Regional convergence of income and education: 
Investigation of distribution dynamics*

Jørn Rattsø and Hildegunn E. Stokke
Department of Economics, Norwegian University of Science and Technology
jorn.rattso@svt.ntnu.no, hildegunnes@svt.ntnu.no

Abstract

Recent US studies find that regional education levels diverge and that this can explain the decline of income convergence. We challenge the suggested relationship between movements in the distributions of income and education based on Norwegian data. We apply Kernel density functions and Markov chains and offer a test of co-movements in the distributions of education and income. Education levels converge and are equalized across the country, and this process coincides with income convergence. However, the test indicates that transitions in the income and education distributions are basically unrelated. The education level increases in large cities with limited income growth and the income growth is strong in regions with continued low education level.

JEL codes: C14, I21, O15, R11, R12
Key words: Income convergence, education convergence, smart cities, human capital, Markov chain, kernel density function

Date: May 27, 2013

* The project is funded by the Demosreg program of the Norwegian Research Council. We appreciate data handling by Jo Jakobsen, discussions at the 2011 ERSA conference, the 2011 RSA conference, the 2011 EEA conference, the 2011 UEA meeting, the 2011 International Atlantic Economic Conference, the NUPI workshop on ‘Barriers to Entry, Internationalisation and Technology’ and at staff seminars at Brown University, the Kiel Institute for the World Economy, Norwegian University of Science and Technology, Statistics Norway, the University of Bergen and the University of California Irvine, and comments from Rolf Aaberge, Nathaniel Baum-Snow, Frank Bickenbach, Hans Bonesrønning, Jan Brueckner, Fredrik Carlsen, Tony Champion, Torberg Falch, Oded Galor, Amihai Glazer, Holger Gorg, Vernon Henderson, Magne Mogstad, Bjarne Strøm, Isabel Tecu, and Abigail Wozniak.
1. Introduction

Income levels expand hand in hand with education levels. The correlation certainly is true at the national level – rich countries have high levels of education. Education increases individual labor productivity and thereby income. In addition, education is assumed to generate positive externalities and then the social return is larger than the private. The cross-country data embody large heterogeneity in institutions and technology, and the relationship between education and income has been investigated in more detail at the regional level. Rauch (1993) innovated the literature by looking at the geographic concentration of highly educated in cities. Using US metropolitan data he shows that the average level of education has an independent effect on the wage level and thereby confirms the external effect of sharing knowledge. The dynamics of the relationship have not been much studied.

Recent evidence at the regional level is summarized by Gennaioli et al. (2013) and Glaeser and Resseger (2010). ‘Skilled cities’ or ‘smart cities’ (Glaeser and Saiz, 2004, Shapiro, 2006, and Winters, 2011) are seen as engines of growth. Berry and Glaeser (2005) find that education levels diverge across cities in the US, and argue that divergence of education levels is related to decline of income convergence. In a study of OECD countries, Wolff (2000) finds convergence in schooling levels and observes that it corresponds to convergence in labor productivity levels. We question the suggested relationship between movements in the distributions of income and education based on Norwegian data.

The dynamic process of accumulating education and income generation is studied for 89 NUTS-4 regions. Based on Kernel density functions and Markov transition matrices we find evidence of both income and education convergence among regions. Pearson and Likelihood Ratio tests are used in the literature to investigate time stationarity. We suggest that the same test design can be used to study co-movements in the distributions of income and education. The test shows that the transitions in the income distribution are independent of whether regions are moving up or down in the distribution of relative education levels.

We see this distribution analysis as an empirical challenge to the literature on education and economic growth. In our data the education level increases in large cities with limited income growth. Interestingly, the young and highly educated concentrate in large cities with high
income level, but high income growth in the urban regions does not follow. At the other end of the income distribution income growth is strong in regions with fairly low education level.

Our results should be understood in the context of institutions and characteristics in Norway. This is a welfare state with a large public sector, strong trade unions and income generation influenced by oil and other resource oriented industries (electric power, salmon). The distribution of income is relatively equal, and as summarized by Oreopoulos and Salvanes (2011), the monetary return to education is lower relative to other high-income countries. The lack of income growth in high skilled cities may result from a compressed wage structure and low return to education. High income growth in the periphery is probably related to resource based industries such as salmon farming, electric power generation and oil service.

