The Role of Income Distribution in Financing Human Capital Investment

V.K.Chetty
Basanta K.Pradhan

2020

IEG Working Paper No. 394

सत्यमेव परमो धर्मः

The Role of Income Distribution in Financing Human Capital Investment*

V K Chetty
Boston University, USA

Basanta K Pradhan
Institute of Economic Growth, Delhi

Abstract

We study relations between parents' income and children's college graduation, using data for 20.5 million children, born during 1978-83 in the U.S. or authorized immigrants, and their parents, mobility matrices for 741 U.S. commuting zones and Markov chain model for income distribution. Graduation rates are significantly and causally related to parents' income shares. In the short run, graduation rate increases when low income share increases for compensated decreases in middle or high income shares. But, in the long run, increase in middle income share compensated by a decrease in low or high income share increases the college graduation rate. In the long run, the middle income class has the dominant effect: a 1% rise in middle income share compensated by a 1% reduction in high income share increases college graduation rate by 0.626%. Impact of low income share is expectedly positive on the high school dropout rate. It falls with a compensated rise in middle or high income share both in the short and long run. Steady state analysis also indicates: (1) economic growth is positively correlated with investment in human (though at a decreasing rate) and social capital; and (2) surprisingly tax progressivity increases economic growth. College graduation rate is negatively associated with fraction foreign born, fraction divorced, fraction black and fraction religious both in short run and in long run. Middle income group is the main driving force, and income distribution and economic growth are intertwined in the determination of investment in human capital. Short-run regressions have causal interpretation for effects of income distribution as data for parents' and children's income distributions are separated by as long as 30 years. Directional consistency of steady state and short run results support the hypothesis that effects of parents' incomes on children's education are exogenous.

JEL Codes: I24, E22, J62

Key words: Human Capital, Parents Income, Income Distribution

*A previous version of the paper was accepted for presentation at the Royal Economic Society Annual Conference, which was to be held at Queen's University, Belfast (UK), during 6th and 8th April 2020. The authors would like to acknowledge Raj Chetty, Jonathan Skinner and Bruce Zellner for their valuable comments and suggestions.

1. Introduction

Technological progress and accumulation of physical capital in the last century have led to significant economic growth in the U.S. During the forty years, spanning from 1959 to 2000, per capita personal real disposable income increased by 167 per cent, and the poverty rate has halved for all Americans (Bureau Of Economic Analysis, (2018b)). But, the inter-generational distribution has been uneven. Share of aggregate income of the fifth quintile increased from 41.1 to 47.7 per cent while, correspondingly, the shares of the remaining four quintiles decreased(Bureau Of Economic Analysis, (2018a)). In more recent years, from 2000 to 2010, the poverty rate has increased by 2 per cent for all Americans, in spite of the increase in per capita income from $31,524 to $35,685 during the same period (Census(2017)).

Recent literature also points in the same direction. Probability of children exceeding parents' income decreased from 92 per cent for children born in 1940 to a coin flip of only 50 per cent for children born in 1984, (Chetty *et al*(2017)).While accumulation of capital has contributed to growth, the benefits have accrued more to the rich, perhaps due to lack of human capital for middle and low income groups. Hence an interesting question is: How to increase the human capital in the American economy?

Capital often needs large investments requiring groups of individuals to finance and share both risks and benefits. Mechanisms to finance capital accumulation depend on the nature of capital. Some capital such as bridges and land are geographically fixed with limited markets for their services. Human capital including education and health are embodied in individuals. Human capital cannot be traded but it is easily movable with international markets for their services. Ownership of physical capital and associated risk is transferable through financial markets. Investment in physical capital can be privately financed by stocks and bonds or publicly financed by taxes and bonds.

Features of human capital consisting of education and health are profoundly different. The gestation period and durability are much longer and returns are too risky, exacerbated by rapidly changing technologies. It is embodied in individuals. Unlike physical capital, human capital cannot be sold in financial markets; only services of human capital can be sold in markets. Lack of markets for human capital assets increases the investment risk and could lead to under investment.

Children cannot borrow money to go to school. Parents cannot borrow using children's future earnings as collateral. Hence the "wall street" cannot help to finance human capital. Only parents and the "main" street must bear the risk and finance education. Moreover parents have most information about their children and their monitoring will be more effective than the markets, thereby reducing the impacts of both adverse selection and moral hazard. Returns through services are distributed over the life of the capital increasing the risk of obsolescense and recovering the invested resources.

Due to unrestricted interstate commerce and possibilities to move, public investment by state governments in human capital will be very limited- a negative consequence of free trade! Hence, investment in human capital must be financed from own savings or through secured loans. This increases the role income distribution in market economies.

When there is unequal distribution of income, poor parents cannot pay of their children's education. With less investment in education, children become poorer leading to a "vicious circle" of poverty. As Charles Darwin aptly observed centuries ago, "Man accumulates property and bequests it to his children, so that the children of the rich have an advantage over the poor in the race for success" (Darwin(1871)). Thus equitable income distribution will help to increase growth, creating a"virtuous circle" of distribution and growth.

Therefore it is imperative to investigate the short and long run relations between economic growth,income distribution and accumulation of human capital and test the hypothesis that investment in human capital and economic growth is positively related to the share of middle income. In the short run, capital is fixed limiting the supply elasticity of goods and services. Similarly, short and long run income elasticities of demand for goods and services will also different. Hence it will be useful to estimate short and long run relations.

Variations in the size distribution of income in the 741 commuting zones in the U.S., the corresponding mobility matrices and several commuting zones' characteristics such as high school dropout rates and college graduation rates in the publicly available data set provided by the 'equality of opportunity' project are useful to estimate both short and long run relations and test this hypothesis.

2. Model and Methods

a. Model

Since investment in education depends both on (micro)family characteristics and (macro) policies, our model involves two components, one for investment behavior at the micro level and the other for aggregate short and long run relations between human capital and income distributions at the level of commuting zones (CZ), i.e., at a macro level.

