Income Mobility in America

Manu Navjeevan
Adviser: Prof. Laurence Ales

May 3, 2018

Abstract

This paper looks at both intragenerational and intergenerational income mobility in America. The paper is divided in two parts. The first part of the paper uses social security data to find that intragenerational income mobility in America has fallen by half over the past 50 years. I then consider which statistical model of income are consistent with the observed decline in income mobility. This paper finds that the AR(1) and ARIMA models do not do a good job of explaining mobility rates without using unreasonable parameters. However, the results from these models suggest that lowering the variance in shocks to income from year to year is more important in explaining falls in mobility than changes in year-to-year persistence of income. The model of income dynamics proposed by Guvenen (2016) does a better job of generating mobility rates seen in the data, but also provides somewhat limited explanation for falling mobility rates in our analysis. The results from the Guvenen model also suggest that variation in annual income shocks is significant in explaining falling mobility rates. In the second part I examine income mobility looking for significant county-level covariates to Chetty (2016) estimates on the causal effect of counties on income mobility. The analysis finds that social capital, income shares, and policy variables such as households on social security and size of public assistance in a county are significant in explaining variance in county level mobility rates. This paper also estimates the causal effects of housing rents as a percent of income using instrumental variable regression and estimates that a 1% increase in average housing rents as a percent of income in a county leads to a -0.21 change in the causal effect of spending an additional year in the county on percentile income rank at age 26. This paper also uses results from regressions to predict which counties would have the highest mobility rates in the future. Finally, the relationship between intergenerational mobility rates and inter-county migration is studied. The analysis finds that higher mobility rates are not strongly related to increased migration

to a county, suggesting that movers have limited knowledge of which counties lead to better outcomes.

Acknowledgement

This thesis is in partial fufillment of the honors requirement for Dietrich College at Carnegie Mellon University. This completed document would not have been possible without help from my adviser, Prof. Laurence Ales, nor without the support of Dietrich College and undergraduate economics department at the Tepper School of Business.

1 Introduction

Rising income inequality is an increasingly salient issue in American political and cultural discourse. By a number of indices, inequality has risen sharply in the last half century (Corak, 2013). The popular Gini index of inequality has increased from the mid-30's in the 1980's to upwards of 40 today. And, compared to other developed countries, the U.S does worse on most inequality measures. This is not necessarily a bad sign by itself. A nation may be willing to accept higher levels of inequality so long as people are still getting what they "deserve." One part of this is one's ability to move up upon entering the workforce, either by acquiring new skills or working hard in one's initial position. The rest of this theis will refer to this as intragenerational income mobility. The second part of this concept of fairness is equal opportunity given one's family background. This will be refered to as intergenerational income mobility. So long as everyone starts out on an even playing field and people can move up the income ladder, society may not have as much trouble with inequality. Unfortunately, a growing literature (Corak, 2013) on both types of income mobility shows that this is not the case. This paper seeks to study the dynamics of income mobility.

The first part of this thesis studies models of income dynamics to identify mechanisims by which intra-generational income mobility may fall over time. Differences in how income mobility responds to changes in the AR(1) and ARIMA parameters and in the income generating mechanism proposed by Guvenen (2016) are then examined. The results from simulating these models suggest that the variance of year-to-year shocks in income are important in explaining the decline in intragenerational income mobility over the years. The variance of the year-to-year shocks do more to explain the fall in income mobility than the persitence. These results also expose some of the problems in our standard models of income dynamics. Under reasonable parameter variations, simulations of the AR(1) and ARIMA models could not replicate the decline in intragenerational income mobility over the years that we see in the data. Simulating the model proposed by Guvenen (2016) generated better results but still could not explain the steep decline in mobility rates observed in the data.

The second part looks at the intergenerational county income mobility estimates from Chetty (2016) and seek to identify significant covariates of mobility rates. I find that policy variables such as the number of households on public assitance income, industry composition, and rents are important in explaining differences in income mobility rates through the years. Using these results, this paper predicts how the geography of intergenerational income mobility may have changed over the years. This paper predicts that the most mobile counties in America are in the Mid-Northern and Mid-Western regions of the country. Finally, this paper examines the relationship between intergenerational income mobility and inter-county migration and finds that counties that produce better outcomes for children are not migrated at a significantly higher rate than poorer perfoming counties. This suggests either that Americans are unaware of the differences in mobility rates between counties or that there are other barriers to moving to high mobility counties that are not observed in our analysis.

1.1 Prior Literature

Shorroks (1978) observes that people at the very top or bottom of the income distribution tend to not stay there for long. Looking at inequality indices, then, income vectors summed over multiple years should yield lower inequality than income vectors in the individual time periods. Kohen et al. (1965) looks at U.S data and finds that the Gini index falls by 4-7% compared to individual years when summed over two years, and 9-11% when summed over three years. It follows that the degree to which inequality falls over multiple year periods gives a measure of how mobile an economy is. Shorroks formalizes this metric of mobility, shown in section 3.1.

To simulate income dynamics, the AR(1) model covered in section 3.2 is a natural starting point. Heathcote (2003) notes that the ARIMA(1,1,1) and ARIMA(0,1,1) models are generally suggested as improvements to the AR(1) model for modeling income dynamics. He compares the ARIMA(0,1,1) model to the AR(1) model and finds that, for reasonable time periods, they generate roughly similar results. For longer periods of time the variance of ARIMA(0,1,1) increases without bound, making it difficult to incorporate into structural models. The ARIMA model is covered in section 3.2.

Guvenen (2016) notes that the yearly changes to income in the typical AR(1) model are distributed significantly differently from what is observed in data. Specifically, real changes to log wages are distributed with significantly higher kurtosis (sharpness of peak) than predicted by the AR(1) model. Guvenen proposes his own model for income dynamics which is covered in section 3.3. He finds that this model generates inequality metrics closer to what is seen in the data, but still somewhat unable to explain the long right tail observed. Further, this model better exlpains the higher earnings growth kurtosis observed in the data.

In studying intergenerational income mobility, this paper uses the results and county level mobility estimates from Chetty (2016). Chetty uses tax data to analyze life outcomes of people who moved between counties at various ages to identify the causal effect of spending an extra year growing in a county on lifetime earnings. Sharkey (2014) looks through the literature on the *Moving to Opportunity* program, which gave low income families in various cities cash vouchers to move better areas. Most studies Sharkey reviews find that the children of *Moving to Opportunity* participants who moved to better areas had significantly better life outcomes than their non-moving counterparts. Fauth *et al.* examine the effect of a court-mandated mobiliy program in Yonkers, New York, wherein low-income black and latino adults in poor and segregated neighborhoods were randomly chosen to relocate to townhomes in middle-income neighborhoods. They found that, over time, movers were less likely to be receiving welfare payments and more likely to be employed than non-movers. The large effect of neighborhoods on mobility is further documented by Rothwell and Massey (2015) who estimate that neighrborhoods effects are half as important as parental effects. Using PSID data, their paper estimates that children born into a low quartile neighborhood would make $500,000 more over the course of a lifetime if they were relocated at birth to a high quartile neighborhood.

On the policy side, Butcher (2017) conducts a review of the literature surrounding welfare and transfer programs to low-income families. Butcher notes that programs such as Medicare and child-care are consistenly found to be significantly effective in increasing life outcomes for children. Further, early intervention in these policies is crucial. Butcher finds that interventions before the age of 5 have significantly larger effects than interventions that take place afterward and in some cases these interventions can erase the gap between disadvantaged adn advantaged children. In a similar vein, Chetty *et al.*(2011) analyze project STAR, a Tenesee state experimental program where students and teachers were randomly assigned to classrooms within their schools from kindergarten to third grade. They find that changes in claroom enviornment at this young age have a significant impact on lifetime earnings, furthering the consensus that early enviornment has large impacts on outcomes.

2 Intragenerational Income Mobility

2.1 Trends in Intragenerational Mobility

In analyzing models of income dynamics, it is important to know if they generate metrics of inequality and mobility consistent with what we see in the data. Further, it is important to know how these metrics have changed over time so that we can analyze how the models are performing in describing income dyanmics or analyze which parameter changes generate patterns of mobility that we see in the data. To this end, graphs of inequality and income mobiltiy over time are given in Figures 1 and 2.

