UNIVERSITY OF TECHNOLOGY SYDNEY

ECONOMICS DISCIPLINE GROUP

UTS BUSINESS SCHOOL

WORKING PAPER NO. 17 January 2014

A Brief History of Equality

Geoffrey Brennan
Gordon Menzies
Michael Munger

ISSN: 2200-6788
http://www.business.uts.edu.au/economics/

A Brief History of Equality

Geoffrey Brennan
Australian National University and Duke University

Gordon Menzies
University of Technology, Sydney
gordon.menzies@uts.edu.au
(corresponding author)

Michael Munger
Duke University

Abstract

We explicate an iron law of intergenerational transmission of income dispersion. The same mechanism that limited income disparities, as population and prosperity increased through much of the early industrial revolution, will now sharply exaggerate inequality. The reason is that, for the first time in human history, richer parents are having fewer surviving children. Moreover, the effects of this fact in a setting like the current, where average family size is small and economic growth is strong, are quite marked. The social contract implicit in free market liberalism may require ongoing policy intervention to lean against the scolding winds of inequality.

"It is the great multiplication of the productions of all the different arts, in consequence of the division of labour, which occasions, in a well-governed society, that universal opulence *which extends itself to the lowest ranks of the people*." [our emphasis]

Smith (1776) I.1.10

"The rich… are led by an invisible hand to make nearly the same distribution of the necessaries of life which would have been made had the earth been divided into equal portions among all its inhabitants"

Smith TMS IV.I.10

The aim of this paper is to propound and briefly explicate what we shall term the "iron law of Intergenerational Transmission of income Dispersion" (ITD). In so doing we will show that the optimism of Adam Smith about income distribution, though well founded, was idiosyncratic to his own time and place. As with the iron laws of Smith's contemporaries (eg. Malthus (1798)),[1] ours depends on an underlying mechanism that is simple, and systematic -- and "invisible" in the Smithian sense that its effects emerge independently of the intentions (perhaps even of the knowledge) of the participating agents.

Our iron law postulates a relationship between the income elasticity of family size and the Gini coefficient (or comparable measure of income inequality) in subsequent periods. Its operation depends only on the familiar fact that individuals care more for their own children than they do for others' children. Specifically, parents make transfers throughout their lives (as well as at death) to their own offspring. However, because the total of such transfers within any family has to be divided among the number of children, the larger the number of children the smaller the average transfer to each, ceteris paribus.

Now, if the number of surviving offspring is *positively* related to parental income—as was true in Smith's era, and was true for much of human history until the second half of the C18th – then richer families will have a larger number of effective claimants on their lifetime inter-generational transfers. So, while the wealthier family transfers more wealth, the transfer will be divided across more people. The net effect of this process across the whole economy will be towards greater equality over time. Incomes will not converge entirely, of course, but tendencies toward increasing inequality are sharply damped.

If, however, the number of surviving offspring is *negatively* related to parental income, then richer families will have fewer claimants on their intra-family transfers than will poorer families, and this will be a force for greater inequality in subsequent generations.

It is important to emphasize that the transfers we have in mind include not just material resources but other scarce 'resources' like parental attention and energy. These resources are in turn associated with the acquisition of human capital. An extensive sociological literature attests that children from small families are more likely to excel in education than children from larger families (independent of any effects of parental economic status[2]). Employing a meta-analysis of no less than eleven US national surveys in Blake (1989), Alexander and Cherlin (1990) outline the advantages that accrue to children in relatively small families:

> "… such youngsters tend to go further through school….; they benefit from enhanced verbal skills; they and their parents tend to hold higher expectations for their ultimate

[1] Malthus devoted a key portion of *an Essay on the Principle of Population* to what he called the Iron Law of Population, which is his famous assertion that growing population would increase the supply of labour and lower wages, entrenching poverty. Marx (1875) discusses the origin of 'Iron' in the phrase, attributing it to Lassalle and, originally Goethe (1783).

[2] To which family size may be correlated.

> educational attainment …; they are more likely to engage in a rich variety of intellectually and culturally broadening activities at home (e.g., to have music and dance lessons, to travel abroad, to be read to by their parents early in life) and at school (e.g., they are more likely to be active in student government, on the school newspaper, in drama groups and the like); and they seem to be more confident of their academic ability, even more so than would be expected from their superior test performance."   Op. cit. pg. 344

Nor are such advantages confined to the developed world. According to Dang (2013), the recent burgeoning of private tutoring among Vietnamese families (a better measure of human capital investment than enrolments) is far more prevalent among smaller families, holding other factors constant.

