C.E.M.A.
Virrey del Pino 3210
1426 Capital Federal
Tel.: 552-3291/9313/7771

* ISBN Nº 987-99363-3-7
Queda hecho el depósito que marca la Ley 11.723

DURATION DEPENDENCE IN THE SEQUENTIAL MIGRATION MODEL

Carola Pessino
Noviembre 1993
Nº 93

I. Introduction[1]

At the early stages of an individual's working career a decision has to be made on what type of job and where geographically to work. Sometimes, previous investment in human capital narrows the choice to a few jobs and locations, some other times this choice is wider. But, in most cases the individual has a narrow information on the characteristics of a given job and/or location and on his/her own traits that could be matched to those. This implies that there is a lot to "learn" at the initial stages in order to find the best possible match between worker and job-location. These ideas have been formalized in the job mobility context by Jovanovic (1979), Johnson (1980) and Miller (1984) among others. One of the implications found in this literature is that individuals faced with the initial choice of say two jobs, will sample the one with more learning opportunities first, even if wage differentials are not favorable. It is then possible by "experiencing" the job to accumulate and update previous information and conditional on it decide to quit or not the chosen job. This "risky" job has a premium over differential rates of return conditional on the option to quit if it turns to be an undesirable (ex-post) match.

What is the connection between these theories of job-mobility and migration? : there are several stylized facts of the migration process that have not been successfully explained by the existing "human capital" theory of migration. These facts include the existence of high rates of return migration, strong positive correlation between in and out migration rates in given locations, and migration concentrated at ages below those implied by a present value calculation of returns minus costs. McCall and McCall (1987) and Pessino (1989,1991) showed that these facts can be explained in the context of learning theories as those used in the job mobility literature. Imperfect information is more prevalent in a move between locations than between just jobs in the same location and the empirical consequences are more clear cut. It is rather ambiguous to talk about "return" from a job to a previous one, than to talk and do empirical analysis over returning from a location.

This paper presents to my knowledge the first formal empirical test for one of the implications of the learning-matching literature, namely, if some people move for the first time for learning reasons, the conditional probability of remigration will increase with duration in the chosen location for some time and decrease thereafter. There is no reason to expect this pattern to occur in the alternative theory; that of preplanned mobility.

Pessino (1991) provided initial empirical evidence in favor of the sequential migration model. This paper extends the implications of her model to account for structural duration dependence. The purpose of this extension is to provide further implications for the empirical analysis of individual migration histories. Section II presents the extension of the two-period model of sequential migration to M periods. This extension is necessary to show the pattern of dependence of the conditional probability of remigration (the hazard) on duration. Section III proves that the hazard rate of remigration exhibits first positive duration dependence and then negative duration dependence. That is, the learning model implies a special functional form for the probability distribution of remigration times. Section IV discusses the empirical implications of this result. A natural test that arises from this result is that the hazard rate of migration for the first time will be different from the hazard rate of remigration (the return move) if bayesian learning is present. A second implication is derived from the comparison of this hazard with the hazard rate of remigration in a preplanned model of migration. If people "preplan" in advance a fixed time of returning to their original location, the hazard rate in this case will present duration dependence that is the result mainly of pure heterogeneity in the population. Section V derives the likelihood function for the model if the analyst had available "ideal" duration of residence data. The actual duration data with the different types of censoring is described in Section VI. There, the likelihood function used in the analysis is derived taken into account the format of the data. Section VII presents the empirical findings using individual migration histories from Peru. The analysis concentrates in the flows of migration from Lima and into Lima taking into account the order of the migration event. Section VIII concludes with a brief summary.

II. Sequential Migration Model: M periods.

To show the dependence of the hazard rate of remigration on duration it is necessary to extend the two period model of Pessino (1991) to an M period model, where $m=1,\ldots M$. This is an M-priod two-armed bandit problem with normal arms. The simple decision rules IN_1 and IN_2 for the two period model get very messy in the M period case. For this reason, it is desirable to get a recursive expression for the M period decision rule. There are a number of results that are needed to show this recursion.[2]

The structure of the model is analogous to the one presented in Pessino (1991). The potential wages in the two locations $i=1,2$ can be written as:

$$X_{im} = \theta_i + \sigma_i \epsilon_{im} , \qquad (1)$$

where

$$\begin{aligned} \epsilon_{im} &\sim N(0,1) \\ \theta_i &\sim N(\mu_i,\rho_i^2) \\ X_{im}/\theta_i \text{ and } I_t &\sim N(\theta_i,\sigma_i^2) \end{aligned} \qquad (2)$$

and where $I_1 = 0$; $I_2 = z_1,\tau(0)$; $I_3 = z_1,z_2,\tau(z_1)$, etc., are the information sets available at the moment of deciding which location to choose. X_{im} in equation (1); the wage in location i at time period m; is a random variable that follows a normal distribution with unknown mean θ_i and known variance σ_i^2. Individuals are assumed to have a prior distribution on the parameter θ_i which is also assumed to be normal with mean μ_i and variance ρ_i^2.

The posterior distribution of θ_i, $P(\theta_i/I_n)$, when all of the n observations were taken in location i is normal with mean $\mu_i(n)$ and variance $\rho_i(n)$:

$$\theta_i/I_n \sim N(\mu_i(n),\rho_i(n)) \tag{3}$$

where

$$\mu_i\,(n) = \frac{\sigma_i^2\mu_i + \rho_i^2 \sum_{i=1}^{n-1} z_i}{\sigma_i^2 + (n-1)\rho_i^2} \quad \text{for } n > 1 \text{ and } \mu_i(1) = \mu_i \tag{4}$$

and

$$\rho_i\,(n) = \frac{\sigma_i^2\,\rho_i^2}{\sigma_i^2 + (n-1)\rho_i^2} \tag{5}$$

The marginal unconditional distribution of X_{im} given that location i has been sampled n times is:

$$X_{im}\,(n) \sim N(\mu_i(n),\frac{\sigma_i^2\,(\sigma_i^2 + n\rho_i^2)}{\sigma_i^2 + (n-1)\rho_i^2}) \tag{6}$$

The optimal value of the problem is the following:

$$V^m(P(\theta_1),P(\theta_2)) = \sup E(\sum_{m=1}^{M} z_m) \tag{7}$$

where $P(\theta_1)$ and $P(\theta_2)$ are the posterior distributions in the respective locations when no observations were taken; i.e, they are identical to the prior distributions of θ_1 and θ_2.

Let V_i^m denote the worth of selecting location i initially and then continuing with the optimal strategy, that is :

$$V_i^m(P(\theta_1),P(\theta_2)) = \mu_i + E[V^{m-1}[P(\theta_1/I_2 = z_1,i),\ P(\theta_2/I_2 = z_1,i)/P(\theta_1),P(\theta_2)]] \tag{8}$$

We can now state the following lemmas:

LEMMA 1. $V^m(P(\theta_1),P(\theta_2)) = \max\ [V_1^m(\cdot),V_2^m(\cdot)]$

The individual will choose location 1 if the worth of choosing in the first period location 1 and continue optimally exceeds the worth of choosing first location 2 and then proceed optimally.

LEMMA 2. An optimal strategy is given by

(i) Go to location 1 if $V_1^m(\cdot) > V_2^m(\cdot)$

(ii) Go to location 2 if $V_1^m(\cdot) < V_2^m(\cdot)$.

Now, instead of using as state of the problem, the posterior distribution of the process state, it is preferable to use the posterior moments of the normal distribution and characterize V_1^m and V_2^m by using the following recursions:

$$V_1^m(\mu_1,\mu_2,\rho_1^2,\rho_2^2) = \mu_1 + \int V^{m-1}\left(\frac{\mu_1\sigma_1^2}{\sigma_1^2+\rho_1^2} + \frac{X_{11}\rho_1^2}{\sigma_1^2+\rho_1^2},\mu_2,\frac{\rho_1^2\sigma_1^2}{\sigma_1^2+\rho_1^2},\rho_2^2\right) dP(X_{11})$$

We can now standardize X_{11} using the distribution defined in (6) to obtain:

$$V_1^m(\mu_1,\mu_2,\rho_1^2,\rho_2^2) = \mu_1 + \int V^{m-1}\left(\mu_1 + \frac{\rho_1^2}{(\sigma_1^2+\rho_1^2)^{1/2}} z_1,\mu_2,\frac{\rho_{11}\sigma_1^2}{\sigma_1^2+\rho_1^2},\rho_2^2\right) d\Phi(z_1)$$

where $X_{11} = \mu_1 + (\sigma_1^2+\rho_1^2)^{1/2} z_1$ and Φ is the cdf of the standard normal. In an analogous way, we can get $V_2^m(\cdot)$.

The interpretation of the above equation is the following. The value of choosing location 1 when there are m periods to go is a function of the current information on the two locations: the prior means μ_1 and μ_2 and the prior variances ρ_1^2 and ρ_2^2. This equals the prior mean in location 1 plus the value of proceeding optimally when there are m-1 periods to go as a function of the posterior mean and variance in location 1 when one observation has been taken integrated over all possible values of the normal random variable z_1. A similar interpretation can be given to the value of choosing first location 2.

Now, it is possible to define the advantage of location 1 over location 2 in an m period problem as:

$$\Delta^m(\mu_1,\mu_2,\rho_1^2,\rho_2^2) = V_1^m(\mu_1,\mu_2,\rho_1^2,\rho_2^2) - V_2^m(\mu_1,\mu_2,\rho_1^2,\rho_2^2) \tag{9}$$

and so, location 1 is selected when there are m periods to go iff $\Delta^m > 0$. Let's additionally define:

$$\Delta^m(\cdot)^+ = \max(0,\Delta) = \Delta^m(\cdot)\, 1(\Delta^m > 0)$$

$$\Delta^m(\cdot)^- = \max(0,-\Delta) = \Delta^m(\cdot)\, 1(\Delta^m < 0)$$

where

$$1(\Delta^m > 0) = 1 \text{ if } \Delta^m \geq 0$$

$$1(\Delta^m > 0) = 0 \text{ if } \Delta^m < 0$$

Let V_{12}^m be the optimal value of beginning sampling location 1 and then move to location 2 and continue optimally and V_{21}^m be the optimal value of beginning sampling location 2 and then move to location 1 and continue optimally. Note that since the order of receiving information is irrelevant, $V_{12}^m = V_{21}^m$.[3] Therefore, to get the recursion for $\Delta^m(\cdot)$, V_{12}^m can be subtracted and V_{21}^m added to (9):

$$\Delta^m(\cdot) = V_1^m(\cdot) - V_{12}^m(\cdot) + V_{12}^m(\cdot) - V_2^m(\cdot)$$

so, we get

$$\Delta^m(\mu_1,\mu_2,\rho_1^2,\rho_2^2) = \int \Delta^{m-1}\left(\mu_1 + \frac{\rho_1^2}{(\sigma_1^2+\rho_1^2)^{1/2}}z_1,\mu_2, \frac{\rho_{11}\sigma_1^2}{\sigma_1^2+\rho_1^2},\rho_2^2\right)^+ d\Phi(z_1)$$
$$+ \int \Delta^{m-1}\left(\mu_1,\mu_2 + \frac{\rho_2^2}{(\sigma_2^2+\rho_2^2)^{1/2}}z_1,\rho_2^2, \frac{\rho_{22}\sigma_2^2}{\sigma_2^2+\rho_2^2}\right)^- d\Phi(z_1) \tag{10}$$

Through the following lemma, $\Delta^m(\cdot)$ can be reparameterized to depend on the difference in means $\delta = \mu_1 - \mu_2$.

LEMMA 3. If $\mu_1' - \mu_2' = \mu_1^* - \mu_2^*$ then $\Delta^m(\mu_1',\mu_2',\rho_1^2,\rho_2^2) = \Delta^m(\mu_1^*,\mu_2^*,\rho_1^2,\rho_2^2)$ so that $\Delta^m(\mu_1,\mu_2,\rho_1^2,\rho_2^2) = \Delta^m(\mu_1 - \mu_2,0,\rho_1^2,\rho_2^2) = \Delta^m(0,\mu_1 - \mu_2,\rho_1^2,\rho_2^2) = \Delta^m(\delta,\rho_1^2,\rho_2^2)$.

Then, the recursion for the M period problem can be reparametrized as:

$$\Delta^m(\delta,\rho_1^2,\rho_2^2) = \int \Delta^{m-1}\left(\delta + \frac{\rho_1^2}{(\sigma_1^2+\rho_1^2)^{1/2}}z_1, \frac{\rho_{11}\sigma_1^2}{\sigma_1^2+\rho_1^2},\rho_2^2\right)^+ d\Phi(z_1) \tag{11}$$

$$+ \int \Delta^{m-1}\left(\delta + \frac{\rho_2^2}{(\sigma_2^2+\rho_2^2)^{1/2}} z_1, \rho_1^2, \frac{\rho_{22}\sigma_2^2}{\sigma_2^2+\rho_2^2}\right) d\Phi(z_1)$$

In the two period model, Pessino (1991) showed that Δ^m is an increasing function of δ for given variances, and an increasing function of ρ_1^2, ceteris paribus. By induction, similar statements for the M period model can be proved. These lemmas are stated below.

LEMMA 4. $\Delta^m(\delta,\rho_1^2,\rho_2^2)$ is strictly increasing in δ.

LEMMA 5. $\Delta^m(\delta,\rho_1^2,\rho_2^2)$ is strictly increasing in ρ_1^2 and strictly decreasing in ρ_1^2.

The following lemma states that there is a unique δ^* such that $\Delta^m = 0$.