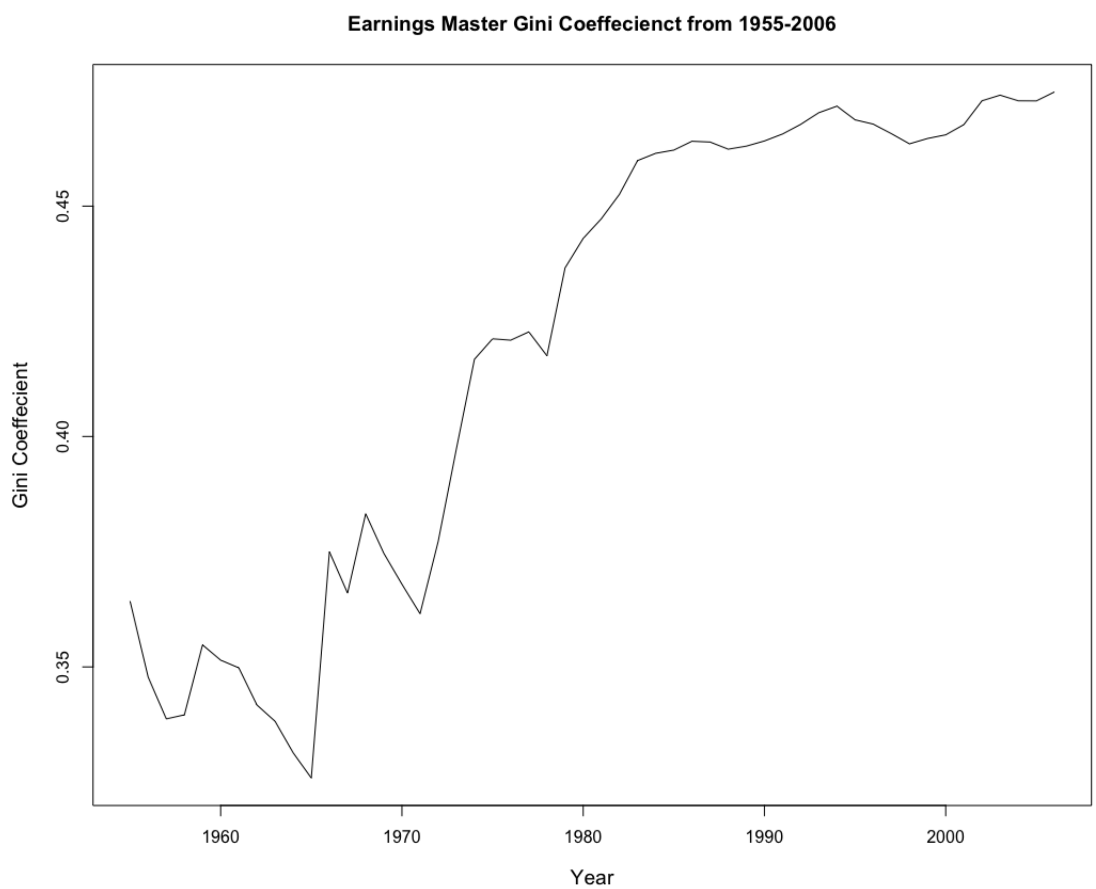

Figure 1: Earnings Master Gini index of income inequality

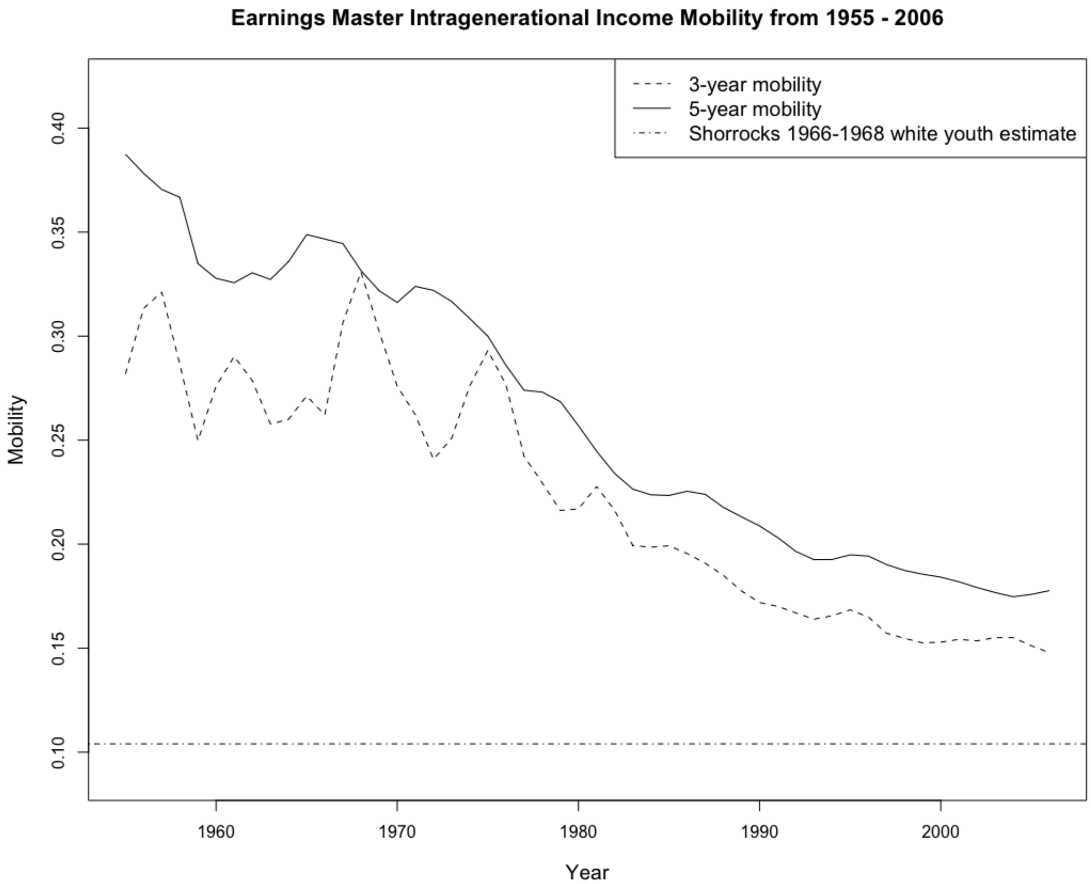

Figure 2: Intragenerational Income Mobility

The raw data for Figures 1 and 2 are collected from the 2006 Social Security Earnings public master file. The data contains social security taxable income data for a 1% sample of all Social Security numbers with positive income in any year from 1951 to 2006. For privacy, income is top coded at the taxable maximum so that incomes at the maximum are coded at the maxible taxable for social security purposes and bottom coded so that incomes between \$0 and \$100 annually are averaged and people with income in that range have their income set to that number. The Gini coeffecient [1] of the incomes in this dataset

[1]The Gini coeffecient G over a n-vector of incomes Y is given $G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i - y_j|}{2n\sum_{i=1}^{n} y_i}$. The Gini coeffecient can also be calculated as the area under the Lorenz curve, which plots on the y-axis the proportion of total

are calculated and given in Figure 1.We can see that levels of income inequality rose steeply from 1970 to 1980 and continued to grow from 1980 on, but at a significantly lower rate.

The Shorrocks (1978) mobility metric is used to calculate intragenerational mobility rates, which is given by $M = 1 - R$ where for consecutive income periods $0, \dots, m$, R is the rigidity ratio:

$$R = \frac{I[Y(t_0, t_m)]}{\sum_{k=1}^{m} \frac{1}{m} I[Y(t_{k-1}, t_k)]}, \tag{1}$$

where $Y(t_i, t_j)$ represents the vector of total income from time period i to time period j, inclusive. Generally I can be any measure of income inequality but this paper specifically uses the Gini index. Roughly, this index should measure regression to the mean from year to year. One would generally expect that people at the bottom of the income distribution will move up a bit from one year to the next whereas people at the top should move down slightly from one year to the next. Thus, multi-year income vectors should give lower inequality indices than their yearly components. The Shorrocks index compares these to index mobility. If incomes in a given year are independent from the year before and drawn from a distribution with the same variance (that is to say the economy is perfectly mobile), then inequality over the multiple year period should approach 0 and so yield a rigidity metric of 0 in expectation. Conversely, if the income vector in a given year is exactly the same as the income vectors in the years prior then the multi year inequality will be the same as the year-to-year inequality and yield a rigidity of 1. The easy interpretability is the first primary benefit of using the index. The second is that the index can detect income swaps because inequality is calculated over an income vector that consists of the same people from year to year.

Figure 2, on page 5, shows how income mobility has been falling in the U.S over the past 40 years. The solid line at the top of the chart represents the mobility metric calculated in moving 5-year periods. For example the mobility rate in 1960 is calculated using the income vectors from years 1956 - 1960 inclusive. The dashed line represents the same mobility metric but calculated in moving 3-year periods. The line at the bottom (at Mobility = 0.104) is the mobility metric for whites calculated by Shorrocks (1968). It is notable that this line is significantly lower than the solid or dashed lines. This is probably due to the truncation caused by the top coding of income in the data. This significantly

income earned by the bottom x% of the population.

lowers the measured inequality and raises the mobility metric. An attempt was made to roughly quantify how this truncating effects the mobility metric to get a better sense of how real mobility rates have changed over the last half century, however, no model of income dynamics generated the large gap in income mobility estimates that we see between the Earnings Master data and Shorrocks (1968).

As with inequality we see a large change (albeit in the opposite direction) in mobility rates from 1970 to 1980. However, unlike in the graph of income inequality, this decline generally continues from 1980 onwards. There are many possible reasons for this drop: a greater focus on skilled labor could make it harder for the labor class to get high paying jobs, capital buildup after World War Two is putting increasingly large shares of wealth in the hands of a few, and so on. Investigating the particulars of what exactly has caused falling mobility rates and to what extent, however, would be beyond the scope of this thesis and is left to future research to identify. Instead models of income dynamics, specifically the AR(1), ARIMA(1,0,1), and the model proposed in Guvenen (2016) are broadly examined to see how mobility rates generated by these models respond to changes in parameters. By analyzing how the mobility metric responds to changes in parameters, we hope to infer what features of the economy may have caused the decline in mobility seen in Figure Figure 2.

2.2 AR(1) Parameter Changes

The first model considered is the 1st degree autoregressive model of income dynamics. That is the model specified:

$$Y_{t+1} = c + \varphi Y_t + \epsilon_t, \forall t$$

where $c, \varphi \in \mathbb{R}$ and ϵ is a normally distributed error term. With $|\varphi| < 1$ this model is wide sense stationary with expectation $E(Y_t) = \frac{c}{1-\phi}, \forall t$. When φ is varied, the value of c is adjusted so that the expected income in any given year is $\sim \$60,000$, roughly the average income in the U.S. The model is then specidied with values of φ in the range of 0.85 to 0.95 as suggested by Heathcote (2005).

The graph of how 3-year mobility rates compare to values of φ can be found below in Figures 3 and 4. Figure 3 gives mobility rates without truncating incomes at $\$94,500$. To generate the figures an AR(1) process is simulated1000 times for each φ and AR(1) shock standard deviation pair. Each time the simulation is run on a random sample of 10,000 incomes whose initial values were drawn from a log-normal distribution with log mean 11 and log standard deviation 0.25. The AR(1) shocks were then initially drawn from a normal distribution with mean 0 and a standard deviation that was varied. Before calculating inequality indexes log values were exponentiated to obtain the real simulated income values.

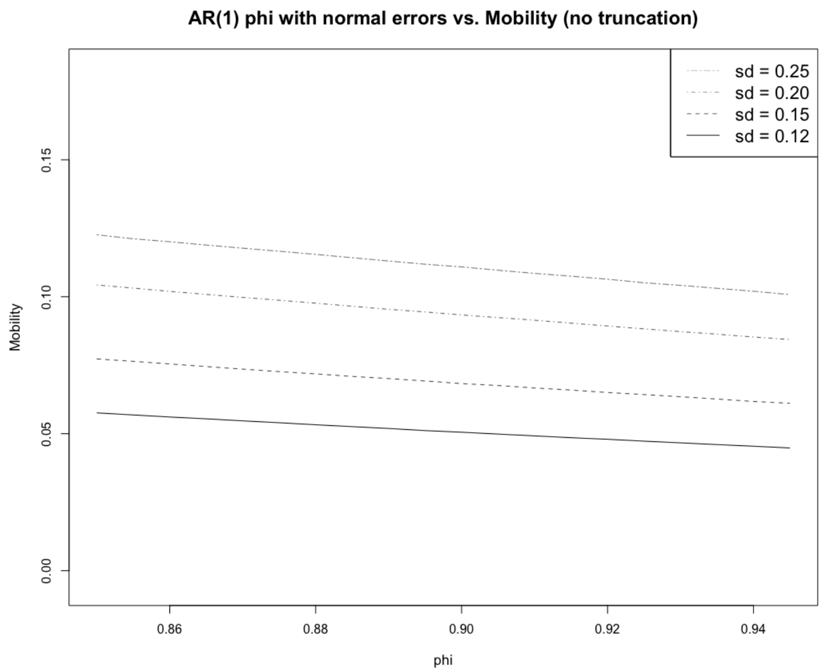

Figure 3: AR(1) Mobility no truncation

From $\varphi = 0.84$ with shock standard deviation 0.25 to $\varphi = 0.95$ with shock standard deviation 0.12, our mobility rates all lay within a range of 0.07. This means that paramater changes to the AR(1) model could only explain about 30% of the large decline in mobility we see in Figure 2. In order to recover the mobility rates calculated by Shorroks in 1998 one would need to use a φ value of close to 0.85 and a standard deviation on the shocks of close to 0.25, both of which are on the extreme end of the spectrum. Much more likely would be an observed φ value of 0.90 and standard deviation of 0.15 which gives us a mobility rate of 0.07 in both the non-truncated model and truncated model. This

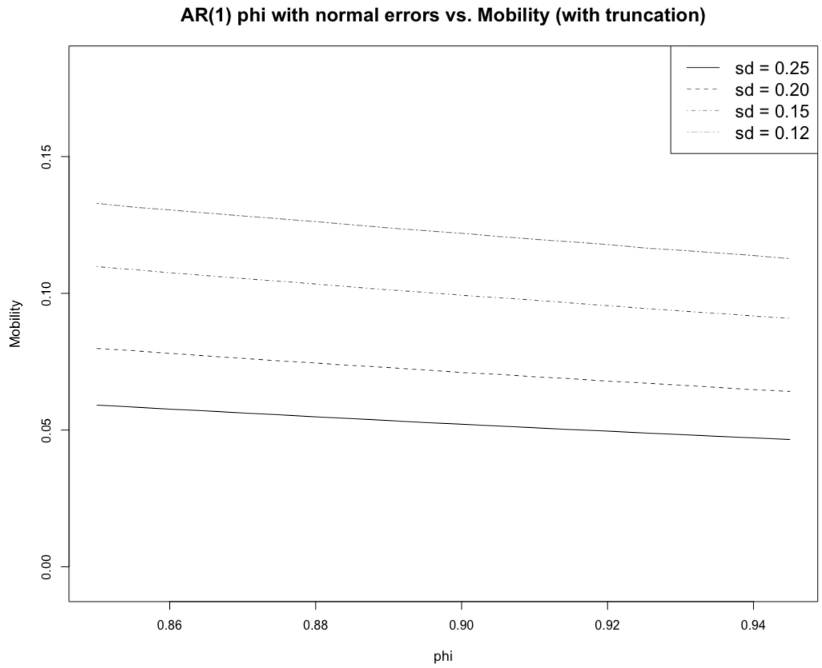

Figure 4: AR(1) Mobility truncated

is significantly lower than the mobility rate observed by Shorroks in 1968 and gives little room for falling mobility in the years following. More concerningly, the AR(1) model seems to give similar mobility rates regardless of truncation. This all suggests that people aren't swapping income ranks in the AR(1) model as often as we'd expect in real life. It is possible that with some changes to the initial income vector specification mobility rates could be obtained somewhat closer to our observed rates in Figure 2, but given the wide variation already used with parameters, it's unlikely we'd get terribly closer.

