MONASH 11/94

MONASH UNIVERSITY

AUSTRALIA

MODELLING THE PROBABILITY OF YOUTH UNEMPLOYMENT IN AUSTRALIA: 1985-1988

Mark N. Harris

Working Paper No. 11/94

July 1994

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0752 3

MODELLING THE PROBABILITY OF YOUTH UNEMPLOYMENT IN AUSTRALIA: 1985-1988

Mark N. Harris

Working Paper No. 11/94

July 1994

DEPARTMENT OF ECONOMETRICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

MODELLING THE PROBABILITY OF YOUTH UNEMPLOYMENT IN AUSTRALIA: 1985 - 1988

Mark N. Harris*
Dept. of Econometrics
Monash University
Clayton
Vic., 3186

ABSTRACT

This paper attempts to explain how particular personal characteristics affect the probability of unemployment of Australian youth. For this purpose the Australian Longitudinal Survey (1985-1988) is utilised and a Univariate Equicorrelated Random Effects Probit model applied to the data. The Survey appears to be affected by endogenous attrition, the source of which was found to be nationality and education levels. These processes were accounted for in the estimation procedures. This study appears to be the first attempt to analyse this particular data as a panel data set in a Random Utility Discrete Choice context. Results indicate that age, education and financial housing commitments exert a positive influence on the probability of unemployment. Also, there is evidence to suggest that the disabled are discriminated against, and that reservation wages exert a strong negative effect.

KEY WORDS

Panel data, attrition bias, labour supply functions, work-force discrimination, fixed and random effects models, discrete data and univariate probit.

J.E.L Classifications: C23, C25, J24.

Jun-94

---

* I am grateful to Professor P. Dixon, Centre of Policy Studies, Monash University, for kind use of the data. I am also grateful to Dr. L. Mátyás, Dr. T. Fry, Associate Professor K. McLaren and Ms A. Hosking, who provided helpful comments on earlier versions of this paper. Any remaining errors are my own.

1. INTRODUCTION.

Attention in Australia has once more turned to the issue of unemployment with the recent release of the Federal Government's White Paper *Working Nation*[1]. Currently unemployment in Australia is estimated to be 11.1%.[2] This rate is of great social and political concern. Quantifying the causes of unemployment appears imperative, especially for policy makers.

The youth are potentially the most vulnerable to unemployment. This is unfortunate, since an early history of unemployment is likely to have adverse effects on an individual's subsequent employment prospects. This employment "scar" may mentally affect an individual (a *work ethic* in reverse) and make him/her less attractive in some sense to potential employers.

An *aggregate* quantitative analysis of youth unemployment would not appear appropriate, as it is an individual's particular characteristics that will affect his/her willingness to supply labour, or make him/her relatively "employable". A qualitative study based on *individual* data may be more appropriate. The Australian Longitudinal Survey (ALS) provides us with such a data set, focussing on the youth of Australia (persons aged between 16 and 25).

The survey contains a plethora of information on the individuals' personal and family attributes. It was first undertaken in 1985 and then subsequently annually until 1991. However, it only consistently tracks the same individuals until 1988 (see Section 3 below). Initially the data set consisted of 8998 respondents, which fell to 6151 by 1988.

The plan of this paper is as follows. In Section 2, panel data sets and panel models are discussed. Section 3 describes the data set used. The empirical results are given in Section 4. Finally, some conclusions are drawn in the final section.

---

[1] May, 1994.
[2] Source: Australian Bureau of Statistics, March 1994.

2. PANEL DATA SETS AND PANEL MODELS.

2.1. Panel Data Sets.

A major problem with many econometric models using aggregated data can be traced directly to the aggregation itself. For example, econometric results may paradoxically tend to contradict the underlying economic theory (Hsiao [1986]).

Panel data sets avoid this problem by observing individuals (countries, firms, households, persons, or indeed any other economic unit) not only at one particular point in time (i.e. cross-sectional data), but also the same individuals over time. If N is defined as the total number of individuals, T as the number of time periods and K as the number of explanatory variables, then the data set immediately assumes the enlarged dimensions of $(NT \times 1)$ observations on the dependent variable and $(NT \times K)$ on the X-matrix.

Thus the very nature of panel data sets increases the richness of information available to the applied researcher. This increased information potentially yields more correctly specified econometric models, avoiding aggregation bias and providing better inferences about hypotheses of interest (see Hsiao [1985]).

Survey panel data sets are typically expensive and fraught with practical problems. A major difficulty arises from the fact that respondents often fail to reply for the full duration of the time period (often this can be attributed to individual lethargy, movement away from usual place of abode and so on). This process is known as *panel attrition*. As a result the initial cross-section component of the data sets tend to be relatively "large" which diminishes fairly rapidly over time. Consequently, the time-series component tends to be relatively "short". Cost reasons may also influence the duration of the survey.

2.2. Panel Models: General.

A convenient starting block is the *Classical Linear Regression* model (CLR). For each individual, i, there are T observations on the dependent variable in question. Stacking these individual vectors yields a dependent variable vector of dimension $(NT \times 1)$. Similarly stacking the observations of the individuals' specific X-matrix (X_i), yields a full X-matrix of dimensions $(NT \times K)$. That is, let:

$$(1) \qquad y = \begin{pmatrix} y_{11} \\ \cdot \\ \cdot \\ y_{1T} \\ \cdot \\ \cdot \\ y_{N1} \\ \cdot \\ \cdot \\ y_{NT} \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} x_{11}^2 & \cdot & \cdot & x_{11}^K \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{NT}^2 & \cdot & \cdot & x_{NT}^K \end{pmatrix}.$$

Just as with the CLR model, y is assumed to be a linear function of the X-matrix, with unknown regression parameters, β. That is:

$$(2) \qquad y_i = \alpha + X_i\beta + \varepsilon_i,$$

where: ε is a $(T \times 1)$ vector of error terms (for each individual).

It is likely that α will vary either across individuals (or groups of "similar" individuals) and /or over time. These are respectively known as *individual effects* and *time effects*. Indeed, if there were no individual or time effects then there is said to be *perfect homogeneity* and usual regression techniques would be appropriate, yielding consistent and efficient estimates of the parameter vector, β.

Individual and time effects are generally incorporated into the framework of (2) in one of two ways. *Fixed effects* models involve including the above effects directly into the structural part of the model specification, whereas *random effects* models include them as additional random disturbances.

2.3. Panel Models: Qualitative Dependent Variables.

The key question of interest in this study is to examine the likelihood of an individual being in a particular state in the labour market (employed or unemployed).[3] Thus it is the *probability* of a particular state occurring that is of interest. Therefore the usual OLS techniques are not appropriate, primarily for the reason that predicted values are not bounded to lie in the zero-one interval.

For the time being it is assumed that the available data is cross-sectional. Now, by defining R, as an index of "risk" of employment, then for each individual this will be a function of their personal attributes, x_i, with associated coefficients, β (as well as any demand side factors).[4] If, for simplicity, this function is assumed to be linear then:

$$(3) \qquad R_i = x_i\beta + u_i,$$

where: R_i is the risk of employment faced by the ith individual,
x_i is a row vector of personal characteristics,
β a column vector of coefficients corresponding to x_i
and u_i are error terms, representing random individual behaviour.

The observed realisation, Y_i, of the latent variable, R_i, is:

$$(4) \qquad Y = \begin{cases} 1 \text{ if } R > 0, \\ 0 \text{ if } R \leq 0. \end{cases}$$

Defining the two states as employed ($Y = 1$) and unemployed ($Y = 0$) respectively, then if there is a *positive* chance of an individual being employed (as determined by his/her characteristics), then that individual is more likely to be employed than unemployed. Conversely, a *negative* risk of employment and the individual is more likely to be unemployed. The probability that an individual will be employed is thus:

[3] And also the extent to which identified attributes are policy variant, that is under the control of the authorities.