LEMMA 6. There exists a unique δ^* such that $\Delta^m(\delta^*,\rho_1^2,\rho_2^2) = 0$ and that satisfies

if $\delta \geq \delta^*$ sample location 1

if $\delta < \delta^*$ sample location 2

That is, δ^* is the differential reservation wage and δ is the current difference in posterior means. This lemma follows from the uniform continuity of $\Delta^m(\delta,\rho_1^2,\rho_2^2)$ in δ. For a proof of the above results see Fahrenholtz (1982).

In next section, these results are used to prove the dependence of the hazard rate on duration of stay in a given location.

III. The Hazard Function and its Dependence on Duration

This Section shows that the hazard rate of remigration first increases with duration and then decreases monotonically after some time.

In this model the individual sequentially optimizes the remaining stream of wages after observing realizations of productivity in the chosen location. It is assumed in this Section that the individual chooses in the first period to sample location 1 so that he updates each period the expected productivity in location 1. If location 1 was chosen at the beginning of his working career is because there was a positive wage differential in location 1 and/or a payoff to learning by experiencing location 1; i.e. the index function in equation (11) was positive. Suppose that this individual has already observed n realizations of productivity in location 1. In this case, the decision rule at time n+1 will dictate to exit this location if $\Delta^{m-n} < 0$, where the arguments of Δ^{m-n} are now the updated mean and variance after n periods of experience in location 1. The first time n after the initial decision that Δ^{m-n} becomes negative is called the remigration time.[4] At this point it is useful to rewrite the model in terms of the hazard function. The hazard function is a conditional density of exit times from a location given the length of time spent in that location. We want to find out what is the dependence of this conditional probability of leaving location 1 on the number of periods n spent sampling that location. From equation (11) and lemma 6 this is the probability that the current difference in posterior means is less than the differential reservation wage given duration of stay in the current location. Let's define $\delta(n) = \mu_1(n) - \mu_2(n)$ as the difference (after n sampling periods in location 1) between the posterior mean in location 1 and the fixed mean in location 2. Let's denote by $\delta^*(n)$ the differential reservation wage when n observations have been already taken in location 1. Then, If $\delta(n) < \delta^*(n)$ the individual exits location 1.

Denote by $G(\delta(n + 1);\delta,n)$ the conditional probability of $\delta(n + 1)$ given $\delta(n) = \delta$. Note that $\delta(n + 1)$ can be written recursively as a function of $\delta(n)$ and $\rho_1(n)$ the posterior variance in location 1 that given the fixity of μ_2 it is also the posterior variance of the difference in productivities. The recursion is the following:

$$\delta(1) = \mu_1 - \mu_2$$

$$\delta(2) = \frac{\sigma_1^2}{\sigma_1^2+\rho_1^2}\,\mu_1 + \frac{\rho_1^2}{\sigma_1^2+\rho_1^2}\,X_{11} - \mu_2$$

$$= \frac{\rho_1(2)}{\rho_1(1)} (\mu_1 - \mu_2) + (1 - \frac{\rho_1(2)}{\rho_1(1)}) X_{11}$$

$$\delta(n + 1) = \frac{\rho_1(n + 1)}{\rho_1(n)} \delta(n) + (1 - \frac{\rho_1(n + 1)}{\rho_1(n)}) X_{1n} \tag{12}$$

So, $\delta(n + 1)$ conditional on $\delta(n) = \delta$ follows a normal distribution with mean δ and the following variance:

$$\text{var}(\delta(n + 1)/\delta(n) = \delta) = (1 - \frac{\rho_1(n + 1)}{\rho_1(n)})^2 [\rho_1(n) + \sigma_1^2]$$

$$= \frac{\rho_1(n + 1)\, \rho_1(n)}{\sigma_1^2} \tag{13}$$

So, we can see that the variance of $\delta(n + 1)$ conditional on $\delta(n) = \delta$ decreases with n because $\rho_1(n)$ and $\rho_1(n + 1)$ decrease with n. In conclusion $G(\delta(n + 1);\delta,n)$ is a normal distribution representing the distribution of the wage differential at the next observation of productivity in location 1 given that $\delta(n) = \delta$. In this dynamic programming model when there are m periods to go and the individual has not yet sampled any of the locations the decision rule Δ^m depends on the values of the prior means and variances, $\delta(1) = \delta$, $\rho_1(1) = \rho_1^2$ and $\rho_2(1) = \rho_2^2$. Given that at each successive period the individual reoptimizes with his new information, the decision rule after n observations in location 1 Δ^{m-n} can be written as depending on $\delta(n)$, $\rho_1(n)$ and $\rho_2(n)$.

In this model, the probability that the worker remigrates after sampling n times location 1 and observing $\delta(n) = \delta$ is the probability that the new wage differential is less than $\delta^*(n + 1)$, the reservation wage differential with a sample size of n + 1. So, the hazard rate of remigration $h(\delta,n)$ is:

$$h(\delta,n) = G(\delta^*(n+1);\delta,n) \tag{14}$$

Note that

$$\frac{\partial h(\delta,n)}{\partial \delta} < 0$$

The theoretical remigration rate holding sample size constant is decreasing in the wage differential because the higher the current wage differential the less likely that any future wage differential will be below the reservation wage as a consequence of the positive autocorrelation in the wage process implied by the model.

To show how $h(\delta,n)$ changes with duration in location 1, it is necessary to show the dependence of the differential reservation wage on n. This is shown in the following lemma.

LEMMA 7. $\delta^*(n)$ is non decreasing in n.

Proof:

Define by $IN(n) = \Delta^{m-n}(\delta^*(n),\rho_1(n),\rho_2(1)) \equiv 0$ where in IN(n) we include explicitly the dependence of $\rho_1(n)$ on n. Note that $\rho_1(1) = \rho_1^2$.

Differentiate IN(n) with respect to n to obtain:

$$\frac{\partial IN(n)}{\partial n} = \frac{\partial \Delta}{\partial \delta^*}\frac{\partial \delta^*}{\partial n} + \frac{\partial \Delta}{\partial \rho_1}\frac{\partial \rho_1(n)}{\partial n} = 0$$

and this implies that $\partial\delta^*(n)/\partial n > 0$; the differential reservation wage increases with n, and when $\partial\rho_1(n)/\partial n = 0$ we will have that $\partial\delta^*(n)/\partial n = 0$. This is because $\partial\Delta/\partial\rho_1 > 0$ by lemma 5, $\partial\rho_1(n)/\partial n < 0$ by (5), and $\partial\Delta/\partial\delta > 0$ by lemma 4.

The reservation differential wage that makes one indifferent between staying or returning increases with duration precisely because the posterior variance of the unknown mean decreases, so that the learning motive that provided the incentive to move in the first place dissipates as time goes by. As duration increases, the prediction of future productivity in location 1 becomes increasingly more precise. In the limit, the worker's average productivity θ_1 is known with certainty, provided, of course, that he/she has not yet decided to leave. Once remigrated, the worker will not return.

Now, we can establish the behavior of the hazard rate with respect to n. Differentiate equation (14) with respect to n:

$$\frac{\partial h(\delta,n)}{\partial n} = \frac{\partial G}{\partial \delta^*}\frac{\partial \delta^*}{\partial n} + \frac{\partial G}{\partial n} \qquad (15)$$

The change in the hazard rate of remigration attributable to an increase in the sample size is the sum of two effects corresponding to the two terms on the right hand side of equation (15). The first term is the change due to the change in the differential reservation wage. Because G is a distribution function, $\partial G/\partial \delta^* > 0$ and by lemma 7, $\partial \delta^*/\partial n > 0$, this effect is always positive but will diminish to zero as n becomes large as a consequence of the convergence of the posterior variance of θ_1 to zero. The second term is the change attributable to the decrease in the variance of the next wage induced by the increase in sample size. Because $G(\cdot)$ is the normal distribution function, a decrease in variance reduces its value to the left of the mean and increases its value to the right. So the second term in equation (15) is negative if $\delta > \delta^*(n + 1)$. And for long enough durations, given that most people with $\delta < \delta^*(n + 1)$ would have already left, this term will be negative. Summing up, for short durations both terms in equation (15) will be positive. As time goes by, the first term converges to zero and the second becomes negative. This is the main result, namely: the hazard rate of remigration increases with duration and after some time, enough to learn about productivity in location 1, it turns to depend negatively on duration. This result is not new in the learning literature. It has been proved by Jovanovic (1979) and Mortensen (1985) in the context of job matching models. The proof of the result in this section followed the proof presented by Mortensen adapted to the two-armed bandit problem.[5]

IV. Empirical Implications

The result of last Section presents a number of interesting empirical implications for the dynamics of migration. I will analyze household data from Peru, and specifically movements in and out of Lima. I will refer to the empirical implications making explicit reference to this case; however, the implications of this model are much more general.

Lima-Callao is the area of Peru that concentrates the highest proportion of the population and that according to the 1981 Census had the highest rate of in and out migration. I will analyze movements in and out of Lima from other areas. Other areas include both Other Urban areas and Rural areas.[6] The analysis of movements in and out of Lima is twofold. Movements into Lima (that I will sometimes call movements from "Other") are divided into primary moves and return moves. Primary moves are made by people beginning the process in "Other" and return moves by people beginning the process in Lima but having moved to "Other" in their first event. Movements from Lima (that I sometimes will call movements into "Other") are also divided into movements made by primary migrants and by secondary migrants. Primary migrants from Lima are those individuals who begin their working career in Lima and make a move to "Other" for the first time. Secondary migrants are people born in "Other" but who had made a first move into Lima so they are at risk of remigrating from Lima.

What has the theory of sequential migration to say about the hazards; i.e the conditional probability of exiting Lima or "Other" given duration of stay, for primary and secondary migrants?. How much time do individuals spend in "Other" areas or Lima before they decide to migrate?. How much time do individuals spend in Lima before deciding to remigrate (if ever) again?. How does the duration of stay in each region vary across individuals?. The length of stay in each location and the determinants of whether or not to exit given the past history of migration are key ingredients of a sequential migration process.

The following is a list of empirical implications of the model:

(i) Positive and eventually negative duration dependence is implied by this model. Only young workers will take the risk of migrating for the purpose of learning. In this way they can cut off the left tail of the wage distribution and have the option to remigrate if the match is not satisfactory. The length of time before negative duration dependence settles in is an empirical matter. However, opportunity costs and specific human capital costs will dictate that the period should not be that long in years. This functional form for the hazard rate is especially pertinent to secondary movements out of Lima; that is the hazard rate of remigration from Lima. I do not expect the same type of result (or not as strong) for the hazard rate of migration from "Other" or from Lima for the first time because it is less likely that persons beginning their careers in those places have so much to learn as those that have moved to a different place. If one views θ_i as a vector of location-worker characteristics, the elements corresponding to location specific amenities will trivially be better known by the individual born in location i. On this account, the hazard of migration for the first time will not present an initial positive slope or if positive not that important because less people will be expected to fail after beginning their working career in the better known location than in the distant and unknown one.

(ii) Is remigration preplanned or is the remigration that often follows an initial migration episode a consequence of unsuccessful labor market outcomes?. Previous work on the determinants of multiple moves have emphasized an alternative view of the remigration decision. The preplanned model of migration claims that migrants return because they have planned that in advance; it does not depend on the outcome of migration.[7] The main distinction between the preplanned model of migration and the sequential model is that the former is deterministic in its extreme version. The individual that migrates does so knowing in advance that he/she will return at some prespecified date. The reason for this behavior is presumably the desire to accumulate some amount of assets that is not possible in the origin destination given the absence or incompleteness of capital markets. So, at the individual level there is no stochastic process governing the conditional probability of returning as compared to the sequential migration model. It is only when an assumption is made about the distribution of dates/target wealth of these individuals in the population that an aggregate hazard rate possibly duration dependent emerges as a result of pure heterogeneity in the population. It is the distribution of different fixed dates of departure that creates spurious duration dependence. This claim is illustrated by considering the sequence of probabilities of terminating a location match at specified duration levels, derived from sequentially maximizing an intertemporal utility function subject to income and information constraints. We called this sequence $\{h(\delta,n)\}$. Contrast this setup with a very simple model in which a population of identical agents stay in the location according to the fixed time $\{i=1,..n\}$ that takes to accumulate their target wealth. Suppose that the proportion of people that accumulate wealth during i periods is hi. Then, clearly, the second model can generate duration dependence only as a consequence of the various fixed dates of departure. In the first model, each individual probability is duration dependent. Compared to the hazard function generated by the sequential migration model, the hazard for a preplanned model will not present a decreasing hazard at long enough durations.[8] Furthermore, controlling for observed and unobserved heterogeneity, the individual hazard rate is predicted to be approximately degenerate at some fixed target date. Only, when we aggregate across individuals the preplanned model can generate duration dependence.

(iii) The sequential migration model predicts a positive correlation between in and out migration flows in given locations, usually in prosperous or more developed ones. By analyzing the serial order of migration events out of each location, Lima and "Other", the empirical model can help to discriminate between the two explanations given for this phenomenon. In the sequential model, new migrants with poor realizations from the wage distribution are more likely to leave in a subsequent period. In contrast, static models presume that this phenomenon is due to cross-section heterogeneity, that different people with different skills will enter a location and some others will decide to leave it at the same time. This issue is analyzed in Section 7 by comparing the flows in and out of Lima for first time migrants and secondary migrants.