These results don't change significantly when lognormal instead of normally distributed shocks are used. In this case our model is given

$$Y_{t+1} = c + \varphi Y_t + (\epsilon_t - 1), \forall t$$

where ϵ_t is distributed lognormally with varying standard deviation. Our full shock is given $(\epsilon_t - 1)$ to compensate for the fact that our lognormally distributed ϵ_t must have positive mean, in this case one. Given that in the lognormal distribution smaller magnitutude and negative shocks ($\epsilon_t < 1$) are significantly more likely to occur than large magnitutude and positive ones, income shocks in real life may closer match the lognormal distribution.

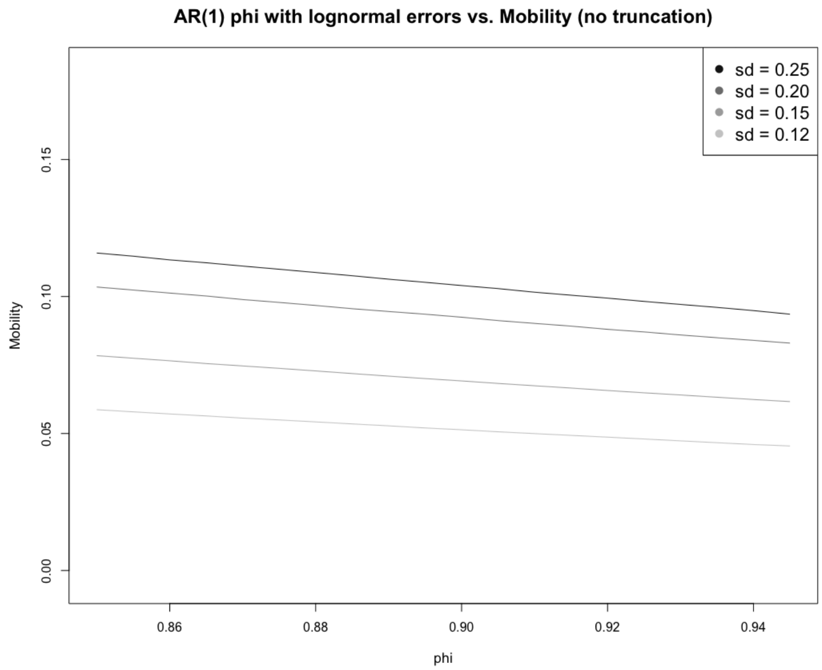

Figure 5: AR(1) Mobility with lognormal errors and no truncation

Unfortunately, as seen in Figures 5 and 6, lognormal errors provide no improvement on the estimation. The mobility rates here are again nowhere near those observed in the data and truncation again seems to provide no different mobility rates than the the non-truncated model (the average difference between the truncated and nontruncated mobility rates is 0.007). This suggests that the problem is with the mechanics of the AR(1) model itself rather than the distribution of the shocks.

All together this is not completely unexpected. That the AR(1) model does a poor job of explaining changing income distributions is well documented in prior literature (Heathcote, 2005). Our results further this conclusion and show the degree to which the AR(1) model fails. The results also give a mechanism by which the AR(1) fails by suggesting that AR(1) simulations do not demonstrate income rank swaps at nearly the rate at which they happen in real life. One reason for this could be that the AR(1) shocks are uncorrelated from year to year. It is more plausible that there is an inherent (non-random) reason that people's

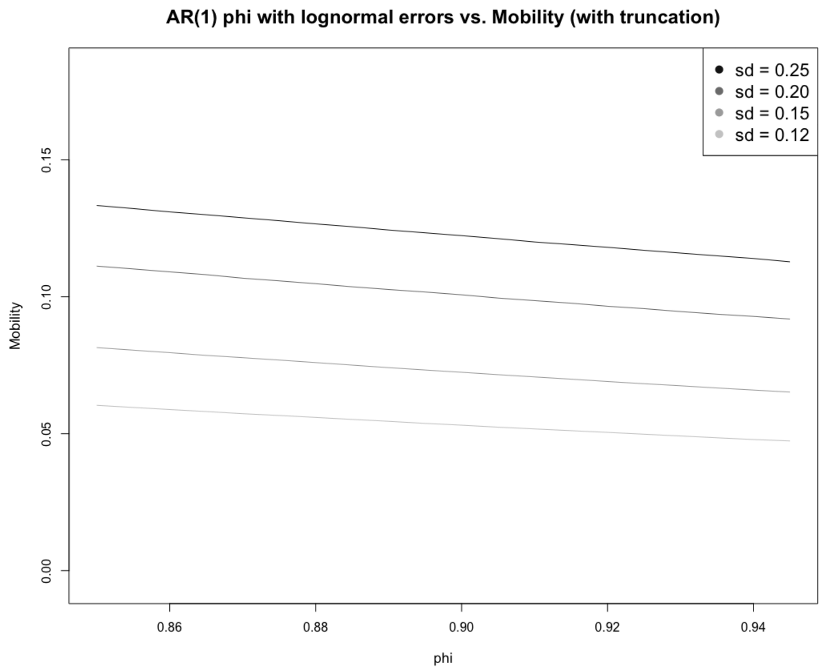

Figure 6: AR(1) Mobility truncated with lognormal errors

incomes change from year to year. This means that if someone's income rises in one time period, it is more likely to rise in subsequent time periods and vice versa. Allowing for correlated shocks would increase movement in the rank distribution and thus could allow the AR(1) model to better replicate observed mobility rates. Another problem is that the shocks in the model are independently and identically distributed across the income scale (homoskedastic and with mean zero). This means that someone at the top of the income distribution faces the same probability of earning $1000 more in the next time period as someone at the bottom of the income distribution. There are a couple reasons to believe this is not the case in actuality. First, someone earning more probably sees greater variance in their incomes from year to year than someone earning less. The average worker is extremely unlikely to double his income from one year to the next, but it is significantly more plausible that someone at the very top of the income distribution increases his income by $60,000 in a year. Secondly, the mobility metric itself is based on the observation that people at the bottom of the income distribution tend to move up and people at the top of the income distribution tend to move down. This means that one should expect shocks to be positive at the bottom of the distribution and negative at the top. Finally, that there is no difference between mobility rates before and after truncation suggests that there are not enough people making above $94,500 to make a large difference in the inequality metric despite the relatively high $60,000 mean income. This suggests that the AR(1) model does not introduce enough variability into our model.

However, the AR(1) model results may still be able to provide some insight into falling mobility. In increasing the value of φ from 0.885 to 0.95 the mobility rate falls by around 25% holding constant the stardard deviation of the shocks. By contrast, if φ is held constant, the mobility rate can be reduced by over 50% by reducing the standard deviations of the shocks in both the truncated and non-truncated models. This suggests that the persistance and size of these shocks is more important in explaining decreases in mobility rates than the the the AR(1) parameter φ. That is, the model suggests that the variance of the year-to-year income shocks matters more than the expected persistence of income from year to year with respect to income mobility. When lawmakers look for policy initiatives to raise mobility rates, it may be more helpful to focus on policies that help people gain more income than policies that redistribute income.

2.3 ARIMA(1,0,1) Parameter Changes

As noted in the previous section, one of the problems with using the AR(1) process to model income dynamics is that errors are uncorrelated over time periods. That is someone who experienced as growth in income in time period one is no more likely to continue growing income in time period two. One way around this is using an ARIMA (Autoregressive Integreated Moving Average) model. The general ARIMA(i,0,j) model is given:

$$Y_t = \beta^T \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-i} \end{bmatrix} + \alpha^T \begin{bmatrix} \epsilon_{t-1} \\ \epsilon_{t-2} \\ \vdots \\ \epsilon_{t-j} \end{bmatrix} + \epsilon_t$$

where β and α are 1 x i and 1 x j vectors of coeffecients, respectively. The ARIMA(1,0,0) model is equivalent to the AR(1) model studied above.

Because the ARIMA model integreates errors from past terms, it should do a better job of replicating real mobility rates if this was significant issue with just AR(1) modeling.

The results for simulating the model with varying alpha and beta are given below. We do not change the standard deviation of the shocks in this simulation, but given the ARIMA model's structural similarity to the AR(1) model, we would presume that this would yield similar results. The results from varying paramaters are given in Figure 7.

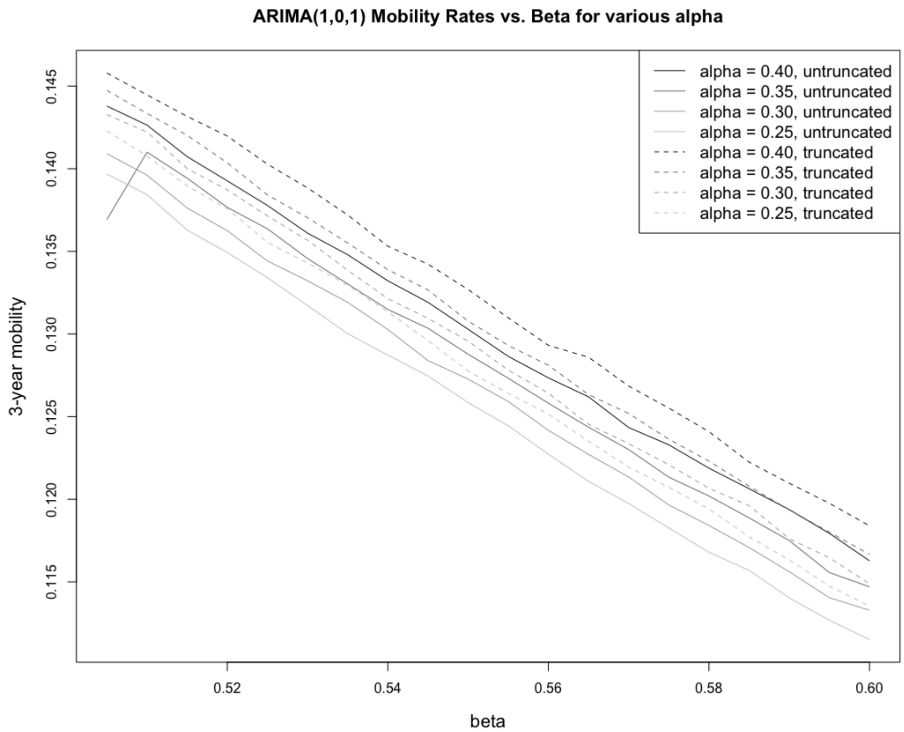

Figure 7: Mobility rates generated by ARIMA model

The ARIMA model generates mobility rates closer to those that we see in the data than the AR(1) model. The maximum mobility rate generated by our untruncated ARIMA model under normal parameter variation was around 0.145 compared to 0.12 for the AR(1) model. nder a shock standard deviation of 0.12, the model generates mobility rates close to what we see in the data for today. Unfortunately, the ARIMA model also does not generate the decline in mobility seen in the data under parameter variation. The range of mobility values in our ARIMA simulation is 0.032 compared to 0.07 in the AR(1) model. This means that the parameter variation in the ARIMA simulation could only explain 14% of the decline in mobility rates seen in the data. Further, the ARIMA model also does not generate enough of a right tail, as evidenced by the small differences between the truncated and non-truncated mobility rates. This means the model is not particularly helpful in explaining the difference between the estimated mobility in this paper and the estimated mobility found by Shorrocks. This is not to say the ARIMA model is useless here, however. The fact that mobility rates were more somewhat more accurate under this model suggests that incoporating past shocks is an important part of modeling income dynamics. The model we study in the next section does this and we examine the results.

