REFERENCES

1. Australian Bureau of Statistics, The Labour Force, Ref. 6.20, Ref. 6.22 (Canberra).

2. Australian Bureau of Statistics, Labour Mobility, Ref. 6.43 (Canberra).

3. Burton, J.F. and Parker J.E., "Interindustry Variations in Voluntary Labor Mobility", (Industrial and Labor Relations Review, January 1969).

4. Congalton, A.A., "Status and Prestige in Australia", Studies in Australian Society, (F.W. Cheshire Publishing Pty. Ltd., Melbourne 1969).

5. McIntosh, M.K. and Granek, C.B., "A Socio-Economic Index for Occupations", Research Memorandum, Bachuroo Module, December 1976, (mimeo), pp.17.

6. Naphtali, J.A., McIntosh, M.K. and Williams, Lynne S., "A Cross-Sectional Analysis of Inter-Occupational Mobility in Australia", Impact of Demographic Change on Industry Structure in Australia, Working Paper B-06, Industries Assistance Commission, Melbourne, February 1978, (mimeo), pp.86.

7. Parnes, H.S. and Spitz, R.S., "A Conceptual Framework for Studying Labor Mobility", (Monthly Labor Review, November 1969).

8. Parnes, H.S., "Labor Force Participation and Labor Mobility", (Review of Industrial Relations Research, Volume 1, May 1970).

9. Stoikov, V. and Raimon, R.L., "Determinants of Differences in the Quit Rate Among Industries", (American Economic Review, December 1968).

IMPACT OF DEMOGRAPHIC CHANGE ON INDUSTRY STRUCTURE IN AUSTRALIA

A joint study by the Australian Bureau of Statistics, the Department of Employment and Industrial Relations, the Department of Environment, Housing and Community Development, the Department of Industry and Commerce and the Industries Assistance Commission

CONFIDENTIAL : Not for quotation without prior clearance from the author; comments welcome

A TWO-STAGE APPROACH TO THE MODELLING OF INTER-OCCUPATIONAL MOBILITY IN AUSTRALIA

by

Lynne S. Williams
Monash University

Preliminary Working Paper No. BP-11 Melbourne March 1978

The views expressed in this paper do not necessarily reflect the opinions of the participating agencies, nor of the Australian Government.

IMPACT Project, Industries Assistance Commission, 608 St. Kilda Road, Melbourne, Victoria 3004 Australia
Area Code 03 Telephone 51 8611

The above analysis supports our claim that we have failed to identify the underlying structure of the occupation changing process. To do this will involve specification of both the demand and supply functions for specific occupations, and the way they interact to yield the actual mobility that occurs in any one period - that is, the dependent variable in the reduced form.

We therefore derive

$$F^2_{16} = \frac{9798.3620/2}{18400.9397/16}$$

$$= 8.5199$$

From statistical tables, F^2_{16} (0.05) = 3.63.

Since 8.5199 > 3.63, we reject the null hypothesis that a common set of regression coefficients exist to explain the Stage 1 process in both periods.

Stage 2

$p = 2, \quad k = 5, \quad m = 18$

$r'r_{1972} = 734.2243$

$r'r_{1975} = 3937.4448$

Summing over both samples,

$r'r_{TOTAL} = 4671.6691$

$s's = 13537.7927$

yielding

$$F^5_{26} = \frac{8866.1236/5}{4761.6691/26}$$

$$= 9.8688$$

Again, statistical tables yield F^5_{26} (0.05) = 2.50, so we reject the null hypothesis of insignificantly different regression coefficients for the Stage 2 procedure also.

C O N T E N T S

$$F^{k(p-1)}_{p(m-k)} = \frac{(S_1+S_3)/k(p-1)}{S_4/p(m-k)}$$

where $S_1 + S_3 = s's - r'r$

$S_4 = r'r$

and $s's$ = residual sum of squares for the combined sample

$r'r$ = sum of residual sum of squares for the individual samples

p = no. of individual samples

k = no. of explanatory variables

m = no. of observations in the combined sample

For reasons advanced in the text, our 'best' estimating equations are those which have been corrected to take account of heteroskedasticity. Thus we will only test for structural change between the weighted estimating equations. Recall also that analysis is restricted to males.

<u>Stage 1</u>

$p = 2, \quad k = 2, \quad m = 18$

$r'r_{1972} = 7785.2290$

$r'r_{1975} = 10615.7107$

Summing over both samples,

$r'r_{TOTAL} = 18400.9397$

$s's = 28199.3017$

APPENDIX 3 : A Structural Test of each Stage of the Two-Stage Procedure

Our aim is to test the equality of the sets of regression coefficients over equations related to two different time periods, viz. 1972 to 1975. If the functional form of the equations is correctly specified and stable, significance of the appropriate F-statistic (that is, rejection of the null hypothesis that equality holds from a statistical point of view) would imply that the parent regression coefficients differ between the two time periods. The interpretation offered in the text above for such significant differences is that the estimated equation is neither structural form nor reduced form, and therefore subject to the instability introduced by omitted variables.

Here we are concerned with testing whether or not the two 'samples' could reasonably have come from the same 'parent' population. That is, have there been significant changes in the sets of coefficients between the two periods for which data exists.

The appropriate F-test[1] is given by

1. See Johnston, J., Econometric Methods, Second Edition (McGraw-Hill, New York 1972) pp. 192-207.

A TWO-STAGE APPROACH TO THE MODELLING OF INTER-OCCUPATIONAL MOBILITY IN AUSTRALIA

by

Lynne S. Williams*
Monash University

1. INTRODUCTION

Occupational mobility is one facet of labour market behaviour that affects both the supply and demand structure of occupations. Knowledge of mobility patterns is important in estimating needs for new workers and consequent occupational training requirements. To individual workers, occupational mobility may be a means to advance towards a career goal or simply a way to choose a better job from among many occupations. Further if a worker finds himself involuntarily unemployed from his optimal job preference, occupational mobility may be one means whereby he can obtain alternative employment.

This paper builds on ideas proposed by Naphtali, McIntosh and Williams[1] who use a reduced form to explain net mobility between occupations in the twelve months preceding both November 1972 and December 1975. Their approach involves regressing net mobility on all the explanatory variables

* The author would like to thank Alan A. Powell for extensive comments and assistance. Useful contributions were also made by Dean Parham and Rowen Craigie.

1. Naphtali, McIntosh and Williams (1978).

expected to be relevant, but overall their results are highly unsatisfactory. The estimated equations (year and sex specific) have very little explanatory power, and individual coefficients are often insignificant, of the wrong a priori expected sign, unstable over time, or some combination of these characteristics.

In an attempt to overcome some of these problems, a more formal structure on the process of occupation changing has been imposed. Mobility is still interpreted from the worker's point of view, but it is hypothesized that the decision to change jobs can be broken into two independent stages: that of leaving an original occupation,[1] and that of deciding which new occupation to enter. Different explanatory variables influence the two stages. The probability of moving from occupation i to occupation j is then viewed as the product of the probabilities first of moving out of i and then of moving into j. It is realized that the decision concerning which occupation to enter may not be taken independently of one's original occupation, but in this exploratory analysis, possible interdependencies of this type are ruled out.