(iv) Do education and family background variables have similar effects on the hazard rates for the first and subsequent movements?. Do they differ by the serial order of the spell and by location?. The theory of sequential choice predicts that wage differentials and information differentials are the two key ingredients of the first migration event. For subsequent migration events, we have to add the realization of the wage in the chosen location that serves to update both wage and informational differences. How then does education affect the conditional probability of migration and remigration?. Most of the literature on return migration expects to find a negative effect of education on the probability of remigration. In the sequential model, the effect is not that clear. Education affects the information factor in two ways. On the one hand, better educated individuals are expected to have better information previous to a move, so on this account they will tend to migrate less to acquire information than the less educated. On the other hand, more educated people are expected to learn faster and this will tend to increase the payoff to learning by moving. So these two effects work in the opposite direction and so it is an empirical question how education affects the likelihood of the different types of moves.

V. Likelihood Function for Ideal Data

With these issues in mind, a continuous time semi-Markov model of the migration process will be implemented. As in existing models of discrete choice, the probability that an individual will opt for a particular location depends on observed characteristics of the location and the decision maker. This discrete time model is generalized to continuous time by regarding the time between moves as a random variable whose expectation may depend on the current state occupied and the observed explanatory variables. Given wage realizations, the prospective migrant faces a discrete choice problem - whether to stay in the given location or move (back) to another one.

The individual is assumed to select that location that yields the highest expected present value of income given current information, the current location and the length of the current spell. The semi-Markov specification allows for the presence of duration of residence effects. The hazard rate in a particular spell will depend on the duration of stay in the current location. In contrast, a Markov model specifies the probability that an individual will change locations as depending only on the location currently occupied. Previous residence history plays no role on subsequent choice under the Markovian hypothesis.

The econometric model proposed in this Section can be regarded as a reduced form resulting from behavioral models. Reduced form results can serve to rule out some potential structural models, but cannot distinguish between others. The maximization model implies a sequence of (conditional) probabilities that an individual will migrate given that he had not previously done so.

I will assume in this Section that the investigator has access to "ideal" migration data. By that

I mean that the completed durations of stay in each location are available for a random sample of the population.

Locations are divided for the analysis into "Lima" and "Other" areas, denoted L and O respectively. The migration history of each individual will be governed by the exit time distribution from "Lima", $f_L(t_L)$ and the exit time distribution from "Other", $f_O(t_O)$. As usual, it is convenient to express the analysis in terms of hazard functions, $h_L(t_L)$ and $h_O(t_O)$. The hazard function is the conditional density of exit times from a location given the length of time already spent in that location. Let S = L,O then the hazard rate in location S is:

$$h_S(t_S) = \frac{f_S(t_S)}{1 - F_S(t_S)} \qquad (16)$$

where $1 - F_S(t_S)$ is the probability that the duration of residence exceeds t_S, i.e, the survival function which can be expressed in terms of the hazard as:

$$1 - F_S(t_S) = S_S(t_S) = \text{Exp}\left[- \int_0^{t_S} h_S(U)dU\right] \qquad (17)$$

From (16) and (17) the density of exit times can be expressed in terms of the hazard function as:

$$f_S(t_S) = h_S(t_S) \,.\, \text{Exp}\left[- \int_0^{t_S} h_S(U)dU\right] \qquad (18)$$

Duration dependence exists if $\partial h_S/\partial t_S \neq 0$. If $\partial h_S/\partial t_S > 0$ there is positive duration dependence. If $\partial h_S/\partial t_S < 0$ there is negative duration dependence. If $\partial h_S/\partial t_S > 0$ for some $t_S < t^*$ and $\partial h_S/\partial t_S < 0$ for some $t_S > t^*$ the hazard is nonmonotonic, first increasing and then decreasing.

With complete duration data, the individual likelihood function is the product of the density of exit times from each location. By multiplying this term over individuals one obtains the likelihood function for the migration data. Let $i = 1,\ldots,N$ index individuals and $L_{ji} = 1,\ldots,m_{Li}$, $O_{ji} = 1,\ldots,m_{Oi}$ the number of total completed spells in Lima and "Other" respectively for each individual. Then the likelihood function is:

$$L = \prod_{i=1}^{N} \left[\prod_{L_{ji}=1}^{m_{Li}} f_{L_{ji}}(t_{L_{ji}}{}^{i}) \prod_{O_{ji}=1}^{m_{Oi}} f_{O_{ji}}(t_{O_{ji}}{}^{i})\right]$$

This general parametrization permits behavioral coefficients to differ depending on the location and the serial order of the spell. Such shifts in coefficients have been termed "occurrence dependence" by Heckman and Borjas (1980).

This "ideal" migration data is generally unavailable for migration histories. The likelihood function and the empirical analysis in general must be adjusted to match the available data. In next Section, I discuss how the duration and transition data are recorded and how to specify the likelihood function for the actual data.

VI. Likelihood Function for Actual Duration Data

The data set from Peru has retrospective information on migration events. However, some of these events and their respective durations are incompletely recorded.

The Survey reports for each individual (15 years of age or older) their birthplace, previous place of residence and current residence. It also reports when the last move was made and the age at which the individual first left his/her birthplace. So, the data set includes at most three locations for each individual: the place of birth, the last place of residence and the current place. The possible patterns of residence for each individual are the mutually exclusive sequences: O-L-O, O-L-L, O-O-O, L-O-L, L-O-O and L-L-L. For example O-L-O means that the individual began the process in "Other", the last place of residence was "Lima" and currently resides in "Other"; O-L-L means that the individual began the process in "Other", moved to "Lima" and stay there till the time of the Survey; and O-O-O means that the individual never moved to Lima.

Figure 1 shows the possible patterns of lifetime migration for each individual originating in "Lima" and in "Other" as recorded in the data set.[9] The first part of the figure displays the possible paths of the 76 percent of the men that began their migration history in "Other" and the second gives the path of the 24 percent of the men who began their migration history in "Lima". The upper branch of each node (L) indicates a move to Lima, and the lower branch (O) indicates a move to "Other". I will consider in the empirical analysis a first spell out of "Other" (FO) with 3134 individuals, of whom 27 percent migrated to Lima and a second spell out of "Lima" (SL) where out of 840 individuals 29 percent remigrated to "Other". I will also consider a first spell in "Lima" (FL) of 983 individuals, of whom 17 percent exited. The second spell out of "Other" (SO) will begin with these 171 individuals at risk, of whom 82 percent remigrates to "Lima".

As reported in figure 1, two is the maximum number of moves reported in the data for which the locations are known. If the individual made actually more than two moves the intermediate locations are not known. Only the total number of moves are recorded but not the duration, origin or destination for all of the events. Given these facts, the larger the number of moves the person reported, the less accurate the information. Younger cohorts have actually moved less number of times than older cohorts. This augmented with the fact that any economic theory of migration (and especially the learning model) applies to the young implies that the data for

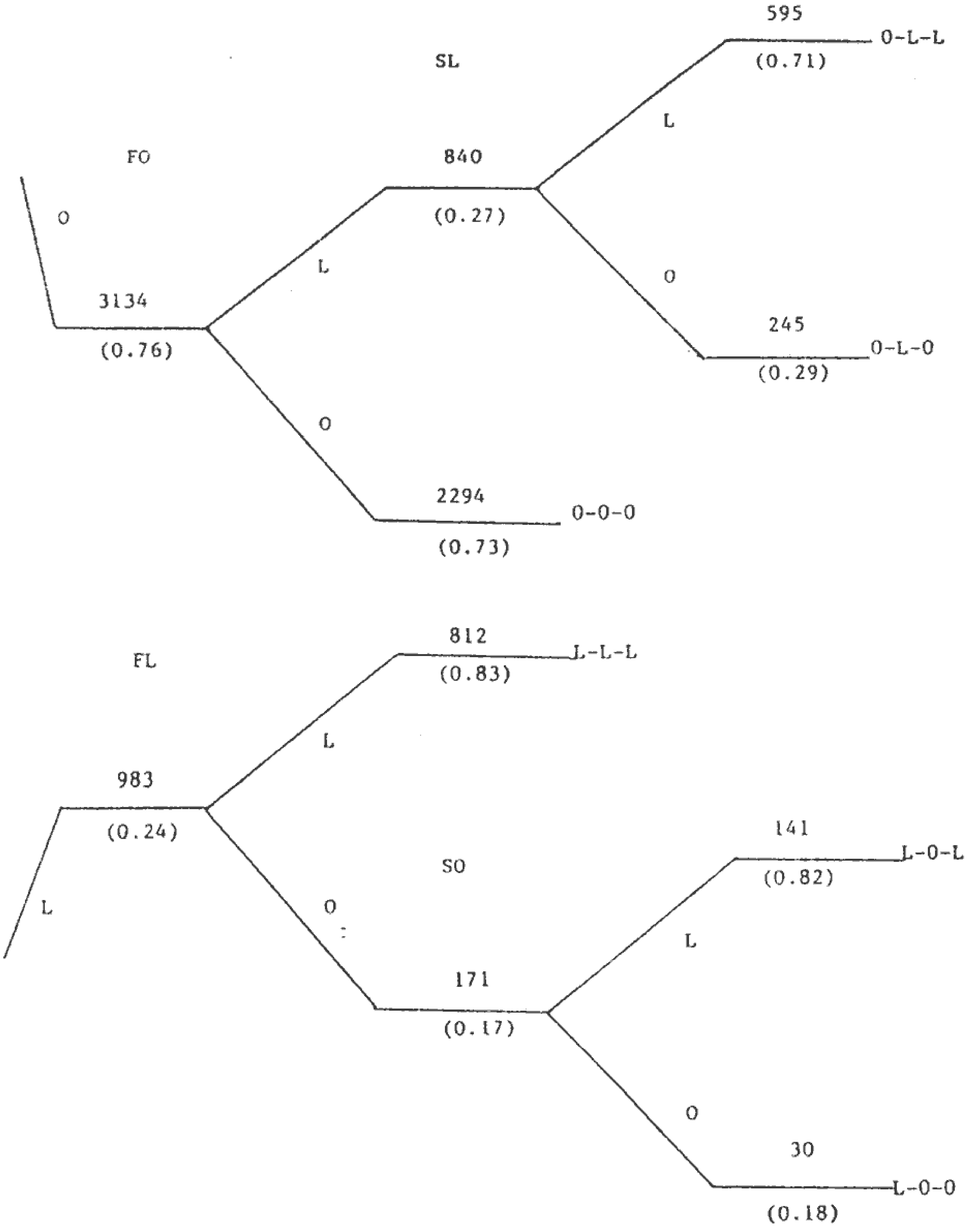

Figure 1.--Observed lifetime movements in and out of Lima for sample of Peruvian males

younger cohorts is the most accurate for any meaningful interpretation of the theory. If some individuals from older cohorts perform more than two moves, the last move recorded might have been performed at an older age and the moves following the first move from his/her birthplace are lost. This second move that is lost for some individuals is the crucial move for implications on the learning-matching model. However, even for individuals that made more than two moves, we know the maximum amount of time they spend in the second location.

In the discussion that follows, I will concentrate on the information available for movements originating in "Other". The characteristics of the duration data for "Lima" are symmetric to this case. Considering all the possible migration events originating in "Other", the following complete or censored information is available:

(1) For transitions "O-O-O" , we know the migration history up to censoring on the right. The information available is that the individual survived till the current date of the Survey without making a transition into Lima. This is accurate as long as the individual performed two or less moves in his/her lifetime.

(2) For "O-L-L" transitions we have complete duration information for the first spell in "Other" and right censored information for duration in "Lima". The same qualifications apply as in (1) if the person made more than two moves.

(3) For "O-L-O" individuals, the timing and duration information is not complete in any spell if the person reported making more than two moves. In this case (and in L-O-L) we know that a first movement occurred but not exactly at what point in a given interval.[10]

The likelihood function of last Section has then to be modified to take into account the information available.

The last aspect to consider is the time at which the process begins. The ideal is to have a date (say, age of the individual) at which the individual begins his working career. However, that information is not provided by the Survey. In the empirical analysis, I experimented with a beginning age of 12 years old which is the age at which most individuals end primary school. If the individuals made a move before age twelve, they were treated as though they began their migration history at the new location. The case O-L-O was treated somewhat differently. Of the 245 individuals that performed this sequence, 19 percent reported leaving their birthplace at an age less than 12 years old. Given that the variable "age at which left birthplace" provides a lower bound for the interval at which migration took place, I decided to keep these individuals in the current sequence instead of moving them to the "L O O" category.

A key distinction in the sequential model of migration is that spells in different locations differ in terms of their order. A different hazard will characterize movements out of Lima and out of "Other" and first or second spells out of these locations. In particular, for the hazard function for the second spell out of Lima, the remigration decision, it is expected that a non-monotone hazard, first increasing and then decreasing characterizes the process.

With these issues in mind, I will derive the likelihood function for the first spell out of "Other", the first spell out of "Lima" is analogous. The following terms contribute to the likelihood function:

i) Right censored observations, the sequence "O-O-O", they contribute the probability of survival more than the censoring time:

$$S_o^i(t_o^i) = \text{Exp}\left[- \int_0^{t_s} h_o^i(U)dU\right] \tag{20}$$

where t_o^i is the total duration of stay in "Other" (i.e, age at Survey minus twelve years old).

ii) Uncensored observations, the "O-O-L" transition; its contribution to the likelihood is the density of exit times:

$$f_O^i(t_O^i) = h_O^i(t_O^i) \cdot \text{Exp}\left[-\int_0^{t_O^i} h_O^i(U)dU\right] \tag{21}$$

and, finally,

(iii) interval censored observations, the sequence "O-L-O". The upper limit of this interval is the timing of the last migration event, call it t_{OU}^i. The lower limit t_{OL}^i is the age at which they left their birthplace minus 12 years. This term is:

$$F_O(t_{OU}^i) - F_O(t_{OL}^i) \tag{22}$$

that is the difference between the cdf for duration at t_{OU}^i and the cdf at t_{OL}^i.

