CENTRE FOR DEVELOPMENT STUDIES

WORKING PAPER NO.3[illegible]

THE MARKET FOR THE EDUCATED IN KERALA:

Determinants of the Waiting Period

Chandan Mukherjee

Centre for Development Studies
Ulloor, Trivandrum 695011
Kerala, (India)

Introduction

The most important section of the educated job seekers in the State of Kerala consists of those with educational attainment ranging from matriculation to the graduate level. As on 30th June, 1973, this section constituted 92.6% of the total educated job seekers on the live registers of the employment exchanges in the State.[1] As a considciable proportion of the job seekers in this section are found to be registrants of the employment exchanges, data on these registrants can be expected to reflect fairly accurately the experiences of these job seekers.[2]

A sample survey of these educated registrants was conducted by Centre for Development Studies, Trivandrum, during April, 1973. Two of the eleven districts of Kerala viz. Trivandrum and Calicut were chosen for the purpose of the survey.[3]

Apart from the live register there is also a deed register at each of the employment exchanges, of persons who fail to renew their registration from time to time. In both exchanges random samples were drawn directly from stocks of registration cards of both of the live and dead registers of persons whose educational attainment ranged from Matriculation to Graduate level. The reason for drawing samples from dead registers too was that, a transfer from live to dead register need not necessarily imply

---

1. 'Kerala Economic Review,' 1973, Government of Kerala.

2. 'Poverty Unemployment and Development Policy', Vol.2, Appendix G, Centre for Development Studies, Trivandrum. (Mimed).

3. The reason for the choice of those two districts are explained in 'Poverty Unemployment and Development Policy' [2]

that the registrant is now employed and not seeking work; it may be simply due to a failure to renew the registration in time. Moreover persons with completed waiting periods were expected to be found in greater numbers on dead registers; the estimation of average waiting periods is possible only on the basis of such completed waiting periods.[4]

The main form of employment disired by the educated is the salaried one. Accordingly, duration of employment was defined with respect to salaried jobs. Using this criterion of salaried employment the experiences of the registrants were studied. The results of the study incorporated in "Poverty Unemployment and Development Policy", can be summarised as follows.

1. There is no significant difference in the experiences of the registrants of the two exchanges.[5]

2. Waiting periods are found to be inversely correlated with level of education.

---

4 The distinction between live and dead registers was found to be of technical nature,..."the distinction between live and dead register is irrelevant to a certain extent from the standpoint of the market for the educated. While the proportion of the currently employed persons is larger on the dead register than on the live register, the proportion of unemployed persons on the dead register is quite substantial: the proportion of wholly or partly unemployed registrants among those whose registrations have lapsed during the year preceding April 1973 was 43 per cent in Trivandrum and 56 per cent in Calicut" (Refer 2).

5 ".......71 per cent of the registrants surveyed in Trivandrum had no jobs even after spending varying lengths of time in search of employment. The average time spent by them at the date of survey i.e. April, 1973, was 40.4 months. The remaining 29 per cent who had been employed at one time or other had an average first waiting period, [period from completion of education to first salaried job] of 36.6 months. In Calicut the average time spent by those who never had a salaried job – constituting, again, 71 per cent of the total registrants – was 38.6 months and the average first waiting period of those who had some experience of salaried employment i.e. the remaining 29 per cent – was 34 months. There is thus practically no difference in the severity of the situation in the two districts". (Refer [2]).

3. Waiting periods are found to be inversely correlated with family income.

4. Waiting periods (average) are significantly lower for registrants belonging to households of salaried white-collar earners.

5. Caste did not prove to be an important factor influencing the duration of unemployment.

There are two technical points to be noted in the context of this analysis. First, over two thirds of the registrants had already spent more than 3 years without a job at the time of the survey. So, the average waiting period for the registrants, as a whole is likely to be much higher than the estimate based on the data on those who have had a job at the time of the survey. Secondly, the analysis was based on one variable at a time. Thus, although the results of the analysis suggests the hypothesis that apart from the level of education there are other socio-economic factors influencing the duration of unemployment, it does not confirm so. It is because one cannot say how far one factor contributes independently of other factors.

This paper attempts to complete the previous analysis in this respect. It has been attempted to verify the presence of discrimination in the market for the educated by estimating average duration of unemployment and features of the distribution of the waiting periods in different socio-economic groups.