[4] Note that here none of the variables are observed over time, *i.e.* more than once.

(5) $$Prob(Y_i = 1) = Prob(R_i > 0) = Prob(u_i > -x_i\beta).$$

Differing assumptions about the distributional form of u_i, give rise to different functional forms for these probabilities. If they are assumed to independent random drawings from an Extreme Value distribution, then a Logit model results, whereas a Probit model arises from an assumption of (standard) normality. Probabilities yielded from both models are very similar, apart from regions close to either unity or zero (since at these regions the cumulative normal distribution tends to its limit at a faster rate than the logistic distribution).

These models are appropriate when an individual is observed only once. As soon as the economic units are observed over time, these models need to be augmented to allow for any potential individual heterogeneity. For these purposes, there exist discrete choice counterparts to the regression fixed and random effects models outlined above.

i. Fixed Effects Models for Discrete Data.

Fixed effects models generally assume that the differences between economic units can be captured by differences in the constant term. It is also possible to allow the *slopes* to vary across individuals, however this severely increases the complexity of the problem and is therefore less popular than the former.

Equation (3) now becomes:

(6) $$R_{it} = \alpha_i + x_{it}\beta + u_{it}.$$

That is, the basic linear model is augmented by an additional constant term, α_i, for each individual. The index, t, indicates that individuals are now observed over time. The probability that individual i is employed in time period t, P_{it}, will now be:

(7) $$Prob(Y_{it} = 1) = Prob(u_{it} > -x_{it}\beta - \alpha_i) = \int_{-x_{it}\beta}^{\infty} f(u_{it})du_{it} = 1 - F(-x_{it}\beta - \alpha_i).$$

Note that the α's are individual specific and by definition *constant*, *i.e.* time invariant.

From (7) it is clear that both the α's and the β vector require estimation. The data set utilised in this study is fairly "typical" in that it contains a large cross-sectional component (the first wave, 1985, contains 8998 respondents) but a small time series component (T = 4). This creates two inter-related problems. Firstly, N separate α parameters need to be estimated. This is likely to be computationally time-consuming and unlikely to be a parsimonious representation of the data. Secondly, only t = 1,..,T observations on Y_{it} are available for inference concerning α_i. Increasing the cross-sectional sample size gives no additional information, and indeed, increases the number of parameters to be estimated. Conventional estimation by Maximum Likelihood (MLE) provides *consistent estimators only as T tends to infinity*. Thus for "small" T, MLE of α_i will be inconsistent. Moreover, as β and α are not independent, then MLE of β will be similarly inconsistent.[5]

Finally, if α_i are in fact random, but treated as fixed, then a significant loss of information is imposed. At best there will be a loss of efficiency in the estimation of $\underline{\beta}$. At worst, the fixed effects estimators will be inconsistent.[6] For these reasons, estimation of a Fixed Effects model was not considered.

ii. Random Effects Models for Discrete Data.

Fixed effects models inherently assume that differences across individuals can be represented by parametric shifts in their behavioural regression functions. However, in certain instances it may be more realistic to view these differences as simply being randomly distributed across cross-sectional units. These random disturbances can be rationalised as factors specific to the individual (or group of "like" individuals) that cannot be explicitly entered into the model. For example, they may relate to factors that are unobservable, or data which is simply not available to the researcher.

Thus the α terms of (6) are no longer constant, but are a random sampling from a univariate distribution, G, indexed by a finite number of parameters, δ. The log-likelihood for the random effects model is then:

---

[5] See Hsiao (1986) chapter 7.
[6] *Ibid.*

(8) $$log\ L = \sum_{i=1}^{N} log \int \Pi_{t=1}^{T} F\left(\underline{x}_{it}\ \underline{\beta} + \alpha_i \right)^{y_{it}} \left[1 - F\left(\underline{x}_{it}\ \underline{\beta} + \alpha_i \right) \right]^{1-y_{it}} d\,G\left(\alpha_i \mid \underline{\delta} \right),$$

which is a function of $\left(\underline{\beta}', \underline{\delta}'\right)$ parameters. Maximisation of (8) yields both consistent and efficient parameter estimates (under weak regularity conditions) for both β and δ.

If the distributional form for F is assumed to be standard normal and that for G is similarly normal, but with variance σ_α^2, then this specification yields a Probit model of the form:

(9) $$Y_{it} = 1 \quad if \quad x_{it}\beta + \alpha_i + u_{it} > 0.$$

As the α and u terms are independent and normally distributed, then the composite "error" term, $\varepsilon_{it} = \alpha_i + u_{it}$, is similarly normally distributed, with mean $\underline{0}$ and variance, $Var(\varepsilon_{it}) = \sigma_\alpha^2 + \sigma_u^2$. However, the Butler and Moffitt (1982) *equicorrelated* model additionally allows for equivalent (hence the name) correlation between successive (composite) error terms over time for each individual.[7] That is $Corr^2(\varepsilon_{it}, \varepsilon_{is}) = \rho^2 = \sigma_\alpha^2 / (\sigma_u^2 + \sigma_\alpha^2)$, $i \neq s$. Under this formulation the log-likelihood function of (8) becomes:

(10) $$log\, L = \sum_i log\, L_i,$$

where: $$L_i = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} e^{-\varepsilon_i^2/2} \prod_t \Phi(r_{it}\ z_{it}) d\varepsilon_i,$$

$$r_{it} = 2Y_{it} - 1$$

and: $$z_{it} = (\ x_{it}\beta + \rho\varepsilon_i) / (1-\rho^2)^{1/2}.$$

Make the change of variable $u_i = \varepsilon_i / \sqrt{2}$, (10) then becomes:

[7] The following draws heavily on the Greene (1992), pp. 439-440.

(11) $$L_i = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-u_i^2} \prod_t \Phi(r_{it}\, w_{it}) du_i,$$

where: $w_{it} = \left(x_{it}\beta + 2\rho\, u_i \right) / \left(1-\rho^2\right)^{1/2}$.

Now letting:

(12) $$\gamma = \beta / \left(1-\rho^2\right)^{1/2}$$

and:

(13) $$\theta = \sqrt{2\rho} / \left(1-\rho^2\right)^{1/2},$$

then w_{it} in (11) can be written:

(14) $$w_{it} = x_{it}\gamma + \theta\, u_i.$$

Parameter estimates are obtained by maximising the log-likelihood function over γ and θ. Thus the score vector is:

(15) $$\frac{\partial L_i}{\partial \psi} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-u_i^2} \left(\sum_t r_{it}\, \lambda_{it} \right) \left(\prod_t \Phi(r_{it}\, w_{it}) \right) (Z_{it}) d\, u_i$$

where: $\psi = (\gamma', \rho)$

$Z_{it} = (x_{it}, u_i)$

and $\lambda_{it} = \phi(r_{it}\, w_{it}) / \Phi(r_{it}\, w_{it})$, and parameter estimates are obtained by setting (15) equal to zero.

Once γ and θ are computed then the original parameters can be recovered such that:

$$(16) \quad \rho^2 = \theta^2 / (2 + \theta^2)$$

and:

$$(17) \quad \beta = (1 - \rho^2)^{1/2} \gamma.$$

The problem with this specification is that it involves the evaluation of *T*-dimension integrals. Butler and Moffitt (1982) condition on the permanent component of the likelihood function to reduce the expression to a single integral, whose integrand is a product of one normal density, and *T* differences of normal cumulative distribution functions, for which there are available highly accurate approximations. They suggest Gaussian quadrature, requiring Hermite integration, for the evaluation of this single integral. Indeed, this is the algorithm employed by LIMDEP[8] the estimation package used in this study, which uses an eight point quadrature.