Let i refer to individuals, j to commuting zones and n_j to the number individuals in j . The dichotomous variable, $c_{i,j}$,= 1, if the individual is a college graduate and 0 otherwise. Similarly, $s_{i,j} = 1$,if the individual is a high school graduate and 0 otherwise. Parents income is grouped into three categories, low, middle and high and denoted by three indicator variables, $l_{i,j}$, $m_{i,j}$ and $h_{i,j}$. Probability of graduating from college depends on several factors including income of parents and other demographic and socio-economic variables

relating to the commuting zone,z_j.The linear regression model,without an intercept, for college graduation is:

$$c_{i,j} = \alpha_j l_{i,j} + \beta_j m_{i,j} + \gamma_j h_{i,j} + \delta_j z_j + \epsilon_{i,j} .....$$

Summing over i and dividing by n_j, we have the following aggregate relation for commuting zone j,

$$c_{.j} = \alpha_j l_{.j} + \beta_j m_{.j} + \gamma_j h_{.j} + \delta_j z_j + \epsilon_{.j} .....(1)$$

Note that

$$l_{.j} = \frac{\sum_{i=1}^{n_j} l_{i,j}}{n_j},$$

is the fraction of parents in the low income group. Similarly, $m_{.j}$ and $h_{.j}$ are the fractions of parents in the middle and high income groups.

If the commuting zone characteristics are adequately represented by the variables z_j , we can assume that the regression coefficients for the income categories are the same for CZs. We can then estimate the parameters using the cross sectional variations among fractions of college graduates and shares of income in CZs.

This is a short run, dynamic relation between children's human capital formation and parents incomes. We estimate this relation using data relating to parents incomes, their children's educational attainments and the CZ characteristics.

Suppose this dynamic process continues over time and reaches a steady state. In a steady state, income distributions will be the same over time and among parents and children. Hence, income distributions of parents and children of the same time period (hence not their children) can be used to estimate long run relations between fraction of college graduates and income shares using the same regression model (2) and aggregate crosssection data at the CZ level. The long run model is

$$c_{.t} = \alpha_t l_{.t} + \beta_t m_{.t} + (\gamma_t) h_{.t} + \delta_t z_j + \epsilon_{.t} = \alpha_t + (\beta_t - \alpha_t) m_{.t} + (\gamma_t - \alpha_t) h_{.t} + \delta_t z_j + \epsilon_{.t} .....(2)$$

Note that $c_{.j}$ is the fraction of college graduates in the commuting zone and hence an estimate of the (marginal) probability of children from the commuting j graduating from college. Similarly, $l_{.j}$, $m_{.j}$ and $h_{.j}$ are conditional probabilities of belonging to the low, middle and high income respectively. Then the probability of graduating from college is a weighted average of the conditional probabilities, $l_{.j}$, $m_{.j}$ and

$h_{.j}$,of belonging to low, middle or high income groups. In the linear regression model, the weights are the regression coefficients.

Probabilities of belonging to an income group can be estimated as steady state or long run probabilities of belonging to the income groups using the mobility matrix and the steady state formula as indicated below in equation (2).

The difference between models (1) and (2) is that (1) relates to individual behavior requiring micro data while in (2) both dependent and independent variables are aggregates and statistical relation exist even if they do not exactly correspond the cohorts of parents and children. The income distributions in a steady state are the same for both parents and children. Hence the aggregate model in (2) can be used to estimate the aggregate effect of policies. Even if the aggregate relation is based on *irrational* behavior of individuals as in Houthakker(1955) and Becker(1962), it will still be useful for policy purposes as long as aggregate proportions respond to policies.

Any stable relationship between policy and behavioral variables such as income tax, interest, poverty, unemployment and savings rates, public expenditures, income distribution and mobility, can be estimated using aggregate variables and can be used to study the effects of policies. For example, a fiscal policy that reduces the size of the poor group from 20% to 10% is better than another policy that reduces only to 15%, even without further micro details about the 5% of the group that benefit from the policies.

We model income distribution among successive generations as a stochastic process using the finite Markov chain. (see Solow(1951), Loury(1981) and Aiyagari(2002)) and calculate the long run shares of three income groups (See Appendix Section 2).

b. Methods

The mobility matrix provides the conditional transition probabilities. The marginal income distribution associated with the mobility matrix provides the short run shares of the income classes. A mobility matrix is regular if all the elements, when raised to some power, are positive. For regular Markov chains, the transition probabilities converge and become independent of the initial states over time. This is the unique long run steady state for the size distribution of income. The steady state income distribution is the unique solution to a set of linear equations (see Taylor and Karlin 1998, Theorem 1.1, p.204). We used the R-functions "solve" and "Markov" (given in the Appendix) to calculate the long run shares.

Transition matrices are conditional probabilities and refer to individuals and do not require a marginal distribution. A regular transition matrix determines a unique steady state distribution. Since the converse is

not true, a steady state income distribution may correspond to more than one transition matrix.

Technologies and behavioral relations may change over time. Even under such circumstances, aggregation over a large number of individuals may have some regularities (Houthakker(1955),Becker(1962)). Even if individuals switch between income groups, the fraction of individuals in any income group will converge in the long run.[1] Hence it may not be possible to make inferences about individual behavior from aggregated models.

While estimates from cross section data are often interpreted as long run estimates, Markov chain model provides a formula to calculate long run (steady state) size distribution of income using the short run estimates from the cross section data.

The population of parents and children are divided into three groups using the five quintiles from the Opportunity Insight data set: the first quintile is designated as low income group; the next three quintiles are combined into the middle income group and the fifth quintile as the high income group. Here income intervals are used to designate the groups. In the short run, the income distributions may differ across generations. In the steady state they converge to one distribution. Note that the dollar intervals of income groups may change over time. For example the income limits of middle income group is distinctly higher in 2010 than those in 1980. Therefore the dollar intervals in steady state will be greater than the intervals in 2010.

In the literature, income distribution is described by observed or calculated fractions of individuals in fixed income intervals. Thus, the middle income share is the fraction of the population with incomes in intervals corresponding to the definition of middle income such as the intervals defined by the second, third and fourth quintile of the national income distribution. Then, by definition, 60% of the national population will belong to the middle income group. However, for any subgroup of the population such as females or Hispanics, share of population in the middle income group may vary from 0 to 1.