Methodology

Based on the findings of the study referred to above we can assume that the distributions of the waiting periods are the same for both districts; i.e. we assume that the two samples are drawn from the same population independently. Thus we can put both the samples together and consider them as a single sample.

The average waiting periods are found to be inversely related with levels of education. So, socio-economic groups should be considered at different levels of education in order to verify if socio-economic back ground has any bearing on the average waiting period independently of the level of education. We have defined two levels of education - first level as matriculation and second one as higher than matriculation.

At each of the two levels of education we have classified the registrants in two ways - one according to the economic status of the registrant's families and the other according to the social status of the same. We have chosen 'occupation of the main earner of the registrant's family' and 'the monthly income of the registrant's family' as the indices of the registrant's social and economic backgrounds respectively. occupations have been divided into two classes, viz., wage-occupation and non-wage-occupation. Again according to family income, registrants have been divided into two groups with monthly income of Rs.350.00 as the upper bound for low income group.

Conceptually, we can regard the waiting time to the first salaried job after the completion of education as similar to the 'time to failure of an item' after it is put to a life test. This enables us to employ the well-known methods of statistical analysis for life testing experiments, particularly for the study of the distribution. In this section we shall accordingly use, the 'time to failure' and the 'waiting time to first salaried jobs' (or 'period of unemployment'), in the equivalent sense.

We now formulate the problem in terms of a life testing experiment as follows.

Let $W_1, W_2, \ldots\ldots, W_n$ be the true failure times of n times put to life test. $W_1, W_2, \ldots\ldots W_n$ are assumed to be independent random variables having a common distribution $F(.)$; $F(x) = [\text{prob.}\ (W \leq x)]$. The period of observation of the ith item will be limited by an amount T_i, i.e. W_i, are censored on the right by T_i and what will be observed are only $X_i = \text{Min.}\ (W_i, T_i)$. T_i are not random variables but specified beforehand.

In the present context -

W_i : Waiting time of the ith registrant to the first salaried job after completion of education.

T_i : Period from completion of education to the date of survey (April '73) for the ith registrant.

This kind of structure is essential because it is impossible to observe a cohort of registrants until all of them secure jobs. What we do observe is whether or not (as on the date of survey) a registrant has had a job sometime after the completion of education. If a registrant reports experience of employment, the corresponding W could be ascertained (where W is the period of unemployment); otherwise we observe an 'incomplete waiting time' T which measures the time spent by the registrant till the date of survey since completing education. Although T gives an incomplete information about the waiting time, its importance in the estimation of the mean waiting time has already been indicated in the previous section. So, the formulation of the problem is aimed at incorporating the incomplete information too.

A formal method of analysis of the data in each group is constructed in the following lines.

We shall try to approximate the failure time distribution $F(t)$ by examining the failure rate function $r(t)$. Failure rate is defined as, $r(t) = f(t)/W(t)$; $t > 0$, where $F(t)$ is assumed to be differentiable $F'(t) = f(t)$ and $U(t) = 1-F(t)$.

Case 1: Let $r(t) = 1/a$; 'a' - finite positive constant, i.e. the failure rate is constant and $f(t)$ is exponential with parameter 'a'.

We have $g(t) = -\log_e U(t) = t/a; \quad 0 < t < \infty$ ..... (1)

So, if $g(t)$ plotted against t approximates well to a straight time passing through the origin, we may conclude that $F(t)$ can be well approximated by an exponential distribution with mean 'a'.

The well known maximum likelihood estimator[6] (m.l.e.) of 'a' is given by

$$\hat{a} = \sum_{i=1}^{r} x_i / r$$ where r is the number of failures.

Case 2: Let $r(t) = a + b.t$; a, b are finite constants, i.e. the failure rate is monotonic with a constant increasing or decreasing rate.

6 'The sampling distributions of an estimate arising in life testing' Bertholomew (1963), Technometrics; Vol.5.