3. THE DATA.

The data set used in this study is from the ALS from years 1985 to 1988. Later years were not considered for the following reasons. Firstly, there is quite severe *natural* attrition.[9] Secondly, there is *enforced* attrition, in that from 1989 onwards, respondents are voluntarily retired from the panel - immediately instigating attrition bias into the sample from these years onwards. Thirdly, there were changes to the method of interview due to cost considerations. From 1985 to 1988 the interviews were on a person-to-person basis. In 1989 the interview was self-completion (mail out, mail back) and the final years witnessed a further change to telephone interview. Finally, the later surveys witness a significant decline in the range and number of

[8] Econometric Software, Inc., New York.
[9] For example, the original data set of 1985 had 8998 respondents, which fell to 6151 by the 1988 survey.

questions asked. Therefore it may be impossible to track an individual's response to a particular question past 1988.

The survey data and sampling procedures are described in detail in Harris (1993). The key characteristics are summarised below.

3.1. The Australian Longitudinal Survey.

The *Australian Labour Force Survey* (ALS) is a survey focusing on the youth labour market of Australia, defined as persons aged between 16 and 25 years of age on September 21st, 1985. The sample is a stratified multi-stage (three-stage) equal probability sample.

The primary objective of the above study was to collect information concerning the dynamics of the youth labour market. This necessarily involved repeating the survey over successive periods, yielding information that would not otherwise be available in the more usual cross-sectional sources. Such a procedure defines the ALS as a *panel* study in that the same individuals are tracked over time. The data is stored by the Social Science Data Archives, Australian National University, Canberra, A.C.T..

The sample was designed to adequately represent all Australians in this particular age cohort. Due to the particular population distribution within Australia, the survey was designed only to be representative of Australians not in "sparsely" populated regions.[10]

3.2. Dimensions of the Data Set.

The 1985 survey contains some 800 responses to questions by approximately 9,000 respondents. Even in the later survey, where the number of respondents decreases somewhat, the total number of responses is still very large. This study utilises only a relatively small proportion of all the available information (approximately 30 responses are used) and only considers those respondents who remained in the survey up until 1988, yet this still yields a total X-matrix of dimensions $(NT \times K)$, that is over 700,000 elements![11]

---

[10] Where a sparsely populated area is defined as one where the population density is 2% or less.
[11] Indeed, this is paradoxically somewhat problematic as the handling/estimation of such large data sets becomes very unwieldy - a problem not often encountered in applied econometrics.

3.3. *Variables Available within the Data Set.*

The ALS focuses primarily on labour market dynamics, hence labour market variables form the bulk of the questionnaire. However, there are also background variables, and variables indirectly related to the dominating labour market theme. Finally, there are also general demographic variables.

The labour market variables partly consist of: the number and length of any employment and unemployment spells; the nature of the employment and method and amount of payment received; attitudes towards workplace conditions (job satisfaction in general); trade union membership; workplace/job training; the nature of any job search; and the level of reservation wages, to name but a few.

Related variables include: level of educational achievement; any health disabilities; and the nature of the transition from school to the workplace. Background and demographic variables include: age; sex; racial origin; country of birth; marital status; partner's employment status/income/educational attainment; number of dependent children; and place and nature of occupancy.

3.4. *Data Description.*

Summary descriptive statistics for some of the variables retained/used in this study are reported in Appendix I, Tables 2 to 5. The (original) coding of these variables is reported in Table 1, Appendix I. Due to the categorical nature of many of these variables, their description is quite often aided by histograms. Examples of these can be found in Charts 1 to 6 in Appendix II.

Note that although in some instances there may appear to be a large number of *missing observations*, in many cases these will not be "true" missing values but simply not applicable. For example, if the respondent is still at school, his/her educational achievement will be recorded as missing.

Several pertinent facts can be drawn from the above charts and statistics. Over the four years the data set is subject to quite severe *attrition*, falling from 8998 individuals in 1985 to 6151 in 1988. As the study tracks the same individuals, it is not surprising that average age increases over the years. Moreover, as a consequence of the cohort aging, the number of marriages (and de-facto marriages) and the number of children increases over the sample (taking into account the attrition). These *growing-up* effects are also present in the accommodation variable. This variable exhibits a modal category shift from *boarder* to *renter*, as the young adults leave the parental abode and move (in the first instance) into rented premises.

The split between male and female and the place of residence appears to diminish fairly evenly. Individuals in each employment category also appear to fall fairly evenly across the years, although the summary statistics for employment show a slight increase, indicating a slight rise in employment. The rise in employment is presumably attributable to those entering the labour force as they finish their studies/schooling. The movement away from full-time schooling would also explain the rise in the mean of the part-time/full-time study category (which marks a movement away from full-time study). Finally, Chart 7 illustrates that although the lower levels of highest educational achievement (*i.e.* Grade/Year 12 or below) decline with general attrition rates, the higher levels exhibit a relative rise as the aging youth progress on to, and complete, their higher education (see also Section 4.2 *vi* below).

4. EMPIRICAL ANALYSIS.

4.1. Variables Considered Important in Determining Employment.

i. Human capital.

A positive relationship between the probability of employment and educational achievement is expected. Greater knowledge of labour markets, more experience and a perceived more conscientious attitude towards work by employers, all increase with age. Therefore, once more a positive relationship is expected to exist between the probability of employment and a person's age.

ii. Financial Commitments.

The extent to which an individual has financial commitments is expected to exert a positive influence on his/her job search if unemployed, or to discourage voluntary withdrawal of labour services if within employment. Variables considered within this context are housing commitments, marriage and the number of dependent children.

iii. Place of Residence.

Labour market theory suggests that, at least in the medium run, any regional imbalances in the level of unemployment rates will, via competition and labour mobility, be erased (up to the cost of inter-regional migration differentials). However, it is quite possible that in the short run one's place of abode (in the sense of an urban or rural dwelling) will affect employment prospects. Indeed, in a survey primarily focussing on *supply-side* factors, such a variable could provide for a valuable *demand-side* effect.

iv. Inherited Personal Characteristics.

There are several personal characteristics recorded in the survey which may affect one's employment prospects. For example, Aborigines may be subject to racial discrimination, as may other racial and minority groups (for example the physically disabled).

v. Reservation Wages and Unemployment Benefit.

The lowest wage an individual is prepared to work for, **relative** to the level of unemployment benefit he/she would be entitled to is expected to exert a strong negative effect on an individual's willingness to supply labour services. Indeed, along with *place of residence*, this provides an additional demand-side variable.

vi. Partner's Employment Status.

There is evidence to suggest that partners exhibit similar labour market status, especially for childless couples.[12] This phenomenon may arise from complementary leisure times, or simply that partners may tend to have similar attitudes towards the *work ethic*.

---

12 See for example Bureau of Labour Market Research (1986).

4.2. Issues in Estimation.

i. Re-coding of Variables.

Due to the categorical nature of most of the explanatory variables, they were re-coded into a series of zero-one dummy variables if there were no natural ordering in the series. For example, if the *marital status* variable was entered as it is coded in Appendix I, Table 1, it is not possible to make any distinction between the inherent sub-categories (married, single, divorced *etc.*). To avoid this problem variables were sub-divided into zero-one indicator variables indicating married or not, single or not, and so on. In this way it is possible to ascertain the significance and direction of causality between the various sub-categories and the dependent variable.

Note that if an original variable is mutually exclusive and exhaustive, then one of these constructed variables must be excluded from the estimation procedure, as otherwise perfect multicollinearity would result. The omitted category is generally the least common one, or one not thought to be influential in determining the probability of employment *a priori*.

