Intergenerational Correlation Curves:

Evidence from PSID

William Nilsson

Department of Applied Economics, University of the Balearic Islands
Ctra Valldemossa Km 7,5, E-07122 Palma de Mallorca, Spain
Phone: +34 971 17 13 77, Fax: +34 971 17 23 89
william.nilsson@uib.es

Abstract: Correlations curves are used to measure the degree of association of father's and son's earnings locally, i.e. at different positions in the earnings distributions. The local correlation is nonlinear, and can be very different from local elasticities from nonparametric, or quantile regression. Elasticities should not be used to measure the local degree of association.

Keywords: Intergenerational mobility; nonlinear; nonparametric; correlation curve

JEL: C14, D63, J62

1 Introduction

The purpose of this study is to introduce correlation curves to measure intergenerational income correlation. Estimating intergenerational income elasticity, i.e. regressing father's log income on the child's adult log income, does not, in general, measures the degree of association. Interpreting the elasticity as a correlation coefficient is only correct as long as the variance in income is the same for the two generations. Solon (1992) clearly discusses this issue, but, while the reference is widely used, this insight seems to have lost in attention over time. In addition, the dispersion can be different locally, which makes it very dubious to interpret the elasticity as local degree of association. For example, Eide & Showalter (1999) use quantile regression and state that their objective is "to look for differences in explanatory power of father's earnings across the conditional distribution of son's earnings". They also refer to the results as "correlation", but the elasticities obtained do not measure explanatory power and, only in special cases, is an interpretation as a correlation valid. Björklund et al (2012), use a linear spline regression across fathers' fractiles and find elasticities of 0.896 for incomes and 0.447 for earnings, for the top 0,1% of the income distribution of the fathers. They are, in general, careful to refer to the elasticities as measures of income transmission, but they do ask "Why is it that the intergenerational association is so strong in the top?". Again, the elasticity does not measure the degree of association, and while they make an important effort to explain the reasons for the high transmission, its explanation may simply be due to a longer upper tail in the son's income distribution which can inflate the elasticity. It seems to be important to measure the *degree of the association* locally, at either different parts of the father's or son's income distribution, and a correlation curve is introduced in this study.

Data from the Panel Study of Income Dynamics (PSID) is used to illustrate how correlation curves can be informative. Data from PSID has been used extensively to study intergenerational income transmission. See for example, Solon (1992), Eide & Showalter (1999), Murtazashvili (2012) and Gouskova et al. (2012). While several methodological caveats (i.e. life cycle bias, how to measure permanent income, sample selection etc.) are present when intergenerational income transmission is estimated, these issues are completely disregarded in this study. The purpose is not to obtain the perfect estimate, rather to clarify the importance of distinguishing intergenerational income transmission from intergenerational income association, and show how the correlation curve can be used to obtain flexible estimates of the latter. Details concerning the literature on intergenerational income mobility can be found in the surveys Solon (1999) and Blanden (2011).

2 Method

Bjerve & Doksum (1993) introduced correlation curves, $\rho(x)$, which measures the degree of association locally at different values of the covariate X. Their proposed correlation curve is invariant to changes in the origin and scale and is $-1 \leq \rho(x) \leq 1$ for all x. It is accordingly easy to interpret and a suitable alternative to the correlation coefficient for nonlinear models. Their correlation curve is defined as,

$$\rho(x) = \frac{\sigma_x m'(x)}{\left[\{\sigma_x m'(x)\}^2 + \sigma^2(x)\right]^{1/2}}, \qquad (1)$$

where $m(x) = E(Y \mid X = x)$, where X in this case is the logarithm of income of the fathers/parents and Y is the logarithm of income of the children at adult age. $m'(x)$ is the slope of the regression function and can be used as a local measure of the intergenerational income transmission. $\sigma^2(x) = \text{var}(Y \mid x)$ is the residual variance. σ_x is the standard deviation of X. The correlation curve can be estimated parametrically, as showed in Blyth (1994), or, by using a more flexible nonparametric estimation technique.