The remainder of the paper is structured as follows: in section 2 the proposed model is described at the theoretical level; the methodology and results of the different stages of estimation are described and presented in section 3; in section 4 conclusions are given and implications for future research are summarised.

---

1. See, for example, Stoikov and Raimon (1968), Burton and Parker (1969), Parnes and Spitz (1969), Parnes (1970).

Because skill barriers between occupations are directly related to source occupations, it appears that their influence cannot be adequately captured in a two-stage process to describe occupational change, where the two stages are assumed independent. However, in so far as we are aware of the limitations of $Y_{.j}$, and because the averaging process employed results in it being appropriate to some source occupations, we have retrained $Y_{.j}$ as an explanatory variable in the Stage 2 estimating equation.

Males 1972

Weighted

$$(6'''')\quad M_{.j} = \underset{(29.5003)}{0.0579} + \underset{(0.5550)}{0.0074}\, W_{.j} - \underset{(-6.0964)}{0.0022}\, U_{.j} + \underset{(2.2177)}{0.0296}\, S_{.j} + \text{residual} \qquad \bar{R}^2 = 0.9767$$

Males 1975

Unweighted

$$(7''')\quad M_{.j} = \underset{(10.4840)}{0.0448} + \underset{(0.2457)}{0.0091}\, W_{.j} - \underset{(-0.5596)}{0.0006}\, U_{.j} + \underset{(1.0926)}{0.0365}\, S_{.j} + \text{residual} \qquad \bar{R}^2 = 0.4182$$

Weighted

$$(7'''')\quad M_{.j} = \underset{(9.5308)}{0.0412} + \underset{(0.1310)}{0.0047}\, W_{.j} + \underset{(0.2027)}{0.0002}\, U_{.j} + \underset{(0.4288)}{0.0153}\, S_{.j} + \text{residual} \qquad \bar{R}^2 = 0.7599$$

Comparison of these equations to those in the text (where $Y_{.j}$ is included) show that the equations in which $Y_{.j}$ is ommitted have slightly less explanatory power. This is to be expected, since in so far as the $Y_{.j}$ are high, they do reflect the difficulty of movement between occupations requiring extensive retraining. However, this same averaging procedure disguises the relative ease of movement that occurs between occupations of similar skills, thus making the interpretation of the $Y_{.j}$ coefficient very difficult.

The coefficient itself is badly determined, as reflected in the fact that it is insignificant for both 1972 and 1975. Further, there is a significant change in its magnitude from positive to negative over the two years.

2. THEORETICAL OUTLINE

It is hypothesized that the process of moving from occupation i to occupation j involves two stages, namely,

(i) the "decision" (whether voluntary or forced) to leave occupation i ;

(ii) having left one's previous occupation, the decision to move into a particular occupation j .

It is further assumed that these two stages are independent, so that, for $i \neq j$,

Pr {moving from occupation i to occupation j}

= Pr {moving out of occupation i} · Pr {moving into occupation j} .

Stage 1

Total movement out of occupation i is regarded as being influenced primarily by a weighted average of the past and present rates of unemployment in i relative to the corresponding weighted average in all other occupations. That is, if unemployment in the rest of the market has been consistently low relative to occupation i , movement out of i is expected to be high. Alternatively stated, if unemployment in i has been relatively high, movement out of i will be greater as more people are laid off, or as they voluntary go to other occupations in an attempt to obtain higher job security.

Thus,

(1) $M_{i.} = f(R_{i.}) + \text{error}$

where

$M_{i.}$ = gross rate[1] of movement out of i to all other occupations ($M_{i.} = \sum_{j \neq i} M_{ij}$, where M_{ij} is the rate of movement from i to j);

$R_{i.}$ = the weighted unemployment rate in i relative to all other occupations.

The specific functional form and method of estimation of equation (1) will be investigated below.

Stage 2

This stage of the decision process involves a comparison of relevant variables for the actual destination occupation with composite variables representing all other possible occupations. The arguments entering this stage include both economic and social factors such as relative wages, rates of unemployment, years of training required and a measure of social status. *A priori*, movement is expected to be greater to areas of higher relative wages and social standing; acting as opposing forces, however, will be factors such as high unemployment rates (implying not only lower job security prospects but also smaller probabilities of getting jobs), and length of training required to enter a particular

---

1. The "gross rate" of movement out of i is defined as the total number of persons leaving i within a twelve month period divided by the number of persons in i during that period.

To illustrate the ambiguity of $Y_{.j}$, consider total movement into the occupation denoted 'Lecturers and Teachers'. The number of part-time years required to move from each occupation to that of 'Lecturers and Teachers' are estimated to be:

Professional White Collar	2
Skilled White Collar	4
Semi and Unskilled White Collar	6
Skilled Blue Collar - Metal and Electrical	6
- Building	6
- Other	6
Semi and Unskilled Blue Collar	6
Rural	6 .

The calculated $Y_{.j}$ values for 1972 and 1975, where j denotes 'Lecturers and Teachers', are 5.6012 and 5.5414 part-time years respectively. Yet the majority of movement into the teaching group comes from the professional and skilled white collar sectors, movements requiring relatively little retraining. Therefore the high value of $Y_{.j}$ does not adequately represent the sectors which provide the majority of the corresponding $M_{.j}$, making it difficult to interpret exactly what the coefficient on $Y_{.j}$ does represent.

Both equations (6) and (7) [where (7) is (6) corrected for heteroskedasticity] were re-estimated ommitting the variable $Y_{.j}$ with the following results :

Males 1972

Unweighted

(6''') $$M_{.j} = \underset{(25.9075)}{0.0602} + \underset{(0.9586)}{0.0155}\, W_{.j} - \underset{(-4.2141)}{0.0025}\, U_{.j} + \underset{(3.0778)}{0.0438}\, S_{.j} + \text{residual} \qquad \bar{R}^2 = 0.9119$$

APPENDIX 2 : The Years of Training Variable

The number of part-time years required to retrain for a specific occupation relative to an average required to retrain for all other occupations appears as an explanatory variable at Stage 2 of the decision making process. That is, it is considered relevant in determining which destination occupation to enter. Although it is recognised that certain skill barriers do exist between the various occupations, these barriers are directly related to the source occupations.

Recall that equation (6) takes the form

(6) $$M_{.j} = C_1 + C_2 W_{.j} + C_3 U_{.j} + C_4 S_{.j} + C_5 Y_{.j} + \text{error} .$$

The variables are as defined in the text, and although the coefficients C_1, C_2, C_3 and C_4 have an unambiguous meaning, the interpretation of C_5 is far more complex. This is because it is the only variable which is related to qualifications required in the source occupation.

