Earnings Mobility and Origin Dependence:

What can twins say together with nonparametric econometrics?

William Nilsson*

Department of Economics, Umeå University, Sweden and
Centre for Research in Welfare Economics (CREB), Barcelona, Spain

Abstract: This study focuses on earnings immobility and its relation to the origin in terms of both the social background and the initial position in the earnings distribution. Twin data is used to reveal the importance for the common background for immobility. A nonparametric technique is used to study if the immobility varies over the distribution. The results indicate strong immobility, an important effect of the background, and that these effects vary over the distribution. For the male monozygotic sample, the social background accounts for 71-88 percent of the immobility in deciles 3 to 7, where the background is found to be most important. The common background has its strongest impact in deciles 6 to 10 for the female sample, where these effects accounts for 66-77 percent of the immobility. Comparing results for monozygotic and dizygotic twins also indicate that genes play an important role in income immobility.

JEL: C14, D63, I31

Keywords: earnings mobility; nonparametric; twins

Umeå Economic Studies 716, 2007

---

* I wish to thank seminar participants at ECINEQ 2007 held in Berlin. Financial support from Centre Cournot (Robert Solow Fellowship) is gratefully acknowledged. All data, except for the identification of the twin sample, came from Statistics Sweden (SCB). The twin information comes from the Swedish Twin Registry (STR). The Swedish Twin Registry is supported by grants from the Swedish Research Council. To contact the author, write to william.nilsson@econ.umu.se.

1 Introduction

The focus of the study is to empirically analyze the importance of two different dependencies on the origin for earnings immobility in the society. First, how important is the origin in terms of social background and genetics for earnings immobility? Secondly, how important is the origin in terms of the initial position in the earnings distribution for immobility? If, for example, high earnings in one year often are followed by high earnings the next year, is this due to the unobserved family background? Or, is it common that once a high position is achieved, the probability is high to maintain this position, regardless of background factors? Both the degree of immobility and the importance of background factors can differ depending on where the individual is in the earnings/income distribution. Dickens (2000) and Jarvis & Jenkins (1998) found, for example, that mobility varies over the distribution with British data. The purposes of this study are to investigate how important the background is for immobility and if immobility and the importance of the background are different over the earnings distribution.

The empirical analysis is made using data on Swedish twins. With information on twins it is possible to identify the origin dependence based on the background. The position in the distribution for one twin is assumed to not, in itself, affect the twin siblings' position the following period. Any correlation than may be found is, accordingly, due to the similarity between the (identical) twins that has its origin in a very common social background and identical genes. The foundation of the idea is the well-established method of using sibling correlation to investigate the importance of shared background factors for a certain outcome (Solon, 1999). The assumption is that, if shared factors are important, the siblings will show a strong resemblance in the outcome. In this study the correlation between the earnings for twins in period t, and their twin siblings' earnings in period t - 1 is compared to the immobility for twins between period t - 1 and period t. If the correlation between the twin siblings is close to the correlation of earnings in different periods for the same twins, the immobility is largely due to family specific heterogeneity. If the difference between the measures is large, serial correlation of earnings, due to, for example, the labor market situation is the main reason for immobility. This is a new and innovative idea to measure how much of a measure of immobility that is explained by family specific heterogeneity. The idea is first implement for Spearman rank correlation, and is later extended to allow for differences over the earnings distribution. A nonparametric estimation technique is used to allow for different immobility and different relative importance of the background over the income distribution. The idea is to compare how an improved rank in period t - 1 would affect log earnings in t - 1 and t, allowing for different effects at different initial ranks. The relation between the rank and income in the same period captures income disparities. Comparing that relation with the connection between rank and earnings for a later period captures earnings mobility. If the benefit in a later period for an increased rank in t - 1 is similar to the effect of an improved rank in the earnings distribution in t - 1 this would indicate a very immobile situation. Differences between the two effects are expected due to both positional movements and if income grows at different rate in different parts of the distribution. These explanations are separated in the empirical analysis.

This paper contributes to the literature in the following ways: First, a twin method is developed and applied to identify the effect of the common background as explanation for immobility. Secondly, a nonparametric technique is used to allow for different immobility over the distribution. Further, the nonparametric technique also captures if the importance of the background varies over the distribution. Another advantage with the method, apart from allowing different effects over the distribution, is that it has not an inherited effect depending on the disparity of the initial distribution as other immobility measures based on re-ranking.

The method on how twin data can be used in mobility analysis and the nonparametric technique are described in section 2. The data is explained in section 3 and the results are presented in section 4. Concluding remarks are offered in section 5.

2 Method

The literature of social mobility contains a large amount of different measures of mobility or immobility. Spearman rank correlation coefficient is used to illustrate how twin data can reveal the permanent part of immobility. The method is, of course, not only applicable to that measure and it is later applied to a more flexible measure, in which it is possible to distinguish different effects depending on the initial position in the distribution.

A simple, but useful measure of immobility is Spearman rank correlation;

$$\rho_{SPEARMAN} = \frac{Cov(R_{sit-1}, R_{sit})}{\sqrt{Var(R_{sit-1})Var(R_{sit})}} \tag{1}$$

R_{sit} refers to the rank of income for twin s, (s = 1, 2), in twin pair i, (i = 1,…,N), at time t and R_{sit-1} the rank of income at time t - 1. Of course the measure can be calculated for time periods further apart. The subindex s takes the value 1 for the first twin and 2 for his/her twin sibling. Note that R_{sit} has N as a maximum and if s = 1 the rank refers to the position in the income distribution for the *first* set of twins. The reason to calculate the correlation of the rank, instead of income, is to reduce the importance of outlier (Atkinson *et al*. 1992, page 30). A $\rho_{SPEARMAN}$ close to one would indicate a very immobile society. To find out the part of immobility that can be explained by the common background, including the genes, $\rho_{SPEARMAN}$ is first calculated for s = 1. This measure is then compared to ρ_{TWIN}, where R_{1it-1} is replaced by R_{2it-1} in equation (1), while R_{1it} is kept as before. All shared characteristics, including genes, which matters for the rank will contribute to ρ_{TWIN}. The immobility that can be found from time t - 1 to time t is, however, likely to not only be based on the social background. $\rho_{SPEARMAN}$ is expected to be higher than ρ_{TWIN} as the first also includes immobility due to, for example, education, health status and the labor market situation. The share of immobility due to the background is finally calculated as $\rho_{TWIN} / \rho_{SPEARMAN}$.

