Oxford Poverty & Human Development Initiative (OPHI)
Oxford Department of International Development
Queen Elizabeth House (QEH), University of Oxford

OPHI WORKING PAPER NO. 33

Comparing Economic Mobility with Heterogeneity Indices: an Application to Education in Peru

Gaston Yalonetzky*

December 2009

Abstract

The long literature on intergenerational transmission of well-being has largely been driven by concerns for inequality of opportunity and the persistence of low levels of wellbeing among certain social groups. A comparative strand of this literature seeks to compare indicators of these transmission mechanisms, i.e. mobility regimes, across societies, regions or time. In this paper I contribute to this literature by suggesting an additional way of comparing mobility regimes with indices of heterogeneity across distributions based on a traditional homogeneity test of multinomial distributions, which is helpful to compare discrete-time transition matrices. The indices measure the degree of dissimilarity between two or more transition matrices controlling for population size and the dimensions of the matrix. The indices provide a good alternative to between-group comparisons based on linear parametric models (chiefly OLS) in which either slope coefficients are compared directly or group dummy variables are interacted with parameters from the models. They also provide complementary information to comparisons based on summary indicators of transition matrices. An application to educational mobility in Peru shows that the transition matrices of males and females are more similar among the youngest cohorts of adults.

Keywords: Intergenerational mobility.

JEL classification: D30, J62.

* Oxford Poverty and Human Development Initiative (OPHI), Department of International Development, University of Oxford

This study has been prepared within the OPHI theme on Multidimensional Poverty Measurement.

OPHI gratefully acknowledges support for its research and activities from the Government of Canada through the International Development Research Centre (IDRC) and the Canadian International Development Agency (CIDA), the Australian Agency for International Development (AusAID), and the United Kingdom Department for International Development (DFID).

ISSN 2040-8188 ISBN 978-1-907194-17-7

Acknowledgements

I would like to thank Sabina Alkire, Casilda Lasso de la Vega, Maria Emma Santos, Maria Ana Lugo, Suman Seth, Cesar Calvo, Rodrigo Escudero, Miguel Angel Carpio and seminar participants at Universidad de Piura and OPHI for very helpful comments and fruitful discussions. I am also grateful to Lorenzo Oimas for providing me with the database.

Oxford Poverty & Human Development Initiative (OPHI)
Oxford Department of International Development
Queen Elizabeth House (QEH), University of Oxford
3 Mansfield Road, Oxford OX1 3TB, UK
Tel. +44 (0)1865 271915 Fax +44 (0)1865 281801
ophi@qeh.ox.ac.uk http://ophi.qeh.ox.ac.uk/

1 Introduction

The analysis of the intergenerational transmission of welfare has a long history. Economists and sociologists have looked at the intergenerational resemblance of income, wealth, education, occupation and multidimensional categories like social class.[1] From a theoretical perspective, economists have debated on the factors ostensibly behind the intergenerational transmission of welfare outcomes, ranging from inherited abilities and other family background endowments[2] to credit and insurance market imperfections affecting human capital investment differentially across populations.[3] Such concern is at the core of a classical theme in studies of intergenerational transmission of welfare: the degree of equality of opportunity, and/or the degree of (in)dependence of people's wellbeing outcomes from family background and other circumstances for which they should not be held accountable. In that sense it has been advocated as a criterion to judge the degree of justice in a society (e.g. Roemer, 1998). Higher equality of opportunity has traditionally been regarded as a desirable social trait; and even said to attenuate the negative welfare effects of economic inequality.[4] Intergenerational economic mobility has also been studied with an interest in looking at the intergenerational persistence of poverty[5], and regarding its effects on long-term inequality.[6] Van de Gaer, Schokkaert and Martinez (2001) also suggest considering intergenerational mobility in terms of equalization of life chances, i.e. not just the degree to which children's expected outcomes depend on parental background but the degree to which they are predictable.

More generally, studies comparing the degree of inequality and mo-

[1] See for instance the following studies. For income: Solon (1992, 2002); Chadwick and Solon (2002). For wealth: Charles and Hurst (2003); Asadullah (2006). For education: Hertz *et al.* (2007); Holmlund *et al.* (2006); Behrman *et al.* (2001); Gaviria (2007); Pasquier-Doumer (2003). For occupation: Cogneau and Gignoux (2005); Di Pietro and Urwin (2003); Hayes and Miller (1991). For social class: Erikson and Goldthorpe (2002); the several works of Wright quoted in Crompton (1998); Benavides (2002).

[2] For instance, see Becker, 1993; chapter 7 and its supplement. Also see Piketty (1999) for a good overview.

[3] For instance, see Galor and Zeira, 1993; Banerjee and Newman, 1993; Moav and Mazoz, 1999. For a review of the theoretical debate in Economics see Piketty, 1999.

[4] Kuznets (1966) suggested that among two societies with the same size distribution the one with higher mobility would be more desirable. Rosen (1992) argued that relatively high income mobility should reduce the concern for economic inequality.

[5] For a review of this literature see, for instance, Corcoran (1995).

[6] For instance, Benabou and Ok (2001) have proposed to rank societies in terms of their degree of progressivity which is a measure of convergence, as in the Economic Growth literature, but whose effect they also measure with intergenerational changes in Gini coefficients.

bility across societies and across different subgroups within societies abound and have a long history.[7] One motivation for such comparisons is the old concern in the social sciences for the welfare outcomes and opportunities among groups of societies facing discrimination or another source of socioeconomic or geographic disadvantage.[8] In the intergenerational mobility literature gender concerns appear for instance in Hayes and Miller (1991) who study occupational mobility in Ireland and in Pasquier-Doumer (2003) who studies educational mobility in Peru; while a focus on ethnicity is found, for instance, in the studies of the impact of educational mobility on earnings inequality in Brazil by Bourguignon *et al.* (2007), and by Cogneau and Cignoux (2005).

Similarly the methodologies found in the empirical literature are wide-ranging, including regression-based approaches, log-linear models for contingency tables, and applications of indices and statistical inference on Markov chain models. In this paper I contribute methodologically to the quantitative analysis of economic mobility by suggesting the use of a family of heterogeneity indices based on the statistic of a traditional test of homogeneity of multinomial distributions, which is useful to compare the degree of dissimilarity across transition matrices.[9] The heterogeneity indices are innovative, first, in that they enable the comparison of transition matrices element-by-element as opposed to comparisons of summary mobility indices derived from them.[10] Secondly,

[7] A famous empirical debate in the literature is about whether the U.S. exhibits higher economic mobility than several European countries even though it is characterized by higher inequality. For examples of empirical evidence in favour of and against this claim see Piketty (1999), Bjorklund and Janti (2000), and Benabou and Ok (2001). Among recent research in developing countries it is worth mentioning, for instance, work on educational mobility and earnings inequality in Brazil (Cogneau and Cignoux, 2005; Bourguignon *et al.*, 2007), on occupational and educational mobility in Peru (Pasquier-Doumer, 2003; Benavides, 2002), on educational mobility in Malaysia (Lillard and Willis, 1994), on educational mobility and inequality in Mexico (Binder and Woodruff, 2002),and on occupational mobility in Ghana, Uganda, Cote d'Ivoire, Guinea and Madagascar (Bossuroy *et al.* , 2007).

[8] In the Economics literature, an early example of this concern is the seminal work of Becker (1971) on the economics of discrimination. Nowadays the studies on gender and ethnicity discrimination are plentiful. For a broad interdisciplinary review see, for instance, Corcoran (1995) who documents comparative studies in the context of the "culture of poverty" debate in the United States.

[9] Let be the value that variable x takes at time t and assume the range of potential values that x takes can be partitioned into s non-overlapping intervals. A typical element of a transition matrix (e.g. in row i and column j) is the probability of observing x_t in the interval i conditional on having observed x_{t-k} (where k is an integer number) in interval j.

[10] For examples of statistical inference and empirical applications of comparisons of mobility indices based on transition matrices see, for instance, Trede (1999), Fields (2001), Formby *et al.* (2004).

they provide a non-parametric alternative for comparing the mobility regimes of different populations to comparisons based on linear parametric models of intergenerational mobility, usually OLS estimations, in which either slope coefficients are compared directly or group dummy variables are interacted with parameters from the models.[11]

The heterogeneity indices are meant to be most informative in, at least, two empirical situations. Firstly, to assess the actual degree of heterogeneity of mobility regimes when summary indices of transition matrices or slope coefficients from regression models are not different with statistical significance among a group of samples. In this context the heterogeneity indices are helpful by answering the following question: "according to the summary indicator (or the slope coefficient) the two (or more) mobility regimes seem to exhibit the same degree of persistence (or any other concept of mobility that the indicator is capturing), but actually how homogeneous are these mobility regimes?" Secondly, the heterogeneity indices help to pinpoint the sources of heterogeneity in a mobility regime when the latter is represented by a transtition matrix. A particular subfamily of these indices has a property of additive decomposability which enables the estimation of the relative contribution of every column (or row) heterogeneity to total matrix heterogeneity. The importance of this feature is apparent in the empirical application of this paper in which the contributions to total heterogeneity of transition matrices of educational attainment are estimated for every conditional distribution of educational attainment, i.e. for every column.[12]

I offer an empirical application of the heterogeneity index for transition matrices to intergenerational mobility of education in Peru, a country with the highest coefficient of correlation between parental and offspring's education among a sample of 42 countries (Hertz *et al.* 2007). I compare the transition matrices linking fathers' education to adult sons and daughters separately and for several cohorts of offspring. With the indices I document an interesting decrease in the degree of heterogeneity between the mobility matrices of men and women, which points to a homogenization of the total effect of the fathers' education on their offspring across gender.