Movement will be greatest between occupations requiring the least retraining, and vice versa. Yet this distinction is lost in an equation such as (6), since total movement into occupation j is regressed on a variable reflecting the average amount of retraining required from all other occupations, when in fact most of the movement comes from those few occupations requiring least retraining. The averaging process we adopt to derive $Y_{.j}$ results in the loss of this information, and further makes the coefficient on $Y_{.j}$ badly determined.

occupation - - the longer the training period required, the smaller the expected movement to that occupation.

Thus,

(2) $$M_{.j} = g(W_{.j}, U_{.j}, S_{.j}, Y_{.j}) + \text{error}$$

where

$M_{.j}$ = gross rate[1] of movement into occupation j from all other occupations ($M_{.j} = \sum_{i \neq j} M_{ij}$, where M_{ij} is as previously defined);

$W_{.j}$ = the wage rate in occupation j relative to a weighted average of average wages in the other occupations;

$= \frac{W_j - W_.}{W_.}$ where W_j = the average wage rate in occupation j

and $$W_. = \frac{\sum_{i \neq j} L_i W_i}{\sum_{i \neq j} L_i}$$

where L_i = labour force in occupation i ;

$U_{.j}$ = the unemployment rate in occupation j relative to a weighted average of the other occupations;

$= \frac{U_j - U_.}{U_.}$ where U_j = the unemployment rate in occupation j

$$U_. = \frac{\sum_{i \neq j} L_i U_i}{\sum_{i \neq j} L_i} .$$

1. Gross rates into occupation j are defined analagously to gross rates out of i . See previous footnote.

$S_{.j}$ = the socio-economic index of occupation j relative to a weighted average of those of other occupations;

$= \frac{S_j - S_.}{S_.}$ where S_j = the socio-economic index assigned to occupation j

and $S_. = \frac{\sum_{i \neq j} S_i}{n}$

where n = (no. of occupations -1) .

$Y_{.j}$ = a weighted average of the number of years of part-time training[1] required to move to occupation j from all alternative source occupations;

$= \frac{\sum_{i \neq j} L_i Y_{ij}}{\sum_{i \neq j} L_i}$.

The proposed theory can thus be summarised in terms of three empirically testable hypotheses (which assume for simplicity that both equations (1) and (2) will be estimated linearly) :

H_1 : $E(M_{ij}) = E(M_{i.}) \times E(M_{.j})$, $i \neq j$;

H_2 : $E(M_{i.}) = a + bR_{i.}$, $b > 0$;

H_3 : $E(M_{.j}) = c_1 + c_2 W_{.j} + c_3 U_{.j} + c_4 S_{.j} + c_5 Y_{.j}$,

$c_2, c_4 > 0$, $c_3, c_5 < 0$;

1. Because this study deals with persons who change occupations without leaving the work force the relevant unit for the training variable is part-time years. That is, people retrain part-time without actually leaving the work force.

DATA FOR FEMALES :

Occupation	November 1972			December 1975		
	$M_{i.}$	$R_{i.}$	$\ell_{i.}$	$M_{i.}$	$R_{i.}$	$\ell_{i.}$
Professional White Collar	0.032200	0.902093	26.1736	0.040026	0.499809	37.9151
Lecturers & Teachers	0.037246	0.382853	62.4688	0.030423	0.286716	62.9894
Skilled White Collar	0.055096	0.378679	98.9253	0.033416	0.332520	85.0330
Semi and Unskilled White Collar	0.037056	1.436113	181.2562	0.022744	1.350875	160.0475
Skilled Blue Collar - Metal and Electrical	0.236504	1.222139	51.0404	0.214372	0.730498	54.1072
Skilled Blue Collar - Building	0.293883	0.000001	39.5588	0	0.000001	0.0600
Skilled Blue Collar - Other	0.132946	0.826419	87.4449	0.061264	1.157603	54.2230
Semi and Unskilled Blue Collar	0.056696	1.129765	170.7635	0.038904	1.217869	148.7158
Rural	0.044037	0.489510	54.8208	0.041202	0.728464	54.5301

SOURCES : as for males.

APPENDIX 1 : Some Female Results

As reported in the text, the results of estimating equations (3), (4) and (5) using female data were highly unsatisfactory but for completeness are recorded below :

Females 1972

(3) $$M_{i.} = \underset{(2.1503)}{0.1429} - \underset{(-0.6990)}{0.0532}\, R_{i.} \qquad \bar{R}^2 = 0.0652$$

(4) $$\ell n\, M_{i.} = \underset{(-10.7717)}{-2.8181} - \underset{(-1.9472)}{0.1099}\, \ell n\, R_{i.} \qquad \bar{R}^2 = 0.3514$$

(5) $$\ell n\, M_{i.} = \underset{(-18.1262)}{-2.9479} - \underset{(-1.4602)}{0.1271}\, \ell n\, R_{i.} \qquad \bar{R}^2 = 0.9426$$

Females 1975

(3) $$M_{i.} = \underset{(0.9301)}{0.0385} - \underset{(0.4325)}{0.0216}\, R_{i.} \qquad \bar{R}^2 = 0.0260$$

(4) $$\ell n\, M_{i.} = \underset{(-12.1957)}{-3.1596} - \underset{(-4.0210)}{0.2241}\, \ell n\, R_{i.} \qquad \bar{R}^2 = 0.6979$$

(5) $$\ell n\, M_{i.} = \underset{(-19.5610)}{-3.3896} - \underset{(-0.5707)}{0.1864}\, \ell n\, R_{i.} \qquad \bar{R}^2 = 0.9524$$

A cursory glance at these figures shows that the explanatory variable has the wrong a prioi expected sign in both years, that in both cases this variable is insignificant, and that weighting the estimating equation by the theoretically derived weighting factor reduces the apparent significance of the explanatory variable (equation (4) to equation (5)). Further, there is no evidence of stability across the two years.

where

M_{ij} = gross rate of movement from occupation i to occupation j , E is the expected value operator, and other variables are as previously defined.

The sequence of testing involves investigating H_2 and H_3 first, then using these results to test H_1 . By formulating the theory in terms of these three hypotheses, it is possible to reduce partially the severity of the data limitations on the analysis.

Data used

Data for the total number of workers transferring between designated source and destination occupations is only available for the twelve months preceding November 1972 and December 1975. This data was collected by the Australian Bureau of Statistics (ABS) as part of the Labour Force Surveys for these two years, and summaries using the IMPACT occupational breakdown have been made available to the IMPACT project. Because this data was only collected at two points in time, estimation is restricted to cross-section methods utilizing variation across occupations, although the analysis can be split by sex. The IMPACT 9-major occupation classification yields 72 data points from which to estimate 72 parameters. Yet, by adopting H_1 , H_2 and H_3 , the parameters to be estimated are reduced to 7. That is, H_1 reduces the number of parameters from 72 to 18 (9 + 9); H_2 takes the first 9 and reduces them to 2 ; and H_3 the second 9 and reduces them to 5.