This measure concerns mobility in terms of positional movements. Studies of mobility do, however, usually have a welfare motive and how the income develops could also be of interest. If someone loses in rank, it is still possible that his/her welfare is improved if his/her real income is higher than before. Beenstock (2004) distinguish between rank mobility and quantity mobility. He uses the Gini regression coefficient to measure quantity immobility and the results from Woden and Yitzhaki (2005) to decompose the measure in rank mobility, disparity and growth. A change in quantity mobility could, accordingly, be due to changes in rank mobility, disparity and growth.

Using the above mentioned measures summarize the mobility in the society with single measures and these do not take into account that the mobility could be different depending on the initial position in the income distribution. While twin data can be used to reveal the relative importance of the background for the immobility, these measures cannot take into account variation over the distribution. Since the importance of the background can vary over the income distribution it is desirable, to measure both mobility and the importance of the background at different places in the income distribution. In this study this is done with a local linear nonparametric regression. The idea is to study how an improved rank in the initial period would affect income in the same period and compare this to the effect on income in a later period. Consider that the rank is normalized to be between 0 and 100, and we are interested in how the income would improve if an individual improved the rank with 1 percentile. Improving the rank with 1, would improve the income the same year, but with how much will depend on the initial position in the distribution. An improved rank for someone in the top tail of the distribution would likely mean a substantially improved income. This is expected because the density in the distribution is lower and, hence, the difference between the incomes in each improved rank is important. The density is also low in the lower tail of the distribution, and an improved rank would also be quite beneficial in terms of higher income. The increased income from rank improvements are hereafter called the *distributional effect* as it depends on the initial position and the shape of the distribution.

Keeping this in mind, we can focus on how we expect the improved rank to manifest itself in terms of improved income in a later period. Now, the answer will also depend on the expected mobility in each position in the distribution. If the income is expected to grow very fast in a particular part of the distribution this will also matters for an ongoing effect of an improved rank. Depending on the purpose of the study, it could be desirable to separate these motives. This issue is discussed later.

The income in time t, is explained by the normalized rank in t-1;

$$\ln y_{sit} = m(r_{sit-1}) + \varepsilon \tag{2}$$

$m(r_{sit-1})$ is an unrestricted functional form and the estimator of the derivative, which is the main interest, is;

$$\gamma(r_{st-1}) = \frac{\sum_{i=1}^{n}(r_{sit-1})(\ln y_{sit})K((r_{sit-1} - r_{st-1})/h)}{\sum_{i=1}^{n}(r_{sit-1})^2 K((r_{sit-1} - r_{st-1})/h)} \tag{3}$$

A normal kernel is used with the bandwidth, h.[1] $\gamma(r_{st-1})$ is hereafter labeled *expected long-run rank effect*. The nonparametric estimation technique allows the effect to vary depending on the initial rank in the distribution. Knowing the expected extra income that an individual would gain in the next period has to be compared to the gain that would occur in the same period, i.e. the distributional effect. If we only are interested in immobility and not growth, the distributional effect has to be adjusted. If the growth is very high in a particular part of the distribution, moving away from that rank will also mean that another, less favorable, path of growth is adopted. A pro-poor growth is, for example, the case where the income for the lowest ranks grows relatively more than for the rest of the distribution. Note that, the traditional view of pro-poor growth is considered where different groups are compared for the different years. In Jenkins & Van Kerm (2006) pro-poor growth is analyzed individually, i.e. the income growth for the initially poor is studied. It is likely that regression towards the mean occur, i.e. "being unlucky twice is unlikely", and the income for the initially poor is, accordingly, likely to grow fast at the same time as the group of poor is partially changed. Van Kerm (2006) also shows this with a nonparametric technique for 10 European countries. These positional movements are re-ranking and something that should be captured as mobility in this study. To calculate the expected growth (without positional movements) the income in t for individual j positioned in rank 1 is deducted by income for individual i in t – 1 in rank 1. This is done for rank $1,\ldots,R_{sNt}$. To calculate the *growth adjusted distributional effect* each individual in t – 1 is assumed to get the income that the individual on the same rank in t has. The growth adjusted distributional effect is then identified by calculating;

[1] See appendix for a specification of the normal kernel. The bandwidth used is based on Scott' rule of thumb (Scott (1992), page 152), $\hat{h} = \hat{\sigma}(n)^{-1/5}$, where $\hat{\sigma}$ is the estimated standard deviation, n is the number of individuals.

$$\ln y_{st-1}^{adj} = m(r_{st-1}) + \varepsilon \tag{4}$$

The derivate illustrate the expected improvement of the log income from an advance in rank with one, and when the income would grow as the distribution evolves to the next period. A measure of immobility in each position in the distribution is illustrated if the results for the derivates of equation (2) are compared with those of equation (4). Consider the case where no re-ranking takes place. In that case $\ln y_{sit-1}^{adj} = \ln y_{sit}$ and the two results will coincide perfectly. If, on the other hand, an improved rank in t – 1 does not influence the income in t at all, $\gamma(r_{st-1})$ in equation (3) will be zero for all initial positions. If the mobility, in terms of re-ranking, is very low in a particular part of the distribution, the expected long-run rank effect will be closer to the growth adjusted distributional effect in that particular position. In other parts of the distribution, where re-ranking is common, an increased rank is no guarantee for an increased income in the future. The two results will, accordingly, be more distant apart. In these examples, the effect of growth was not considered to be interesting and the immobility would rather be measured without it. It is, nevertheless, interesting to comment the case if the distributional effect not is adjusted for growth. If income is measured in logarithmic form, the growth will only matter if it is different over the distribution. If the growth would be the same all over the distribution, $m(r_{sit-1})$ would shift upwards, but its shape would be conserved, and hence, its derivative would not change. If income is not used in logarithmic form, $m(r_{sit-1})$ would shift upwards with the percentage of uniform growth. In that case the increase in income would be different over the distribution and the shape would change. Accordingly, its derivative would also change, i.e. with the percentage of uniform growth.