Let's define the following dummy variables:

$\lambda_i = 1$		if observation is right censored
$= 0$		otherwise;
$d_i = 1$		if observation is interval censored
$= 0$		otherwise.

Then, the log-likelihood function for the first spell out of "Other" is:

$$L = \sum_{i=1}^{N}(1 - d_i)(1 - \lambda_i)\log f_O^i(t_O^i) + \sum_{i=1}^{N}(1 - d_i)\lambda_i \log S_O^i(t_O^i) + \sum_{i=1}^{N} d_i(1 - \lambda_i)\log[F_O^i(t_{OU}^i) - F_O^i(t_{OL}^i)] \tag{23}$$

The following terms contribute to the likelihood for the second spell out of Lima (the likelihood for the second spell out of "Other" is analogous):

(i) right censored observations for individuals with the sequence "O-L-L":

$$S_L^i(t_L^i) = \text{Exp}\,[-\int_0^{t_L} h_L^i(U)dU] \tag{24}$$

(ii) uncensored observations, for individuals with the sequence "O-L-O" and who performed exactly two movements in their lifetime:

$$f_L^i(t_L^i) = h_L^i(t_L^i)\,.\,\text{Exp}\,[-\int_0^{t_L^i} h_L^i(U)dU] \tag{25}$$

and, finally

(iii) interval or left censored observations, for individuals with the sequence "O-L-O" and who performed more than two movements in their lifetime. These individuals exit Lima at the timing of the last migration event but they enter at some point after the age they left their birthplace. The duration of this spell is at most the age they last migrated minus the age at which they left their birthplace. Their contribution to the likelihood is:

$$F_L^i(t_{LU}^i - t_{LL}^i) \tag{26}$$

Then, the log-likelihood for the second spell out of Lima is:

$$L = \sum_{i=1}^{N}(1 - d_i)(1 - \lambda_i)\log f_L^i(t_L^i) + \sum_{i=1}^{N}(1 - d_i)\lambda_i \log S_L^i(t_L^i) + \sum_{i=1}^{N} d_i(1 - \lambda_i)\log[F_L^i(t_{LU}^i - t_{LL}^i)] \tag{27}$$

To actually implement the estimation of the model I need to specify distributional assumptions for the hazards and the way that covariates influence them. For people born or beginning the migration process in "Lima", we called the conditional probability that the individual will migrate

$h_L(t_L)$ and for people born in "Other", $h_O(t_O)$.[11] In the case of a stationary economic environment and no behavioral duration dependence, the hazard rates are independent of t_S and the duration distribution is exponential; i.e, the hazard rate is constant, $h_S(t_S) = hS$, $S = O,L$.

One way to test for the absence of duration dependence is to assume a Weibull distribution of exit times. The Weibull hazard function is:

$$h_S(t_S) = \alpha t_S^{\gamma-1} \tag{28}$$

The exponential distribution is obtained by restricting $\gamma = 1$. For $\gamma < 1$ the hazard decreases with duration. For $\gamma > 1$ the hazard increases with duration.

A distribution that allows for positive and then negative duration dependence is the log-logistic distribution with parameters $\gamma > 0$ and $\alpha > 0$:

$$h_S(t_S) = \frac{\alpha t_S^{\gamma-1}}{1 + t_S^{\gamma}\, \alpha/\gamma} \tag{29}$$

For $\gamma > 1$, the hazard first increases and then decreases with duration. If $0 < \gamma \leq 1$ the hazard function decreases with duration.

To introduce covariates in the analysis, I assume, as has been the practice in most of this literature, that:

$$\alpha_i = \mathrm{Exp}(-\beta x_i) \tag{30}$$

where x_i is a vector of socioeconomic characteristics of person i. Here, any element of x_i

influences negatively the hazard both in the Weibull and the log-logistic case. Hence, a positive coefficient means that the hazard decreases. In the Weibull case, the resulting hazard will be both of the proportional hazard form and accelerated lifetime form. In the proportional hazard specification, the effect of the regressors is to multiply the hazard function itself by a scale factor. The regressors in the accelerated failure time model have the effect to rescale the time axis. In the log-logistic case, the resulting hazard is of the accelerated failure time form.

The variables used to explain the hazard variation across individuals and spells, to account both for observed heterogeneity and occurrence dependence, are in the first specification that I implement years of schooling and parents' years of schooling. It has been usually expected that own and parents education will positively affect the hazard of leaving "Other"; the less advantageous region; in the first spell. The conventional reason for this result is that highly educated individuals and with better family background will profit more from their human capital investment and endowments and use them in a more productive environment. However, if informational factors are in part determined by educational variables, the effect of education on the probability of migration can be ambiguous. For those that experience a second spell in "Lima", having moved from "Other" in the first transition, it has also been usually expected that family background and schooling variables affect negatively or not at all the hazard of exiting "Lima". The more schooling attributes a person possesses the more probable will be that the individual will decide to stay, but a bad draw or bad "luck" in the draw of the wage distribution may cause him/her to perform a new transition even if he/she has adequate attributes in terms of productivity. On the other hand, if more educated individuals are more prone to migrate for learning reasons than less educated individuals, they will also be more prone to return. These facts will try to be elucidated in the empirical section.

There are other covariates that can influence the decision whether to move or not. Such variables as household size and marital status that are thought to increase the costs of migration are excluded from the analysis because only the current value is known.

Wages have not been included in this initial specification. The reason, being that wages are endogenous to the migration process and are only observed for right censored observations. This first specification, omitting wages, but including some of its determinants can be thought as a reduced form estimation. In a subsequent step, I include predicted wages as an additional covariate. This extension aids in identification of the duration parameter. By not conditioning on wages, an average over all possible wages of the hazard function is being estimated up to the control that educational variables can provide. This tends to bias the results towards negative duration dependence. On the other hand, by entering wages in the estimation, the net effect of education on the hazard of migration and remigration can be estimated.

Some of the individuals will never move either in the first or second spell out of each location. If the control for the effects of explanatory variables is insufficient, heterogeneity will remain and will lead to misleading inferences, especially about duration dependence. In general, not controlling for unobserved heterogeneity leads to downward biased estimates of duration dependence. This is because, as time elapses "movers" will exit the current state or location and the proportion of "stayers" will rise. This will show as a decline in the hazard function over time. To account for heterogeneity, I use the Heckman and Singer (1984) specification of the mover-stayer model with two points of support. I assume that a proportion (1-p) of the sample who have not moved yet will never move, while the proportion p remain at risk of moving. This specification was also used by Walker (1986) and by David, Mroz and Wachter (1985) in studies of duration and timing of births.

Incorporating unobserved heterogeneity, the terms of the likelihood function get transformed in the following way:[12]

(i) $S_s^*(tS) = pS_s(t_s) + (1 - p)$

(ii) $f_s^*(t_s) = pf_s(t_s)$

(iii) $p[F(t_{sU}) - F(t_{sL})]$

and $pF(t_{sU} - t_{sL})$ for the second spell.

VII. Empirical Results

The first and second spell out of Lima and "Other" were estimated for the Weibull and log-logistic specification with and without unobserved heterogeneity correction.

To account for cohort effects and minimize the impact of the biases introduced by the data in the estimated hazards, the sample was divided by cohorts. The sample size does not allow for narrow definitions of cohorts. Hence, I implemented the model using two definitions, the first one divides the sample in three cohorts in the age groups 15-29, 30-44 and 45-65. The second one refines the cohort interval to ten years and so consists of four cohorts: 15-24, 25-34, 35-44 and 45-54. Preliminary results showed that a Wald test for stability of estimates across cohorts was rejected, so I decided to continue the analysis with these two cohort definitions. With these two cohort definitions I am able to compare how sensitive the results are to the width of the cohort interval. The covariates used were years of schooling, YRSCHL, and mother years of schooling, MYRSCHL.[13]

Table 1 summarizes the results for the Weibull specification for each cohort and spell. For the first spell out of Lima the exponential model cannot be rejected at the 1 percent level of significance, except for cohorts 45-65 and 45-54 which exhibit negative duration dependence. For cohorts 15-29 and 25-34, for the first spell out of "Other" and for the cohorts 15-29, 45-65

TABLE 1

MIGRATION PROCESS WITH WEIBULL DURATION DEPENDENCE^

Cohort	15-29	30-44	45-65	15-24	25-34	35-44	45-54	55-65
				(a) First spell Lima (FL)				
Intercept	3.863	4.992	4.886	4.850	4.097	5.156	4.816	5.372
	(6.941)	(12.436)	(10.921)	(5.089)	(8.780)	(9.730)	(8.703)	(5.982)
MYRSCHL	-0.072	-0.058	-0.014	-0.155	-0.079	0.0136	-0.019	-0.035
	(-1.167)	(-1.728)	(-0.269)	(-2.444)	(-2.177)	(0.269)	(-0.286)	(-0.397)
YRSCHL	0.099	-0.047	-0.074	0.091	0.059	-0.106	-0.087	-0.038
	(1.705)	(-1.483)	(-1.631)	(0.896)	(1.424)	(-2.535)	(-1.572)	(-0.464)
γ	0.968	0.913	0.596	1.297	0.992	0.772	0.525	0.829
	(6.957)	(8.580)	(5.924)	(4.589)	(7.963)	(5.901)	(4.784)	(3.392)
Log-L	-253.4	-366.3	-181.2	-114.1	-322.1	-179.1	-110.6	-69.4
N	459	336	188	266	351	178	117	71
				(b) First spell Other (FO)				
Intercept	4.672	4.710	4.629	4.704	4.845	4.648	4.541	4.813
	(21.075)	(33.342)	(36.193)	(14.933)	(23.038)	(27.829)	(28.502)	(21.623)
MYRSCHL	-0.087	-0.033	-0.037	-0.066	-0.072	-0.022	-0.029	-0.046
	(-4.283)	(-1.750)	(-1.488)	(-1.938)	(-3.849)	(-0.852)	(-0.881)	(-1.155)
YRSCHL	-0.061	-0.105	-0.105	-0.077	-0.089	-0.103	-0.109	-0.099
	(-2.551)	(-7.804)	(-6.961)	(-2.105)	(-4.537)	(-6.290)	(-5.365)	(-4.271)
γ	1.094	0.799	0.694	1.214	0.968	0.757	0.694	0.716
	(14.898)	(18.938)	(17.972)	(10.582)	(15.370)	(14.858)	(14.427)	(10.609)
Log-L	-858.5	-1524.5	-1492.9	-441.9	-959.0	-977.1	-944.0	-546.7
N	902	1185	1047	528	783	776	647	400

TABLE 1 (continued)

Cohort	15-29	30-44	45-65	15-24	25-34	35-44	45-54	55-65
				(c) Second Spell Lima (SL)				
Intercept	2.362	2.881	3.355	1.931	3.037	2.706	3.531	3.194
	(6.549)	(12.273)	(15.195)	(3.277)	(8.675)	(10.282)	(11.768)	(9.312)
MYRSCHL	0.105	0.038	0.0001	0.225	0.034	0.050	-0.105	-0.040
	(2.085)	(1.016)	(0.003)	(2.114)	(0.872)	(0.980)	(-0.726)	(-0.588)
YRSCHL	-0.013	-0.008	0.018	0.008	-0.033	0.006	0.0005	0.040
	(-0.303)	(-0.266)	(0.541)	(0.100)	(-0.863)	(0.172)	(0.012)	(0.764)
γ	0.615	0.425	0.322	0.655	0.510	0.394	0.262	0.399
	(6.502)	(7.087)	(5.103)	(4.518)	(6.149)	(5.599)	(3.478)	(3.744)
Log-L	-176.9	-349.2	-251.6	-77.4	-221.0	-227.1	-140.6	-109.5
N	197	342	301	100	222	217	193	108
				(d) Second spell Other (SO)				
Intercept	0.493	1.405	3.316	0.857	0.966	1.467	2.481	5.480
	(0.819)	(3.674)	(5.322)	(0.906)	(1.949)	(2.882)	(3.132)	(3.303)
MYRSCHL	-0.006	-0.065	-0.046	0.221	-0.048	-0.095	-0.040	0.025
	(-0.150)	(-1.482)	(-0.847)	(2.746)	(-1.024)	(-1.423)	(-0.683)	(0.113)
YRSCHL	0.031	0.025	-0.154	-0.189	0.029	0.042	-0.082	-0.386
	(0.484)	(0.636)	(-2.506)	(-1.622)	(0.583)	(0.746)	(-1.067)	(-1.947)
γ	0.801	0.704	0.454	1.085	0.799	0.689	0.383	0.942
	(7.062)	(8.068)	(4.396)	(4.405)	(7.906)	(5.647)	(3.548)	(2.495)
Log-L	-61.5	-118.3	-50.8	-16.0	-90.3	-67.4	-33.0	-15.4
N	50	83	38	22	69	42	24	14

^ Asymptotic normal statistics in parentheses.

15-24 and 55-65 for the second spell out of "Other", the evidence of no duration dependence cannot be rejected. Cohort 15-24 for the first spell out of "Other" is the only one that shows significant evidence in favor of positive duration dependence. All the other cohorts in the remaining transitions show evidence of negative duration dependence.