In the next section the heterogeneity indices are introduced. A subsection compares it with other comparison methods, chiefly tests on slope coefficients of linear regression models and comparisons based on summary indicators from transition matrices. The following section is

[11] For examples of regression-based analyses of intergenerational mobility see, for instance, Solon (1992, 2002) and Chadwick and Solon (2002), both focusing on income.

[12] Or row, depending on how the transition matrix is constructed.

an empirical application to educational mobility in peru, after which the paper concludes.

2 The heterogeneity indices

To define the indices formally, let's start by assuming that a population can be partitioned into a vector of groups of individuals/observations, denoted by $G := \{1, ..., C\}$. The absolute frequency of people belonging to group g, i.e.$g \in G$, is N^g. Individuals from every group can attain a certain value of an outcome (e.g. income or education). All possible values of the outcome are in the vector $R := \{1, ..., A\}$.[13] The probability of attaining a given value of the outcome (e.g. $.k \in O$) conditional on being in group g is: p^g_k. The corresponding absolute frequency of people being of group g and attaining an outcome value k is N^k_g.

The indices are based on a test of homogeneity among multinomial distributions (e.g. see Hogg and Tanis, 1997) that produces the following statistic, Q:

$$Q = \sum_{g=1}^{C} \sum_{\alpha=1}^{A} N^g \frac{(p^g_\alpha - p^*_\alpha)^2}{p^*_\alpha}, \tag{1}$$

where p^*_α is a weighted average of all the group-specific probabilities for state α in which the weights are given by the share of each sample size on the total sum of them.[14] p^*_α is calculated the following way:

$$p^*_\alpha \equiv \sum_{g=1}^{C} p^g_\alpha \frac{N^g}{\sum_{g=1}^{C} N^g} = \frac{\sum_{g=1}^{C} N^g_\alpha}{\sum_{g=1}^{C} N^g}. \tag{2}$$

The weighted average probability performs the comparison of the probabilities across the different group samples. The closer the respective probabilities across samples, the more the weighted average probability resembles each and every of its constituting probabilities (in (2)) and therefore the closer to zero the statistic in (1) is. The null hypothesis of the test is that the C distributions are homogenous, i.e. identical in a statistical sense. Formally: $H_0 : p^1_\alpha = p^2_\alpha = \ldots = p^C_\alpha \; \forall \alpha = 1, ..., A$.

The statistic in (1) has an asymptotic chi-square distribution with $(C-1)(A-1)$ degrees of freedom under the null hypothesis of homogeneity. The statistic also has a maximum value which depends on the number of groups, the number of states (e.g. the categories of multidimensional outcomes) and on each of the group's sample sizes. The

[13] In the case of continuous variables discretization would be needed.

[14] An alternative likelihood ratio statistic is asymptotically identical and has the following form: LR=$2\sum_{g=1}^{C} \sum_{\alpha=1}^{A} N^g \log \left[\frac{N^g_\alpha}{N^g} \frac{\sum_{g=1}^{C} N^g}{\sum_{g=1}^{C} N^g_\alpha} \right]$

maximum value is easily found by noticing that the homogeneity test of multinomial distributions yields the same statistic as the Pearson goodness-of-fit one:

$$Q = \sum_{g=1}^{C} \sum_{\alpha=1}^{A} N^g \frac{(p_\alpha^g - p_\alpha^*)^2}{p_\alpha^*} = \sum_{g=1}^{C} \sum_{\alpha=1}^{A} \frac{\left(N_\alpha^g - \sum_{\alpha=1}^{A} N_\alpha^g \frac{\sum_{g=1}^{C} N_\alpha^g}{\sum_{g=1}^{C} N^g} \right)^2}{\sum_{\alpha=1}^{A} N_\alpha^g \frac{\sum_{g=1}^{C} N_\alpha^g}{\sum_{g=1}^{C} N^g}}. \quad (3)$$

Intuitively one can bring together all the conditional probability vectors, i.e. the distributions of outcomes conditional on a given group, to form a contingency table. In such table N_k^g would be the observed frequency of individuals from a group "g" exhibiting a level k of the outcome; whereas the expected frequency for "g" and k under the null hypothesis of lack of association between groups and outcomes would be given by the expression $\sum_{\alpha=1}^{A} N_\alpha^g \frac{\sum_{g=1}^{C} N_\alpha^g}{\sum_{g=1}^{C} N^g}$ (see for instance, Everitt, 1992). Therefore, (3) can be expressed as:

$$Q = \sum_{g=1}^{C} \sum_{\alpha=1}^{A} \frac{(O_\alpha^g - E_\alpha^g)^2}{E_\alpha^g} \quad (4)$$

Where the O stands for observed and the E for expected frequency. Cramer (1946) showed that the maximum for an expression like (4) is $\min(C-1, A-1)N$ where N stands for the sample size of the contingency table. The corresponding maximum for the statistic (1) is:

$$Q_{\max} = \min(C-1, A-1) \sum_{g=1}^{C} N^g. \quad (5)$$

Thus combining (1) and (5) an heterogeneity index is defined as:

$$H = \frac{Q}{Q_{\max}}. \quad (6)$$

This index fulfils axioms of population invariance[15] and scale invariance.[16] It is also normalized in order to take the value of 0 when the samples under comparison (i.e. the conditional probability vectors) are

[15] That is, if every individual in society is replicated n times, the value of the index remains unaltered.

[16] That is, if the measurements of outcomes are altered proportionately (or additively) in the same way as the boundaries of the partitions of outcomes are altered (i.e. the boundaries that determine whether for one individual $\alpha = k$), then the index's value remains unaffected.

identical. And it takes the value of 1 with maximum association between groups and outcomes. As mentioned, perfect (positive or negative) correlations are just examples of maximum association. In the context of the heterogeneity index (and in general, of contingency tables analysis) maximum association has three related meanings depending on whether $C < A$, $C > A$ or $C = A$. When $C < A$ (more outcome categories or states than groups) maximum association means that for any arbitrary partition of the sets G and R into non-overlapping subgroups then:

$\forall k \in G, R_k \subset R : R_k \rightleftarrows k$

$\wedge$

$R_1 \cup R_2 \cup \ldots \cup R_C = R,$

where R_k is a subset of R made of all those outcome elements attained by group k with positive probability. Maximum or perfect association means therefore that for every group there is a vector of outcomes which is a subset of the outcome vector and is only attainable by that group. For instance, if type g_1 is associated with outcomes α_3 and α_4 (i.e. that there exists a positive probability of being in outcomes α_3 or α_4 conditional on being in group g_1), then no other group is associated with those categories, and similarly if group g_2 is associated with outcomes α_5 and α_6 then group g_1 is not associated with those latter outcomes. The concept of maximum association is not a concept of perfect predictability because if a group is associated with more than one outcome grids (as in the aforementioned examples) then one cannot perfectly predict the final outcome (e.g. it could be either α_3 or α_4 if the group is g_1) although one can accurately predict that someone in group g_1 never attains outcomes α_5 or α_6.

When $C > A$ (more groups than outcomes) the roles of groups and outcome values are reversed, so maximum association means:

$\forall \alpha \in R, G_\alpha \subset G : G_\alpha \rightleftarrows \alpha$

$\wedge$

$G_1 \cup G_2 \cup \ldots \cup G_A = G,$

where G_α is a subset of G made of all those types who attain outcome α with positive probability. Maximum or perfect association means in this case that for every outcome state, or value/category, there is a vector formed by all and only the groups that attain that specific outcome state. Any other subset of groups can not attain that outcome and/or any other outcome is associated with a different, non-overlapping subset of groups.

When $C = A$ maximum association implies that every group is associated exclusively with only one outcome and the reverse holds true: every outcome is associated exclusively with only one group:

$\forall \alpha \in R, k \subset G : k \rightleftarrows \alpha$

The heterogeneity index can be used to compare the degree of heterogeneity between two (or more) mobility matrices in two ways. First define x_t as a variable whose values can be classified into S categories or *states*[17], such that state $i \in \{1, ..., S\}$. The elements of a transition matrix are the probabilities of observing variable x in a given state in the present period conditional on observing it in a given state in a prior period. The probability of observing x in state i in period t, conditional on having observed it in state j in period $t-k$ (where k is just an integer) is defined as: $P_{i|j} \equiv \Pr(x_t = i \mid x_{t-k} = j)$. A conditional probability vector of a transition matrix is one of its columns (or rows, depending on the assortment) which contains the probabilities of being observed in all different states in the present conditional on having been in one specific state in a prior period. Formally, $V_j' = \left(P_{1|j}, P_{2|j}, \ldots, P_{S|j}\right)'$. The probability of observing x in state i in period t *and* in state j in period $t-k$ is:

$$P_{ij} \equiv \Pr(x_t = i \wedge x_{t-k} = j) = P_{i|j} P_{.j},$$

where $P_{.j} \equiv \Pr(x_{t-k} = j) = \sum_{i=1}^{S} P_{ij}$ is the initial probability of being in state j in period t-k.[18]

A first way in which the heterogeneity index can be used to compare two or more mobility regimes is to apply it to joint distributions of parental and offspring's outcomes. Applying (1) through (6) the index would be:

$$H = \frac{\sum_{g=1}^{C} \sum_{i=1}^{S} \sum_{j=1}^{S} N^g \frac{\left(P_{ij}^g - P_{ij}^*\right)^2}{P_{ij}^*}}{\min\left(C-1, S^2-1\right) \sum_{g=1}^{C} N^g}, \tag{7}$$

where the superscripts g on the probabilities denote the probabilities from the joint distribution corresponding to each compared group. Notice that (7) does not compare the transition matrices themselves but the bivariate joint distributions, which are both affected by the posited underlying transition matrices and the initial distributions, $P_{.j}$. Therefore, for instance, two groups may have the same transition matrices but (7) may exhibit a high degree of heterogeneity if the initial distributions are sufficiently different. Or conversely the transition matrices may be significantly different but (7) may not reflect such heterogeneity if those differences are compensated by the initial distributions in a way that renders the joint distributions homogeneous.