In other instances, where the original variable contains a natural ordering (*e.g.* the *number of dependent children* and *reservation wage*) no re-coding was necessary. Here one can conclude that a (significantly) positive coefficient on the *number of dependent children* variable, for example, implies that as the number of children increases so does the probability of employment (if the dependent variable is coded as 1=employed and 0 otherwise).[13]

Missing values may have been recorded for variables simply because a question was not relevant and therefore not asked. A distinction was made in the re-coding of variables between true missing values and those simply not appropriate and recorded as missing. For instance *partner's employment status* assumes a value of 1 if the respondent has a partner **and** that the partner is employed, zero if the respondent has no partner or a partner who is not employed and missing otherwise.

[13] This kind of procedure is standard practice for models with discrete data, see Miller (1989). Note also, that inclusion of variables containing a natural ordering, assumes that their affect on employment is the same for all incremental changes in this (these) variable(s).

ii. Definition of the Age Variable.

The age variable was entered in a non-linear (exponential) form in accordance with existing knowledge of the relationship between age and the incidence of unemployment.[14] Thus age was entered in the form; age=$exp^{(-0.1*\text{age})}$. Due to this parameterisation, a negative relationship between age and employment is expected.

iii. Allowing Individuals to Enter & Leave the Labour Force.

This study is interested in the binary outcome of an individual's employment status. However, if the only individuals selected are those that are either in work or unemployed, then this would not allow for individuals who were at some stage either discouraged or not in the workforce, to enter the workforce at a later stage. A good example of this would be students who would not initially be in the labour force. By removing these individuals from the sample, it is possible that attrition bias would be instigated into the parameter estimates.

To allow movement into (and out from) the labour force, individuals were removed from the sample if, and only if, they were not in the labour force for the full duration of the survey. For others who experienced only periods outside of the labour force, their *x*-matrix was *dummied-out* during these periods and therefore not allowed to influence the parameter estimates.

iv. Defining the Reservation Wage.

The reservation wage was defined as the ratio of the reservation wage to the appropriate level of unemployment benefit. The latter varied according to age, marital status and number of children, over each of the years considered.[15]

v. Parametric Shifts in the Probit Functions for Male & Females.

It was decided that the differences between male and female labour supply functions would not be adequately represented by "shift" variables. Therefore a separate

[14] See for example Miller (1989).
[15] Source: Department of Social Security, Outreach & Information Services.

equation for each sex was estimated, allowing for "slope" coefficients to vary across sex (see also point *vi.* below).

vi. Balanced & Unbalanced Data Sets & Panel Attrition.

Estimation is facilitated by use of a *balanced* data set. That is, the number of periods observed must be the same for each individual. For estimation purposes this necessitates removal of the those respondents initially in the survey who later drop out. This de-selection process can be considered either endogenous or exogenous. If the latter is the case, then it is valid to remove the cases not observed for the full duration, with the only complication being the potential loss of information. However, if the de-selection process is endogenous, it is likely that "similar" groups (similar in the sense of possessing particular characteristics) will de-select themselves from the sample. This instigates selectivity bias into subsequent parameter estimates if not accounted for.

A method of examining the endogeneity of the attrition is to define an indicator variable as unity if the respondent remains in the survey throughout, and zero if he/she is not. A probit model is then used to see if the attrition can be attributed to any of the exogenous variables considered in the wider study (*i.e.* that of the probability of employment). As the latter will only be observed in the first period for those respondents who de-select themselves, only a cross-section probit is required. If variables are positively significant, then this suggests that respondents who have these particular characteristics are more likely to remain in the survey and therefore be over-represented (and *vice versa* for negatively significant variables).

If this is the case, then (to avoid biased estimators) it is necessary to estimate separately on the identified sub-groups. Indeed, this was found to be the case, with the identified sub-groups being Australian born/non-Australian born and high education/low education for both males and females.[16] Note that this implies that the disproportionate decline within the education variable noted above (see Section 3.4 above), cannot be solely attributed to an aging and more educated sample population. Note also that the (exponential) age variable was generally significantly positive, however being a quasi-time trend, this was interpreted as the extent of "natural" (*i.e.* exogenous) attrition. Finally, those respondents who were originally in the labour market and possessed a "low education" status, and at a later stage (within the survey)

[16] "Low education" defined as Year 11 or below.

recorded a "high educational" status were removed from the data as they could not consistently be considered as possessing either a "low" or "high" educational status.[17]

At this juncture it was decided to concentrate solely on the Australian born groups as the sample sizes for the non-Australian born groups were too small. Thus often there were insufficient responses to adequately base inference on.

vii. Pre-Test Bias & Preferred Specifications.

To simultaneously avoid pre-test bias and inclusion of insignificant variables, equations were estimated once and insignificant variables were removed in the second (and final) stage of estimation. Likelihood ratio tests were employed to test the validity of these implied restrictions.

4.3. The Empirical Results.

The empirical results for the estimated male and female high and low education equicorrelated random effects probit equations are reported in Tables 6 and 9 below.

The partial derivative of the probability of employment with respect to an independent variable is difficult to interpret when the latter is categorical. This process is aided however, by the re-coding procedure employed in this study (see Section 4.2 *i*). In this way, a direct comparison of both the signs and magnitudes of estimated coefficients of variables (that are coded as indicator variables) is possible. For example, one can now directly compare the significance of living in the city relative to living in a rural area, as opposed to a conclusion solely as to the significance of the composite *place of residence* variable.

[17] See the section on *Issues in Estimation* above for the treatment of individuals while studying.

Table 6: Maximum Likelihood Parameter Estimates and Summary Statistics for the Equicorrelated Random Effects Probit Model: ***High Education Australian Born Males*** *(ALS 1985 - 1988, Number of Time Periods = 4, Number of Individuals = 1549). Second Stage Estimation Results Arise from Removal of Insignificant Variables in the Initial Estimation.*

	First Estimation		*Second Estimation*	
Variable	***Coefficient***	***t-statistic***	***Coefficient***	***t-statistic***
Constant	1.0444	2.619	1.797	10.205
Age (=$exp^{(-0.1*age)}$)	-4.4490	-3.607	-5.048	-4.316
Married	-0.2391	-1.343	-	-
Divorced/ Separated	-0.5667	-1.894	-	-
City Dwelling	0.1558	1.895	-	-
Rural Dwelling	0.2509	1.725	-	-
Buying House	0.7734	3.807	0.7448	4.221
Rent Free Accommodation	0.2056	2.227	0.1881	2.126
Renting Accommodation	0.3800	4.376	0.3693	4.377
Aboriginal/ Torres Strait	0.9188	0.787	-	-
Western Origin	0.4326	1.207	-	-
Asian Origin	0.3753	0.481	-	-
Year 12	0.1409	1.778	-	-
Degree	0.6419	3.598	0.5221	3.183
Diploma	0.3985	3.414	0.2890	2.835
Trade Qualification	0.1517	1.403	-	-
Partner Employment Status	0.5335	2.596	0.3406	2.356
Relative Reservation Wage	-0.5953	-16.329	-0.5868	-16.227
Disabled	-0.3216	-3.099	-0.32981	-3.315
Number of Children	0.1951	1.050	-	-
Correlation coefficient, $\rho_{\varepsilon_{it},\varepsilon_{is}}$	0.7445	29.034	0.7478	29.824
Max. Log-Likelihood	-1021.422		-1030.055	
Likelihood Ratio Statistic	17.266			
Critical value, $\chi^2_{10,0.05}$	18.307			

Note: *Coefficients are asymptotically (standard) normally distributed.*

Table 7: Maximum Likelihood Parameter Estimates and Summary Statistics for the Equicorrelated Random Effects Probit Model: ***Low Education Australian Born Males*** *(ALS 1985 - 1988, Number of Time Periods = 4, Number of Individuals = 619). Second Stage Estimation Results Arise from Removal of Insignificant Variables in the Initial Estimation.*