Diagrammatically,

$$72 \xrightarrow[H_1]{} 9 + 9(18) \xrightarrow[H_2, H_3]{} 2 + 5(7) \quad .$$

Wage data on an occupation specific basis has been obtained from the ABS Income Distribution Surveys. Unfortunately, these were only conducted for the periods 1968-69 and 1973-74. However, it is argued that workers are more likely to react to wages in previous periods, so that 1968-69 wage data is more relevant to job changing decisions taken in 1972 than the (possibly unknown) wages holding in 1972. The same justification is advanced for including 1972-73 wages as an explanatory variable for mobility in 1975.

Unemployment data on an occupation specific basis is available from the ABS Labour Force Surveys for every year from 1968 except 1969 and 1975. A measure based on Congalton's[1] scale has been derived as a proxy for the social status variable. Estimates of the length of training required to effect transfers between major IMPACT occupational groups have been prepared by Craigie.[2]

Mobility rates are recorded in tables 1 and 2, while tables 3 and 4 contain the data matrices for the explanatory variables.

1. A.A. Congalton (1969) and modified to fit the IMPACT occupational breakdown by McIntosh and Granek (1977).

2. These estimates were obtained from work undertaken in the development of the education sub-module of BACHUROO. For a discussion of the appropriateness or otherwise of the length of training variable as an explanatory variable, see Appendix 2.

FOR THE TWELVE MONTHS PRECEDING DECEMBER 1975

Skilled Blue Collar - Building	Skilled Blue Collar - Other	Semi and Unskilled Blue Collar	Rural Workers	$\hat{M}_{i.}$
0.001842	0.001961	0.002071	0.001697	0.043188
0.001519	0.001616	0.001707	0.001399	0.035597
0.001707	0.001817	0.001919	0.001572	0.040015
0.001867	0.001987	0.002098	0.001719	0.043753
0.001839	0.001957	0.002067	0.001694	0.043105
	0.002033	0.002148	0.001760	0.044784
0.001999		0.002248	0.001842	0.046869
0.002218	0.002361		0.002043	0.051990
0.002047	0.002178	0.002301		0.047979
0.042660	0.045403	0.047958	0.039295	

TABLE 6 : PREDICTED MOBILITY RATES FOR MALES

$\hat{M}_{ij}$	Professional White Collar	Lecturers & Teachers	Skilled White Collar	Semi and Unskilled White Collar	Skilled Blue Collar - Metal and Electrical
Professional White Collar		0.001590	0.001733	0.001917	0.001839
Lecturers & Teachers	0.001245		0.001429	0.001580	0.001516
Skilled White Collar	0.001400	0.001473		0.001777	0.001704
Semi and Unskilled White Collar	0.001530	0.001611	0.001756		0.001863
Skilled Blue Collar - Metal and Electrical	0.001508	0.001587	0.001730	0.001914	
Skilled Blue Collar - Building	0.001566	0.001649	0.001798	0.001988	0.001907
Skilled Blue Collar - Other	0.001639	0.001725	0.001881	0.002081	0.001996
Semi and Unskilled Blue Collar	0.001819	0.001914	0.002087	0.002308	0.002214
Rural Workers	0.001678	0.001766	0.001926	0.002130	0.002043
$\hat{M}_{.j}$	0.034979	0.036812	0.040138	0.044396	0.042584

3. METHODOLOGY AND EMPIRICAL RESULTS

3.1 Stage 1

This stage deals with modelling the decision to move out of occupation i. Assume equation (1) has the functional form

(3) $$M_{i.} = a + bR_{i.} + \text{error} , \quad i=1,\dots,9 .$$

The construction of the indices $R_{i.}$ proceeded as follows. First, since the decision to change occupations is relatively long term in nature, it was felt that relative job security should be measured by a weighted average of the unemployment histories of different occupations over the 5 years (1968-72 and 1971-75) prior to, and including the survey years (1972 and 1975). A truncated lag scheme with the following weighting factors was adopted:

$$W_t = \frac{\beta \sum_{\tau=0}^{t} (1-\beta)^{\tau}}{\sum_{\tau'=0}^{4} \left[\beta \sum_{\tau=0}^{t} (1-\beta)^{\tau} \right]} , \quad 0<\beta<1 , \quad t=0,\dots,4 .$$

The parameter β was arbitrarily set at 0.5, resulting in the following weights :

$w_0 = 0.2403$; $w_1 = 0.2326$; $w_2 = 0.2171$;

$w_3 = 0.1861$ and $w_4 = 0.1240$.

The relevant average unemployment rate in occupation i for each of the two survey years was then calculated as[1]

$$N_i = \frac{\sum_{t=0}^{4} w_t U_{i,(x-t)}}{\sum_{t=0}^{4} w_t L_{i,(x-t)}} \quad \text{where} \quad x = 1972,\ 1975 \ .$$

the computation of the indices was completed by forming

$$R_{i.} = \frac{N_i}{\sum_{j \neq i} \left[N_j \times \frac{L_j}{\sum_{k \neq i} L_k} \right]} \ ,$$

where

N_i = average weighted unemployment rate in occupation i ;

$U_{i,t}$ = the number of unemployed in occupation i at time t ;

and $R_{i.}$ and L_i as previously defined.

Initial estimation of equation (3) involved a simple linear regression using the standard ordinary least squares technique. This was done, by sex, for each of the two survey years. The results for males were as expected, with relative unemployment having the correct sign in

1. Information concerning unemployed by occupation was not available for the years 1969 and 1975, so once the weights had been calculated they were standardised such that they summed to one. That is,

$w_{68} + w_{70} + w_{71} + w_{72} = 1$ and $w_{69} = 0$ for the 1972 data,

and $w_{71} + w_{72} + w_{73} + w_{74} = 1$ and $w_{75} = 0$ for the 1975 data.

FOR THE TWELVE MONTHS PRECEDING NOVEMBER 1972

Skilled Blue Collar - Building	Skilled Blue Collar - Other	Semi and Unskilled Blue Collar	Rural Workers	$\hat{M}_{i.}$
0.003971	0.003646	0.003485	0.002588	0.051866
0.002346	0.002518	0.002407	0.001787	0.035815
0.003195	0.003429	0.003278	0.002434	0.048777
0.003676	0.003946	0.003772	0.002801	0.056126
0.003505	0.003762	0.003596	0.002671	0.053515
	0.004298	0.004109	0.003051	0.061139
0.004327		0.004440	0.003297	0.066066
0.004731	0.005078		0.003604	0.072230
0.003992	0.004285	0.004096		0.060950
0.065497	0.070305	0.067202	0.049902	