Income distributions in the short run can be determined empirically from observed data. However, for the long run or steady states, income distributions must be determined as limits of short run income distributions. Mobility matrices describe the transitions from a given income group, such as the middle income, to each possible income group from low to high. The transition probabilities are then used to calculate the limiting distribution of any given group of individuals into the fixed income intervals. This distributes individuals

[1]For example, in a model with only two income groups, poor and rich, individuals will be switching between the two groups with statistical regularity. Suppose the transition probabilities for upward and downward mobilities are 0.6 and 0.3 respectively. Then, the steady state probabilities for the poor and rich states are $0.3/0.9 = 1/3$ and $0.6/0.9 = 2/3$. An individual in the poor group can remain poor in the next period or move up and return to poor state later in one, or two or more periods. The mean period of return to the poor state is $1 \div (1/3) = 3$ periods (See Appendix Section B for details). The model is consistent with the old adage, "From shirt sleeves to shirt sleeves in three generations" or as the Scottish say "The father buys, the son builds, the grandchild sells, and his son begs."

into an exclusive and exhaustive set of income intervals. Thus it provides a distribution of the individuals as well as the distribution of the income. In other words, steady state transition matrix provides both long run shares of the population and the total income of the group.

Further, it should be noted that the distributions refer to both population and income. Group size may change when people shift from one group to another. When income of middle income group increases, within the defined limit, the size of the group increases in terms of income, making an impact on the number and the fraction of graduations in each group. As previously mentioned group limits in terms of income levels may keep changing with time; however, at each point of time the groupings are defined in those fixed income limits.

Further, it should be noted that the distributions could refer to both population and income. But the data from mobility matrices refer to only changes in population shares. Group size may change when people shift from one group to another. When income of middle income group increases, within the defined limit, the share of the group increases in terms of income, making an impact on graduation rates. As previously mentioned group limits in terms of income levels may keep changing with time; however, at each point of time the groupings are defined in those fixed income limits.

For any regular transition probability matrix, the long run shares can be calculated approximately by raising the matrix to a sufficiently high power. An exact calculation of the shares can be determined by solving a system of linear equations. (See Taylor and Karlin (1998, Theorem1.1, p.204)). An R-function to compute long rung run shares and some discussion of two simple transition matrices are given in the Appendix Section C.

3. Data

We use data from (Chetty et al. 2018)relating to 20.5 million children in the 1978-83 birth cohorts who were (1) born in the U.S. or are authorized immigrants who came to the U.S. in childhood and (2) whose parents were also U.S. citizens or authorized immigrants. The units of observation for the statistical model are 741 commuting zones in the U.S. Commuting zone population varies widely from 881 to 17.8 million with a mean of 410,811 and a median of 108,337. Descriptive measures of variables used in the analysis are given in Tables 1 to 3. Large variations among the 730 commuting zones within the U.S. help to estimate statistical relations and identify significant variables.[2] Definitions of variables and data sources are given in Table 7, Section 4 of the Appendix.

[2]For example, Gini ratios vary from 0.2 to 0.85 between the commuting zones.

Table 1.Selected Descriptive Statistics of 741 Commuting Zones		
Variable	Mean	Std.Dev.
Gini Ratio	0.406	0.081
College Graduation Rate	0.389	0.145
High School Dropout Rate	0.041	0.023
College Graduates per 10,000	1363.747	444.105
Fraction Black	0.078	0.122
Racial Segregation	0.130	0.100
Fraction Religious	0.546	0.167
Social Capital Index	0.172	1.295
Fraction Foreign Born	0.041	0.050
Fraction AdultDivorced	0.097	0.018
Teacher Student Ratio	16.508	2.429

Table 2.Income Distribution of Parents and Children			
Group	Income_Shares	Mean	Std.Dev.
Parents			
	Low	0.225	0.093
	Middle	0.654	0.076
	High	0.121	0.066
Children			
	Low	0.190	0.056
	Middle	0.607	0.033
	High	0.203	0.059
Steady State			
	Low	0.176	0.057
	Middle	0.595	0.038
	High	0.229	0.070

Summary measures of income distributions among low, middle and high income groups for parents, children and steady states are given in Table 2. Middle income population share (for parents in 1980) decreases from 65.4% to 60.7% for children in 2010. The estimated steady state distribution (same in steady states for all generations making spatial income distributions among parents and children invariant through time) is 59.5%.

Group	*Table 3. College Graduation, High School Dropout and Income Distribution of Parents and Children*				
	Income	Col.Grad.Rate		HS.Drop.Rate	
Parents					
		corr.	t.value	corr.	t.value
	Low	-0.409	-9.709	0.411	9.762
	Mid	0.257	5.760	-0.487	-12.09
	High	0.301	6.838	-0.035 #	-0.766
Children					
	Low	-0.363	-8.443	0.588	15.739
	Mid	0.154	3.363	-0.130	-2.831
	High	0.254	5.707	-0.474	-11.655
Steady State					
	Low	-0.261	-5.865	0.602	16.352
	Mid	0.120	2.618	0.0242 #	0.526
	High	0.145	3.175	-0.488	-12.128
			,	# Not Sig.	Rest High. Sig.

In Table 3, simple correlations between income shares and college graduation and high school dropout rates are given for the three income groups of parents and children. There is significant negative correlation between fraction of poor income parents and college graduates rates. This correlation is again significant and positive for high school dropout rates. Opposite relations hold for middle and high income shares.

In the 'equality of opportunity' project, Chetty(2014) have provided the quintile-quintile transition matrices of child and parent incomes for 21.3 million children and parents using the 1980-85 birth cohorts for all commuting zones (See Appendix Table 4 for the definition of variables).