	First Estimation		*Second Estimation*	
Variable	***Coefficient***	***t-statistic***	***Coefficient***	***t-statistic***
Constant	1.294	1.824	1.2118	6.176
Age (=*exp*$^{(-0.1*\text{age})}$)	-3.848	-2.0847	-3.8792	-3.096
Married	-0.3587	-2.352	-0.4200	-3.756
Divorced/ Separated	-0.4087	-1.829	-0.4505	-2.056
City Dwelling	-0.0033	-0.039	-	-
Rural Dwelling	0.2346	1.613	0.2208	1.610
Buying House	0.3484	1.880	0.4316	2.742
Rent Free Accommodation	-0.2264	-1.465	-	-
Renting Accommodation	-0.0800	-0.638	-	-
Aboriginal/ Torres Strait	-0.7978	-1.139	-	-
Western Origin	0.0357	0.053	-	-
Asian Origin	1.7842	0.010	-	-
Year 10	0.3607	3.041	0.3734	3.151
Year 11	0.8483	5.651	0.8426	5.779
Partner Employment Status	0.8660	3.551	0.9083	4.105
Relative Reservation Wage	-0.6233	-14.022	-0.6317	-14.646
Disabled	-0.3913	-3.779	-0.37287	-3.658
Number of Children	-0.0401	-0.418	-	-
Correlation coefficient, $\rho_{\varepsilon_{it},\varepsilon_{is}}$	0.7569	26.818	0.7598	27.258
Max. Log-Likelihood	-744.8353		-751.5396	
Likelihood Ratio Statistic	13.4086			
Critical value, $\chi^2_{7,0.05}$	14.0671			

Note: *Coefficients are asymptotically (standard) normally distributed.*

Table 8: Maximum Likelihood Parameter Estimates and Summary Statistics for the Equicorrelated Random Effects Probit Model: ***High Education Australian Born Females*** *(ALS 1985 - 1988, Number of Time Periods = 4, Number of Individuals = 1494). Second Stage Estimation Results Arise from Removal of Insignificant Variables in the Initial Estimation.*

	First Estimation		*Second Estimation*	
Variable	*Coefficient*	*t-statistic*	*Coefficient*	*t-statistic*
Constant	2.1869	4.198	2.4867	13.438
Age (=$exp^{(-0.1*age)}$)	-7.2771	-5.593	-8.8389	-7.922
Married	-0.2996	-1.506	-	-
Divorced/ Separated	-0.5158	-2.320	-0.5107	-2.455
City Dwelling	0.1819	2.100	0.1850	2.428
Rural Dwelling	0.0531	0.420	-	-
Buying House	0.4221	2.599	0.2930	2.144
Rent Free Accommodation	-0.0281	-0.290	-	-
Renting Accommodation	0.1323	1.398	-	-
Aboriginal/ Torres Strait	-0.5094	-0.905	-	-
Western Origin	-0.0577	-0.126	-	-
Asian Origin	0.0325	0.050	-	-
Year 12	0.0842	1.022	-	-
Degree	0.2959	1.647	-	-
Diploma	0.0977	1.006	-	-
Trade Qualification	0.0945	0.525	-	-
Partner Employment Status	0.4864	2.294	0.2040	1.980
Relative Reservation Wage	-0.3047	-12.465	-0.3025	-12.709
Disabled	-0.3472	-3.326	-0.3458	-3.373
Number of Children	-0.4469	-5.996	-0.4764	-6.611
Correlation coefficient, $\rho_{\varepsilon_{it},\varepsilon_{is}}$	0.6831	19.984	0.6855	21.279
Max. Log-Likelihood	-1073.068		-1079.686	
Likelihood Ratio Statistic	13.236			
Critical value, $\chi^2_{11,0.05}$	19.6751			

Note: *Coefficients are asymptotically(standard) normally distributed.*

Table 9: Maximum Likelihood Parameter Estimates and Summary Statistics for the Equicorrelated Random Effects Probit Model: ***Low Education Australian Born Females*** *(ALS 1985 - 1988, Number of Time Periods = 4, Number of Individuals = 566). Second Stage Estimation Results Arise from Removal of Insignificant Variables in the Initial Estimation.*

	First Estimation		*Second Estimation*	
Variable	***Coefficient***	***t-statistic***	***Coefficient***	***t-statistic***
Constant	0.9255	2.102	1.4635	5.880
Age (=$exp^{(-0.1*age)}$)	-5.8802	-3.612	-6.0539	-4.019
Married	-0.7116	-3.562	-0.6343	-3.364
Divorced/ Separated	-0.2095	-0.806	-	-
City Dwelling	-0.0754	-0.780	-	-
Rural Dwelling	0.1547	0.909	-	-
Buying House	0.3882	1.584	-	-
Rent Free Accommodation	-0.1478	-0.690	-	-
Renting Accommodation	0.0271	0.143	-	-
Aboriginal/ Torres Strait	0.0853	0.180	-	-
Western Origin	0.6095	1.791	-	-
Asian Origin*	-	-	-	-
Year 10	0.2913	2.019	0.3195	2.341
Year 11	0.5725	3.533	0.6263	4.063
Partner Employment Status	0.8233	4.065	0.8771	4.552
Relative Reservation Wage	-0.4069	-10.973	-0.4037	-11.518
Disabled	-0.2980	-2.629	-0.3119	-2.875
Number of Children	-0.2182	-3.058	-0.2097	-3.010
Correlation coefficient, $\rho_{\varepsilon_{it},\varepsilon_{is}}$	0.7713	23.881	0.7840	25.392
Max. Log-Likelihood	-587.0839		-594.3821	
Likelihood Ratio Statistic	14.5964			
Critical value, $\chi^2_{8,0.05}$	15.5073			

* *No observations recorded*

Note: *Coefficients are asymptotically (standard) normally distributed.*

i. The Male Equations.

The constant term for both low and high education males is significantly positive (Tables 6 and 7). An interpretation of this is that the constants are related to the *natural rate of employment* for these respective youth groups. As one might expect, this rate appeared to be higher in the high education group.[18] Both groups also exhibit increasing probability of employment with age, as expected *a priori*. Again, this proxy for experience tends to be more pronounced in the high education group.

The marital status of the high education group did not appear to influence the probability of employment. However, in the low education group the *married* indicator variable is significantly negative, indicating that marriage has a negative effect on the males' employment prospects, or indeed on his willingness to work. Also, the *divorced* indicator is similarly significantly negative, indicating that once the financial commitments of marriage are released, these males were less likely to work.

Somewhat surprisingly, males did not appear to embrace parenthood with an increased drive for employment, as the *number of dependent children* variable was comprehensively insignificant in both educational groups.

Demand for high education labour did not appear to vary between rural and city locations as these variables were insignificant. Demand for low education labour however, appeared to be stronger in rural locations. Employment in rural locations, being predominantly agricultural and service orientated, may well be more suited to those less well educated. These effects may however, be confused by commuting.

In terms of type of accommodation, the financial commitment of buying a house appeared to exert a strong positive effect on employment in both groups. Moreover, this was the only influence the type of accommodation had on the low education group. In the high education group, renting also exerted an impact again due to its financial commitment, but to a lesser degree than buying. Finally, somewhat surprisingly, living rent free also exerted a positive impact to employment prospects in this group, albeit a very small one.

Previous studies (see Miller [1989] for example), have found evidence of racial disadvantage in the workplace *using the same data set*. However, once the effects of

[18] Without considering the full functional form for the probabilities, this *natural rate of employment* is not bounded to lie in the zero-one interval (see Section 4.3, iv below). Suffice it to say that a higher value for the constant implies greater employment prospects.

education have been removed *none of the racial indicator variables were significant*. That is, the Aboriginal racial group, for example, were not discriminated against. Relatively high unemployment rates may simply be attributable to low educational achievement.