If the growth is pro-poor, the adjusted distributional effect will be lower than the distributional effect for those individual with an initial low rank. The reason is that the distributional effect does not include that those on a low initial position are moving away from a favorable growth path. The expected long-run rank effect compared to the distributional effect, would indicate a lower immobility for that group compared to if the growth effect was deducted. The distinction made in Beenstock (2004) between rank mobility, disparity and growth is not applicable with the present method. This approach does not distinguish disparity and growth, as pro-poor (pro-rich) growth, in fact, means a reduced (increased) disparity.

An advantage of this method to illustrate immobility is that immobility is set into relation to the initial income distribution and its disparity. With measures of mobility based on re-ranking, mobility has been found to be higher in Germany than the US, while the opposite is found if mobility is measured with average income change (Jenkins & Van Kerm, 2006).[2] It is expected that mobility measures based on re-ranking tend to give societies with a higher dispersion a lower mobility. The higher dispersion makes each rank wider, and thus, a larger change in income is required to improve the rank. This also explains why "the data cloud tends to thicken at the top and bottom ends of the 45° line, implying that mobility diminishes at the top and bottom ends of the earnings distribution" when earnings rank in 1983 is plotted against rank in 1995 in Beenstock (2004). The same pattern was found for British data in Dickens (2000) and the author precisely comments that the result "may be expected, since the dispersion of wages is higher at these points of the distribution, particularly in the top decile". While re-ranking is an attractive concept to capture mobility, it is difficult to draw inference if it is not set in relation to an initial dispersion in the society. This is also why international comparison of mobility based on these measures is complicated. The measure in this study does not have an inherit effect of the disparity in the initial income distribution which makes the method easier to apply when interest is to compare mobility over time or mobility for difference countries.

To implement the twin method to identify the part of immobility that is due to the common background r_{2it-1} is included instead of r_{1it-1} in equation (2). $\gamma(r_{2t-1})$, hereafter labeled *expected long-run twin rank effect*, indicates how an improved rank for one twin would be accompanied with an improved ln(income) for his twin sibling depending on the rank for the first twin. Of course, the observed effect is not causal, but only due to the common social environment and genetics. The effect is allowed to be different depending on the position, as it is possible that an improved rank for someone in an

[2] Aaberge *et al*. (2002) compares inequality and income mobility in the US and the Scandinavian countries. They do not find any "*positive* relationship between inequality and mobility", when mobility is measured as inequality-reducing rank-changes.

already high position can be related differently to the social background compared to someone in the middle or lower part of the distribution. The share of the expected long-run twin rank effect of the expected long-run rank effect is twin heterogeneity, and explains the relative importance of the background for the immobility. Note that twin heterogeneity refers to characteristics that the twins *share* and how these differ compared to the rest of the population. Anything that makes twins more similar compared to the rest of the population will contribute to make the expected long-run twin rank effect different from zero and accordingly twin heterogeneity different from zero.

3 Data

The empirical analysis is done with twin data from the Swedish Twin Register linked with administrative taxation data. The population is Swedish twins born between 1949 and 1958. A reference sample consisting of a 10 percent random sample of the Swedish population within the same age range is also available. The twin sample is divided into subsamples of male and female monozygotic respective dizygotic twins. The reference sample is also split into a male and a female sample. Data on the immigrant population are excluded from the reference samples to make them more similar to the twin samples.

The samples of monozygotic twin include 774 male and 906 female twin pairs. Only same sex twins are included in the samples of dizygotic twins. 1257 male dizygotic twin pairs and 1226 female dizygotic twin pairs are available. The reference samples consist of 47 250 male and 45 670 female individuals. The income variable used for the analysis is earnings, incomes from self-employment, and social work related benefits. The latter includes unemployment insurance, sick leave benefits, parental leave benefits etc. Taxes are not deducted. The income is deflated to the price level of 2001. Sample statistics for 1999 and 1994 are summarized in Table 1.

[Table 1 about here]

The income variable is collected from taxation data and should give less problems of measurement error compared to survey data. The taxation data does, of course, not include incomes from the underground economy and all conclusions about immobility refer to taxable incomes. The male twin samples seem to have a slightly higher income than the reference sample. It is, however, important to remember that the inclusion of individuals in the twin samples is restricted to cases where both twins are alive and have positive earnings. It is, accordingly possible that at least some twins (with low income) were dropped due to a twin sibling that was not alive or did not have any incomes.

4 Results

This results section consists of calculation of Spearman rank correlation for the years 1994 to 1999 for both male and female monozygotic samples. Thereafter results from the nonparametric method are shown for monozygotic, dizygotic and reference samples for both male and female individuals.

[Table 2, about here]

Table 2 includes Spearman rank correlation for the same twin and also correlation between twin siblings. The upper half of the table, i.e. above the diagonal of 1.0, consists of results for the male sample, while the results for the female sample are included below the diagonal. The income immobility from one year to another is estimated to about 0.91-0.94 for the male sample and 0.87-0.92 for the female sample. The rank correlation from one year to another for twin siblings' incomes is about 0.53-0.59 for the male sample and 0.41-0.48 for the female sample. If the length of the period is extended so that incomes in 1994 and 1999 are compared the immobility is reduced to about 0.80-0.82 for the male sample and 0.71 for the female sample. The rank correlations for twin sibling incomes' for the same period are 0.51-0.56 and 0.39-0.40 for the male respective female sample. Accordingly, a quite important share, i.e. 64-69 percent for the male sample and 55-57 percent for the female sample, of the immobility is based on a common background for the twins. These numbers do not take into account that both the immobility and the common background can vary over the distribution.