As is the case with other "iron laws", the ITD clearly does not capture every factor that goes into the determination of the income distribution in any period. Changes in: factor prices; the age at which children become 'productive' (and more generally the time profile of marginal productivity); the public/private mix in human capital acquisition; and the extent and nature of (other) public transfers; will all exercise some influence on the income distribution. In fact, at any given point in time, one or more of these might exert a larger influence than does the ITD mechanism. However, the iron law is constant and ineluctable, exercising its influence at the margin, generation after generation. What sets it apart among the various sources of inequality at any point is less its empirical weight than its relentlessness.

Of course, any systematic influence on the distribution of income (or lifetime well-being) is important. But there is a particular story to tell about the time-path of the income elasticity of family size – a story that bears on our title. Until the last 150 years, human population has been regulated largely by the premature death (before the age of five) of actual births. Higher income *countries* have had higher populations, because higher income protects children from the effects of malnutrition and the vagaries of disease. And this inter-country phenomenon has its intra-family analogue. Wealthier parents have been better placed than poorer families to shield their children from a variety of deadly threats and crippling conditions.[3] In short, for most of human history, supply-side effects dominated population size: numbers of births emerged as an incidental consequence of sexual activity, while numbers of surviving children depended on the capacity to thwart forces for premature death – a capacity that was a positive function of income. As a result, each of the children of richer families received a smaller proportion of their parents' lifetime transfers than did children of poorer families. As long as the income elasticity of fertility was greater than zero, ITD produced an ever-expanding 'middle' class, and perhaps justified Smith's rather extravagant optimism about distributional questions evident in the epigraph.

But as the demographic transition developed and contraception techniques became more widely available, almost all wealthy countries witnessed both sharply reduced rates of infant mortality and lower birth rates. Demand-side factors came to dominate supply-side factors in the determination of population (and family size), and consequently the general relation between family size and income changed. Specifically, the sign on the income elasticity reversed: on average, richer families have *fewer* children than poorer. What this means is that the very process that once limited inequality actually tended sharply to magnify income disparities. In the developed world in particular, but increasingly elsewhere as well (except where family size is legally regulated), the most likely picture of the future

[3] Again, Adam Smith makes the point nicely: "…among the inferior ranks of people…the scantiness of subsistence …set(s) limits to the further multiplication of the human species; and it can do so in no other way than by destroying the greater part of the children which their fruitful marriages produce. The liberal reward of labour by enabling them to provide better for their children and consequently to bring up a greater number naturally tends to widen and extend those limits." Smith (1776) I.viiii.39/40

seems to be one of ever-increasing *in*equality. And as we shall show, in a period of relatively low population growth, the effect sharply concentrates wealth in relatively few hands, causing (under plausible parameter values) dramatic social changes in just a few generations.

Hence our title.

Of course, other factors may intervene to offset the ITD process – though it is worth emphasizing that that possibility is no more likely than the reverse. Unless there are changes in the income elasticity of family size, the period of greater income equality among the human species will turn out to be have been a short-lived product of a special set of dynamic conditions. There is no obvious means by which these conditions would tend to reassert themselves.

The plan of the paper is as follows. In section 1 we illustrate the effects of ITD in a simple Taylor series expansion of income per child. In section 2 we trace the sign of a key variable in the ITD – the covariance between children and income – through history, drawing on recent scholarship. In section 3 we derive some analytic results about intergenerational inequality. In section 4 we compare simulated ITD changes in Gini coefficients to movements across actual economies to gain a sense of the ITD's importance. Section 5 concludes.

One final methodological point is in order. The dynamics of the ITD process are simple. And our exposition here is designed to make that fact evident. Accordingly, we have focused *directly* on the connection between the income elasticity of family size and the subsequent period Gini coefficient. It would have been possible to attempt to derive a model of family size based on an explicit behavioural microfoundations. Then the elasticity of family size itself could be derived from more fundamental parameters. But any such model would have to include: the desire for sex; the probability of sexual activity producing offspring; the probability of children born surviving given income constraints; the natural time limits on aggregate number of births for any mother; as well as the opportunity cost of parental time, and the demand for children as such. By dint of such a model, one may be able to 'explain' the demographic transition itself – a task that has been famously and fascinatingly pursued by Gary Becker in his various collaborations. But any such exercise would, we think, background precisely what it is that we wish to foreground – namely the effects which that income elasticity of family size will have on the future level of income dispersion over successive generations.[4] We think that this is an important relation in its own right and operates independently of how the particular value of the income elasticity comes about. Thus our focus is on the consequences of the dynamic process, rather than its behavioural microfoundations

