Dependency Ratio and Future Education Level Improvement: by analyzing the preschool education ratio

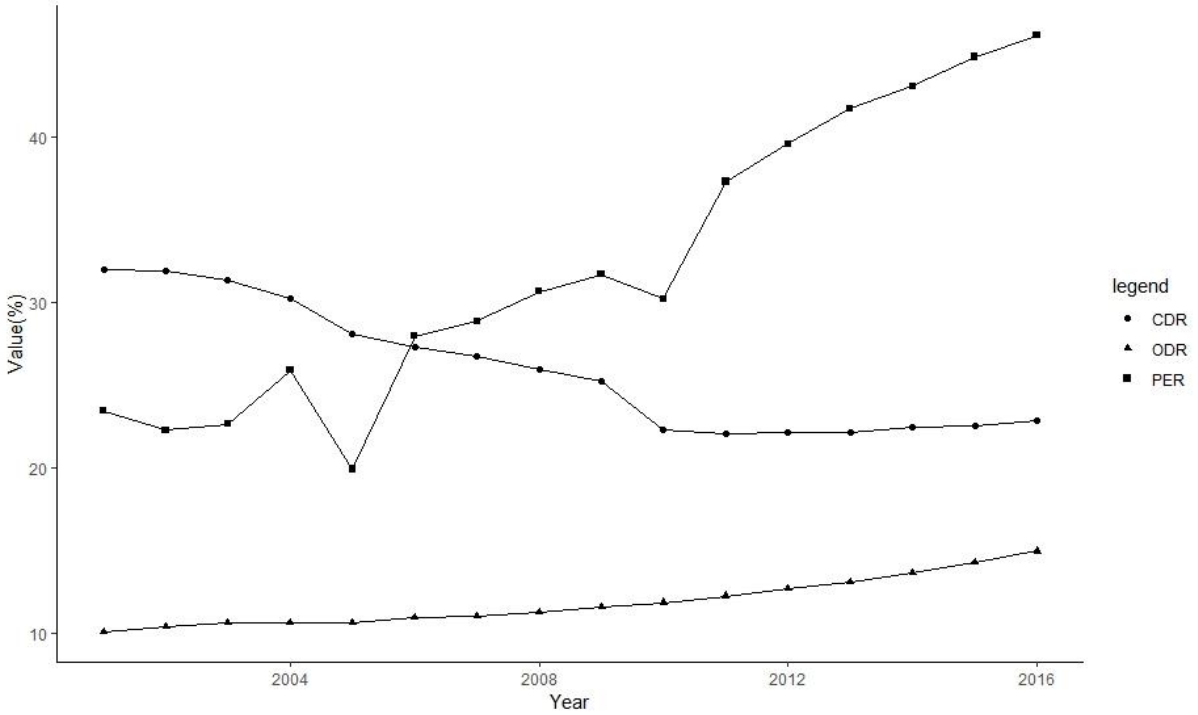
By ZENG SIDONG AND ZHAO DELIN*

Using the panel data regression, we estimate the causal effect of dependency ratios on work force's willingness to have children receive more and better education by analyzing the preschool education ratio. Our estimates imply that an increase in the children dependency ratio will pull down the preschool education ratio while an increase in the old dependency ratio will raise it up. The decreasing effect of the children dependency ratio is smaller for the regions in which real GDP per capita is higher.

I. Background

Figure 1 suggests that from 2001 to 2016, the children dependency ratio (***CDR*** for short) has been showing a downward trend and old dependency ratio (***ODR*** for short), on the contrary, has been trending upward. The age structure in China is now experiencing a tremendous change, and how this change will affect the future education level in China is of our particular interest. Here, we focus on one key factor of boosting the future education level and cultural quality of the whole society — the willingness of work force to have their children receive more and better education. In this passage, this willingness is reflected by preschool education ratio (the proportion of children of the appropriate age for preschool education to actually receive preschool education, ***PER*** for short).