The results confirm the expectations about the sign and significance of the effect of covariates for some of the cohorts. For most of the cohorts at risk of moving for the first time out of Lima, there is no significant effect of educational variables. If there is any effect, the results imply that better educated individuals are at higher risk of exiting Lima for the first time. Comparing to the first transition out of "Other", years of schooling exert a positive and significant effect at the 1 percent level for all cohorts. Mother years of schooling exerts a positive effect for the younger cohorts only. These results confirm the findings of Pessino (1991): people that move into urban areas are more educated and tend to have better family background than those that stay. For those individuals that made a first move into Lima and came back to "Other" areas, mother's years of schooling exerts a negative and significant effect on the likelihood of returning for the younger cohorts. For cohort 15-29, however, the effect of own years of schooling has negative sign. So, for the younger cohorts better family background increases the likelihood of a first migration event to "Lima" but decreases the likelihood of returning. The effect of the covariates and the duration parameters are not very precisely estimated for the second spell out of "Other" given the small sample size for each cohort. If there is any effect from the own and mother's education covariates, it is to increase the probability of a transition out of "Other" back to Lima. Analyzing this reverse case, where people migrate first from the "developed" region into "Other" areas, the results are not surprising. The reason they moved in the first place seems to have little to do with the sequential model.

Table 2 presents the results of estimating the spells for the Weibull case accounting for

TABLE 2

MIGRATION PROCESS WITH WEIBULL DURATION DEPENDENCE WITH MOVER-STAYER HETEROGENEITY CONTROL

Cohort	15-29	30-44	45-65	15-24	25-34	35-44	45-54	55-65
				(a) First spell Lima (FL)				
Intercept	3.863 (6.941)	4.992 (12.436)	3.975 (5.205)	4.077 (1.149)	4.097 (8.780)	5.156 (9.730)	4.028 (3.959)	4.834 (3.289)
MYRSCHL	-0.072 (-1.167)	-0.058 (-1.728)	-0.039 (-0.557)	-0.169 (-1.497)	-0.079 (-2.177)	0.0136 (0.269)	-0.035 (-0.449)	-0.415 (-1.910)
YRSCHL	0.099 (1.705)	-0.047 (-1.483)	-0.085 (-1.398)	0.109 (0.678)	0.059 (1.424)	-0.106 (-2.535)	-0.102 (-1.419)	0.225 (1.261)
γ	0.968 (6.957)	0.913 (8.580)	0.721 (4.362)	1.359 (2.737)	0.992 (7.963)	0.772 (5.901)	0.607 (-3.547)	1.618 (3.127)
δ	No Admit	No Admit	-0.618 (-0.880)	-0.026 (-0.004)	No Admit	No Admit	-0.346 (-0.302)	-1.218 (-3.621)
Probab. of Stayer			0.65	0.51			0.59	0.77
Log-L	-253.4	-366.3	-180.7	-114.1	-322.1	-179.1	-110.5	-67.0
				(b) First spell Other (FO)				
Intercept	4.357 (15.180)	4.503 (23.571)	4.387 (25.605)	4.704 (14.933)	4.657 (18.192)	4.546 (18.379)	4.370 (20.217)	4.584 (14.975)
MYRSCHL	-0.104 (-3.854)	0.021 (0.700)	0.001 (0.029)	-0.066 (-1.938)	-0.081 (-3.535)	0.053 (1.422)	-0.021 (-0.572)	0.053 (1.046)
YRSCHL	-0.075 (-2.546)	-0.167 (-7.911)	-0.145 (-6.445)	-0.077 (-2.105)	-0.094 (-4.243)	-0.195 (-7.697)	-0.124 (-4.930)	-0.189 (-5.399)
γ	1.227 (12.509)	0.969 (16.144)	0.774 (15.447)	1.214 (10.582)	1.029 (12.875)	0.970 (13.756)	0.733 (12.374)	0.864 (10.009)
δ	0.366 (0.832)	0.147 (0.791)	0.662 (2.010)	No Admit	1.116 (1.570)	-0.022 (-0.129)	1.287 (1.639)	0.160 (0.529)
Probab. of Stayer	0.30	0.46	0.34		0.25	0.50	0.22	0.46
Log-L	-855.2	-1517.2	-1489.6	-441.9	-958.2	-967.2	-943.3	-542.9

TABLE 2 (continued)

Cohort	15-29	30-44	45-65	15-24	25-34	35-44	45-54	55-65
				c) Second spell Lima (SL)				
Intercept	1.924 (3.498)	1.964 (3.693)	1.378 (4.054)	1.405 (1.742)	1.796 (3.102)	1.859 (2.773)	0.862 (1.807)	1.767 (3.358)
MYRSCHL	0.129 (1.883)	0.055 (1.035)	-0.063 (-0.696)	0.319 (2.054)	0.055 (1.007)	0.061 (0.841)	-0.105 (-0.726)	-0.126 (-1.049)
YRSCHL	-0.145 (-2.127)	-0.014 (-0.348)	0.056 (1.021)	-0.062 (-0.447)	-0.036 (-0.770)	0.002 (0.041)	0.087 (0.943)	0.081 (1.039)
γ	1.037 (6.208)	0.534 (4.649)	0.596 (4.502)	0.800 (4.123)	0.723 (4.834)	0.491 (3.484)	0.648 (3.677)	0.637 (3.222)
δ	-0.415 (-2.000)	0.031 (0.053)	-0.993 (-5.454)	0.177 (0.244)	-0.297 (-0.897)	0.173 (0.203)	-1.244 (-6.663)	-0.617 (-1.734
Probab. of Stayer	0.60	0.49	0.73	0.46	0.57	0.46	0.78	0.65
Log-L	-173.2	-348.7	-249.4	-76.9	-219.6	-226.8	-138.4	-108.6
				d) Second spell Other (SO)				
Intercept	0.985 (1.198)	1.405 (3.674)	3.316 (5.322)	-1.877 (-1.473)	0.420 (0.714)	1.467 (2.882)	2.481 (3.127)	5.480 (3.303)
MYRSCHL	-0.125 (-2.272)	-0.065 (-1.482)	-0.046 (-0.847)	-0.030 (-0.350)	-0.126 (-2.314)	-0.095 (-1.423)	-0.040 (-0.683)	0.025 (0.113)
YRSCHL	-0.016 (-0.207)	0.025 (0.636)	-0.154 (-2.506)	0.089 (0.580)	0.069 (1.223)	0.042 (0.746)	-0.082 (-1.067)	-0.386 (-1.947)
γ	1.115 (7.048)	0.704 (8.068)	0.454 (4.396)	2.229 (3.962)	1.231 (7.754)	0.689 (5.647)	0.383 (3.548)	0.942 (2.495)
δ	2.428 (4.291)	No Admit	No Admit	1.846 (2.971)	2.092 (4.721)	No Admit	No Admit	No Admit
Probab. of Stayer	0.08			0.14	0.11			
Log-L	-57.4	-118.3	-50.8	-13.25	-82.8	-67.4	-33.0	-15.4

^ Asymptotic normal statistics in parentheses.

heterogeneity and Table 3 the results for the log-logistic case also accounting for mover-stayer heterogeneity.

To prevent p from falling outside the (0,1) interval, a logistic specification was adopted for the mover probability, such that $p = \exp(\delta)/(1+\exp(\delta))$.

The results change in the expected manner when the heterogeneity parameter is admitted in the estimation: the duration parameters increase in size and in some cases the direction of duration dependence is changed towards the positive side. As it is apparent from Table 2, for the first transition out of Lima, the exponential model is rejected in favor of negative duration dependence for cohorts 45-65, 35-44 and 45-54 at the 5 percent level of significance. For the first spell out of "Other", the exponential model is rejected in favor of the Weibull with positive duration dependence for the youngest cohorts. It is rejected in favor of negative duration dependence for cohorts 45-65 and 45-54. The results about the magnitude of γ show a clear pattern by order of cohort for the first spell out of "Other" areas. The younger the cohort, the larger γ. Is this attributable to pure cohort effects or to the way the migration history is recorded? Given the retrospective aspects of the data, it is likely that the events of migration are recorded better for younger cohorts and by definition both the length of the interval for interval censored information and their incidence will be smaller.

The effect of the covariates on the conditional probability of a move out of both regions are unchanged by the introduction of control for unobserved heterogeneity. The heterogeneity parameter is admitted only for three of the cohort categories for the first spell in Lima and for all cohorts, except one for the first spell in "Other". Panels (c) and (d) of Table 2 refer to the second spell in Lima and "Other" areas respectively. For the second spell in Lima, negative duration dependence cannot be rejected for most older cohorts and the exponential model cannot be rejected for younger cohorts. For the second transition out of "Other", there is evidence of

TABLE 3

MIGRATION PROCESS WITH LOGLOGISTIC DURATION DEPENDENCE WITH MOVER-STAYER HETEROGENEITY CONTROL^

Cohort	15-29	30-44	45-65	15-24	25-34	35-44	45-54	55-65
			(a) First spell Lima (FL)					
Intercept	3.828 (6.291)	5.034 (11.269)	4.316 (5.193)	4.797 (4.748)	4.047 (7.887)	5.167 (8.866)	4.382 (3.621)	5.399 (3.136)
MYRSCHL	-0.078 (-1.685)	-0.069 (-1.729)	-0.044 (-0.547)	-0.165 (-2.414)	-0.085 (-2.107)	0.016 (0.276)	-0.037 (-0.431)	-0.476 (-1.718)
YRSCHL	0.104 (1.664)	-0.045 (-1.257)	-0.088 (-1.385)	0.104 (0.969)	0.067 (1.452)	-0.111 (-2.318)	-0.101 (-1.389)	0.202 (0.906)
g	1.003 (6.971)	0.990 (8.728)	0.753 (3.665)	1.346 (4.605)	1.055 (8.079)	0.826 (5.989)	0.623 (3.032)	1.753 (2.914)
d	No Admit	No Admit	-0.038 (-0.030)	No Admit	No Admit	No Admit	0.474 (0.190)	-1.051 (-2.624)
Probab. of Stayer			0.51				0.38	0.74
Log-L	-253.4	^T-366.6	-180.7	-114.1	-322.2	-179.7	-110.5	-67.5
			(b) First spell Other (FO)					
Intercept	4.726 (15.415)	4.877 (22.964)	4.706 (25.980)	4.832 (13.595)	5.009 (18.605)	4.818 (17.836)	4.649 (20.423)	5.025 (14.278)
MYRSCHL	-0.132 (-3.855)	0.022 (0.569)	-0.014 (-0.400)	-0.073 (-1.788)	-0.088 (-3.237)	0.035 (0.702)	-0.028 (-0.692)	0.029 (0.443)
YRSCHL	-0.087 (-2.599)	-0.189 (-6.127)	-0.168 (-6.159)	-0.092 (-2.198)	-0.107 (-4.043)	-0.205 (-5.689)	-0.143 (-4.753)	-0.231 (-4.459)
g	1.378 (10.772)	1.062 (13.027)	0.858 (13.571)	1.303 (10.682)	1.119 (10.813)	1.055 (11.428)	0.802 (10.789)	0.991 (8.701)
d	0.778 (1.353)	0.687 (1.927)	1.293 (2.158)	No Admit	2.508 (0.876)	0.424 (1.447)	2.731 (0.850)	0.471 (1.183)
Probab. of Stayer	0.31	0.33	0.22		0.08	0.40	0.06	0.38
Log-L	-853.4	-1516.8	-1486.8	-442.5	-957.1	-968.6	-941.9	-541.7

TABLE 3 (continued)

Cohort	15-29	30-44	45-65	15-24	25-34	35-44	45-54	55-65
				c) Second spell Lima (SL)				
Intercept	1.853 (2.134)	1.833 (2.798)	0.768 (1.167)	0.873 (0.656)	1.873 (2.743)	1.561 (1.916)	0.076 (0.081)	1.222 (1.284)
MYRSCHL	0.270 (2.626)	0.076 (1.089)	-0.076 (-0.581)	0.396 (2.148)	0.101 (1.232)	0.078 (0.807)	-0.104 (-0.462)	-0.150 (-0.918)
YRSCHL	-0.266 (-2.269)	-0.023 (-0.436)	0.105 (1.192)	-0.230 (1.257)	-0.094 (-1.175)	-0.002 (-0.025)	0.131 (0.978)	0.145 (1.160)
g	1.906 (5.097)	0.655 (3.769)	0.802 (3.664)	2.367 (3.309)	0.962 (4.083)	0.644 (2.792)	0.877 (2.903)	0.834 (2.669)
d	-0.451 (-2.348)	0.302 (0.420)	-0.866 (-3.784)	-0.609 (-2.201)	-0.173 (-0.475)	0.306 (0.359)	-1.151 (-5.187)	-0.479 (-1.099)
Probab. of Stayer	0.61	0.43	0.70	0.65	0.54	0.42	0.76	0.62
Log-L	-169.1	-348.3	-248.9	-74.2	-218.7	-226.4	-138.5	-108.3
				d) Second spell Other (SO)				
Intercept	-1.410 (-1.200)	0.468 (0.594)	3.150 (3.409)	-2.085 (-0.977)	-1.112 (-1.188)	1.029 (1.117)	1.643 (0.986)	6.261 (2.863)
MYRSCHL	-0.090 (-0.959)	-0.125 (-1.783)	-0.056 (-0.621)	0.159 (0.693)	-0.173 (-2.310)	-0.194 (-1.623)	-0.040 (-0.348)	0.163 (0.594)
YRSCHL	0.100 (0.900)	0.076 (1.165)	-0.203 (-2.178)	-0.093 (-0.318)	0.154 (1.757)	0.072 (0.839)	-0.115 (-0.984)	-0.602 (-2.032)
g	2.079 (5.427)	1.189 (4.953)	0.676 (4.537)	2.768 (3.482)	2.301 (6.305)	1.059 (5.459)	0.725 (2.077)	1.275 (2.599)
d	2.429 (4.030)	No Admit	No Admit	1.969 (2.676)	2.114 (4.555)	No Admit	2.654 (0.915)	No Admit
Probab. of Stayer	0.08			0.12	0.11		0.07	
Log-L	-54.3	-115.1	-50.75	-14.2	-78.4	-67.6	-32.3	-15.5

^Asymptotic normal statistics in parentheses.

negative duration dependence for cohorts 30-44 and 45-65 of the first group and for cohorts 35-44 and 45-54 of the second group. The evidence of positive duration dependence for cohort 15-24 cannot be rejected.