In this case we have*

$$U(t) = \exp.(-a.t - b.t^2/2);\ o < t < \infty ,$$

$$\text{and}\quad g(t) = -\log_e U(t) = a.t. + \frac{b}{2}.\ t^2;\quad o < t < \infty \quad .....(2)$$

So like case 1, if g (t) plotted against t approximates well to a convex or a concave curve of the type $c.t.+d.t^2$ we may conclude that F can be well approximated by a distribution having a density given by,

$f(t) = (a+b.t)\exp.(-a.t.-\frac{b}{2}.t^2);\ o < t < \infty$. The mean 'm' of this distribution is easily verified** to be

$$m = \tfrac{1}{2}.(2\pi/b)^{\frac{1}{2}}.\ \exp.(a^2/2.b)$$

The maximum liklihood estimates of a and b can be obtained by solving the following maximum likelihood equations***

$$\sum_{i=1}^{n} (a+b.X_i)^{-1}.\ K_i = \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} (a+b.X_i)^{-1}.\ X_i.K_i = \tfrac{1}{2}\sum_{i=1}^{n} X_i^2 \quad .... (3)$$

where
$$K_i = 1 \text{ if } W_i \leq T_i$$
$$= 0 \text{ if } W_i > T_i \quad .... (4)$$

To verify how adequately g(t) can be approximated by the theoretical curves (straight line or monotonic curve through the origin) derived from the two forms of failure rate functions considered above, we propose to compute the Co-efficient of Determination (R^2) of a least square fit of (1) or (2) to g(t). If R^2 is satisfactorily high the corresponding distribution can be taken as a good approximation of F(t).

* See Appendix 1
** See Appendix 2
*** See Appendix 3

After estimating the parameters of the distribution (suggested by the graphical and least square fit studies) we can also examine the goodness of fit of the theoretical distribution by comparing the observed number of failures with the expected number of failures. The expected number of failures can be obtained as follows:

The observed number of failures $= \sum_{i=1}^{n} K_i$, where K_i has already been defined earlier. K_i are random variables following binomial distributions such that,

$$\text{Prob.}\ (K_i = 1) = \text{Prob.}\ (W_i \leq T_i) = F(T_i)$$

Thus, expected number of failures $= \sum_{i=1}^{n} F(T_i)$, where

$$F(T_i) = \sum_{i=1}^{n} 1-\exp.\ (-T_i/a) \qquad \text{for case } 1$$

$$= \sum_{i=1}^{n} 1-\exp.\ (-a.t- \frac{bt^2}{2}) \qquad \text{for case } 2$$

Now, the sample tail function $U(t)$ is not directly available from the data as the failure times are not known for all the sample units. We propose to take the maximum likelihood estimator of the tail function $U(t)$ as suggested by Kaplan and Meir[7] (1958). The m.l.e of $U(t)$ from censored data is derived as follows:

Define m intervals; $I_k = (h_{k-1}, h_k]$ with end points

$0 = h_1 < h_2 < \ldots\ldots < h_m < \infty$. The conditional probability q_k of failure in the intervals I_k is estimated as -

[7] 'Non-parametric estimation from incomplete observations' - Kaplan, E.L, and Meir, P. (1958), J.Amer. Statist.Assoc. 53:457-481.

$$\hat{q}_k = D_k / N_k \; ; \quad k = 1,2,\ldots\ldots,m \quad \text{where}$$

$$D_k = \sum_{i=1}^{n} L_i$$

$$N_k = \sum_{i=1}^{n} L_i'$$

and $L_i = 1$ if $x_i \in I_k$ and $T_i > h_k$

$= 0$ otherwise,

$L_i' = 1$ if $x_i > h_{k-1}$ and $T_i > h_k$

$= 0$ otherwise.

Thus, $\hat{U}(h_k) = \hat{p}_1 . \hat{p}_2 . \ldots\ldots \hat{p}_k \; ; \; \hat{p}_k = 1 - \hat{q}_k$.

We have analysed the data in each group (socio-economic) at each of the two education levels following the above method. The results are presented and in the next section.

Results and inferences

Results of the analysis of the data have been summarised in Table 1. It can be seen from this table that Constant failure rate model fits adequately to the periods of unemployment in each of the four Non-wage occupation groups. Co-efficient of determination[8] (discussed in previous section) is greater than 0.92 in each of those cases. Same degree of adequacy of fit of Increasing (at constant rate) failure rate model to the Wage occupation-low income-low education group data, is observed. Co-efficient of determination in this case is as high as 0.99.