2.4 Guvenen (2016) Model

Since neither the AR(1) or ARIMA(1,0,1) models do a particularly good job of explaining changes in mobility patterns over time, this paper turns to the model of income dynamics proposed by Guvenen in *What Do Data on Millions of U.S Workers Reveal about Life-Cycle Earnings Dynamics*(2016). The model is based of the observation that earnings growth is negatively skewed and that that the distribution earnings growth has very high kurtosis (the fourth moment of the distribution, this is to say the peak is very sharp). Guvenen notes that fitting values to his model gives an objective value more than three times lower than that of the AR(1) process, indicating significantly better fit of the model. This section examines this model to see if it generates more realistic mobility metrics and whether parameter changes in the model can do a better job of explaining the long term decline in mobility seen in Figure 2.

It is prudent to spend some time here going over the model and interpreting the various parameters before adjusting them. The model is presented:

$$Y_t^i = (1 - v_t^i)\exp(g(t) + \alpha^i + \beta^i t + z_{1,t}^i + z_{2,t}^i + \epsilon_t^i) \tag{2}$$

$$z_{1,t}^i = \rho_1 z_{1,t-1}^i + \eta_{1,t}^i \tag{3}$$

$$z_{2,t}^i = \rho_2 z_{2,t-1}^i + \eta_{2,t}^i \tag{4}$$

$$\tag{5}$$

The shocks to the model are distributed for $j = 1, 2$ as follows:

$$n_{jt} \sim \begin{cases} -p_{jt}^i \mu_j & \text{with pr. } 1 - p_{jt}^i \\ N((1 - p_{jt}^i \mu_j), \sigma_j^i) & \text{with pr. } p_{jt}^i \end{cases}$$

$$\log(\sigma_j^i) \sim N(\log(\bar{\sigma}_j) - \tilde{\sigma}_j^2/2, \tilde{\sigma_j})$$
$$z_{j0}^i \sim N(0, \sigma_j^i \sigma j 0)$$
$$v_t^i \sim \begin{cases} 0 & \text{with pr. } 1 - p_{vt}^i \\ \min\{1, Expon(\lambda)\} & \text{with pr. } p_{vt}^i \end{cases}$$

In the model, t is normalized age $t = (age - 24)/10$, g(t) is a quadratic polynomial in age that represents the life-cycle of earnings. The vector (α^i, β^i) is a randomly drawn from a multivariate normal distribution with zero mean and estimated covariance matrix that allows for variability in the life-cycle component of earnings for different people in the population. This model is simulated in R and present the mobility metrics obtained. Baseline parameters are obtained from the guvenen paper. The quadratic lifetime component of earnings is estimated using the 2015 ACS estimate and is given:

$$g(t) = 11.76947 - 5.5706t + 3.469t^2$$

Using the parameters reported in the paper, the 3-year truncated mobility rate in the model is estimated as 0.1366 wheras the 3-year non-truncated mobiliy rate in the model is estimated as 0.0616. This larger gap of 0.075 in mobility rates before and after is more consistent with what we'd expect given the difference between the Shorrocks paper and my estimates from the Social Security dataset, but still fails to explain the large difference between the two. None of the models of income dynamics this paper looks at are able to explain the large difference between the two, which we conclude is probably due to the long right tail observed in the income distribution that is not replicated by our models. It's relevant, however, to note that the Guvenen model already is giving mobility rates much closer to what is observed in the data, which is a good sign for it's ability to explain changes in mobility rates over time.

The model is examined by changing the standard deviations of the AR(1) shocks (z_1 and z_2). Changes to the life-cycle component to earnings were also examined but found to have little impact on mobility rates. The simulation allows each of the shocks' standard deviations to deviate by 0.1 in either direction and the results are reported in Figure 8.

In the Guvenen model, increases in AR(1) shock standard deviations lead to lower mobility rates. This could be because in the Guvenen model shocks accumulate over time,

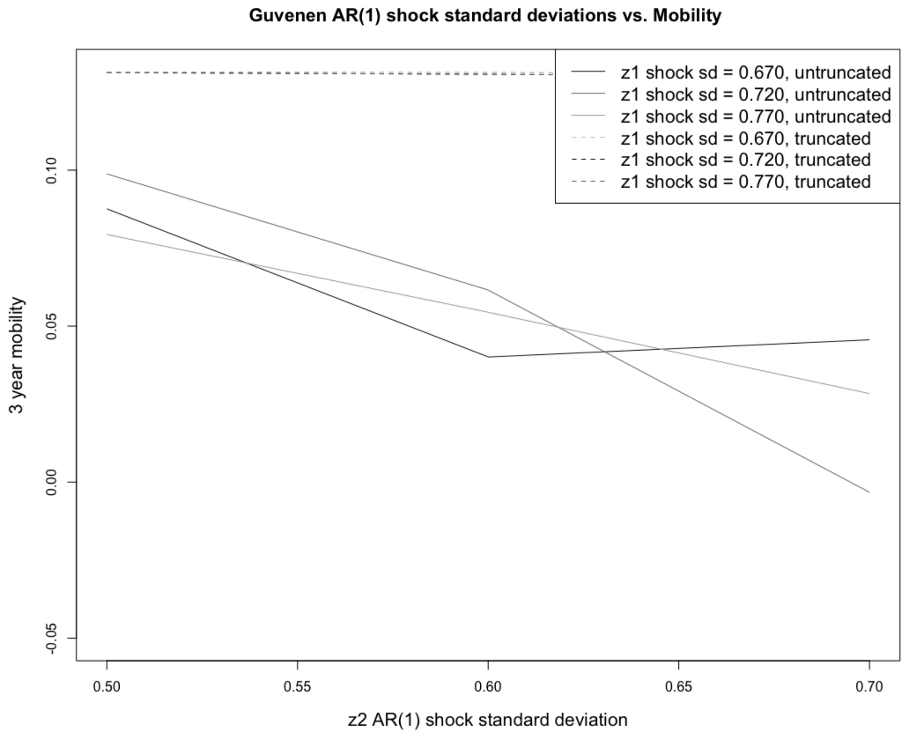

Figure 8: Changes in standard deviations to AR(1) shocks in the Guvenen model and their effects on mobility rates

meaning that a strong initial shock could lead to sustained higher growth. Further, the standard deviation of the shock to z_2 seems to have a higher impact on mobility rates than the standard deviation of the shcok to z_1. It is also notable that the effects of truncation are clearly more severe here. Whereas changes to the standard deviation of the AR(1) shocks cause significant variance in the non-truncated mobility metric, the truncated mobility metrics for each standard deviation level are virtually indistinguishable. This suggests that much of the inequality generated by the Guvenen model is in the top end of the distribution. Though it may be representative of the actual income distribution, it fails to explain the difference in mobilty rates between our calculations and the Shorrocks paper.

2.5 Concluding Remarks

Between the models studied,the Guvenen (2016) model does the best job of generating mobility rates similar to what is observed in the data, probably due to it's incorporation of the lifetime earnings component of earnings and the the accumulation of shocks over time. This suggests that changes in the importance of initial shocks (education, etc.) in an individuals lifetime earnings are important in explaining the long-term decline in income mobility. Income mobility estimates generated using the AR(1) and ARIMA(1,0,1) processes do not change much when we vary the income persistence parameters (φ, β, α) but are more responsive in changes to the standard deviations of the shocks. This suggests that falling income mobility rates can also be attributed to how much people's income vary from year to year in absolute terms. These results provide an explanation for falling mobility which is somewhat consistent with the Guvenen model, which is that each person's starting position (starting shock) has become more important over the last half-century. A large positive shock is likely to accumulate over time, leading to consistenly higher incomes, and is unlikely to be reversed by large negative future shocks. This suggests that good initial conditions such as family background or education have become more important over the past half-century in securing high incomes.

However, our findings show that none of the models can fully explain the decline in intragenerational income mobility over the past 50 years. This is probably due to each model's documented difficult in generating the long tails we see at the top of the income distribution. Regardless, this is an area of future study and studying models that do better jobs of generating these long tails could lead to better insights as to why mobility has been falling.

3 Intergenerational Mobility

This next section considers intergenerational mobility, how income levels transfer from parents to children. Prior research (Rothwell and Massey, 2015; Chetty, 2016) shows that the area a child grows up in has a significant effect on their future earning. Chetty (2016) provides estimates of the causal effect of growing up in a county on percentile earning rank at age 26 for children born between 1980 and 1986 in the bottom and top quartiles. These estimates are compiled for most of the over 3,000 counties in the U.S and measure the effect of spending a year extra growing up in a specifc country compared to the mean. We combine the county-level causal estimates with county level data from the census, USDA, and the Equality of Opportunity Project to look for significant covariates for county level mobility. By comparing Chetty estimates to county level variables, this theis identifies some potential features of counties with higher mobility rates that make them produce better outcomes than counties with low mobility rates. Chetty estimates are also compared to migration rates to investigate whether families move to places with higher mobility rates.

The results from these comparasions, which are done via linear regression, show that policy variables such as percentage of county residents on public assistance income, percentage of people working in manufacturing and retail, and the average percentage that rents take out of income are important in explaining differences in mobiltiy rates between counties. Further, levels of these variables in earler years, when children are young, explain more of variance in county variables than the levels of the variables in later years. These results are then used to make predictions on which counties would show the highest mobility rates for children born in 2000 and children born in 2015. These predictions can be compared to real results in the future to verify the predictive power of our variables. Finally, the relationship between mobility rates and inter-county migration is examined. Regression results do not show a significant relationship between the two, which may motivate future research on why families move and on whether families are aware of differences in mobiltiy rates between counties.

3.1 Data

Demographic data (racial shares, single parent household, income, etc.) for each county which are obtained from the 1980, 1990, and 2000 censuses and the 2010 and 2015 ACS 5-year estimates via http://www.socialexplorer.com. Notably in addition to this county level data on natural amenities is obtained from the USDA. The main catergories of variables and their definitions are presented below to provide clarity in interpreting regression results later.

Data from county level mobility is obtained from the online appendix to Chetty's *Where is the Land of Opportunity?* (2016) (http://www.equality-of-opportunity.org/data/). Online data table 3 is used, specifically looking at the variable 'pct_causal_p25_kr26' which is refered to generally as "mobility" and measures the causal effect of spending an extra year growing up in a county on earnings at age 26 for children born between 1980 and 1986 into the 25th percentile of income.

Descriptive statistics for these variables are given in the appendix in Tables 2, 3, and 4.

3.1.1 Variable Definitions

1. Demographic Variables: Racial Shares (broken down into white, black, asian, american indian, and other by percentage of total population), population density (people per square mile), percentage of students dropping out of school, percentage of households with one parent, percentage of population that is male, percentage of population with complete college and high school education, social capital index (only available some years), racial segregation (measured by entropy indices and isolation indices)

2. Economic Variables: Median cash rent as a percent of income, percentage of population below the poverty line, Gini income index for bottom 99% (only available some years), percentage of workers employed in manufacturing, percentage of workers employed in retail

3. Policy Variables: Percentage of workers employed in public administration (proxy for size of government), percentage of workers employed in social assistance roles (proxy for size of welfare)

Demographic and policy variables were transformed from the raw census data to normalize them across counties as a percetage of total population.Racial segregation indices are not directly obtained from the census data. Instead, racial shares by census tracts and the isolation index

$$\sum_{i=1}^{n}(\frac{x_i}{X_c}\frac{y_i}{t_i}),$$

are used to calculate racial segregation. Here X_c is the minority population (black) of the county, y_i is the majority population (white) of the census tract i, x_i is the minority population of census tract i, and t_i is the total population of census tract i. A variety of other indices, including the entropy index used by Chetty,are calculated the isolation index was found to be the most effective in predicting mobility rates.

