Parallel Session 2B: Economic mobility and vulnerability
Monday, 25th August, Afternoon

Paper Prepared for the 30th General Conference of
The International Association for Research in Income and Wealth

Portoroz, Slovenia, August 24-30, 2008

Socioeconomic mobility as change in dependence across welfare attributes: an
application to assortative mating in Peru

Gaston Yalonetzky

For additional information please contact:

Gaston Yalonetzky
Queen Elizabeth House, University of Oxford
3 Mansfield Road
Oxford OX1 3TB
gaston.yalonetzky@qeh.ox.ac.uk

This paper is posted on the following website: http://www.iariw.org

Socioeconomic mobility as change in dependence across welfare attributes: an application to assortative mating in Peru

Gaston Yalonetzky[1]
University of Oxford
This draft: July 2008

Outline

I introduce the concept of socioeconomic mobility as the change in the degree of dependence between two variables whose correlation is meaningful, such as the education of partners. Such multivariate conception entails a departure from traditional univariate analysis of mobility. In order to measure how the joint dynamics of two variables affects the degree to which they correlate to one another, I propose an index suited for discrete variables based on one originally proposed by Bartholomew (1982), analog to the symmetric-movement indices of Fields and Ok (1996, 1999). I propose indices for both short-term and long-term co-dependence. A criterion to compare two distribution in terms of which one leads to higher concentration of co-dependent variables is proposed, which is useful, direct test with analytically derivable standard deviations. For an empirical application I compare mobility in educational assortative mating between male indigenous household heads and their non-indigenous counterparts in Peru. I ask four empirical questions: are the two mobility regimes statistically homogeneous?; which mobility regime exhibits more persistence in terms of reproducing better the initial joint distribution?; which mobility regime leads to a higher degree of educational homogamy?; and within each mobility regime, is it the case that the more homogamous the parents the more homogamous the sons? To answer the third question I also propose a multidimensional extension of one of Shorrocks indices. I find heterogeneity between the indigenous and non-indigenous transition matrices. The indigenous sample exhibits more persistence but the difference is not statistically significant. The non-indigenous mobility process leads to relatively higher long-term co-dependence in the education of the partners with statistical significance and finally I do not find evidence in any of the samples of higher homogamy among the sons of more homogamous parents.

---

[1] Contact information: gaston.yalonetzky@qeh.ox.ac.uk, cell phone: (0044)7854968059

1. Introduction

Several meanings of socioeconomic mobility have been discussed in the Economics and Sociology literature. Some meanings, like mobility as time-dependence, or as positional movements, focus on the degree of persistence of the variable across time and at different parts of the variable's distribution.[2] Indices measuring these meanings do not necessarily follow an axiomatic or welfare-based approach (e.g. as explained by van de Gaer et. al., 2001) although they can be used to measure, for instance, mobility as equality of opportunity in the sense of the degree of dependence of children's outcomes on parental background. Other meanings of mobility are more complex in that they require mobility to be associated with another outcome worthy of concern based on either axiomatic or welfare criteria. Examples of these meanings are mobility as progressivity, i.e. as an inter-generational equalizer of the expected value of outcomes (Benabou and Ok, 2001); or mobility as equalizer of welfare attributes/outcomes (Shorrocks, 1978b; Maasoumi Zandvakili, 1986) or equalizer of long-term outcomes (Fields, 2002; Fields et. al. 2007).

In this paper I introduce the concept of socioeconomic mobility as the change in the degree of dependence between two variables whose correlation is meaningful, such as the education of partners or their ethnicity. By contrast to the aforementioned meanings of economic mobility, this conception is multivariate. It entails a departure from traditional analysis of mobility in which just one variable (at a time) is followed across time. In order to measure how the joint dynamics of two variables affects the degree to which they correlate to one another, I propose an index suited for discrete variables based on one originally proposed by Bartholomew (1982) and resembling the symmetric-movement indices of Fields and Ok (1996, 1999). These precursor indices measure the distance, i.e. the difference, between a variable in its current state and in a past state. In the index I propose distance is measured between different but correlated variables. I propose indices for both short-term and long-term concentration, i.e. co-dependence, of welfare characteristics and also propose a criterion to compare two distribution in terms of which one leads to higher concentration of co-dependent variables, which is a test with analytically-derivable standard errors.

As an application I investigate the dynamics of educational assortative mating in Peru looking at mobility differences between matrices of indigenous male heads and non-indigenous ones. The evolution of assortative mating has been a topic of interest for long time both in Economics and Sociology (e.g. see Becker, 1993; Kalmijn, 1991a, 1991b, 1994, 1998; Schwartz and Mare, 2005). In terms of its relevance beyond the interest for its own sake, at least since Plato thinkers and scientists have been concerned with the impact of assortative mating on intergenerational transmission of welfare and in general with the distribution of welfare and living conditions in society.[3]

The empirical study of assortative mating has traditionally focused on cohorts and relied on the pooling of cross-sectional data sources and the use of log-linear models[4] (e.g.

---

[2] For a discussion of these meanings see Fields et. al. 2007.
[3] For recent empirical applications of the impact of assortative mating on income mobility see, for instance, Ermisch et. al., 2006, who find a significant effect of marital sorting on income mobility in Germany; or Blanden, 2005, who performs a similar assessment and finds similar qualitative results for the UK.
[4] For the rationale and implementation of log-linear models see Bishop et. al., 1975; Everitt, 1992; Agresti, 2002.

Schwartz and Mare, 2005). Only a handful of studies (e.g. Kalmijn, 1991a) have relied on retrospective data to assess the impact of parental background on assortative mating among the offspring. This paper assesses how parental assortative mating itself is related to children's assortative mating. By connecting parental to children's assortative mating I estimate a model of assortative mating dynamics and look at its consequences on the degree of long-term homogamy in a society.

More specifically I construct bi-dimensional transition matrices which link up the final joint distribution of education of male household heads and that of their respective spouses with the initial joint education distribution of the heads' fathers and respective mothers. Then I compare the respective matrices of indigenous Peruvian household heads with those of non-indigenous Peruvian heads seeking to answer four empirical questions: are the two mobility regimes statistically homogeneous?; which mobility regime (indigenous or non-indigenous) exhibits more persistence (i.e. is likely to reproduce better the initial joint distribution?); which mobility regime leads to a higher correlation[5] between the educational levels of the spouses, i.e. a higher degree of so-called homogamy in education; and within each mobility regime, is it the case that the more homogamous the parents the more homogamous the sons?

The first question is answered using multinomial tests as suggested by Anderson and Goodman (1957). I reject homogeneity of the bi-dimensional transition matrices. For the second question I propose an extension of the Shorrocks index of persistence (Shorrocks, 1978a, first section) and find that the transition matrix of indigenous heads exhibits more persistence but the difference is not statistically significant. For the third question, I apply the proposed indices of concentration-inducing mobility based on one of Bartholomew´s original persistence indices as well as the concentration comparison criterion and find with statistical significance that the mobility regime of indigenous heads leads to relatively lower correlation of educational outcomes between spouses. Finally, for the fourth question I test for the monotonicity of the transition matrices that connect the absolute values of the differences in educational levels between the head's father and the head's mother with the absolute values of the differences in educational levels corresponding to the head itself and his spouse. No evidence of monotonicity is found in either sample therefore a higher degree of parental homogamy is not necessarily conducive to relatively higher offspring homogamy in the Peruvian samples.

The rest of the paper is organized as follows. The next section lays out the conceptual framework in which I explain the estimation of first-order multidimensional Markov chains and the implementation of homogeneity tests on them. The section then explains the multidimensional Shorrocks index used to asses the relative persistence of the compared samples. Thereafter the main part of the section presents the indices of short-term and long-term concentration of variables along with the criterion used to test whether one sample leads to higher co-dependence of variables relative to another one. The next section discusses the nature, advantages and limitations of the dataset used for the empirical application and then there is a section of results with the quantitative answers for the empirical questions. Finally the paper has a concluding section with further discussion on the reach and limitations of this study.

---

[5] Throughout the paper I refer to correlation, co-dependence and concentration indistinguishably.

2. Conceptual framework and methodological results

a. Multidimensional Markov chains

Markov chains represent a time-series dataset as a vector of probabilities of being in a certain state conditional on past information of the observations (e.g. see Anderson and Goodman, 1957; Luenberger, 1978; Hamilton, 1994). With discrete variables the states are naturally defined by the values the variable can take. However states can be merged or split further depending on the variable's measurability. The narrower the width and greater the number of states the more informative the model is, assuming a large enough sample size. For a given sample size, a more refined model of conditional probabilities comes at the cost of efficiency in the estimation of the transition probabilities.