Bjerve & Doksum (1993) discuss the properties of the correlation curve. These are very similar to the correlation coefficient, but an important difference is that, in general, $\rho_{XY}(\cdot) \neq \rho_{YX}(\cdot)$. $\rho_{XY}(x)$ is used if we want the measure the local degree of association at different positions of the income distribution of the fathers. $\rho_{YX}(y)$ is used if the interest is on the degree of association at different position of the adult children's income distribution.

Local polynomial regression (Fan, 1993), with a normal kernel, is used to estimate the regression function, its derivative, and the residual variance, to be able to estimate the correlation curve. The optimal bandwidth for the second derivative is found with the data-driven procedure suggested by Fan & Gijbels (1995). A simple algorithm was used to drop extreme outliers in the tails of the income distribution of x, when $\rho_{XY}(x)$ was estimated, and y, when $\rho_{YX}(y)$ was estimated. These few observations would otherwise force the optimal bandwidth to be the first sufficiently large bandwidth to avoid singularity problems due to sparse data. Pointwise bootstrap confidence intervals are included as suggested in Nilsson & del Barrio (2012).

3 Data

The source for the data is the Panel Study of Income Dynamics (PSID). For simplicity, the data was downloaded from the Data archive of Journal of Applied Econometrics and was used in Murtazashveli (2012). The multiple sons sample with 404 observations is used in the analysis. A more detailed description can be found in Murtazashveli (2012). The upper part of table 1 includes summary statistics.

[Table 1 about here]

Notice that the difference between percentile 10 and 20 is much larger for the sons compared to the fathers. If percentiles 5 and 15 are also calculated the larger dispersion in the lower tail of the distribution is evident. The difference in log earnings for the sons at percentile 10 compared to 5 is 0.80 and at percentile 15 compared to 10 is 0.27. The corresponding numbers for the fathers are 0.41 and 0.13. A locally higher dispersion will inflate the elasticity in this part of the distribution, and the elasticity will not be close to a local measure of the degree of association.

4 Results

The lower part of table 1 includes results from ordinary least squares (OLS) and quantile regression, which are used as departures for the analysis. The elasticity from OLS is higher than the correlation coefficient, since the standard deviation is higher for the son's earning. Using quantile regression provides elasticity at percentile 10 that is substantially higher than in other parts of the distribution. It is found to be as high as 0.59. A correlation curve is used to evaluate if the degree of association is high, or if only the transmission is high in this position of the son's distribution.

[Figure 1 about here]

Figure 1 includes a scatter plot with the regression function and a correlation curve. The dotted curve is the median of the correlations curve from the bootstrap replications, and the dashed curves are 95% confidence intervals. The solid grey curve is the corresponding elasticity. The elasticity is above 2 at a log earnings between 9.4 and 9.6 for the father.

This is certainly an extreme intergenerational transmission and, being above 1, indicates that it cannot be used to measure the local degree of association. The correlation curve underlines that the degree of association is about 0.8 when the log earnings of the father is just below 9.6, where its highest value is found. This is below percentile 20 in the earnings distribution of the fathers. The extreme dispersion in the sons' distribution, at this part of the distribution, has made the elasticity to reach very extreme values.

The scatter plot identifies four observations with son's log earnings below 6.5, and three of these are attached with log earnings of a father close to 9.2. As a sensitivity analysis these observations were dropped.

[Figure 2 about here]

The quantile regression at the 10 percentile is suddenly dropped to 0.33 and it is almost the same as in percentile 20, where it is 0.32. Decile 1 is still the decile where the elasticity is found to the highest. The point estimate of the elasticity from OLS is also reduced, from 0.352 to 0.286. The nonparametric estimate of the elasticity is now reduced to above 0.8, when father's log earnings are between 9.4 and 9.6. The correlation curve is just below 0.6 at the same position. All of the local measures are found to be sensitive to the exclusion of only four observations; the local nonparametric elasticity is, at this position of the distribution, reduced with more than 50%, while the correlation curve is reduced with just above 25%.The elasticity for the first decile when quantile regression is used is also sensitive, and the elasticity is almost reduced with 50% compared to what was found earlier.