1. Divided We Stand

Consider an economy where workers' income draws in generation *t+1* depend positively upon the income of their parent in generation *t*, divided by the total number of children *N* in their family in generation *t*. Equation (1) shows this relationship for a single family.

$$Y_{t+1} = f\left(\frac{Y_t}{N_t}\right) \qquad f'>0 \tag{1}$$

We are abstracting from a number of things here to make a point. We refer to a single (asexual) parent as do Becker and Tomes (1979). We are letting income stand in for the whole resource pool potentially available for children, including time and attention. We ignore the effects of macroeconomic

[4] We eschew the modeling exercise partly because there seem to be a number of puzzles involved here – for example, the sign of the income elasticity of family size changed (apparently) long before the widespread availability of contraceptive devices made possible a strict separation between sexual activity and childbearing.

growth, or any other factor that determines the function *f*, such as parents' taste for children's welfare,[5] on the earning capacity of subsequent generations. For our purposes, this is without loss of generality since a Gini coefficient would be unaffected by scaling up all $t+1$ incomes by the same factor.[6]

In an earlier age, the earning capacity of children often depended on their physical strength, partly a function of their size and vigour. These in turn were related to the share of food that adults had had access to as children, because food itself was relatively homogenous for all but the wealthiest families. In a stylized representation, in generation *t* the children consume an undifferentiated commodity called nutrition and each gains strength $f(Y_t/N_t)$. When they become adults in generation $t+1$, they no longer need to eat, start working, and devote themselves to feeding their children before they die. In the modern world, the most natural justification for $f(Y_t/N_t)$ lies in the division of resources available for education, as we flagged in the introduction. We linearize Y_{t+1} around the means of workers' income and offspring, *y* and *n*, and take the variance of this linearization as a measure of inequality in $t+1$.[7] Equation (2) refers to the variance for the whole economy, calculated over the distribution of individual family draws from (1).

$$\begin{aligned} Inequality_{t+1} &= V(Y_{t+1}) \approx \left(\frac{f'}{n}\right)^2 \left[\left(\frac{1}{n}\right)V(Y_t) + \left(\frac{-y}{n^2}\right)^2 V(N_t) + 2\left(\frac{-y}{n^2}\right)\left(\frac{1}{n}\right)Cov(Y_t, N_t)\right] \\ &= \left(\frac{f'}{n}\right)^2 \left[V(Y_t) + \left(\frac{y}{n}\right)^2 V(N_t) - 2\left(\frac{y}{n}\right)Cov(Y_t, N_t)\right] \qquad (2) \end{aligned}$$

Noting that $V(Y_t)$ is inequality in generation *t*, (2) is a difference equation in inequality. The possibility of a steady state will depend upon the magnitude of f'/n. For ease of discussion we let $f'/n=\phi$ where ϕ is a fixed parameter. Integrating $f'/n=\phi$ with respect to Y/N we write (1) with a linear *f*:[8]

$$Y_{t+1} = f\left(\frac{Y_t}{N_t}\right) = \phi n\left[\frac{Y_t}{N_t}\right]. \qquad (3)$$

Scaling up each child's income by *n* in (3) rules out implausible ever-declining incomes over generations for the case when *N* exceeds one, and neatly meets a subtle mathematical objection to (2).[9] Furthermore, the experience over the last two centuries can be parameterized by $\phi>1$. If n/N is

---

[5] An altruistic parent coalition ('parents') consumes p_t of income Y_t leaving children c_t. Let parents have c.e.s. preferences $U=[\alpha c_t^{\gamma} + (1-\alpha) p_t^{\gamma}]^{1/\gamma}$ where α is the taste for children's welfare. Since the relative price for an income division problem is unity, optimal c_t is $Y_t/[1+\{(1-\alpha)/\alpha\}^{1/\gamma}]$ and clearly if there are N_t children, each child receives $Y_t/N_t[1+\{(1-\alpha)/\alpha\}^{1/\gamma}]$. In the limit generous parents give children everything (α→1) and ungenerous parents give them nothing (α→0) but for any value of α, a child's resource is a function of the ratio Y_t/N_t , as in (1).

[6] The Gini coefficient measures the area between the Lorenz curve and a line of hypothetical perfect equality, expressed as a percentage of the maximum area under the line. Thus a Gini coefficient of 0 represents perfect equality (all citizens the same), while a coefficient of 100 implies perfect inequality (one citizen possesses all the wealth in the whole society).