Hence a second way to use the heterogeneity index is to apply it directly to the conditional probability vectors of the transition matrices.

[17] Should the variable be continuous it would have to be discretized.

[18] Likewise the probability of being in state i in period t is defined as: $P_{i.} \equiv \Pr(x_t = i) = \sum_{j=1}^{S} P_{ij}$

In such context the homogeneity test (upon which the heterogeneity index is based) was first proposed by Anderson and Goodman (1957). It compares conditional probability vectors across samples individually and then aggregates the respective statistics into a statistic which has an asymptotic chi-square distribution with $(C-1)S(S-1)$ degrees of freedom; where C is the number of compared samples/groups.[19] Since each one of the statistics which make for the total statistic has a specific maximum value (as described above) then a family of indices of heterogeneity for transition matrices based on the multinomial distribution test and fulfilling axioms of normalization, scale invariance and population invariance is:

$$F_H^M := \left\{ \begin{array}{c} H_\eta^M \mid H_\eta^M \equiv \left[\sum_{j=1}^S w_j \left(H_{V_j}\right)^\eta\right]^{\frac{1}{\eta}} \equiv \left[\sum_{j=1}^S w_j \left(\frac{Q}{Q_{\max}}\right)_{V_j}^\eta\right]^{\frac{1}{\eta}} \forall \eta \in \mathbb{R}/\left\{0\right\}, \\ H_\eta^M \mid H_\eta^M \equiv \prod_{j=1}^S \left(H_{V_j}\right)^{w_j} \equiv \prod_{j=1}^S \left(\frac{Q}{Q_{\max}}\right)_{V_j}^{w_j} \; ; \; w_j > 0 \forall j \wedge \sum_{j=1}^S w_j = 1 \end{array} \right\}, \tag{8}$$

where:

$$H_{V_j} \equiv \left(\frac{Q}{Q_{\max}}\right)_{V_j} = \frac{\sum_{g=1}^C \sum_{i=1}^S N_{.j}^g \frac{\left(P_{i|j}^g - P_{i|j}^*\right)^2}{P_{i|j}^*}}{\min\left(C-1, S-1\right) \sum_{g=1}^C N_{.j}^g}, \tag{9}$$

$N_{.j}^g$ is the sample size of the conditional probability vector of group g conditioned on past state j, and the superscripts g on the probabilities denote the probabilities from the conditional probability vectors, and hence transition matrices, corresponding to each compared group. The family F_H^M is characterized by indices which aggregate the heterogeneity indices for specific conditional probability vectors (compared across C matrices), i.e. H_{V_j}, using functions with constant elasticity of substitution, also known as weighted generalized means. Such aggregation choice is common in the inequality literature and other topics of distributional analysis.[20] Notice firstly that any member of the family F_H^M reaches its maximum value of 1, denoting maximum heterogeneity among matrices, if and only if $H_{V_j} = 1 \forall j$. Likewise any member of F_H^M reaches its minimum value of 0, denoting perfect homogeneity among matrices, if and only if $H_{V_j} = 0 \forall j$.

Notice also that w_j manages the relative importance of the heterogeneity among probability vectors conditioned on state j. Several choices

[19] Let Q_j be the statistic for a test of homogeneity of the jth column of C transition matrices, then Anderson and Goodman's (1957) statistic is: $\sum_{j=1}^S Q_j$

[20] See e.g. Atkinson (1970), Foster and Shneyerov (1999, 2000), Foster *et al.* (2005), Foster and Szekely (2008).

for these weights are possible. Three natural choices are the following: $P_{.j}, \overline{P_{.j}}, \frac{1}{S}$, where $\overline{P_{.j}}$ stands for the ergodic distribution of the transition matrix. The first choice is the initial distribution. Coupled with $\eta = 1$, this choice transforms (8) back into (9). Therefore it is not appropriate if the focus is on comparing transition matrices. The second choice is appealing[21] since it weights the contribution of heterogeneity from each conditional probability vector by the long-term relative size of each conditioning state. However in empirical applications it requires that the transition matrix is regular.[22] For the empirical application of this paper I use the third option which has the both the practical appeal of easy calculation and the conceptual appeal of not attaching weights that favour the contribution of certain initial states over others.

The choice of η determines the impact of different magnitudes of probability-vector heterogeneity on H_{η}^{M} and the degree of substitution (compensation) between the contributions to total heterogeneity of the respective conditioning initial states. For $\eta > 1$ initial states with greater heterogeneity contribute more to H_{η}^{M}. By contrast, when $\eta < 1$ initial states with less heterogeneity contribute more to H_{η}^{M}. Again, to avoid favouring any specific initial states ($j \in \{1, ..., S\}$), in the empirical application of this paper I use $\eta = 1$. Another substantial advantage of setting $\eta = 1$ is that the ensuing indices gain the property of *additive decomposability*, i.e. that the index can be split into additive subelements.[23] In empirical applications this property is helpful in tracking the contributions of each conditional probability vector to the transition matrices' total heterogeneity. Thereby an initial-state-neutral index of heterogeneity of transition matrices using $w_j = \frac{1}{S}$ and $\eta = 1$ is the following:

$$H_{1}^{M} \equiv \frac{1}{S} \sum_{j=1}^{S} H_{V_j} \equiv \frac{1}{S} \sum_{j=1}^{S} \left(\frac{Q}{Q_{\max}} \right)_{V_j}. \qquad (10)$$

For exposition purposes, since in some applications values for (7) and/or (8) may lie far from unity[24], an alternative set of indices stem-

[21] It has been used in some applications in the economic mobility literature. For instance see the Atkinson-Dardanoni condition described in Formby *et al.* (2004).

[22] For a description of the regularity property see Luenberger (1979).

[23] In the traditional inequality literature additive decomposability means that the inequality measure "can be expressed as a weighted sum of the inequality values calculated for population subgroups plus the contributin arising from differences between subgroup means" (Shorrocks, 1980, p. 613). In the context of the heterogeneity index additive decomposability is only meant to imply that the total value of the transition matrix index is the sum of the values of the heterogeneity indices for the comparisons of the respective conditional probability vectors.

[24] Pearson's coefficient of contingency also has a similar empirical trait (Everitt,

ming from a monotonic transformation of (7) and (8) respectively, and whose ordinal rankings are perfectly consistent with them, is:

$$H^{\surd} \equiv \sqrt{H} \tag{11}$$
$$H_{\eta}^{\sqrt{M}} \equiv \sqrt{H_{\eta}^{M}}$$

2.1 Relationship to other approaches to mobility comparisons

2.1.1 Comparisons of slope coefficients from OLS-based models

The use of OLS-based models for group comparisons of economic mobility is very popular. These models have the advantages of being adequate for continuous variables (e.g. income), easy to implement, and allow for the inclusion of several control variables. However some of the most popular linear parametric techniques, like OLS models, conflate much information from the joint distribution of parental and offspring's welfare outcome into a handful of parameters. Therefore the indices described above can help to improve the comparison when the question is whether two (or more) mobility regimes are homogeneous by zooming in parts of the joint distribution which might be sources of heterogeneity but may not be detected properly by comparisons of coefficients across groups and/or by the use of dummy variables to represent the effect of belonging to a certain group.

The two key relationships between comparisons of slope coefficients from linear parametric models like OLS and the heterogeneity indices in the way they asses dissimilarity between mobility regimes are the following:

First, whenever the index is applied to joint distributions (e.g. to the P_{ij} elements , (7)) every time $H = 0$ then tests of homogeneity of parameters fail to reject the null hypothesis of homogeneity (e.g. slope coefficients are identical or coefficients of dummy variables are zero). The reverse however is not true. As I show below, when parametric tests cannot reject the null of homogeneity one should not infer homogeneity of the joint distributions. Therefore in this situation one should rely on the index (and its underlying statistic). $H = 0$ is a sufficient but not necessary condition to ensure that the parametric tests fail to reject the null hypothesis.

Second, it follows logically from the first relationship that whenever the tests of homogeneity of parameters reject the null of homogeneity

1992, p. 54-5).

then $H \neq 0$. However the reverse is not true. $H \neq 0$ does not ensure that the tests of homogeneity of parameters reject the null hypothesis. The rejection of homogeneity by the parametric tests is a sufficient but not necessary condition to render $H \neq 0$.

These same conclusions apply when heterogeneity indices are applied to transition matrices, i.e. (8), as long as the initial distributions across the compared samples are identical.[25] Otherwise these relationships do not hold and all combinations of testing outcomes and indices' values are possible because the coefficients in the parametric models depend on the joint distribution of the dependent and the independent variables, which in turn depends both on the initial distribution and the transition matrix.