Figure 1.Reaching American Dream: Variations among States

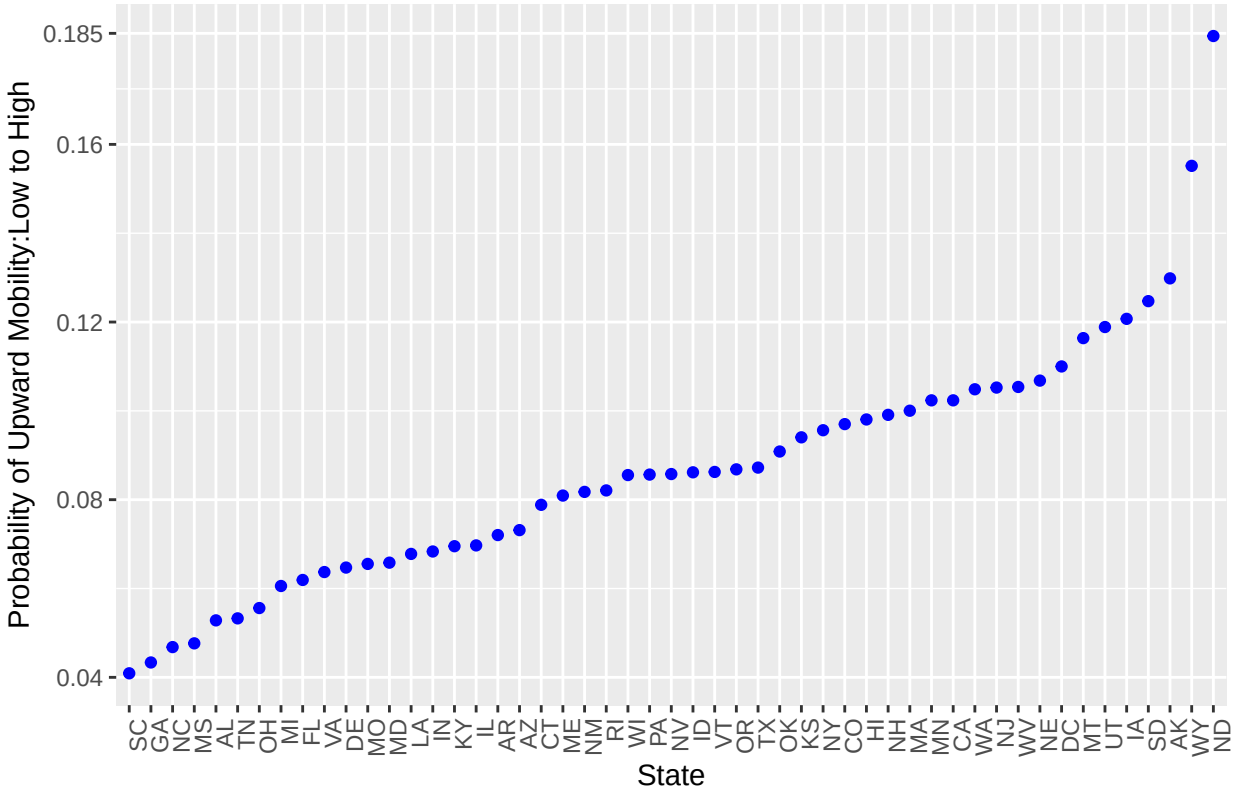
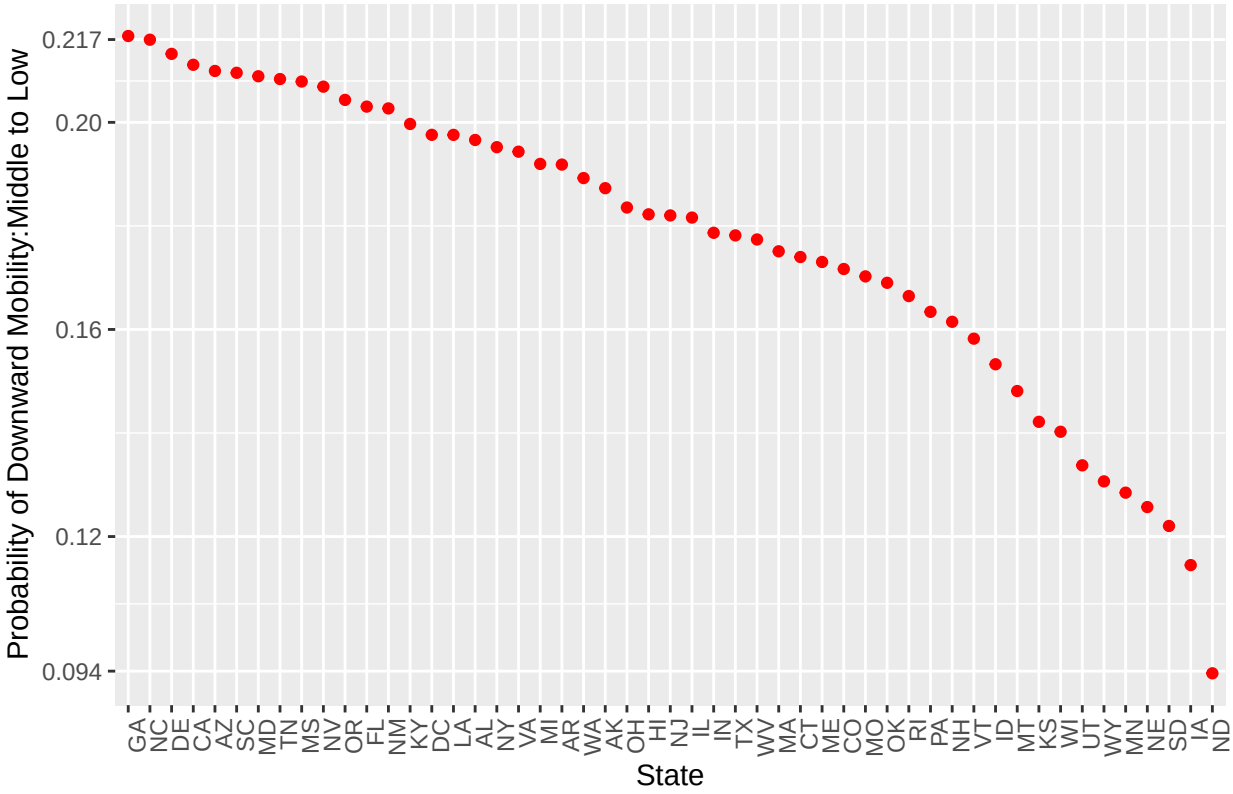


Figure 2. Shrinking Middle Class: Variations among States



The transition (mobility) matrices, available at community zone level, provide the probability that the child has family income in a given quintile of the national child income distribution (in the terminal year) conditional on having parents with family income in a given quintile of national parent income distribution (in the initial year). We have aggregated these transition probabilities to the state level using commuting zone population as weights.

In Figure 1, the conditional probabilities of reaching the fifth quintile from the first quintile (the American Dream) across the states are shown. It varies from a minimum of 4% in South Carolina to a maximum of 18.5% in North Dakota. In 25 states, the probability is less than 8.6%. The top five states are: Utah, Iowa, South Dakota, Arkansas and North Dakota while South Carolina, Georgia, North Carolina, Missouri and Alabama are at the bottom.

Along the similar lines, in Figure 2, probabilities of moving down from the middle class (quintiles 2, 3 and 4) to low income group (quintile 1) are shown. The maximum downward mobility is 21.7%, in Georgia, while the minimum, 9.4%, is in North Dakota. North Dakota is performing much better than any other state in both the measures.