The next subsection looks at certain variables that Chetty proposed as significant in explaining variance in county level mobility rates and confirms that these are significant covariates. The next subsections consider, in order, polivy variables, industry variables, and rents to show that they are significant in explaining variances in mobility rates. The subsection examining rents also instruments on natural amenties to isolate the causal effect of higher rents on moblity rates. This thesis then uses these results to make predictions on how counties may effect income levels at age 26 for children born in 2000 and 2015 and considers the relationship between mobility rates and inter-county migration.

3.2 Replicating Chetty's Results

At the end of his paper, Chetty proposes a group of variables that he claims are significant in explaining income mobility, namely: racial shares, single parent households, the Gini measure of inequality, social capital. Here mobility is regressed against the Chetty variables to confirm that they are significant in explaining variances in mobility rates and examine the results. Chetty's results are extended by considering variables from 1980 and 2000 (Chetty only considered variables around 2000). This is not possible for all variables (social capital and Gini by county are only available),but hopefully the abridged results may still provide some insight.

The individual regressions can be found in the appendix, but most variables were confirmed to be significant covariates with the notable exception of the entropy segregation index. Instead, the entropy index listed above is found to be the most significant in explaining variance in mobility rates. The lack of significance in our results could well be due to the coarseness of our metric. Large metropolitan areas may have many census tracts per county, under which our isolation metric may provide some sense of the segregation in each county. However in smaller counties with only a few or sometimes only one census tract, our index may not be fine enough geographically to give a good sense of segregation.

Table 1: Estimated Chetty Regression 2000

	(1)	
	pct_causal_p25_kr26	
entropy_2000	-0.108	(-1.71)
gini_chetty	-1.708***	(-12.24)

topinc_chetty	0.865***	(3.78)
oneparent_2000	-3.674***	(-22.12)
ski_1997	0.113***	(18.64)
white_2000	-1.358***	(-10.88)
black_2000	-1.358***	(-10.81)
amind_2000	-1.566***	(-7.97)
asian_2000	-4.250***	(-6.78)
constant	3.218***	(22.74)
N	2740	
R^2	0.681	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Approximated Chetty Regression 1990

	(1)	
	pct_causal_p25_kr26	
entropy_1990	-0.104	(-1.73)
oneparent_1990	-5.390***	(-32.63)
ski_1990	0.127***	(24.35)
white_1990	-1.348***	(-9.08)
black_1990	-1.129***	(-7.62)
amind_1990	-1.069***	(-5.35)
asian_1990	-1.798**	(-2.98)
constant	2.800***	(18.35)
N	2844	
R^2	0.689	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Approximated Chetty Regression 1980

	(1)	
	pct_causal_p25_kr26	
oneparent_1980	-6.731***	(-24.20)
white_1980	-0.315	(-1.95)
black_1980	-0.650***	(-3.83)
amind_1980	-0.742**	(-3.28)
asian_1980	4.904***	(6.36)
constant	1.649***	(9.95)
N	2844	
R^2	0.619	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We start by regressing the variables proposed be Chetty (2016) against the county mobility metric and assesing the results. The regression tables in Tables 1,2, and 3 show that the Chetty proposed variables are important in explaining income mobility. Looking at the R^2 value, we find these variables explain $> 60\%$ of the variance in income mobility rates across county, which is more than expected for a random group of demographic characteristics. For the most part, the signs on the coeffecients are interpretable and make sense. Higher rates of segregation, inequality, and single parent households are associated with lower mobility, which is expected given results mentioned in the prior literature section. Conversely, social capital (an index measuring civic participation, religious service attendance, etc.) is positively correlated with income mobility. Most of the variables are significant at the $\alpha = 0.001$ level, but since the data most large counties in the U.S rather than a small random sample, we caution against reading too much into this.

Oddly, all races are negatively correlated with mobility rates with the exception of asian americans in 1980. This paper cannot propose any concrete way to interpret this, especially given that all racial shares are expressed as a percentage of the total population. With some probability this is just an adjustment making up for a high constant term, but this is could be good question for future investigation.

It is also important to note here that the 1990 regression explains more variance than the

2000 regression and does so without the inequality or social capital variables, both of which are important standalone covariates with mobility. The 1980 regression explains less of the variance than the 1990 regression but is additionally missing the segregation variable. The higher explanatory power of variables in older years is consistent with results from Butcher (2017) that broadly show that a child's enviornment in early years is very important in explaining future earnings. However, this result is also confirmed at a macro level and shows that demographic characteristics of a county effect have effects on it's mobility and earnings rates that are lagged by 20 years or so. In terms of establishing a causal relationship, I would guess that variables in 1980 and 1990 would have the highest causal effect on mobility rates, and that the high explanatory seen by demographic variables in 2000 would be due to demographic characteristics of counties being auto-correlated through the decades. Recognizing that changes to county demographics or programs may only have large effects on earnings or mobiltiy rates is important when developing policy and in not evaluating the sucess or failure of a policy too early.

3.3 Policy Variables

The effect of the policy variables mentioned abov is now considered. The results of regressing the size of social assistance, the size of public administration, the percentage of people on social security income, and the percentage of people on public assistance income against mobility are presented here.

Table 4: Policy Regression 2000

	(1)	
	pct_causal_p25_kr26	
socialassist_2000	1.093***	(4.86)
pubadmin_2000	-1.885***	(-5.34)
hhssinc_2000	1.864***	(11.12)
hhpainc_2000	-9.382***	(-15.15)
constant	-0.137*	(-2.11)
N	2844	
R^2	0.147	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Policy Regression 1980

	(1)	
	pct_causal_p25_kr26	
socialassist_1980	-0.387*	(-2.28)
pubadmin_1980	-1.165***	(-4.36)
hhssinc_1980	2.081***	(13.43)
hhpainc_1980	-7.065***	(-33.42)
constant	0.326***	(5.95)
N	2844	
R^2	0.372	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Policy Regression 2000

	(1)	
	pct_causal_p25_kr26	
socialassist_2000	1.093***	(4.86)
pubadmin_2000	-1.885***	(-5.34)
hhssinc_2000	1.864***	(11.12)
hhpainc_2000	-9.382***	(-15.15)
constant	-0.137*	(-2.11)
N	2844	
R^2	0.147	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The policy variables: percentage of people working in social assitance roles, percentage of people working in public administration, and percentage of people on public assistance or social securit income on are regressed against our mobility metric. The goal is to asses whether these variables are significant in explaining variance in income mobility and if the explanatory effect varies based on the year. Looking at the R^2 values in Tables 4,5, and 6 we see that policy variables in 1980 and 1990 can explain 31-37% of the variance in mobility rates, wheras variables in 2000 can only explain 14% of this variance. The percentage of people in social assistance and the percentage of households on social security income were positively associated in 2000 and 1990 with higher mobility rates while the percentage of people in public administration roles and percentage of of households on public assistance income were negatively associated with mobility. The percentage of people in social assistance roles was slightly negatively associated with 1980. Age is a lurking variable that could potentially explain the positive association with social security recipients. Older communities are probably more affluent with lower crime rates, both positive factors for children's outcomes. Further, these communities may be less likely to be urban, which could also lead to better outcomes for children. Poor enviornment also may explain the negative coeffecients on pubadmin and hhpainc, poorer and more urban communities both probably have larger governments and more poeple on public assitance incomes. These enviornmental lurking variables probably explain why the socialassist coeffecient is negative in 1980 as well. This being said, the positive coeffecient on socialassist in 2000 and 1990 is worth looking into more in future research. Prior literature (Butcher, 2017) already suggests that social assistance programs can be effective in raising children's incomes so it is plausible that there is a causal association between social assistance and mobility.

More importantly, the difference in explanatory power between the various year is much more apparent here. The policy variables in 2000 explain little of the variance in mobility rates, whereas the same variables measured in 1990 and 1980 have much higher explanatory power. It is somewhat possible that because of changes in how welfare is admistered policy variables just weren't as influential on mobility rates in 2000 as they were in 1980 and 1990, but it is more plausible that this is due to the importance of early enviornment on life outcomes. If Chetty was looking for significant covariates for mobility and only looking at variables in 2000 he would not have been able to identify the variables mentioned above.

3.4 Industry Regressions

Given that much of the concern over falling mobility rates in America has to do with the loss of stable middle class jobs in fields like manufacturing, this paper looks at how industry shares of manufacturing and retail are associated with mobility rates. The hope is to get a sense of how children who may have grown up in traditionally working class areas like Pittsburgh may have been affected by the decline in manufacturing. The results from the regression are displayed below in Tables 7,8 and 9.

Table 7: Industry Regressions 2000

	(1)	
	pct_causal_p25_kr26	
manufacturing_2000	-1.077***	(-9.22)
retail_2000	-3.183***	(-4.81)
constant	0.783***	(8.79)
N	2844	
R^2	0.038	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Industry Regressions 1990

	(1)	
	pct_causal_p25_kr26	
manufacturing_1990	-1.954***	(-18.06)
retail_1990	-3.383***	(-8.59)
constant	1.167***	(14.23)
N	2844	
R^2	0.123	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Industry Regressions 1980

	(1)	
	pct_causal_p25_kr26	
manufacturing_1980	-1.887***	(-18.84)
retail_1980	-2.052***	(-5.08)
constant	0.960***	(11.84)
N	2844	
R^2	0.144	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The differences in explanatory power for industry shares in 2000 compared to earlier years are immediately notable here. The industry variables in 2000, with an R^2 of 0.038, hardly posess any explanatory power above random association whereas the variables in 1990 and 1980, which explain between 12 and 14% of the variance in county level mobility rates are moderately sucessful in explaining mobility rates. Again this is probably due to the importance of early enviornment on life outcomes.

The negative coeffecients on manufacturing and retail are as expected and git with the popular narrative on the decline of the middle class. Holding all else constant, a 1% in the fraction of people employed in manufacturing in a county in 1980 would be associated with a 0.019 unit decrease in the mobility metric. Over 18 years of growing up in such a county, this would mean at 0.34% decrease in income, which is not insignificant. The national average for percentage of people employed in manufacturing was $\sim$ 21% in 1980 whereas the percentage of people employed in manufacturing in the Detroit area in 1980 was $\sim$ 31% (McDonald, Bernstein 2013), which means that children in Detroit and counties similar to in in manufacturing output could expect to have income rank 3% lower at age of 26 compared to their similar peers in other counties. This gap between expected mobility rates is even more striking when looking at retail shares of employment. A 1% increase in the fraction of people employed in retail in 1990 is associated with a 0.03 decrease in our mobility metric, which would amount to an approximately 0.6% decrease in percentile income rank at age 26. Children growing up in counties with a 10% higher share of people employed in retial compared to the national average could then expect to rank 6% lower

Table 10: Regression of average rents a percentage of income against county mobility

	(1)	
	pct_causal_p25_kr26	
rentav	-0.0731***	(-11.75)
hhinc_1980	0.0000337***	(11.93)
hhinc_1990	-0.0000411***	(-9.36)
hhinc_2000	0.0000248***	(6.68)
_cons	0.629***	(5.12)
N	1752	
R^2	0.222	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

in income at age 26 compared to their comparable peers in other counties.

Obviously this isn't a direct causal effect. It's unlikely that manufacturing itself depreciates the earnings outlooks for children growing up around it. Rather, it is more likely that the the unemployment caused by declining manufacturing jobs nationwide leads to unfavorable effects such as lower graduation rates and higher crime. Regardless, the results above show how large these effects can be. The causal effect of these changes should be the subject of future research and policy interventions.

3.5 Rents & Disposable Income

Parents with high disposable income can afford better education resources for their children which lead to better life outcomes. Given this, one may expect places with higher rents to produce worse outcomes for children by eating into the disposable income of parents. To test this, the average rent as percentage of income in 1980, 1990, and 2000 and is initially regressed against the mobility metric adding household income in each year as a control. The simple regression results are displayed in Table 10.