In a uni-dimensional first-order Markov chain model the distribution of one variable conditional on one prior state of it is estimated. By contrast, in a multidimensional first-order Markov chain the joint distribution of two variables conditional on an initial joint state is estimated. In both first-order cases the models assume that all path-dependence information is condensed in the last period, i.e. the probability of being in a given state in the present depends only on the state occupied in the immediately prior period. Let $M_{j|i}(t)$ be the transition matrix of a first-order Markov chain in period t. Its typical element is the probability of being in state j at period t conditional on having been in state i in the immediate past, i.e. period t-k,[6] denoted by $p_{j|i}(t)$. Assuming that the number of states is s, its respective log-likelihood function and maximum likelihood estimator of probabilities is:

$$L = \sum_{t=1}^{T}\sum_{j=1}^{s}\sum_{i=1}^{s} N_{ji}(t)\log p_{j|i}(t) \quad , \quad p_{j|i}(t) = \frac{N_{ji}(t)}{N_{.i}(t)}$$

(1)

Where $N_{ji}(t)$ is the absolute frequency of observations of the sample who find themselves in state j at time t and in state i at time t-1; $N_{.i}(t)$, is the number of observations in state i at time t-1; so $N_{.i}(t) = \sum_{j=1}^{s} N_{ji}(t)$. The total sample size, N, is

$$N = \sum_{t=1}^{T}\sum_{j=1}^{s}\sum_{i=1}^{s} N_{ji}(t)$$

Analogously the typical element of a first-order bi-dimensional Markov chain transition matrix in period t, $M_{ij|gh}(t)$, e.g. the probability of being in joint state i of variable y_1 and j of variable y_2 at period t conditional on having been in joint state g of variable y_1 and h

---

[6] For $1 \le k \le t-1$. The choice of time interval in discrete Markov models, usually constrained by the type and quantity of data available, is non-trivial for two reasons. First, because it may or may not coincide with the natural time or rate of potential transition of the observed units, which might even exhibit different transition propensies (and realized transitions) in the same time interval. Secondly, unless the choice of time interval and the natural rate of potential transition are identical (or assumed identical), then some discrete Markov models may not be compatible with the actual data. For further elaboration and examples see Singer and Spilerman (1976, p. 451-2).

of variable y_2 in the immediate past, i.e. period t-1 (or t-k) is $p_{ij|gh}(t)$. Assuming that the number of states for variables y_1 and y_2 is, respectively, s1 and s2, its respective log-likelihood function and maximum likelihood estimator is:

$$L = \sum_{t=1}^{T}\sum_{j=1}^{s_2}\sum_{i=1}^{s_1}\sum_{h=1}^{s_2}\sum_{g=1}^{s_1} N_{jihg}(t)\log p_{ij|gh}(t) \quad , \quad p_{ij|gh}(t) = \frac{N_{jihg}(t)}{N_{..hg}(t)}$$

Where $N_{jihg}(t)$ is the absolute frequency of units of the sample who find themselves in states i and j at time t and in states g and h at time t-1; $N_{..gh}(t)$, is the number of units/people in states g and h at time t-1; so $N_{..gh}(t) = \sum_{i=1}^{s_1}\sum_{j=1}^{s_2} N_{jihg}(t)$. Total sample size,

N, is $N = \sum_{t=1}^{T}\sum_{h=1}^{s_2}\sum_{g=1}^{s_1} N_{..hg}(t) = \sum_{t=1}^{T}\sum_{h=1}^{s_2}\sum_{g=1}^{s_1}\sum_{j=1}^{s_2}\sum_{i=1}^{s_1} N_{jihg}(t)$

As with a uni-dimensional Markov chain, in a multi-dimensional Markov chain an equilibrium joint distribution can be calculated from the respective transition matrix. Such equilibrium distribution is worthy of interest in empirical applications (e.g. see Quah, 1997; Fafchamps and Desmet, 2005; Hites, 2007) and so it is in this paper's analysis of long-term concentration of characteristics. Therefore it is worth mentioning how the equilibrium distribution, representing the long-term distribution, is calculated in the multidimensional framework.

In the first-order case the equilibrium distribution is calculated by re-presenting the multidimensional Markov chain as a uni-dimensional one in which each new state is a combination of states from the original multidimensional matrix. For instance, a matrix of n dimensions, each of which having s_i states, has $s_1 s_2 \ldots s_n$ states when it is re-written as a uni-dimensional matrix (e.g. see Hegre and Fjelde, 2007). Then the usual procedure for estimation of an equilibrium distribution for a first-order Markov process can be applied (e.g. see Fafchamps and Desmet, 2005). The regularity of the matrix (see Luenberger, 1978) is a sufficient but not necessary condition for having an equilibrium and ergodic distribution.[7]

Homogeneity tests on mutlti-dimensional Markov chains

A population homogeneity test is a standard multinomial test (e.g. see Hogg and Tanis, 1997). It can take the form of a likelihood ratio or a Pearson chi-square: both are asymptotically equivalent.[8] The test provides an answer to the question whether two or more population samples come from the same common multinomial distribution. Applied to Markov models these tests tell whether respective vectors of transition probabilities are different (or not) with statistical significance across different subsamples.

In the Pearson chi-square form the test for first-order stationary bi-dimensional Markov chains is the following:

---

[7] Sennot (1985) provides criteria to determine the non-ergodicity of a multidimensional Markov chain.
[8] Anderson and Goodman (1957, p. 106)

$$(1)\ \chi^2 = \sum_{w=1}^{W} \sum_{j=1}^{s_2} \sum_{i=1}^{s_1} \sum_{h=1}^{s_2} \sum_{g=1}^{s_1} \frac{N_{..hg}^{w} \left(p_{ij|gh}^{w} - p_{ij|gh}^{\mathrm{hom}} \right)^2}{p_{ij|gh}^{\mathrm{hom}}}$$

$$(2)\ p_{ij|gh}^{\mathrm{hom}} = \frac{\displaystyle\sum_{w=1}^{W} N_{jihg}^{w}}{\displaystyle\sum_{w=1}^{W} N_{..hg}^{w}} = \sum_{w=1}^{W} p_{ij|gh}^{w} \frac{N_{..hg}^{w}}{\displaystyle\sum_{w=1}^{W} N_{..hg}^{w}}$$

Where the index w denotes the compared population sub-groups. In the chi-square statistic (1) the transition probabilities of each sub-group are compared against the transition probability (2) which is a pooled estimate under the null hypothesis that the samples under comparison are homogeneous. Under such null the chi-square statistic, for first-order bi-dimensional chains, has a limiting chi-square distribution with (W-1) $s_1 s_2$ ($s_1 s_2$-1) degrees of freedom.

The treatment of the degrees of freedom in multinomial tests deserves carefulness because as the number of states and the order of the chain increases the sample size gets overstretched thinner, leading to the appearance of several (or just more) zeros in the transition hyper-surfaces (matrices, cubes, hyper-cubes, etc.). When all compared samples exhibit zeros in the same position or grid of the transition matrix then the corresponding element of the sums in the likelihood ratio (or the Pearson chi-square) statistic becomes indeterminate. The presence of zeros in the contingency table and transition matrix literature has been extensively treated.[9] In the case of homogeneity tests, when the problem of indeterminate elements arises there are two basic approaches.

One approach is to treat the zeros as inadequate sources of evidence reasoning that those grids/positions could have had probability mass had the sample size been large enough. The statistics literature calls these sampling zeros (e.g. see Bishop et. al., 1975; Everitt, 1992) and has proposed different techniques to deal with them, especially when the interest is in estimating log-linear models (Everitt, 1992, p.136). In the application of homogeneity tests Billingsley (1961) suggested subtracting from the degrees of freedom for every indeterminate element in the sum of the statistic. This latter option is appealing if one does not want to favor the null hypothesis of homogeneity by treating zeros common to the subsamples as a true homogenizing feature. On one extreme if all zeros common to all subsamples are treated as sampling zeros and Billingsley's suggestion is taken up then an upper bound of discounting of the degrees of freedom is established: for a given value of the statistic (be it the likelihood ratio or the Pearson chi-square), this approach yields the lowest possible p-value.

Another approach is to treat the zeros as contributors of evidence toward homogeneity, since the zeros are common to all subsamples under comparison. The reasoning is that no matter how large the sample size is those grids may never have any probability mass in them.[10] These zeros are referred to as structural zeros in the statistics literature

---

[9] For an introduction see Everitt, 1992.

[10] A classical example from the medicine literature is a contingency table which classifies causes of bleeding according to gender. The entry for menstruation and male would always be empty. So if two of these tables (say from different countries) are to be compared, the corresponding element for that entry will be indeterminate, see Everitt (1992, p. 106-107).

(Everitt, 1992). In such a case one may still want to discount degrees of freedom after structural zeros if the interest is to focus on homogeneity testing within the parts of the multinomial distributions that have probability mass. Alternatively, as done by Collins (1973), all common structural zeros can be considered in the test as sources of homogeneity in which case the statistics are computed without accounting for the indeterminate elements and the degrees of freedom are the same as usual. This latter approach yields, on the other extreme, a lower bound of discounting of degrees of freedom, i.e. the highest possible p-value for a given statistic. Any combination of the two extremes is also possible. I report results based on the two aforementioned extremes.

b. Multidimensional Shorrocks index

The index developed in the first section of Shorrocks (1978a) is based on a formula for the average time spent in a state of a transition matrix originally developed by Prais (1955). It is a first-order, transition matrix index which measures mobility as departures from the transition matrix's diagonal. In the uni-dimensional context it is defined by [11]:

$$D i \big| i = \frac{s - \sum_{i=1}^{s} p_{i|i}}{s-1}$$

Applied to continuous variables, the Shorrocks index, whether applied to a quantile matrix or not, is insensitive to mobility due to changes in the variable which are not large enough to generate a state transition. And for the same reason it experiences discontinuous variations when a change in the variable is large enough to generate a state transition.[12] Therefore it may understate mobility understood in terms of path-dependence[13], and this understatement will depend on the magnitude and number of states of the transition matrix.[14] Shorrocks (1978a) studied four desirable axiomatic properties that the Prais index may fulfill: normalization (N), monotonicity (M), perfect immobility (PI) and perfect mobility (PM). In the multidimensional index that I propose below I check whether they fulfill these desirable properties as well.

Following Shorrocks (1978a) (N) means that $0 \leq \Pr i \big| i \leq 1$; (M) means that if all the off-diagonal probabilities of a transition matrix are at least as great as those of another matrix

---

[11] The index is indirectly based on the sum of the eigenvalues of the matrix since the latter add up to the trace which features in the index's numerator.

[12] By contrast, continuous indices like those of Fields and Ok (1996, 1999) do capture better path-dependence and movement in the variable of interest. However because of their continuity, they do not differentiate between more and less meaningful transitions (e.g. crossing the poverty line). Indices like the one in the first section of Shorrocks (1978a) are better tailored to pick up such so-regarded meaningful transitions. This difference has been discussed by Fields and Ok (1999, 462-3.)