To illustrate these points I resort to the following simple examples:

Let y and x have a bivariate multinomial distribution such that, for instance, following previous notation $P_{ij} \equiv \Pr(y = y_i \wedge x = x_j) = P_{i|j}P_{.j}$ and let there be two groups: A and B. If a simple linear regression model with an intercept and one slope coefficient is fitted separately for A and B, such that:

$$y_i^g = \alpha^g + \beta^g x_j^g + \varepsilon_{ij}^g \quad \forall g = A, B$$

(where ε_{ij}^g is a random disturbance)), then parameter estimates are:

$$\widehat{\beta^g} = \frac{\sum_{i=1}^S \sum_{j=1}^S P_{ij}^g x_j^g \left(y_i^g - \sum_{i=1}^S P_{i.}^g y_i^g \right)}{\sum_{i=1}^S \sum_{j=1}^S P_{ij}^g x_j^g \left(x_j^g - \sum_{j=1}^S P_{j.}^g x_j^g \right)}$$

$$\widehat{\alpha^g} = \sum_{i=1}^S P_{i.}^g y_i^g - \widehat{\beta^g} \sum_{j=1}^S P_{.j}^g x_j^g$$

If the index (7) is estimated for A and B and $H = 0$ then $\widehat{\alpha^A} = \widehat{\alpha^B}$ and $\widehat{\beta^A} = \widehat{\beta^B}$ since $P_{ij}^A = P_{ij}^B\ \forall i, j$. The reverse is not true. Consider the simpler case where the marginal distributions of A and B are identical, i.e. $P_{i.}^A = P_{i.}^B\ \forall i \wedge P_{.j}^A = P_{.j}^B\ \forall j$. In such case the slope coefficients can only potentially differ in $\sum_{i=1}^S \sum_{j=1}^S P_{ij}^g x_j^g y_i^g$. With $S > 2$, it is easy to find examples in which $\widehat{\alpha^A} = \widehat{\alpha^B}$ and $\widehat{\beta^A} = \widehat{\beta^B}$ do not imply $H = 0$. Similarly, in the case of identical marginal distributions, $\widehat{\beta^A} \neq \widehat{\beta^B}$ requires $H \neq 0$. But the reverse is not true since, as in the last example, $\sum_{i=1}^S \sum_{j=1}^S P_{ij}^A x_j^A y_i^A = \sum_{i=1}^S \sum_{j=1}^S P_{ij}^B x_j^B y_i^B$ does not require $P_{ij}^A = P_{ij}^B\ \forall i, j$. Also, as mentioned above, if and only if $P_{.j}^A = P_{.j}^B\ \forall j$

[25] They also hold if the ergodic distributions are identical.

then $H_\eta^M = 0$ implies $\widehat{\alpha^A} = \widehat{\alpha^B}$ and $\widehat{\beta^A} = \widehat{\beta^B}$, since $H\eta^M = 0$ means that $P_{i|j}^A = P_{i|j}^B \forall i, j$ (and the initial distributions are assumed to be identical). But the reverse is not true: $\widehat{\alpha^A} = \widehat{\alpha^B}$ and $\widehat{\beta^A} = \widehat{\beta^B}$ do not imply $H_\eta^M = 0$ even when $P_{.j}^A = P_{.j}^B \ \forall j$.

The same conclusions are warranted if the model estimated is the following:

$$y_i = \alpha + \gamma I\left(g = A\right) + \beta x_j + \varepsilon_{ij}$$

where I is an indicator function that takes the value of 1 if the statement in parenthesis is true or 0 if it is false. The superscripts A and B on the probabilities indicate that the latter were estimated only using the sample of the respective group. The coefficient of the dummy variable is:

$$\gamma = \frac{\sum_{i=1}^S P_{i.}^A y_i - \sum_{i=1}^S P_{i.} y_i}{w_B} + \left(\sum_{j=1}^S P_{.j} x_j - \sum_{j=1}^S P_{.j}^A x_j\right) \left(\frac{\sum_{i=1}^S \sum_{j=1}^S P_{ij} y_i x_j - \sum_{j=1}^S P_{.j} x_j \sum_{i=1}^S P_{i.}^B y_i + \frac{w_A}{w_B}\left[\sum_{i=1}^S P_{i.} y_i - \sum_{i=1}^S P_{i.}^B y_i\right]}{w_B \sum_{j=1}^S P_{.j} x_j^2 - \sum_{j=1}^S P_{.j} x_j \sum_{j=1}^S P_{j.}^B x_j + w_A\left[\sum_{j=1}^S P_{.j} x_j - \sum_{j=1}^S P_{.j}^A x_j\right]}\right)$$

where w_g is the percentage of the total population size contributed by group g. In this case notice that a condition like $P_{i.}^A = P_{i.}^B \ \forall i \wedge P_{.j}^A = P_{.j}^B \ \forall j$ suffices to render $\gamma = 0$, but such condition does not require $H = 0$, therefore $\gamma = 0$ does not imply $H = 0$. Similarly, $\gamma = 0$ does not imply $H_\eta^M = 0$ even when $P_{.j}^A = P_{.j}^B \ \forall j$. However $H_\eta^M = 0$ does imply $\gamma = 0$ if and only if $P_{.j}^A = P_{.j}^B \ \forall j$.[26]

Finally the same conclusions are accrued in another simple model in which heterogeneity is meant to be captured by different slope coefficients on x:

$$y_i = \alpha + \beta^g x_j + \varepsilon_{ij} \quad \forall g = A, B$$

The respective coefficients of groups A and B are:

[26] Because $H^M = 0$ and $P_{.j}^A = P_{.j}^B$ render $P_{i.}^A = P_{i.}^B$ (since $P_{i.}^g = \sum_{j=1}^S P_{i|j}^g P_{.j}^g$). In turn $P_{i.}^A = P_{i.}^B$ and $P_{.j}^A = P_{.j}^B$ suffice to render $\gamma = 0$.

$$\widehat{\beta^A} = \frac{\left[\sum_{j=1}^S P_{.j}^B x_j^2 - w_B \left(\sum_{j=1}^S P_{.j}^B x_j\right)^2\right] \left[\sum_{i=1}^S \sum_{j=1}^S P_{ij}^A y_i x_j - \sum_{i=1}^S P_{i.} y_i \sum_{j=1}^S P_{.j}^A x_j\right]}{D}$$
$$+ \frac{w_B \sum_{j=1}^S P_{.j}^A x_j \sum_{j=1}^S P_{.j}^B x_j \left[\sum_{i=1}^S \sum_{j=1}^S P_{ij}^B y_i x_j - \sum_{i=1}^S P_{i.} y_i \sum_{j=1}^S P_{.j}^B x_j\right]}{D}$$

$$\widehat{\beta^B} = \frac{\left[\sum_{j=1}^S P_{.j}^A x_j^2 - w_A \left(\sum_{j=1}^S P_{.j}^A x_j\right)^2\right] \left[\sum_{i=1}^S \sum_{j=1}^S P_{ij}^B y_i x_j - \sum_{i=1}^S P_{i.} y_i \sum_{j=1}^S P_{.j}^B x_j\right]}{D}$$
$$+ \frac{w_A \sum_{j=1}^S P_{.j}^A x_j \sum_{j=1}^S P_{.j}^B x_j \left[\sum_{i=1}^S \sum_{j=1}^S P_{ij}^A y_i x_j - \sum_{i=1}^S P_{i.} y_i \sum_{j=1}^S P_{.j}^A x_j\right]}{D}$$

where:

$$D = \left[\sum_{j=1}^S P_{.j}^A x_j^2 - w_A \left(\sum_{j=1}^S P_{.j}^A x_j\right)^2\right] \left[\sum_{j=1}^S P_{.j}^B x_j^2 - w_B \left(\sum_{j=1}^S P_{.j}^B x_j\right)^2\right]$$
$$- w_A w_B \left(\sum_{j=1}^S P_{.j}^B x_j \sum_{j=1}^S P_{.j}^A x_j\right)^2$$

Again it is easy to ascertain that whenever $H = 0$ then $\widehat{\beta^A} = \widehat{\beta^B}$ but the reverse is not true. Therefore the OLS model is not informative as to the underlying heterogeneity of the mobility regimes when the slope coefficients are equal. Similarly whenever $\widehat{\beta^A} \neq \widehat{\beta^B}$ then $H \neq 0$, but the reverse is not true. In addition, similar statements, as in the above examples, apply to H_η^M. That is: $H_\eta^M = 0$ implies $\widehat{\beta^A} = \widehat{\beta^B}$ if and only if $P_{.j}^A = P_{.j}^B$ $\forall j$, but the reverse is not true even when initial distributions are identical. In conclusion, when the tests based on these linear regression models reject homogeneity of the parameters one can reject homogeneity of the underlying joint distributions but the estimates of the parameters themselves should not be used to measure the degree of heterogeneity.[27]

[27] One such heterogeneity index for pair-wise comparisons based on slope coefficients could be:

$$I_{A,B} = \frac{\left|\widehat{\beta^A} \frac{\sigma_x^A}{\sigma_y^A} - \widehat{\beta^B} \frac{\sigma_x^B}{\sigma_y^B}\right|}{2}$$

As the linear parametric models become richer by adding more covariates, particularly higher-order polynomials of the parental explanatory variable,[28] these discrepancies between the heterogeneity indices and the OLS-based comparisons should narrow down since, for a given S-size of the joint distribution, there is less freedom in terms of heterogeneous probabilities available to render the OLS coefficients homogeneous.

2.1.2 Comparisons of summary indicators from transition matrices

The economic mobility literature features extensive examples of summary indicators from transition matrices used to rank mobility regimes in terms of concepts of mobility which the indicators are meant to capture.[29] Van de Gaer, Schokkaert and Martinez (2001) explain that some summary indicators are based on a welfarist approach, others on an axiomatic approach, whereas some others have a more direct, mathematical, *ad hoc* appeal. They mention three concepts of mobility that can be analyzed with transition matrices: mobility as movement, i.e. any departures from perfect path-dependence; mobility as equality of opportunity in terms of the association between parental and offspring's outcome; and mobility as equality of life chances which is a measure of unpredictability of offspring's outcome (as in Parker and Rougier, 2001).

Usually a test of summary indicators has as its null hypothesis the equality of the values of a given summary indicator for two or more samples.[30] When the null hypothesis is rejected not only is homogeneity rejected but also the summary indicator permits ranking the two or more groups according to the underlying mobility concept. The heterogeneity indices, on the other hand, embody a homogeneity test (of multinomial distributions) and are not meant to provide a ranking of individual distributions. Instead they measure the degree of dissimilarity between two or more joint distributions or transition matrices. They can be used to rank groups of matrices according to their degree of dissimilarity (as in the empirical example of this paper).