* Zeng Sidong: Mathematical Statistics, Department of Statistics, School of Economics, Xiamen University. Student ID: 15220162202483; Zhao Delin: Mathematical Statistics, Department of Statistics, School of Economics, Xiamen University. Student ID: 15220162202520.



(Figure 1)

Here are several reasons why preschool education ratio can reflect the willingness of work force:

The preschool education resources are far from scarcity. Kindergartens, for example, are capable of providing sufficient and high-quality enrollment. Comparing to high schools and universities that have strict restrictions on the number of enrollments, parents can always find a kindergarten for their children if they want.

Unlike the high schools and universities that have some requirements of enrollment's competence, such as their grades of the entrance examination, kindergartens will let the children in regardless of their capacity, as long as their supporters are willing to pay for the tuition.

As the kind of education all teens must take, compulsory education cannot reflect the willingness of work force.

Whether a child will take the preschool education or not depends entirely on his or her supporters since they are too young to think for themselves.

II. Data

Data used here comes from *China Statistical Yearbook,* which is panel data for all Chinese provinces except Xinjiang, Hong Kong, Macao and Taiwan, and for year from 2001 to 2016.

A. Variable Introduction

Variable	Definition
*Preschool Education Ratio (**PER**)*	the proportion of children of the appropriate age for preschool education to actually receive preschool education
*Children Dependency Ratio (**CDR**)*	The number of population aged 0-14 / the number of population aged 15-64 (work force population)
*Old Dependency Ratio (**ODR**)*	the number of population aged 65 and over / the number of population aged 15-64
*Compulsory Education Ratio (**CE**)*	The proportion of work force that have finished primary education or secondary education
*Illiteracy Ratio (**IR**)*	The illiteracy rate of work force
*Inflation Rate (**INF**)*	CPI index based on 2001 (price level in 2001 is 100)
*Level of Real GDP Per Capita (**GDP**)*	Real GDP per capita index based on 2001(real GDP per capita in 2001 is 100)
*Consumption Rate (**CR**)*	Final consumption expenditure of households / gross regional domestic product

(Table 1)