[8] The co-efficient of determination is unity in case of a perfect fit of the model to the da a.

Mean waiting period in a group is estimated by using maximum likelihood estimators of the parameters of the distribution which is found to fit well (judged by Co-efficient of determination) to the group data. In case of the distribution with increasing failure rate the parameters were estimated as $\hat{a} = 0.00135140$; $\hat{b} = 0.00007146$. In case of the distributions with constant failure rate the estimated parameters are the means of the distributions shown in Table 1 as MW. The goodness of fit of these models has been well assured by the extreme closeness of the expected number (determined by the respective model) of persons who have had jobs sometime after completion of education to the observed number of the same, in different groups. The maximum deviation of the observed number from the expected number is only 5% (in the case of Non-wage occupation-high income-low education group).

Before we begin to interprete the results obtained, we must note here that we are analysing the data compiled from a random sample of registrants of employment exchanges. Thus, whatever inference we draw is valid only for the registrants (classified as SSLC, PDC and B.A./B.Sc.) of the employment exchanges of Kerala. But, as it is indicated in Poverty Unemployment and Development Policy[2] that a fairly high proportion of job seekers (with education as mentioned earlier) do register in the employment exchanges of the state, the inferences drawn here can cautiously be extended to the whole population of educated job seekers (belonging to the section under consideration) in the state.

Table 1

Summary of the results from the analysis of the data

Occupation of the main earner of the family \ Levels of education and family income	Level of education: Matriculation		Level of education: Above Matriculation	
	Monthly family income ≤ Rs.350	Monthly family income > Rs.350	Monthly family income ≤ Rs.350	Monthly family income > Rs.350
Wage Occupation	FR: Increasing at constant rate CD = 0.9896 O = 18 E = 17.8 MW = 150 n = 88	(*)	(**)	(*)
Non-Wage Occupation	FR: Constant CD = 0.9267 O = 104 E = 104.2 MW = 181 n = 365	FR: Constant CD = 0.9646 O = 27 E = 25.7 MW = 108 n = 66	FR: Constant CD = 0.9951 O = 34 E = 32.7 MW = 78 n = 97	FR: Constant CD = 0.9948 O = 22 E = 22.3 MW = 39 n = 45

FR = Failure rate of the waiting time distribution

CD = Co-efficient of determination of least square fit to the tail function.

O = Observed number of persons got jobs

E = Expected number of persons got jobs.

MW = Estimated mean waiting time in months.

n = Number of sampled registrants.

(*) There were only two registrants from wage earners family with monthly family income more than Rs.350.

(**) Out of 23 registrants in this group 13 registrants completed their education less than one month before the survey. Of the rest 10 registrants 3 had got jobs with average waiting period of 19 months. The other 7 registrants (out of the above 10) were looking for jobs for 31.9 months on an average. This is an interesting observation by itself but a discussion on this is, however, beyond the scope of this paper.

From Table No.1, it can be seen that, while the bottom most socio-economic group with lower education has a waiting time distribution with increasing failure rate, all other groups (wherever the distribution could be analysed) have constant failure rate for the same The failure rate property of a distribution can be interpreted as follows.

Consider the function $g(x,t) = \{F(t+x) - F(t)\} / \{1-F(t)\}$ $g(x,t)$ is the mathematical probability of - 'given that a person is looking for a job unsuccessfully for t units of time, he would get it before (t+x) units of time'. t and x are both assumed to be non-negative Now if the distribution of the waiting periods to salaried jobs has an increasing failure rate it would imply that $g(x,t)$ is increasing with t for any given positive finite x i.e. the longer a registrant looks for a job unsuccessfully the more is his chance of getting one in near future Similarly, a constant failure rate would imply that $g(x,t)$ is remaining the same for all t for a given positive finite x i.e. however long one is searching for a job, the chance of his getting a job within near future remains the same. This is illustrated in Table No.2. Here we have taken x as 12 months and computed $g(x,t)$ for different groups.