---

or

$$\overline{I_{A,B}} = \left|\widehat{\beta^A}\frac{\sigma_x^A}{\sigma_y^A} - \widehat{\beta^B}\frac{\sigma_x^B}{\sigma_y^B}\right|$$

if the slope coefficients are not expected to be negative.

[28] Gaviria (2007), for instance, estimates regressions of offspring's education as a function of parental education in which the latter variable enters as a second-order polynomial.

[29] Some classic examples of mobility measures and related concepts are in Shorrocks (1978), Sommers and Conlisk (1979), Bartholomew (1982), Parker and Rougier (2001) and van de Gaer et al. (2001).

[30] The statistical tools used to test differences between groups in summary indicators have been developed by Trede (1999) and Formby et al. (2004).

In this section I do not go through the extensive list of summary indicators. Because these summary indicators are the result of linear and non-linear combinations of the transition probabilities the claims made about the connection between the heterogeneity indices and the OLS estimates also hold when comparing results of tests on summary indicators and the heterogeneity indices. That is, whenever $H = 0$ any test of summary indicators fails to reject the homogeneity of the indicators. The reverse is not true. And whenever these tests reject homogeneity then $H \neq 0$, but the reverse is not true. Two examples illustrate the point:

First, consider the trace index (Shorrocks, 1978):

$$T^g = \frac{S - \sum_{i=1}^{S} P_{i|i}^g}{S-1} \quad \forall g = A, B$$

Whenever $H = 0$ between A and B, it follows that $T^A = T^B$. However the reverse is not true. Firstly, $\sum_{i=1}^{S} P_{i|i}^A = \sum_{i=1}^{S} P_{i|i}^B$ can be achieved with different individual values for the respective diagonal probabilities. Secondly, even if $P_{i|i}^A = P_{i|i}^B \; \forall i \in \{1, ..., S\}$, there can be substantial heterogeneity between the respective off-diagonal probabilities. The same conclusions applied when considering H_η^M instead of H.

Secondly, consider the second-largest eigenvalue index (Sommers and Conlisk, 1979):

$$E^g = 1 - |\lambda_2^g| \;\; \forall g = A, B$$

where λ_2^g is the second largest eigenvalue of the transition matrix of group g. The same result follows: whenever $H = 0$ between A and B, $E^A = E^B$, but the reverse is not true. To illustrate further imagine $S = 2$ such that:

$$M^g = \begin{pmatrix} P_1^g & P_2^g \\ 1 - P_1^g & 1 - P_2^g \end{pmatrix}$$

Since transition matrices are stochastic matrices, the largest eigenvalue is equal to unity.[31] In this example with $S = 2$ the second largest eigenvalue is: $\lambda_2^g = P_1^g - P_2^g$. Hence $\left|\lambda_2^A\right| = \left|\lambda_2^B\right|$ does not guarantee $H = 0$, but $\left|\lambda_2^A\right| \neq \left|\lambda_2^B\right|$ implies $H \neq 0$. The same relationship holds when H is replaced by H_η^M. Therefore a similar conclusion as in the previous subsection follows: a test of summary indicators should not be

[31] A stochastic matrix is a square matrix in which either all its rows or all its columns add up to one. Whenever both rows and columns add up to one then the matrix is bi-stochastic, as is the case with quantile transition matrices.

used to assess the degree of heterogeneity whenever its null hypothesis cannot be rejected.

3 Empirical application: educational mobility of men and women in Peru

In this section I illustrate the usefulness of the heterogeneity index in mobility analysis with an application to intergenerational mobility of education among adult men and women in Peru. The heterogeneity index allows me to show that the association between father's education and adult male's education increasingly starts to resemble that between father's education and adult female's education in Peru among younger cohorts of adults. In other words the index provides (necessary but insufficient) evidence to claim that the effect of the father's education on sons have become less different from that of father's education on daughters.[32] A similar assessment could be performed by fitting parametric models as Solon (1992) does for income or Behrman *et al.* (2001) and Gaviria (2007) do for education. They all run regressions of offspring's outcome on a polynomial form (e.g. linear, quadratic, etc.) of the same outcome for the parents. One could run these models for male and females separately and then test the equality of the slope parameters. However, unlike the non-parametric option proposed in this paper, these parameters do not fully capture the likely differential impact of different levels of parental outcome on the offspring's.

Peru is an interesting case-study, first, because it has the highest coefficient of correlation between parental and offspring's education among a sample of 42 countries (Hertz *et al.*, 2007).[33] However, as other developing countries it underwent a dramatic socioeconomic transformation during the 20th century to such an extent that it can be hypothesized to exhibit heterogeneity in the mobility regimes across cohorts and in several welfare outcomes beyond education, like occupation, living standards or general social status indicators. Indeed by 1940, date of the earliest census of the 20th 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 and Cueto, 2000). Simultaneously, starting incipiently in the 1920s

[32] The association of maternal education with offspring's education could also be studied. For other studies focused on intergenerational transmission of education see, for instance, Berhman *et al.*, 2001; Pasquier-Doumer, 2003; Holmlund *et al.*, 2006; Gaviria, 2007; Hertz *et al.* (2007).

[33] The correlation coefficients estimated by Hertz *et al.* (2007) are simple averages of coefficients for ten offspring cohorts within an age range from 20 to 69 years.

and 1930s the construction of state school facilities boomed between the 1940s and the 1960s (Portocarrero *et al.*, 1988).

These and other major changes in both demand and supply-side factors led to increasing levels of literacy and educational attainment.[34] They also increasingly weakened the links between parental and offspring education[35] which motivates the question as to whether a lower degree of heterogeneity between the educational mobility matrices of adult males and females (both with respect to their father's education) should be expected among the youngest cohorts. Figures 1 through 8 show an apparent increase in resemblance between the educational distributions of male and women among younger cohorts, which points to a genuine weakening of the first-order stochastic dominance of males' distributions over females', tested using Anderson's (1996) multiple-contrasts test (Table 2).[36] Is it the case that the links between parental and offspring education have also become more similar between adult males and females? This question is answered with the heterogeneity index.

3.1 Data

The dataset is the Peruvian 2001 Household National Survey (ENAHO) with information for 16,515 households (INEI, 2001). Information on the education of household heads, spouses and respective offspring is available in an educational module. For the education of the parents of heads and spouses there is a special module on "Household perception" which includes 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.[37] Matching categories were defined for constructing the respective variable in which the tertiary categories include technical and university education. The educational variable therefore has the following ordered categories: no education (value equal to 1); incomplete primary (equal to 2); complete primary (equal to 3); incomplete secondary (equal to 4); complete secondary (equal to 5); and any post-secondary education accomplished (which includes incomplete and

[34]For instance, illiteracy rates fell from 59% in 1940 to 11% in 1993.

[35]See Yalonetzky (2008, Chapter 4).

[36]The z-statistics decrease with the youth of the cohorts although in principle this decrease could both be due to genuine increase in resemblance or to a sample size effect.

[37]Therefore the analysis is performed in levels as opposed to years of education.

complete technical and university tertiary and has a value of 6).

The adult sample is comprised of household heads, spouses and adult offspring cohabiting with parents. A minimum age of 25 was defined for being an adult. People who reported being studying were excluded from the sample to dispose of censored observations.[38] It is still possible, however, that young people who claimed to be not undertaking any education at the interview, and thus made it into the dataset, were just caught in the middle of a break from studying.

With respect to cohorts, even though the overall Peruvian sample size is larger than others for developing countries, which allows a more refined cohort analysis, it is still not large enough to define cohorts by year of birth. Therefore I clustered together some years in order to define cohorts, most of them being five-years[39], as shown in Tables 1a and 1b. The sample sizes for every cohort and conditional probability vector are in the same tables.

Even though few Living Standard Measurement Surveys in developing countries are as extensive with regard to retrospective information on parents of household heads and spouses as the Peruvian ENAHO 2001[40], I do not have information on the age at which parents of household heads and spouses gave birth to them. Such information is only available for adults found living in the same households as their parents. Therefore, because I want to account for the information from household heads and spouses in order to have the most representative sample of all adults, I cannot 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 used this dataset or others lacking the same information.[41]

Similarly controlling for changes in the quality of education along time and its interaction with the attained level of education would be interesting.[42] Unfortunately the database does not have information on

[38] In other words, the value of the education variable for them was at least the one they reported since their education was still ongoing at the time of the interview.

[39] Hertz *et al.* (2007) also work with five-year cohorts.

[40] Some exceptions, among developing countries, include the five African datasets used by Bossuroy *et al.* (2007), the Brazilian Household Survey used by Bourguignon et al. (2007) and Cogneau and Cigneaux (2005), the Bangladeshi dataset of rural households from the national census used by Assadulah (2006) and the datasets for developing countries listed by Hertz *et al.* (2007). The special ENNIV module used by Benavides (2002) is another relevant exception, although the latter has a significantly smaller sample size.

[41] For instance, Pasquier-Doumer (2003).

[42] For instance, Calonico and Nopo (2007) provide evidence of private-public education gaps in earnings and other labor market outcomes in Peru.

the type of school attended by the parents of the household head and their spouses. For similar reasons the rest of the literature does not account for these factors in studies related to Peru, as well as other developing countries with a paucity of similar data in the retrospective question modules.

3.2 Results

Table 3 shows the homogeneity test results for the eight cohorts of the Peruvian sample. The second column indicates clear rejection of the null hypothesis of homogeneity between the transition matrices of males and females for all cohorts although there is a non-monotonic reduction of the p-values among the youngest cohorts. Interestingly, the third column shows that the rejection of the null hypothesis is mostly due to heterogeneity stemming from the probability vectors (the matrices' columns) conditioned on the lowest levels of parental education. In fact the homogeneity of the conditional probability vectors related to the highest level of parental education can never be rejected at 95 per cent of confidence. Likewise in several cohorts there is no evidence to reject that the probability vectors conditioned on parents having attained complete or incomplete secondary education are homogeneous.

