in demographic analysis and is considered to be a breakthrough 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 accommodating 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 (\(w_{x1}^{AB}\)) 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 \(x1\)-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 \(x1\)-specific rates \(r_{ik}(x)\) as follows.

\[
r_{ik}^{s}(.) = \sum_{x} r_{ik}(x) \cdot V_{i}(x)
\]

(13)

where

\[
V_{i}(x) = \frac{E_{i+}'(x)}{E_{i+}'(+)}
\]

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 \(x1\)-specific rates \(r_{ik}(x)\) are not affected by the method of
adjustment as stated earlier, and the method of adjustment produces changes in the confounding factors, i.e. the weights $V^\Lambda_i(x)$, 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 is defined as

$$\text{DSR}_{ik} = \sum_x r_{ik}(x) V^S(x)$$  \hspace{1cm} (14)$$

where

$$V^S(x) = \frac{E_{++}^S(x)}{E_{++}(+), \text{ such that } \sum_x V^S(x) = 1.$$  

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

It may be noted that while $V^\Lambda_i(x)$ 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 $V^\Lambda_i(x)$ 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.

<table>
<thead>
<tr>
<th>State factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome factor/Confounding factor</td>
<td>1</td>
<td>2</td>
<td>Total</td>
<td>1</td>
<td>2</td>
<td>Total</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>50</td>
<td>50</td>
<td>25</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>40</td>
<td>50</td>
<td>15</td>
<td>60</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>92</td>
<td>100</td>
<td>2</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>157</td>
<td>200</td>
<td>67</td>
<td>133</td>
<td>200</td>
</tr>
<tr>
<td>Crude rates</td>
<td>.215</td>
<td>.335</td>
<td>.500</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For $k=1$ the rate $r_{11}(x) = E_{11}(x)/E_{1+}(x)$, 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 $x_i$-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 $x_i$-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

<table>
<thead>
<tr>
<th>State factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome factor/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11.6009</td>
<td>46.4037</td>
<td>10.5411</td>
</tr>
<tr>
<td>3</td>
<td>5.8140</td>
<td>66.8605</td>
<td>5.2770</td>
</tr>
<tr>
<td>Total</td>
<td>47.0709</td>
<td>42.7432</td>
<td>95.8026</td>
</tr>
</tbody>
</table>

$E_{ik}^*(x) = E_{ik}(x)/W_{xi}^{AB}$, 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 $V_1^A(x)$ vary according to the pattern of the $x_1$-specific rates. The proportions are identical only in the states where the $x_1$-specific rates are identical. For instance, since the rates $r_{x1}(x)$ in states 1 and 2 are identical, the weights $V_1(x)$ associated with states 1 and 2 are also identical.

Table 2.b. Frequencies $E_1^*(x)$ and proportions $V_1(x)$ purged of AB interactions by state.

<table>
<thead>
<tr>
<th>State factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1+*(x)</td>
<td>59.3120</td>
<td>0.3122</td>
<td>53.8502</td>
</tr>
<tr>
<td>V1(x)</td>
<td>58.0046</td>
<td>0.3053</td>
<td>52.7055</td>
</tr>
<tr>
<td>E2+*(x)</td>
<td>72.6745</td>
<td>0.3825</td>
<td>65.9630</td>
</tr>
<tr>
<td>V2(x)</td>
<td>189.9911</td>
<td>1.0000</td>
<td>172.5187</td>
</tr>
</tbody>
</table>

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 $x_1$-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 (6.7%, -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 adjustment 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
(Ei(x) with i = year, k = employment status, x = age)

<table>
<thead>
<tr>
<th>Age</th>
<th>1969</th>
<th>1971</th>
<th>1973</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>unempl.</td>
<td>empl.</td>
<td>total</td>
</tr>
<tr>
<td>14-19</td>
<td>668</td>
<td>5,122</td>
<td>5,790</td>
</tr>
<tr>
<td>20-34</td>
<td>764</td>
<td>18,581</td>
<td>19,365</td>
</tr>
<tr>
<td>35-49</td>
<td>413</td>
<td>19,155</td>
<td>19,568</td>
</tr>
<tr>
<td>50-64</td>
<td>310</td>
<td>14,250</td>
<td>14,560</td>
</tr>
<tr>
<td>65+</td>
<td>58</td>
<td>2,508</td>
<td>2,566</td>
</tr>
</tbody>
</table>

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 Vj(x), when the
adjusted rates $r_{ik}^*(.)$ are expressed as a weighted average of the $x_1$-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.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounding factor</td>
<td>Frequencies $E_{i+}^*(x)$</td>
<td>Proportions $V_i(x)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-19</td>
<td>5934.2011</td>
<td>6325.9433</td>
<td>5632.9519</td>
<td>0.0926</td>
<td>0.1059</td>
<td>0.1000</td>
</tr>
<tr>
<td>20-34</td>
<td>21284.8980</td>
<td>19437.7550</td>
<td>18582.9200</td>
<td>0.3322</td>
<td>0.3254</td>
<td>0.3299</td>
</tr>
<tr>
<td>35-49</td>
<td>19857.9260</td>
<td>17926.4380</td>
<td>16869.8730</td>
<td>0.3099</td>
<td>0.3001</td>
<td>0.2995</td>
</tr>
<tr>
<td>50-64</td>
<td>14775.7260</td>
<td>13866.4440</td>
<td>12910.3730</td>
<td>0.2306</td>
<td>0.2322</td>
<td>0.2292</td>
</tr>
<tr>
<td>65+</td>
<td>2216.4637</td>
<td>2171.7222</td>
<td>2336.5362</td>
<td>0.0346</td>
<td>0.0364</td>
<td>0.0415</td>
</tr>
<tr>
<td>All ages</td>
<td>64069.2156</td>
<td>59728.3035</td>
<td>56332.6541</td>
<td>0.9999</td>
<td>1.0000</td>
<td>1.0001</td>
</tr>
</tbody>
</table>

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 \( r_{11}(x) \) 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 occurred 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 $V^S(x)$

<table>
<thead>
<tr>
<th>Age</th>
<th>1969</th>
<th>1971</th>
<th>1973</th>
<th>Standard $V^S(x)$ a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-19</td>
<td>.0936</td>
<td>.0985</td>
<td>.1062</td>
<td>.0961</td>
</tr>
<tr>
<td>20-34</td>
<td>.3131</td>
<td>.3243</td>
<td>.3507</td>
<td>.3187</td>
</tr>
<tr>
<td>35-49</td>
<td>.3164</td>
<td>.3059</td>
<td>.2857</td>
<td>.3112</td>
</tr>
<tr>
<td>50-64</td>
<td>.2354</td>
<td>.2348</td>
<td>.2212</td>
<td>.2351</td>
</tr>
<tr>
<td>65+</td>
<td>.0145</td>
<td>.0365</td>
<td>.0344</td>
<td>.0390</td>
</tr>
<tr>
<td>All ages</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0001</td>
</tr>
</tbody>
</table>