[7] Policy discussions sometimes conflate inequality, which refers to the outcomes of the income distribution, with mobility, which refers to the chance an individual (or their children) have to move around the distribution. In an earlier version of this paper, we included a discussion of how the ITD mechanism bears on the relation between the size of the Gini coefficient and the degree of social mobility (recently highlighted by the 'Great Gatsby' curve in Kreuger (2012)). That there is an interesting relation here, and that the ITD has relevance to that relation, seem to us to be indisputable. But the details are complex enough to require a more focused and independent treatment than can be provided here.

[8] For simplicity, $[1+\{(1-\alpha)/\alpha\}^{1/\gamma}]$ is missing in (3) though it could easily be incorporated into ϕ.

[9] As discussed in Appendix 1, the pdf of children's incomes is scaled by family size, and the variance in (2) must be calculated over *that* distribution, not the original distribution of income. Thus, if a cohort of parents accounted for, say, the bottom decile of income, and on average the relationship between income and family size is negative, the cohort of their children will account for *more* than the bottom decile because they have relatively few children.

approximately unity, (3) says children are richer than their parents if ϕ exceeds unity, and, income is stagnant across generations if ϕ is equal to unity. We rewrite (2) using (3) for *f*.

$$V(Y_{t+1}) == \phi^2 \left[V(Y_t) + \left(\frac{y}{n}\right)^2 V(N_t) - 2\left(\frac{y}{n}\right) Cov(Y_t, N_t) \right] \tag{4}$$

Working our way through the RHS of (4), ϕ – our secular intergenerational growth in wages – connects ITD to growth theory. Higher values of ϕ will increase inequality across generations. Given that strong economic growth is quite common in the modern world, equation (4) reminds us that we cannot simply grow our way out of the inequality magnified by the ITD.

It is clear from the first term in the brackets that current inequality (high $V(Y)$) implies more future inequality. Thus, those who propose tackling future inequality via current policy interventions focused on income redistribution (for example, Krueger (2012) and Corak (2012)) can make a demographic argument, using the ITD.

Less obviously, coming to the second term, increased variation in children numbers per family also increases inequality, but it is scaled up by the square of y/n. The explanation for this is mathematically intuitive. Since N is a random variable, a low mean value carries with it the possibility of a near-zero draw on N – a tail event of super-rich children in the next generation – which would dramatically increase inequality. Thus the second term of (4) implies a *Forbes kids' effect* of division.[10]

This is rather different from the popular narrative of concern about the implications for inequality of large low-income families. According to (4) an extra child born into a low income family with an already high n is unimportant for the transmission of inequality to the next generation, since y/n is already small. Economically, all the children in the family are already relatively poorly endowed, so one more makes little difference to the future earnings of each, at least as far as the ITD is concerned.[11] This is quite different from a wealthy family having one less child, since the average is already less than two. Relatively small changes in child-bearing by the wealthy have large effects; small changes by the poor are much less impactful.

Although departing from the strict variables of the model, focussing on the number of children in families is a reasonable representation of the underlying dynamics. Consider the human capital effects on children of decisions one can make about a particular child – the choice of school, the place of abode, or the communities that the family could become involved in. All of these parental decisions affect a child's wealth, at the margin. However, from an economic perspective the rates of return from these investments are likely small compared to the return from a decision to refrain from having an extra child, vying for parental attention and resources. If rich families act in this way, the resultant cohort of Forbes kids will have a big impact on inequality in the next generation leading to a 'circulation of elites' (Pareto, 1935).

So far, we have focussed on the first two terms in equation (4), ignoring any covariance between N and Y. But if two random variables are negatively correlated their ratio will be more volatile than if they are positively correlated. Negative correlation implies that a large numerator draw will be divided by a small denominator draw, on average. Positive correlation means that the effects on the ratio of draws of the denominator will tend to be offset by draws of the numerator. Thus, in (4) a positive covariance will reduce the variance in generation t and therefore the inequality in generation $t+1$, whereas a negative

[10] Our own expression: Forbes (http://www.forbes.com/forbes-400/) compiles lists of the super-rich.

[11] This is not to say that this is unimportant for utility, though. At very low levels of income, even small further reductions can have huge utility effects. Nevertheless, speaking outside the model, in a world with many other sources of disadvantage, the impact of an extra child in a low income family is likely to be swamped by other factors.