TABLE 5 : PREDICTED MOBILITY RATES FOR MALES

$\hat{M}_{ij}$	Professional White Collar	Lecturers & Teachers	Skilled White Collar	Semi and Unskilled White Collar	Skilled Blue Collar - Metal and Electrical
Professional White Collar		0.002572	0.002787	0.003333	0.001988
Lecturers & Teachers	0.001314		0.001924	0.002301	0.001373
Skilled White Collar	0.001790	0.002418		0.003134	0.001870
Semi and Unskilled White Collar	0.002059	0.002783	0.003016		0.002152
Skilled Blue Collar - Metal and Electrical	0.001964	0.002653	0.002876	0.003439	
Skilled Blue Collar - Building	0.002434	0.003031	0.003285	0.003928	0.002344
Skilled Blue Collar - Other	0.002424	0.003276	0.003550	0.004245	0.002533
Semi and Unskilled Blue Collar	0.002650	0.003581	0.003881	0.004641	0.002769
Rural Workers	0.002236	0.003022	0.003275	0.003916	0.002337
$\hat{M}_{.j}$	0.036693	0.049580	0.053734	0.064255	0.038338

both years. However, whilst significantly different from zero in the first period, this variable's coefficient failed a significance test when estimated from data for the second period. This suggests that a greater level of sophistication is needed to represent the structural form adequately. Female results showed relative unemployment entering with a negative but insignificant coefficient. It is highly optimistic to expect any realistic results from the female data until it is split at least into two groups, 'married' and 'others', as the important decision making variables are likely to vary between these two groups. Further, the structural explanation of the behaviour of females is likely to differ from the one appropriate for males. Thus from now on analysis is restricted to the explanation of male mobility.

All the coefficients of determination ($\bar{R}^2$) obtained were on the low side, the best result being 42.57% for males in 1972. A plot of the raw data revealed that the relationship might be logarithmic in form. Therefore, equation (3) was respecified as :

(4) $$\ell n\, M_{i.} = \tilde{a} + \tilde{b}\, \ell n\, R_{i.} + \text{error}, \quad i=1,\ldots,9 .$$

Further at a theoretical level, there are good reasons for giving greater weights to observations involving larger rates of mobility. A correction of this type enables the existence of heteroskedasticity to be taken into account. Thus the equation to be estimated became

(5) $$\ell_{i.}\, M_{i.} = a^* + \ell_{i.}\, \tilde{b}\, \ell n\, R_{i.} + \text{error}, \quad i=1,\ldots,9 .$$

where $\ell_{i.}$ were the actual weights used and were derived in the following manner:

Recall that H_2 can be rewritten as

$$E(M_{i.}) = \sum_{j \neq i} E(M_{ij}) = \phi_i(R_{i.}) , \quad i,j=1,\ldots,9 .$$

Taking natural logs, and assuming ϕ_i is a linear function (as in equation (4)), one obtains

$$\ell n\, E(M_{i.}) = a^* + \tilde{b}\, \ell n\, R_{i.} .$$

Now, $E(M_{i.}) = \mu_{i.}$ where $\mu_{i.}$ = the population proportion parameter, $0 \leq \mu_{i.} \leq 1$.

That is, $M_{i.}$ is a sample estimate of the corresponding (unknown) population proportion $\mu_{i.}$, with variance

$$\text{Var}\,(M_{i.}) = \frac{\mu_{i.}\,(1-\mu_{i.})}{L_i} .$$

But

$$\text{Var}\,(\ell n\, y) \approx \left(\frac{\partial\, \ell n\, y}{\partial y}\right)^2 \cdot \text{Var}\,(y) = \frac{1}{y^2}\,\text{Var}\,(y) \quad \text{for any } y .$$

So,

$$\text{Var}\,(\ell n\, M_{i.}) \approx \frac{1}{M_{i.}^{\,2}} \cdot \frac{\mu_{i.}(1-\mu_{i.})}{L_i} ,$$

and

$$\widehat{\text{Var}\,(\ell n\, M_{i.})} \approx \frac{1}{M_{i.}^{\,2}} \cdot \frac{M_{i.}(1-M_{i.})}{L_i}$$

$$= \frac{1 - M_{i.}}{M_{i.}L_i} .$$

TABLE 4 : DATA MATRIX FOR MALES, DECEMBER 1975

Occupation	$R_{i.}$	$W_{.j}$	$U_{.j}$	$S_{.j}$	$Y_{.j}$	$\ell_{i.}$	$\ell_{.j}$
Professional White Collar	0.554210	0.433490	-0.871209	-0.5315	6.6542	55.1015	2207.8461
Lecturers & Teachers	0.114114	0.271369	-0.432454	-0.3671	5.5414	48.5458	2102.2187
Skilled White Collar	0.296961	0.302397	-0.646109	-0.1572	2.0613	137.4464	3269.8040
Semi and Unskilled White Collar	0.615575	-0.069558	2.464832	0.1989	0.0000	183.2535	3327.0623
Skilled Blue Collar - Metal and Electrical	0.545336	-0.068307	-0.934824	0.1200	4.0000	144.2144	4373.7251
Skilled Blue Collar - Building	0.745257	-0.156855	-1.000000	0.1530	4.0000	124.5526	2392.1315
Skilled Blue Collar - Other	1.081410	-0.245524	-0.534906	0.3538	4.0000	93.3515	1270.4475
Semi and Unskilled Blue Collar	2.523142	-0.245888	12.671379	0.3665	0.0000	244.3470	5124.8115
Rural	1.309607	-0.232634	-0.943848	-0.0438	0.0000	150.2221	2793.1325

SOURCES : $R_{i.}$, $U_{.j}$ } unpublished data from ABS <u>The Labour Force 1971, 1972, 1973, 1974</u> (Ref. No. 6.22) Canberra, 1973-76.

$W_{.j}$ unpublished data from ABS <u>Income Distribution Survey 1973-74</u> (Ref. No. 17.6) Canberra, 1974.

The values for these variables are based on working estimates supplied by the ABS for research purposes only. As such the data need not conform to the usual standards of reliability set by the ABS in its publications and care should be taken in applying them to uses other than those for which they were intended.

TABLE 3 : DATA MATRIX FOR MALES, NOVEMBER 1972

Occupation	$R_{i.}$	$W_{.j}$	$U_{.j}$	$S_{.j}$	$Y_{.j}$	$\ell_{i.}$	$\ell_{.j}$
Professional White Collar	0.452097	0.515893	4.613166	-0.5315	6.6832	58.2102	2195.4627
Lecturers & Teachers	0.069376	0.221471	0.292589	-0.3671	5.6012	33.4966	1222.6656
Skilled White Collar	0.331427	0.345994	0.603931	-0.1572	2.0731	147.5217	2811.7742
Semi and Unskilled White Collar	0.674541	-0.080987	-0.737506	0.1989	0.0000	196.5279	2897.1095
Skilled Blue Collar-Metal and Electrical	0.529794	-0.060119	10.461642	0.1200	4.0000	167.3626	3790.5367
Skilled Blue Collar - Building	1.039960	-0.136090	-1.000000	0.1530	4.0000	136.7887	1922.9001
Skilled Blue Collar - Other	1.539692	-0.101491	-0.410031	0.3538	4.0000	288.3785	1104.5604
Semi and Unskilled Blue Collar	2.417977	-0.309576	-0.321402	0.3665	0.0000	268.8902	4167.0114
Rural	1.023333	-0.180987	2.116352	-0.0438	0.0000	162.9670	2768.9949

SOURCES : $R_{i.}$, $U_{i.}$ } unpublished data from ABS The Labour Force 1968, 1970, 1971, 1972 (Ref. No. 6.22) Canberra, 1970-74.