4. Empirical Results

a. Short Run Relations

Using linked data for 20.5 million children and their parents, we estimate the following regression models. The dependent variables are proportions of the following variables at the CZ level: college graduation, high school dropout,labor force participation, teen age abortion, single mothers, adults married, adults divorced and religious persons. Independent variables include parents income ranks, their family characteristics such as family structure, measures of health and social capital and some CZ characteristics. Estimates of regression coefficients and p-values for all regression models are given in the Appendix-section 1.

Results of the regression model of college graduation are given in Appedix A. Variables relating to income distribution are the intercept(relates to low income class i.e. the lowest quintile) and the coefficients of proportions of middle income (three middle income quintiles) and high income (the fifth quintile). As pointed out in the model section, the intercept is the marginal effect of the low income parents on college graduation and the coefficients of middle income and high income parents are the *additional* effects.

Note that the income distribution in each CZ is based on exogenously specified income intervals determined by the national income distribution. Hence, increasing the share of one group, say the middle income group, holding the population share as well as per capita income level of the high income group constant, will require more income. This is feasible since the incomes vary with CZs with variations in income distributions. Thus the CZ level variations in income and its distribution helps study the effects of growth and distribution. We can interpret the regression coefficient of the middle income share as the marginal effect increasing the share of middle income with constant high income share while the low income decreases, but that is regressive. In other words, this represents the effect of economic growth when the entire benefit of growth goes to the middle income group.

The estimated short run effects of incomes on college graduation are: 0.478 for low income, 0.478+0.317=0.795 for middle income and 0.478+0.297=0.775 for high income parents. In other words, if 1 % of low income parents are moved to the middle income, the proportion of college graduates will increase by 0.00317 or there will be a 0.317% increase in the college graduates. College graduation rate is also expectedly positively associated with teacher-student ratio and college tuition fees, and negatively associated with fraction foreign born, fraction divorced, fraction black and fraction religious, both in short run and in long run.

Effects of income distribution on the high school dropout rates is given in Appendix Table 5. Estimated short run effects are : low income: 0.099, middle income: -0.001 and high income: 0.069. Increasing the middle

income share by decreasing either the low income or the high income share decreases the high school drop put rates. All effects are statistically highly significant.

The impact of middle class on human capital investment is higher than even the high income group. If income shares of all three groups increase due to economic growth, the positive impact of middle income group will dominate. High school dropout rate is also positively associated with racial segregation, fraction divorced and fraction black, and negatively associated with income growth and fraction religious, both in short run and in long run.

The impact of middle income group is similar to that of the high income group on graduation rate though in the long run it is more than double. The reason could be the inelastic supply side effects. Though middle income group demand more enrollments, the space is limited as the demands of high income group are preferentially met. For high school dropout rate the relative values of coefficients are exactly the reverse; in short run the middle income impact is about three times of the high income group impact, and in the long run the coefficients are almost equal. In the short run drug problem may be a bigger issue for the high income group kids forcing them to drop out in large numbers in the short run. Whereas middle income kids may be dropping out to take up jobs in the long run.

b. Long Run (Steady State) Relations

Statistical analysis presented in the previous section provide evidence for significant effects of parents income distribution (income distribution in the short run) and other characteristics of CZs on high school and college education. Since there is a long time lag of 15 to 20 years between childrens' graduation and parents' incomes, we can assume that the observed relations are causal and not merely correlations. Hence parental income is one of the causal factors for childrens' educational outcomes. Parents' in the low income percentiles will have little savings to finance childrens' education, while parents' in the high income group will have more than adequate savings to finance their (two?) children. Children of rich parents' are more likely to accumulate more human and social capital and are more likely to move up in the income ladder. When capital helps to accumulate more capital, the rich will have a better chance to climb up the income ladder while the poor may remain poor or may even become poorer in the long run.

According to the OECD report OECD(2018), the number of generations it would take for those born in the low-income families (bottom 10% of the income distribution) to approach the mean income in their country varies from 2 for Denmark to 11 for Columbia. The expected number of generations for the U.S. according to this study is 5.

Parents' incomes affect their childrens' education and hence their incomes. With rapidly changing technologies, the dynamic relation of incomes across generations is more likely to be stochastic. Causal relations of interventions with a single response can be estimated using randomized experiments. When responses are many and distributed over many generations, quasi-experimental studies over many generations are not feasible. Long run estimates of economic policies have to be based on time series of observational data and models appropriate to analyse such non-experimental data.

We estimate some policy relevant relationships using mobility matrices, observational data and a Markov chain model for the American economy. In a steady state, income distributions and relations among variables do not vary. Hence, income distributions of parents' of any generation will have the same relation to income distributions of children of all generations. This helps to estimate relations between income distributions and education using data relating to the same generation, i.e., cross section data.

Statistical relations estimated using aggregate data at the CZ level may or may not correspond to the observed micro relations at the household level. Parents' and their children may not have the same incomes. But they will have the same *probability* of being in any income quintile. Even such relations are useful to compare policies. A transfer or a public investment policy that increases the probability of being in the middle or high income group, is always desirable.

Using data for the years 1940 to 1985 from Chetty et al (2017) and the formula for steady states for Markov chain models (see Appendix, section 3), we calculated the long run shares for low, middle (consisting of second, third and fourth quintiles) and high income groups as follows. Transition probabilities between income groups for 741 commuting zones are taken from the same source. The steady state income distribution is the unique solution to a set of linear equations Taylor(1998) (see Theorem 1.1, p.204). We used the R-functions "solve" and "Markov" (given in the Appendix- section 3) to calculate the long run shares.

We have estimated (population weighted) OLS linear regressions to test four hypotheses. As noted earlier, savings and investment behavior of the middle income group is likely to play a crucial role in the accumulation of human capital.

Our first hypothesis is that the college graduation is positively related to the share of middle income group in the total population and that the middle income group contributes more than low and high income groups after adjusting for other covariates related to costs, race and social capital. The estimated regression model is given in Appendix Table 6. The coefficient of low income share is statistically insignificant. Only middle and high income shares are statistically significant. If 1 % of low income parents are moved to the middle income, the college graduation rate increases by 0.0107 + 0.00047 = 0.01117 or by 1.117%. The long run

response is almost 3.5 times the short run response, 0.317%. The corresponding increase for high income group is 0.487%, 43.6% of the middle income effect.