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 $E_{i*k}^*(x)$, we also purged out the three factor interaction $w_{i*k}^{ABD}$. Dividing $E_{i*k}^*(x)$ by $w_{i*k}^{ABD}$ and denoting the purged frequencies so obtained by $E_{i*k}^{**}(x)$, we have

$$E_{i*k}^{**}(x) = w_{i,k}^A w_{i,k}^B w_{i,k}^D w_{i,k}^{AD} w_{i,k}^{BD}$$  \hspace{1cm} (15)

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

$$V_i^*(x) = \frac{w_{i,k}^A \sum_k w_{i,k}^D \cdot w_{i,k}^{BD} \cdot w_{i,k}^{AD}}{\sum_x w_{i,k}^A \sum_k w_{i,k}^D \cdot w_{i,k}^{BD} \cdot w_{i,k}^{AD}}.$$  \hspace{1cm} (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 $r_{i*k}^*(x)$,

$$r_{i*k}^*(x) = \frac{w_{i,k}^D w_{i,k}^{BD} w_{i,k}^{AD}}{\sum_x w_{i,k}^D \cdot w_{i,k}^{BD} \cdot w_{i,k}^{AD}} = E_{i*k}^{**}(x)/E_{i*}^{**}(x).$$  \hspace{1cm} (17)

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

$$r_{i*k}^{**}(x) = \sum_x r_{i*k}^*(x) V_i^*(x).$$  \hspace{1cm} (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 $r_{i*k}(x)$ 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 empirical 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.

<table>
<thead>
<tr>
<th>State factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E_{1+}^{**}(x)$</td>
<td>$v_{1+}^*(x)$</td>
<td>$E_{2+}^{**}(x)$</td>
</tr>
<tr>
<td>1</td>
<td>67.31</td>
<td>0.33655</td>
<td>76.31</td>
</tr>
<tr>
<td>2</td>
<td>63.72</td>
<td>0.31860</td>
<td>63.72</td>
</tr>
<tr>
<td>3</td>
<td>68.97</td>
<td>0.34485</td>
<td>68.97</td>
</tr>
<tr>
<td>All categories</td>
<td>200.00</td>
<td>1.00000</td>
<td>200.00</td>
</tr>
</tbody>
</table>

Source: Table 2.a and Appendix 1.
References


Appendix 1

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

\[
\log E_{jk}(\chi) = U + U^A_x + U^B_x + U^D_x + U^{AB}_{x_1} + U^{AD}_{x_1} + U^{BD}_{x_1} + U^{ABD}_{x_1}
\]

Overall Mean \( U = 3.233 \)

<table>
<thead>
<tr>
<th>U_1^A = .2174</th>
<th>U_1^B = -.0607</th>
<th>U_1^C = 4.254</th>
</tr>
</thead>
<tbody>
<tr>
<td>U_2^A = -.0272</td>
<td>U_2^B = -.1567</td>
<td>U_2^C = -4.254</td>
</tr>
<tr>
<td>U_3^A = -.1902</td>
<td>U_3^B = .2174</td>
<td></td>
</tr>
</tbody>
</table>

Two Factor Interactions

<table>
<thead>
<tr>
<th>AB</th>
<th>AD</th>
<th>BD</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.1702</td>
<td>.6187</td>
<td>-.4485</td>
</tr>
<tr>
<td>-.1468</td>
<td>.3526</td>
<td>-.2038</td>
</tr>
<tr>
<td>.3190</td>
<td>-.9713</td>
<td>.6523</td>
</tr>
<tr>
<td>.4254</td>
<td>.4254</td>
<td>-.2127</td>
</tr>
<tr>
<td>-.0367</td>
<td>.0367</td>
<td>-.2127</td>
</tr>
<tr>
<td>-.2127</td>
<td>.2127</td>
<td>.2127</td>
</tr>
</tbody>
</table>

Three Factor Interactions (ABD)

<table>
<thead>
<tr>
<th>AB</th>
<th>AD</th>
<th>BD</th>
</tr>
</thead>
<tbody>
<tr>
<td>.2126</td>
<td>-.2126</td>
<td>.2126</td>
</tr>
<tr>
<td>-.0184</td>
<td>.0184</td>
<td>.0184</td>
</tr>
<tr>
<td>-.1943</td>
<td>.1943</td>
<td>.1944</td>
</tr>
<tr>
<td>-.4254</td>
<td>.4254</td>
<td>.3887</td>
</tr>
<tr>
<td>.3887</td>
<td>.3887</td>
<td>.3667</td>
</tr>
</tbody>
</table>

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.

$\text{UNITS} 18$
$\text{DATA} Y$
$\text{FACTOR} A 3 \ B 3 \ D 2$
$\text{CALC} A=\%GL(3,6); \ B=\%GL(3,2); \ D=\%GL(2,1)$
$\text{INPUT}$
$\text{ERROR} P$
$\text{CALC} A2=\%EG(A,1)-\%EG(A,2); \ A3=\%EG(A,1)-\%EG(A,3); \ A4=\%EG(A,1)-\%EG(A,4); \ D2=\%EG(D,1)-\%EG(D,2)$
$\text{CALC} P1=A2*B2; \ P2=A2*B3; \ P3=A3*B2; \ P4=A3*B3$
$\text{FIT} A2+A3+B2+B3+D2+P1+P2+P3+P4+P5+P6+P7+P8+Q1+Q2+Q3+Q4$
$\text{DISPLAY} \text{MERT}$
$\text{STOP}$

GLIM 3.11 (C) 1977 ROYAL STATISTICAL SOCIETY, LONDON

<table>
<thead>
<tr>
<th>CYCLE</th>
<th>DEVIANCE</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.4394E-11</td>
<td>0</td>
</tr>
</tbody>
</table>