$W_{.j}$ unpublished data from ABS Income Distribution Survey 1968-69 (Ref. No. 17.17) Canberra, 1969.

The values for these variables are based on working estimates supplied by the ABS for research purposes only. As such, the data need not conform to the usual standards of reliability set by the ABS in its publications and care should be taken in applying them to uses other than those for which they were intended.

$S_{.j}$ McIntosh,M.K., and Granek,C.B., A Socio-Economic Index for Occupations, Research Memorandum, Bachuroo Module, December 1976, (mimeo), pp.17.

$Y_{.j}$ Craigie, R. see p.6.

$\ell_{i.}$, $\ell_{.j}$ } weighting factors as defined in the text.

Hence the appropriate weighting factor is $\sqrt{\frac{M_{i.} L_i}{1-M_{i.}}}$.

Although the $\bar{R}^2$'s for equation (5) were very high, because the variables have been weighted, these statistics do not have their usual interpretation. In this case,

$$\bar{R}^2 = 1 - \frac{\sum \left(\frac{e_{i.}}{\ell_{i.}}\right)^2 \Big/ (n-k)}{\sum \left(\frac{y_{i.}}{\ell_{i.}}\right)^2 \Big/ (n-1)}$$

where

$\sum \left(\frac{e_{i.}}{\ell_{i.}}\right)^2$ = residual sum of squares,

$\sum \left(\frac{y_{i.}}{\ell_{i.}}\right)^2$ = total sum of squares,

n = total number of observations,

and

k = number of explanatory variables.

That is, the $\bar{R}^2$'s are deceptively high. A more accurate assessment of the explanatory power of the weighted variables is given by the correlation between actual $\ell n\, M_{i.}$ and the value of $\ell n\, M_{i.}$, $(\widehat{\ell n\, M_{i.}})$, predicted from the weighted variables regression equation. i.e. $r(\widehat{\ell n\, M_{i.}}, \ \ell n\, M_{i.})$. Similarly, comparison of the relative explanatory powers of equations (3) and (5) requires calculation of

$r(\hat{M}_i, M_i)$. The corresponding coefficients of determination are designated "quasi - $\bar{R}^2_{45}$" and "quasi - $\bar{R}^2_{35}$" and are reported where appropriate.

Actual results[1] for the three equations described are:

(t - statistics in brackets)

Males 1972 :

Unweighted

(3') $$M_{i.} = 0.0346 + 0.0189\, R_{i.} + \text{residual}$$
(3.7140) (2.2780) $\bar{R}^2 = 0.4257$

(4') $$\ln M_{i.} = -2.8641 + 0.4527 \ln R_{i.} + \text{residual}$$
(23.8758) (4.0759) $\bar{R}^2 = 0.7035$

Weighted

(5') $$\ln M_{i.} = -2.8023 + 0.1975 \ln R_{i.} + \text{residual}$$
(-44.6186) (2.0592) $\bar{R}^2 = 0.9771$

quasi - $\bar{R}^2_{35} = 0.6602$

quasi - $\bar{R}^2_{45} = 0.7035$

Males 1975 :

Unweighted

(3") $$M_{i.} = 0.0347 + 0.0108\, R_{i.} + \text{residual}$$
(3.5599) (1.2198) $\bar{R}^2 = 0.1753$

(4") $$\ln M_{i.} = -3.0786 + 0.3086 \ln R_{i.} + \text{residual}$$
(-18.3638) (1.7525) $\bar{R}^2 = 0.3050$

1. Corresponding results for females are not given in the body of the text, but can be found in Appendix 1.

TWELVE MONTHS PRECEDING DECEMBER 1975

Skilled Blue Collar - Building	Skilled Blue Collar - Other	Semi and Unskilled Blue Collar	Rural Workers	$M_{i.}$
0	0	0.004684*	0.001171*	0.015224
0	0	0.002405*	0	0.024048
0.001404*	0.001404*	0.008110	0.003288	0.037015
0.002041	0.001535*	0.024821	0.003071*	0.053066
0.001942*	0.001165*	0.017761	0.002584	0.033265
	0.001679*	0.032400	0.003358*	0.053582
0.004301*		0.029722	0.009060	0.075337
0.006479*	0.003324		0.007958	0.046192
0.003799	0.000633	0.036677		0.058486
0.047885	0.066268	0.046941	0.046565	

SOURCE : unpublished data from surveys partially reported in :

(1) ABS Labour Mobility February 1976 (Ref No. 6.43) Canberra 1976;

(2) ABS The Labour Force 1975 (Ref. No. 6.22) Canberra 1976.

TABLE 2 : MOBILITY RATES FOR MALES FOR THE

M_{ij}	Professional White Collar	Lecturers & Teachers	Skilled White Collar	Semi and Unskilled White Collar	Skilled Blue Collar - Metal and Electrical
Professional White Collar		0.003513*	0.003513*	0.002342*	0
Lecturers & Teachers	0.004810*		0.007214*	0.009619*	0
Skilled White Collar	0.003839	0.000468*		0.015494	0.003007
Semi and Unskilled White Collar	0.003793	0.000768*	0.012395		0.004642
Skilled Blue Collar - Metal and Electrical	0.000777*	0.000777*	0.003963	0.004295	
Skilled Blue Collar - Building	0.000839*	0.000839*	0.004197*	0.006912	0.003358*
Skilled Blue Collar - Other	0	0	0.010752*	0.012902*	0.006451*
Semi and Unskilled Blue Collar	0.001739*	0	0.005920*	0.012463	0.008309*
Rural Workers	0.001266	0	0.004432	0.005599	0.006081
$M_{.j}$	0.040290	0.021642	0.045969	0.054136	0.030949

* The m_{ij} values for these cells have been derived by the author to complete the table for the purposes of this analysis only. They have no official status with either the ABS or the IMPACT project. The large standard errors attached to the estimates make them highly unreliable for assessments of individual occupations or for other purposes.

$$M_{ij} = \frac{m_{ij}}{L_i}$$

where m_{ij} = gross movement from occupation i to occupation j, $i \neq j$

L_i = labour force in occupation i.