Prior literature has docmunted a relationship between higher disposable parental income and better life outcomes for children. Parents with more disposable income can buy educational materials such as books and afford to spend more time nurturing children

Table 11: IV Regression of rents against mobility, instrumenting on Natural Amenities Index

	(1)	
	pct_causal_p25_kr26	
rentav	-0.0640*	(-2.38)
hhinc_1980	0.0000345***	(9.67)
hhinc_1990	-0.0000421***	(-8.00)
hhinc_2000	0.0000254***	(6.14)
_cons	0.465	(0.94)
N	1752	
R^2	0.221	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(Butcher, 2017), so we expect a positive causal relationship between lower household rents as a percentage of income and county mobility rates. With an R^2 value of 0.222 and a negative coeffecient of -0.0731 on the rentav variable, the regression in Table 10 confirms that household rents as a percentage of income are important in explaining variance in county mobility rates. Moving forward, we seek to identify the causal, rather than just the correlational, effect of higher rents.

As mentioned earlier, this paper uses data from the USDA on natural amenties index. This index is used to instrument and recover a first order estimate of the causal effect of rents on mobility. The logic is that while better climate may increase rents, the weather probably doesn't otherwise effect future otcomes for children. The result of the instrumental variable regression is given in Table 11.

Average household income are added as a control since higher incomes associated with rents could be a confounding variable. After this, our estimated causal effect is -0.0650 which is slightly lower in magnitude than the coeffecient of -0.0731 in the initial non-instrumental variable regression. This negative effect is what expected, which is good news for the soundness of the statistical model. Tellingly, the R^2 coeffecient (0.222) for the instrumental variable regression is only slightly higher than the R^2 in the non-IV regression

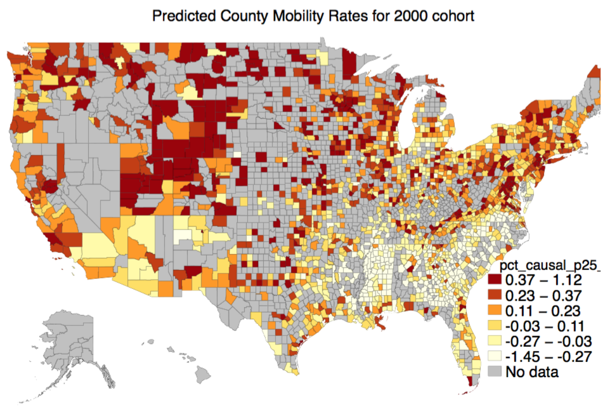

Figure 9: County Mobility Predictions for 2000 cohort

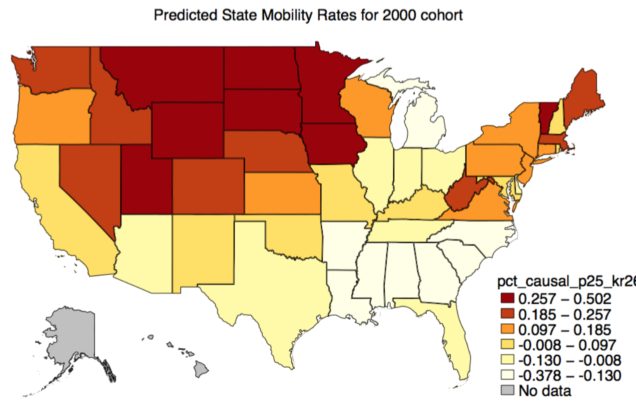

Figure 10: State Mobility Predictions for 2000 cohort

(0.221), suggesting that the natural amenties index does not have much explanatory power of it's own and is a good instrument. OV

3.6 Predicting Future Mobility

As mentioned before, the Chetty mobility estimates are for the 1980-1986 cohort. However, one may be interested in what mobility rates would look like in counties today. Specifically, one may want to know which counties are likely to produce higher mobility rates and which counties are not. For obvious reasons, it would be impossible to directly measure now the effect of counties on earnings for people born only in the past couple decades. However, the results from regressions above can be used to make predictions. To do so, each of the variables are normalized by standard deviations from the mean, regressions are re-run using these standardized variables, and normalized variables in later years are then used to predict mobility rates and rankings. To provide the most accuracy, the isolation index of segregation is used to produce The results are in Figures 9, 10, 12, and 13.

Figure 9 gives predictions for individual county mobility rates whereas Figure 10 makes predictions by using state demographic rates. Both are using data from the year 2000 and so should be predictions for expected mobility rates of children born around 2000. These figures are compared to the Chetty 1980 mobility estimates in Figure 11 with the note that the coloring in the Chetty map is inverse of the coloring in our maps.

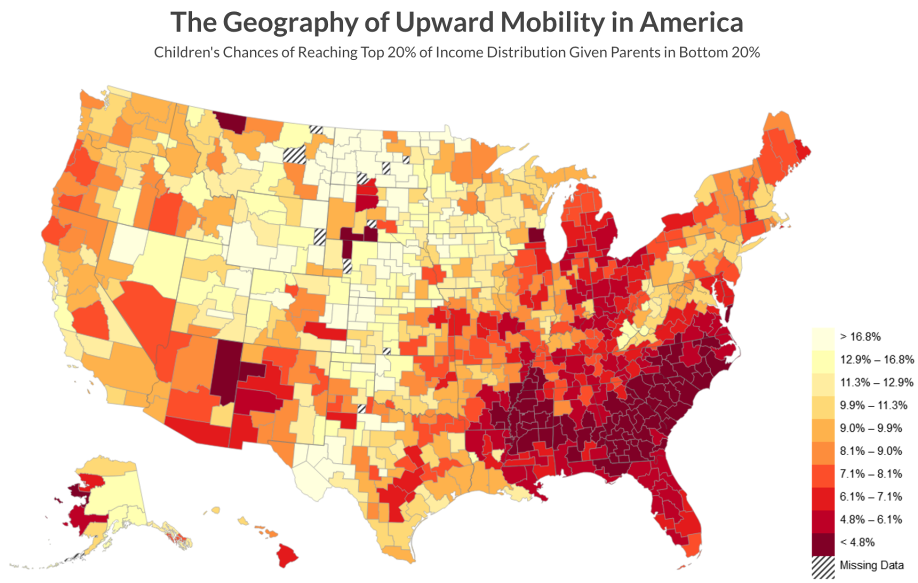

Figure 11: Chetty Mobility Estimates (source: http://equality-of-opportunity.org)

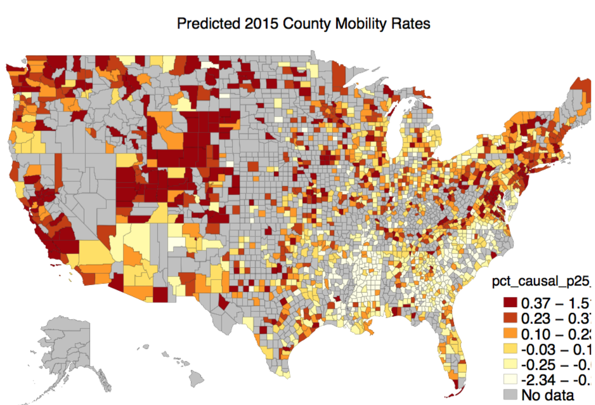

Figure 12: County Mobility Predictions for 2015 cohort

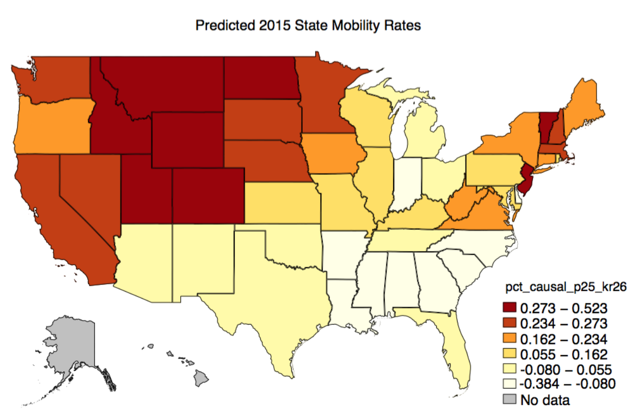

Figure 13: State Mobility Predictions for 2015 cohort

Table 12: Predicted County Mobility Ranks

Predicted 2000 Rank	Chetty Rank	County
1	4	Crook County, Wyoming
2	6	Fallon County, Montana
3	62	Millard County, Utah
4	12	Mercer County, North Dakota
5	496	Bath County, Virginia
6	96	Lincoln County, Wyoming
7	26	Emery County, Utah
8	270	Los Alamos County, New Mexico
9	9	Yuma County, Colorado
10	259	Toole County, Montana
...	...	...
Predicted 2015 Rank	**Chetty Rank**	**County**
1	51	Scott County, Kansas
2	1	Rio Blanco County, Colorado
3	48	Johnson County, Wyoming
4	312	Custer County, South Dakota
5	35	Dawson County, Montana
6	223	Teton County, Wyoming
7	4	Crook County, Wyoming
8	18	Weston County, Wyoming
9	204	Routt County, Colorado
10	5	Duchesne County, Utah
...	...	...

Mobility rates for children born in 2015 are given in Figures 12 and 13. The most mobile counties remain in the Midwest for all time period, with counties in the south also consistently performing poorly. Counties on the coasts performed about average on mobility metrics in the original Chetty estimates and predict that they stay about average in our 2000 and 2015 cohort predictions. Given earlier results, this paper presumes that much of the relatively poor performance of coastal counties on mobility metrics has to do with the high cost of living in those regions.

There have been some significant changes to our expected mobility map, however. Over time, there seems to be a move towards stronger mobility rates in the North-midwetern region, with states like Montana, Wyoming, and the Dakotas consistently having high predicted mobilities. In 2000 the top states for predicted mobility are Wyoming, North Dakota, Montana, South Dakota, and Utah. In 2015 they are Wyoming, North Dakota, Vermont, Colorado, and Idaho. By contrast the top states for mobility in 1980 paper are North Dakota, Wyoming, South Dakota, Utah, and Iowa. As noted before there is some consistency in the top states being midestern in all time periods (with the exception of Vermont in 2015), but there is still movement in rankings. The Spearman rank correlation between counties in 1980 and 2000 by mobility is 0.73 wheras the rank correlation between counties in 1980 and 0.5852. This is not altogether unexpected given changing demographics over the last 45 years, but it is still notable the degree to which the ranks have changed. Finally, our results predict that mobility rates have gotten more equitably distributed over the decades. The standard deviation of the original 1980 mobility metric is 0.4766 compared to a predicted mobility stadard deviation of 0.3407 in 2000 and 0.3758 in 2015. If future data collection on life outcomes of the 2000 and 2015 cohorts confirm these results, the reduction in mobility variance should be a topic of future research.

3.7 Migration and Mobility Rates

This last part looks at how mobility rates may affect migration. Migration could be a powerful stabilizing force for mobility rates. The logic is that high mobility rates in a county may encourage families to immigrate to the county, putting strain on local resources and lowering mobility rates. This would provide a long-run equilibrium where, under no cost of moving, homogenous mobility rates may be achieved across counties.

To study the relationship between migration rates and mobility, data from the census on the number of people that have moved to a county in a given year is used. I don't include people that have moved from out of the country in the past 5 years since immigrants may move for different reasons than residents. The number of people moving to a county is expressed as a percentage of total population to control for size effects. Then the percentage of people new to a county is regressed against the mobility metric , weighting by total

Table 13: Estimated linear relationship between mobility, inequality and migration rates

	(1)		(2)	
	movedto_2000		movedto_2000	
pct_causal_p25_kr26	0.0178^{***}	(5.03)		
gini_chetty			-0.0608^{**}	(-3.10)
_cons	0.202^{***}	(116.39)	0.228^{***}	(28.41)
N	1752		1736	
R^2	0.014		0.005	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

population. The results are displayed in Table 12.