Y-VARIATE Y
ERROR POISSON LINK LOG

LINEAR PREDICTOR

<table>
<thead>
<tr>
<th>ESTIMATE</th>
<th>S.E.</th>
<th>PARAMETER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.233</td>
<td>3.233</td>
</tr>
<tr>
<td>2</td>
<td>.2724E-01</td>
<td>.799E-01</td>
</tr>
<tr>
<td>3</td>
<td>1.902</td>
<td>1.902</td>
</tr>
<tr>
<td>4</td>
<td>1567</td>
<td>995OE-01</td>
</tr>
<tr>
<td>5</td>
<td>.2174</td>
<td>7469E-01</td>
</tr>
<tr>
<td>6</td>
<td>.4254</td>
<td>6105E-01</td>
</tr>
<tr>
<td>7</td>
<td>3526</td>
<td>1220</td>
</tr>
<tr>
<td>8</td>
<td>.2038</td>
<td>1023</td>
</tr>
<tr>
<td>9</td>
<td>.9713</td>
<td>1779</td>
</tr>
<tr>
<td>10</td>
<td>.3023</td>
<td>1155</td>
</tr>
<tr>
<td>11</td>
<td>.369E-01</td>
<td>799E-01</td>
</tr>
<tr>
<td>12</td>
<td>.3887</td>
<td>1019</td>
</tr>
<tr>
<td>13</td>
<td>.2127</td>
<td>995OE-01</td>
</tr>
<tr>
<td>14</td>
<td>.4254</td>
<td>7469E-01</td>
</tr>
<tr>
<td>15</td>
<td>.1839E-01</td>
<td>1220</td>
</tr>
<tr>
<td>16</td>
<td>.369E-01</td>
<td>1023</td>
</tr>
<tr>
<td>17</td>
<td>.1944</td>
<td>1779</td>
</tr>
<tr>
<td>18</td>
<td>.3887</td>
<td>1155</td>
</tr>
</tbody>
</table>

SCALE PARAMETER TAKEN AS 1.000

<table>
<thead>
<tr>
<th>UNIT</th>
<th>OBSERVED</th>
<th>FITTED</th>
<th>RESIDUAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>25.00</td>
<td>.1990E-11</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>25.00</td>
<td>.1918E-11</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>50.00</td>
<td>.2110E-11</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>30.00</td>
<td>.2713E-11</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>25.00</td>
<td>.1421E-11</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>25.00</td>
<td>.1705E-11</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>10.00</td>
<td>.4494E-13</td>
</tr>
<tr>
<td>8</td>
<td>24</td>
<td>40.00</td>
<td>.1887E-11</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>15.00</td>
<td>.9907E-12</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>60.00</td>
<td>.2202E-11</td>
</tr>
<tr>
<td>11</td>
<td>25</td>
<td>25.00</td>
<td>.1137E-11</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
<td>25.00</td>
<td>.1279E-11</td>
</tr>
<tr>
<td>13</td>
<td>38</td>
<td>38.00</td>
<td>.3136E-11</td>
</tr>
<tr>
<td>14</td>
<td>92</td>
<td>92.00</td>
<td>.3271E-11</td>
</tr>
<tr>
<td>15</td>
<td>22</td>
<td>20.00</td>
<td>.4994E-07</td>
</tr>
<tr>
<td>16</td>
<td>23</td>
<td>23.00</td>
<td>.8660E-12</td>
</tr>
<tr>
<td>17</td>
<td>50</td>
<td>50.00</td>
<td>.130AE-11</td>
</tr>
<tr>
<td>18</td>
<td>50</td>
<td>50.00</td>
<td>.1708E-11</td>
</tr>
</tbody>
</table>
Appendix 3

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

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

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
ZGM A2 A3 B2 B3 D2 P5 P6 P7 P8 Q1 Q2 Q3 Q4

<table>
<thead>
<tr>
<th>ESTIMATE</th>
<th>S.E.</th>
<th>PARAMETER</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.232</td>
<td>5313E-01</td>
<td>%GM</td>
</tr>
<tr>
<td>2.732E-01</td>
<td>7165E-01</td>
<td>A2</td>
</tr>
<tr>
<td>1.92E-01</td>
<td>7915E-01</td>
<td>A3</td>
</tr>
<tr>
<td>1568</td>
<td>7541E-01</td>
<td>B2</td>
</tr>
<tr>
<td>-2172</td>
<td>6602E-01</td>
<td>B3</td>
</tr>
<tr>
<td>-4283</td>
<td>6684E-01</td>
<td>D2</td>
</tr>
<tr>
<td>3885</td>
<td>7391E-01</td>
<td>P5</td>
</tr>
<tr>
<td>7101E-01</td>
<td>P7</td>
<td></td>
</tr>
<tr>
<td>6688E-01</td>
<td>P8</td>
<td></td>
</tr>
<tr>
<td>1856E-01</td>
<td>9021E-01</td>
<td>Q1</td>
</tr>
<tr>
<td>3678E-01</td>
<td>9110E-01</td>
<td>Q2</td>
</tr>
<tr>
<td>-1941</td>
<td>9116E-01</td>
<td>Q3</td>
</tr>
<tr>
<td>3885</td>
<td>9687E-01</td>
<td>Q4</td>
</tr>
</tbody>
</table>

SCALE PARAMETER TAKEN AS 1.000

UNIT OBSERVED FITTED RESIDUAL
1 30 29.65 1975E-02
2 30 29.65 1975E-02
3 27 26.92 4664E-03
4 27 26.92 4664E-03
5 39 39.13 -1795E-02
6 39 39.13 -1795E-02
7 12 11.61 1740E-02
8 12 11.61 1740E-02
9 11 10.54 8696E-03
10 42 12.16 7410E-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 5 5.284 3155E-02
17 61 60.69 9286E-03
18 26 28.03 1937E-02
19 26 28.03 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.

$\text{UNITS 30}$
$\text{DATA Y}$
$\text{FACTOR A 5 B 3 D 2}$
$\text{CALC A=}XGL(5,6) : B=}XGL(3,2) : D=}XGL(2,1)$
$\text{INPUT 1}$
$\text{LOOK Y}$
$\text{YVAR Y}$
$\text{ERROR P}$

$\text{CALC A2=}XEQ(A,1)-XEQ(A,2) : A3=}XEQ(A,1)-XEQ(A,3)$
$\text{CALC A4=}XEQ(A,1)-XEQ(A,4) : A5=}XEQ(A,1)-XEQ(A,5)$
$\text{CALC B2=}XEQ(B,1)-XEQ(B,2) : B3=}XEQ(B,1)-XEQ(B,3)$
$\text{CALC B4=}XEQ(B,1)-XEQ(B,4) : D2=}XEQ(D,1)-XEQ(D,2)$


$\text{CALC P5=}A4*B2 : P6=}A4*B3 : P7=}A5*B2 : P8=}A5*B3$.