B. Variable's Statistical Description

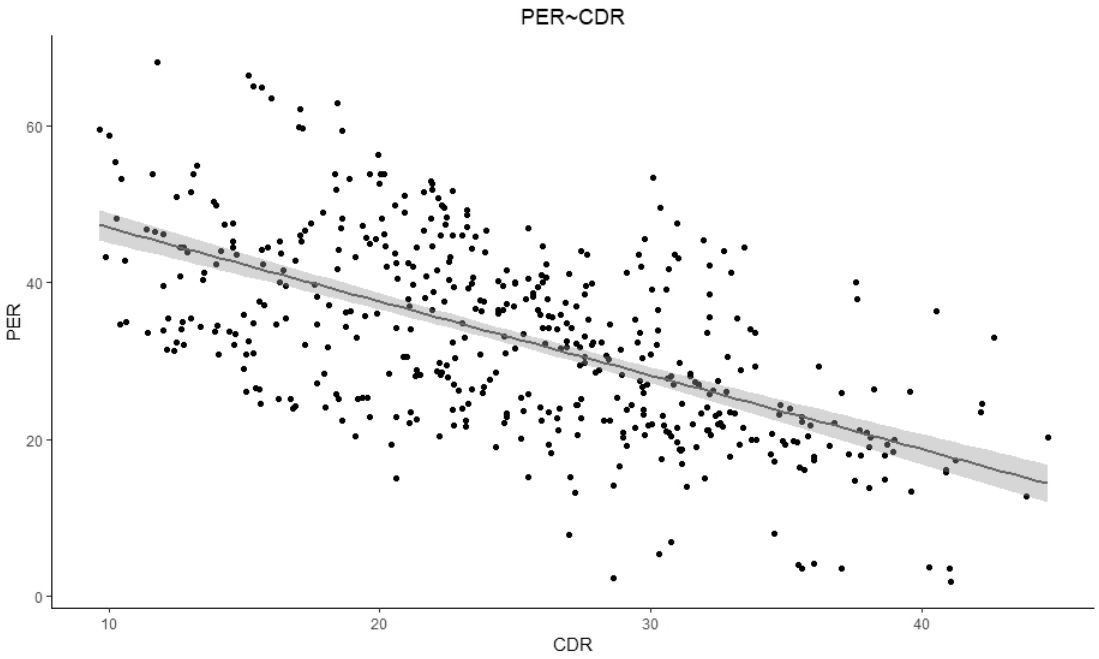
Statistical Description of Variables

Variable	Unit	Mean	Min	Max	SD
PER	%	32.92422	1.78	68.23	12.0752
CDR	%	24.95575	9.64	44.65	7.57397
ODR	%	12.39185	6.71	21.88	2.57685
IR	%	5.966521	0.20	55.10	7.33374
CE	%	67.80769	23.7	82.77	10.7228
INF	%	120.5236	98.2	170.7	16.5508
GDP	%	252.3740	100	734.7	123.397
CR	%	36.54216	22.9	65.15	6.39911

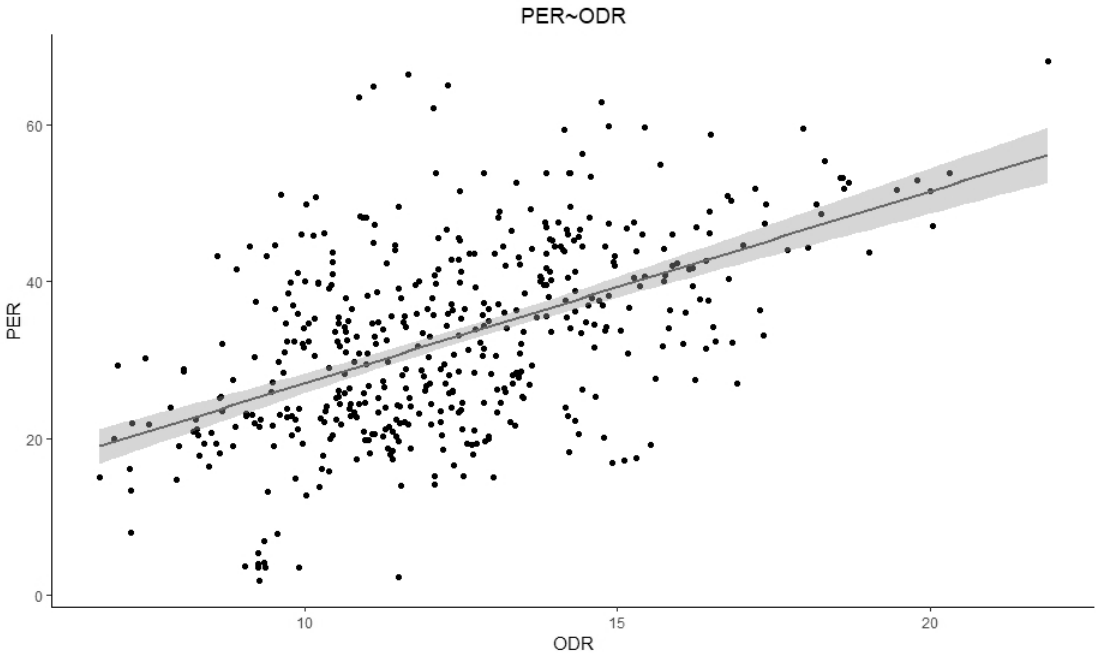
(Table 2)

III. Model

A. Linear Model:



(Figure 2)



(Figure 3)