[13] That is, the extent to which current economic well-being is determined by the past (see Fields et. al., 2007, p. 107).

[14] Its standard deviation, as shown by Schluter (1998), Trede (1999) and Formby et. al. (2005) is:

$$\sigma\left(D i \big| i\right) = \frac{\sqrt{\sum_{i=1}^{s} \frac{p_{i|i}\left(1-p_{i|i}\right)}{N_{.i}}}}{s-1}$$

then the value of the mobility index for the former one should be accordingly higher; (PI) states that for a perfectly immobile transition matrix, i.e. an identity matrix, the value of the index should be zero and perfect (PM) establishes that the value for the index should be one for a transition matrix exhibiting perfect mobility, the latter understood as equality of the constituting conditional probability vectors, i.e. perfect unpredictability or path-independence.

The multidimensional index proposed in this paper inherits all these latter properties. But in a multidimensional setting the diagonal has $s_1 s_2 \ldots s_n$ elements as opposed to just s_i. The bi-dimensional Shorrocks index takes the following form:

$$(3) \quad Dij|ij = \frac{s_1 s_2 - \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} p_{ij|ij}}{s_1 s_2 - 1}$$

The index (3) takes the value of 0 when perfect persistence across time takes place therefore satisfying (PI) and partially (N), i.e. since it has a lower bound at zero. The index also satisfies (M) but in order to satisfy (PM) a restriction on the set of possible contingency tables, similar to the one imposed by Shorrocks (1978a), must be imposed. The restriction, which in the case of Shorrocks (1978a) implies focusing on matrices with quasi-maximal diagonals[15], is that the matrices under comparison should fulfill the following condition:

$$(4) \quad \mu_{ij} p_{ij|ij} \geq \mu_{kl} p_{kl|ij} \quad \forall kl \neq ij$$

Where $\mu_{11}, \ldots, \mu_{s_1 s_2}$ are positive.

Condition (4) ensures that:

---

[15] According to Shorrocks (1978, p. 117) a matrix with quasi-maximal diagonal is one in which:

$\mu_i p_{i|i} \geq \mu_j p_{j|i} \quad \forall i, j$. Where $\mu_1, \ldots, \mu_s$ are positive. This condition, in the context of Shorrock's analysis, ensures that so-called periodical matrices are excluded from consideration. Periodical matrices are such that the distribution of the variable of interest experiences shifts of probability mass from one tail to another and vice versa after every transition with corresponding reversals in the relative density of each tail. Such matrices do not have equilibrium distributions and are never observed in empirical applications of households economics since they would imply significant reversals of ranks (which could be too stark in the case of intra-generational mobility). Besides as Shorrocks points out, they imply a higher degree of predictability of past conditions than a matrix whose conditional probability vectors are identical, even though an index which fulfils (M) would rank the periodical matrix as exhibiting more mobility than the matrix with identical conditional probability vectors which conceptually should be regarded as exhibiting perfect mobility and be attributed the highest value of the index accordingly.

$$(5) \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} p_{ij|ij} \geq 1\ ^{16}$$

Therefore (N) is fulfilled by the index (3). (PM) is also satisfied when perfect mobility in terms of perfect time-independence is attained. Moreover when the index (3) is equal to one, (5) holds as an equality which therefore implies that:

$$(6)\ p_{kl|ij} = \frac{\mu_{kl}^{-1}}{\displaystyle\sum_{k=1}^{s_1} \sum_{l=1}^{s_2} \mu_{kl}^{-1}}$$

In other words the index (3) will be equal to one if and only if perfect mobility in the form of identical conditional probability vectors is present. Such condition was named by Shorrocks (1978a) strong perfect mobility (SPM) and is satisfied by the index (3) when condition (4) holds. Now its variance is:

$$\mathrm{var}\, Dii|ii = \frac{\displaystyle\sum_{i=1}^{s_1} \sum_{j=1}^{s_2} p_{ij|ij}\left(1 - p_{ij|ij}\right)\Big/ N_{..ij}}{\left(s_1 s_2 - 1\right)^2}$$

In some applications, as in this paper's, the appearance of zeros in the initial joint distribution is likely for both theoretical and sampling reasons. In order to ensure that the bi-dimensional Shorrocks index is bounded between zero and one the following adjusted index is proposed:

$$ADii|ii = \frac{s_1 s_2 - e - \displaystyle\sum_{i=1}^{s_1} \sum_{j=1}^{s_2} p_{ij|ij}}{s_1 s_2 - e - 1}$$

Where e is the number of empty cells in the initial joint distribution.[17] The adjusted bi-dimensional Shorrocks index will be used in the application to assortative mating in Peru to test which of the two population subsamples, indigenous or non-indigenous, exhibits higher persistence in terms of the proneness to replicate the initial distribution.

---

[16] (4) implies:

$$p_{ij|ij} \sum_{k=1}^{s_1} \sum_{l=1}^{s_2} \mu_{kl}^{-1} \geq \mu_{ij}^{-1} \sum_{k=1}^{s_1} \sum_{l=1}^{s_2} p_{kl|ij} = \mu_{ij}^{-1} \to p_{ij|ij} \geq \frac{\mu_{ij}^{-1}}{\displaystyle\sum_{k=1}^{s_1} \sum_{l=1}^{s_2} \mu_{kl}^{-1}} \to \sum_{i=1}^{s_1} p_{ij|ij} \geq \frac{\displaystyle\sum_{i=1}^{s_1} \sum_{j=1}^{s_2} \mu_{ij}^{-1}}{\displaystyle\sum_{k=1}^{s_1} \sum_{l=1}^{s_2} \mu_{kl}^{-1}} = 1$$

[17] Its variance is:

$$\mathrm{var}\, ADii|ii = \frac{\displaystyle\sum_{i=1}^{s_1} \sum_{j=1}^{s_2} p_{ij|ij}\left(1 - p_{ij|ij}\right)\Big/ N_{..ij}}{\left(s_1 s_2 - e - 1\right)^2}$$

c. Indicators of co-dependence of welfare attributes in bidimensional Markov chains

I introduce the concept of socioeconomic mobility as the change in the degree of dependence between two different variables whose correlation is meaningful, such as the education of partners or their ethnicity. This conception of mobility provides a criterion for comparing bi-variate distributions in terms of the relative degree of (de) concentration that they induce in the two variables.

In the following analysis I focus on square bi-variate joint distributions because the indices of short-term and long-term co-dependence hereby proposed apply to this type of distributions. By square bi-variate distributions I mean, narrowly, that the two co-moving variables have similar units and ranges.[18]

The indices are based on an univariate index developed by Bartholomew (1982) originally for discrete, ordered, categorical variables which takes the following form:

$$(7)\quad B_u \equiv \frac{\sum_{i=1}^{s} a_i \sum_{j=1}^{s} p_{j|i} |i-j|}{s-1}$$

Where s is the number of states and ai is a probability weight.[19] The index takes value of zero when there is perfect auto-co-dependence or auto-correlation and a value of one at the extreme of furthest distance or lack of auto-co-dependence possible. Different indices can be derived from (7) when the probability weights take different values.[20] An important one ensues when the probability weight is the initial distribution of the variable. In that case the Bartholomew index is a measure of autocorrelation on the joint distribution of the variable. Another important index is the one in which the probability weights are given by the equilibrium distribution. In that case the Bartholomew index in (7) is measuring long-term auto-correlation. A similar criterion for measuring auto-correlation, i.e. by looking at the absolute value of distances, was proposed by Fields and Ok (1996, 1999) for continuous variables in order to measure so-called symmetric movement.[21]

---

[18] Like in applications to assortative mating in education, occupation, ethnicity or religion.

[19] The original index has the following form: $B_G \equiv \sum_{i=1}^{s} a_i \sum_{j=1}^{s} p_{j|i} |m_i - m_r|$ where m_i and m_r are magnitudes corresponding to the states i and j (Field and Ok, 1996, p. 363). The empirical application of this paper is based on a proposed index based on (7), the latter being a specific version of B_G as rendered by Formby et. al. (2004).

[20] If the probability weights are independent from $p_{j|i}$ then (7) satisfies Shorrock's axioms of strong immobility (SI), and monotonicity (M) (see Shorrocks, 1978a, p. 1015).

[21] Their absolute change index in Fields and Ok (1999, p. 357-8) is:

$$FO_A \equiv \frac{\sum_{n=1}^{N} |y_n^{final} - y_n^{initial}|}{N}$$

Where N is the number of observations. Their relative change index is:

I hereby propose to measure co-dependence using the distance criterion as in Bartholomew (1982) and Fields and Ok (1996, 1999) but instead of looking at auto-correlations the focus is on how transitions affect both short-term and long-term co-dependence between two variables whose correlation is of socioeconomic interest and whose units and ranges are homogeneous.

The first proposed index measures co-dependence in the bi-dimensional transition matrix and takes the following form:

$$(8) \quad B_b \equiv \frac{\sum_{g=1}^{s}\sum_{h=1}^{s} a_{gh} \sum_{j=1}^{s}\sum_{i=1}^{s} p_{ij|gh} |i-j|}{s-1}$$

Where the a_{gh} are probability weights as in (7). Considering properties similar to those proposed by Shorrocks (1978) I now propose some desirable axiomatic properties for an index of co-dependence, $C\left(M_{ij|gh}(t)\right) = C(M)$, and assess whether the index (8) fulfills them:

(N) Normalization: $0 \leq C(M) \leq 1$

(MPD): Maximum positive co-dependence: whenever $p_{ij|gh} > 0$ if and only if $i = j, \quad \forall g, h$, then $C(M) = 0$

(MND): Maximum negative co-dependence: whenever $p_{ij|gh} > 0$ if and only if $|i-j|$ is maximized $\forall g, h$, then $C(M) = 1$

The axiom (MPD) establishes that a matrix whose conditional joint distributions have probability mass only where the attributes are perfectly positively correlated should be ranked by the index as exhibiting maximum positive co-dependence. Certainly several different matrices may exhibit this property. The importance of this axiom is that an index that fulfills it ranks all and only those matrices as having the exact same degree of positive co-dependence. Similarly if the index satisfies (MND) it ranks all and only those matrices whose conditional joint distributions have probability mass only where the attributes appear with maximum distance as having the exact same degree of negative co-dependence. The index (8) satisfies (N), (MPD) and (MND).