$\text{FIT A2+A3+A4+A5+B2+B3+D2+P1+P2+P3+P4+P5+P6+P7+P8+P9+P10+P11+P12}$

$\text{DISPLAY MERT}$

$\text{STOP}$

<table>
<thead>
<tr>
<th>ESTIMATE</th>
<th>S.E.</th>
<th>PARAMETER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 7.531</td>
<td>9061E-02</td>
<td>%GM</td>
</tr>
<tr>
<td>2 -4825</td>
<td>1143E-01</td>
<td>A2</td>
</tr>
<tr>
<td>3 -4953</td>
<td>1334E-01</td>
<td>A3</td>
</tr>
<tr>
<td>4 -2019</td>
<td>1458E-01</td>
<td>A4</td>
</tr>
<tr>
<td>5 1.664</td>
<td>3112E-01</td>
<td>A5</td>
</tr>
<tr>
<td>6 -1271</td>
<td>1193E-01</td>
<td>B2</td>
</tr>
<tr>
<td>7 3786E-01</td>
<td>1342E-01</td>
<td>B3</td>
</tr>
<tr>
<td>8 -1.158</td>
<td>9061E-02</td>
<td>D2</td>
</tr>
<tr>
<td>9 4874E-02</td>
<td>1305E-01</td>
<td>P1</td>
</tr>
<tr>
<td>10 8966E-01</td>
<td>1649E-01</td>
<td>P2</td>
</tr>
<tr>
<td>11 2735E-01</td>
<td>1754E-01</td>
<td>P3</td>
</tr>
<tr>
<td>12 -1263E-01</td>
<td>1924E-01</td>
<td>P4</td>
</tr>
<tr>
<td>13 1995E-01</td>
<td>1924E-01</td>
<td>P5</td>
</tr>
<tr>
<td>14 -7022E-02</td>
<td>2097E-01</td>
<td>P6</td>
</tr>
<tr>
<td>15 1203E-01</td>
<td>4086E-01</td>
<td>P7</td>
</tr>
<tr>
<td>16 -6380E-02</td>
<td>4694E-01</td>
<td>P8</td>
</tr>
<tr>
<td>17 -1.182</td>
<td>1134E-01</td>
<td>P9</td>
</tr>
<tr>
<td>18 2140</td>
<td>1334E-01</td>
<td>P10</td>
</tr>
<tr>
<td>19 2350</td>
<td>1458E-01</td>
<td>P11</td>
</tr>
<tr>
<td>20 -2840</td>
<td>3112E-01</td>
<td>P12</td>
</tr>
<tr>
<td>21 1.413</td>
<td>1193E-01</td>
<td>P13</td>
</tr>
<tr>
<td>22 -4825</td>
<td>1143E-01</td>
<td>P14</td>
</tr>
<tr>
<td>23 2069E-01</td>
<td>1905E-01</td>
<td>G1</td>
</tr>
<tr>
<td>24 1692E-01</td>
<td>1649E-01</td>
<td>G2</td>
</tr>
<tr>
<td>25 913E-02</td>
<td>1754E-01</td>
<td>G3</td>
</tr>
<tr>
<td>26 3186E-01</td>
<td>1924E-01</td>
<td>G4</td>
</tr>
<tr>
<td>27 -9619E-02</td>
<td>1924E-01</td>
<td>G5</td>
</tr>
<tr>
<td>28 -2521E-01</td>
<td>2097E-01</td>
<td>G6</td>
</tr>
<tr>
<td>29 2677E-01</td>
<td>4086E-01</td>
<td>G7</td>
</tr>
<tr>
<td>30 -9083E-01</td>
<td>4694E-01</td>
<td>G8</td>
</tr>
</tbody>
</table>

$\text{SCALE PARAMETER TAKEN AS} 1.000$
### Appendix 5

#### Table 5.a. Log-linear parameters of "age-time interaction" $U_{x.t}^{AB}$

<table>
<thead>
<tr>
<th>Age</th>
<th>Years (States)</th>
<th>1969</th>
<th>1971</th>
<th>1973</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-19</td>
<td></td>
<td>-.0246</td>
<td>-.0638</td>
<td>.0884</td>
</tr>
<tr>
<td>20-34</td>
<td></td>
<td>-.0946</td>
<td>.0049</td>
<td>.0897</td>
</tr>
<tr>
<td>35-49</td>
<td></td>
<td>-.0147</td>
<td>.0273</td>
<td>-.0126</td>
</tr>
<tr>
<td>50-64</td>
<td></td>
<td>-.0126</td>
<td>.0196</td>
<td>-.0070</td>
</tr>
<tr>
<td>65+</td>
<td></td>
<td>.1465</td>
<td>.0120</td>
<td>-.1585</td>
</tr>
</tbody>
</table>

Source: Appendix 4.

#### Table 5.b. "Age-time interaction" parameters $w_{x.t}^{AB}$ of the multiplicative model.

<table>
<thead>
<tr>
<th>Age</th>
<th>Years (States)</th>
<th>1969</th>
<th>1971</th>
<th>1973</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-19</td>
<td></td>
<td>.9757</td>
<td>.9382</td>
<td>1.0924</td>
</tr>
<tr>
<td>20-34</td>
<td></td>
<td>.9097</td>
<td>1.0049</td>
<td>1.0938</td>
</tr>
<tr>
<td>35-49</td>
<td></td>
<td>.9854</td>
<td>1.0277</td>
<td>.9875</td>
</tr>
<tr>
<td>50-64</td>
<td></td>
<td>.9875</td>
<td>1.0198</td>
<td>.9930</td>
</tr>
<tr>
<td>65+</td>
<td></td>
<td>1.1578</td>
<td>1.0121</td>
<td>.8534</td>
</tr>
</tbody>
</table>

Source: Table 5.a.
### Appendix 6

**Matrix M - crude and standardized rates**

<table>
<thead>
<tr>
<th>Standard used $V^S(x)$</th>
<th>States</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1969</td>
<td>1971</td>
<td>1973</td>
</tr>
<tr>
<td>1969</td>
<td>.0361</td>
<td>.0633</td>
<td>.0513</td>
</tr>
<tr>
<td>1971</td>
<td>.0368</td>
<td>.0643</td>
<td>.0522</td>
</tr>
<tr>
<td>1973</td>
<td>.0380</td>
<td>.0663</td>
<td>.0539</td>
</tr>
</tbody>
</table>

*Obtained by using the age-compositions as a standard for the prevalence rates in each year.*

*Source. Table 3.*
<table>
<thead>
<tr>
<th>Price</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF 230</td>
<td>US$ 6.00</td>
</tr>
<tr>
<td>BF 300</td>
<td>US$ 8.00</td>
</tr>
<tr>
<td>BF 230</td>
<td>US$ 6.00</td>
</tr>
<tr>
<td>BF 350</td>
<td>US$ 9.00</td>
</tr>
<tr>
<td>BF 300</td>
<td>US$ 8.00</td>
</tr>
<tr>
<td>BF 230</td>
<td>US$ 6.00</td>
</tr>
<tr>
<td>BF 230</td>
<td>US$ 6.00</td>
</tr>
<tr>
<td>BF 230</td>
<td>US$ 6.00</td>
</tr>
</tbody>
</table>

**IPD-WORKING PAPERS 1983**

1983-1. R. Lesthaeghe: "A Century of Demographic and Cultural Change in Western Europe: An Exploration of Underlying Dimensions".