I turn now to Table 3, where the log-logistic model is presented with the mover-stayer unobserved heterogeneity correction. For the first spell in Lima, there is no clear pattern for the γ parameter. For most cohorts the null hypothesis that $\gamma = 1$ cannot be rejected at the 5 percent level except for cohort 45-54. For most of the cohorts for the first spell out of "Other" in panel (b) of Table 3 there is evidence that $\gamma > 1$, except for cohorts 45-65, 45-54 and 55-65. Only cohort 15-29 exhibits significant positive duration dependence. But γ is slightly greater than 1, meaning that for very small durations the hazard increases and then decreases monotonically.

The results for the second spell in Lima are in panel (c) of Table 3; there is evidence of first positive duration dependence and then negative duration dependence for the younger cohorts. The magnitude of γ is larger than that for the first spell in "Other". Again, γ turns to be less than 1 for older cohorts. The fact that for younger cohorts the evidence in favor of inver[illegible] shaped distribution cannot be rejected, points to the fact that duration intervals for younger cohorts tend to be more precisely estimated, given the disposition of the data. Moreover, relying upon the theory, for sufficiently long durations, negative duration dependence will emerge.

The Weibull and log-logistic are non-nested duration models. Therefore, there is no straightforward statistical method to judge which model best fits the data. The informational criterion proposed by Schwarz (1978) can be used as a first approximation to check the validity of the models. This method is not directly applicable to these distributions because they are not members of the exponential family. The Schwarz criteria is to choose the model for which

$$\text{Log-Lik} - 1/2\ m\log(n)$$

is the largest, where Log-lik is the log-Likelihood for the model, m is the number of parameters of the model and n is the number of observations. For the Weibull and log-logistic models, m and n coincide for each cohort, so the criteria is to choose the maximum value of the log-likelihood for each cohort and transition. The Weibull model performs slightly better for the first transition out of Lima, where for most of the cohorts the evidence of no duration dependence cannot be rejected. For the rest of the transitions the log-logistic model performs almost uniformly better than the Weibull model. In particular, this is true for those transitions that exhibit non-monotone hazard behavior under the log-logistic specification.

In light of this evidence, the discussion that follows will be based on the log-logistic model for each transition. Based on the log-logistic results, the hazard rates by cohort for the different transitions will be compared. This provides with tests of the sequential theory of migration based on the behavioral differences among transitions. In particular, the comparisons will try to take into account the weaknesses of the data and isolate the results that are robust to them. I will concentrate the analysis on the first cohort group composed of ages 15-29, 30-44 and 45-65.

In this study, the second transition out of Lima is of particular importance, given that it summarizes the conditional probability of remigration given duration of stay after the first transition took place. Figure 2 presents the plot of the hazard rate for the second transition out of Lima as a function of duration for all three cohorts. The hazard rates for each cohort were evaluated at the mean of the explanatory variables. For the first cohort, 15-29, the hazard rate is non-monotone. The peak occurs after one year of stay when it reaches 15 percent, decreases slightly during the second year and then decreases sharply. After five years of residence, the hazard rates declines to a level of less than 2 percent. This accords with the fact that return migrants come back early to their origin places or not at all as the learning model will predict. In contrast, for cohort 30-44 the average hazard rate is monotone decreasing and lower at durations less than five in comparison to the hazard rate for cohort 15-29. A similar pattern with

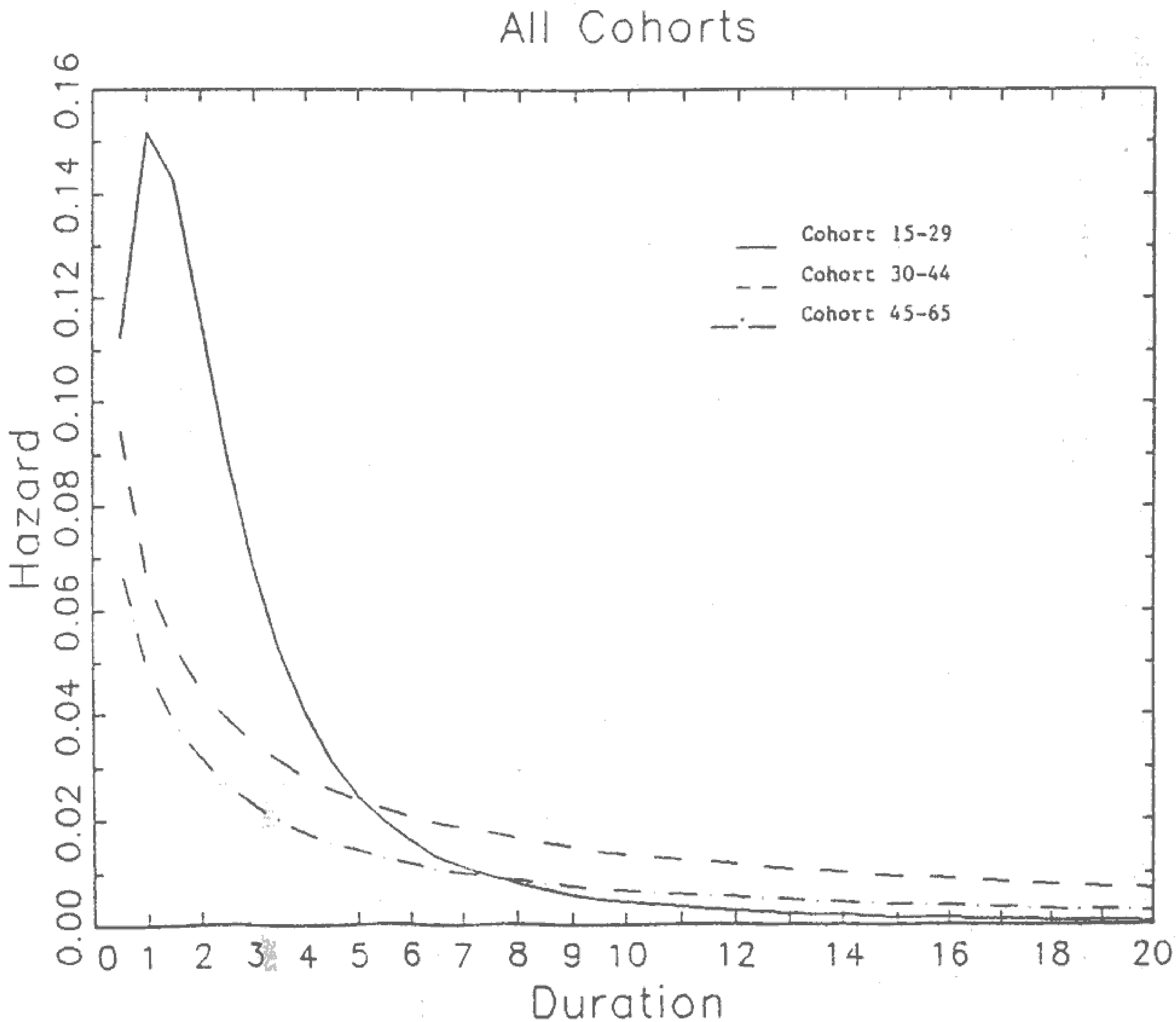

Figure 4 .--Log-logistic hazard functions for second transition from Lima

still lower hazard occurs for cohort 45-65. The peak hazard rate for cohort 30-44 is about 10 percent and for cohort 45-65, 7 percent. My initial presumption for these results was that for older cohorts who have performed on average more number of moves the probability that the second move is a typical "learning" move is low so that the duration of stay in the second location is really much lower than the upper bound recorded in the data. However, further sensitivity analysis indicated a second factor contributing to the monotone decreasing hazard result. When the model admits heterogeneity control and $\delta < 0$ (that is the probability of being a stayer is more than 50 percent), the hazard tends to be first increasing and then decreasing independent of the age of the cohort.[14] Uncontrolled heterogeneity tends to bias the duration parameters towards the negative side. The result that is robust to uncontrolled heterogeneity and data censoring is the level of the average hazard rate. For the first year of residence in Lima, it tends to fluctuate between 7 and 15 percent and it diminishes sharply (to less than 2 percent) after five years of residence. What is then the percentage of migrants that survive after migrating into Lima?. In Pessino (1991) it was demonstrated, assuming no prior expected wage differentials among locations, that the likelihood of remigration should be around 50 percent. That is, 50 percent of the primary migrants will end up returning to the origin place. If there exists a positive wage differential, the likelihood of return for each individual will be lower. Figure 3 presents the Survival function for the three cohorts 15-29, 30-44 and 45-65. After one year of stay in Lima, the survival fraction is similar for all cohorts, ranging from 85 percent to 95 percent. However, after 2 years, the differences in survival among cohorts increase, such that 75 percent of cohort 15-29 survives, 82 percent of cohort 30-44 and almost 90 percent of the oldest cohort survive. After 5 years of residence, 65 percent of cohort 15-29 survives and this rate stays constant for longer durations. For the second oldest cohort, survival is overall of 70 percent and for the third oldest cohort of 80 percent. So, on average 20 to 35 percent of the

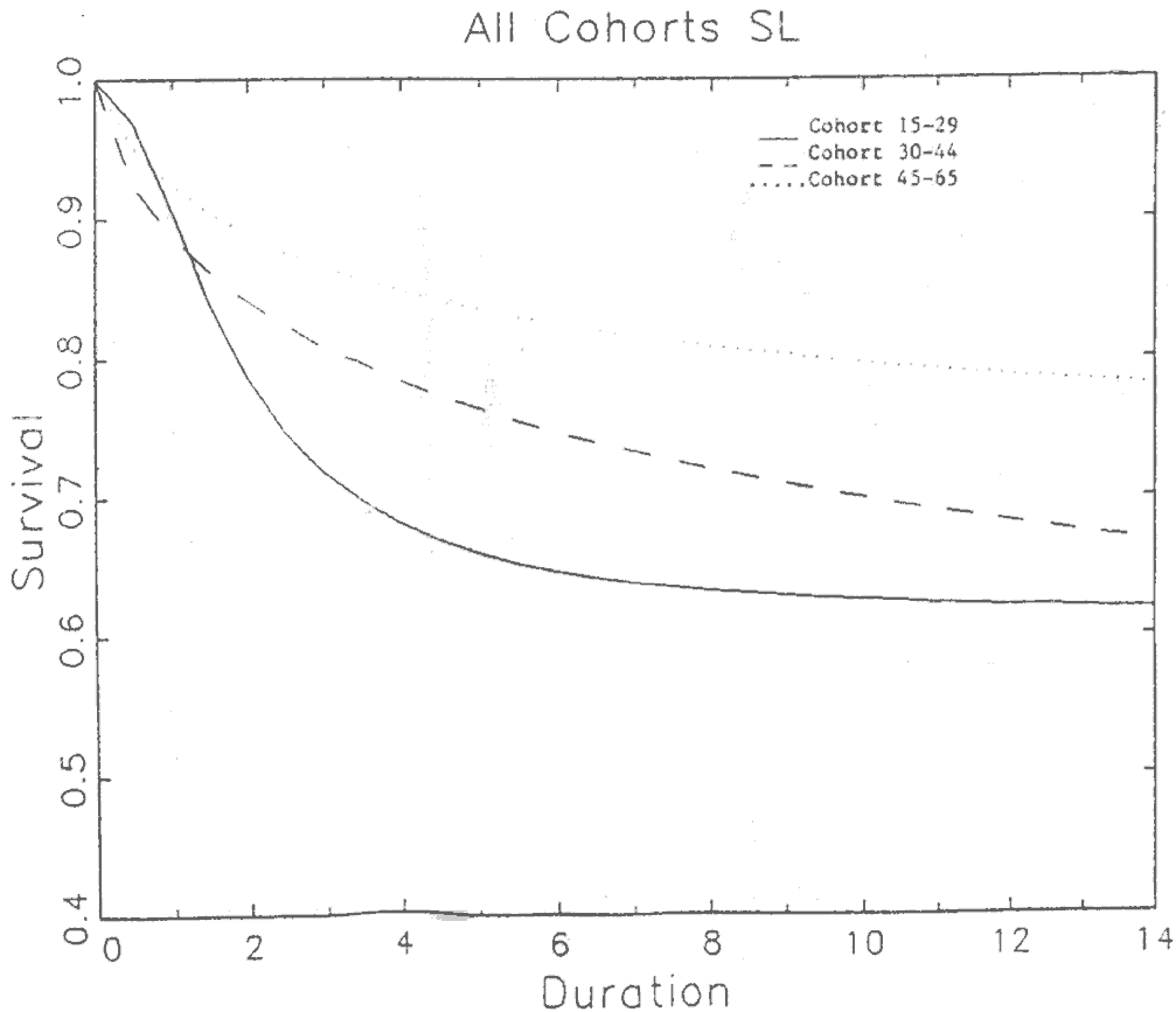

Figure 3.--Log-logistic Survivir functions for second transition from Lima

initial migrants into Lima leave and they tend to do so heavily in the first two years after their initial migration.