Unfortunately, mobility rates explain little of the variance in immigration rates for counties ($R^2 \approx 0.06$). Further, the estimated coeffecient on mobility is small and positive, providing evidence for a weak linear relationship between the two as well. Interestingly, the regression of inequality against migration realizes a lower R^2 value of 0.005. The results of this regression are given at the bottom of Figure 13. This suggests that people are not making moving decisions based on the more immediately obervable inequality either.

4 Conclusion

For modeling intragenerational income dynamics, this paper should prompt further research into structural models of income dynamics. The AR(1) and ARIMA(1,0,1) models studied here have obvious shortcomings in explaining current levels of income dynamics and cannot explain falling income mobility without unreasonable parameter changes. The Guvenen model studied in section 3.4 does a better job of explaining current mobility rates by allowing shocks to acculumlate over time and incorporating the effect of age on earnings, but also struggles to explain the downward trend in mobility. All models studied have documented problems in explaining the high levels of inequality observed in the data. A possible avenue of exploration is incorporating wealth into structural models. The models studied do not differentiate income earned from work and from rents despite the fact that income from these sources may evolve differently with time. Given that the wealth distribution is highly unequal, incorporating it into models of dynamics could explain the income distribution's long right tail. Stronger structural models could better help inform policy by providing clearer answers as to why mobility has been falling over the last half-century.

This thesis on provides a starting place for future research on policies that can be implemented at a local level to increase mobility rates. It's clear that diversity of industry and amount of public assistance available in a region are associated with higher mobility rates. Combined with prior research that suggests that these variables can have impacts on children's future earnings, this provides basis for more experimental policy on the local level aimed at providing assistance to lower income households and caution against overreliance on individual industries. Controlling rents also seems to be an important factor in increasing mobility rates. Given increasing concerns about gentrification in urban areas, this could be a good focus for policy. Finally, it seems important to understand the relationship between mobility and migration. It is possible that people understand the tradeoffs and the cost of moving is to high to justify the gains. There is also some evidence from the *Moving to Opportunity* project that minorty families may be resistant to moving to areas with larger majority white populations due to fears of discrimination, even if these areas produce better outcomes. This is a larger societal problem but it is possible that there are economic interventions that could alleviate this. If people are not moving to areas because they are unaware of the importance of neighborhood effects on future earnings then effort should be made to disseminate this information.

References

[1] C Brown, VA Freedman, N Sastry, KA McGonagle, FT Pfeffer, RF Schoeni, and F Stafford. Panel study of income dynamics, public use dataset. *Ann Arbor, MI: University of Michigan*, 2014.

[2] Kristin Butcher. Assessing the long-run benefits of transfers to low-income families. In *Hutchins Center on Fiscal and Monetary Policy at Brookings' conference "From Bridges to Education: Best Bets for Public Investment," held at Brookings*, 2017.

[3] Raj Chetty, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly Journal of Economics*, 126(4):1593–1660, 2011.

[4] Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. Technical report, National Bureau of Economic Research, 2016.

[5] Michael Compson. The 2006 earnings public-use microdata file: an introduction. *Soc. Sec. Bull.*, 71:33, 2011.

[6] Miles Corak. Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, 27(3):79–102, 2013.

[7] Rebecca C Fauth, Tama Leventhal, and Jeanne Brooks-Gunn. Seven years later: Effects of a neighborhood mobility program on poor black and latino adults' well-being. *Journal of health and social behavior*, 49(2):119–130, 2008.

[8] Fatih Guvenen, Fatih Karahan, Serdar Ozkan, and Jae Song. What do data on millions of us workers reveal about life-cycle earnings risk? Technical report, National Bureau of Economic Research, 2015.

[9] Jonathan Heathcote. Fiscal policy with heterogeneous agents and incomplete markets. *The Review of Economic Studies*, 72(1):161–188, 2005.

[10] Douglas S Massey and Nancy A Denton. The dimensions of residential segregation. *Social forces*, 67(2):281–315, 1988.

[11] Jonathan Rothwell and Douglas Massey. Geographic effects on intergenerational income mobility. *Economic Geography*, 91(1):83–106, 2015.

[12] Patrick Sharkey. Neighborhoods, cities, and economic mobility. *RSF*, 2016.

[13] Anthony Shorrocks. Income inequality and income mobility. *Journal of Economic Theory*, 19(2):376–393, 1978.

5 Appendix

5.1 Descriptive Statistics for County Level Intergenrational Mobility Covariates

Table 14: Descriptive Statistics for 1980 variables

Variable	Mean	Std. Dev.	Min.	Max.
total_pop_1980	916201.503	1586150.987	3763	7477503
male_1980	0.485	0.013	0.412	0.625
white_1980	0.828	0.139	0.15	0.999
black_1980	0.121	0.126	0	0.842
amind_1980	0.005	0.02	0	0.657
asian_1980	0.014	0.02	0	0.217
highschool_1980	0.506	0.064	0.237	0.674
college_1980	0.168	0.063	0.028	0.478
dropout_1980	0.133	0.048	0.013	0.387
labforce_1980	0.473	0.042	0.263	0.673
manufacturing_1980	0.226	0.091	0.011	0.615
retail_1980	0.162	0.021	0.066	0.272
socialassist_1980	0.162	0.038	0.069	0.444
pubadmin_1980	0.053	0.035	0.013	0.36
hhinc_1980	38655.304	7738.883	14635	66344
hhearn_1980	0.818	0.055	0.508	0.969
hhssinc_1980	0.251	0.059	0.071	0.573
hhpainc_1980	0.078	0.036	0.014	0.337
rentpct_1980	0.014	0.002	0.006	0.022
poverty_1980	0.129	0.062	0.028	0.524
N		1752		

Table 15: Descriptive Statistics for 1990 variables

Variable	Mean	Std. Dev.	Min.	Max.
total_pop_1990	1032378.264	1833856.139	3103	8863164
male_1990	0.487	0.013	0.443	0.598
white_1990	0.799	0.151	0.137	0.999
black_1990	0.125	0.128	0	0.862
amind_1990	0.007	0.022	0	0.718
asian_1990	0.029	0.037	0	0.291
highschool_1990	0.549	0.061	0.283	0.744
college_1990	0.211	0.079	0.037	0.534
dropout_1990	0.113	0.04	0	0.331
labforce_1990	0.772	0.029	0.611	0.897
manufacturing_1990	0.176	0.073	0.014	0.537
retail_1990	0.17	0.022	0.072	0.281
socialassist_1990	0.169	0.037	0.069	0.432
pubadmin_1990	0.048	0.029	0.013	0.29
hhinc_1990	42326.347	10599.939	13153	79496
hhearn_1990	0.809	0.056	0.531	0.963
hhssinc_1990	0.256	0.06	0.072	0.539
hhpainc_1990	0.074	0.036	0.013	0.303
rentpct_1990	26.295	2.274	14.7	35.1
poverty_1990	0.825	0.085	0.319	0.981
N		1752		

Table 16: Descriptive Statistics for 2000 variables

Variable	Mean	Std. Dev.	Min.	Max.
total_pop_2000	1121880.156	1935803.696	2837	9519338
male_2000	0.49	0.012	0.426	0.614

Continued on next page...

... table 16 continued

Variable	Mean	Std. Dev.	Min.	Max.
white_2000	0.745	0.165	0.131	0.993
black_2000	0.127	0.131	0	0.865
amind_2000	0.008	0.023	0	0.747
asian_2000	0.038	0.045	0	0.308
highschool_2000	0.555	0.075	0.276	0.747
college_2000	0.252	0.092	0.054	0.637
dropout_2000	0.099	0.038	0	0.333
labforce_2000	0.497	0.043	0.315	0.66
manufacturing_2000	0.139	0.066	0.007	0.486
retail_2000	0.117	0.015	0.041	0.269
socialassist_2000	0.201	0.039	0.096	0.471
pubadmin_2000	0.048	0.025	0.013	0.426
hhinc_2000	45887.717	11071.864	17813	85708
hhearn_2000	0.811	0.052	0.559	0.951
hhssinc_2000	0.249	0.057	0.08	0.544
hhpainc_2000	0.034	0.02	0.004	0.153
rentpct_2000	25.446	2.153	14	38
poverty_2000	0.161	0.077	0.021	0.524
N		1752		

5.2 Regressions of Covariates against Chetty (2016) County Level Mobility Estimates

5.2.1 (Approximate) Chetty Regressions

Table 17: Estimated Chetty Regression 2000

	(1)
	pct_causal_p25_kr26

entropy_2000	-0.108	(-1.71)
gini_chetty	-1.708***	(-12.24)
topinc_chetty	0.865***	(3.78)
oneparent_2000	-3.674***	(-22.12)
ski_1997	0.113***	(18.64)
white_2000	-1.358***	(-10.88)
black_2000	-1.358***	(-10.81)
amind_2000	-1.566***	(-7.97)
asian_2000	-4.250***	(-6.78)
constant	3.218***	(22.74)
N	2740	
R^2	0.681	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Approximated Chetty Regression 1990

	(1)	
	pct_causal_p25_kr26	
entropy_1990	-0.104	(-1.73)
oneparent_1990	-5.390***	(-32.63)
ski_1990	0.127***	(24.35)
white_1990	-1.348***	(-9.08)
black_1990	-1.129***	(-7.62)
amind_1990	-1.069***	(-5.35)
asian_1990	-1.798**	(-2.98)
constant	2.800***	(18.35)
N	2844	
R^2	0.689	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Approximated Chetty Regression 1980

	(1)	
	pct_causal_p25_kr26	
oneparent_1980	-6.731***	(-24.20)
white_1980	-0.315	(-1.95)
black_1980	-0.650***	(-3.83)
amind_1980	-0.742**	(-3.28)
asian_1980	4.904***	(6.36)
constant	1.649***	(9.95)
N	2844	
R^2	0.619	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2.2 Individual Chetty Regressors

Table 20: Gini Regression

	(1)	
	pct_causal_p25_kr26	
gini_chetty	-3.478***	(-25.88)
constant	1.547***	(30.04)
N	2740	
R^2	0.342	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 21: Top Income Regression

	(1)	
	pct_causal_p25_kr26	
topinc_chetty	-1.876***	(-6.97)

constant	0.392***	(14.41)
N	2740	
R^2	0.034	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 22: Entropy Regressions

	(1)		(2)	
	pct_causal_p25_kr26		pct_causal_p25_kr26	
entropy_2000	-0.0268	(-0.27)		
entropy_1990			0.0346	(0.34)
constant	0.246***	(3.93)	0.207**	(3.13)
N	2844		2844	
R^2	0.000		0.000	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 23: Single Parent Regressions

	(1)		(2)		(3)	
	pct_causal_p25_kr26		pct_causal_p25_kr26		pct_causal_p25_kr26	
oneparent_2000	-5.425***	(-51.43)				
oneparent_1990			-5.890***	(-47.54)		
oneparent_1980					-7.269***	(-38.22)
constant	1.736***	(56.38)	1.530***	(54.47)	1.409***	(45.77)
N	2844		2844		2844	
R^2	0.584		0.611		0.606	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 24: Black Share Regressions

	(1)		(2)		(3)	
	pct_causal_p25_kr26		pct_causal_p25_kr26		pct_causal_p25_kr26	
black_2000	-2.238***	(-47.46)				
black_1990			-2.276***	(-47.14)		
black_1980					-2.257***	(-47.62)
constant	0.439***	(44.37)	0.438***	(44.39)	0.436***	(44.18)
N	2844		2844		2844	
R^2	0.392		0.395		0.390	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2.3 Policy Regressions

Table 25: Policy Regression 2000

	(1)	
	pct_causal_p25_kr26	
socialassist_2000	1.093***	(4.86)
pubadmin_2000	-1.885***	(-5.34)
hhssinc_2000	1.864***	(11.12)
hhpainc_2000	-9.382***	(-15.15)
constant	-0.137*	(-2.11)
N	2844	
R^2	0.147	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 26: Policy Regression 1990