1983-4. S. Becker, A. Chowdhury, S. Huffman: "Determinants of Natural Fertility in Matlab, Bangladesh".


1984-3. C. Vanderhoeft: "Life Tables and Relational Models: A Unified Approach, and Applications in Demography".

1984-4. R. Lesthaeghe: "Demographic Renewal and Problems of Social Organization".

1984-5. S. Becker: "A Response Bias in the reporting of Month of Birth in Pregnancy History Surveys".

1985-1. R. Lesthaeghe and F. Eelens: "Social Organization and Reproductive Regimes: Lessons from sub-Saharan Africa and historical Western Europe".


1985-6. Stanley Wijewickrema: "Childlessness in Belgium and Flanders".

1985-7. R. Lesthaeghe: "Value Orientations, Economic Growth and Demographic Trends - Toward a Confrontation?"

1985-8. Fernando Rajulton and Stan Wijewickrema: "An Enquiry into the Two Basic Assumptions of Multistate Increment-Decrement Life Tables".
1986-1. F. Rajulton: "Marriageability and divorceability: a simulation of the unobservables through the conditional gaussian diffusion process."


1986-4. R. Lesthaeghe, G. Kaufmann, D. Meekers: "The Nuptiality Regimes in Sub-Saharan Africa: Data Files NUPFILE 1 and NUPFILE 2".

1986-5. Hwa Young LEE and Fernando RAJULTON: "A Semi-Markovian Multistate Analysis of Sexual Union Formation in Flanders, Belgium".

1986-6. Zeng Yi: "Fertility and Household Dynamics in China".

IPD-WORKING PAPERS 1987

1987-1. C. Vanderhoeft : "Hidden heterogeneity in composite link models : further developments".


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".

A CRITIQUE OF CLOGG'S METHOD OF RATES ADJUSTMENT *

M. Rafiq Shah, Ph.D.

Interuniversity Programme in Demography
Vrije Universiteit Brussel

Abstract

Clogg (1978) has developed a method for the adjustment 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 calculated 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 occurrence/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 occurring in the $i$-th state that fall in the $k$-th class of $D$ and the $x$-th level of $A$ is the $\chi_i$-specific rate of the $k$-th class of $D$:

$$r_{ik}(x) = E_{ik}(x)/\sum_{x}E_{ik}(x) = E_{ik}(x)/E_{ik}(+)$$

(1)

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

$$r_{ik}(.) = \sum_{x}r_{ik}(x) = E_{ik}(+)/E_{ik}(+)$$

(2)

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

$$r_{ik}(.) = \sum_{x}r_{ik}(x).V_{i}^{A}(x)$$

(3)

where $V_{i}^{A}(x) = E_{ik}(x)/E_{ik}(+)$. The model that describes the data fully in a contingency table is the saturated model:

$$E_{ik}(x) = W_{i}^{A}.W_{i}^{B}.W_{i}^{D}.W_{i}^{AB}.W_{i}^{AD}.W_{i}^{BD}.W_{i}^{ABD}.$$ 

(4)
The usual constraints on the parameters are:

\[ \prod_x w_x^A = \prod_x w_x^B = \prod_x w_x^D = \prod_x w_{x1}^{AB} = \ldots = \prod_x w_{xik}^{ABD} = \ldots = 1 \]

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 (\( w_x^A, w_x^B, w_x^D \)), first order interaction effects (\( w_{x1}^{AB}, w_{xk}^{AD}, w_{ik}^{BD} \)) 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, \( w_{x1}^{AB} \) represents 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 differences in D.

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

\[ \log E_{ik}(x) = U + U_x^A + U_k^B + U_k^D + U_{x1}^{AB} + U_{xk}^{AD} + U_{ik}^{BD} + U_{xik}^{ABD} \]  \hspace{1cm} (5)

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

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

\[ \sum_x U_x^A = \sum_x U_x^B = \sum_k U_k^D = \sum_x U_{x1}^{AB} = \ldots = \sum_x U_{xik}^{ABD} = \ldots = 0 \]

An immediate advantage of these models is that one can easily express the \( x_i \)-specific rates \( r_{ik}(x) \), the crude rates \( r_i(x) \) 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:

\[ r_{ik}(x) = \frac{W_x^A W_i^B W_k^D W_{x1}^{AB} W_{xk}^{AD} W_{ik}^{BD} W_{xik}^{ABD}}{\sum_k W_x^A W_i^B W_k^D W_{x1}^{AB} W_{xk}^{AD} W_{ik}^{BD} W_{xik}^{ABD}} \]
Clogg's Method of Adjustment

Due to the fact that the two factor interaction $\omega_{x_1}^{AB}$ 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 adjustment 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 adjustment based on purging and rescaling is explained in the following subsections.
1. **Purging of Undesired Interaction**

The interaction $w_{x\alpha}^{AB}$ 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:

$$E_{ik}^*(x) = E_{ik}(x)/w_{x\alpha}^{AB} = w_A w_i w_B w_k w_{xk} w_{ik} w_{xik} w_{ikxik}^{-1}$$

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

$$E_{ik}^{**}(x) = w'w_A w_i w_B w_k w_d w_{xk} w_{ik} w_{xik}^{-1}$$

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 $E_{++}(+)/E_{++}^*(+)$. The rescaled frequencies $E_{ik}^{'}(x)$ so obtained are:

$$E_{ik}^{'}(x) = E_{ik}^*(x)[E_{++}(+)/E_{++}^*(+)]$$

$$= w'w_A (w_B^{-1}) w_d w_k w_{xk} w_{ik} w_{xik}^{-1}$$

$$= w'A w_i w_B w_k w_{xk} w_{ik} w_{xik}^{-1}$$
where
\[
(w_i^B)^{-1} = w_i^* \prod_{jk} \left( \frac{E_{ik}^+(x)}{E_{ik}^*(x)} \right)^{1/(RK)}
= \left[ \frac{E_{ik}^+(x)}{E_{ik}^*(x)} \right] \left[ \frac{E_{ik}^+(x)}{E_{ik}^*(x)} \right] w_i^B.
\]

3. Adjustment

The adjusted rate \( r_{ik}^{s*}(x) \) 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:

\[
\begin{align*}
r_{ik}^{s*}(x) &= \frac{\sum_x E_{ik}^*(x)}{\sum_{xk} E_{ik}^*(x)} \\
&= \frac{E_{ik}^*(+)}{E_{ik}^*}(+) = \frac{E_{ik}^*(+)}{E_{ik}^*}(+)
\end{align*}
\]