Substituting middle income group share by (1 – low income share – high income share), we have,

$$cgrate09 = 1.023 - 1.07 low.inc.share - 0.626 high.inc.share$$

$$-0.787 ccddrpout912 + 0.008 TeacherStudentRatio - 0.337 FracForeignBorn$$

$$-3.031 FractionAdultsDivorced + 0.015 CollegeTution - 0.100 FracBlack - 0.261 FracReligious$$

One of the noteworthy determinants of college graduation rate is the population share of the middle income group (see the first version of the estimated coefficients in Appendix Table 6).

The equation with low and high income shares is even more interesting. Increasing the share of high income or the low income or both at the cost of the middle income share reduces the fraction of college graduates. Most of the savings of the high group may be going to real estate, gold and stock markets.

Our estimates indicate that increasing the share of both low and high income groups by decreasing the middle income group decreases the graduation rate. Thus decreasing both the tails of the income distribution and increasing the middle income share is always more efficient for increasing the supply of college graduates in the population. Therefore policies are required to lift people from low income group to increase the size or the middle income group. This may require taxing the high income group.

Chetty(2018) have reported, "Atlanta and Charlotte have had exceptionally high rates of job and wage growth over the past two decades, yet have among the lowest rates of upward mobility for children who grow up there. These cities achieve high rates of economic growth despite offering local residents limited prospects for upward income mobility by importing talent – i.e., attracting talented individuals to move in and fill high-paying jobs." For example, median household income increased by 67 % from $39,751 to $66,533 with a job growth rate of 50% while share of individuals below the federal poverty line *increased* from 8.186% in 1990 to 12.475% in 2010, i.e., high growth is neither generating enough human capital nor reducing poverty.

Economic growth by *importing human capital* without change in income distribution favoring the resident middle income group will have a small effect on human capital formation. High positive effects of middle income shares imply that increasing the share of low income by decreasing either middle or high income will decrease the fraction of college graduates. Hence, it is social welfare improving if individuals from low income group can be pushed up to middle income group.

Another interesting result is that a 1% drop in the high school dropout rate increases the college graduation rate by 0.787%.

While the share of population and income of the middle class is central to increasing human capital, there are other variables which are equally significant in terms of their effect on the graduation rate; however, they are not amenable to economic policies. For example, the coefficient of the fraction of divorced adults is especially large. But this is a variable largely determined by social structure. Similarly, religiosity is negatively correlated with college graduation rate.

Our second hypothesis is that the high school dropout rates are positively related to the share of low income group. The estimated regression model is given in Appendix Table-7 below.

Low income share is positively related to high school dropout rates, while middle and high income shares are negatively related, supporting our hypothesis. An increase in the middle income share can indirectly accelerate the graduation rate by reducing the school dropout rate. Income growth without a change in income distribution has almost no effect on high school dropout rates. Again, fraction of black students and religiosity are positively related to high school dropout rates.

As in the case of the regression model for college graduates, decreasing both low and high income shares increase high school graduation rates.

The third regression model is for the stock of human capital which is crucial for determining both growth and distribution of income. The estimated regression coefficients are given in Table 8 below.

Again the middle income share is an important determinant of the stock of human capital. From Table 6,A 1% increase in the middle income share increases the number of college graduates approximately by 111 per 10,000 undergraduates or 111*0.1363=15.1293 per 10,000 population. For an equivalent increase, the household income per capita should be increased by $111 \times 0.01363/0.052 = 290.$ per capita or 2,909,481 for 10,000 population. This implies a cost of $192,308 per college graduate.

From Table 8,a 1% increase in the high school dropout rate decreases the number of graduates by 18. As before, religiosity is negatively related to stock of human capital.

In the fourth regression model, we estimate a relation between household income per capita and various types of private and public capital.

This is a quadratic relation with respect to the proportion of college graduates, i.e. per capita income rises but at a decreasing rate as human capital increases in the economy. The relationship is similar with respect to health capital. Social capital impact is positive. This result shows the importance of balancing the stocks

of various types of capital. Public policy, as reflected here in terms of government expenditure and income tax progressivity, which positively affects the distribution parameters in the economy, influences the per capita income in a similar fashion.

5. Discussion

Productive activities generate functional distribution of incomes. Savings and investment behavior of consumers and producers determine their size distribution and the supply of and demand for physical and human capital. Technological change accompanied by consistent investment in physical and human capital increases the income level. Inadequate investment in human capital decreases income especially for the middle income working class leading to less savings and less investment in human capital and thus creating a downward spiral.

Arthur Okun(1975,p80-81) has remarked concerning the modern U.S. economy that " . . . The most important consequence (of an imperfect loan market) is the inadequate development of the human resources of the children of poor families-which I would judge, is one of the most serious inefficiencies of the American Economy today."

In their extensive study of the race between education and technology, Goldin and Katz(2009) have shown that "in the race between technological change and education, education ran faster during the first half of the (last) century and technology sprinted ahead of limping education in the last 30 years". Therefore policies to increase human capital to catch up with the technological progress are necessary. According to them, two factors appear to be holding back the educational attainment of many American youth: (1) Lack of college readiness of youth who drop out of high school and (2) financial access to higher education especially lack of family resources and financial aid and educational loans. Our results are consistent with their findings and indicate that by increasing the share of the middle income group we can address both the problems.

Two significant results emerge from the analysis of the results reported in Tables 6 and 9. First,if 1 % of low income parents are moved to the middle income, the college graduation rate increases by 1.117%. Second, for optimal growth of per capita income, number of college graduates should be increased from the median value of 12.6 to 25.3 per 10,000 persons. This supports one of the suggestions of Goldin and Katz (2009): "A modest increase in tax rates at the very top end of the income distribution can provide revenue to fund payroll tax relief for lower wage workers." However, we suggest that fiscal policy must be used only to make short run changes to income distribution i.e. to increase middle income share in a graduated way to taper off such that the economy will function free of fiscal intervention in the long run.

When resource requirements for human capital were less, financing school education through local taxes and decentralized decisions was both feasible and efficient. Though income inequality among regions led to inequality in per pupil expenditures, and perhaps had some effect on the quality of college graduates, it did not result in shrinking middle class. This might be due to the fact that resources for high school education were adequate for most jobs in the first half of the last century. But modern higher education requires large and risky investments, both in physical and human capital, resulting in financing problems.