The trend across cohorts in the degree of heterogeneity between males' and females' transition matrices is portrayed in Figure 9 and Table 4. Going from oldest to youngest, the degree of heterogeneity reaches a peak at the cohort born between 1947 and 1951. Thereafter it declines monotonically as younger cohorts are considered. The bootstrapped confidence intervals suggest that the index's values for contiguous cohorts may not be different with statistical significance although such statistical significance is attained as cohorts are compared to others farther away in time. Considering the statistically different values of the indices for the oldest and youngest cohorts and the trend described above, the evidence points to an actual decrease in the degree of heterogeneity in the mobility processes between adult men and women in Peru.

Finally Table 5 shows the contributions of heterogeneity between specific conditional probability vectors to total heterogeneity of the mobility matrices. This decomposition is a convenient property of index (8).[43] Generally the contributions of the probability vectors conditioned on the lowest levels of parental education tend to be the greatest whereas those of the vectors conditioned on the highest levels of parental education tend to be the lowest, across cohorts. For instance, the contribution

[43] The contributions of the comparison of conditional probability vectors conditioned on state j is: $W_{V_j} \equiv \frac{H_{V_j}}{SH_1^M}$. Hence from (10): $\sum_{j=1}^{S} W_{V_j} = 1$.

of the conditional probability vector with the lowest parental education is never *below* 25% whereas that of the conditional probability vector with the highest parental education is never *above* the same percentage mark.

4 Conclusions

The heterogeneity indices proposed in this paper belong in the family of indices based on transition matrices.[44] Unlike indices based on the information of one transition matrix, the heterogeneity indices summarize information from two or more. The reason is that the other indices rank mobility matrices in terms of conceptual criteria related to the degree or nature of the association between the variables in rows and those in columns, whereas the heterogeneity indices rank pairs (or larger groups) of matrices in terms of how similar they are within each pair or group. The indices provides a non-parametric alternative to comparing slope parameters from regression models in the traditional intergenerational mobility literature.

In the empirical application to educational mobility matrices in Peru the heterogeneity index proved useful to show an interesting decline in the degree of heterogeneity between mobility matrices of adult males and females, both describing the relationship between their education and their fathers', among the youngest cohorts. The analysis suggests that the reduction in heterogeneity is led by an increase in the relative resemblance of the matrices' probability vectors conditioned on the lowest levels of parental education. Therefore the heterogeneity index implies that the differences in the causes that generate an association between parental and offspring's education are becoming less pronounced between male and female offspring. Whether this decreasing trend will continue is an open and interesting research question for the future. The adult people considered in this paper were all born in 1976 or before, i.e. before or right at the beginning of a period of stagnation in the growth of per capita GDP in Peru. The educational attainment of most of them was not completely damped by the effects of such stagnation on the incentives, opportunities and constraints to invest in education. Similarly most of them were too old to reap any benefits from improvements in educational facilities and the expansion of tertiary education during the 1990s.

[44] For examples see Shorrocks (1978), Sommers and Conlisk (1979), Bartholomew (1982), Trede (1999), Fields (2001), van de Gaer et al. (2001), Parker and Rougier (2001), Formby et al. (2004).

References

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

[2] Anderson, G. (1996), 'Nonparametric tests of stochastic dominance in income distributions', *Econometrica*, 64(5): 1183-1193.

[3] Asadullah, N. (2006), "Intergenerational economic mobility in rural Bangladesh", manuscript.

[4] Atkinson, A.A.B. (1970), "On the measurement of inequality", *Journal of Economic Theory*, 2: 244-63.

[5] Banerjee, A. and Newman, A. (1993), "Occupational choice and the process of development", *Journal of Political Economy*, 101(2): 274-298.

[6] Bartholomew, D.J. (1982), *Stochastic models for social processes*, Chichester: John Wiley.

[7] Becker, G. (1971), *The economics of discrimination*, Chicago University Press, Second Revised Edition.

[8] Becker, G. (1993), *A treatise on the family*, Harvard University Press, Second Edition.

[9] Behrman, J., Gaviria, A. and Szekely, M. (2001), "Intergenerational mobility in Latin America", *Economia*, 2(1): 1-44.

[10] Benabou, R. and Ok, E. (2001), "Mobility as progressivity: ranking income processes according to equality of opportunity", NBER Working Paper No. W8431.

[11] Benavides, M. (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.

[12] Binder, M. and Woodruff, C. (2002), "Inequality and intergenerational mobility in schooling: the case of Mexico", *Economic Development and Cultural Change*: 250-267.

[13] Bjorklund, A. and Jantti, M. (2000), "Intergenerational mobility of socio-economic status in comparative perspective", *Nordic Journal of Political Economy*, 26: 3-32.

[14] Bossuroy, T., Cogneau, D. and Hiller, V. (2007), "Intergenerational mobility in five African countries", mimeograph.

[15] Bourguignon, F., Ferreira, F. and Menendez, M. (2007), "Inequality of opportunities in Brazil", *Review of Income and Wealth*, 53(4): 585-618.

[16] Calonico, S. and Nopo, H. (2007), "Where did you go to school? Private-public differences in schooling trajectories and their role on earnings", *Well-being and social policy*, 3(1): 25-46.

[17] Chadwick, L. and Solon, G. (2002), “Intergenerational income mobility among daughters”, *The American Economic Review*, 92(1): 335-344.

[18] Charles, K.K. and Hurst, E. (2003), “The correlation of wealth across generations”, *The Journal of Political Economy*, 111(6): 1155-1182.

[19] Cogneau, D. and Cigneaux, J. (2005), “Earnings inequalities and educational mobility in Brazil over two decades”, Document de travail, DIAL.

[20] Contreras, M. and Cueto, C. (2000), *Historia del Peru contemporaneo*, Lima: Red para el desarrollo de las ciencias sociales en el Peru.

[21] Corcoran, M. (1995), “Rags to rags: poverty and mobility in the United States”, *Annual Review of Sociology*, 21: 237-67.

[22] Cramer, H. (1946), *Mathematical methods for Statistics*, Princeton: Princeton University Press.

[23] Crompton, R. (1998), *Class and stratification. An introduction to current debates*, Polity Press, 2nd edition.

[24] Di Pietro, G. and Urwin, P. (2003), “Intergenerational mobility and occupational status in Italy”, *Applied Economics Letters*, 10: 793-797.

[25] Erikson, R. and Goldthorpe, J. (2002), “Intergenerational inequality: a sociological perspective”, *The Journal of Economic Perspectives*, 16(3): 31-44.

[26] Everitt, B.S. (1992), *The analysis of contingency tables*, Monographs on Statistics and Applied Probability, 45, Second Edition: Chapman & Hall.

[27] Fields, Gary (2001), *Distribution and Development: a new look at the developing world*, MIT Press.

[28] Formby, J., Smith, J. and Zheng, B. (2004), “Mobility measurement, transition matrices and statistical inference”, *Journal of Econometrics*, 120: 181-205.

[29] Foster, J. and Shneyerov, A. (1999), " A general class of additively decomposable inequality measures", *Economic Theory*, 14(1): 89-111.

[30] Foster, J. and Shneyerov, A. (2000), "Path independent inequality measures", *Journal of Economic Theory*, 91: 199-222.

[31] Foster, J. and Szekely, M. (2008), "Is economic growth good for the poor? Tracking low income using general means", *International Economic Review*, 49(4): 1143-72.

[32] Foster, J., Felipe Lopez-Calva, L. and Szekely, M. (2005), "Measuring the distribution of human development: methodology and an

application to Mexico", *Journal of Human Development and Capabilities*, 6(1): 5-25.

[33] Galor, O. and Zeira, J. (1993), "Income distribution and Macroeconomics", *Review of Economic Studies*, 60: 35-52.

[34] Gaviria, A. (2007), "Social mobility and preferences for redistribution in Latin America", *Economia*, 8(1): 55-96.

[35] Hayes, B. and Miller, R. (1991), "Intergenerational mobility in the Republic of Ireland. Does gender make a difference?", *The Sociological Quarterly*, 32(4): 621-635.

[36] 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.

[37] Hogg, R. and Tanis, E. (1997), *Probability and statistical inference*, New Jersey: Prentice Hall.

[38] Holmlund, H., Lindahl, M. and Plug, E. (2006), "Estimating intergenerational schooling effects", manuscript.

[39] INEI (2001), *Encuesta Nacional de Hogares (ENAHO) IV trimestre: condiciones de vida y pobreza*, Lima: INEI.

[40] Lillard, L. and Willis, R. (1994), "Intergenerational educational mobility. Effects of family and state in Malaysia", *JournalofHumanResources*, 29(4): 1126-1166.

[41] Luenberger, D. (1979), *Introduction to dynamic systems: theory, models, and applications,* New York, Chichester: Wiley.

[42] Kuznets, S. (1966), *Modern Economic Growth: rate, structure and spread*, Yale University Press, New Haven.

[43] Maoz, Y. and Moav, O. (1999), "Intergenerational mobility and the process of development", *The Economic Journal*, 109: 677-697.

[44] Mooney, C. and Duval, R. (1993), *Bootstrapping. A nonparametric approach to statistical inference*, Series: Quantitative applications in the social sciences, 95: A Sage University Paper.

[45] Parker, S. and Rougier, J. (2001), "Measuring social mobility as unpredictability", *Economica*, 68: 63-76.

[46] Pasquier-Doumer, L. (2003), 'L'evolution de la mobilite scolaire intergenerationelle au Perou depuis un siecle', Document de travail 2003/02, DIAL.

[47] Piketty, T. (1999), "Theories of persistent inequality and intergenerational mobility", in Atkinson, Anthony and Francois Bourguignon (editors), *Handbook of Income Distribution*, Vol 1., Elsevier Science: 430-476.