The distribution analysis is outlined in section 2. Section 3 shows the evidence of regional income convergence in Norway. Section 4 analyzes the accumulation of the education level across regions. The relationship between movements in the distributions of income and education is investigated in section 5. Concluding remarks are offered in section 6.

2. Distribution analysis

We take benefit of long time series of education and income and investigate the distributions across 89 NUTS-4 regions in Norway. The regions are aggregated from municipalities and are defined by common labor markets. The average region has about 40-50,000 inhabitants. We have collected a measure of regional income per capita based on tax data for the adult population. Personal income in the tax statistic includes wage income, social security and personal capital income. Income from firms is hard to locate at this level of disaggregation, and regional GDP measures are not available. The data cover all years during the period 1972-2008 and it follows that we have 3,293 observations of per capita incomes. We measure the level of education in each region as the share of the adult population (16 years and older) with tertiary education, including both short higher education (college level, up to 4 years in duration) and long higher education (university level, more than 4 years in duration). The data cover the single year 1970 and all years during the period 1980-2008. In the analysis the

---

1 Data source: Statistics Norway, Table 06983 (Persons 16 years and older by level of education).
income and education levels are measured relative to the average levels of income and education across all regions in each year.

On average across the 89 regions income per capita grows by 2.7% annually, and the share of the adult population with tertiary education increases from 6% to 21% during the past four decades. But there are significant differences across regions both with respect to income and education levels. At the beginning of the time period studied relative incomes vary from 0.65 to 1.69, while the relative level of education varies from 40% of the average to four times higher than the average. The differences across regions are decreasing over time. In 2008, the relative income levels lie in the range 0.85-1.52, while the share of the adult population with tertiary education varies from 60% of the average to twice the average. The raw data indicates a common convergence of both income and education levels.

We apply distribution analysis to capture heterogeneous processes with different growth paths from different starting points. There is a large literature applying Markov chain transition probability matrices to study income convergence across regions and countries. Quah (1993a, 1993b, 2001) developed this methodology, more recently applied and extended by Kremer et al. (2001). The basics of the method are presented by Shorrocks (1978). The discussion below relates to the distribution of income, but we also apply the method to investigate the distribution of education levels across regions.

The whole range of relative per capita income is divided into a finite number of N mutually exclusive income groups and in this analysis we follow the convention of working with five groups (N=5). For each region we get a sequence of variables describing the income group of that region at time \( t \). The sequences are considered as independent realizations of a single homogeneous Markov chain with finite group space N. The assumption of a finite first order Markov chain implies that the probability of being in a specific income group at time \( t \) only depends on the group of the previous period (and not earlier periods). The probability of moving from group \( i \) to group \( j \) from period \( t-1 \) to period \( t \) is described by \( p_{ij}(t) \). The probability is estimated based on observations of how regions move between income groups over time. The number of regions moving from group \( i \) to group \( j \) from period \( t-1 \) to \( t \) is measured by \( n_{ij}(t) \). The total number of regions moving from group \( i \) from period \( t-1 \) to \( t \) is measured by \( n_i(t-1) = \sum_j n_{ij}(t) \). The Markov chain can be reduced to a product of five mutually
independent multinomial distributions (one for each row $i$ of the transition matrix). For each
time period $t$, the distribution function is:

$$ f(n_y(t)) = \prod_{i=1}^{5} f_i(n_y(t)) = \prod_{i=1}^{5} \left( \frac{n_y(t-1)!}{n_y(t)!} \prod_{j=1}^{5} P_{ij}^{n_y(t)} \right) $$

(1)

The transition probabilities can be estimated by maximizing the log likelihood of the $T$
multinomials above with respect to $p_{ij}$:

$$ f(n_y) = \prod_{t=1}^{T} f(n_y(t)) $$

(2)

Given the constraint that the sum of $p_{ij}$ over all $j$ is 1, the maximum likelihood estimator is
simply the relative frequency of transitions:

$$ \hat{p}_{ij} = \frac{n_{ij}}{n_i} = \frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{t=1}^{T} n_i(t-1)} $$

(3)

where $n_{ij}$ and $n_i$ are the sums of the observed frequencies over all transition periods.