Comments and Comparisons

The method of adjustment presented by Clogg (1978) displays a useful application of the saturated models in demographic analysis and is considered to be a breakthrough 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 accommodating 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 \( w_{x1}^{AB} \) 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 \( r_{ik}^{*}(x) \) with the crude rate \( r_{ik}(x) \) 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 \( r_{ik}^{*}(x) \) could be expressed as a weighted average of the \( xi \)-specific rates \( r_{ik}(x) \) as follows.

\[
\begin{align*}
  r_{ik}^{*}(x) &= \sum_x r_{ik}(x) \cdot V_i(x) \\
  &= \sum_x r_{ik}(x) \cdot \frac{E_{i+}(x)}{E_{++}(x)} \\
  &= \frac{E_{i+}(x)}{E_{++}(x)}
\end{align*}
\]

where

\[
V_i(x) = \frac{E_{i+}(x)}{E_{++}(x)}
\]

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
adjustment as stated earlier, and the method of adjustment 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

\[
\text{DSP}_{ik} = \sum_{x} r_{ik}(x) V^{S}(x) \tag{14}
\]

where

\[
V^{S}(x) = \frac{E_{x}^{S}(x)}{E_{x}^{S}(x)}, \text{ such that } \sum_{x} V^{S}(x) = 1.
\]

A comparison of (13) and (14) reveals that whereas \( r_{ik}(x) \) is common in both \( r_{ik}^{S}(x) \) and \( \text{DSP}_{ik} \), 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 \( r_{ik}^{S}(x) \) and \( \text{DSP}_{ik} \) 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.

<table>
<thead>
<tr>
<th>State factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E_{ik}(x)$</td>
<td></td>
<td></td>
<td>$r_{ik}(x)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcome factor/ Confounding factor</strong></td>
<td>1</td>
<td>2</td>
<td>Total</td>
<td>1</td>
<td>2</td>
<td>Total</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>25</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>40</td>
<td>50</td>
<td>15</td>
<td>60</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>92</td>
<td>100</td>
<td>2</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>43</td>
<td>157</td>
<td>200</td>
<td>67</td>
<td>133</td>
<td>200</td>
</tr>
<tr>
<td><strong>Crude rates</strong></td>
<td>.215</td>
<td>.335</td>
<td>.500</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For $k=1$ the rate $r_{11}(x)=E_{11}(x)/E_{1+}(x)$, 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 $E_{ik}\hat{(*)}$ 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 $x_i$-specific rates $r_{ik}\hat{(*)}(x)$ of the $k$-th category of $D$ are equal to the observed rates $r_{ik}(x)$ as shown in Table 1. For instance, $r_{11}\hat{(*)}(1)$ based on Table 2.a is $r_{11}\hat{(*)}(1)=E_{11}\hat{(*)}(1)/E_{1+}\hat{(*)}(1)=29.656/59.312=0.50$, 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 $x_i$-specific rates of any category of $D$ are independent of the AB interactions (see equation (5)).

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

<table>
<thead>
<tr>
<th>State factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome factor/</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Confounding factor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11.6009</td>
<td>46.4037</td>
<td>10.5411</td>
</tr>
<tr>
<td>3</td>
<td>5.8140</td>
<td>66.6605</td>
<td>5.2770</td>
</tr>
<tr>
<td>Total</td>
<td>47.0709</td>
<td>42.7432</td>
<td>95.8026</td>
</tr>
</tbody>
</table>

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

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 \( x_1 \)-specific rates. The proportions are identical only in the states where the \( x_1 \)-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 \( E_{ij}^+(x) \) and proportions \( V_i(x) \) purged of AB interactions by state.

<table>
<thead>
<tr>
<th>State factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounding factor</td>
<td>( E_1^+(x) )</td>
<td>( V_1(x) )</td>
<td>( E_2^+(x) )</td>
</tr>
<tr>
<td>1</td>
<td>59.3120</td>
<td>0.3122</td>
<td>53.8502</td>
</tr>
<tr>
<td>2</td>
<td>58.0046</td>
<td>0.3053</td>
<td>52.7055</td>
</tr>
<tr>
<td>3</td>
<td>72.6745</td>
<td>0.3825</td>
<td>65.9630</td>
</tr>
<tr>
<td>Total</td>
<td>189.9911</td>
<td>1.0000</td>
<td>172.5187</td>
</tr>
</tbody>
</table>

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 \( x_1 \)-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 \( x_i \)-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 adjustment 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 \( i=\text{year}, k=\text{employment status}, x=\text{age} \)

<table>
<thead>
<tr>
<th>Age</th>
<th>1969</th>
<th>1971</th>
<th>1973</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>unempl.</td>
<td>empl.</td>
<td>total</td>
</tr>
<tr>
<td>14-19</td>
<td>668</td>
<td>5122</td>
<td>5790</td>
</tr>
<tr>
<td>20-34</td>
<td>784</td>
<td>18581</td>
<td>19365</td>
</tr>
<tr>
<td>35-49</td>
<td>413</td>
<td>19155</td>
<td>19568</td>
</tr>
<tr>
<td>50-64</td>
<td>310</td>
<td>14250</td>
<td>14560</td>
</tr>
<tr>
<td>65+</td>
<td>58</td>
<td>2508</td>
<td>2566</td>
</tr>
<tr>
<td>Total</td>
<td>2233</td>
<td>59616</td>
<td>61849</td>
</tr>
</tbody>
</table>

Crude unemployment rate (percent)

<table>
<thead>
<tr>
<th>Year</th>
<th>1969</th>
<th>1971</th>
<th>1973</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>3.61</td>
<td>6.43</td>
<td>5.39</td>
</tr>
</tbody>
</table>

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 $r_{ik}^{e}(.)$ are expressed as a weighted average of the $x_1$-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 $r_{ik}^{e}(.)$ 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.