Depending on the probability weights different indices can be derived. I hereby propose three of them:

$$FO_R \equiv \frac{FO_A}{\dfrac{\sum_{n=1}^{N}\left(y_n^{initial}\right)}{N}}$$

- $a_{gh} = p_{gh}$: where p_{gh} is the initial joint distribution of y_1 and y_2. This index measures co-dependence of the two variables in the joint distribution of initial and final values of y_1 and y_2. It is also a measure of current co-dependence based on the initial distribution; hence it measures co-dependence in the short term. Because current co-dependence, as shown in (8), depends both on the transition matrix *and* on the initial distribution, comparisons of co-dependence may be problematic since the initial distribution may explain part or all of the prospective differences in co-dependence between two samples. For such reason this index is not suitable to compare co-dependence in the long term as departures from the initial distribution take place.

- $a_{gh} = 1/s^2$: This index measures co-dependence in the short term by setting a uniform initial distribution thus allowing for comparison of how conducive toward co-dependence is one transition matrix with respect to another one. In the long-term this index is not suitable as there is less justification to homogenise the initial distributions as transitions change current distributions toward long-term stationary values.[22]

- $a_{gh} = p_{gh}^*$: Where p_{gh}^* is the equilibrium probability of being in states g of y_1 and h of y_2. This index measures co-dependence in the long term and therefore is helpful for stating among compared samples which ones lead to higher co-dependence in the long term.

Even though the last index is not affected by changing joint distributions like the first two, it is inconvenienced by not having analytically derivable standard errors for inference. By contrast, the first two indices have readily derivable standard errors.[23]

---

[22] Provided that regularity conditions hold. For regularity conditions on transition matrices see Luenberguer (1978).

[23] For the standard deviation of the first index notice first that:

$$B_b \equiv \frac{\sum_{g=1}^{s}\sum_{h=1}^{s} a_{gh} \sum_{j=1}^{s}\sum_{i=1}^{s} p_{ij|gh} |i-j|}{s-1} = \frac{\sum_{g=1}^{s}\sum_{h=1}^{s}\sum_{j=1}^{s}\sum_{i=1}^{s} \theta_{ijgh} |i-j|}{s-1}$$

Where $\theta_{ijgh} = \Pr\left(y_1(t)=i, y_2(t)=j, y_1(t-1)=g, y_2(t-1)=h\right)$ is the probability of the joint distribution across variables and time. Therefore the standard deviation takes the following form:

$$\sigma_\theta = \frac{\sqrt{R'WR}}{s-1}$$

Where R is a column vector of s^4 dimension whose values are the ordered distance values, $|i-j|$, ranging from 0 to s-1, repeated four times. And W is a square $s^4 \times s^4$ matrix with the following elements:

$$W = \begin{cases} \dfrac{\theta_{ijgh}\left(1-\theta_{ijgh}\right)}{N} & if \quad i=a \wedge j=b \wedge g=c \wedge h=d \\ \dfrac{-\theta_{ijgh}\theta_{abcd}}{N} & otherwise \end{cases}$$

The standard deviation of the second index has a similar form:

Yet there is a sufficient condition with which one can state that a sample features higher co-dependence in both the short and the long term than another sample. This condition basically ensures that any index derived from (8) has a lower (higher) value for one of the samples. If the samples are A and B, and A induces more co-dependence the condition can be written as:

$$(9) \quad \sum_{j=1}^{s}\sum_{i=1}^{s} p_{ij|gh}^{A}\left|i-j\right| \leq \sum_{j=1}^{s}\sum_{i=1}^{s} p_{ij|gh}^{B}\left|i-j\right| \quad , \quad \forall g,h \in \left[1,s\right]$$

The advantage of this condition is that it can be tested with traditional multi-contrast procedures (e.g. see Stoline and Udry, 1979). The disadvantage is that it requires testing s^2 contrasts. As I hereby show there is a better sufficient condition that ensures long-term co-dependence, can be tested with multiple-contrast inference procedures and yields much less contrasts.

To attain this condition it is necessary first to re-write the bi-dimensional transition matrix as a uni-dimensional one. In this new matrix the states are defined by the distances $\left|i-j\right|$. Therefore, there are now s discrete states ranging from values of 0 to s-1. Such re-writing implies that certain distances are being equally regarded. For instance, in an application to assortative mating of education, the distance generated by the fact that a husband with 12 years of education is married to a wife with 8 years of education is given the same consideration as the distance generated by a husband with 2 years of education married to a wife with 6 years of education. For the purpose of comparing samples in terms of how conducive to homogamy they are this equal consideration of different cases is sensible. For other research questions differentiated treatment of such hypergamous and hypogamous examples is warranted.

In the new uni-dimensional transition matrices the probability of being in a given current distance state is conditioned on the distance state occupied in the past. A typical probability of the new transition matrix is:

$$\sigma_{B} = \frac{\sqrt{R'ZR}}{s-1}$$

Where now Z is a square $s^4 \times s^4$ matrix with the following elements:

$$Z = \begin{cases} \dfrac{p_{ij|gh}\left(1-p_{ij|gh}\right)}{N_{gh}} & if \quad i=a \wedge j=b \wedge g=c \wedge h=d \\[4mm] \dfrac{-p_{ij|gh}p_{ab|gh}}{N_{gh}} & if \quad \left(i\neq a \vee j\neq b\right)\wedge\left(g=c \wedge h=d\right) \\[4mm] 0 & if \quad g\neq c \vee h\neq d \end{cases}$$

$$p_{h|g} \equiv \Pr\left(s(t) = h \,\middle|\, s(t-1) = g\right)$$

Where s(t) is the distance state occupied in the present period and s(t-1) is the distance state occupied in the past period. The second stage consists of suggesting a new Bartholomew index for this new uni-dimensional transition matrix:

$$(10) \quad B_{bu} \equiv \frac{\sum_{g=0}^{s-1} a_g \sum_{h=0}^{s-1} p_{h|g} h}{s-1} = \frac{\sum_{g=0}^{s-1} a_g \overline{h}_g}{s-1}$$

Where $\overline{h}_g$ is the expected value of the distance variable, h, in the current period conditional on a past value g of the same variable measuring the original distance. The a_g are weighting probabilities.

Considering properties similar to those proposed for the index (8), (N), (MPD) and (MND) are also desirable for (10) but the latter two properties require the following re-statement:

(UMPD): Maximum positive co-dependence: whenever $p_{h|g} > 0$ if and only if $h = 0, \quad \forall g$, then $C(M) = 0$

(UMND): Maximum negative co-dependence: whenever $p_{h|g} > 0$ if and only if $h = s-1$ $\forall g$, then $C(M) = 1$

The axiom (UMPD) establishes that a matrix whose conditional probability vectors have probability mass only where the attributes are perfectly positively correlated should be ranked by the index as exhibiting maximum positive co-dependence. Similarly if the index satisfies (UMND) it ranks all and only those matrices whose conditional probability vectors have probability mass only where the attributes appear with maximum distance as having the exact same degree of maximum negative co-dependence. The index (10) satisfies (N), (UMPD) and (UMND).

In this context, the sufficient condition for a sample A to lead to higher short-term and long-term concentration than a sample B is the following:

$$(11) \quad \overline{h}_g^A \leq \overline{h}_g^B \quad , \quad \forall g \in [0, s-1]$$

The contrasts in (11) are now s, as opposed to s^2. This condition then ensures that for any sample-specific a_g:

$$(12) \quad B_{bu}^A \leq B_{bu}^B$$

Specifically when $a_g = p_g^*$, where p_g^* is the equilibrium distribution of the new uni-dimensional matrix whose states are distance states:

$$\text{(13) } B_{bu} \equiv \frac{\sum_{g=0}^{s-1} p_g^* \sum_{h=0}^{s-1} p_{h|g} h}{s-1} = \frac{\sum_{h=0}^{s-1} p_h^* h}{s-1}$$

Which is basically the proposed long-term index of concentration, since h is measuring the distance $|i-j|$. The index in (13) is a normalized, weighted average of the distance between the two variables where the weights are the long-term probabilities. Therefore, the index (13) is also stationary. By (12) and (13) sample A exhibits higher co-dependence than sample B.