Figure 2 shows the relationship between preschool education ratio and children dependency ratio while Figure 3 the relationship between preschool education ratio and old children. It is natural to speculate from the two scatter plots that preschool education ratio is linearly correlated with the two dependency ratios. One simple linear model is

therefore constructed as follows:

$$\widehat{\boldsymbol{PER}}_{\boldsymbol{i}} = 30.42157 - 0.74411 \times \boldsymbol{CDR}_{\boldsymbol{i}} + 1.70051 \times \boldsymbol{ODR}_{\boldsymbol{i}} + \boldsymbol{u}_{\boldsymbol{i}}$$

(2.91110) (0.05681) (0.16698)

$$\boldsymbol{SER} = 8.841, \quad \bar{\boldsymbol{R}}^{\boldsymbol{2}} = 0.4639$$

*(The p-value of these two regressions (**CDR** and **ODR**) are both less than 0.01)*

B. Panel Data Regression:

Considering the characteristics of time series data, the unit root test and co integration test are needed here to avoid pseudo-regression which the instability of economic variables may cause and to make sure the validity of the regression results.

The unit root test of all variables:

Unit root test of variables

Variables	Test Statistics (*p*-value)		Results
	ADF	PP	
PER	-7.2942(<0.01)***	-447.83(<0.01)***	None
CDR	-7.2333(<0.01)***	-516.36(<0.01)***	None
ODR	-6.3365(<0.01)***	-451.57(<0.01)***	None
IR	-6.8500(<0.01)***	-442.45(<0.01)***	None
CE	-6.1924(<0.01)***	-503.05(<0.01)***	None
INF	-7.1656(<0.01)***	-268.54(<0.01)***	None
GDP	-10.785(<0.01)***	-554.35(<0.01)***	None
CR	-6.1072(<0.01)***	-532.48(<0.01)***	None

(Table 3)

The results suggest all variables are stationary, so there is no need for co integration test.

Here is some analysis of the regression results in Table 4.

Column (1) uses the pooling model and is just the same to the simple linear model above.

It is natural to think that provinces differ greatly from each other and each province should have its unique intercept, so individual fixed effect model may be preferred than the pooling model in this case. Column (2) is the result of the individual fixed effect model using the same explanatory variables as the regression in Column (1). The *F*-statistic of testing "Province effects = 0" is 4.4382 with a *p*-value of less than 0.01, so the province effects are significant at the 1% level. "Time effects = 0" is also tested here with a *p*-value of 0.1362, suggesting time effects are insignificant at the 10% level.

Column (3) is the result of adding two control variables related to the education level of work force into the regression in Column (2). On the one hand, they will affect the procreation conception of the masses, and in turn affect the society's children

dependency ratio, on the other hand, they will affect work force's willingness to support the old, and in turn affect the old dependency ratio. They can also raise parents' awareness of the importance of education for their children and parents may be more willing to have their children receive more and better education. That is, they are important omitted variables.

Dependent variable: PER (Preschool Education Ratio)

Regressor	(1)	(2)	(3)	(4)	(5)
CDR	-0.74411***	-1.22212***	-0.52572***	-0.56247***	-0.63351***
	(0.05681)	(0.07781)	(0.08086)	(0.10039)	(0.10458)
ODR	1.70051***	1.85307***	1.46720***	0.92874***	1.05654***
	(0.16698)	(0.19115)	(0.15984)	(0.24435)	(0.24946)
IR			-0.88882***	-0.35690***	-0.35960***
			(0.07206)	(0.80301)	(0.08262)
CE			-0.48985***	-0.23780***	-0.20540***
			(0.04470)	(0.05953)	(0.06092)
INF				-0.11401*	-0.13972**
				(0.05134)	(0.05231)
GDP				0.00613	0.01299
				(0.01412)	(0.01436)
GDP×CDR				0.00248***	0.00258***
				(0.00031)	(0.00031)
GDP×ODR				-0.00096***	-0.00130
				(0.00071)	(0.00073)
CR					0.14452*
					(0.06290)
(Intercept)	30.42157***				
	(2.91110)				
Province effects?	No	Yes	Yes	Yes	Yes
Time effects?	No	No	No	No	No