Another important implication of interest in the sequential migration model concerns if we can pool together the different transitions by location and/or by order of the transition. I will compare then the hazard for the second spell out of Lima with : a) the first transition out of Lima, b) the first transition out of "Other" and c) the second transition out of "Other". This comparison can improve our understanding of the observed positive correlation between in and out migration in given locations. In particular, is the positive correlation observed a fact of the correlation between first and second migration events (compare a) and c)) for the same individuals or it is a consequence of cross-section correlation (compare a) and b))?. Figure 4 plots together the hazard rate for the first and second transition out of Lima. The first transition out of Lima (FL) would correspond to the "cross-section" positive correlation; some people leave Lima because they have higher productivity in other areas. The second transition out of Lima (SL) identifies the "sequential" positive correlation; these individuals entered Lima in a first migration event and then they decided to remigrate if the match was not favorable. Although the first transition to Lima exhibits also a non-monotone hazard, it is difficult to see in the figure when compared to the hazard for the second transition out of Lima. The reason is that the hazard for the first spell in Lima is very low and after 1 year (after 12 years of age) it is fairly constant at 1 percent. If we add this to the fact that the original population in Lima is much lower than that in "Other" we see that the bulk of "Out of Lima movements" are made by previous migrants to the area. Pessino (1991) showed using Peruvian Census data, that in and out migration correlation is an important event in Peru in several Departments. This same event was recorded for other countries (see Greenwood (1975)), included the U.S.

Figure 5 plots together the hazard rate for the first spell out of "Other" (FO) and the second

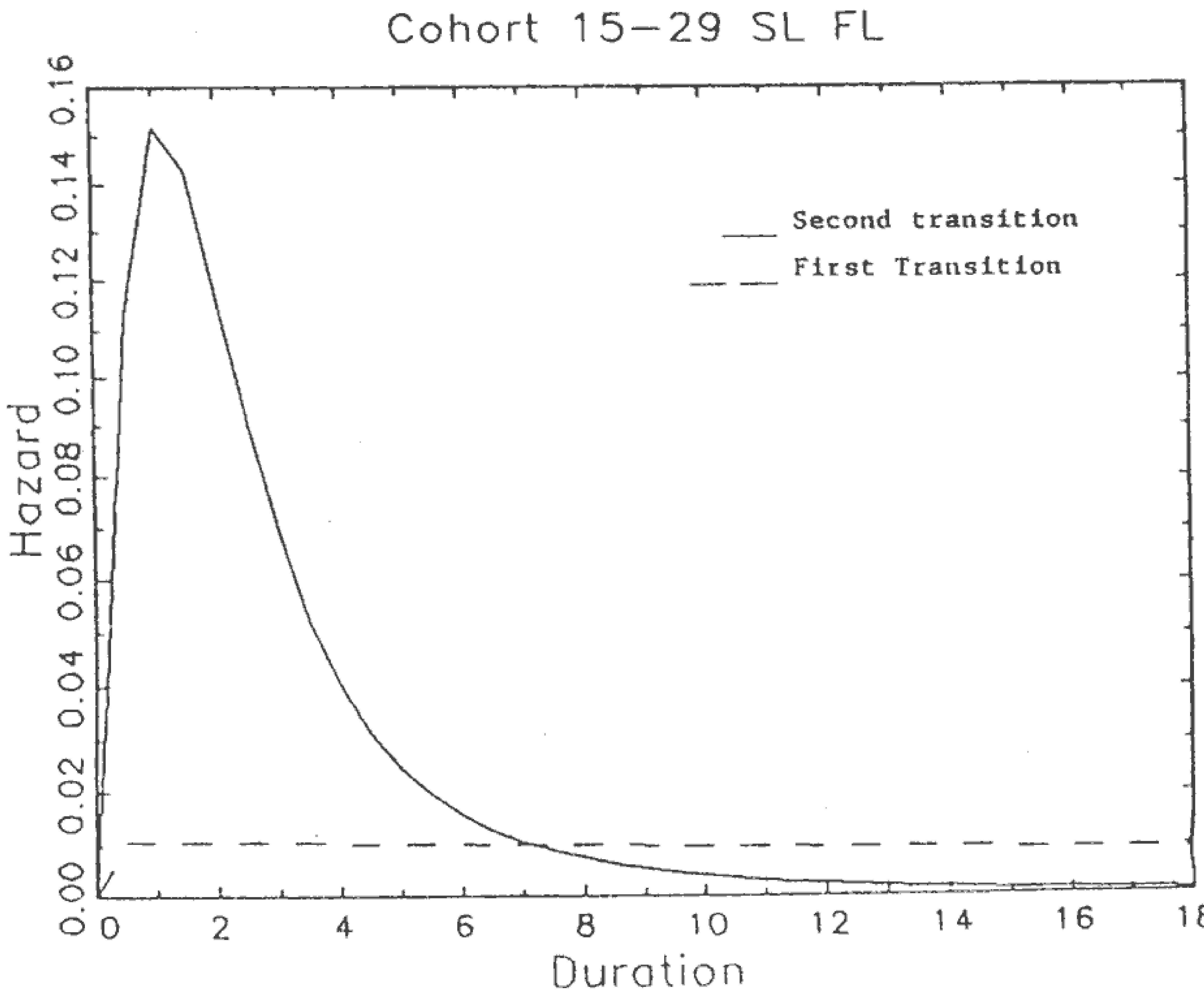

Figure 4 --Log-logistic hazard functions for first and second transition from Lima

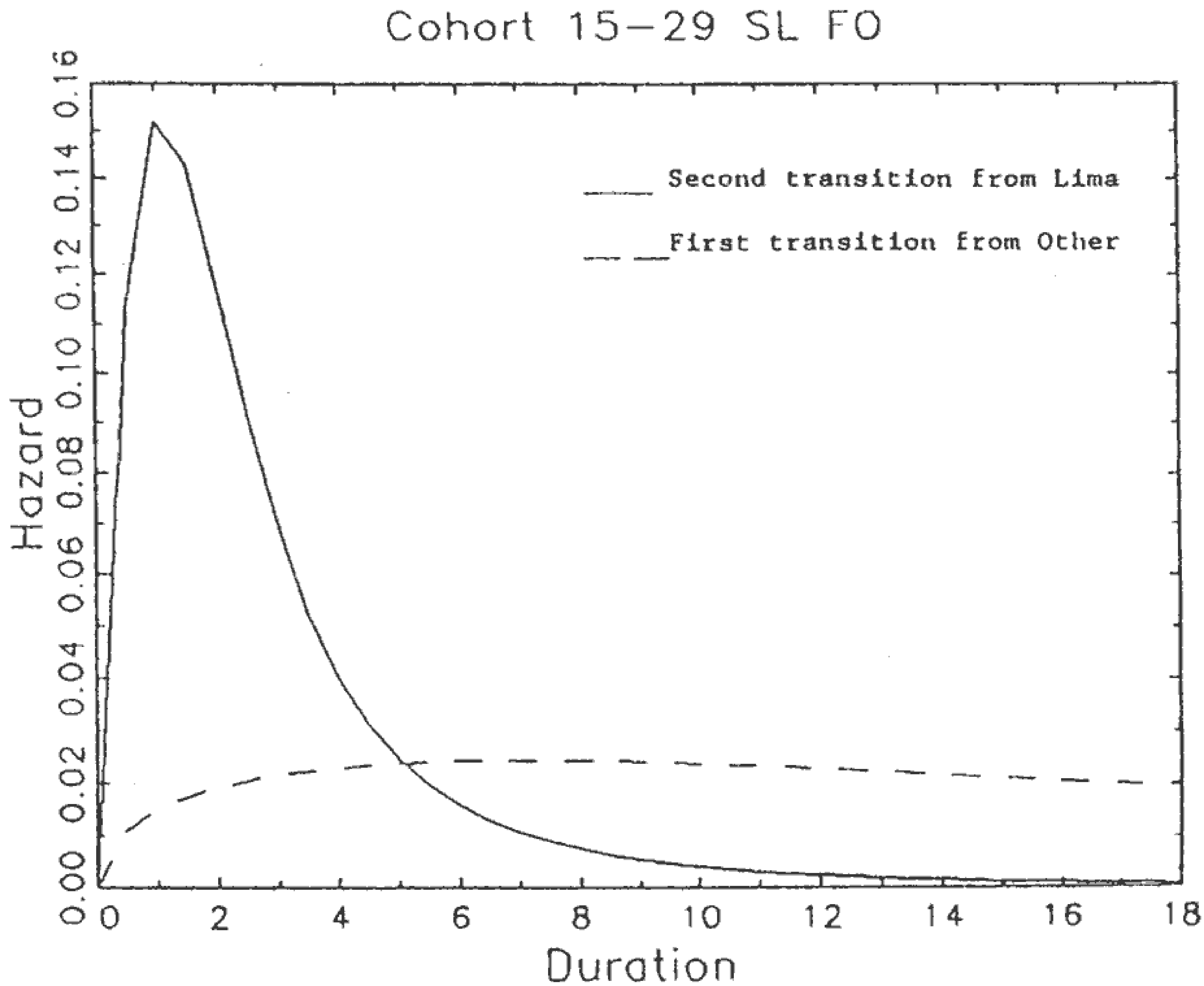

Figure 5.--Log-logistic hazard functions for second transition from Lima and first from Other.

spell out of Lima (SL) for cohort 15-29. This figure relates the first and the subsequent move made by those that migrated the first time to Lima. Pooling together migration events without taking into account the order would lead to misleading conclusions, not only on the effect of the covariates but on the duration and timing of the events.[15] For the first transition out of "Other", the hazard rate fluctuates around 2 percent a year and is fairly constant afterwards. It is expected that the pattern of first positive and then negative duration dependence applies especially to the hazard rate of remigration from Lima. Those that moved initially from "Other" to Lima are more at risk of moving again (to "Other") in comparison with those that are risk of migrating for the first time from Lima or from "Other". Those that are at risk of moving for the first time have better information on at least some of the unknown components of the location where they were born than those that are at risk of remigration.

The last comparison in figure 6 is for the second spell out of Lima (SL) and the second spell out of "Other" (SO) for Cohort 15-29. Although both distributions exhibit non-monotone hazard functions, the hazard rate of remigration is much higher for the second spell out of "Other" (that is, for those that have made a first move out of Lima). As mentioned before, these moves cannot be interpreted as typical learning moves. Most primary moves originated in Lima are apparently job-related moves and end up with a return.

To gauge the importance of education and the viability of the duration dependence result, predicted wage differentials were introduced as an additional covariate in the hazard functions for the different spells. Wages are only observed in the current place of residence and they are simultaneously determined with the migration history of each individual. Moreover, the theory of sequential choice implies that the current difference in posterior means or wages is the relevant measure to consider. For these reasons, I enter predicted and not actual wage differentials as covariates. To partially account for the revision in expectations, given that first time migrants

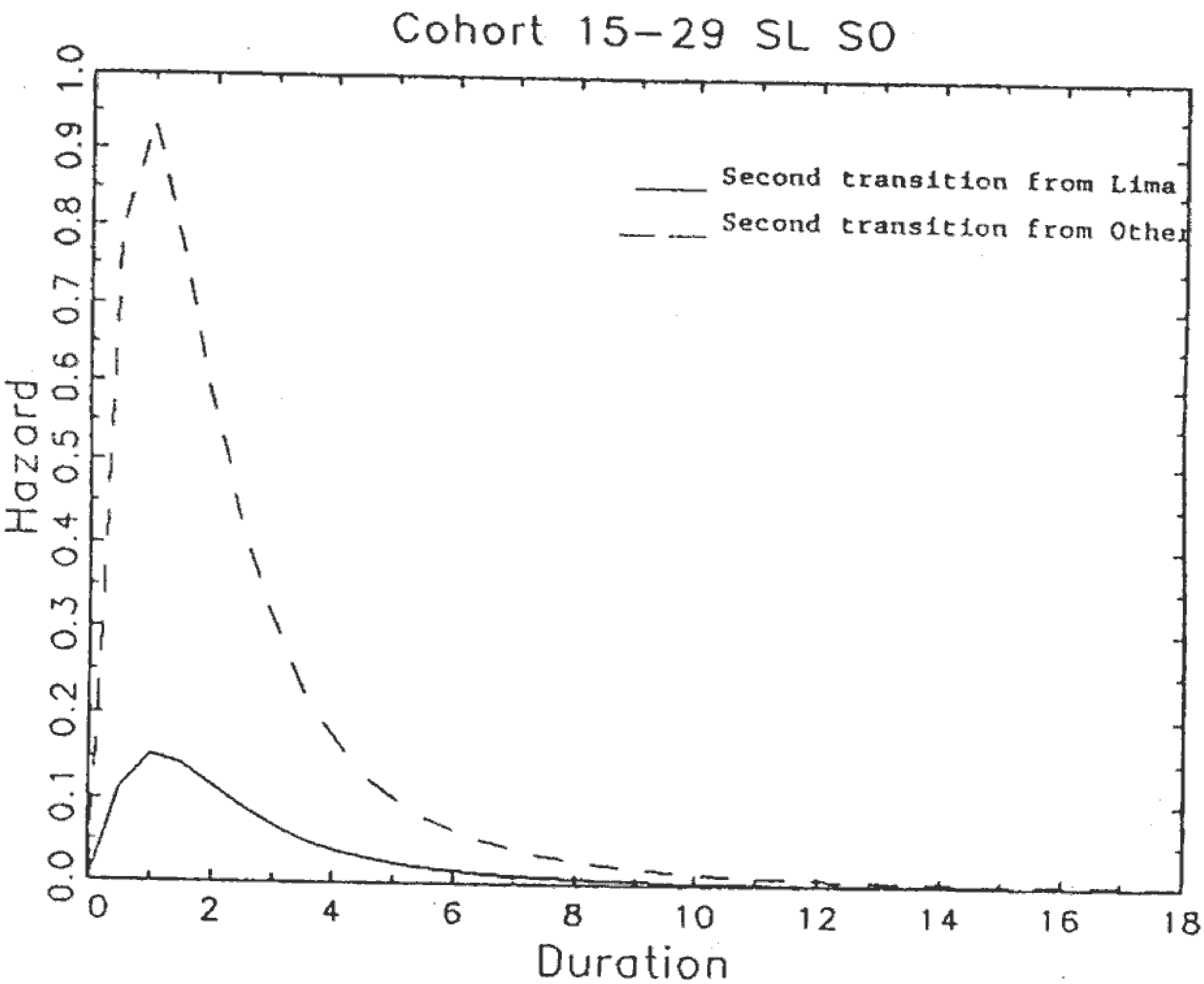

Figure 6.--Log-logistic hazard function for second transition from Lima and Other