Continuous case generalization

The previous analysis can be generalized from discrete ordered categorical variables to continuous variables, as long as the latter share the same units and range of values. Take co-dependent continuous variables y_1 and y_2, and let $\underline{y}$ and $\overline{y}$ be respectively the lower and upper bounds of the two variables' range. In the bi-dimensional setting the Bartholomew-based index for continuous variables would be:

$$\text{(14) } C_b \equiv \frac{\int_{\underline{y}}^{\overline{y}}\int_{\underline{y}}^{\overline{y}} a\left(y_1^a, y_2^a\right)\left(\int_{\underline{y}}^{\overline{y}}\int_{\underline{y}}^{\overline{y}} f\left(y_1, y_2 \big| y_1^a, y_2^a\right)|y_1 - y_2| dy_1 dy_2\right) dy_1^a dy_2^a}{\overline{y} - \underline{y}}$$

Where $\int_{\underline{y}}^{\overline{y}}\int_{\underline{y}}^{\overline{y}} a\left(y_1^a, y_2^a\right) dy_1^a dy_2^a = 1$ and $\int_{\underline{y}}^{\overline{y}}\int_{\underline{y}}^{\overline{y}} f\left(y_1, y_2 \big| y_1^a, y_2^a\right) dy_1 dy_2 = 1$. Therefore the analogue to sufficient condition (9) for a sample A to exhibit higher co-dependence than a sample B is:

(15)
$$\int_{\underline{y}}^{\overline{y}}\int_{\underline{y}}^{\overline{y}} f^A\left(y_1, y_2 \big| y_1^a, y_2^a\right)|y_1 - y_2| dy_1 dy_2 \le \int_{\underline{y}}^{\overline{y}}\int_{\underline{y}}^{\overline{y}} f^B\left(y_1, y_2 \big| y_1^a, y_2^a\right)|y_1 - y_2| dy_1 dy_2 \quad \forall y_1^a, y_2^a \in \left[\underline{y}, \overline{y}\right]$$

As with the discrete case, a better sufficient can be found if the bi-dimensional matrix is expressed as a new uni-dimensional matrix where the states are the differences between the two variables. In such case the Bartholomew-based index is:

$$\text{(16) } C_{bu} \equiv C_b \equiv \frac{\int_0^{\overline{y}-\underline{y}} a\left(h^a\right)\left(\int_0^{\overline{y}-\underline{y}} f\left(h \big| h^a\right) h \, dh\right) dh^a}{\overline{y} - \underline{y}}$$

Where $\int_0^{\overline{y}-\underline{y}} a\left(h^a\right) dh^a = 1$ and $\int_0^{\overline{y}-\underline{y}} f\left(h \big| h^a\right) dh = 1$. The sufficient condition analogous to condition (11), for a sample A to lead to higher co-dependence than a sample B is:

$$\text{(17) } \int_0^{\overline{y}-\underline{y}} f^A\left(h \big| h^a\right) h \, dh \le \int_0^{\overline{y}-\underline{y}} f^B\left(h \big| h^a\right) h \, dh \quad \forall h^a \in \left[0, \overline{y} - \underline{y}\right]$$

Finally notice that both (14) and (16) satisfy the translation invariance axiom of Fields and Ok (1996, p. 352) and the scale invariance axiom of Fields and Ok (1999, p. 457), which means that the indices' values do not change when affine transformations are applied to y_1 and y_2 $C_b\left(y_1, y_2, \overline{y}, \underline{y}\right) = C_b\left(ky_1 + l, ky_2 + l, k\overline{y} + l, k\underline{y} + l\right)$ and $C_{bu}\left(y_1, y_2, \overline{y}, \underline{y}\right) = C_{bu}\left(ky_1 + l, ky_2 + l, k\overline{y} + l, k\underline{y} + l\right)$.

d. Comparing bivariate distributions in terms of co-dependence of welfare attributes

The contrasts of condition (11) can be used to test, based on (12), the null hypothesis that two samples, A and B, lead to the same degree of long-term co-dependence against the alternative hypothesis that one of them, e.g. say A, leads to higher co-dependence both in the short and the long term:

$$(18) \quad \begin{array}{ll} Ho: & \overline{h}_g^A = \overline{h}_g^B \quad , \quad \forall g \\ Ha: & \overline{h}_g^A \leq \overline{h}_g^B \quad , \quad \forall g \end{array}$$

Because $\overline{h}_g$ is a weighted sum of random variables multinomially distributed, it is asymptotically distributed as normal. Its variance is:

$$(19) \quad \sigma_{\overline{h}_g}^2 = H'VH$$

Where H is a column vector of s dimension whose values are the ordered distance values, h, ranging from 0 to s-1. V is a square sxs matrix with the following elements:

$$(20) \quad V = \begin{cases} \dfrac{p_{i|g}\left(1 - p_{i|g}\right)}{N_g} & if \quad i = j \\ \dfrac{-p_{i|g}\,p_{j|g}}{N_g} & if \quad i \neq j \end{cases}$$

Where N_g is the number of observations with an initial distance equal to g. Employing the respective standard deviations derived from (19) and (20), the test of (18) are performed for the s contrasts of (11) with z-scores using the special critical values for the studentized maximum modulus distribution with infinite degrees of freedom developed by Stoline and Udry (1979).[24] The critical value depends on the number of contrasts which in this case depends on the original number of states of the variables whose distances are being considered for co-dependence analysis.

---

[24] For an earlier application in Economics of the tables of the studentized maximum modulus distribution to contrasts of means of Bernoulli-distributed variables see Anderson (1996).

e. Testing the monotonicity of a transition matrix

The last question of this paper is whether sons of relatively more homogamous parents tend to be more homogamous themselves. It turns out that, working with the uni-dimensionalized distance matrices, an answer to that question can be provided by testing a sufficient condition, i.e. whether the distance transition matrix is monotonic. In transition matrices whose states are ordered in terms of a magnitude, monotonicity means that the conditional probability vectors corresponding to higher state values stochastically dominate (in the first order) conditional probability vectors corresponding to relatively lower state values, so that the expected current value of the variable in question conditional on the higher initial state is higher than the expected current value conditional on a lower initial state, *regardless of the magnitudes of the states*. Such latter result was first introduced, along with others, into the Economics literature by Dardanoni (1995).[25] Formally, a uni-dimensional transition matrix, representing a first-order Markov process, is monotonic if and only if:

$$(21) \sum_{s=1}^{s} \sum_{i=1}^{s} p_{i|j} \geq \sum_{s=1}^{s} \sum_{i=1}^{s} p_{i|k} \quad \forall j > k, \quad k = 1, ..., s-1; \; j = 2, ..., s$$

In a transition matrix where states denote a distance between welfare attributes of spouses, from perfect homogamy to perfect heterogamy in a specific context, monotonicity provides a sufficient condition for the Markov process to lead to higher homogamy among the observations whose initial conditions reflect more homogamy.

In order to test for monotonicity, I perform tests of first-order stochastic dominance on pairs of conditional probability vectors from the transition matrix. I start with the first vectors with the lowest value for the initial distance state versus the adjacent second vector and then test the second vector versus the consecutive third and so on. The null hypothesis is that for every pair there is no stochastic dominance and it is tested against the alternative hypothesis that the probability vector conditional the lower initial state is stochastically dominated by the one with the relatively higher initial state. The actual first-order stochastic dominance tests are based on contrasts from (21) following Anderson (1996). If there are s states then s-1 contrasts are tested simultaneously.[26] The critical values of the t statistics are drawn from the maximum modulus distribution with infinite degrees of freedom developed by Stoline and Udry (1979). For the null hypothesis to be rejected it is necessary that equation (21) for a pair (j, k) of states holds as a statistically significant inequality at least for one of the s-1 contrasts. If the cumulative probabilities from (21) cross with statistical significance then there is indetermination: the evidence supports neither stochastic dominance nor homogeneity of the two cumulative distributions.

Finally, in order to state that a matrix is monotonic the test on (21) has to hold for every pair under comparison. If first-order stochastic dominance can not be proven for one of the pairs then there is no evidence that the more homogamous the parents are the more homogamous the children will be.

---

[25] The monotonicity property features prominently in Benabou and Ok (2001) who consider these matrices to perform their mobility analysis in terms of their relative degree of progressivity.

[26] Because $\sum_{i=1}^{s} p_{i|j} = \sum_{i=1}^{s} p_{i|k} = 1$

3. Data

Peru is an interesting case-study because, as other developing countries, it underwent a dramatic socioeconomic transformation during the 20th century to such an extent that it ought to exhibit, first, heterogeneity in the mobility regimes across cohorts, in several attributes like education, occupation, living standard or general social status indicators. Secondly, differential growth rates in educational attainments across gender and ethnicities, as happened in Peru during the aforementioned period of transformation, should have an effect in the patterns of assortative mating across different population groups, e.g. among indigenous and non-indigenous adults.

By 1940 (date of the earliest census of that century) the country's population was mostly rural (65%) and living in the highlands (63%). By 1993 (date of the latest census of that century) cities held the majority of the population and the coast had become the major region of population settlement (Contreras y Cueto, 2000). Simultaneously, starting incipiently in the 1920s and 1930s the construction of state school facilities boomed between the 1940s and the 1960s (Portocarrero et. al., 1988). These major changes in both demand-side and supply-side factors brought increasing levels of literacy and educational attainment in general (e.g. illiteracy rates fell from 59% in 1940 to 11% in 1993). More importantly they increasingly weakened the links between parental and offspring education. As I show in another paper (Yalonetzky, 2008) uni-variate educational mobility as measured by one of the Bartholomew indices[27], increased during the aforementioned periods, although not always monotonically, for men and women in urban and rural areas, and also for indigenous versus non-indigenous heads.

The dataset is the Peruvian 2001 Household National Survey (ENAHO) with information for 16,515 households. Education for household heads, spouses and respective offspring are available in an educational module. For the education of the parents of heads and spouses there is a special module on "Household perception" including retrospective questions on education, language and ethnicity characteristics of the parents and grandparents of heads and spouses. Parental education of the head and spouse is available in terms of the following levels (not in years): no education, incomplete primary, complete primary, incomplete secondary, complete secondary, incomplete technical tertiary, technical complete tertiary, incomplete university tertiary, complete university tertiary. [28] Matching categories were defined for constructing the respective variable in which the tertiary categories include technical and university education.

Ethnicity is a multidimensional phenomenon, as has been acknowledged elsewhere (e.g. see Valdivia, 2002), and so is ethnic identification in applied, quantitative research.