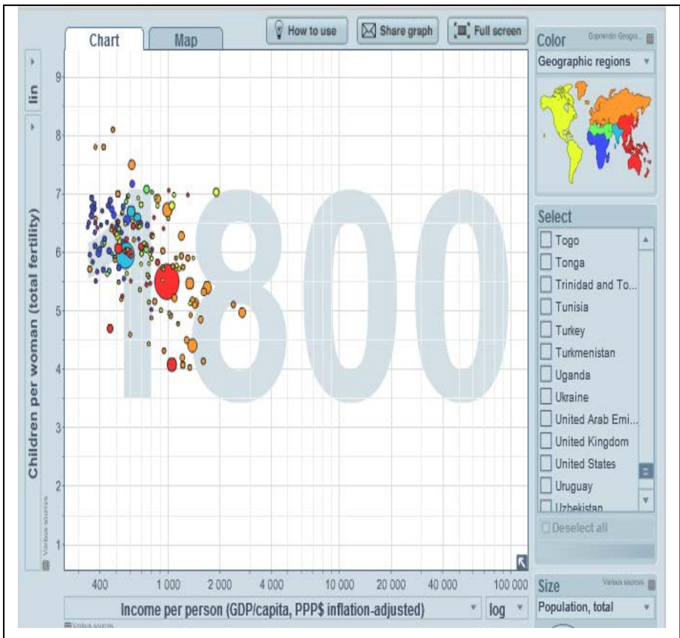
covariance will lift inequality in $t+1$. All other things equal, inequality will be greater when the covariance between N and Y is negative, inequality will take a midrange when they are uncorrelated, and inequality will be attenuated or damped when the covariance is positive.

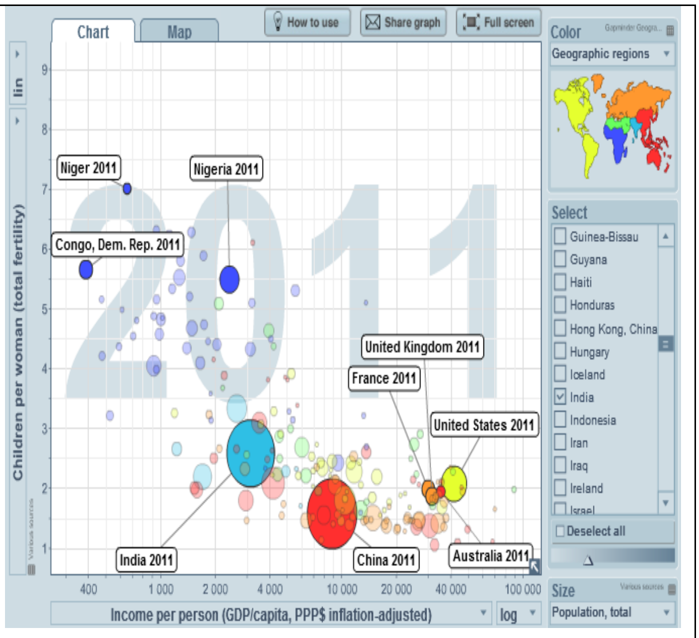
This is more than just a mathematical observation. The dependence of inequality on this covariance forges a powerful conceptual link between our concept of ITD and the "Economics of the Family" literature (Becker (1981)). The latter has a well-developed set of explanations for the relationship between incomes and numbers of offspring, a matter under consideration all the way back to Malthus (1798). But before formalizing this link, we need to acquaint ourselves with the history of the covariance term in (4).

2. A Brief History of Covariance

Our history of covariance must travel backwards in time, following the path of scholarship which, until recently, had not reached consensus about the situation prior to the industrial revolution. These snapshots from 'gapminder' show a negative relationship across countries in recent history:

Figure 1: Income and Number of Children



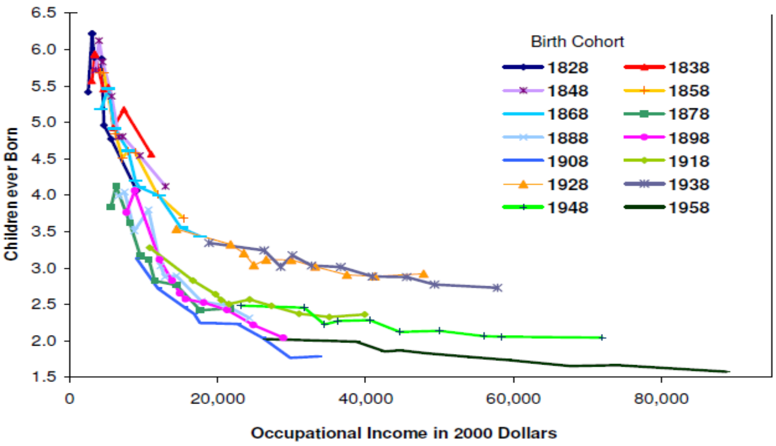


Source: http://www.gapminder.org/world

with the same pattern evident *within* the US (Jones et al. (2008), Jones and Tertilt (2008)).[12]

[12] Figures 1 and 2, and the forthcoming Table 1, refer to fertility rates, whereas the relevant quantity for our theorizing is surviving children, since a child must survive long enough to accumulate human capital or receive a bequest. In the modern era (Table 1) the two quantities are likely to be close. However, further back in history, a large discrepancy between the two is evident. Fortunately, when we come to discuss the pre-Industrial Revolution era, we use available data on surviving children, which is enough to secure the central claims of this paper.