[48] Portocarrero, F., Beltran, A. and Zimmerman, A. (1988), Inver-

siones publicas en el Peru (1900-1968): una aproximacion cuantitativa, Cuadernos de Investigacion, Lima: CIUP.
[49] Rosen, H. (1992), Public Finance, Irwin, Homewood.
[50] Shorrocks, A. (1978), "The measurement of mobility", *Econometrica*, 46(5): 1013-24.
[51] Shorrocks, A. (1980), "The class of additively decomposable inequality measures", *Econometrica*, 48: 613-25.
[52] Solon, G. (1992), "Intergenerational income mobility in the United States", *The American Economic Review*, 82(3): 393-408.
[53] Solon, G. (2002), "Cross-country differences in intergenerational earnings mobility", *The Journal of Economic Perspectives*, 16(3): 59-66.
[54] Sommers, P.M. and Conlisk, J. (1979), "Eigenvalue immobility measures for Markov chains", *Journal of Mathematical Sociology*, 6: 253-76.
[55] Stoline, M. and Ury, H. (1979), "Tables of the studentized maximum modulus distribution and an application to multiple comparisons among means", *Technometrics*, 21(1): 87-93.
[56] Trede, M. (1999), "Statistical inference for measures of income mobility", J*ahrbucher fur Nationaloknomie und Statistk*, 218: 473-90.
[57] Van de Gaer, D., Schokkaert, E. and Martinez, M. (2001), "Three meanings of intergenerational mobility", *Economica*, 68: 519-37.
[58] Yalonetzky, G. (2008), *Essays on Economic Mobility*, DPhil Dissertation.

5 Tables

Table 1a: Sample size per cohort and father's educational level. Adult males

	Educational level						
Cohorts	1	2	3	4	5	6	Total
1: 25-29	308	699	387	190	207	160	1951
2: 30-34	397	721	392	122	176	103	1911
3: 35-39	424	751	388	96	164	77	1900
4: 40-44	405	637	353	83	119	60	1657
5: 45-49	395	569	276	44	97	54	1435
6: 50-54	362	390	258	27	76	28	1141
7: 55-59	304	295	170	25	49	28	871
8: 60+	1018	740	399	62	121	51	2391

Table 1b: Sample size per cohort and father's educational level. Adult females

	Educational level						
Cohorts	1	2	3	4	5	6	Total
1: 25-29	393	745	452	179	282	164	2215
2: 30-34	458	797	429	148	225	111	2168
3: 35-39	512	746	464	119	172	93	2106
4: 40-44	459	612	388	83	142	76	1760
5: 45-49	459	481	298	52	111	49	1450
6: 50-54	416	355	235	30	61	28	1125
7: 55-59	374	250	174	24	53	26	901
8: 60+	994	563	325	37	97	48	2064

Table 2: First-order stochastic dominance test. Ho: educational distribution of females – educational distribution of males=0. Z-scores

Cumulatives	Cohorts							
	1	2	3	4	5	6	7	8
1	5.84582	7.40242	9.46915	12.1482	12.3445	13.4571	13.8392	21.3325
1+2	8.28466	8.54794	9.47119	11.1629	10.9819	10.5909	9.71518	12.5452
1+2+3	7.70468	7.09466	8.10759	9.04718	8.09397	9.23039	7.18209	8.78771
1+2+3+4	5.55559	3.15636	5.60473	5.85672	6.10475	8.23213	5.97175	6.71522
1+2+3+4+5	0.52411	0.70856	1.86764	2.08552	5.28624	6.33594	5.38104	5.56447

The critical values for the studentized maximum modulus distribution for 6 contrasts, infinite degrees of freedom and 98% of confidence, are -3.143 and 3.143 (Stoline and Ury, 1979, p. 88). Since there are 5 contrasts in the table, the critical values are slightly lower (Stoline and Ury do not offer values for 5 contrasts but they do for 3 contrasts which for the same degrees of freedom and confidence are equal to -2.934 and 2.934). According to Anderson's (1996) the null hypothesis of homogeneity in educational distributions is rejected for all cohorts in favour of the alternative hypothesis of first-order stochastic dominance by males' distribution. Notice though that the z-scores decrease even though sample sizes are not decreasing monotonically (Tables 3.1a and 3.1b), which suggests also a relative decrease in the distributions' heterogeneity.

Table 3: Homogeneity test results. $H_0 : P_{i|j}^{Males} = P_{i|j}^{Females}$ $\forall i, j = 1, \ldots, S$

Cohorts	Statistic	P-value	Homogeneous conditional probability vectors*
1: 25-29	136.619	1.49E-15	4, 5, 6
2: 30-34	140.3578	3.33E-16	3, 4, 5, 6
3: 35-39	208.0513	1.53E-28	4, 6
4: 40-44	262.5193	5.71E-39	5, 6
5: 45-49	269.084	3.02E-40	4, 5, 6
6: 50-54	274.6629	2.47E-41	5, 6
7: 55-59	259.038	2.7E-38	4, 5, 6
8: 60+	583.3616	8.2E-104	4, 6

*At 95% of confidence. The numbers are the educational level of the father that defines the conditional probability vector. For instance a 6 means that the distribution of education of males conditional on having a father with some tertiary education is not statistically different from that of females with same parental background for the corresponding cohort-row of the table.

Table 4: Heterogeneity indices and confidence intervals

Cohorts	Heterogeneity index	Confidence intervals by bootstrapping method*			
		Percentile		Bias-corrected percentile	
1: 25-29	0.028385	0.026062	0.045681	0.022108	0.03216
2: 30-34	0.032098	0.029	0.052706	0.027081	0.036386
3: 35-39	0.043956	0.040095	0.070407	0.036295	0.050088
4: 40-44	0.068266	0.062886	0.101572	0.058379	0.074128
5: 45-49	0.075425	0.069807	0.131446	0.061533	0.083077
6: 50-54	0.118398	0.105941	0.187736	0.078992	0.135558
7: 55-59	0.11434	0.110712	0.186168	0.092141	0.116986
8: 60+	0.091209	0.088747	0.142327	0.076713	0.095227

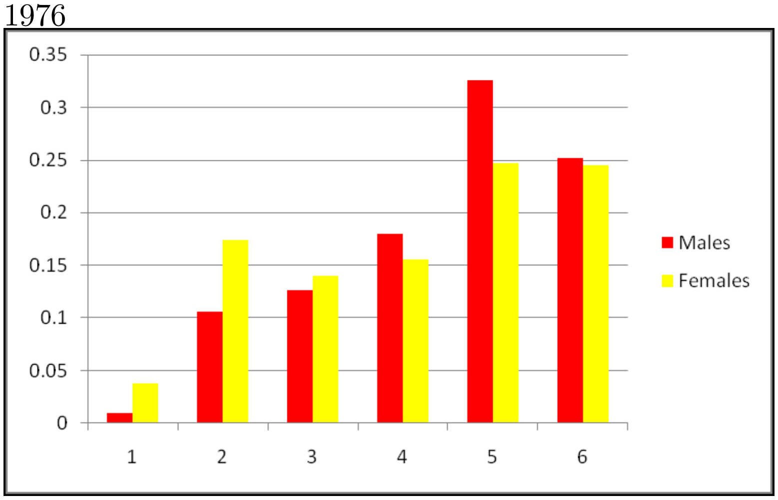
500 re-samplings were performed. For technical details on the bootstrapping methods see Mooney and Duval (1993).

Table 5: Relative contributions of the conditional probability vectors to heterogeneity per cohort*

Cohort	Conditioning educational level					
	1	2	3	4	5	6
1: 25-29	0.365145	0.208725	0.183754	0.087355	0.049659	0.105362
2: 30-34	0.392784	0.159208	0.063662	0.102916	0.058888	0.222542
3: 35-39	0.408207	0.140704	0.12129	0.044422	0.209301	0.076077
4: 40-44	0.414784	0.145745	0.044376	0.195389	0.064316	0.135389
5: 45-49	0.449315	0.118602	0.072876	0.120027	0.079635	0.159547
6: 50-54	0.269498	0.136145	0.076701	0.295353	0.10422	0.118084
7: 55-59	0.326836	0.179905	0.101407	0.127324	0.052796	0.211732
8: 60+	0.343354	0.203014	0.075342	0.14	0.15656	0.081729

*The rows add up to 1.