Registrants with matriculation as qualification from non-wage earners' low income families are found to have the highest average waiting time of 161 months. The non-wage occupation consists of self-employment (traders, tailors, cultivators etc.) and different types of salaried jobs (teachers, clerks, peons, small technicians etc.). The staggering magnitude of this average is not all that unexpected as it has been already found that in both the districts of Trivandrum and Calicut

71 per cent of the total registrants were still looking for jobs at the time of the survey after completing education more than 3 years before on an average. Particularly in this group there was 5.2 per-cent of the registrants on the look out for a salaried job unsuccessfully for more than 100 months with an average of 122 months. This means for those registrants average waiting time is more than 122 months.

Table 2

Probabilities of - 'a registrant would get a job within an year given that he has already spent t years searching for one' - in different socio-economic groups

t	Matriculation			Above Matriculation	
	Monthly family income $\leq$ Rs.350		Monthly family income $>$ Rs.350	Non-wage Occupation	
	Wage Occupation	Non-Wage Occupation	Non-wage Occupation	Monthly family income $\leq$ Rs.350	Monthly family income $>$ Rs.350
0	.02114	.06415*	.10516*	.14260*	.26486*
1	.03116				
2	.04108				
3	.05089				
4	.06061				
5	.07023				
6	.07023				
7	.08917				
8	.09849				
9	.10772				

* These probabilities do neither improve nor deteriorate with 't' so that the chance of getting a job within a year remains the same irrespective of the time spent in in search for one.

Note: For the first column:

$$g(x,t) = 1-\exp.\left\{-ax-\frac{b}{2}(x^2+2tx)\right\}$$

where parameters 'a' and 'b' are already explained.

For the second column:

$$g(x,t) = 1-\exp.\,(-x/a)$$

Where again parameter 'a' has been explained earlier.

This seems to be well reflected in the large magnitude of the mean and constant failure rate of the distribution of waiting time. Because these propertise imply a low chance of success (estimated as 0.06415) in the first year of search for a salaried job which does not improve in succeeding years.

Registrants from non-wage earners' high income families with the same education level as the registrants in the previous group discussed above, have a considerably lower (more than 6 years difference) average waiting time (108 months) than the previous group. But the magnitude of this average by itself is very high for similar reasons discussed in the previous case. The registrants in this socio-economic group do not differ from those in the previous group except in that — they are from families with higher income. This comparative advantage of lower waiting time of the higher income group is found in the case of higher education level also. Although higher level of education lowers the waiting time considerably for both the socio-economic groups, the average waiting time of registrants from higher income families is half of that of those from low income families. Thus it seems that family income, which is taken as an index for the economic status of the registrant's family, has a considerable bearing on the waiting time to salaried jobs. It must be added that even the registrants from the non-wage earners' higher income families with higher level of education do have to spend on an average $3\frac{1}{2}$ years of unemployment.

It is interesting to note that registrants with matriculation from wage earners' families (which are naturally low income families) have lower average waiting time than those from non-wage earners'-low income families. While registrants in the former group have to wait 150 months on an average for a job, those from the latter group have to wait for 181 months on an average for the same (see Table 1). Analytical reason for this lies in the finding that the former group has an increasing failure rate while the latter a constant failure rate. Thus, for a registrant from the former group, the chance of getting a job (although comparatively lower) improves over time while for a registrant from the latter group it remains the same, leading to a better average waiting time in the long run. But what causes this particular socio-economic group to have an increasing failure rate for the waiting time distribution while all other socio-economic groups have a constant failure rate for the same? It can be seen from Table 2 that the chance of getting a job in the first year is considerably lower (less than half) for the bottom-most socio-economic group than any other group. Only after 4 years of search for a job a registrant from the wage earners'—low income group has a better prospect than a registrant from non-wage earners'-low income group with the same level of education (matriculation). It can be seen from Table No.3, that the only difference between these two groups just mentioned is that registrants with backward castes (Muslims, Ezhavas, Other backward castes, Scheduled Castes/Scheduled Tribes etc.) are proportionately higher in wage occupation groups than in the non-wage occupation group. All of these backward castes enjoy reservation facilities in Government Employment. This might partly explain the fact that chance of finding job improves with time (increasing failure rate) for this group.