will have ordinarily less information than secondary migrants I estimate separately the expected wages for them. That is, for people at risk of migrating for the first time from Lima and "Other" (FL and FO) I use all of the individuals that reside currently in Lima or "Other" to estimate wages separately in each region. For people at risk of returning (SL and SO) I estimate the wage functions using only observations on individuals that have moved at least once from their initial place of residence.[16] The list of variables used to estimate wages are age (AGE), age squared (AGESQ), years of schooling (YRSCHL), parents' years of schooling (MYRSCHL and FYRSCHL) and a dummy variable that is equal to one if the person resides in rural areas in the estimation of wages in "Other" (RURAL). Table 4 presents the estimation of the wage equations for each group. The first and second columns show the wage equations in Lima and "Other" using the whole sample, columns three and four using only individuals with a previous movement from Lima, and columns four and six using only those with a first movement from "Other". In general, the coefficients of the wage equations have the expected Mincerian signs and magnitudes; however, the estimates for the selected samples are less precisely estimated. These estimates were used to compute W_L - W_O, the predicted wage differentials for each of the categories (W_L refers to predicted wage in Lima and W_O refers to predicted wage in "Other"). For first spells I use age fifteen to predict wages and for the second spells I use age thirty for a of the individuals. In addition, for FL transitions I estimate W_O without introducing the dumm for rural-urban location since previous to moving I do not know if they will end up in an urba or rural destination. This dummy variable turns out to be the identifying restriction, since wage are dependent on the same covariates used in the hazard functions. Table 5 shows the log-logistic hazard functions introducing predicted wages. These estimations are showed for the first group of cohorts only. For the first spell in Lima in panel (a) I do not find any significant effect of predicted wages on the hazard. However, for the first spell in "Other" I find a positive and

TABLE 4

WAGE EQUATIONS IN LIMA AND OTHER AREAS FOR ALL INDIVIDUALS AND FOR PRIMARY MIGRANTS IN THEIR CURRENT PLACE OF RESIDENCE

	All individuals		Migrants Lima		Migrants Other	
	Lima	Other	Lima	Other	Lima	Other
Intercept	-1.033	-1.074	-0.662	-3.603	-0.897	0.028
	(-5.498)	(-5.216)	(-1.040)	(-2.131)	(-2.676)	(0.039)
AGE	0.077	0.076	0.044	0.194	0.078	0.011
	(7.407)	(7.165)	(1.329)	(2.059)	(4.491)	(0.290)
AGESQ	-0.0007	-0.0008	-0.0003	-0.0019	-0.0007	-0.00003
	(-5.395)	(-6.232)	(-0.795)	(-1.724)	(-3.351)	(-0.077)
YRSCHL	0.077	0.088	0.102	0.061	0.057	0.088
	(11.534)	(11.747)	(5.576)	(0.950)	(5.781)	(4.199)
MYRSCHL	0.023	0.044	-0.004	0.024	0.049	0.0685
	(2.896)	(3.580)	(-0.215)	0.388	(3.765)	(2.370)
FYRSCHL	0.018	0.018	0.039	0.421	0.015	0.0161
	(2.529)	(1.683)	(1.999)	(0.848)	(1.276)	(0.615)
RURAL	.	-0.707	.	1.032	.	-0.255
		(-13.283)		(2.236)		(-1.734)
ADJ R2	0.221	0.283	0.312	0.425	0.228	0.274
F-Stat	88.753	168.630	12.276	4.571	36.205	16.389
D. of F	5,1543	6,2563	5,135	6,23	5,595	6,238

Note._ t-statistics in parentheses

TABLE 5

MIGRATION PROCESS WITH LOGLOGISTIC DURATION DEPENDENCE WITH MOVER-STAYER HETEROGENEITY CONTROL AND PREDICTED WAGES^

	(a) First spell Lima (FL)			(b) First Spell Other (FO)		
Cohort	15-29	30-44	45-65	15-29	30-44	45-65
Intercept	3.097 (1.445)	5.297 (2.987)	4.794 (1.313)	5.813 (15.307)	5.213 (18.153)	4.903 (19.546)
MYRSCHL	-0.026 (-0.171)	-0.088 (-0.685)	-0.078 (-0.294)	-0.180 (-5.145)	0.003 (0.091)	-0.030 (-0.792)
YRSCHL	0.143 (1.122)	-0.059 (-0.599)	-0.117 (-0.525)	-0.134 (-3.838)	-0.207 (-6.340)	-0.174 (-6.220)
WL - WO	1.064 (0.356)	-0.379 (-0.155)	-0.703 (-0.133)	-1.333 (-4.527)	-0.466 (-1.809)	-0.277 (-1.115)
g	1.004 (6.983)	0.990 (8.726)	0.759 (3.568)	1.363 (11.517)	1.068 (13.357)	0.852 (13.584)
d	No Admit	No Admit	-0.076 (-0.062)	1.132 (1.717)	0.672 (2.005)	1.398 (2.113)
Probab. of Stayer			0.48	0.24	0.34	0.20
Log-L	-253.4	-366.5	-180.7	-842.5	-1515.1	-1486.2

	(c) Second spell Lima (SL)			(d) Second spell Other (SO)		
Cohort	15-29	30-44	45-65	15-29	30-44	45-65
Intercept	2.073 (1.466)	2.519 (4.179)	1.411 (1.506)	-2.518 (-1.682)	0.841 (1.198)	2.947 (3.079)
MYRSCHL	0.265 (2.497)	0.020 (0.258)	-0.157 (-0.990)	-0.067 (-0.699)	-0.172 (-2.273)	-0.056 (-0.617)
YRSCHL	-0.282 (-1.929)	-0.050 (-0.885)	0.082 (0.881)	0.168 (1.325)	0.081 (1.321)	-0.197 (-2.045)
WL - WO	-0.205 (-0.187)	-1.233 (-1.879)	-1.128 (-0.897)	1.122 (1.155)	-1.017 (-1.435)	1.214 (1.011)
g	1.926 (4.982)	0.723 (4.542)	0.858 (3.989)	2.184 (5.638)	1.208 (7.617)	0.686 (4.499)
d	-0.456 (-2.382)	0.092 (0.226)	-0.926 (-4.746)	2.329 (4.291)	No Admit	No Admit
Probab. of Stayer	0.61	0.48	0.72	0.09		
Log-L	-169.0	-346.0	-248.5	-53.7	-114.1	-50.2

^Asymptotic normal statistics in parentheses.

significant effect of W_L - W_O on the hazard. That is, as expected, positive wage differentials in Lima attract migrants from other areas. Now, the educational effect can be interpreted as net of the wage effect and isolating its effect on information. Years of schooling continues to be an important determinant of the transition from "Other" to "Lima".

For the second spell in Lima and the second spell in "Other" in panels (c) and (d) of Table 5, I do not find a significant effect of WL - WO on the probability of returning. As expected, the results on duration dependence and the now net effect of educational variables for the second spell in Lima continue to hold when controlling for expected wage differentials.

VIII. Conclusions

The main finding in this paper is that transitions in and out of Lima are behaviorally different and tend to accord with the predictions of the sequential migration theory. The finding that the hazard rate of remigration is first increasing and then decreasing for the main learning move (the second spell In Lima) is a contribution to the learning-matching literature. To my knowledge, this prediction has not yet been empirically implemented in the job matching literature and even less in the migration literature.

References

Berry, D. A., and Fristedt, B. *Bandit Problems: Sequential Allocation of Experiments.* London: Chapman and Hall, 1985.

David, Paul.; Mroz Thomas.; and Wachter, K. "Rational Strategies of Birth-spacing and Fertility Regulation in Rural France During the Ancien Regime." Stanford Project on the History of Fertility Control Working Paper No. 14, 1985.

Fahrenholtz, Steven K. "Normal Bayesian Two-Armed Bandits." Ph.D. Dissertation, Iowa State University, 1982.

Flinn, C. J., and Heckman, J. J. "Models for the Analysis of Labor Force Dynamics." *Advances in Econometrics* 1 (1982) : 35-95.

Ginsberg, Ralph B. "Timing and Duration Effects in Residence Histories and Other Longitudinal Data:II-Studies of Duration Effects in Norway, 1965-1971." *Regional Science and Urban Economics* 9 (1979) : 369-392.

Greenwood, Michael J. "Research on Internal Migration in the United States: A Survey." *Journal of Economic Literature* 13, no.2 (1975): 397-433.

Grootaert, C. and A-M. Arriagada, *The Peruvian Living Standards Survey: An Annotated Questionnaire*, Washington: The World Bank, 1986.

Harris, John, and Todaro, M. P. "Migration, Unemployment and Development: A Two-Sector Analysis." *American Economic Review* 60 (1970): 126-142.

Heckman, James J., and Borjas, G. J. "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence." *Economica* 47 (1980): 247-283.

Heckman, James J., and Singer, B. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data." *Econometrica* 52 (1984): 271-320.

Johnson, William R. "A Theory of Job Shopping." *Quarterly Journal of Economics* 92 (1978): 261-277.

Jovanovic, Boyan. "Firm-Specific Capital and Turnover." *Journal of Political Economy* 87 (1979): 972-990.

McCall, B. P., and McCall, J. J. "A Sequential Study of Migration and Job Search." *Journal* *of Labor Economics* 5, no.4 pt.1 (1987): 452-476.

Miller, Robert A. "Job Matching and Occupational Choice." *Journal of Political Economy* 92 (1984): 1086-1120.

Mortensen, Dale T. "Job Search and Labor Market Analysis." In *Handbook of Labor Economics*, edited by O. Ashenfelter and R. Layard. Amsterdam: North-Holland Publishing, 1986.

PERU, Censo Nacional de Poblacion y Vivienda, Lima: INE, 1981.

Pessino, Carola., "Sequential Migration Theory and Evidence from Peru." *Journal of Development Economics* 36 (1991): 55-87.

Schwarz, G. "Estimating the Dimension of a Model." *The Annals of Statistics* 6 (1978): 461-64.

Sjaastad, L. "The Costs and Returns of Human Migration." *Journal of Political Economy* Supplement 70, no. 5 (1962): 80-93.

Tunali, Insan. "Isolating the Use of a Decision on the Outcome: Some Formal Illustrations and Evidence from a Joint Migration Earnings Structure." Paper prepared for the 1985 Winter Meetings of the Econometric Society, 1985.

________. "Migration, Earnings, and Selectivity: Reconciling Theory and Fact." Working Paper No. 366, Cornell University, 1986.

Walker, James R. "An Empirical Investigation of the Timing and Spacing of Births in Sweden: The Effects of Changing Economic Conditions." Ph.D. dissertation, University of Chicago, 1986.

Footnotes

1. This article is a revised version of Chapters 5 and 6 of my Ph.D Dissertation. I wish to thank Gary Becker, James J. Heckman, and Thomas Mroz for their advice. I also wish to thanks workshop participants at The University of Chicago, The University of North Carolina at Chapel Hill and Yale University for their helpful comments. The World Bank Living Standard Survey Division kindly provided the data for this study.

2. For detailed proofs of some of the results in the text see Farenholtz (1982).

3. This statement is not true if the rate of discount β is different from 1. For simplicity, the rate of discount was not introduced in the dynamic optimization problem. The important results are not affected with its introduction, but proofs will get more complicated.

4. I call the exit time the remigration time for illustrative purposes and because it is the most likely situation in this model. That is, the implicit assumption is that the individual was born in location 2 and at the start of her/his working career chooses location 1. So, the decision to quit location 1 is termed here the remigration decision. If the individual was born in location 1 and chooses to work in location 1 at the outset, the decision to quit that location is termed the migration decision.

5. The reason I chose the two armed-bandit formulation instead of the formulation used by the above authors is that it makes more explicit the different learning opportunities in the two locations and stresses both the initial decision rule and the quit or remigration decision rule in this context. It is the natural way to extend the migration model between two locations with perfect information to imperfect information about the characteristics of those locations.

6. I leave for another opportunity the braking up of "Other" into its components.

7. Tunali (1985) discusses these issues and sets up an empirical test of the difference between these two models based on discrete choice rules with and without multiple selection. The test relies on the difference between the joint probability of choosing location 1 and 2 and the sequential probability of first choosing location 1 and then location 2.

8. In the preplanned model of migration, one can predict positive duration dependence if the fixed dates of departure are for example uniformly or normally distributed in the population: if everybody is expected to leave after reaching their target wealth, the probability of exiting increases with duration after some minimum period spent accumulating assets.

9. I used the same subsample as in Pessino (1991) paper: it consists of male wage and self-employed persons in the age group 15-65 who reported positive remuneration and positive hours worked in their main occupation during the week prior to the survey. This subsample consists of 4,195 men, that accounts for about 70% of the male labor force in the 15-65 age group.

10. In the estimation, I account for the possibility of interval censoring for all the individuals