---

[27] The one deemed M2 by Formbi et. al. (2003). Let m be the number of states and θ_{ii} the joint probability of being in initial state i and in final state i, the index is defined as follows:

$$B = \frac{m\left(1 - \sum_{i=1}^{m} \theta_{ii}\right)}{m-1}$$ .Typically B=0 denotes perfect immobility (an identity matrix), and for a matrix characterized by time-independence B=1 although the reverse is not necessarily true (in fact it is also possible that B>1 without time-independence).

[28] For this reason I work with education in levels instead of years.

Considering this aspect, and in order to perform comparisons between indigenous and non-indigenous people I used two indicators of ethnicity available in the dataset: by mother tongue and by self-identification. This information was only available to household heads and spouses. Those who answered quechua, aymara or "other indigenous language" as their mother tongue, were categorized as indigenous according to the first definition. Similarly, those who answered "indigenous from the Amazon", "Quechua" and "Aymara" to the question "For your ancestors and according to your customs you regard yourself as" in the same module were deemed indigenous.[29]

The adults sample is comprised of household heads and spouses. Adult offspring cohabiting with parents are excluded because no attempt was made to impute ethnicity to them. A minimum age of 16 was defined for being an adult, because it was the age of the youngest household head. People who said to be studying were excluded from the sample to dispose of censored observations but it is still possible that some young people found in recess between two periods of studies made it into the dataset.

Even though few Living Standard Measurement Surveys in developing countries have as much retrospective information on parents of household heads and spouses as the Peruvian ENAHO 2001[30] I do not have information on the age at which parents of households heads and spouses gave birth to them. Such information is only available for adults found living in the same households as their parents. Therefore I can not control for parental cohort effects and/or the effects of life-cycle patterns of household resource allocation on the inter-generational transmission of education. This limitation is also present in other studies of Peru which have resorted to this dataset or others lacking the same information.

In this paper I study the changes in assortative mating of male household heads with respect to their parents and compare indigenous versus non-indigenous heads. As it turns out in table 3.1, the average educational level has increased from older to younger cohorts for the population subgroups involved in this paper's analysis, i.e. indigenous male heads, non-indigenous male heads and their respective spouses. The cohort averages tend to be the largest for non-indigenous males heads and the smallest for the spouses of indigenous males heads. Interestingly, the predominance in the averages of male indigenous heads over the spouses of non-indigenous heads depends on the cohort. Another interesting observation is that the educational attainment gaps between heads and their spouses do not seem to have changed significantly across cohorts. For indigenous, the gap is about 0.7 favouring male heads whereas for non-indigenous it is about 0.5 also favouring male heads. Such gaps suggest the possibility of heterogamy in both samples but potentially less so among non-indigenous. Whether relatively more homogamy is expected of the non-indigenous sample in the long term, assuming that current trends prevail is one of this paper's research questions which is answered in the next section.

---

[29] A discussion of the interaction and overlap of the two definitions is available upon request to the author.

[30] Some exceptions include the datasets used by Hertz et. al. (2007), the five African datasets used by Bossuroy et. al. (2007), the Brazilian Household Survey used by Bourguignon et. al. (2003) and Cogneau and Cigneaux (2005) and the Bangladeshi dataset of rural households used by Assadulah (2006). Also the special ENNIV module used by Benavides (2002), although the latter has a significantly smaller sample size.

Table 3.2 offers information on cross-sectional trends in the degree of perfect homogamy[31] across cohorts for indigenous and non-indigenous male household heads. Such information considers surviving marriages and hence might be affected by selective marital dissolution. The sample of non-indigenous heads exhibits a higher percentage of perfect homogamy than that of indigenous heads in all age cohorts but one. By contrast, the comparison of perfect homogamy between the parents of indigenous and non-indigenous heads reveals a more complex outlook. For the older cohorts it seems that fathers of indigenous exhibit relatively larger perfect homogamy whereas for the younger cohorts even the opposite is observed in some cases. In both samples the degree of perfect homogamy is remarkably higher for fathers than for the heads. Therefore the ratios of son-to-father homogamy are always below unity. With one cohort exception, the ratios are higher among the non-indigenous samples. This cross-sectional information reveals higher perfect homogamy among the youngest cohort of non-indigenous and a history of relatively higher perfect homogamy among non-indigenous for most cohorts. Whether this current trend continues toward the future considering the current patterns of intergenerational transmission of education and assortative mating can only be answered, under certain interpretative assumptions, with information that links individual heads with their respective parental background information.

I focus on young couples up to 35 years old primarily because of the patent cross-cohort breaks in the individual transmissions of education (which I have documented in Yalonetzky, 2008) and also because by focusing mostly on relatively newlyweds selective marital separation is less of a problem.[32] The data include 1,118 couples in which the male partner is indigenous and 1,696 in which the male partner is non-indigenous all belonging to the aforementioned age bracket (each with information on the education of the parents of both heads and spouses).

In the next section I report results to the four empirical questions considering different state specifications of the transition matrices. Such considerations are due to concerns about the size of the compared samples. First, I consider three states based on the educational levels in which state 1 stands for no education, state 2 represents incomplete primary and state 3 includes complete primary and any higher achievement. Arguably this specification compresses the educational-level space significantly but it can be used to study educational transitions with an emphasis on the region of educational poverty or destitution. Then I consider five states in which the first two states of the former specification are kept but now a state 3 stands for complete primary, a state 4 represents incomplete secondary and state 5 refers to complete secondary or more. In this specification complete secondary is compressed with tertiary education due to concerns about the relatively small sample size for tertiary achievement. Then six states are considered in which the first four are the same as in the just mentioned specification, then the fifth stands just for complete secondary and the sixth state includes any additional tertiary education irrespective of whether the tertiary degree was finished or not.[33] Finally, when working with the uni-variate distance state matrices I consider also a cap on the maximum possible distance because the latter usually show up, if at all, with few observations. I set a limit of four educational levels of distance as the maximum

---

[31] I measure the degree of perfect homogamy as the percentage of all couples in which both spouses have exactly the same educational level.

[32] Schwartz and Mare (2005), though, do report evidence about changes in homogamy in the United States using both data from newlyweds and from surviving marriages.

[33] I have information on whether a tertiary degree was finished but the size of the samples of those who did not finish is too small.

(even though with six educational levels the real maximum with these ordered, categorical variables would be five).

4. Empirical results

Homogeneity tests: do indigenous and non-indigenous mobility processes resemble each other?

The results for the homogeneity tests are in table 4.1. In both three specifications homogeneity between indigenous and non-indigenous samples can not be rejected when all the common zeros are regarded as source of evidence favoring homogeneity. However when all those zeros are discounted from the degrees of freedom homogeneity is rejected with 99% of confidence in the matrices with 5 and 6 specified states, but still homogeneity can not be rejected the three-state matrices. That is evidence of heterogeneity between indigenous and non-indigenous in the parts of the Markov chains where there is non-zero probability mass in the more detailed transition matrices. Interestingly the p-values are significantly lower in the state specifications where the sample sizes are stretched thinner. This result makes sense though because transition toward the top state of the 3-state specification matrix, measuring complete primary education, has been significantly high for both heads and their spouses in both samples.

Homogeneity tests were also performed on the uni-dimensionalized matrices whose states are the absolute educational distances between heads and their respective spouses. The results are in table 4.4 showing evidence of heterogeneity in the Markov processes that model the transmission of educational distance from fathers to sons between indigenous and non-indigenous samples, for the three state specifications.

Multidimensional Shorrocks index tests: which mobility process replicates better the initial conditions?

The test results on the multidimensional Shorrocks indices are in table 4.2. In all three state specifications the sample of non-indigenous exhibits higher mobility but the difference is never statistically significant. The indices' values for both samples are in fact very close to unity. I attribute this result to the significant increase in educational attainment across cohorts and from parents to children in Peru during the 20th century which renders the conditional probabilities of replicating initial relatively low joint educational levels very low or even empty.

Co-dependence tests: which mobility process leads to a higher degree of co-dependence in the short and the long term?

The Bartholomew-based indices for the bi-dimensional matrices are in table 4.3. They were calculated for the aforementioned weighting probabilities: initial distributions, uniform distributions, discounted uniform distributions and ergodic distributions. For all the three state specifications and for all weighting probabilities, the results indicate that the mobility process of the non-indigenous leads to higher co-dependence of the educational levels of heads and spouses both in the short term and in the long term. The differences are statistically significant with the exception of the differences between the indices calculated with the two uniform distributions as weights in the case of three educational levels.

In table 4.5 the Bartholomew-based indices have been calculated for the uni-dimensionalized matrices. The point estimates must be different because the distance

states of the uni-dimensionalized matrix are aggregations of states of the bi-dimensional matrix. The sample of non-indigenous exhibits a higher propensity toward relative homogamy vis-à-vis that of indigenous as measured by the Bartholomew-based indices, both in the short term and in the long term, although no confidence intervals are reported for the indices based on ergodic distributions. The differences are again statistically significant with the exception of the comparisons based on both uniform distributions in the specification with six educational levels.

The results on the sufficient condition for the co-dependence comparison are in table 4.6. The point estimates of the weighted averages (e.g. see condition (11)) are all lower for non-indigenous with a three state specification. With three contrasts the z-statistic for a one-tailed test with p-value of 0.01 is equal to -2.934 (Stoline and Ury, 1979, p. 88) which, following the criterion laid out by Anderson (1996), implies that in the sample of indigenous more assortative mating in education is expected both in the short and in the long term, with statistical significance since the z-score for the initial distance zero is higher in absolute value than the critical value.

When five states are specified the weighted average for the initial state with the greatest distance has a different sign to the others. However it is not statistically significant since with 98% of confidence the null hypothesis of equality of averages for that initial state can not be rejected, considering that for five contrasts the z-statistic is near -3.143 and 3.143 at every tail for one-tailed p-values of 0.01.[34] On the other hand, because two of the remaining negative z-scores have absolute values higher than -3.143 I reject homogeneity of the weighted averages in favor of the alternative hypothesis, again, of higher co-dependence in the sample of non-indigenous.