Table 3

Caste composition and average salary of the first job secured in different socio-economic groups

Occupation of the main earner of the family. / Level of education and family income	Level of Education: Matriculation — Monthly family income ≤ Rs.350	Matriculation — Monthly family income > Rs.350	Above Matriculation — Monthly family income ≤ Rs.350	Above Matriculation — Monthly family income > Rs.350
Wage Occupation	C = 79 S = 120	-	-	-
Non-wage Occupation	C = 62 S = 128	C = 44 S = 154	C = 58 S = 183	C = 43 S = 230

C : Proportion (%) of backward caste registrants.

S : Monthly salary (Rs) of the first job.

On the whole, it appears that there are factors other than level of education that have an effect on the waiting time.

Acknowledgement

The author is deeply indebted to Professor N.Krishnaji who made the initial formulation of the problem and provided guidance throughout the preparation of this paper.

Appendix 1

Consider a differentiable distribution function $F(.) = [\text{Prob.}\ (X \leq x)]$.

The failure rate function $r(x)$ is defined as, $r(x) = f(x)/U(x)$ where

$f(x) = F(x)$ and $U(x) = 1-F(x)$.

Suppose $r(x) = a + b.x$ where a and b are finit

$$\text{Then,}\quad f(x)/U(x) = a+b.x$$

$$\text{ie.}\quad -d \log_e U(x) = (a+b.x)dx$$

$$\text{ie.}\quad -\log_e U(x) = a.x + \frac{b}{2}.x^2$$

$$\text{ie.}\quad U(x) = \exp.(-a.x-\frac{b}{2}.x^2)$$

$$\text{ie.}\quad f(x) = (a + b.x).\exp.(-a.x-\frac{b}{2}.x^2)$$

Appendix 2

Mean 'm' of the distribution whose density is given by,

$f(x) = (a+b.x).\exp.\ (-a.x-\frac{b}{2}.x^2)$; $0 < x < \infty$, is derived as follows:

$$m = \int_0^\infty x.f(x)\ dx$$

$$= \int_0^\infty x.\ (a.+b.x).\exp.\ (-a.x-\frac{b}{2}.x^2)\ dx$$

$$= -x.\exp.\ (-a.x-\frac{b}{2}.x^2)\Big|_0^\infty + \int_0^\infty \exp.\ (-a.x-\frac{b}{2}.x^2)\ dx \quad \text{(integrating by parts)}$$

$$= \int_0^\infty \exp.\ (-a.x-\frac{b}{2}.x^2)\ dx$$

$$= \exp.\ (a^2/2b)\ .\int_0^\infty \exp.-(\sqrt{\frac{b}{2}}.x + \frac{a}{\sqrt{2b}})^2\ dx$$

$$= \frac{1}{2}.\ \exp.(a^2/2b).\ (\pi/a)^{\frac{1}{2}}$$

Appendix 3

Suppose $W_1, W_2, \ldots\ldots, W_n$, are the true failure times of n items

$T_1, T_2, \ldots\ldots, T_n$ are pre-assigned numbers.

$$\text{Define } X_i = \text{Min.}\,(W_i, T_i)$$

$$K_i = 1 \text{ if } W_i \leq T_i$$

$$= 0 \quad \text{otherwise}$$

We assume that $K_i = 1$ for at lease one 'i'.

Then the likelihood function is given by,

$$L = \prod_{i=1}^{n} K_i \,.\, (a+b x_i).\ \exp.\ (-ax_i - \frac{b}{2} x_i^2).\quad \prod_{i=1}^{n} (1-K_i).\ \exp.\ (-ax_i - \frac{b}{2} x_i^2).$$

Thus, $L^* = \log_e L$

$$= \sum_{i=1}^{n} \left\{ \log_e (a+bx_i) - ax_i - \frac{b}{2} x_i^2 \right\}.\ K_i$$

$$+ \sum_{i=1}^{n} (-ax_i - \frac{b}{2} x_i^2)\ .\ (1-K_i)$$

So, we get the maximum likelihood equations as follows:

$$\frac{L^*}{a} = \sum_{i=1}^{n} \left\{ (a+bx_i)^{-1}\ .\ K_i \right\} - \sum_{i=1}^{n} X_i = 0$$

$$\frac{L^*}{a} = \sum_{i=1}^{n} \left\{ (a+bX_i)^{-1} . X_i\ .\ K_i \right\} - \frac{1}{2} \sum_{i=1}^{n} X_i^2 = 0$$

In this paper the above equations are solved by using Newton-Raphson method.