China Economic Review 31 (2014) 379–391

Contents lists available at ScienceDirect

China Economic Review

Decomposing the rich dad effect on income inequality using instrumental variable quantile regression

Zaichao DU [a,*], Renyu LI [a], Qinying HE [b,c,**], Lin ZHANG [a]

[a] Research Institute of Economics and Management, Southwestern University of Finance and Economics, Chengdu, China
[b] School of Economics and Management, South China Normal University, Guangzhou, China
[c] Scientific Laboratory of Economic Behaviors, South China Normal University, Guangzhou, China

ARTICLE INFO

Article history:
Received 16 September 2013
Received in revised form 11 June 2014
Accepted 12 June 2014
Available online 19 June 2014

JEL classification:
D31
E24
J62

Keywords:
Intergenerational inequality
Counterfactual decomposition
Composition effect
Income structure effect
Instrumental variable quantile regression

ABSTRACT

In this paper we evaluate the relative importance of the two main channels, namely the composition effect and the income structure effect, through which the paternal income affects children's income inequality. Using data on 2677 pairs of father and children from China Health and Nutrition Survey (CHNS), we construct the counterfactual income of children from poor families if they had the same characteristics as children from rich families. We propose an instrumental variable quantile regression-based method to solve the endogeneity problem and decompose the rich dad effect on income inequality into the composition effect and the income structure effect. We find that the composition effect explains at least 80% of the income difference at any quantile, and it explains all the income difference at the top four deciles. Income structure effect has a significant impact only at quantiles between 20% and 40%, where it explains about 20% of the income difference.

1. Introduction

Income inequality has become a more and more important problem in China. The income share held by the top 10% households in China has exceeded 57%.[1] Meantime, about 50% of Chinese households have no savings for the year 2010.[2] How to reduce the income inequality has become one of the biggest problems for stimulating domestic consumptions and economic growth in China.

One important source of income inequality is through intergenerational transmission, which has drawn the attention of many researchers, e.g. Becker and Tomes (1986), Zimmerman (1992), Solon (1999), and Erikson and Goldthorpe (2002), just to name a few. Meanwhile, as evidenced by the large popularity and controversy of "competition of the father" and "the rich second generation", intergeneration inequality is also of great concern to the general public.

This paper studies the relative importance of the two main channels, namely the composition effect and the income structure effect (cf. Firpo, Fortin, & Lemieux, 2007; Mata & Machado, 2005), through which the paternal income affects children's income inequality. Most of the literature on intergenerational inequality has been focusing on intergenerational income elasticity (IIE), see Solon (1999) and Björklund and Jäntti (2009) for an excellent survey. However these studies fail to answer how the income of the parents

* Correspondence to: Z. Du, RIEM, Southwestern University of Finance and Economics, 55 Guanghuacun Street, Chengdu 610074, China. Tel.: +86 28 8735 3797.
** Correspondence to: Q. He, School of Economics and Management, South China Normal University, Guangzhou Higher Education Mega Center, Guangzhou 510006, China. Tel.: +86 20 39310177.
E-mail addresses: duzc@swufe.edu.cn (Z. Du), qinying_he@hotmail.com (Q. He).

[1] *Research Report of China Household Finance Survey 2012.* The same number for US in 2009 is only 40.6%.
[2] *Research Report of China Household Finance Survey 2012.*

http://dx.doi.org/10.1016/j.chieco.2014.06.007
1043-951X/

affects that of the children. One exception is Lefgren, Lindquist, and Sims (2012), who study the effects of the father's human capital and financial resources on children's income. Our study combines two strands of research on the transition mechanism of intergenerational inequality. On the one hand, Bowles and Gintis (2002), Shea (2000), Mayer (2002), Dahl and Lochner (2012) and many others have found that parents with higher income can invest more on their children's human capital, which brings higher income for their children. On the other hand, Conlisk (1974), Ruhm (1988), Björklund and Jäntti (2009), Zhang and Eriksson (2010) and Li, Meng, Shi, and Wu (2012) have found that children from different socioeconomic family backgrounds, mainly characterized by the income of the parents, may face different work opportunities (or different returns to the covariates), which may also cause income inequality for the children.