As with three states when six states are specified all weighted averages are lower for non-indigenous. However the z-scores do not warrant rejecting the null hypothesis of homogeneity in weighted averages at 98%. The greatest z-score in absolute value corresponds to the initial distance state of zero and is equal to -2.16 which falls within -2.378 and -2.091, i.e., the critical values for 90% and 80% of confidence. Therefore there is not sufficient evidence to reject the null hypothesis in this case but I suspect that, considering the point estimates, I would reject homogeneity with a larger sample size.

Monotonicity tests: is it the case that in any of the two samples under comparison the more homogamous the fathers the more homogamous the sons?

In tables 4.7 through 4.9 I report results for the monotonicity tests aimed at ascertaining whether more homogamous fathers are related to more homogamous sons. It turns out that for the three state specifications I can not conclude monotonicity of the respective matrices for both indigenous and non-indigenous. In the case of the three states, one of the comparisons does not provide evidence of first-order stochastic dominance for non-indigenous, whereas in the case of indigenous homogeneity of the conditional probability vectors can not be rejected. In the case of the five states specification, there is evidence of first-order stochastic dominance in the comparisons of the first two pairs of conditional probability vectors of the non-indigenous sample (counting from the lowest initial state) but the stochastic dominance relationship gets reversed thus leading to rejection of monotonicity. In the case of indigenous, homogeneity can not be rejected in the comparison between the first pair of conditional probability vectors. Then in the case

---

[34] The tables of Stoline and Udry (1979) do not have values for 5 and other numbers of contrasts.

of six states, the comparison of the first pair is reported in table 4.9. No stochastic dominance can be ascertained in either sample, therefore monotonicity is rejected too.

Evidence from estimations with limits to maximum distance

From tables 4.10 to 4.13 I report results on the same tests but instead of capping the maximum educational level I cap the maximum distance. I set the maximum distance between spouses' education at 4, thereby any couple whose educational gap is above 4 is attributed an educational gap of 4 nevertheless. The results from this new estimation coincide with the previous ones. Table 4.10 reports that homogeneity is rejected at 95% of confidence but not at 99%. In table 4.11 and 4.12 there is evidence that the sample of non-indigenous leads to higher short-term and long-term co-dependence of spouses' education. In table 4.13 the results for first-order stochastic dominance tests of the first pair of conditional probability vectors for both samples are reported. I only report the first pairs because already it is clear that there is no stochastic dominance relationship present in them and therefore the conclusion is that neither matrix is monotonic.

5. Conclusions

In this paper I sought to contribute both to the economic mobility literature and to the household economics literature, regarding assortative mating. Firstly I introduced the notion of socioeconomic mobility as the change in the dependence across welfare attributes. This notion suggests thinking about mobility in terms of the co-movement across time of variables whose correlation is meaningful for social science. From a measurement perspective it implies looking at multivariate trajectories as opposed to univariate ones which has been the case in the mobility literature. This is not to say that there is a lack of interest in looking at changes in the correlation of certain variables. That interest exists. But usually this analysis is performed by tracking cohorts along time, whereas in this paper I have proposed looking at multivariate mobility following the same units, e.g. households, dynasties, individuals, etc. along time.

For such purpose, on top of the already existing literature on log-linear models, I suggested the adaptation of old indices from the mobility literature, one of Shorrocks' indices and one of Bartholomew's, to a multivariate setting. With the extended Shorrocks index I sought to measure the degree to which the joint distribution of potentially correlated variables persists through time. With the adaptation of the Bartholomew indices I aimed at measuring and comparing the degree to which a mobility process leads to relatively more co-dependence of the variables both in the short term and in the long term. I focused that part of the analysis on bi-variate distribution and related it both to the work of Bartholomew (1982) and to that of Fields and Ok (1996, 1999). The use of longitudinal datasets and questionnaires with retrospective questions allows for the estimation of equilibrium distributions thus enhancing long-term analysis, which is not possible with just tracking cohorts from pooled cross-sections.

I also suggested a sufficient condition that enables the comparison of two bi-variate mobility processes in terms of the degree of co-dependence that they lead to and provides a multiple-contrast statistical test of relative co-dependence. With such test it is possible to state with sufficiency whether one distribution entails higher co-dependence than another one both in the short term and in the long term.

In order to generally test whether more homogamous parents have more homogamous children I proposed the use of monotonicity tests on uni-dimensionalized transition matrices of bi-variate distributions. This emphasis on monotonicity links this paper's work to that of Dardanoni (1993) and Benabou and Ok (2001).

I applied these techniques to analyze assortative mating in education across samples of indigenous males and non-indigenous males in Peru. I found that the transition matrix of non-indigenous males exhibits slightly less persistence as measured by the Shorrocks index although the difference is never statistically different. Regarding relative co-dependence I found evidence of relatively higher short-term and long-term co-dependence in the education of spouses among non-indigenous. As for monotonicity, the lack thereof in both samples suggests that among Peruvian couples sons of relatively more homogamous parents are not more likely to be relatively more homogamous than their peers of relatively less homogamous parents.

The latter empirical illustration brings about questions worth considering for further research. For instance, is the lack of monotonicity to be expected as a general rule or is it an otherwise unusual feature of the Peruvian case? Is there any particular reason why the non-indigenous samples tend to be more homogamous beyond a faster increase in the education of non-indigenous females vis-à-vis indigenous ones?

Methodologically, the multivariate analysis of mobility should be further expanded to account for more than two variables, for continuous variables and combinations of continuous and discrete variables, for higher-order Markov processes and multi-period setting, among other aspects. Several potential applications are interesting, ranging from assortative mating in education and other variables like income or ethnicity, to the more general mobility analysis of multidimensional welfare both at the micro-level (e.g. following households or individuals) and at the macro-level (e.g. following countries or regions).

Descriptive statistics

Table 3.1. Average educational levels by age cohorts

Age cohorts	Indigenous		Non-indigenous	
	Male head	Spouse of male head	Male head	Spouse of male head
16-29	3.914992	3.243144	4.301232	3.904698
30-34	4.012132	3.26779	4.460285	3.997785
35-39	3.798825	3.078275	4.495677	4.056663
40-44	3.852542	2.926276	4.444227	3.985792
45-49	3.423146	2.571721	4.286192	3.782554
50-54	3.214765	2.379487	4.163408	3.551613
55-59	2.89589	2.138264	3.7	3.325
60-onward	2.407984	1.703325	3.119136	2.652833

Table 3.2. Degree of perfect educational homogamy by age cohorts

Age cohorts	Indigenous			Non-indigenous		
	Male head	Father of male head	Ratio: son/father	Male head	Father of male head	Ratio: son/father
16-29	36.39	53.18	0.684279804	39.87	52.61	0.757840715
30-34	32.65	56.47	0.578183106	38.65	60.4	0.639900662
35-39	36.42	59.05	0.616765453	36.31	59.04	0.615006775
40-44	31.76	58.35	0.544301628	37.98	59.44	0.638963661
45-49	31	65.62	0.472416946	38.61	59.04	0.653963415
50-54	33.59	66.68	0.50374925	35.71	63.31	0.564049913
55-59	32.8	71.43	0.459190816	38.63	61.42	0.628948225
60-onward	37.09	74.95	0.494863242	42.52	65.83	0.645906122

Results

Table 4.1. Homogeneity tests for the multi-dimensional matrices

Educational levels	3	5	6
Chi-square statistic	59.83617	405.8265	454.9623
Degrees of freedom	72	600	1260
P-value	0.435237	1	1
Empty cells	20	344	872
Discounted degrees of freedom	52	256	388
P-value	0.212573	6.97E-09	0.01068

Table 4.2. Bi-dimensional Shorrocks indices

Educational levels	3 educational levels	5 educational levels	6 educational levels
Non-indigenous	0.974049	0.981123	0.981296
Indigenous	0.968979	0.967008	0.963658
z-score	0.383811	0.806978	1.296825

Table 4.3. Bartholomew indices for the multi-dimensional matrices

	Choice of a_g	Initial distribution	Uniform distribution	Uniform distribution discounted	Ergodic distribution
3 educational levels	Non-indigenous	0.11704	0.142899	0.142899	0.047323
	Indigenous	0.179235	0.161164	0.161164	0.07289
	z-score	-9.14861	-1.03116	-1.03116	
5 educational levels	Non-indigenous	0.174823	0.136372	0.14823	0.096004
	Indigenous	0.218433	0.184865	0.20094	0.140494
	z-score	-8.42829	-5.56707	-5.56707	
6 educational levels	Non-indigenous	0.170873	0.131817	0.1438	0.112478
	Indigenous	0.19212	0.143196	0.171835	0
	z-score	-6.19883	-2.55586	-5.14567	

Table 4.4 Homogeneity tests for the uni-dimensional matrices

Educational levels	3	5	6
Chi-square statistic	54.56075	76.84756	40.02574
Degrees of freedom	6	20	30
P-value	5.69E-10	1.34E-08	0.104367
Empty cells	0	3	9
Discounted degrees of freedom	6	17	21
P-value	5.69E-10	1.38E-09	0.007383

Table 4.5 Bartholomew indices for the uni-dimensional matrices

	Choice of a_g	Initial distribution	Uniform distribution	Uniform distribution discounted	Ergodic distribution
3 educational levels	Non-indigenous	0.11704	0.121592	0.121592	0.115079
	Indigenous	0.179338	0.181936	0.181936	0.179894
	z-score	-6.24603	-3.76346	-3.76346	
5 educational levels	Non-indigenous	0.174823	0.158477	0.158477	0.172737
	Indigenous	0.21847	0.210871	0.210871	0.21821
	z-score	-5.11501	-2.34875	-2.34875	
6 educational levels	Non-indigenous	0.170873	0.157979	0.157979	0.169689
	Indigenous	0.192129	0.194863	0.194863	0.192536
	z-score	-3.02775	-1.34766	-1.34766	