The results from the nonparametric method are included in Figures 1 and 2 and also summarized in Table 3 and 4.[3] To take advantage of all observations both $\ln y_{1t} = m(r_{2t-1}) + \varepsilon$ and $\ln y_{2t} = m(r_{1t-1}) + \varepsilon$ are estimated. In the figures $\gamma(r_{1t-1})$ and $\gamma(r_{2t-1})$ are summarized with an average for each rank. Only the growth adjusted distributional effect, expected long-run rank effect and the expected long-run twin rank effect are included in the figures. Results from other relevant estimations, including confidence interval, are summarized in each decile and included in Tables 3 and 4. The figures are used to show the general pattern while the magnitude of the effects is discussed on basis of summarized measures for different deciles included in the tables. Figure 1 includes the effects for the male monozygotic and dizygotic samples.

[Figure 1, about here]

It is clear that both the long-run rank effect and the expected long-run twin rank effect are increasing in the tails of the distribution. As expected, the adjusted distributional effect also increases in the tails of the distribution. This captures the lower density, and thus, the higher benefits of an increased rank, due to the disparities in the initial distribution. For the monzygotic sample Figure 1 indicates that the immobility seems fairly high over the distribution with a peak at deciles 6 and 7 and lower for deciles 1 and 2. Heterogeneity seems to be more important between deciles 3 and 7, since the curves diverge at the tails of the distribution.

The general tendency concerning immobility is confirmed for the male dizygotic sample. The expected long-run twin rank effect is, however, more distant to the expected long-run rank effect all over the distribution. The expected long-run twin rank effect is, in fact, found to be very low from approximately deciles 2 to 7. The difference from the monozygotic sample indicates an important role for genetics in explaining immobility. So far, only results for the male samples are shown. It is, however, not necessarily that a female sample would show a similar pattern. Results for the female

[3] The bandwidth used was for all nonparametric regressions is $\hat{h} = 1.5\hat{\sigma}(n)^{-1/5}$, where $\hat{\sigma}$ is the estimated standard deviation and n is the number of individuals. The same qualitative conclusions are found for $\hat{h} = \hat{\sigma}(n)^{-1/5}$.

monozygotic and dizygotic samples are included in Figure 2.

[Figure 2, about here]

The immobility is found to be fairly similar to the male sample with the lowest immobility found for approximately decile 1 and 2. For the female monozygotic sample the effect of the common background is found to be particularly strong from decile 6 to 10. Figure 2 shows a very low effect of the common background for the female dizygotic sample. The expected long-run twin rank effect is found to be close to zero, and in fact below zero for the first decile. Again, the results indicate an important genetic effect.

The results from the figures are summarized in Table 3 and Table 4. The tables also include confidence intervals, calculation of measures of immobility and twin heterogeneity, as well as results for the reference samples.

[Table 3, about here]

Table 3 shows that for the male monozygotic sample, expected long-run twin rank effect is significantly different from zero all over the distribution. This contrasts with the results for the dizygotic sample, where the expected long-run twin rank effect is not significantly different from zero for deciles 1 to 4. The reference sample confirms that the immobility is very high from deciles 3 to 10. Immobility is much lower for deciles 1, and the reference sample suggests that the immobility is underestimated for the monozygotic and dizygotic sample for this group. This is, however, expected as the sample size is much larger for the reference case. When estimating the local linear least square regression the bandwidth will cover higher ranks, if the sample size is small. As the effect is U-shaped, this means that the effect is weighted down at the lower tail of the distribution. This also explains why the distributional effect and the growth adjusted distributional effects are much higher for percentile 1 and 2 for the reference sample.

The results indicate an important immobility that also varies over the distribution. Results for the monozygotic sample show that a large part of this immobility is based on the common background, including the genes. For deciles 3 to 7, 71-88 percent of the immobility is explained by the common background. These explanations are also important, although to a less extent, at the extremes of the distribution. For example, for deciles 8 to 10 the corresponding numbers are 62-66 percent.

[Table 4, about here]

The immobility is also found to be high for the female samples. As mentioned earlier, results for the female monozygotic sample show that the common background is especially important for deciles 6 to 10. Between 66 and 77 percent of the immobility is due to the common background for individuals in these deciles. The lower bound for the confidence interval for the female dizygotic sample shows that the expected long-run twin rank effect is not significantly different from zero in any part of the distribution. This only occurs for deciles 4 and 5 for the female monozygotic sample. Genetics seems to play an important role for immobility.

Heterogeneity is for the male monozygotic sample found to be more important at the middle part of the distribution. This is consistent with a hypothesis of true state dependence in the extreme part of the distribution. Being poor could, for example, in itself affect the probability of being poor in later periods through persistent unemployment and health problems. At the same time, once a high position is achieved this could also make future success more likely. An acquired talent could be beneficial for several years, and hence produce immobility in the distribution. The results for the female samples do, however, not follow this explanation. For the female sample heterogeneity is strong for the upper half of the distribution and the difference between the monozygotic and dizygotic samples indicate that genetics is important to explain the pattern. Why do genes seem to play such an important role for immobility at the upper part of the distribution for the female population? A hypothesis is that genetically inherited personality characteristics could be important for the position in the distribution for the female population. Is it possible that even in a country with a long history of female participation in the labor market, genetics could sort the female population into two different groups? One group would enter the labor market with an objective to compete and achieve an important income. This would be the high income earners where the background is important for the immobility. The other group would be satisfied to be a second income earner with more household responsibilities and for this group the background is less important to explain the immobility in the distribution. The reason that this pattern would not occur for the male population is that it is already the norm for the male population to compete on the labor market and genetics is not an important sorting mechanism. The method in the study is new and any explanation of the found patterns tends to be rather speculative and maybe even provocative. The hypothesis outlined is certainly not the only possible explanation of the pattern and it is important to see if the pattern is replicated in other societies and what actually is behind the pattern.