Given the initial distribution of regional income per capita across income groups,
$h(0) = [h_1(0), h_2(0), h_3(0), h_4(0), h_5(0)]$ where $\sum_i h_i(0) = 1$, the distribution after the first
transition period can be calculated as $h(1) = h(0)\Omega$, where $\Omega$ is the estimated $5\times5$ Markov
transition matrix. And similar, the distribution after $k$ transition periods follows as
$h(k) = h(0)\Omega^k$. Given that the matrix is regular, the distribution converges to the limiting
distribution $h^* = \lim_{k \to \infty} h(0)\Omega^k$, which is independent of the initial distribution. This is the
ergodic long-run distribution of regional income levels and is estimated based on the Markov
chain matrix under the assumption that the transition dynamics remain unchanged.

Pearson and Likelihood Ratio tests are used in the literature to investigate time stationarity
(Geppert and Stephan, 2008; Essletzbichler and Kadokawa, 2010). The test divides the entire
sample period into $M$ mutually exclusive and exhaustive subperiods and compares the
transition matrices under each of the $M$ subsamples to the entire sample. The estimators are
obtained based on the distribution function above, and Bickenbach and Bode (2003, p. 369) show that the test statistics have an asymptotic $\chi^2$ distribution. We suggest that the same test design can be used to study co-movements in the distributions of income and education. Instead of looking at different time periods we study different education experiences.

The test investigates whether the income transition probabilities are related to the movement in the education distribution. The entire sample of regions is divided into $M$ mutually exclusive and exhaustive subsamples according to the change in the relative level of education during 1970-2008. The transition matrices under each of the $M$ subsamples are then compared to the entire sample. The following Pearson ($Q$) and Likelihood Ratio ($LR$) test statistics have an asymptotic $\chi^2$ distribution with degrees of freedom equal to the number of independent pairwise comparisons:

$$Q = \sum_{m=1}^{M} \sum_{i=1}^{S} \sum_{j \in A_i} n_{ij|m} \left( \hat{p}_{ij|m} - \hat{p}_{ij} \right)^2 \sim \chi^2 \left( \sum_{i=1}^{S} (a_i - 1)(b_i - 1) \right)$$

$$LR = 2 \sum_{m=1}^{M} \sum_{i=1}^{S} \sum_{j \in A_i} n_{ij|m} \ln \left( \frac{\hat{p}_{ij|m}}{\hat{p}_{ij}} \right) \sim \chi^2 \left( \sum_{i=1}^{S} (a_i - 1)(b_i - 1) \right)$$

$A_i$ is the set of nonzero transition probabilities in the $i$th row of the transition matrix estimated from the entire sample, while $A_{i|m}$ is the set of nonzero transition probabilities in the $i$th row of the matrix estimated from the $m$th subsample. The total number of transitions from group $i$ in subsample $m$ and the total number of transitions from group $i$ to group $j$ in subsample $m$ are given by $n_{ij|m}$ and $n_{ij|m}$, respectively. The degrees of freedom are given in the last parenthesis, where $a_i$ is the number of elements in $A_i$ and $b_i$ is the number of subsamples with a positive number of observations in the $i$th row.

3. Income level convergence

A large literature has analyzed income convergence between regions and countries. The estimation of structural convergence equations derived from growth models using various national and regional samples broadly support income convergence. But the convergence is heterogeneous, often conditional on other determinants of income, and sometimes restricted to convergence clubs with similar initial conditions. And the convergence is slow, often with a
rate of convergence about 2% per year, which implies that the half life of the convergence process is about 35 years. An influential summary of recent econometric analyses of panel data is given by de la Fuente (2002), including his own analysis of Spanish regions. He concludes that panel analyses of regional data indicate income convergence and with a higher convergence rate than the early static analyses. The use of distribution analysis came as a reaction to the shortcomings of the regression literature and they have given more support to income divergence. Magrini (2009) and Sakamoto and Islam (2008) argue that income convergence understood as reduction in the dispersion of income across regions is best analyzed by distribution analysis. Islam (2003) offers an overview of the concepts involved.

To examine how the distribution of regional income per capita develops over time, we compare the estimated kernel density functions for the first year 1972 and the last year 2008, as shown in Fig. 1. The horizontal axis represents income per capita relative to the average level across regions, while the vertical axis gives the density of regions at different relative income levels. The estimated distributions show a clear pattern of convergence over time. Both functions have a single-peak distribution with the majority of regions located close to the average level of income per capita, but the distribution is narrower and more concentrated around the peak in 2008 than in 1972.