	(1)
	pct_causal_p25_kr26

socialassist_1990	0.626**	(3.16)
pubadmin_1990	-1.042***	(-3.82)
hhssinc_1990	2.451***	(15.18)
hhpainc_1990	-6.353***	(-29.34)
constant	-0.0408	(-0.68)
N	2844	
R^2	0.314	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 27: Policy Regression 1980

	(1)	
	pct_causal_p25_kr26	
socialassist_1980	-0.387*	(-2.28)
pubadmin_1980	-1.165***	(-4.36)
hhssinc_1980	2.081***	(13.43)
hhpainc_1980	-7.065***	(-33.42)
constant	0.326***	(5.95)
N	2844	
R^2	0.372	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2.4 Chetty Plus Regressions

Table 28: Chetty and Policy Regression 2000

	(1)	
	pct_causal_p25_kr26	
entropy_2000	-0.101	(-1.66)
gini_chetty	-2.015***	(-14.28)

topinc_chetty	1.512***	(7.20)
oneparent_2000	-4.603***	(-26.68)
ski_1997	0.114***	(17.38)
white_2000	-1.373***	(-10.96)
black_2000	-1.017***	(-7.95)
amind_2000	-1.808***	(-9.47)
asian_2000	-2.545***	(-5.51)
socialassist_2000	0.420**	(2.87)
pubadmin_2000	0.279	(1.20)
hhssinc_2000	1.253***	(10.66)
hhpainc_2000	3.666***	(8.58)
constant	2.897***	(20.55)
N	2740	
R^2	0.715	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 29: Chetty and Policy Regression 1990

	(1)	
	pct_causal_p25_kr26	
entropy_1990	-0.0807	(-1.36)
oneparent_1990	-5.902***	(-32.09)
ski_1990	0.117***	(21.49)
white_1990	-1.338***	(-8.36)
black_1990	-1.066***	(-6.79)
amind_1990	-1.168***	(-5.72)
asian_1990	0.0546	(0.11)
socialassist_1990	0.551***	(3.95)
pubadmin_1990	0.847***	(3.54)
hhssinc_1990	0.918***	(7.99)
hhpainc_1990	0.640**	(3.27)

constant	2.402***	(14.36)
N	2844	
R^2	0.705	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 30: Chetty and Policy Regression 1980

	(1)	
	pct_causal_p25_kr26	
oneparent_1980	-6.501***	(-22.99)
white_1980	-0.732***	(-4.35)
black_1980	-0.782***	(-4.44)
amind_1980	-0.924***	(-3.70)
asian_1980	5.148***	(5.95)
socialassist_1980	0.569***	(3.89)
pubadmin_1980	0.407	(1.60)
hhssinc_1980	1.154***	(9.66)
hhpainc_1980	-1.876***	(-8.53)
constant	1.696***	(9.76)
N	2844	
R^2	0.641	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2.5 Chetty Plus Regressions with Controls

Table 31: Chetty and Policy Regression with Controls 2000

	(1)	
	pct_causal_p25_kr26	
entropy_2000	-0.0721	(-1.24)

gini_chetty	-2.227***	(-14.98)
topinc_chetty	1.894***	(8.62)
oneparent_2000	-4.596***	(-27.51)
ski_1997	0.112***	(17.01)
white_2000	-1.246***	(-10.48)
black_2000	-1.006***	(-8.38)
amind_2000	-1.890***	(-9.97)
asian_2000	-0.726	(-1.67)
socialassist_2000	0.303	(1.77)
pubadmin_2000	0.649**	(2.88)
hhssinc_2000	0.347**	(2.68)
hhpainc_2000	1.900***	(4.15)
density_2000	0.0000157**	(3.25)
hhinc_2000	-0.0000143***	(-11.90)
rentpct_2000	-0.0182***	(-6.36)
college_2000	0.213	(1.37)
constant	4.058***	(24.96)
N	2740	
R^2	0.743	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 32: Chetty and Policy Regression with Controls 1990

	(1)	
	pct_causal_p25_kr26	
entropy_1990	-0.0953	(-1.61)
oneparent_1990	-5.674***	(-32.38)
ski_1990	0.117***	(20.99)
white_1990	-1.113***	(-7.46)
black_1990	-0.932***	(-6.52)
amind_1990	-1.129***	(-5.83)

asian_1990	1.185*	(2.05)
socialassist_1990	0.670***	(3.90)
pubadmin_1990	0.980***	(4.13)
hhssinc_1990	0.324*	(2.45)
hhpainc_1990	0.0239	(0.10)
density_1990	0.0000237**	(2.67)
hhinc_1990	-0.00000867***	(-7.34)
rentpct_1990	-0.0126***	(-5.27)
college_1990	-0.175	(-1.06)
constant	2.970***	(17.96)
N	2844	
R^2	0.720	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 33: Chetty and Policy Regression with Controls 1980

	(1)	
	pct_causal_p25_kr26	
oneparent_1980	-6.906***	(-26.62)
white_1980	-0.650***	(-4.13)
black_1980	-0.716***	(-4.55)
amind_1980	-0.832***	(-3.47)
asian_1980	2.868**	(2.94)
socialassist_1980	0.192	(1.08)
pubadmin_1980	0.455*	(1.98)
hhssinc_1980	1.131***	(7.63)
hhpainc_1980	-1.705***	(-6.33)
density_1980	0.0000372**	(2.84)
hhinc_1980	-0.00000676***	(-3.78)
rentpct_1980	-20.49***	(-4.56)
college_1980	1.051***	(5.12)

constant	2.103***	(10.93)
N	2844	
R^2	0.657	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2.6 Industry Regressions

Table 34: Industry Regressions 2000

	(1)	
	pct_causal_p25_kr26	
manufacturing_2000	-1.077***	(-9.22)
retail_2000	-3.183***	(-4.81)
constant	0.783***	(8.79)
N	2844	
R^2	0.038	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 35: Industry Regressions 1990

	(1)	
	pct_causal_p25_kr26	
manufacturing_1990	-1.954***	(-18.06)
retail_1990	-3.383***	(-8.59)
constant	1.167***	(14.23)
N	2844	
R^2	0.123	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 36: Industry Regressions 1980

	(1)	
	pct_causal_p25_kr26	
manufacturing_1980	-1.887***	(-18.84)
retail_1980	-2.052***	(-5.08)
constant	0.960***	(11.84)
N	2844	
R^2	0.144	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6 Chetty Plus Regressions with Controls, Industry

Table 37: Chetty Plus with Controls, Industry 2000

	(1)	
	pct_causal_p25_kr26	
entropy_2000	-0.113*	(-2.16)
gini_chetty	-2.533***	(-15.64)
topinc_chetty	2.170***	(8.58)
oneparent_2000	-3.956***	(-23.46)
ski_1997	0.107***	(17.35)
white_2000	-0.629***	(-5.34)
black_2000	-0.477***	(-4.12)
amind_2000	-1.529***	(-8.29)
asian_2000	0.534	(1.17)
socialassist_2000	-0.421*	(-2.42)
pubadmin_2000	-1.296***	(-5.74)
hhssinc_2000	-0.00843	(-0.07)
hhpainc_2000	1.328**	(3.13)
density_2000	0.0000122***	(3.50)

hhinc_2000	-0.0000137***	(-11.77)
rentpct_2000	-0.0194***	(-6.94)
college_2000	-0.511***	(-3.42)
manufacturing_2000	-1.579***	(-16.21)
retail_2000	-1.358***	(-4.05)
constant	4.341***	(26.13)
N	2740	
R^2	0.774	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 38: Chetty Plus with Controls, Industry 1990

	(1)	
	pct_causal_p25_kr26	
entropy_1990	-0.126*	(-2.20)
oneparent_1990	-4.647***	(-25.96)
ski_1990	0.106***	(19.71)
white_1990	-0.0469	(-0.33)
black_1990	-0.0887	(-0.66)
amind_1990	-0.510**	(-2.66)
asian_1990	2.455***	(4.41)
socialassist_1990	-0.168	(-1.03)
pubadmin_1990	-0.956***	(-3.93)
hhssinc_1990	-0.0866	(-0.69)
hhpainc_1990	-0.125	(-0.54)
density_1990	0.0000162*	(2.28)
hhinc_1990	-0.00000683***	(-6.02)
rentpct_1990	-0.0137***	(-6.03)
college_1990	-0.935***	(-5.83)
manufacturing_1990	-1.595***	(-19.78)
retail_1990	-1.834***	(-7.89)

constant	2.797***	(18.86)
N	2844	
R^2	0.763	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 39: Chetty Plus with Controls, Industry 1980

	(1)	
	pct_causal_p25_kr26	
oneparent_1980	-5.319***	(-20.77)
white_1980	0.712***	(4.05)
black_1980	0.330	(1.92)
amind_1980	-0.0944	(-0.33)
asian_1980	4.051***	(4.10)
socialassist_1980	-0.588***	(-3.67)
pubadmin_1980	-1.370***	(-5.90)
hhssinc_1980	0.717***	(5.13)
hhpainc_1980	-1.126***	(-4.28)
density_1980	0.0000306**	(2.76)
hhinc_1980	-0.00000246	(-1.51)
rentpct_1980	-29.58***	(-7.58)
college_1980	0.145	(0.76)
manufacturing_1980	-1.733***	(-23.87)
retail_1980	-1.510***	(-5.73)
constant	1.531***	(8.00)
N	2844	
R^2	0.731	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7 Total Regression

Table 40: Total Regression

	(1)	
	pct_causal_p25_kr26	
entropy_2000	-0.0752	(-0.49)
gini_chetty	-2.024***	(-13.82)
topinc_chetty	1.712***	(8.06)
oneparent_2000	-2.097***	(-7.47)
ski_1997	0.0581***	(4.92)
white_2000	-0.166	(-0.65)
black_2000	-0.0701	(-0.17)
amind_2000	-3.238**	(-3.27)
asian_2000	2.676*	(2.29)
socialassist_2000	0.156	(0.58)
pubadmin_2000	-0.712	(-1.77)
hhssinc_2000	-0.646**	(-2.73)
hhpainc_2000	1.487**	(3.15)
density_2000	-0.000222***	(-4.58)
hhinc_2000	-0.00000556*	(-2.11)
rentpct_2000	-0.00723**	(-2.66)
college_2000	-0.941**	(-2.72)
manufacturing_2000	0.113	(0.66)
retail_2000	-0.564	(-1.59)
entropy_1990	0.0138	(0.09)
oneparent_1990	-1.105**	(-3.11)
ski_1990	0.0105	(0.94)
white_1990	-0.226	(-0.72)
black_1990	1.334	(1.93)
amind_1990	4.155***	(3.38)
asian_1990	-2.760	(-1.35)
socialassist_1990	0.193	(0.59)

pubadmin_1990	0.0658	(0.13)
hhssinc_1990	0.575	(1.88)
hhpainc_1990	-0.655	(-1.94)
density_1990	0.000124***	(3.43)
hhinc_1990	-0.00000371	(-1.27)
rentpct_1990	-0.00679**	(-2.99)
college_1990	0.375	(0.74)
manufacturing_1990	-0.286	(-1.40)
retail_1990	-0.567	(-1.78)
oneparent_1980	-1.212**	(-3.24)
white_1980	0.167	(0.54)
black_1980	-1.374**	(-2.80)
amind_1980	-2.105**	(-2.70)
asian_1980	1.158	(0.57)
socialassist_1980	-0.994***	(-3.82)
pubadmin_1980	-0.699	(-1.88)
hhssinc_1980	0.314	(1.44)
hhpainc_1980	0.345	(1.09)
density_1980	0.000145***	(3.39)
hhinc_1980	0.00000198	(0.85)
rentpct_1980	-5.121	(-1.45)
college_1980	0.586	(1.32)
manufacturing_1980	-1.335***	(-8.95)
retail_1980	-0.0629	(-0.21)
constant	3.374***	(15.97)
N	2740	
R^2	0.814	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$