***F*-Statistic and *p*-value Testing Exclusion of Groups of Variables**

Province effects = 0		4.4382	5.8517	5.1561	4.4588
		(<0.01)	(<0.01)	(<0.01)	(<0.01)
Time effets = 0		1.3005	0.9758	0.6022	0.5691
		(0.1362)	(0.5304)	(0.9993)	(0.9999)
$\bar{R}^2$	0.4639	0.4656	0.64423	0.73744	0.73995

These regressions were estimated using panel data for 30 Chinese provinces. These regressions use data for 2001 to 2016. Standard errors are given in parentheses under the coefficients, and *p*-value are given in parentheses under the *F*-statistics. The individual coefficient is statistically at the 。10%, *5%, **1%,***0.1% significance level.

(Table 4)

In regression (4), two more control variables related to the economy are added. Real GDP per capita and the price level reflects people's ability to pay. Ability to pay will affect the work force's ability to support their children and old and in turn their willingness to give birth and support the old. Besides, with the economic development, medical conditions and living conditions are improved and the old's average life span is prolonged as a result, increasing the society's old dependency ratio. Therefore, the two variables are correlated with the dependency ratios. As the ability to pay rises, the work force is able to afford more education for their children, so the two variables are also correlated with preschool education ratio. We are also interested in the partial effect of dependency ratios, so we add two intersection terms (***CDR*×*GDP*** and ***ODR*×*GDP***) into the regression.

Hall combines the rational expectations theory with the life cycle theory and constructs the following random walk model:

$$\boldsymbol{C_t = C_{t-1} + u}$$

In which $\boldsymbol{C_t}$ represents present consumption, $\boldsymbol{C_{t-1}}$ represents previous consumption and $\boldsymbol{u}$ is the random disturbance term. This function indicates one's present consumption is affected by his previous consumption, meaning one's consuming behavior remains stable. If we replace the personal consumption ($\boldsymbol{C}$) with the society's consumption ($\boldsymbol{CR}$), this model will change into:

$$\boldsymbol{CR_t = CR_{t-1} + u}$$

Following this theory, we assume that consuming behavior for each person remains stable for all his lifetime, but one's income is varied in different stages of his life. More specifically, one's consumption ratio will be higher when he is a child or the old than when he has a job since the income differs. For a country or a region, this means the rise in the dependency ratios of one region will increase the consumption ratio of this region. We can add two more variables ***ODR*** and ***CDR*** into the model and it becomes:

$$\boldsymbol{CR_t = \beta_0 + \beta_1 CR_{t-1} + \beta_2 CDR + \beta_3 ODR + u}$$

Now we have known the two dependency ratios are correlated with consumption ratio. Besides, preschool education ratio will definitely be correlated with consumption ratio. Adding this omitted variable into the regression in (4), we finally get the regression in (5).

For regression (3) ~ (5), The *F*-statistic and corresponding *p*-value of testing fixed effects indicate the province effects are significant at the 1% level while time effects are insignificant at the 10% level.

IV. Conclusion

According to our common sense, we may speculate the coefficients of the children dependency ratio and the old dependency ratio will be both negative because the rise in either will increase the burden on the work force, and then they are unable to spare money for more education for their children. The coefficient of ***CDR*** in regression (5) is negative. However, the coefficient of ***ODR*** is positive. There is one reason associated with Chinese characteristics that may account for this unexpected result: In Chinese family, the old enjoy comparatively higher status than those in foreign family. They love their grandchildren, hoping they can receive better education, and work force will respect their determination and opinions. Some of them may even annuity or alimony to assist their schooling. So as the old dependency ratio increases, the preschool education ratio may, on the contrary, increase.

By Column (5), we can also conclude that for a fixed ***GDP*** (denoted as g_0), as ***CDR*** increases by 1%, ***PER*** will decrease by (0.63351 - 0.00258 * g_0)%. This result suggests that the rise in ***GDP*** will diminish the marginal decreasing effect of ***CDR*** on ***PER***.