In the estimation of Markov chains we report 4-year transitions below. The pattern is the same for annual transitions, but 4-year intervals avoid short term fluctuations and have more stable transition paths. We follow the convention of discretization based on a uniform initial distribution of relative incomes across income groups, which gives the following five groups: 1) less than 91% of the average, 2) between 91% and 96%, 3) between 96% and 101%, 4) between 101% and 107%, and 5) more than 107% of the average income across regions. The transition probability matrix is shown in Table 1, and is estimated based on nine 4-year transitions and a total of 801 observations. Most of the estimated transition probabilities are significant. As seen from the binomial standard errors given in parentheses, all the diagonal

2 The density estimates are calculated using a Gaussian kernel with bandwidth set according to Silverman’s rule of thumb; $1.06\sigma B^{-1/2}$, where $\sigma$ is the standard deviation of the data and $B$ is the number of observations. This gives bandwidth equal to 0.075 and 0.0474 for 1972 and 2008, respectively.
and immediately off-diagonal transition probabilities are significant, while the estimates of the probability of moving two income groups during a 4-year period are typically insignificant.

Table 1 about here

The Markov matrix shows income convergence across regions. The distribution of per capita incomes is tending towards a point mass, rather than towards a two-point distribution. Regions in the lowest income group have 24% probability of moving up during a 4-year period, and the high income regions have 19% chance of moving down the distribution. Regions in income groups 2 and 4 have higher probability of moving towards the middle of the distribution than towards the respective ends. This pattern is confirmed by the implied ergodic (long-run) distribution given in the last row of the matrix. Regional incomes go from a uniform distribution initially to a normal distribution in the long run. The peak is located around the average income level, with income groups 2-4 (relative income between 91% and 107% of the average), accounting for about 70% of the regions in the long run. The lowest and the highest income groups are reduced from 20% initially to about 15%. The distribution tends to accumulate in the middle, combined with thinning of both the lower and the higher tail, consistent with income convergence. The stationarity of the distribution is addressed by looking at the second eigenvalue of the transition matrix, which is found to equal 0.89. It follows that the distribution converges to a steady state, and the asymptotic half life of the process (the time it takes to reach half way to the long run distribution) equals 5.9 transition periods, which corresponds to about 24 years.

We investigate the time stationarity (constant transition probabilities over time) of the income transition matrix by applying the Pearson ($Q$) and the Likelihood Ratio ($LR$) tests. The transition probabilities are estimated based on nine 4-year transitions during 1972-2008. The sample period is divided into three subperiods, each containing three 4-year transitions. The transition matrices for each subperiod are then compared to the full period matrix (given in Table 1). With 28 degrees of freedom the critical value at 5% significance level equals 41.3. The test statistics are calculated to $Q = 39.4$ (prob = 0.07) and $LR = 40.1$ (prob = 0.06), which means that the null hypothesis of constant transition probabilities over time cannot be rejected. The contributions to the Pearson test statistic from each transition in the three
subperiods show that the differences in transition probabilities are minor. Transition probabilities seem to be constant over time.3

4. Education level convergence

The evidence about the development of the regional distribution of education is mixed. Recent studies of US regions find that human capital measured by population shares with college education has diverged during the last decades (Wheeler, 2006, Hammond and Thompson, 2010). A German study looking at shares of high-skilled workers concludes that the regional skill composition has converged over time (Sudekum, 2008).4 But there are methodological issues involved in the interpretation of results. In the US studies, the dynamic process is considered in terms of absolute changes in college/skill shares. Education divergence comes out because highly educated cities increase their college/skill shares more in percentage points than peripheral regions with low education levels. In our dataset the cities have higher increases in shares with high education in percentage points, but peripheral regions have higher growth rate of the tertiary share. The analysis of the education convergence below looks at relative education levels, similar to Sudekum (2008).

The development of the distribution of the education level in the regions is first described by estimated kernel density functions for the first year 1970 and the last year 2008, as shown in Fig. 2. The horizontal axis represents the share of the adult population with tertiary education relative to the average share across regions, while the vertical axis gives the density of regions at different relative education levels.5 Both distributions have a single-peak around the average educational level, but over time, the distribution becomes narrower and the peak more

---

3 We have also divided the sample period into two subperiods, and the null hypothesis of time stationarity is still not rejected.