5 Concluding remarks

This study focuses on the origin in terms of both the social and biological background and the significance of the initial position in the income distribution. The results show that immobility varies over the distribution which suggests that summary measures as the Spearman correlation coefficient cannot give an appropriate illustration of the immobility patterns. The reasons for immobility are also found to vary over the distribution, i.e. the background has different relative effect depending on the position in the distribution. Monozygotic twins are found to resemblance more than dizygotic twins and this suggests a genetic explanation for immobility. Separating the data in male and female samples also indicate gender differences in the reasons for immobility measured over the distributions. Immobility is found to be particularly stronger for deciles 3 to 10 for the male samples. For the male monozygotic sample the strongest effect of the common background is found for deciles 3 to 7. Immobility for the female samples is also strongest for deciles 3 to 10. The common background, identified with the monozygotic sample, is, however, found to be relatively more important for deciles 6 to 10.

The introduced nonparametric method has two important advantages. First, the immobility is allowed to be different over the distribution. Secondly, other immobility measures based on re-ranking tend to give lower mobility where the dispersion is larger. Those measures would naturally indicate a lower mobility in the tails of the distribution, in particular in the top tail, where the dispersion is larger. For the immobility measure used in this study the initial dispersion in the distribution is a part of the measure and such natural effect is taken away.

The results in this study are based on individuals born between 1949 and 1958, with the income measured for 1994 – 1999, i.e. the individuals were between 36 and 50 years old. Since both younger and older individuals are not included, it is likely that the mobility is lower than for the overall Swedish population. The sample also excludes the immigrant population and repeating the study for immigrants, and applying a wider age-range, could give interesting results.

The article only separates background factors from a residual immobility due to investments in human capital, the labor market, demographic characteristics etc. These reasons could be studied in more detail. It is, however, crucial to take into account that changes in these variables are sometimes voluntary and failing to deal with endogeneity could give biased results and misleading conclusions (Atkinson, *et al.* 1992). For example, a change in household composition, such as a marital split, could affect mobility, but could also be a consequence of mobility. Other unobserved variables could also affect both household composition and mobility.

References

Aaberge, R. Björklund, A., Jäntti, M., Palme, M., Pedersen, P. J., Smith, N., & Wennemo, T.: Income inequality and Income Mobility in the Scandinavian Countries Compared to the United States. Review of Income and Wealth **48(4)**, 443-469 (2002)

Atkinson, A.B., Bourguignon, F., Morrison, C.: Empirical Studies of Earnings Mobility. In: Atkinson, A.B. (ed.). Fundamentals in Pure and Applied Economics, vol 52, pp. 1-149. Harwood Academic Publishers, Chur (1992)

Beenstock, M.: Rank and Quantity Mobility in the Empirical Dynamics of Inequality, Review of Income and Wealth **50(4)**, 519-541 (2004)

Dickens, R.: Caught in a Trap? Wage Mobility in Great Britain: 1975-1994, Economica **67**, 477-497 (2000)

Jarvis, S. & Jenkins, S.P.: How Much Income Mobility is there in Britain? The Economic Journal **108**, 428-443 (1998)

Jenkins, S.P. & Van Kerm, P.: Trends in income inequality, pro-poor income growth, and income mobility. Oxford Economic Papers **58**, 531-548 (2006)

Scott, D.W.: Multivariate Density Estimation. Wiley, New York (1992)

Solon, G.: Intergenerational Mobility in the Labor Market. In: Ashenfelter, O. and Card, D. (eds.) Handbook of Labor Economics, vol. 3, pp. 1761-1800. Elsevier, Amsterdam (1999)

Ullah, A. & Roy, N.: Nonparametric and Semiparametric Econometrics of Panel Data. In: Ullah, A. and Giles, D.E.A. (eds.) Handbook of Applied Economic Statistics, vol 155, pp. 579-603. Dekker, New York (1998)

Van Kerm, P.: Comparisons of income mobility profiles. IRISS Working Paper 2006-03, CEPS/INSTEAD, Differdange, Luxembourg.

Woden, Q. & Yitzhaki, S.: Growth and Convergence: A Social Welfare Framework. Review of Income and Wealth **51(3)**, 443-454 (2001)

Appendix

For the nonparametric estimations, a gaussian kernel is used;

$$K(r_{st-1}) = K\left(\frac{r_{sit-1} - r_{st-1}}{h}\right) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{r_{sit-1} - r_{st-1}}{h}\right)^2\right\}$$

The variances, $V\gamma(r_{st-1})$, that are used to estimated the confidence band for the nonparametric estimate are estimated as follows (Ullah & Roy, 1998).

$$V\,(\gamma(r_{st-1})\,|\,r_{sit-1}) =$$
$$(Z'(r_{st-1})K(r_{st-1})Z(r_{st-1}))^{-1}Z'(r_{st-1})\Omega(r_{st-1})Z(r_{st-1})(Z'(r_{st-1})K(r_{st-1})Z(r_{st-1}))^{-1}$$

where, $Z(r_{st-1})$ is a $n*2$ matrix $[1 \quad r_{sit-1} - r_{st-1}]$ and $\Omega(r_{st-1}) = K(r_{st-1})\Sigma K(r_{st-1})$, where Σ is a diagonal matrix with, $\tilde{\sigma}_{\varepsilon}^{2}\,(r_{st-1})$ estimated through local linear estimation.

$$(1 \quad 0)(Z'(r_{st-1})K(r_{st-1})Z(r_{st-1}))^{-1}Z'(r_{st-1})K(r_{st-1})\tilde{\varepsilon}$$

where $\tilde{\varepsilon}$ is a vector of local linear squared residuals, $\ln y_{sit} = r_{sit-1}\gamma(r_{st-1}) + \varepsilon_{sit}$. The confidence bands are estimated as +/- two standard errors from each estimated $\gamma(r_{st-1})$.