The final part of the analysis evaluates the quantile regression elasticity to see if a correlation interpretation is valid. Now the correlation curve is estimated reversed to see the local correlation at different positions of the son's earnings distribution.

[Figure 3 about here]

The point estimate of the correlation curve at a log earnings of about 9.4, i.e. at the 1 decile, the correlation curve is just above 0.2. The relation follows an inverted U-shaped pattern, and its maximum is reached with a correlation of 0.5 at log earnings of about 10.0, i.e. above the median. Interpreting the results from the quantile regression for the lowest decile of the son's earnings distribution as a degree of association is clearly incorrect.

5 Concluding remarks

An intergenerational transmission, based on an estimated elasticity, should not be interpreted as an intergenerational degree of association. In fact, there is no reason to do so. A correlation curve is clearly more suitable for answering such question. Using correlation curves indicates the degree of association is particular strong just before the second decile in the earnings distribution of the fathers, and than gradually decreases, and is close to zero at the 8th decile. If the degree of association is evaluated at different positions of the son's earnings distribution, the correlation is particularly strong just after the median, and much lower for the tails of the distribution.

The sample size used in the analysis is very small and many data-related questions (such as measure permanent income, life-cycle bias etc.) are left aside. This prevents drawing general conclusion from the results, but the analysis clearly shows that correlation curves are feasible and highly suitable tools to study intergenerational income correlation at different parts of the distributions. If the interest is on estimating a local *degree of association*, the elasticity, is likely to provide inadequate conclusions. Note that the elasticity is still a relevant measure, but it only answers questions concerning intergenerational income transmission.

Acknowledgments

The author gratefully acknowledges the financial support from the Spanish Ministry of Science and Innovation (grant #ECO2010-20829).

References:

Bjerve, S., Doksum, K., 1993. Correlation Curves: Measures of Association as Functions of Covariate Values. The Annals of Statistics. 21, 890-902.

Björklund, A., Roine, J., Waldenström, D., 2012. Intergenerational top income mobility in Sweden: Capitalist dynasties in the land of equal opportunity? Journal of Public Economics. 96, 474-484. DOI: 10.1016/j.jpubeco.2012.02.003

Blyth S., 1994. Karl Pearson and the correlation curve. International Statistical Review. 62, 393-403.

Blanden, J., 2011. Cross-country rankings in intergenerational mobility: a comparison of approaches from economics and sociology. Journal of Economic Surveys. DOI: 10.1111/j.1467-6419.2011.00690.x

Eide, E.R., Showalter, M.H., 1999. Factors affecting the transmission of earnings across generations: a quantile regression approach. The Journal of Human Resources. 34, 253-267.

Fan, J., 1993. Local Linear Regression Smoother and Their Minimax Efficiency. The Annals of Statistics. 21, 196-216.

Fan, J., Gijbels, I., 1995. Data-Driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation. Journal of the Royal Statistical Society. Series B (Methodological) 57, 371-394.

Gouskova, E., Chiteji, N., Stanford, F., 2010. Estimating the intergenerational persistence of lifetime earnings with life course matching: Evidence from the PSID. Labour Economics. 17, 592-597. DOI: 10.1016/j.labeco.2009.04.009

Murtazashvili, I., 2012. An Alternative Measure of Intergeneracional Income Mobility Based on a Random Coefficient Model. Journal of Applied Econometrics. 27, 1161-1173. DOI: 10.1002/jae.1246

Nilsson, W., del Barrio Castro, T., 2012. Bootstrap Confidence Interval for a Correlation Curve. Statistics and Probability Letters. 82, 1-6. DOI: 10.1016/j.spl.2011.09.001

Solon, G., 1992. Intergenerational income mobility in the United States. American Economic Review. 82, 393–408.