4 Different results regarding education convergence in different countries may be explained by timing. When institutions of higher education are established, we expect an initial period of education divergence followed by convergence with equalization of education levels across the country. Further urbanization may lead to education divergence again with concentration of the highly educated in smart cities.

5 Consistent with Silverman’s rule of thumb the bandwidth is set to 0.1914 and 0.1014 for 1970 and 2008, respectively.
pronounced, indicating convergence with respect to the level of education. Compared to the
distribution of income levels, the variations in the level of education are larger across regions.

Fig. 2 about here

The robustness of this result is investigated by estimating the kernel using absolute shares of
tertiary educated, similar to Hammond and Thompson (2010). The log shares are presented in
Fig. 3 and they show the positive shifts in the education share since 1970 and that the
distribution has narrowed over time. The standard deviation of the log distribution is lower in
2008 than in 1970. We have calculated the coefficient of variation of the absolute education
shares and it is also falling over time. Education divergence is inconsistent with our data,
different from the US studies.

Fig. 3 about here

Furthermore, we investigate the education dynamics through a Markov chain transition matrix
for the period 1970-2008. Since education levels change slowly we focus on decade
transitions and 89 regions imply that transition probabilities are estimated based on 356
observations. Again, we define the range of education groups so that the number of
observations is similar across groups. This gives the following five groups: 1) less than 79%
of the average educational level, 2) between 79% and 90%, 3) between 90% and 100%, 4)
between 100% and 111%, and 5) more than 111% of the average level of education. The
Markov matrix with respect to the level of education is given in Table 2. As seen from the
binomial standard errors given in parentheses, the estimated transition probabilities are mostly
significant.

The transition matrix is consistent with the findings from the kernel functions with
convergence in educational levels across regions. Regions located in the lowest education
group have 25% chance of moving up the distribution during a 10-year period. Regions in
education group 2 are more likely to move towards the middle of the distribution than towards
the lower end. Regions in the highest group have good chances of remaining in this group
(89%). These dynamics imply a movement towards the top end of the distribution, and
education group 5 dominates the long-run distribution with more than 30% of the regions
(given in the last row of the matrix). The distribution of educational levels goes from a uniform distribution initially towards a single-peaked distribution, consistent with convergence in education levels, and shows no tendencies of a bimodal twin peaked distribution. But the transition towards the long run distribution is slow, and based on the second eigenvalue of the matrix it takes more than 80 years to reach half way to the steady state. The peak at the top end of the distribution means that regions that have reached the highest education group are not likely to leave it again. Kremer et al. (2001) document similar dynamics for the world income distribution.

Table 2 about here

5. Test of co-movement in the distributions of income and education

The data analyzed above show large regional heterogeneity with respect to education and income. The expanding education level in peripheral regions is consistent with convergence in the level of income. In this section we investigate the relationship between income transitions and the changes in the relative education levels. Is there a systematic pattern of rising educational attainment in regions moving upward in the income distribution?

Direct observation of the distribution across regions in the early 1970s confirms the expected relationship between education level and income level. Small regions in the periphery have low income and education level, while the large cities have a high share of the adult population with tertiary education and high income level. But the correlation between education level and income level is decreasing over time (raw correlation coefficient 0.80 falling to 0.71), and the correlation coefficient between change in education and change in income is only 0.43. The low correlations indicate that change in education level has not been of much importance for the income transitions.

We investigate co-movements in the distributions of income and education by calculating income transition probability matrices conditioned on the movement in the distribution of education. We rank the 89 regions according to the change in their relative level of education during 1970-2008, and divide the sample into two equal subsamples (top 50% and bottom 50%). Among the top 45 regions, the relative level of education on average increased by 0.13 (from 0.79 to 0.92). This subsample reflects regions with below average level of education
that is gradually moving up in the education distribution. In the other end, the 44 regions with the largest decrease in relative education are dominated by regions with above average level of education that are gradually moving towards the middle of the education distribution (decreasing by 0.13 from 1.21 to 1.08, on average). The estimated Markov matrices for the two subsamples of regions are shown in panels a and b of Table 3.