As pointed out earlier, it involves risks with moral hazards and adverse selection. Risk must be borne by students and their parents or by the public exchequer. In the case of human capital involving fifteen to twenty years of education and thirty to forty years for returns, the risk is much greater. Markets only for its services, and not for the asset itself, adds to the investment risk. For example, if markets exist only for taxi services and not for either permanent or temporary ownerships of automobiles such as zip cars, then the investment in automobile production will go down. Hence, investments in human capital must be financed directly by families with insiders' information about children's abilities and \ or by public resources.

With national and international markets for human capital services, fiscal and financial policies, both at the federal and state levels, are necessary to increase the stock of human capital. When fiscal policies help to increase human capital, the size of the middle income will increase, decreasing the need for income transfers, thus creating a *'virtuous' cycle*. When such an economy reaches a steady state, growth rate and the stocks of various types of capital will be consistent and no further transfers will be needed for increasing human capital. This steady state will be more efficient with lesser effects of adverse selection and moral hazard.

There are a variety of fiscal interventions which decentralizes financing of college education. Income transfers through vouchers can increase human capital only when the recipients have adequate awareness to make use of the vouchers, i.e., *when the recipients have necessary human capital*. To avoid this "catch22" situation, fiscal policies must also allocate resources to educate the poor to make proper use of the vouchers! Financing of college education will also be decentralized when it is through means tested loans and income-contingent repayments.

Our regressions also show that growth and distribution of income are interdependent, i.e., human capital depends on the middle income share and middle income share and per capita income depend on human capital.

In all probability Gandhi foresaw, though in a different setting, that whenever human capital is less, income distribution will be worse.

"Machinery has its place; it has come to stay. But it must not be allowed to displace the necessary human

labor. An improved plough is a good thing. But, if by some chance, one man could plough up by some mechanical invention of his the whole of the land of India, and control all the agricultural produce and if the millions had no other occupation, they would starve, and being idle, they would become dunces, as many have already become. There is hourly danger of many more being reduced to that unenviable state. I would welcome every improvement in the cottage machine, but I know that it is criminal to displace the hand-labor by the introduction of power-driven spindles (Emphasis added)... Gandhi(1925) (Young India, November 5, p. 377). unless one is at the same time ready to give millions of farmers some other occupation in their homes." **"My machinery must be of the most elementary type which I can put in the homes of the millions."** Gandhi(1934) (Harijan, November 2, pp.301-02).

There could be no better means of production (types of capital) than education and health to be put in every home! As human capital is embodied within and spread across individuals (it cannot be even concentrated with a limited group of persons by its very nature), the growth process will throw up a relatively larger size middle class.

Our work may be extended in several directions. First, it will be useful to study the relation between human capital and income distribution using other data sets such as the 'panel survey of income dynamics'.

Second, our results show the crucial roles played by the middle income group in the accumulation of human capital. For the growth of other capitals such as infrastructure and social capital, high and low income groups may be more important. Income distributions and the mobility matrices for 741 commuting zones from the Opportunity Insights Project will be useful to study these questions.

Third,long run relations depend crucially upon mobility matrices. The transition probabilities are determined by a combination of public policies, market forces, initial endowments and nature itself. It will be useful to study the relations between policy instruments and transition probabilities using the "opportunities insights project' data which has large variations.

Appendix

1. Regression Models

Table 4.Regression of College graduation rate-Short Run

Variables	Estimate	Std.Error	p.value
(Intercept)	0.478	0.113	0.000
ccd_drpout912	-0.799	0.314	0.011
parent.mid	0.317	0.123	0.010
parent.high	0.297	0.086	0.001
TeacherStudentRatio	0.005	0.003	0.080
FracForeignBorn	-0.291	0.101	0.004
FractionofAdultsDivorced	-2.927	0.388	0.000
I(CollegeTuition/1000)	0.012	0.002	0.000
FracBlack	-0.060	0.058	0.296
FractionReligious	-0.236	0.045	0.000
Adj.R-sq	0.380	d.f	433.000

Table 5.Regression of High School dropout rate-Short Run

Variables	Estimate	Std.Error	p.value
(Intercept)	0.099	0.012	0.000
parent.mid	-0.100	0.013	0.000
parent.high	-0.030	0.013	0.020
RacialSegregation	0.025	0.009	0.004
FractionofAdultsDivorced	0.160	0.05	0.002
FracBlack	0.013	0.008	0.109
IncomeGrowth2000_2006_10	-0.143	0.075	0.056
FractionReligious	-0.015	0.006	0.008
Adj.Rsq	0.310	df	589.000

Table 6.Regression of college Graduation Rates in 2009:Long Run

Variables	Estimate	Std.Error	p.value
(Intercept)	-0.047	0.201	0.815
ccd_drpout912	-0.787	0.322	0.015
mid.inc.share	1.070	0.25	<0.001
high.inc.share	0.444	0.189	0.019
TeacherStudentRatio	0.008	0.003	0.005
FracForeignBorn	-0.337	0.08	<0.001
FractionofAdultsDivorced	-3.031	0.396	<0.001
I(CollegeTuition/1000)	0.015	0.001	<0.001
FracBlack	-0.100	0.05	0.047
FractionReligious	-0.261	0.05	<0.001
Adj.Rsq	0.382	df	433

Table 7.Regression of High.Sch.Dropout Rates in 2009:Long Run			
Variables	Estimate	Std.Error	p.value
(Intercept)	0.177	0.024	<0.001
mid.inc.share	-0.182	0.031	<0.001
high.inc.share	-0.187	0.023	<0.001
RacialSegregation	-0.011	0.007	<0.001
FractionofAdultsDivorced	0.100	0.048	0.039
FracBlack	0.029	0.007	0
IncomeGrowth2000_2006_10	0.027	0.100	<0.001
FractionReligious	0.008	0.006	0.199

Table 8 .Regression of College graduates per 10,000 Population			
Variables	Estimate	Std.Error	p.value
(Intercept)	-583.916	210.099	0.006
ccd_drpout912	-1817.983	543.817	0.001
HouseholdIncomepercapita	0.052	0.002	<0.001
mid.inc.share	810.691	303.692	0.008
FractionReligious	-347.835	75.226	<0.001
Adj.Rsq	0.554	df	592

Table 9. Regression of per capita Household income on Education, Health, Social and Public Expenditure			
Variables	Estimate	Std.Error	p.value
(Intercept)	18907.222	1348.467	<0.001
col.grads.prop	118036.846	19636.821	<0.001
col.grads.prop.sq	-225198.222	62881.502	<0.001
TotMD.10K	156.766	28.447	<0.001
TotMD.10K.sq	-1.026	0.303	0.001
SocialCapitalIndex	799.532	116.958	<0.001
cgp.LGEP.SITP	0.578	0.221	0.009
Adj.Rsq	0.496	df	689