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

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 \( r_{11}(x) \) based on the calculations from Table 3 and the weights \( V_{1}(x) \) 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 \( V_{1}(x) \) 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 \(-0.0262\) 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 occurred 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 $V_{iA}(x)$ of the population by age and years of observations and standard $V^S(x)$

<table>
<thead>
<tr>
<th>Age</th>
<th>1969</th>
<th>1971</th>
<th>1973</th>
<th>Standard $V^S(x)$ a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-19</td>
<td>.0936</td>
<td>.0985</td>
<td>.1062</td>
<td>.0961</td>
</tr>
<tr>
<td>20-34</td>
<td>.3131</td>
<td>.3243</td>
<td>.3507</td>
<td>.3187</td>
</tr>
<tr>
<td>35-49</td>
<td>.3164</td>
<td>.3059</td>
<td>.2857</td>
<td>.3112</td>
</tr>
<tr>
<td>50-64</td>
<td>.2354</td>
<td>.2348</td>
<td>.2212</td>
<td>.2351</td>
</tr>
<tr>
<td>65+</td>
<td>.0145</td>
<td>.0365</td>
<td>.0344</td>
<td>.0390</td>
</tr>
<tr>
<td>All ages</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0001</td>
</tr>
</tbody>
</table>

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

$$E_{ik}^{**}(x) = w_{xik} A \cdot w_i B \cdot w_k D \cdot w_{ik}^{AD} \cdot w_k^{BD} . \tag{15}$$

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

$$V_i^*(x) = \frac{w_{xik} A \cdot \sum_k w_k D \cdot w_{ik}^{BD} \cdot w_{ik}^{AD}}{\sum_x w_{xik} A \cdot \sum_k w_k D \cdot w_{ik}^{BD} \cdot w_{ik}^{AD}} . \tag{16}$$

Note that since the $x_i$-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 $x_i$-specific rates purged of both AB and ABD interaction by $r_{ik}^*(x)$.

$$r_{ik}^*(x) = \frac{w_k D \cdot w_{ik}^{BD} \cdot w_{xik}^{AD}}{\sum_k w_k D \cdot w_{ik}^{BD} \cdot w_{xik}^{AD}} = E_{ik}^{**}(x)/E_{ik}^{**}(x) . \tag{17}$$

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

$$r_{ik}^{**}(x) = \sum_x r_{ik}^*(x) \cdot V_i^*(x) . \tag{18}$$

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 empirical 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 $E_{1+}**(x)$ and proportions $V_1^*(x)$ purged of both AB and ABD interactions.

<table>
<thead>
<tr>
<th>State factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounding factor x</td>
<td>$E_{1+}**(x)$</td>
<td>$V_1^*(x)$</td>
<td>$E_{2+}**(x)$</td>
</tr>
<tr>
<td>1</td>
<td>67.31</td>
<td>0.33655</td>
<td>76.31</td>
</tr>
<tr>
<td>2</td>
<td>63.72</td>
<td>0.31860</td>
<td>63.72</td>
</tr>
<tr>
<td>3</td>
<td>68.97</td>
<td>0.34485</td>
<td>68.97</td>
</tr>
<tr>
<td>All categories</td>
<td>200.00</td>
<td>1.00000</td>
<td>200.00</td>
</tr>
</tbody>
</table>

Source: Table 2.a. and Appendix 1.
References


Appendix 1

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

\[ \log E_{ijk}(\%) = U + U_A + U_B + U_C + U_{AD} + U_{BD} + U_{AB} + U_{AC} + U_{BC} + U_{ABD} + U_{ACD} + U_{BCD} + U_{ABCD} \]

Overall Mean: \( U = 3.233 \)

Main Effects

<table>
<thead>
<tr>
<th>( U_1^A )</th>
<th>( U_1^B )</th>
<th>( U_1^C )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.2174</td>
<td>-.0607</td>
<td>.4254</td>
</tr>
<tr>
<td>-.0272</td>
<td>-.1567</td>
<td>-.4254</td>
</tr>
<tr>
<td>-.1902</td>
<td>.2174</td>
<td></td>
</tr>
</tbody>
</table>

Two Factor Interactions

<table>
<thead>
<tr>
<th>AB</th>
<th>AD</th>
<th>BD</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.1702</td>
<td>.6187</td>
<td>-.4465</td>
</tr>
<tr>
<td>-.1488</td>
<td>.3526</td>
<td>-.2036</td>
</tr>
<tr>
<td>.3190</td>
<td>-.9713</td>
<td>.6523</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Three Factor Interactions (ABD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.2126</td>
</tr>
<tr>
<td>-.0183</td>
</tr>
<tr>
<td>-.1943</td>
</tr>
</tbody>
</table>

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)
*INPUT I
*LOOK Y
*YVAR Y
*ERROR P
*CALC A2=%EG(A,1)-%EG(A,2) : A3=%EG(A,1)-%EG(A,3) : B2=%EG(B,1)-%EG(B,2) : B3=%EG(B,1)-%EG(B,3) : D2=%EG(D,1)-%EG(D,2)
*CALC G1=A2*B2*D2 : G2=A2*B3*D2 : G3=A3*B2*D2 : G4=A3*B3*D2
*FIT A2+A3+B2+B3+D2+P1+P2+P3+P4+P5+P6+P7+P8+G1+G2+G3+G4
*DISPLAY MERT
*STOP

25 25 50 50 25 25 8 92 2 23 50 50

GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

CYCLE 3
DEVIANCE 4394E-11

V-VARIATE Y ERROR POISSON LINK LOG
LINEAR PREDICTOR
XGM A2 A3 B2 B3 D2 P1 P2 P3 P4 P5 P6 P7 P8 G1 G2 G3 G4

ESTIMATE S.E. PARAMETER
1 3,233 6105E-01 XGM
2 2724E-01 7998E-01 A2
3 1902 1019 A3
4 1507 9950E-01 B2
5 -2174 7469E-01 B3
6 -4254 6105E-01 D2
7 3526 1220 P1
8 -2038 1023 P2
9 -9719 1779 P3
10 -5393 1135 P4
11 3669E-01 7998E-01 P5
12 3887 1019 P6
13 2127 9950E-01 P7
14 -4254 7469E-01 P8
15 -1839E-01 1220 Q1
16 3669E-01 1023 Q2
17 -1944 1779 Q3
18 -3867 1135 Q4

SCALE PARAMETER TAKEN AS 1.000

UNIT OBSERVED FITTED RESIDUAL
1 25 25 .1990E-11
2 25 25 .1918E-11
3 50 50 .2110E-11
4 50 50 .2713E-11
5 25 25 .1421E-11
6 25 25 .1705E-11
7 10 10 .4494E-13
8 40 40 .1887E-11
9 15 15 .9907E-12
10 60 60 .2202E-11
11 25 25 .1137E-11
12 25 25 .1279E-11
13 8 8 .3135E-11
14 92 92 .3271E-11
15 92 92 .4994E-07
16 23 23 .8860E-12
17 50 50 .1305E-11
18 50 50 .1708E-11
Appendix 3