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

Tables and figures titles to be inserted in main text:

Table 1. Summary statistics

	P10	P20	P30	P40	P50	P60	P70	P80	P90		
	Incomes at different deciles									Mean	SD
Son's earnings (1984)	8.99	9.39	9.62	9.80	9.95	10.09	10.24	10.34	10.53	9.78	0.89
Father's earnings (1967)	9.41	9.65	9.91	10.10	10.19	10.35	10.44	10.53	10.75	10.10	0.67
	Quantile regression									OLS	Corr.
N=404	0.593	0.361	0.336	0.316	0.352	0.325	0.246	0.246	0.248	0.352 (0.064)	0.263

Notes: All regression slopes are significantly different from zero at 1% significance level. Standard error is reported in parenthesis for the OLS.

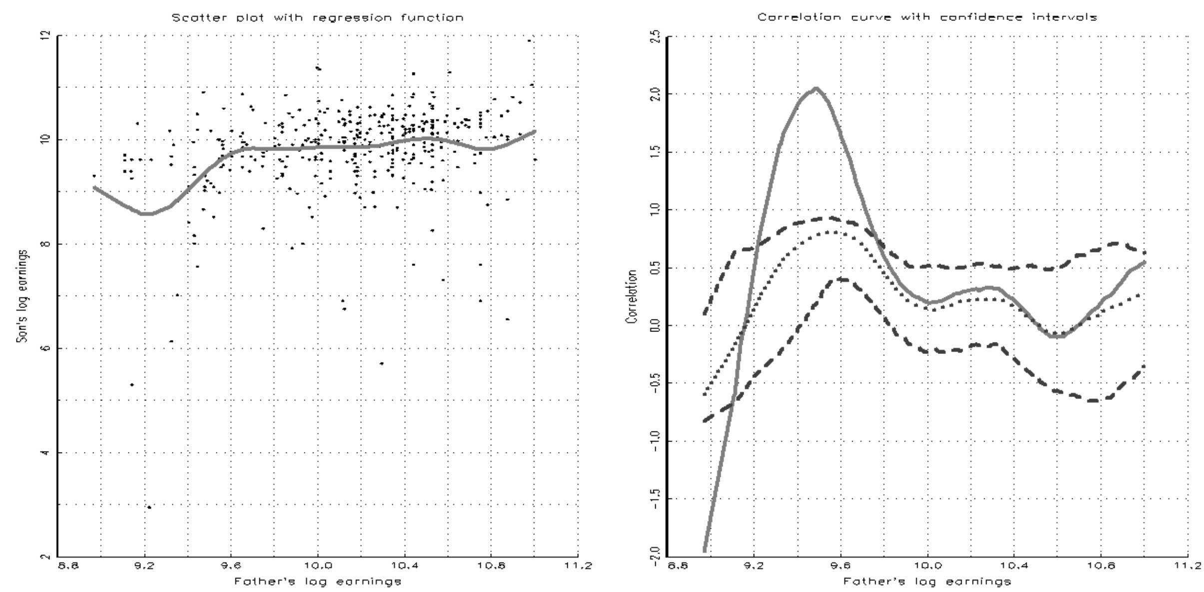

Fig. 1. Scatter plot and correlation curve (and elasticity)