Figure 2: Fertility by Occupational Income in 2000 Dollars



Source: Jones and Tertilt (2008)

Data on income and fertility within other economies is hard to come by, but if we use wealth as a proxy a negative relationship appears in emerging economies, with the fertility rates for some countries at the lower wealth quintiles not dissimilar to the C19th levels in the US from Figure 2.[13]

Table 1: Fertility by Wealth Quintiles

Fertility rates: Total fertility rate						
Country	Survey Year	Household wealth index				
		Lowest	Second	Middle	Fourth	Highest
Sth. Africa	1998	4.8	3.6	2.7	2.2	1.9
Peru	1991-2	7	4.8	3.3	2.5	1.5
Brazil	1996	4.8	2.7	2.1	1.9	1.7
Pakistan	2006-7	5.8	4.5	4.1	3.4	3.0
Ukraine	2007	1.7	1.3	1.3	0.9	1

Source: (ICFI, 2012) http://statcompiler.com/?share=91398D8C3B

In view of the above, it might seem reasonable to assert that the covariance term in (4) is negative, or perhaps near-zero (as for the Ukraine in Table 1). Indeed, until recently, it was though by economic historians that this might even have been the case just prior to the industrial revolution.[14]

However, Clark and Cummins (2010) have used wills from England over the 16th to 19th centuries to make a compelling case that there was a time when the within-economy relationship between income and the numbers of children was positive. They were able to impute estimates of lifetime earnings and numbers of surviving children for over 10,000 wills during a period where real incomes were broadly

[13] Under our ITD hypothesis, the income relationship is likely to be even more striking. As Becker and Tomes (1976) argue, human capital is likely to be dispensed by parents differentially for 'high quality' children, reinforcing income differentials, whereas non-human capital, which is important for household wealth, is likely to be more evenly dispensed.

[14] With regards to England, "The limited and contradictory earlier evidence on the relationship between wealth and fertility in pre-industrial England, and the fact that marriage ages and nuptuality were seemingly similar in 1850 to their earlier levels of many decades, created a false impression that the fertility regime of the mid nineteenth century [the poor having at least as many children as the rich] represented the entire pre-industrial period." pg. 2, Clark and Cummins (2010). See also their footnote 2.

stable. Figure 3 below, which is close to Figure 7 in Clark and Cummins (2010), shows some striking trends.[15]

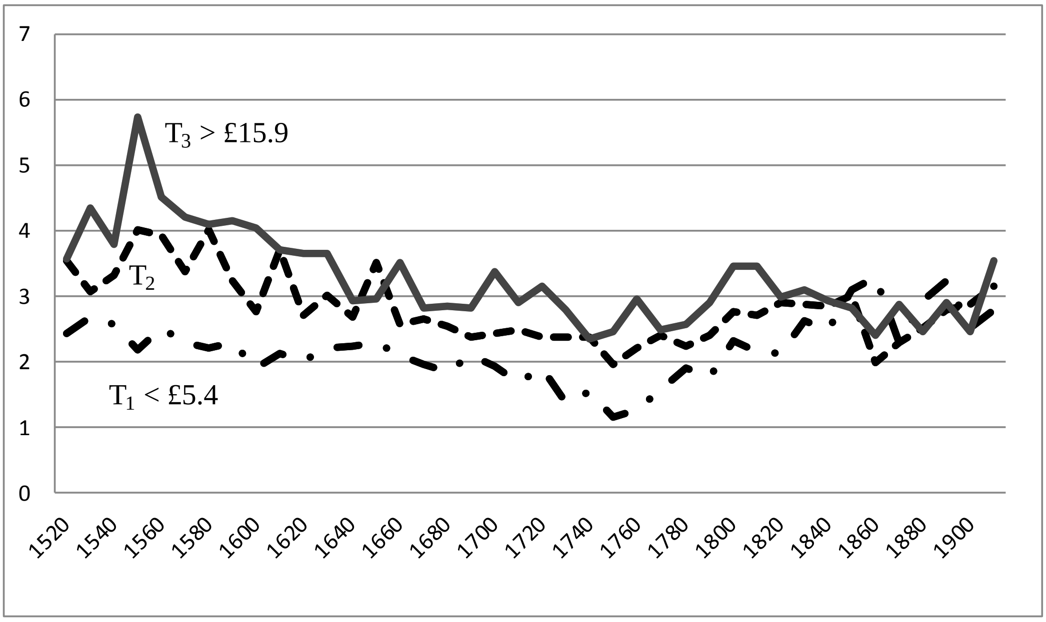
They defined income terciles below, between and above £5.4 and £15.9 (in 1630 prices; about £810 and £2385, respectively, in 2013 equivalents[16]) with their associated dummies T_1, T_2 and T_3.[17] They then divided the period 1520 to 1910 into 10-year sub periods and averaged the number of surviving children (at the time of the testator's death) for each income tercile. The figure suggests an era existed from 1520 up to and including the 1780s where the number of children in the top income tercile was significantly higher than in lower terciles.[18]