Although by splitting Australian males into low and high eduction groups, much of the heterogeneity appertaining to education has been removed, education levels *within* these groups are still important. This is seen in the low education group, with Year 10 achievers enjoying a higher probability of employment than the Year 9 or below base case, but just one more year of schooling (to Year 10) more than doubles the probability of employment. In the high education group, a degree qualification is by the top ranking qualification in terms of employment prospects. A degree increases the probability of employment by nearly double that of a diploma within this group. The *trade qualification* and *Year 12* indicators were insignificant.

As expected, the employment status of the male's partner exerted a positive effect on employment prospects in both groups. The effect was much more pronounced in the low education group. Although both high education males and their partners may exhibit similar workforce status, the level of education may facilitate employment if this becomes necessary due to financial constraints, for example.

Disabled males were significantly disadvantaged in the workplace, irrespective of education. This was presumably a combination of workplace discrimination and the limited opportunities available to the disabled.

In terms of explicit policy variables, the reservation wage relative to the appropriate level of unemployment benefit, had a highly significant negative impact on both groups' willingness to supply labour. This negative effect appears to be irrespective of education levels.

In summary, the results of the male equations are broadly in line with *a priori* expectations, except that the financial commitments of a wife and family did not appear to exert a positive influence on job search intensity.

ii. The Estimated Female Equations.

Again the constant term in both low and high education groups is significantly positive. As with the male equations, the *natural rate of employment* appeared to be higher in the latter group, as might be expected. Again, as with their male counterparts, the probability of female employment increase with age, with the effect being more pronounced in the high education group.

The importance of one's marital status varied with education levels. In the low education group, marital status was unimportant. In the high education group, the *divorced/separated* indicator was significantly negative. Marriages often witness the emergence of parenthood and traditionally, it is the female who will remain at home to nurture the offspring. This explains the significantly negative coefficient on the *number of children* variable for both female educational groups. Somewhat surprisingly, this effect appeared to be stronger in the high education group.[19] Separation combined with stronger parental instincts, appeared to lessen the high education group's willingness to work.

Type of accommodation was uninfluential in determining the low education group's probability of employment. In the other group, as with their male counterparts, the financial commitment of buying a house, did lead to a stronger drive for employment among high education females. However, unlike the former, this was the only significant accommodation variable. Interestingly, the positive effect of buying a house appears much smaller for this female group than for both of the male education groups.[20]

As with both male education groups, the females' racial origin was not influential in determining the probability of their employment. Any superficial workplace discrimination could apparently be explained solely in terms of education levels. There was however, once more significant discrimination against the physically disabled, to an extent irrespective of education levels.

Splitting females into high and low education groups appeared to remove all of the heterogeneity for the former. In the low education group however, female's employment prospects received a boost from completion of both Years 10 and 11, relative to the base case of Year 9 or below. Again, just one more year of schooling

[19] It may be argued that higher educated workers are generally higher paid and can therefore afford child care, *etc.*. The result appears to be psychological one.
[20] Possibly, this can be attributable to the male's traditional status of major *bread winner*.

to Year 11, relative to Year 10, approximately doubled the probability of employment for this group.

As with their male counterparts, the employment status of one's partner is a significant determinant of the probability of both lowly and highly educated females. Again the effect was stronger in the low education group, presumably for the same reasons as outlined in the male case.

Finally, the reservation wage (relative to the appropriate level of unemployment benefit) similarly exerts a strong negative impact on both female education groups' decision to offer their labour services or not. The effect is of slightly smaller magnitude than in the male case.

In summary of the female equations, once more the results are generally in line with *a priori* expectations. The most notable difference between the female equations and those of the males, is the significance of the number of children in the former, suggesting that females were the dominant "full-time" parent.

iii. General Model Specification.

Both models appear to be well specified in terms of t-statistics and "correct" signs of coefficients.[21] All restricted models are "valid" using the likelihood ratio criteria. Finally, the statistical significance of ρ in all equations is a further validation of a correct model specification.[22]

iv. A Method of Model Prediction.

It would be useful to use the estimated equations to predict unemployment rates across standardised groups. In the usual cross-section probit this has the simple form of $P(y=1)=\Phi(X'\beta)$, where Φ is the cumulative standard normal distribution function. However, in the panel probit model, due to the unobserved random effects, the integral does not have a standard closed form, and is therefore very difficult to compute.[23] For this reason this was not undertaken.

---

[21] Asymptotically MLE's follow a standard normal distribution, therefore |t-ratios| > 1.96 indicate significance.
[22] See Greene (1992) p.424.
[23] Although this could be evaluated either by numerical integration or by simulation methods.

5. CONCLUSIONS AND FURTHER WORK.

This has been one of the first (if not the first) attempts at analysing this particular data set within a Random Utility discrete choice panel data setting. Indeed, the results suggest that the data is well-modelled as such.

Endogenous attrition was found and attributed to nationality and educational attainment; overseas respondents were more likely to de-select themselves, as were the "low" educated.

Estimation results suggest that if the authorities wish to influence the level of unemployment, a very powerful tool is the level of unemployment benefit, although this is politically sensitive. A policy aimed at increasing general levels of education will have strong implications for reduced unemployment.

The results also suggests that there is significant discrimination within the job-market, especially towards the physically disabled. In the interests of equity, this is a problem which needs to be addressed. Once educational attainment has been accounted for, there appeared to be no racial disadvantage (among Australian born individuals) in the workplace.

Other factors found to be important in the determination of employment, were place of residence, nature of housing occupancy, one's partner's employment status, marital status and the number of dependent children (for women only).

Although these results are quite promising, one needs to bare several points in mind. Firstly, there has been an inherent ignorance of many demand-side factors. Secondly, there are other potential supply-side variables not considered, for example one's employment history may mark one with an employment *scar* that adversely affects employment prospects.

There also appears to be scope for additional model specifications. For example, a further specification of a random effects probit, in which the individual random effects are now correlated to the individual's characteristics, could be considered. Indeed, if this were the true data generating process, then maximisation of (8) will have the same effect as instigating omitted variable bias into the parameter estimates.

To solve this problem, a distribution for α must be specified which is conditional on $\underline{x}$. Chamberlain (1980, 1984) has suggested a linear decomposition of α such that:

(18) $$\alpha_i = \sum_{t=1}^{T} \underline{a}' \, \underline{x}_{it} + \eta_i ,$$

or:

(19) $$\alpha_i = \underline{a}' \, \underline{x}_i ,$$

where: $\underline{a}' = (\underline{a}'_1, \ldots, \underline{a}'_T)$
and $\underline{x}'_i = (\underline{x}'_1, \ldots, \underline{x}'_T)$.

The log-likelihood function of (8) now becomes augmented by the additional term of (19), such that:

(20)
$$\log L = \sum_{i=1}^{N} \log \int \Pi_{t=1}^{T} F\left(\underline{x}_{it} \underline{\beta} + \underline{a}' \, \underline{x}_i + \eta \right)^{y_{it}} * \left[1 - F\left(\underline{x}_{it} \underline{\beta} + \underline{a}' \, \underline{x}_i + \eta \right) \right]^{1-y_{it}} d\, G\left(\alpha \,|\, \underline{\delta} \right).$$

The function G^* is a univariate distribution for η. If F is standard normal and G^* is the distribution function of a normal random variable with zero mean and variance σ^2_η, this specification yields a counterpart to the Probit of (9) of:

(21) $$Y_{it} = 1 \;\; if \;\; \beta' x_{it} + a' x_i + \eta_i + u_{it} > 0.$$

The only difference between this specification and that outlined above (*i.e.* equation (9)), is the incorporation of the $a' x_i$ term(s) to capture the incidental dependence between the random effects component and the *x*-vector (matrix). Thus the essential characteristics of estimation are the same, however this variant is not yet available in LIMDEP, or indeed in any other package that we are aware of.