We disentangle the composition effect and the income structure effect by counterfactual decomposition. We consider two groups of children, children from rich families and children from poor families. We construct the counterfactual income of children from poor families if they had the same characteristics (or covariates) as children from rich families. The difference between the counterfactual income and the actual income of children from poor families is then purely due to the covariates differences between the two groups of children. Following the literature (Firpo et al., 2007; Mata & Machado, 2005), we call this part the composition effect. The difference between the actual income of children from rich families and the counterfactual income is due to the difference in the returns to the covariates, which is called the income structure effect.

Our methodological contribution is that we extend the counterfactual decomposition method of Mata & Machado, 2005, (MM hereinafter) to cases with endogenous variables. The MM method has been widely used in wage distribution decomposition, and it generalizes the traditional Oaxaca (1973) decomposition of effects on mean wages to the entire wage distribution. However, MM does not consider the endogeneity problem, which will bias the decomposition in cases like here. Our method solves the omitted variable problem by incorporating the IV quantile regression method proposed by Chernozhukov and Hansen (2008). Another nice approach would be the unconditional quantile regression and the decomposition method proposed by Firpo et al. (2007); Firpo, Fortin, and Lemieux (2009), but it is unclear to us how to extend it to the endogenous variable case.

Our empirical analysis using CHNS data finds that the composition effect explains at least 80% of the income disparity between children from rich families and from poor families, and it explains all the income difference at the top 40% quantiles of the distribution. Income structure effect explains about 20% of the income difference at the lower 40% quantiles of the income distribution. Some analyses show that our results are robust to alternative specifications.

Our study sheds some light on the transition mechanism of intergenerational inequality. It gives a better picture of how the composition effect and income structure effect contribute to the income difference between children from rich and poor families. Our results suggest that improving the labor market efficiency can reduce the income inequality to a certain extent, and more importantly, helping children from poor families to get more education can reduce the income inequality substantially.

The rest of this paper is organized as follows. Section 2 describes our econometric model and decomposition method. Section 3 describes our data. Section 4 presents the empirical results. In Section 5 we do some robustness checks and we conclude in Section 6.

2. The econometric model and the decomposition method

In this section, we propose an instrumental variable (IV) quantile regression (QR)-based counterfactual decomposition method. Our method extends MM's method to cases with endogenous variables. One issue with MM's procedure is the omitted variable problem, here ability, which may bias the decomposition when there exists a systematic difference between the ability of children from rich and poor families, see Fortin, Lemieux, and Firpo (2011). Our method solves the omitted variable problem by incorporating the instrumental variable quantile regression method proposed by Chernozhukov and Hansen (2008), which is introduced next.

2.1. An instrumental variable quantile regression (IVQR) model

To allow endogenous control variables and to better describe the different behaviors of people at different income levels, we consider the following IVQR model:

$$\begin{aligned} &Y = D'\alpha(U) + X'\beta(U), \\ &D = f(X, Z, V), \quad U|X, Z \sim Uniform(0, 1), \end{aligned} \tag{1}$$

where Y is the log of the CPI-adjusted disposable yearly income of an individual. U is a scalar random variable that aggregates all of the unobserved factors affecting Y. U follows a uniform distribution on the interval (0,1) as in any QR model. X is a vector of exogenous control variables including experience,[3] experience square, gender, urban, SOE (state-owned enterprise), migrant worker, small business, provincial dummies, time dummies and a constant term. Urban is a binary variable that takes value 1 if one's Hukou is in the urban area and 0 otherwise. Dummy variable SOE equals 1 if an individual works for a state-owned enterprise and 0 otherwise. Variable migrant worker equals 1 if an individual is a migrant worker and 0 otherwise. D is years of education. Because the well-known omitted variables problem, we allow D to be endogenous here. We further assume that D is a functional of X, Z and V, where Z is an IV, and V is an error term affecting D. Here we use the community education index,[4] a measure of the average education level of a community, as the IV, which is

[3] We do not control age here, as experience is a linear function of age. Specifically, here experience = age − years of education − 7.

[4] The community education index is one component of the urbanization index created by Jones-Smith and Popkin (2010) and released by CHNS. The community education index allots a maximum total of 10 points for the average educational attainment of adults more than 21 years old in the community, and a higher score indicates a higher community education level. The scoring algorithms are developed based on distributions in the data, with the goal of having the median score be close to half of the total possible points and with sufficient spread in the scores between the minimum and maximum points.