Figure 1: Educational attainment of Peruvians born between 1972 and 1976



6 Figures

Figure 2: Educational attainment of Peruvians born between 1967 and 1971

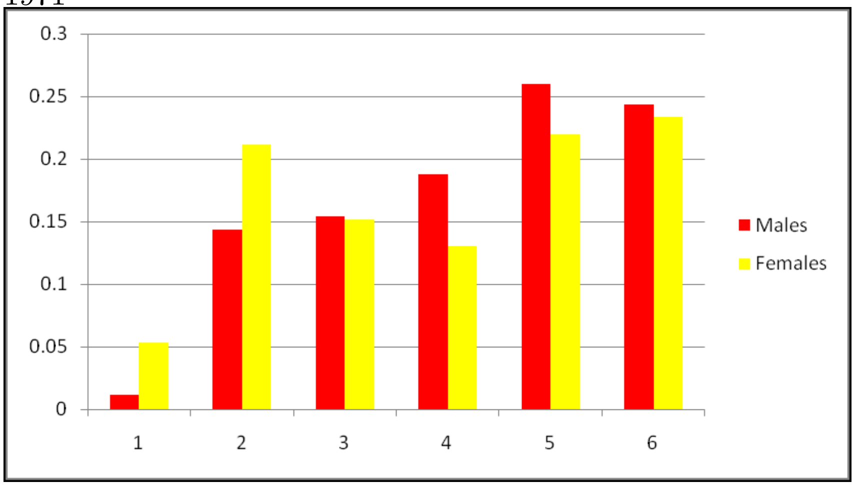


Figure 3: Educational attainment of Peruvians born between 1962 and 1966

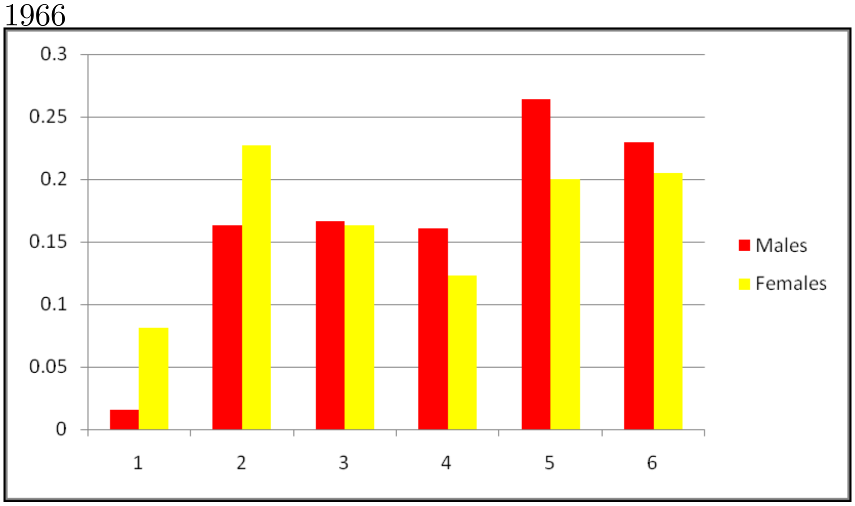


Figure 4: Educational attainment of Peruvians born between 1957 and 1961

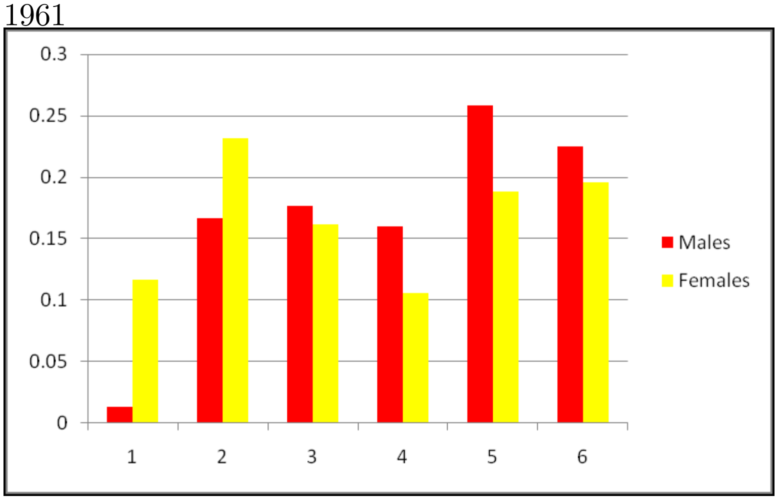


Figure 5: Educational attainment of Peruvians born between 1952 and 1956

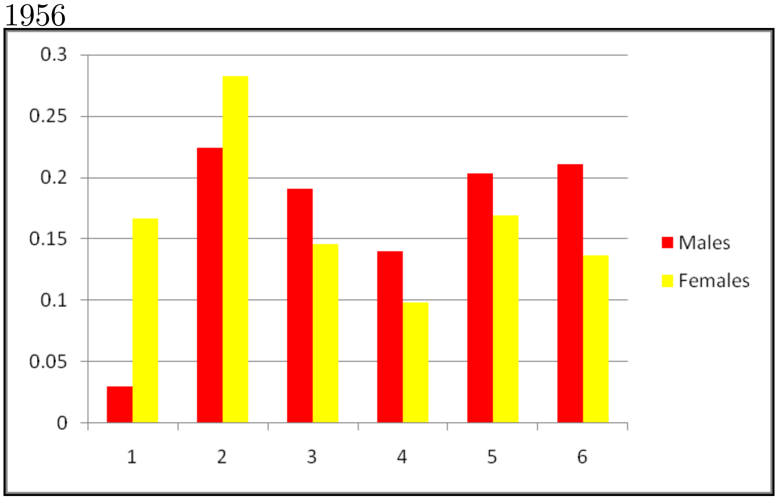


Figure 6: Educational attainment of Peruvians born between 1947 and 1951

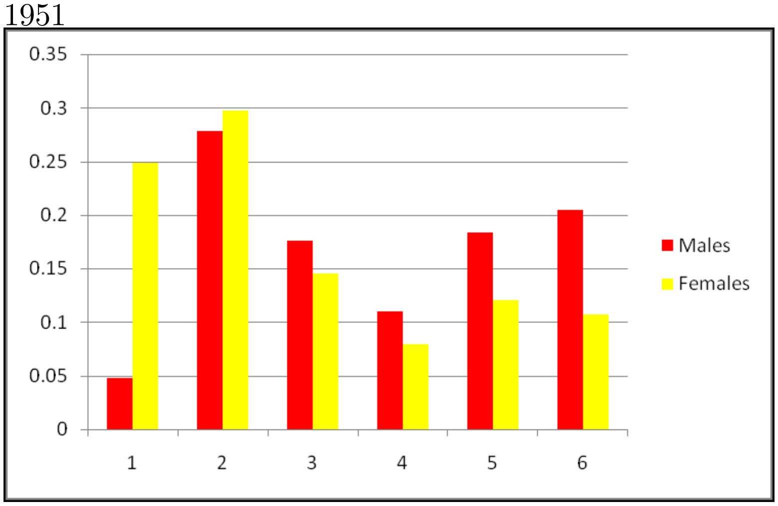


Figure 7: Educational attainment of Peruvians born between 1942 and 1946

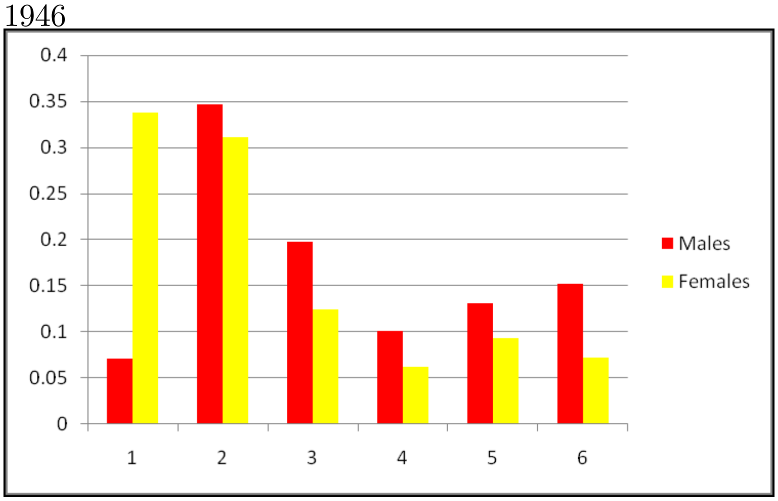


Figure 8: Educational attainment of Peruvians born before 1942

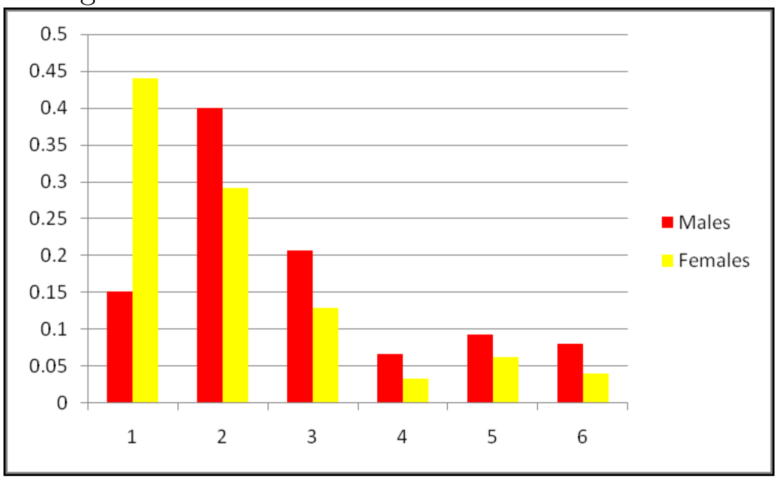


Figure 9: Heterogeneity index of educational mobility matrices across cohorts

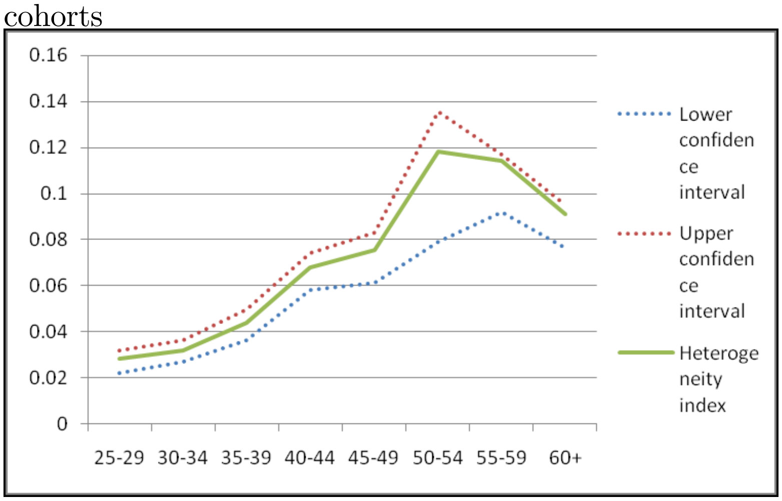


Figure 10: The 95% confidence intervals are estimated using a bias-corrected percentile method of bootstrapping. The horizontal axis measures the age of cohorts in 2001.