Weighted

(5") $\ln M_{i.} = -3.0700 + 0.1223 \ln R_{i.}$ + residual

(-33.9209) (1.0406) $\bar{R}^2 = 0.9586$

quasi - $\bar{R}^2_{35} = 0.3276$

quasi - $\bar{R}^2_{45} = 0.3051$

The two equations to be compared are the initial simple linear specification represented by (3) and the theoretically derived weighted logarithmic form of (5). Comparison of the quasi - $\bar{R}^2_{35}$ with the $\bar{R}^2$ for equation (3) shows that equation (5) yields greater explanatory power in both years. However, comparison between 1972 and 1975 shows little stability in coefficients. For 1972, relative unemployment is significant, but this is not the case in 1975, where this variable's coefficient has fallen by approximately 50% on its 1972 value. Although both constant terms are highly significant, again there is some instability between the two periods. The above analysis suggests that although relative unemployment is important in determining the rate of outflow from a particular occupation, there exist other relevant variables whose influence is currently being captured, if at all, in the constant term[1] and in the coefficient of the unemployment variable. The general lack of stability of the coefficients suggests a failure to model adequately the underlying structural form.[2] This result is not entirely unexpected, as the model has been set up in reduced form, with no reference to the structure of the appropriate demand and supply curves. Future estimation will involve investigation along these lines.

1. The constant term, of course, will only reflect the influence of those omitted variables which are constant over the two cross sections (but which may vary over time).
2. See Appendix 3 for a test of overall stability between the two estimation years.

3.2 Stage 2

Analysis at this stage assumes the worker has left his source occupation, and is now faced with the decision as to which occupation to enter. It is assumed that this decision is made after comparison of the values of the relevant variables pertaining to occupation j relative to an appropriately weighted average of those variables in the rest of the labour market (i.e. what will the individual's position in occupation j be as regards a particular criterion as compared to an average position in the rest of the labour market).

Variables which are considered important in determining this decision include lagged wages, unemployment, some sort of socio-economic index, and length of training required.

Recall that this relationship is initially specified in equation (2). The appropriate functional form is assumed to be linear, giving

$$(6) \qquad M_{.j} = C_1 + C_2 W_{.j} + C_3 U_{.j} + C_4 S_{.j} + C_5 Y_{.j} + \text{error} .$$

Again, however, observations should be weighted according to the magnitude of the flows they represent. Following analysis parallel to section 3.1, the appropriate weighting factor is derived as follows. From H_3 ,

$$E(M_{.j} \mid X_{.j}) = \mu_{.j}(X_{.j})$$

where $\mu_{.j}$ = the population proportion parameter for fixed $X_{.j}$ $(0 \leq \mu_{.j} \leq 1)$,

and $X_{.j}$ = the set of exogenous variables relevant to explaining movement into occupation j .

TWELVE MONTHS PRECEDING NOVEMBER 1972

Skilled Blue Collar - Building	Skilled Blue Collar - Other	Semi and Unskilled Blue Collar	Rural Workers	$M_{i.}$
0	0	0.007218*	0	0.021655
0	0	0	0	0.015198
0.004084*	0.001021*	0.014791	0.004594*	0.047865
0.002374*	0.001978*	0.028112	0.005144	0.064680
0.002775*	0.000793*	0.026678	0.003568*	0.047848
	0.002580*	0.038772	0.005161*	0.067881
0.004335*		0.049855	0.006503*	0.075867
0.009575	0.004519		0.011562	0.061518
0.004345*	0.002173*	0.038460		0.061272
0.069488	0.083566	0.063522	0.053068	

SOURCE : unpublished data from surveys partially reported in :

(1) ABS Labour Mobility November 1972 (Ref. No. 6.43) Canberra 1975;

(2) ABS The Labour Force 1972 (Ref. No. 6.22) Canberra 1974.

TABLE 1 : MOBILITY RATES FOR MALES FOR THE

M_{ij}	Professional White Collar	Lecturers & Teachers	Skilled White Collar	Semi and Unskilled White Collar	Skilled Blue Collar - Metal and Electrical
Professional White Collar		0.004331*	0.007218*	0.002887*	0
Lecturers & Teachers	0.009119*		0.006079*	0	0
Skilled White Collar	0.004084*	0.001531*		0.014696	0.003630*
Semi and Unskilled White Collar	0.001978*	0.001583*	0.017242		0.006268
Skilled Blue Collar - Metal and Electrical	0	0.000396*	0.003568*	0.010068	
Skilled Blue Collar - Building	0	0	0.010185	0.006021*	0.005161*
Skilled Blue Collar - Other	0	0	0.008670*	0.004335*	0.002168*
Semi and Unskilled Blue Collar	0.001002*	0.001002*	0.004977	0.017261	0.011620
Rural Workers	0.000543*	0	0.003802*	0.007061	0.004888*
$M_{.j}$	0.031760	0.048635	0.054756	0.066544	0.038792

* The m_{ij} values for these cells have been derived by the author to complete the table for the purposes of this analysis only. They have no official status with either the ABS or the IMPACT project. The large standard errors attached to the estimates make them highly unreliable for assessments of individual occupations or for other purposes.

$$M_{ij} = \frac{m_{ij}}{L_i}$$

where m_{ij} = gross movement from occupation i to occupation j , $i \neq j$

L_i = labour force in occupation i .

$M_{.j}$ can be regarded as a sample estimate of the corresponding (unknown) population proportion $\mu_{.j}$, with variance

$$\text{Var}\ (M_{.j}) = \frac{\mu_{.j}\ (1-\mu_{.j})}{L_j} \quad \text{for fixed } X_{.j} \ .$$

Since

$$0 \leqslant \mu_{.j} \leqslant 1 , \qquad \frac{\mu_{.j}^{\ 2}}{L_j} \approx 0$$

and

$$\text{Var}\ (M_{.j}) \approx \frac{\mu_{.j}}{L_j} \ .$$

Because there exist only sample estimates for $\mu_{.j}$, the variance of $M_{.j}$ is estimated by

$$\widehat{\text{Var}}\ (M_{.j}) \approx \frac{\hat{\mu}_{.j}}{L_j} = \frac{M_{.j}}{L_j}$$

and the appropriate weighting factor for use in regression analysis of (7) is thus $\sqrt{\frac{L_j}{M_{.j}}}$.

Recall that the analysis has been restricted to cover males only. Empirical results for the weighted and unweighted equations are (quasi - $\bar{R}^2$'s are as previously defined) :

<u>Males 1972</u> :

<u>Unweighted</u>

(6') $M_{.j} = 0.0566 + 0.0078\ W_{.j} - 0.0027\ U_{.j} + 0.0440\ S_{.j} + 0.0014\ Y_{.j}$ + residual

(16.1150) (0.4808) (-4.6711) (3.2937) (1.2941)

$\bar{R}^2 = 0.9379$

Weighted

(7') $$M_{.j} = \underset{(18.9460)}{0.0570} + \underset{(0.3026)}{0.0048}\,W_{.j} - \underset{(-4.4959)}{0.0023}\,U_{.j} + \underset{(2.0313)}{0.0299}\,S_{.j} + \underset{(0.4364)}{0.0006}\,Y_{.j} + \text{residual}$$

$$\bar{R}^2 = 0.9773$$

$$\text{quasi} - \bar{R}^2 = 0.9242$$

Males 1975 :