Finally, there appears the need for further work to enable the estimation of panel probit models without numerical integration methods and so provide an easier closed form for the choice probabilities.

REFERENCES.

Bradbury, B., McRae, I. & Wozybun, L. (1989), "Families and Early Labour Market Experience: An Analysis of Siblings", *The Australian Journal of Statistics*, special volume 31A, August 1898.

Bureau of Labour Market Research (1986), *The First Wave of the ALS: Facts and Figures About Young CES Registrants*, Monograph Series No. 15, Bureau of Labour Market Research, Canberra: APGS.

Butler, J.S. & Moffitt, R. (1982), A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model, *Econometrica*, 50, 3, pp. 761-764.

Chamberlain, G. (1980), An Analysis of Covariance with Qualitative Data, *Review of Economic Studies*, 47, pp. 225-238.

Chamberlain, G. (1984), Panel Data, in *Handbook of Econometrics*, Vol. II, ed. by Z. Griliches and M. Intriligator, pp. 1247-1318, Amsterdam, North Holland.

Greene, W.H. (1993), *Econometric Analysis*, New York: Macmillan.

Greene, W.H. (1992), *LIMDEP Version 6, User's Manual and Reference Guide*, Econometric Software Inc., New York.

Harris, M.N. (1993), The Australian Longitudinal Survey: A Reference Document, Memo, Department of Econometrics , Monash University, Melbourne.

Hsiao, C. (1985), Benefits and Limitations of Panel Data, *Econometric Reviews*, 4, pp.121-174.

Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge: Cambridge University Press.

Keating, P., Prime Minister Australian Federal Government (1994), *Working Nation*, Federal Government of Australia, White Paper, Canberra: APGS.

Kryger, T. (1990), *The Australian Longitudinal Survey: 1985 to 1988 - Dynamics of the Youth Labour Market*, Monograph Series No. 4, DEET, Canberra: APGS.

Larum, J. & Beggs, J.J. (1989), "What Drives Australian Teenage Labour Force Participation?", *The Australian Journal of Statistics*, special volume 31A, August 1898.

Mátyás, L. & Sevestre, P. (eds.) (1992), *Econometrics of Panel Data*, Kluwer Academic Publishers.

Miller, P.W. (1989), "The Structure of Aboriginal and Non-Aboriginal Youth Unemployment", *Australian Economic Papers*, vol. 28, No. 52.

Miller, P. & Volker, P. (1989), "Socioeconomic Influences on Educational Attainment: Evidence and Implications for the Tertiary Education Finance Debate", *The Australian Journal of Statistics*, special volume 31A, August 1898.

Prior, H. & Beggs, J.J. (1989), "Influence of Family Background on the Educational and Labour-Force Outcomes of Year 12 School-Leavers", *The Australian Journal of Statistics*, special volume 31A, August 1898.

Smith, D.J. & Callaghan, A.J. (1989), "Definitional Aspects of Measurement of Unemployment - A Comparative Study of the Australian Longitudinal Survey and Other Sources of Youth Unemployment Data", *The Australian Journal of Statistics*, special volume 31A, August 1898.

Trewin, D., Bode, G., Boal, P. & Newton, D. (1989), "An Automatic Interaction Detection Analysis on Unemployed Youth", *The Australian Journal of Statistics*, special volume 31A, August 1898.

APPENDIX I.

Table 1: Coding of Variables Summarised in Tables 2 - 5 Below (where no numerical equivalent)

Employment Status	1	Employed
	2	Unemployed
	3	Discouraged
	4	Not in Labour Force
	5	Waiting to Start Job
Marital Status	1	De-Facto
	2	Married
	3	Separated
	4	Divorced
	5	Widowed
	6	Single
Sex	1	Male
	2	Female
Disabled	0	Able-Bodied
	1	Physically Disabled
Accommodation	1	Owner
	2	Buyer
	3	Renter
	4	Boarder
	5	Other
Place of Residence	1	Capital City
	2	Other City
	3	County Town
	4	Rural Area
Full/Part Time Study	1	Apprenticeship
	2	Full-Time Study/School
	3	Part-Time Study/School

Note: all missing values are recorded as -9.

Table 2: Summary Statistics for Various Variables, 1985, n = 8998

	Mean	Mode	Standard Deviation	Missing/Not Applicable
Employment Status	1.81	1.00	1.234	0
Marital Status	5.11	6.00	1.725	0
Age	20.216	16.00	2.869	0
Sex	1.498	1.00	.500	0
No. of Children	0.158	0.00	0.497	0
Disabled	0.101	0.00	0.301	564
Accommodation	3.867	4.00	0.984	2
Place of Residence	1.77	1.00	0.978	0
Full/Part Time Study	2.141	2.00	0.746	6749

Table 3: Summary Statistics for Various Variables, 1986, n = 7871

	Mean	Mode	Standard Deviation	Missing/Not Applicable
Employment Status	1.646	1.00	1.152	0
Marital Status	4.916	6.00	1.838	0
Age	21.144	17.00	2.88	0
Sex	1.497	1.00	0.500	0
No. of Children	0.210	0.00	0.581	0
Disabled	0.099	0.00	0.298	298
Accommodation	3.713	4.00	1.015	4
Place of Residence	1.777	1.00	0.974	0
Full/Part Time Study	2.171	2.00	0.746	5825

Table 4: Summary Statistics for Various Variables, 1987, n = 7110

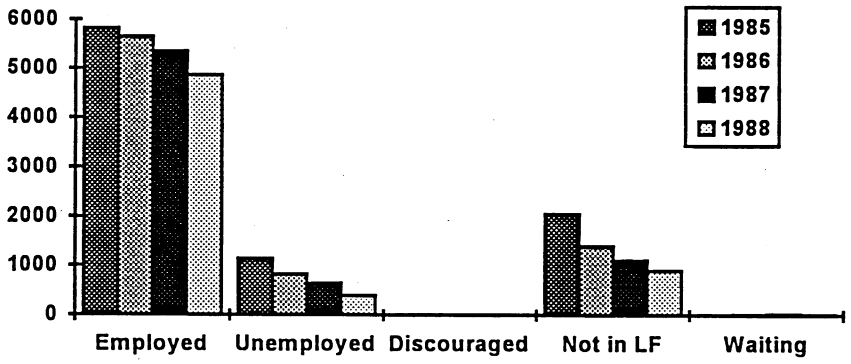
	Mean	Mode	Standard Deviation	Missing/Not Applicable
Employment Status	1.572	1.00	1.106	0
Marital Status	4.694	6.00	1.938	0
Age	22.157	18.00	2.886	0
Sex	1.497	1.00	0.500	0
No. of Children	0.279	0.00	0.681	0
Disabled	0.094	0.00	0.292	244
Accommodation	3.683	4.00	1.292	1
Place of Residence	1.763	1.00	0.968	0
Full/Part Time Study	2.472	3.00	0.635	6119

Table 5: Summary Statistics for Various Variables, 1988, n = 6151

	Mean	Mode	Standard Deviation	Missing/Not Applicable
Employment Status	1.508	1.00	1.071	0
Marital Status	4.445	6.00	2.029	0
Age	23.154	19.00	2.896	2
Sex	1.494	1.00	0.500	2
No. of Children	0.340	0.00	0.758	0
Disabled	0.092	0.00	0.289	169
Accommodation	3.539	3.00	1.295	1
Place of Residence	1.752	1.00	0.948	0
Full/Part Time Study	2.548	3.00	0.615	5460

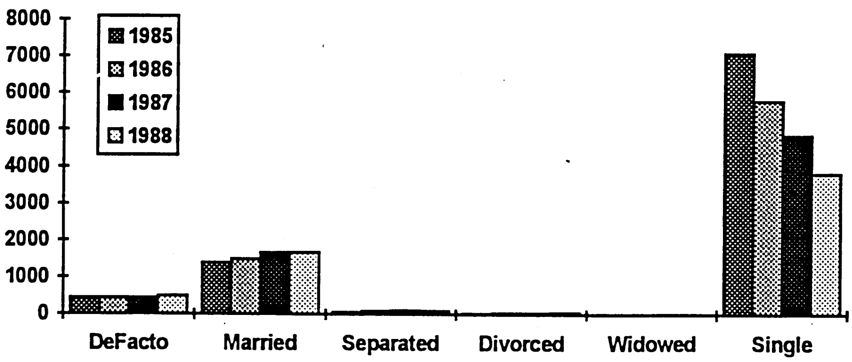
APPENDIX II.