Table 1. Summary statistics for income variable

Sample	Year	Percentile 10	Percentile 30	Percentile 50	Percentile 70	Percentile 90	Mean	Standard deviation	N
Male									
Monozygotic	1994	135103	186019	213530	259288	361617	237126	110911	774
	1999	154801	220526	249527	297926	454610	283750	151826	774
Dizygotic	1994	133974	183283	213100	250805	341660	231259	123194	1257
	1999	156986	213181	248586	294386	432313	282099	175527	1257
Reference	1994	115305	178714	212277	252002	347104	228583	126953	47250
	1999	128436	207155	246702	296796	429375	272642	170329	47250
Female									
Monozygotic	1994	85427	130807	156413	179692	221072	158495	66244	906
	1999	111261	164255	193972	222221	283086	198858	77864	906
Dizygotic	1994	93255	130419	157511	180399	223051	159789	68692	1226
	1999	116383	163841	190959	219546	274800	196755	74761	1226
Reference	1994	88727	131882	158401	184592	231663	162522	67850	45670
	1999	109604	162334	193596	225610	288510	201347	88312	45670

Notes: For twin samples the first set of twins are used for summary statistics. N measures, accordingly, number of twin pairs. Income is measured in Swedish Crowns deflated to the price level of 2001. The exchange rate observed the 31st of December 2001 can be used to get approximated numbers in euros (1 Euro = 9.3029 SEK).

Table 2. Spearman rank correlation; male and female monozygotic samples

		1994		1995		1996		1997		1998		1999	
		twin 1	twin 2	twin 1	twin 2	twin 1	twin 2	twin 1	twin 2	twin 1	twin 2	twin 1	twin 2
1994	*t1*	1.0	0.5660	0.9079	0.5652	0.8571	0.5590	0.8468	0.5305	0.8206	0.4955	0.7961	0.5063
	t2	0.4080	1.0	0.5857	0.9109	0.5565	0.8794	0.5550	0.8524	0.5694	0.8266	0.5622	0.8157
1995	*t1*	0.8710	0.4094	1.0	0.5880	0.9142	0.5732	0.8861	0.5477	0.8500	0.5139	0.8233	0.5291
	t2	0.3833	0.8826	0.4107	1.0	0.5681	0.9262	0.5775	0.8789	0.5865	0.8524	0.5723	0.8382
1996	*t1*	0.7988	0.4222	0.8989	0.4148	1.0	0.5593	0.9251	0.5299	0.8716	0.4955	0.8467	0.5171
	t2	0.4010	0.8255	0.4266	0.8904	0.4305	1.0	0.5729	0.9331	0.5827	0.8994	0.5654	0.8794
1997	*t1*	0.7780	0.4248	0.8606	0.4225	0.9096	0.4298	1.0	0.5467	0.9206	0.5158	0.8803	0.5254
	t2	0.4064	0.7681	0.4322	0.8178	0.4251	0.9158	0.4371	1.0	0.5670	0.9381	0.5465	0.9097
1998	*t1*	0.7607	0.4226	0.8194	0.4263	0.8491	0.4398	0.9069	0.4584	1.0	0.5452	0.9317	0.5573
	t2	0.3998	0.7597	0.4279	0.7908	0.4283	0.8499	0.4477	0.8973	0.4699	1.0	0.5334	0.9259
1999	*t1*	0.7106	0.3938	0.7750	0.4017	0.7923	0.4253	0.8477	0.4505	0.9159	0.4555	1.0	0.5618
	t2	0.4028	0.7183	0.4298	0.7369	0.4396	0.7965	0.4676	0.8393	0.4776	0.9098	0.4779	1.0

Note: Spearman rank correlation for the male sample is included above the diagonal of 1.0, while results for the female sample are included below the diagonal.

Table 3. Summary statistics from nonparametric estimates, male samples

	Male monozygotic sample									
	1	2	3	4	5	6	7	8	9	10
Distributional effect	0.0554	0.0329	0.0181	0.0108	0.0087	0.0091	0.0112	0.0155	0.0224	0.0320
Growth adjusted distributional effect	0.0576	0.0352	0.0194	0.0111	0.0084	0.0089	0.0117	0.0171	0.0253	0.0357
Growth (median)[a]	0.0927	0.1368	0.1675	0.1652	0.1635	0.1519	0.1545	0.1561	0.1824	0.2352
Expected long-run rank effect[b]	0.0325	0.0239	0.0150	0.0087	0.0071	0.0087	0.0114	0.0157	0.0222	0.0307
Expected long-run twin rank effect[c]	0.0136	0.0113	0.0094	0.0073	0.0063	0.0066	0.0081	0.0104	0.0138	0.0191
b($rank_{t2}$) CI – high	0.0248	0.0182	0.0139	0.0108	0.0094	0.0097	0.0111	0.0137	0.0180	0.0256
b($rank_{t2}$) CI – low	0.0024	0.0044	0.0049	0.0039	0.0031	0.0036	0.0051	0.0072	0.0096	0.0126
Immobility[d]	0.5676	0.6856	0.7854	0.7771	0.8486	0.9696	0.9769	0.9198	0.8750	0.8610
Twin heterogeneity[e]	0.4194	0.4750	0.8468	0.6339	0.8814	0.7679	0.7138	0.6666	0.6245	0.6216