V. Model Improving

The dependent variable can also be influenced by hysteresis effect. We want to introduce the lagged variables of ***PER***, ***CDR*** and ***ODR*** into the regression as control variables for ***CDR*** and ***ODR***. The upper bound of the lag order is set three since preschool education in general lasts for three years. The followings are the results of optimum lag order tests:

Lag length Selection for PER

Lag	AIC	HQ	SC	FPE
0	3.985262	3.992144	4.002764*	53.799395
1	3.980715*	3.991038*	4.006968	53.555336*
2	3.981992	3.995756	4.016995	53.623747
3	3.983709	4.000914	4.027463	53.715915

Lag length Selection for CDR

lag	AIC	HQ	SC	FPE
0	2.588837*	2.595719*	2.606339*	13.314281*
1	2.590944	2.601267	2.617196	13.342360
2	2.594305	2.608069	2.629309	13.387285
3	2.597993	2.615198	2.641747	13.436754

Lag length Selection for ODR

lag	AIC	HQ	SC	FPE
0	0.8726887	0.8795707*	0.8901905*	2.3933374
1	0.8735855	0.8839084	0.8998381	2.3954848
2	0.8726242*	0.8863882	0.9076277	2.3931837*
3	0.8758581	0.8930630	0.9196125	2.4009365

("" means the smallest value in the test, and the corresponding lag order is the best order in the test)*

(Table 5)

According to the table, the optimum lag order for ***CDR*** and ***ODR*** is 0 and for ***PER*** is 1. Then we only add the first-order lag of ***PER*** into the panel data regression. The new model is therefore constructed as follows (here we still assume individual fixed effect):

$$\widehat{\boldsymbol{PER}}_{\boldsymbol{it}} = \underset{(0.043278)}{0.222454} \times \boldsymbol{PER}_{\boldsymbol{1,it}} - \underset{(0.093519)}{0.388253} \times \boldsymbol{CDR}_{\boldsymbol{it}}$$
$$+ \underset{(0.162325)}{1.183038} \times \boldsymbol{ODR}_{\boldsymbol{it}} - \underset{(0.075105)}{0.682431} \times \boldsymbol{IR}_{\boldsymbol{it}}$$
$$- \underset{(0.058749)}{0.247148} \times \boldsymbol{CE}_{\boldsymbol{it}} - \underset{(0.120249)}{0.322424} \times \boldsymbol{INF}_{\boldsymbol{it}}$$
$$- \underset{(0.115120)}{0.350543} \times \boldsymbol{GDP}_{\boldsymbol{it}} - \underset{(0.068563)}{0.205668} \times \boldsymbol{CR}_{\boldsymbol{it}} + \boldsymbol{\alpha}_{\boldsymbol{i}} + \boldsymbol{u}_{\boldsymbol{it}}$$

$$\boldsymbol{SER} = 4.30552, \quad \bar{\boldsymbol{R}}^{\boldsymbol{2}} = 0.68602$$

*(**PER_1** represents the first-order lag of **PER**, **α_i** represents the individual fixed effect.)*

VI. Model defects

Generally speaking, real GDP per capita is affected by various factors so it is probable to be an endogenous variable. Once it is endogenous, its estimated coefficient will be biased, so that the coefficient of the intersection term will also be biased. The partial effect of dependency ratios therefore cannot be correctly measured using this model.

REFERENCES

[1]罗光强, 谢卫卫. 中国人口抚养比与居民消费——基于生命周期理论[J]. 人口与经济, 2013(05):3-9.

[2]李祥, 李勇刚. 人口抚养比、房价波动与居民消费——基于面板数据联立方程模型[J]. 经济与管理研究, 2013(01):35-41+68.

[3]**Hall, R. E.** Stochastic Implications of the Life Cycle-permanent Income Hypothesis: Theory and Evidence [J].Journal of Political Economy, 1978, 86 (6)