Table 3 about here

The broad picture is that the development in the relative level of education does not matter much for income transition probabilities. If increased education level is important for upward income transitions, we expect the numbers above the diagonal to be higher in the matrix in panel a than in panel b. But the transition probabilities are roughly similar. Whether a region moves up or down in the distribution of relative educational levels does not affect its chances of catching up with respect to income. When it comes to the lowest income group, almost all the regions belong to the subsample with large increase in relative education. A closer look at the data shows that the average increase in the relative level of education is actually higher among regions that remain in the lowest income group than in regions that are able to catch-up (0.16 vs. 0.12). Rising educational attainment is observed in regions both catching up and falling behind, and is not associated with the income convergence seen in the data. The transition probabilities in the upper end of the distribution are consistent with this view. Income growth has not taken off in high income regions with rising relative education level. And for regions in the subsample with decreasing relative level of education, the decrease is much larger among regions that remain in the top income group than in regions moving down the distribution.

To statistically test for the importance of changes in the relative educational level for the convergence process, we apply Pearson and Likelihood Ratio tests, as explained in section 2. Comparing the two matrices in Table 3 to the matrix for the entire sample of 89 regions (given in Table 1) we get a test statistic of about 22 for both tests. With 13 degrees of

---

6 When comparing the matrices, we exclude transitions with only one observation in the full sample matrix if the single observation is located in a subsample with less than 40 observations in the same income group (giving an unreasonably high probability and large error term). The degrees of freedom are adjusted accordingly. In this case, the transition from group 5 to group 3 is excluded.
freedom, the 5% critical value equals 22.4. The null hypothesis of equal transition probabilities across different developments in relative education cannot be rejected. The contributions to the Pearson test statistic from each transition in the two subsamples are given in Table 4. Differences in transition probabilities are minor, and most error terms are well below 0.5. The relatively large test statistic comes from the inaccurate estimates of transition probabilities in the lowest income group in panel b of Table 3, which is based on only 7 observations. To sum up, rising educational attainment is not related to income catch-up of low income regions, and income growth has not taken off in high income regions with increasing education level.\(^7\)

Table 4 about here

In section 4 we referred to the difference between growth rates of education shares and changes of education shares. Instead of focusing on the movement in the distribution of education (the development in a region’s level of education relative to the average) we consider the change in the absolute level of education, measured as the percentage point increase in the tertiary education share of the adult population during 1970-2008. Given this classification of regions, we perform the same analyses and tests as above, and our main results remain. The null hypothesis of equal transition probabilities across subsamples of regions cannot be rejected. Rising educational attainment is not related to the income convergence movements.

As seen from Table 3, the probability of moving down from the second income group to the lowest group is in fact higher for regions that are catching up with respect to education. Similarly, the probability of remaining in the top income group is highest among regions with large decreases in the relative level of education. These counterintuitive findings might indicate an importance for the level of education in income transitions. The regions moving up the education distribution typically start from below average level of education, while regions moving down the distribution start with above average level of education. We

\(^7\) As a robustness check, we divide the regions into three (rather than two) subsamples according to the change in the relative level of education, and the conclusion still holds. Furthermore, we have done the same analysis at the municipal level. The findings of education and income convergence are confirmed, and the broad picture is that the movements in the income and education distributions are unrelated.
investigate the role of the education stock, but find only a weak relationship between education level and movements in the income distribution. Based on the Pearson and Likelihood Ratio tests, the null hypothesis of equal transition probabilities across subsamples with different levels of education is not rejected. A more detailed analysis of the top 20 ‘smart cities’ confirms that their income advantage is in decline. This is not surprising given the evidence of income convergence among Norwegian regions (section 3) together with high correlation between income and education.

6. Concluding remarks

Recent research addresses the divergence of education levels across regions and the relationship to income convergence. We challenge the suggested relationship between movements in the distributions of income and education using Norwegian data. We use Kernel density functions and Markov chains, and offer a test of co-movements in the distributions of education and income. Education levels converge as the education level is equalized across the country, and this process coincides with income convergence. But further investigation of the relationship between education convergence and income convergence shows that transitions in the income and education distributions are basically unrelated.

It is a puzzle that income levels have been catching up in the periphery during a period of stagnation and even contraction of economic activity. And income levels in ‘smart cities’ with concentration of highly educated have not taken off. The background understanding is that the education level increases in large cities with limited income growth and the income growth is strong in regions with fairly low education level. We cannot rule out that the results follow from special characteristics of Norway with an equal distribution of income and resource-oriented periphery. It is of interest to study whether this pattern is similar in other countries, and in particular our results raise questions about the limited income growth response to concentration of human capital in cities.
References

Berry, C. and E. Glaeser (2005), The divergence of human capital levels across cities, Papers in Regional Science 84, 407-444.