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

```glim
$UNITS 18
$DATA Y
$FACTOR A 3 B 3 D 2
$CALC A=$GL(3,6) : B=$GL(3,2) : D=$GL(2,1)
$INPUT 1
$LOOK Y
$YVAR Y
$ERROR P
$CALC A2=$EQ(A, 1)-$EQ(A, 2)
   B3=$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
$FIT A2+A3+B2+B3+D2+P5+P6+P7+P8+G1+G2+G3+G4
$DISPLAY MAR
$STOP

11.6009  46.4037  10.5411  42.1644  30.6373  30.6373
.8140  66.8605  9.2770  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 G1 G2 G3 G4

ESTIMATE S.E. PARAMETER
  1 3.232 .5313E-01   %GM
  2 7.16E-01  7615E-01   A2
  3 1.792 .7514E-01   A3
  4 1.568 .6602E-01   B2
  5 4.253 .9116E-01   B3
  6 3.679E-01  6854E-01   D2
  7 .3885 .7391E-01   P5
  8 2.126 .7101E-01   P6
  9 4.283 .6638E-01   P7
 10 .1859E-01 .9028E-01   P8
 11 .3679E-01 .9116E-01   G1
 12 .1941 .9116E-01   G2
 13 .3889 .9687E-01   G3
 14 .3985 .9687E-01   G4

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  1738E-02
  6   39  39.13  1738E-02
  7   12  11.61  1740E-02
  8   44  46.41  8696E-03
  9   11  10.54  1490E-02
 10   31  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  3135E-03
 16   61  60.69  9286E-03
 17   26  26.03  1937E-02
 18   26  26.03  1937E-02
```
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.

```plaintext
$UNITS 30
$DATA Y
$FACTOR A 5 B 3 D 2
$CALC A=GL(5.6): B=GL(3.2): D=GL(2.1)
$INPUT
$LOOK Y
$YVAR Y
$ERROR P
$CALC
A2=ZEQ(A,1)-ZEQ(A,2)
A3=ZEQ(A,1)-ZEQ(A,3)
A4=ZEQ(A,1)-ZEQ(A,4)
A5=ZEQ(A,1)-ZEQ(A,5)
B2=ZEQ(B,1)-ZEQ(B,2)
B3=ZEQ(B,1)-ZEQ(B,3)
D2=ZEQ(D,1)-ZEQ(D,2)

$CALC
P6=A4*B3: P7=A5*B2: P8=A5*B3

$CALC

$CALC
G5=A5*B2*D2: G6=A5*B3*D2

$FIT
A2+A3+A4+A5+82+83+D2+Pl+P2+P3+P4+P5+P6+P7+P8+P9+P10+P11+P12
+P13+P14+G1+G2+G3+G4+G5+G6+G7+G8

$DISPLAY MERT
$STOP

GLIM 3.1 (C) 1977 ROYAL STATISTICAL SOCIETY, LONDON

Y-VARIATE Y
ERROR POISSON LINK LOG
LINEAR PREDICTOR

ESTIMATE S.E. PARAMETER
1 7.5512E-02 7GM
2 -8.8051E-01 8GM
3 4.9531E-01 4A2
4 -2.0191E-01 4A3
5 1.6641E-01 5A4
6 -1.1271E-01 6B2
7 3.7601E-01 7B3
8 -1.5161E-01 8B4
9 4.8741E-02 9B5
10 8.9681E-01 10B6
11 -2.7321E-01 11P2
12 1.2631E-01 12P3
13 -1.9351E-01 13P4
14 7.0021E-02 14P5
15 1.2031E-01 15P6
16 1.5831E-01 16P7
17 -1.4001E-01 17P8
18 1.2141E-01 18P9
19 2.2351E-01 19P10
20 2.8401E-01 20P11
21 -1.4131E-01 21P12
22 6.3081E-02 22P13
23 2.0691E-01 23P14
24 1.6811E-01 24P15
25 9.1891E-02 25P16
26 1.1861E-01 26P17
27 9.1911E-02 27P18
28 2.5211E-01 28P19
29 2.6771E-01 29P20
30 -9.0831E-01 30P21
SCALE PARAMETER TAKEN AS 1.000

CYCLE DEVIANCE DF
2 1475E-08 0


1 1.000
```
### Table 5.a. Log-linear parameters of "age-time interaction" $U_{x,t}^{AB}$

<table>
<thead>
<tr>
<th>Age</th>
<th>Years (States)</th>
<th>1969</th>
<th>1971</th>
<th>1973</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-19</td>
<td>1969</td>
<td>-.0246</td>
<td>-.0638</td>
<td>.0884</td>
</tr>
<tr>
<td></td>
<td>1971</td>
<td>.0049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-34</td>
<td>1969</td>
<td>-.0946</td>
<td>.0049</td>
<td>.0697</td>
</tr>
<tr>
<td></td>
<td>1971</td>
<td>.0273</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-49</td>
<td>1969</td>
<td>-.0147</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1971</td>
<td>.0273</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-64</td>
<td>1969</td>
<td>-.0126</td>
<td>.0196</td>
<td>-.0070</td>
</tr>
<tr>
<td></td>
<td>1971</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65+</td>
<td>1969</td>
<td>.1465</td>
<td>.0120</td>
<td>-.1585</td>
</tr>
</tbody>
</table>

Source: Appendix 4.

### Table 5.b. "Age-time interaction" parameters $w_{x,t}^{AB}$ of the multiplicative model.

<table>
<thead>
<tr>
<th>Age</th>
<th>Years (States)</th>
<th>1969</th>
<th>1971</th>
<th>1973</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-19</td>
<td>1969</td>
<td>.9757</td>
<td>.9382</td>
<td>1.0924</td>
</tr>
<tr>
<td>20-34</td>
<td>1969</td>
<td>.9097</td>
<td>1.0049</td>
<td>1.0938</td>
</tr>
<tr>
<td>35-49</td>
<td>1969</td>
<td>.9854</td>
<td>1.0277</td>
<td>.9875</td>
</tr>
<tr>
<td>50-64</td>
<td>1969</td>
<td>.9875</td>
<td>1.0198</td>
<td>.9930</td>
</tr>
<tr>
<td>65+</td>
<td>1969</td>
<td>1.1578</td>
<td>1.0121</td>
<td>.8534</td>
</tr>
</tbody>
</table>

Source: Table 5.a.
## Appendix 6

Matrix M - crude and standardized rates$^a$)

<table>
<thead>
<tr>
<th>Standard used $V^S(x)$</th>
<th>States 1969</th>
<th>States 1971</th>
<th>States 1973</th>
</tr>
</thead>
<tbody>
<tr>
<td>1969</td>
<td>.0361</td>
<td>.0633</td>
<td>.0513</td>
</tr>
<tr>
<td>1971</td>
<td>.0368</td>
<td>.0643</td>
<td>.0522</td>
</tr>
<tr>
<td>1973</td>
<td>.0380</td>
<td>.0663</td>
<td>.0539</td>
</tr>
</tbody>
</table>

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