3. R-Markov Function

Rows of the mobility matrix must be transformed to a long column vector "v"

```
library(matlib)

markov<-function(v){

nrow<-sqrt(length(v))

B<-matrix(v,nrow=nrow,byrow=TRUE)

A.sub<-t(B)[1:nrow-1,]-diag(nrow)[1:nrow-1,]

A<-rbind(A.sub,c(rep(1,nrow)))

b<-c(rep(0,nrow-1),1)

x<-data.frame(solve(A,b))
```

x

}

4. Data Sources and Variables List

Table 10.Definition and Source of Variables

Variable	Definition
Fraction Black	Number of individuals who are black
	alone divided by total population
Racial Segregation	Multi-group Theil Index calculated
	at the census-tract level over four groups:
	White alone, Black alone, Hispanic, and Other
Household Income	Aggregate household income in the 2000 census
per Capita (in thousands)	divided by the number of people aged 16-64
Tax Progressivity	Difference between the top state income tax rate
	and the state income tax rate for individuals
	with taxable income of $20,000 in 2008
Fraction Religious	Share of religious adherents
Fraction of Adults	Fraction of people 15 or older who are divorced
Divorced	
College Graduation Rate	Share of enrollment-weighted undergraduate
(cgrate09)	Proportion of students that complete their degree in
	150% of normal time in 2009
College Tuition	Mean in-state tuition and fees for first-time, full-time undergraduates
(In thousands)	
High School Dropout Rate (ccd_drpout912)	High school dropout rates for the
	2000-2001 school year
IncomeGrowth2000_2006_10	Income growth is calculated as the annual growth rate implied by
	the change in income over the 8 year period
Col.grad.10000	Total number of 25+ years old 4 year college graduates /.....
	10,000 County Population
Student Teacher Ratio	Average student-teacher ratio in public schools

References

Aiyagari, S Rao, Jeremy Greenwood, and Ananth Seshadri. 2002. "Efficient Investment in Children." *Journal of Economic Theory* 102 (2). Elsevier: 290–321.

Becker, Gary S. 1962. "Irrational Behavior and Economic Theory." *Journal of Political Economy* 70 (1). The University of Chicago Press: 1–13.

BureauOfEconomicAnalysis, U.S. 2018a. "Share of Labour Compensation in Gdp at Current National Prices for United." University of Groningen and University of California, Davis, Share of Labour Compensation in GDP at Current National Prices for United States [LABSHPUSA156NRUG], retrieved from FRED, Federal

Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/LABSHPUSA156NRUG, May 15.

———. 2018b. "Real Disposable Personal Income: Per Capita [A229rx0a048nbea], ;" retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/A229RX0A048NBEA.

Census, U.S. Bureau of the. 2017. "Poverty Status of People by Family Relationship, Race, and Hispanic Origin: 1959 to 2016Current Population Survey, Annual Social and Economic Supplements. Table 2." https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-people.html.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2018. "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility." Opportunity Insights. https://opportunityinsights.org/paper/the-opportunity-atlas.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The Fading American Dream: Trends in Absolute Income Mobility Since 1940." *Science* 356 (6336). American Association for the Advancement of Science: 398–406.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States *." *The Quarterly Journal of Economics* 129 (4): 1553–1623. doi:10.1093/qje/qju022.

Darwin, Charles. 1871. "The Descent of Man, 2 Vols." *London* 81: 130–1.

Gandhi, Mahathma. 1925. "On Machinery." *Young India* p.337. https://www.mkgandhi.org/voiceoftruth/machinery.htm.

———. 1934. "On Machinery." *Harijan* p.301-2. https://www.mkgandhi.org/voiceoftruth/machinery.htm.

Goldin, Cludia, and Lawrence F. Katz. 2009. *The Race Between Education and Technology*. Harvard University Press.

Houthakker, Hendrik S. 1955. "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis." *The Review of Economic Studies* 23 (1). JSTOR: 27–31.

Loury, Glenn C. 1981. "Intergenerational Transfers and the Distribution of Earnings." *Econometrica* 49 (4). [Wiley, Econometric Society]: 843–67. http://www.jstor.org/stable/1912506.

OECD. 2018. *BROKEN Social Elevator?: How to Promote Social Mobility*. OECD Publishing Paris.

Okun, Arthur M. 1975. *Equality and Efficiency, the Big Tradeoff*. Brookings Institution Press.

Solow, Robert. 1951. "Some Long-Run Aspects of the Distribution of Wage Incomes." *Econometrica* 19:

333–34.

Taylor, Howard, and Samuel Karlin. 1998. *An Introduction to Stochastic Modeling.* Academic Press.

Recent IEG Working Papers:

Dayal, Vikram and Murugesan, Anand (2020). Demystifying causal inference: ingredients of a recipe. , Working Paper Sr. No.: 393

Dash, Minati (2020). 'Our place in the future': an exploration of chalaki among young men dispossessed by a mining project, Working Paper Sr. No.: 392

Mishra, Ajit and Samuel, Andrew (2020). Does it matter who extorts? Extortion by competent and incompetent enforcers, Working Paper Sr. No.: 391

Sahoo, Pravakar, Ashwani (2020). COVID-19 AND INDIAN ECONOMY: Impact on Growth, Manufacturing, Trade and MSME sector, Working Paper Sr. No.: 390

Das, Saudamini, Jha, Prabhakar and Chatterjee, Archana (2020). Assessing Marine Plastic Pollution in India, Working Paper Sr. No.: 389

Murty, M. N., Panda, Manoj and Joe, William (2020). Estimating Social Time Preference Rate for India: Lower Discount Rates for Climate Change Mitigation and other Long Run Investment Projects* , Working Paper Sr. No.: 388

Nuthalapati, Rao, Chandra Sekhara, Bhatt, Yogesh and Beero, K., Susanto (2020). Is the Electronic Market the Way Forward to Overcome Market Failures in Agriculture? , Working Paper Sr. No.: 387

IEG Working Paper No. 394

INSTITUTE OF ECONOMIC GROWTH

University Enclave, University of Delhi
(North Campus) Delhi 110007, India
Tel: 27667288/365/424
Email: system@iegindia.org