That is, the key point that Clarke and Cummins make is that prior to the early 1800s richer people seem to have more children than poor people, in contrast to the C20th. Not surprisingly, they call this the Malthusian era, but they note that the relationships break down as England moved into the industrial revolution era (which they date in the late 1700s). All income terciles show roughly the same numbers of children from the early 1800s onwards in our Figure 3.

Since this is such an important result, in its own right and for our analysis, we wanted to confirm their findings using their original data. The relevant hypothesis is that the average number of children in each tercile differs in the Mathusian era but not afterwards. To that end we added a Malthus dummy D to Clark and Cummin's tercile dummies T_1, T_2 and T_3. The Malthus dummy is unity up to the 1780s and is zero from 1790.

Figure 3: Surviving Children by Income Tercile

(Essex, Suffolk and Surrey over 1500s -1800s)



Source: Clark and Cummins (2010)

15 We are grateful for the scholarly and generous way in which they provided, and explained, their data to us.

16 http://www.measuringworth.com/ppoweruk/

17 It is sensible to define the terciles with the same figures over the whole period because income is stable. Only in the late C19th does this assumption start to weaken. The reader is referred to Clark and Cummins (2010) for a detailed discussion of their dataset. Suffice it to say that 'surviving children' is a smaller number, at death of the testator, than the numbers appearing in Figure 1. Further, although the sample size is very large (>10,000) it does not satisfy usual standards for random selection, given what can be accomplished in this day and age of central statistical offices. It is, nevertheless, a remarkable and informative dataset.

18 For the 1780s switch date, see Figure 8 of Clark and Cummins (2010).

The basic regression (supressing the error term) is:

$$N=\gamma_1+\gamma_2 T_2+\gamma_3 T_3+\gamma_4 D+\gamma_5 D.T_2+\gamma_6 D\ T_3. \quad (5)$$

This can be rearranged using the fact that $\gamma_1= \gamma_1 (T_1+T_2+T_3)$ and $\gamma_4 D= \gamma_4 D\ (T_1+T_2+T_3)$ to give:

$$N\ =\ T_1\{\gamma_1+\gamma_4 D\}+\ T_2\{\gamma_1+\gamma_2+(\gamma_4+\gamma_5) D\}+\ T_3\{\gamma_1+\gamma_3+(\gamma_4+\gamma_6) D\}. \quad (6)$$

Table 2 shows a battery of tests applied to (6). We run the regression both on the decade data of Figure 3 (cols. 2 and 3) and on the unit record data of every will (cols. 4 and 5). First, we test if the numbers of children in each tercile differ in the Malthusian era by setting $D=1$ in (6) and taking the relevant paired differences (see table notes). The overwhelming conclusion in rows 1 to 3 of the hypothesis tests (see the bottom left of Table 2) is that the number of children in every income tercile differed significantly from the other terciles at the 1 per cent level, establishing a Malthusian (i.e. positive) covariance between *Y* and *N* up to and including the 1780s.

Next, we tested whether this difference in children between the terciles was significantly greater in the Malthusian era compared with afterwards. We derived the tercile differences in the Malthusian era ($D=1$ in (6)) and subtracted the same differences in the post-Malthusian era ($D=0$ in (6)). We tested the significance of the relevant linear combinations of parameters. Here, the verdict from the individual data was that every tercile difference (T_2-T_1, T_3-T_1 and T_3-T_2) fell significantly after the Malthusian era at the 1 per cent level. Apart from the difference T_3-T_2, the evidence from the decade data support the same conclusion.[19]

Table 2: Regression Results

(* and ** are two-sided significance for 5% and 1%. D=1 in Malthusian era)