	Male dizygotic sample									
	1	2	3	4	5	6	7	8	9	10
Distributional effect	0.0792	0.0414	0.0194	0.0111	0.0084	0.0083	0.0103	0.0148	0.0226	0.0339
Growth adjusted distributional effect	0.0691	0.0377	0.0187	0.0109	0.0088	0.0089	0.0111	0.0166	0.0262	0.0402
Growth (median)[a]	0.2290	0.1417	0.1649	0.1475	0.1528	0.1599	0.1694	0.1679	0.1943	0.2306
Expected long-run rank effect[b]	0.0328	0.0223	0.0143	0.0099	0.0080	0.0075	0.0092	0.0150	0.0234	0.0343
Expected long-run twin rank effect[c]	0.0068	0.0044	0.0024	0.0026	0.0034	0.0032	0.0036	0.0056	0.0087	0.0124
b($rank_{t2}$) CI – high	0.0177	0.0108	0.0066	0.0059	0.0065	0.0063	0.0069	0.0093	0.0138	0.0208
b($rank_{t2}$) CI – low	-0.0042	-0.0021	-0.0017	-0.0008	0.0003	0.0000	0.0003	0.0019	0.0036	0.0040
Immobility[d]	0.4783	0.6000	0.7770	0.9118	0.9148	0.8385	0.8291	0.9059	0.8966	0.8548
Twin heterogeneity[e]	0.2069	0.1954	0.1700	0.2624	0.4238	0.4279	0.3897	0.3738	0.3712	0.3614

	Male reference sample									
	1	2	3	4	5	6	7	8	9	10
Distributional effect	0.2214	0.0845	0.0261	0.0138	0.0100	0.0088	0.0097	0.0129	0.0231	0.0578
Growth adjusted distributional effect	0.2229	0.0859	0.0283	0.0152	0.0102	0.0091	0.0104	0.0146	0.0263	0.0627
Growth (median)[a]	0.1312	0.1023	0.1212	0.1462	0.1489	0.1505	0.1560	0.1642	0.1918	0.2220
Expected long-run rank effect[b]	0.0520	0.0421	0.0228	0.0137	0.0099	0.0084	0.0092	0.0130	0.0234	0.0512
Immobility[d]	0.2405	0.5399	0.8094	0.9102	0.9722	0.9239	0.8839	0.8942	0.8893	0.8266

Notes: Nonparametric estimates are summarized with mean in each decile. a) Growth is measured with median to give a more appropriate measure in deciles 1 and 10 where a few observations have a substantial effect on the mean. b) b($rank_{t1}$). c) b($rank_{t2}$). d) Immobility = b($rank_{t1}$)/Growth adjusted distributional effect e) Twin heterogeneity = b($rank_{t2}$)/b($rank_{t1}$).

Table 4. Summary statistics from nonparametric estimates, female samples

	Female monozygotic sample									
	1	2	3	4	5	6	7	8	9	10
Distributional effect	0.0646	0.0390	0.0214	0.0128	0.0093	0.0081	0.0085	0.0108	0.0158	0.0237
Growth adjusted distributional effect	0.0570	0.0347	0.0194	0.0117	0.0085	0.0076	0.0087	0.0108	0.0168	0.0246
Growth (median)[a]	0.2770	0.2377	0.2319	0.2248	0.2119	0.2022	0.2004	0.2121	0.2129	0.2383
Expected long-run rank effect[b]	0.0304	0.0220	0.0140	0.0094	0.0069	0.0060	0.0075	0.0102	0.0139	0.0198
Expected long-run twin rank effect[c]	0.0156	0.0105	0.0046	0.0020	0.0030	0.0046	0.0058	0.0068	0.0096	0.0153
$b(rank_{t2})$ CI – high	0.0250	0.0162	0.0008	0.0052	0.0059	0.0076	0.0089	0.0102	0.0139	0.0219
$b(rank_{t2})$ CI – low	0.0062	0.0047	0.0084	-0.0011	-0.0001	0.0015	0.0026	0.0034	0.0053	0.0088
Immobility[d]	0.5357	0.6391	0.7263	0.7995	0.8205	0.7871	0.8637	0.8859	0.8284	0.8032
Twin heterogeneity[e]	0.5133	0.4721	0.3206	0.2184	0.4254	0.7589	0.7698	0.6644	0.6884	0.7746

	Female dizygotic sample									
	1	2	3	4	5	6	7	8	9	10
Distributional effect	0.0684	0.0390	0.0198	0.0118	0.0092	0.0080	0.0081	0.0100	0.0149	0.0232
Growth adjusted distributional effect	0.0549	0.0328	0.0180	0.0112	0.0085	0.0075	0.0080	0.0105	0.0158	0.0247
Growth (median)[a]	0.2831	0.2316	0.2193	0.2250	0.2124	0.2052	0.2007	0.2036	0.2113	0.2234
Expected long-run rank effect[b]	0.0279	0.0192	0.0110	0.0074	0.0065	0.0057	0.0063	0.0087	0.0130	0.0209
Expected long-run twin rank effect[c]	-0.0017	0.0009	0.0031	0.0029	0.0015	0.0010	0.0017	0.0029	0.0032	0.0044
$b(rank_{t2})$ CI – high	0.0051	0.0056	0.0066	0.0060	0.0044	0.0037	0.0044	0.0058	0.0071	0.0107
$b(rank_{t2})$ CI – low	-0.0085	-0.0037	-0.0004	-0.0002	-0.0013	-0.0018	-0.0011	-0.0001	-0.0007	-0.0018
Immobility[d]	0.5102	0.5888	0.6143	0.6630	0.7681	0.7607	0.7825	0.8289	0.8198	0.8452
Twin heterogeneity[e]	-0.0593	0.0585	0.2945	0.3884	0.2358	0.1710	0.2642	0.3301	0.2488	0.2121

	Female reference sample									
	1	2	3	4	5	6	7	8	9	10
Distributional effect	0.2019	0.0692	0.0211	0.0122	0.0099	0.0087	0.0085	0.0102	0.0162	0.0451
Growth adjusted distributional effect	0.1629	0.0611	0.0208	0.0122	0.0093	0.0080	0.0087	0.0108	0.0177	0.0495
Growth (median)[a]	0.4613	0.2084	0.2127	0.2086	0.2110	0.1997	0.1959	0.2031	0.2060	0.2391
Expected long-run rank effect[b]	0.0516	0.0322	0.0161	0.0092	0.0073	0.0063	0.0069	0.0097	0.0155	0.0375
Immobility[d]	0.3233	0.5682	0.7739	0.7507	0.7891	0.7873	0.7972	0.8935	0.8877	0.7605

Notes: Nonparametric estimates are summarized with mean in each decile. a) Growth is measured with median to give a more appropriate measure in deciles 1 and 10 where a few observations have a substantial effect on the mean. b) $b(rank_{t1})$. c) $b(rank_{t2})$. d) Immobility = $b(rank_{t1})$/Growth adjusted distributional effect e) Twin heterogeneity = $b(rank_{t2})/b(rank_{t1})$.