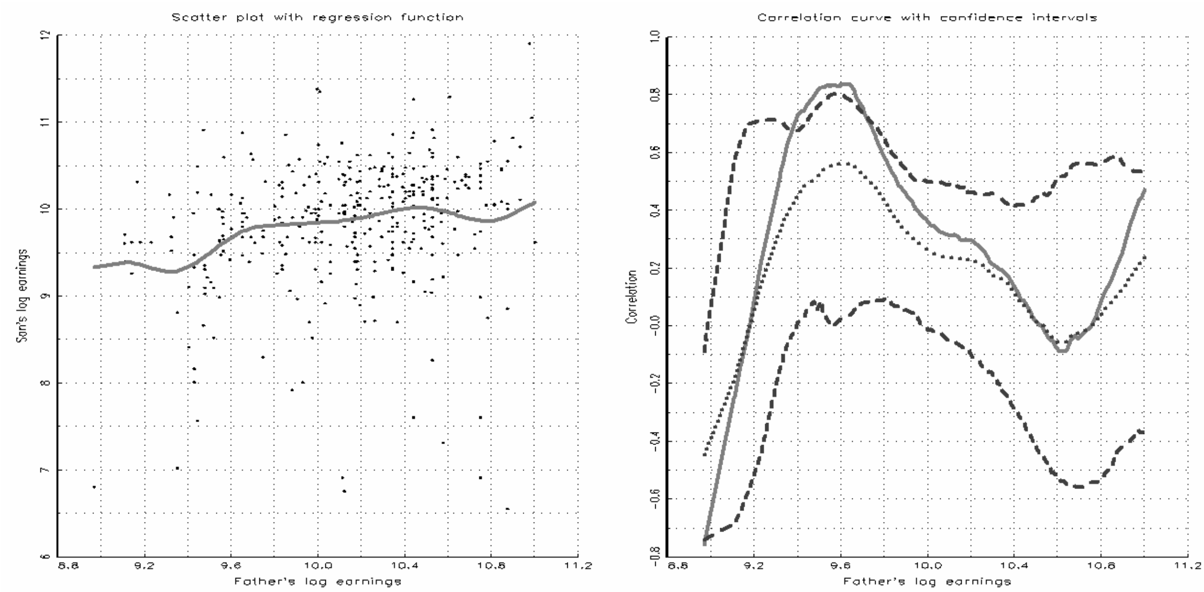

Fig. 2. Scatter plot and correlation curve (and elasticity) when the 4 observations with the lowest son's earnings are dropped.

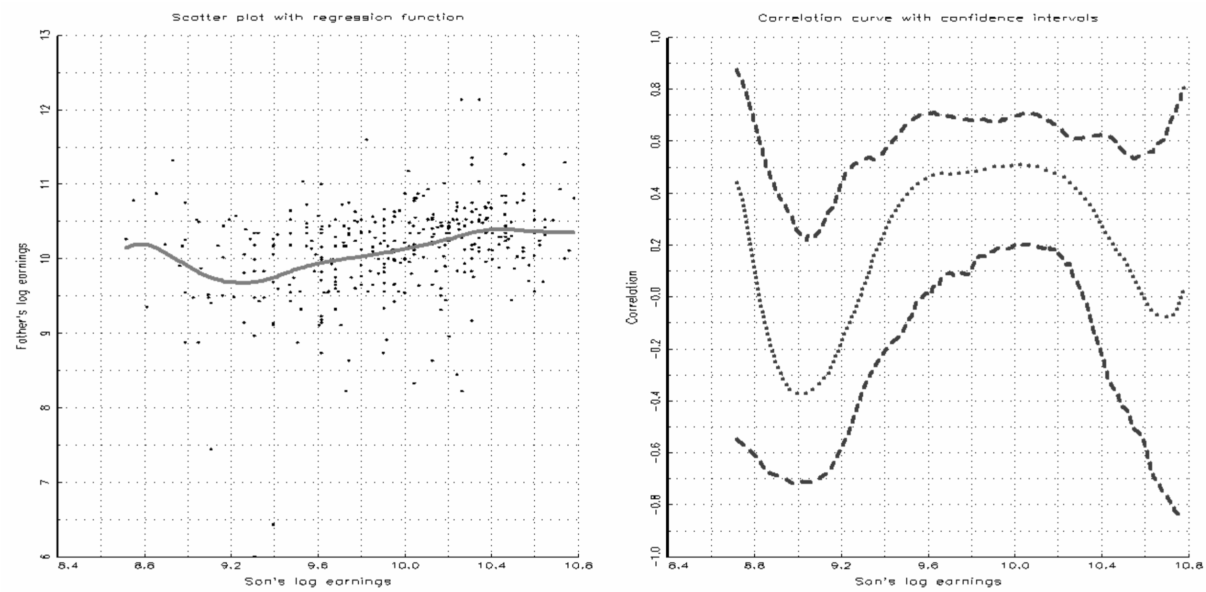

Fig. 3. Scatter plot and correlation curve, given different log earning for the son