<u>*Chart 1: Derived Employment Status*</u>



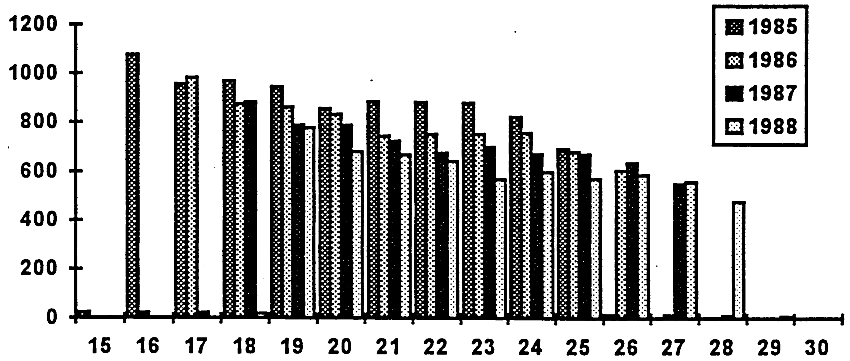
Total sample 1985: 8998, number of missing/not asked observations: 0,
1986: 7871, number of missing/not asked observations: 0,
1987: 7110, number of missing/not asked observations: 0,
1988: 6151, number of missing/not asked observations: 0.

<u>*Chart 2: Marital Status*</u>



Total sample: 1985: 8998, number of missing/not asked observations: 0.
1986: 7871, number of missing/not asked observations: 0,
1987: 7110, number of missing/not asked observations: 0,
1988: 6151, number of missing/not asked observations: 0.

Chart 3: Age



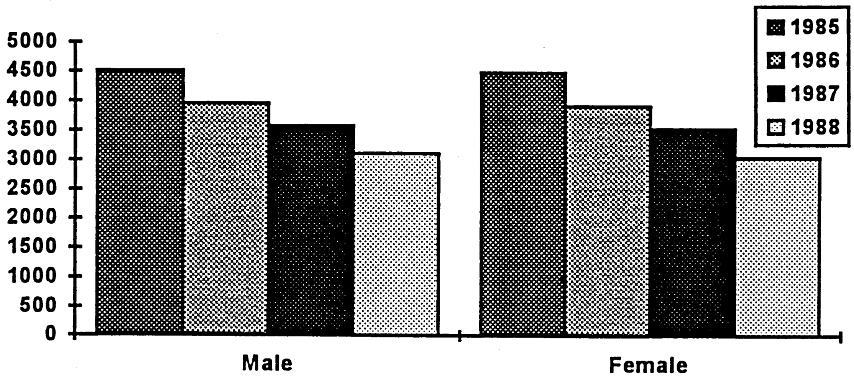
Total sample: 1985: 8998, number of missing/not asked observations: 0.

1986: 7871, number of missing/not asked observations: 0,

1987: 7110, number of missing/not asked observations: 0,

1988: 6151, number of missing/not asked observations: 2.

Chart 4: Sex



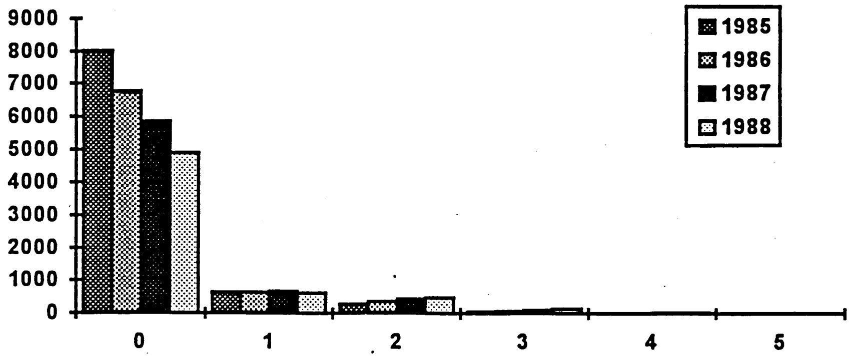
Total sample 1985: 8998, number of missing/not asked observations: 0,

1986: 7871, number of missing/not asked observations: 0,

1987: 7110, number of missing/not asked observations: 0,

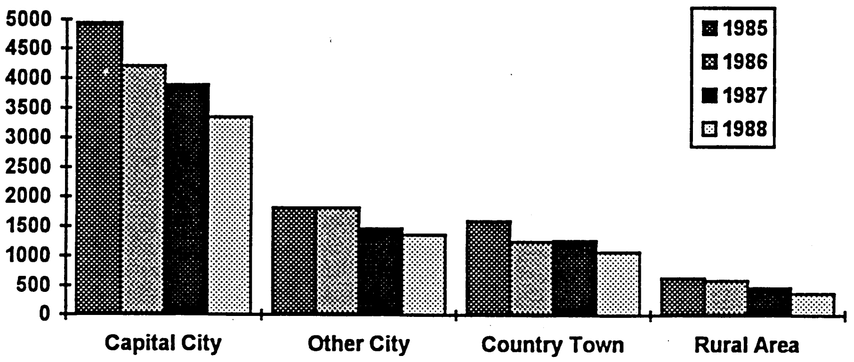
1988: 6151, number of missing/not asked observations: 2.

Chart 5: Number of Dependent Children



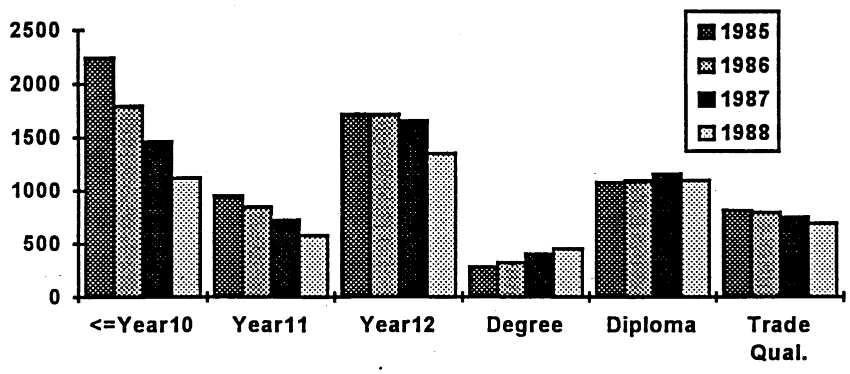
Total sample 1985: 8998, number of missing/not asked observations: 0,
1986: 7871, number of missing/not asked observations: 0,
1987: 7110, number of missing/not asked observations: 0,
1988: 6151, number of missing/not asked observations: 0.

Chart 6: Place of Residence



Total sample 1985: 8998, number of missing/not asked observations: 0,
1986: 7871, number of missing/not asked observations: 0,
1987: 7110, number of missing/not asked observations: 0,
1988: 6151, number of missing/not asked observations: 0.

Chart 7: Educational Attainment



Total sample 1985: 8998, number of missing/not asked observations: 1251,

1986: 7871, number of missing/not asked observations: 510,

1987: 7110, number of missing/not asked observations: 110,

1988: 6151, number of missing/not asked observations: 19.