Gennaioli, N., R. La Porta, F. Lopez-de-Silanes and A. Shleifer (2013), Human capital and regional development, Quarterly Journal of Economics 128, 1, 105-164.


Glaeser, E.L. and M.G. Resseger (2010), The complementarity between cities and skills, Journal of Regional Science 50, 1, 221-244.


Magrini, S. (2009), Why should we analyze convergence using the dynamic distribution approach, Scienze Regionali 8, 1, 5-34.


Figure 1: Kernel density estimates, relative income per capita, 89 regions, 1972 and 2008.

Kernel density estimates, relative income, 1972 and 2008

Density of regions
0 1 2 3 4
Income per capita relative to the average across regions
0 .5 1 1.5 2

1972
2008

Figure 2: Kernel density estimates, relative level of education, 89 regions, 1970 and 2008.

Kernel density estimates, relative education, 1970 and 2008

Density of regions
0 1 1.5 2 2.5 3
Level of education relative to the average across regions
0 0 1 2 3 4

1970
2008
Figure 3: Kernel density estimates, level of education (log), 89 regions, 1970-2008.
Table 1: Markov chain transition probability matrix, income per capita, 4-year transitions, 1972-2008, 801 observations (binomial standard errors in parentheses).

| Income groups | 1 \text{\leq 0.91} & 2 \text{\leq 0.96} & 3 \text{\leq 1.01} & 4 \text{\leq 1.07} & 5 > 1.07 | Obs. |
|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 1             | 76.3 (3.4)      | 23.1 (3.3)      | 0.6 (0.6)       |                 |                 | 160             |
| 2             | 15.6 (2.9)      | 61.3 (3.9)      | 21.3 (3.2)      | 1.9 (1.1)       |                 | 160             |
| 3             | 0.6 (0.6)       | 19.4 (3.1)      | 56.3 (3.9)      | 21.1 (3.2)      | 0.6 (0.6)       | 160             |
| 4             | 1.3 (0.9)       | 24.4 (3.4)      | 63.1 (3.8)      | 11.3 (2.5)      |                 | 160             |
| 5             | 0.6 (0.6)       | 18.0 (3.0)      | 81.4 (3.1)      |                 |                 | 161             |

Initial distribution 20.0
Ergodic distribution 15.0

Table 2: Markov chain transition probability matrix, level of education, decade transitions, 1970-2008, 356 observations (binomial standard errors in parentheses).

| Education groups | 1 \text{\leq 0.79} & 2 \text{\leq 0.9} & 3 \text{\leq 1.0} & 4 \text{\leq 1.11} & 5 > 1.11 | Obs. |
|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 1                | 74.7 (5.2)      | 25.3 (5.2)      |                 |                 |                 | 71              |
| 2                | 5.6 (2.7)       | 73.2 (5.3)      | 19.7 (4.7)      | 1.4 (1.4)       |                 | 71              |
| 3                | 13.9 (4.1)      | 73.6 (5.2)      | 12.5 (3.9)      |                 |                 | 72              |
| 4                | 14.1 (4.1)      | 70.4 (5.4)      | 15.5 (4.3)      |                 |                 | 71              |
| 5                | 11.3 (3.8)      | 88.7 (3.8)      |                 |                 |                 | 71              |

Initial distribution 20.0
Ergodic distribution 3.6
Table 3: Markov chain transition probability matrix, income per capita, 4-year transitions, conditioning on the change in the relative educational level during 1970-2008 (binomial standard errors in parentheses).