Unweighted

(6") $$M_{.j} = \underset{(5.1435)}{0.0485} + \underset{(0.3426)}{0.0144}\,W_{.j} - \underset{(-0.6492)}{0.0009}\,U_{.j} + \underset{(1.0082)}{0.0367}\,S_{.j} - \underset{(-0.4460)}{0.0012}\,Y_{.j} + \text{residual}$$

$$\bar{R}^2 = 0.4457$$

Weighted

(7") $$M_{.j} = \underset{(7.0795)}{0.0505} + \underset{(0.4426)}{0.0144}\,W_{.j} - \underset{(-0.3880)}{0.0003}\,U_{.j} + \underset{(0.3798)}{0.0120}\,S_{.j} - \underset{(-1.5410)}{0.0033}\,Y_{.j} + \text{residual}$$

$$\bar{R}^2 = 0.8505$$

$$\text{quasi} - \bar{R}^2 = 0.3310$$

The results for both years are disappointing. For 1972, although relative unemployment is highly significant and has the expected sign (negative), the variables representing social status and years of part-time training required suggest that people tend to move, ceteris paribus, downwards on the social scale, and into areas requiring the most retraining. The high significance[1] of the incorrectly signed social status variable coupled with the low significance of the relative wages variable (which has its a priori expected sign) further discredits the realism of the estimating form adopted. As in the case of the first stage results, the constant term is still the most significant coefficient. Further, it appears that relative unemployment is still the primary variable driving the system.

1. The level of significance adopted is that of 5 per cent.

(b) A similar approach to the treatment of the relative unemployment variable should be taken in section 3.2 as compared to section 3.1. That is, instead of restricting the worker's decision variable to current unemployment only, a weighted average over the last, say, 5 years should be calculated.

(c) Perhaps some experimentation with non-linear estimating forms should be conducted at the stage 2 level.

It is realized that the optimal way to capture the true characteristics of occupational mobility is to model the process with regard to the appropriate underlying structural form. However, the analysis in this paper has demonstrated that the two-stage approach yields plausible results, especially when the affected cells are omitted. It seems that this approach may therefore provide a satisfactory intermediate step in determining factors affecting occupational mobility, especially as it reduces some of the serious data problems imposed by the relatively small sample size and the associated ABS confidentiality requirements.

4. CONCLUSIONS AND FUTURE RESEARCH

This paper set out to test the primary hypothesis that the decision to move from occupation i to occupation j could be estimated in two independent stages. The explanatory power of the two stages were each relatively high, but stage two results were most unsatisfactory, especially with respect to the signs and significance (or lack thereof) of individual coefficients. Combining stage 1 and stage 2 estimates in order to test the primary hypothesis resulted in rather unsatisfactory results. These, however, may be explained by poor individual cell data.

Before completely rejecting this approach, several possible improvements need to be incorporated :

(a) It is suspected that the $W_{.j}$ variable as used in this paper is subject to a high degree of error. For example, no allowance has been made for the different average numbers of hours worked per week in each occupation. This will obviously have an effect on the weekly average wage, yet in the above empirical analysis the wage rates are proxied by average yearly occupational income. It is further suspected that the mid-points of the 13 income ranges as specified in the Income Distribution Survey do not necessarily reflect mean income for that range. Theoretically it is more realistic to think of a worker as estimating the relative occupational wage over a number of years. However, in so far as occupation specific wage data is only available for 1968/69 and 1972/73, this theory is not empirically testable.

The 1975 results have _a priori_ expected signs on all variables except the socio-economic index, but equation (7) as a whole has less explanatory power than the corresponding estimation for 1972. Both relative wages and unemployment are insignificant, with the constant term again being the only highly significant coefficient.

Comparison over the two years reveals no stability whatsoever across coefficients, again suggesting a failure to capture the true underlying structure of the model.[1]

3.3 Amalgamated results

This sub-section tests the hypothesis that the probability of moving from occupation i to j is equal to the product of the probabilities first of moving out of occupation i and then of moving into occupation j .

Recall, $H_1 : E(M_{ij}) = E(M_{i.}) \times E(M_{.j})$, $i \neq j$.

Estimates for $E(M_{i.})$ and $E(M_{.j})$ are derived in sections 3.1 and 3.2. From these it is possible to estimate $E(M_{ij})$ (and hence M_{ij}) for both 1972 and 1975.[2] If H_1 is true, then the differences between M_{ij} (observed) and M_{ij} (estimated) should be close to zero.

In fact, the results are very poor, the correlation coefficients between estimated and actual M_{ij} values for males being 0.3914 and 0.4861 for 1972 and 1975 respectively. Thus empirical evidence does not appear to support H_1 .

---

1. See Appendix 3 for a test of overall stability between the two estimation years.
2. See tables 5 and 6 for the 1972 and 1975 occupation change specific mobility matrices.

At least some part of the failure of H_1 must be attributed to the poor fit obtained in modelling the $M_{.j}$ stage. Perhaps if a more realistic and theoretically based structure were adopted here (as opposed to the ad hoc linear reduced form specification), consequent results would be improved.

However, it appears highly probable that the bulk of the bad results can be related to the underlying poor data used in the model. Both H_2 and H_3 only use marginal totals in their estimating procedures. That is, they are concerned with total movement out of specific occupations and total movement into specific occupations. Thus the numbers concerned are relatively large and can be accurately supplied by the ABS without violation of confidentiality requirements. To test H_1, however, requires knowledge of the individual inner cells of the from/to occupation matrix. These numbers are much smaller and in many cases cannot be released by the ABS. The method adopted by the ABS to complete the matrix involves adding up the magnitudes in all these cells, dividing the total by the number of these cells, and then substituting this 'average' (or some multiple thereof such that aggregates are still in satisfied) in all cells that would otherwise be regarded as confidential. This procedure for eliminating confidential cells may go a long way to explaining the large discrepancy between the estimated and observed elements for M_{ij}.

In an attempt to ascertain the influence of this data problem on the relationship between the actual and estimated M_{ij} values, the correlation between these was re-calculated omitting the affected cells. This involved eliminating 37 of the 72 observations for 1972, and 33 for 1975, which still left 35 and 39 observations for 1972 and 1975 respectively.

There was a significant increase in the correlation coefficient for 1972, which rose from 0.3914 to 0.5481; the effect was not as dramatic for 1975, where the rise was from 0.4861 to 0.5685.[1] Overall these results support the claim that the data problem is a significant one.

In view of the above, it may be more accurate to use the estimated M_{ij} values (adjusted if necessary to observe the constraints imposed by the marginal sub-totals) as a measure of actual inter-occupational rates of movement for those cells which currently contain "dummy" values. Given a firm prior acceptance of H_1, this approach can be justified in view of the large sampling standard errors associated with at least half of the cells, a problem which may be further exacerbated by the distortions introduced as a result of the confidentiality restrictions observed by the ABS.

1. Given that the data is cross-sectional, these $\bar{R}^2$ values are quite high.