Table 4.6 Sufficient condition on the Bartholomew averages for uni-dimensional matrices. $Ho: \overline{h}_g^A = \overline{h}_g^B$,$Ha: \overline{h}_g^A \leq \overline{h}_g^B$ $\forall g$

Educational levels	Initial distance state	Non-indigenous	Indigenous	z-score
3 educational levels	0	0.219846	0.367316	-5.8031
	1	0.271071	0.338583	-1.89514
	2	0.238636	0.385714	-1.71667
5 educational levels	0	0.688605	0.871589	-4.12517
	1	0.783439	0.885333	-1.66655
	2	0.594937	0.82716	-1.71348
	3	0.333333	0.966667	-3.17058
	4	0.769231	0.666667	0.279076
6 educational levels	0	0.861648	0.961165	-2.16168
	1	0.875252	0.955026	-1.28338
	2	0.816993	0.986842	-1.1818
	3	0.595745	0.942857	-1.75489
	4	0.923077	1	-0.17784
	5	0.666667	1	-0.51355

Table 4.7. z- scores for the monotonicity tests with three educational levels.

$$Ho: \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|j} = \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|k} \quad ,Ha: \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|j} \geq \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|k} \quad j > k$$

	Non-indigenous sample		Indigenous sample	
Cumulative states	0 versus 1	1 versus 2	0 versus 1	1 versus2
0	2.077207	-0.96546	-0.48344	0.268807
0+1	0.45542	-1.19644	-1.20096	1.369933

Table 4.8. z- scores for the monotonicity tests with five educational levels.

$$Ho: \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|j} = \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|k} \quad ,Ha: \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|j} \geq \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|k} \quad j > k$$

	Non-indigenous sample		Indigenous sample	
Cumulative states	0 versus 1	1 versus 2	0 versus 1	1 versus2
0	3.144995	-2.41607	-0.42783	-1.86844
0+1	0.859707	-2.52126	0.600246	-0.24045
0+1+2	-0.94888	0.430897	0.721861	1.544551
0+1+2+3	0.089793	0.328829	0.363172	2.223517

Table 4.9. z- scores for the monotonicity tests with six educational levels.

$$Ho: \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|j} = \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|k} \quad ,Ha: \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|j} \geq \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|k} \quad j > k$$

	Non-indigenous sample	Indigenous sample
Cumulative states	0 versus 1	0 versus 1
0	0.506185	-0.94877
0+1	0.744307	0.305972
0+1+2	-0.59381	0.946634
0+1+2+3	-1.52882	0.094615
0+1+2+3+4	----	---

Table 4.10 Homogeneity tests for the uni-dimensional matrices. Maximum distance equal to 4.

Chi-square statistic	23.57705
Degrees of freedom	12
P-value	0.023208
Empty cells	0
Discounted degrees of freedom	12
P-value	0.023208

Table 4.11 Bartholomew indices for the uni-dimensional matrices. Maximum distance equal to 4.

Choice of a_g	Initial distribution	Uniform distribution	Uniform distribution discounted	Ergodic distribution
Non-indigenous	0.280857	0.121592	0.265516	0.279437
Indigenous	0.317829	0.31678	0.31678	0.317677
z-score	-3.25282	-2.70695	-2.70695	

Table 4.12 Sufficient condition on the Bartholomew averages for uni-dimensional matrices. Maximum distance equal to 4. $Ho: \ \bar{h}_g^A = \bar{h}_g^B \quad , Ha: \ \bar{h}_g^A \le \bar{h}_g^B \quad \forall g$

Initial distance state	Non-indigenous	Indigenous	z-score
0	0.846389	0.956311	-2.45958
1	0.869215	0.949735	-1.31863
2	0.803922	0.960526	-1.15306
3	0.666667	0.934783	-1.61764

Table 4.13. z- scores for the monotonicity tests with five educational levels. Maximum distance equal to 4.

$$Ho: \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|j} = \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|k} \quad , Ha: \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|j} \ge \sum_{s=1}^{s}\sum_{i=1}^{s} p_{i|k} \quad j > k$$

	Non-indigenous sample	Indigenous sample
Cumulative states	0 versus 1	0 versus 1
0	0.506185	-0.94877
0+1	0.744307	0.305972
0+1+2	-0.59381	0.946634

References

Agresti, A. (2002), *Categorical data analysis*, Hoboken, NJ: John Wiley and Sons.

Anderson, T.W., and Leo Goodman (1957), 'Statistical inference about Markov chains', *The annals of Mathematical Statistics*, 28(1): 89-110.

Anderson, Gordon (1996), "Nonparametric tests of stochastic dominance in income distributions", *Econometrica*, 64(5), 1996, p. 1183-1193.

Asadullah, Niaz (2006), "Intergenerational economic mobility in rural Bangladesh", mimeograph.

Bartholomew, D.J. (1982), *Stochastic models for social processes*, London: Wiley.

Becker, Gary (1993), *A treatise on the family*, Cambridge: Harvard University Press.

Benabou, Roland and Efe Ok (2001), "Mobility as progressivity: ranking income processes according to equality of opportunity", Working Paper 8431, Cambridge: NBER.

Benavides, Martin (2002), 'Cuando los extremos no se encuentran: un analisis de la movilidad social e igualdad de oportunidades en el Peru contemporaneo', *Bulletin de l'Institut francais d'Etudes Andines*, 31(3): 473-494.

Billingsley, Patrick (1961), "Statistical methods in Markov chains", *The Annals of Mathematical Statistics*, 32(1): 12-40.

Blanden, Jo (2005), "Intergenerational mobility and assortative mating in the UK", mimeograph.

Bossuroy Thomas, Denis Cogneau and Victor Hiller (2007), "Intergenerational mobility in five African countries", mimeograph.

Bourguignon, Francois, Francisco Ferreira and Marta Menendez (2003), "Inequality of outcomes and inequality of opportunities in Brazil", mimeograph.

Cogneau, Denis and Jeremie Cigneaux (2005), "Earnings inequalities and educational mobility in Brazil over two decades", Document de travail, DIAL.

Dardanoni, Valentino (1995), "Income distribution dynamics: monotone Markov chains make light work", *Social Choice and Welfare*, 12: 181-192.

Ermisch, John, Mario Francesconi and Thomas Siedler (2006), "Intergenerational economic mobility and marital sorting", *The Economic Journal*, 116: 659-679.

Fields, Gary (2002), "Does income mobility equalize longer-term incomes? New measures for an old concept", mimeograph.

Fields, Gary, Robert Duval-Hernandez, Samuel Freije and Maria Laura Sanchez-Puerta (2007), "Intergenerational income mobility in Latin America", *Economia*, 7(2): 101-154.

Formby, John, James Smith and Buhong Zheng (2004), "Mobility measurement, transition matrices and statistical inference", *Journal of Econometrics*, 120, p. 181-205.

Hegre, Havard and Hanna Fjelde (2007), "Democracy depraved. Corruption and institutional change 1985-2004", mimeograph.

Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith and Alina Verashchagina (2007), "The inheritance of educational inequality: international comparisons and fifty-year trends", *The B.E. Journal of Economic Analsysis & Policy*, 7(2): 1-46.

Hites, Gisele (2007), "On the robustness of the twin-peaked ergodic distribution of income across countries", ECARES, mimeograph.

Kalmijn, Matthijs (1991a): "Status Homogamy in the United States," *American Journal of Sociology*, 97:496–523.

Kalmijn, Matthijs (1991b): "Shifting Boundaries: Trends in Religious and Educational Homogamy," *American Sociological Review*, 56:786–800.

Kalmijn, Matthijs (1994): "Assortative mating by cultural and economic occupational status," *American Journal of Sociology*, 100(2):422–452.

Kalmijn, Matthijs (1998): "Intermarriage and Homogamy: Causes, Patterns, Trends," *Annual Review of Sociology*, 24:395-421

Maasoumi, Esfandiar and S. Zandvakili (1986), "A class of generalized measures of mobility with applications", *Economic letters*, 22: 97-102.

Parker, Simon, and Jonathan Rougier (2001), "Measuring social mobility as unpredictability", *Economica*, 68: 63-76..

Prais, S. J. (1955), "Measuring social mobility", *Journal of the Royal Statistical Society*, I(118): 56-66.

Sennott, Linn (1985), "Tests for the nonergodicity of multidimensional Markov chains", *Operations Research*, 33(1): 161-7.

Shorrocks, Anthony (1978a), "The measurement of mobility", *Econometrica*, 46(5): 1013-24.

Shorrocks, Anthony (1978b), "Income inequality and income mobility", *Journal of Economic Theory*, 19: 376-93.

Schwartz, Christine R. and Robert D. Mare (2005): "Trends in Educational Assortative Marriage from 1940 to 2003," *Demography*, 42(4):621-646

Singer, Burton and Seymour Spilerman (1976), "Some methodological issues in the analysis of longitudinal surveys", *Annals of Economic and Social Measurement*, 5/4: 447-74.

Sommers, P.M. and J. Conlisk (1979), "Eigenvalue immobility measures for Markov chains", *Journal of Mathematical Sociology*, 6: 253-276.

Van de Gaer, Dirk, Erik Schokkaert and Michel Martinez (2001), "Three meanings of intergenerational mobility", *Economica*, New series, 68(272): 519-537.

Yalonetzky, Gaston (2008), "The effect of cohort heterogeneity on comparisons of homogeneity and long-term educational prospects: The case of Peru", mimeograph.