Panel a: Top 50% with large increase in the relative educational level (405 observations)

<table>
<thead>
<tr>
<th>Income groups</th>
<th>1 ≤ 0.91</th>
<th>2 ≤ 0.96</th>
<th>3 ≤ 1.01</th>
<th>4 ≤ 1.07</th>
<th>5 &gt; 1.07</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>78.4</td>
<td>20.9</td>
<td>0.7</td>
<td></td>
<td></td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>(3.3)</td>
<td>(3.3)</td>
<td>(0.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>21.9</td>
<td>57.3</td>
<td>18.7</td>
<td>2.1</td>
<td></td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>(4.2)</td>
<td>(5.0)</td>
<td>(4.0)</td>
<td>(1.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.4</td>
<td>21.9</td>
<td>54.8</td>
<td>20.5</td>
<td>1.4</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>(1.4)</td>
<td>(4.8)</td>
<td>(5.8)</td>
<td>(4.7)</td>
<td>(1.4)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.7</td>
<td>27.1</td>
<td>59.3</td>
<td>11.9</td>
<td></td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>(1.7)</td>
<td>(5.8)</td>
<td>(6.4)</td>
<td>(4.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4.2</td>
<td>25.0</td>
<td>70.8</td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>(4.1)</td>
<td>(8.8)</td>
<td>(9.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Initial distribution 37.8 23.7 18.0 14.6 5.9
Ergodic distribution 27.0 25.3 22.1 17.5 8.1

Panel b: Bottom 50% with large decrease in the relative educational level (396 observations)

<table>
<thead>
<tr>
<th>Income groups</th>
<th>1 ≤ 0.91</th>
<th>2 ≤ 0.96</th>
<th>3 ≤ 1.01</th>
<th>4 ≤ 1.07</th>
<th>5 &gt; 1.07</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28.6</td>
<td>71.4</td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>(17.1)</td>
<td>(17.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6.2</td>
<td>67.2</td>
<td>25.0</td>
<td>1.6</td>
<td></td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>(3.0)</td>
<td>(5.9)</td>
<td>(5.4)</td>
<td>(1.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>17.2</td>
<td>57.5</td>
<td>25.3</td>
<td></td>
<td></td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>(4.0)</td>
<td>(5.3)</td>
<td>(4.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.0</td>
<td>22.8</td>
<td>65.3</td>
<td>10.9</td>
<td></td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>(1.0)</td>
<td>(4.2)</td>
<td>(4.7)</td>
<td>(3.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>16.8</td>
<td></td>
<td>83.2</td>
<td></td>
<td></td>
<td>137</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.2)</td>
<td></td>
<td>(3.2)</td>
<td></td>
</tr>
</tbody>
</table>

Initial distribution 1.8 16.2 22.0 25.5 34.6
Ergodic distribution 1.7 19.3 27.9 31.0 20.1
Table 4: Test of co-movements in the distributions of income and education. Contributions of single subsamples to the Pearson test statistics.

<table>
<thead>
<tr>
<th>Change in relative educational level</th>
<th>Income groups</th>
<th>Number of obs.</th>
<th>1 ≤ 0.91</th>
<th>2 ≤ 0.96</th>
<th>3 ≤ 1.01</th>
<th>4 ≤ 1.07</th>
<th>5 &gt; 1.07</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 50% large increase (405 obs)</td>
<td>1</td>
<td>153</td>
<td>0.10</td>
<td>0.32</td>
<td>0.00</td>
<td></td>
<td></td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>96</td>
<td>2.40</td>
<td>0.25</td>
<td>0.28</td>
<td>0.02</td>
<td></td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>73</td>
<td>0.63</td>
<td>0.24</td>
<td>0.03</td>
<td>0.21</td>
<td>0.63</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>59</td>
<td>0.09</td>
<td>0.18</td>
<td>0.14</td>
<td>0.02</td>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>24</td>
<td></td>
<td>0.62</td>
<td>0.31</td>
<td></td>
<td></td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>6.47</strong></td>
</tr>
<tr>
<td>Bottom 50% large decrease (396 obs)</td>
<td>1</td>
<td>7</td>
<td>2.09</td>
<td>7.06</td>
<td>0.04</td>
<td></td>
<td></td>
<td>9.19</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>64</td>
<td>3.60</td>
<td>0.37</td>
<td>0.42</td>
<td>0.03</td>
<td></td>
<td>4.42</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>87</td>
<td>0.55</td>
<td>0.21</td>
<td>0.02</td>
<td>0.18</td>
<td>0.55</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>101</td>
<td></td>
<td>0.05</td>
<td>0.11</td>
<td>0.08</td>
<td>0.01</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>137</td>
<td></td>
<td></td>
<td>0.11</td>
<td>0.06</td>
<td></td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>15.54</strong></td>
</tr>
</tbody>
</table>

Pearson test statistic **22.0**
Critical value at 5% significance level (13 degrees of freedom) **22.4**