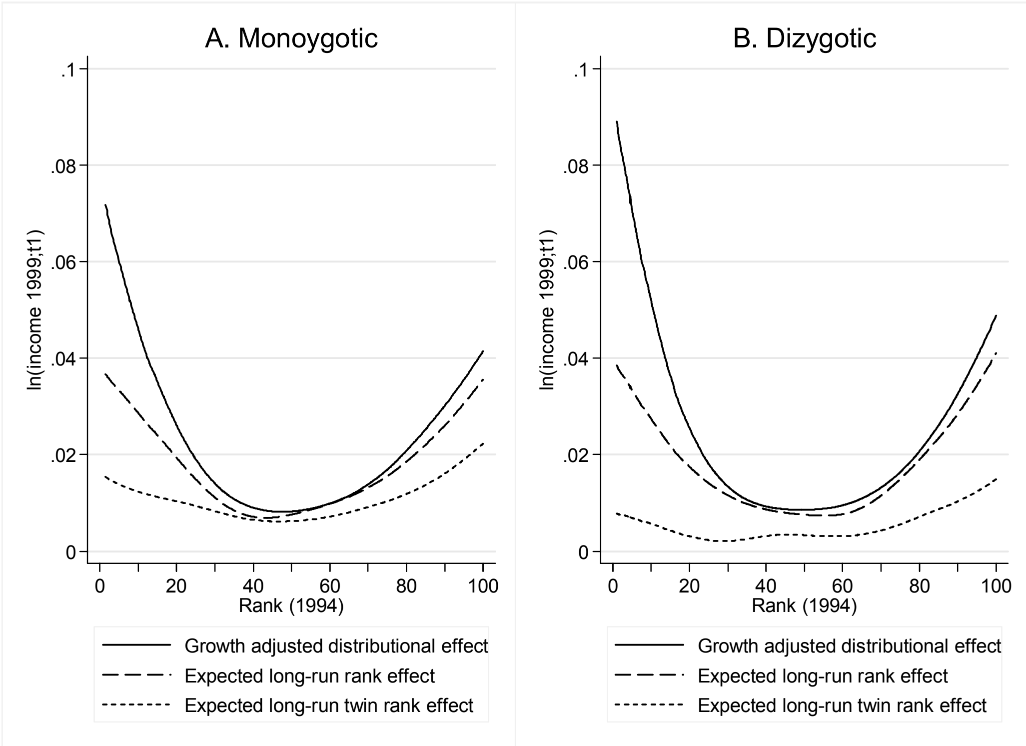

Figure 1. Male monozygotic and dizygotic samples.

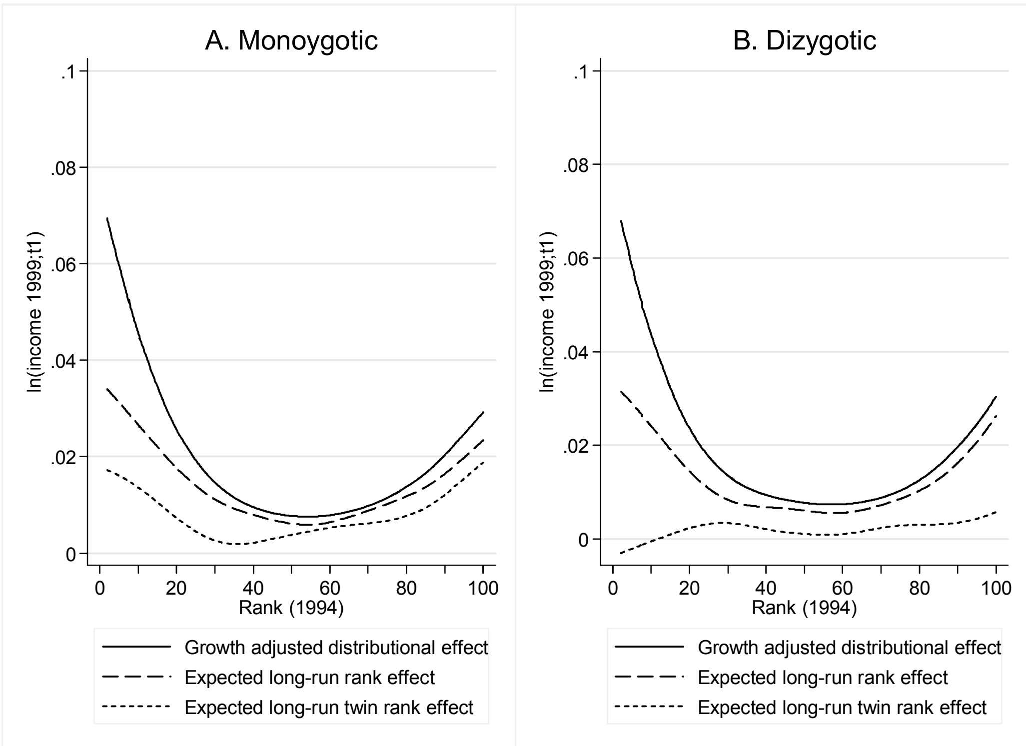

Figure 2. Female monozygotic and dizygotic samples