		Decade		**Individual**		Notes
Var	Coeff	obs=120	(s.e.)	obs=12314	(s.e.)	Tercile dummy coefficient differences:
C	γ_1	**2.5862	0.15	**2.5382	0.08	T_2-T_1= γ_2+ γ_5 D
T_2	γ_2	0.1174	0.22	*0.2180	0.10	T_3-T_1= γ_3+ γ_6 D
T_3	γ_3	0.3603	0.22	**0.3580	0.10	T_3-T_2=[γ_3– γ_2]+ [γ_6– γ_5] D
D	γ_4	** - 0.5746	0.19	** - 0.5348	0.09	The differences in the Malthusian era
D*T_2	γ_5	**0.7538	0.26	**0.4640	0.12	compared to afterwards are
D*T_3	γ_6	**1.0485	0.26	**0.7904	0.12	respectively γ_5, γ_6 and γ_6– γ_5.
Hypothesis Tests: H_1 Coefficient Combinations for F-tests						Hypotheses
γ_2+ γ_5>0		**0.8712	0.15	**0.6820	0.06	N_{T2}>N_{T1} in Malthusian era
γ_3+ γ_6>0		**1.4088	0.15	**1.1484	0.06	N_{T3}>N_{T1} in Malthusian era
[γ_3–γ_2]+[γ_6–γ_5]>0		**0.5377	0.15	**0.4664	0.07	N_{T3}>N_{T2} in Malthusian era
γ_5>0 as above		**0.7538	0.26	**0.4640	0.12	N_{T2}-N_{T1} gap falls after Malthusian era
γ_6>0 as above		**1.0485	0.26	**0.7904	0.12	N_{T3}-N_{T1} gap falls after Malthusian era
γ_6–γ_5>0		0.2947	0.26	**0.3264	0.11	N_{T3}-N_{T2} gap falls after Malthusian era

Notes: All tests are two sided, whereas the requirements of the different alternative hypotheses are one sided. However, when the estimated coefficients have the sign indicated under H_1, and H_0 is rejected with an F-test, these alternatives will certainly be rejected with a one sided t-test from an appropriately reparameterized regression, because the relevant p-value will halve.

[19] Based on this dataset with Clark and Cummin's 1780s cut-off, there is weak evidence of a positive tercile difference in children numbers in the post-Malthusian era. That is, using the individual data and setting D=0 in Table 2, T_2-T_1 (coefficient γ_2) is significant at the 5 per cent level, T_3-T_1 (γ_3) is significant at the 1% level and T_3-T_2 ($\gamma_3 - \gamma_2$, test not shown in Table 2) is not significant. However, the dataset contains only limited datapoints of the post-Malthusian era, and what significance there is almost certainly due to the ambiguity of where to make the cutoff. Looking at Figure 3, it beggars belief that any differences would be significant after, say, the 1840s.

If we are prepared to accept both the existence of a Malthusian era and what seems indubitable from Figure 2 – that by the late C20th the relationship between income and children had become negative – a well-supported narrative about a changing covariance emerges.

The sign of the covariance has undergone a journey from a positive number during Malthusian era, through zero during the industrial revolution, finally settling on a negative sign in what we might call the Modern era. The negative sign of the Modern era emerged in different times and places over the C20th, but its ascendency was assured, according to Becker (1981), by the advent of "the pill" and other reliable contraceptives in the 1960s.

3. ITD-Inequality in Three Eras

Armed with our covariance sign in the modern era (Figures 1 and 2, and Table 1) and the Malthusian era (Table 2) we now incorporate a stylized reduced-form relationship into our analysis. In the simple regression model (7), *u* represents those factors affecting fertility unrelated to income (*Cov(Y,u)=0*) and the sign of β_1 determines the sign of the covariance.

$$N = \beta_0 + \beta_1 Y + u \tag{7}$$

$$Cov\ (N,Y) = \beta_1 V(Y) \tag{8}$$

$$V(N) = \beta_1^{\,2}\, V(Y) + \sigma^2 \quad , V(u) = \sigma^2 . \tag{9}$$

As we have already noted, Becker (1981) argued for a negative value in the modern era when economic growth lifts the female wage, raising the opportunity cost of having children.[20] Becker and Tomes (1976) describe a negative β_1 as a shift from quantity to quality of children, where this shift is based upon a special relationship between the first order conditions of the two.[21]

However, in what follows, the sign of β_1 is more important than the micro-foundations that justify it. That is just as well because as we move across eras, the theoretical justification of the sign of β_1 changes. As flagged in the introduction, we may classify the Malthusian era (positive β_1) as being one where there was excess demand for children. So the short end of the market – supply – determined fertility, through factors such as food, hygiene and workplace danger for child labour. In the modern era, with abated threats to mortality, there has been excess supply of children and the short end of the market (demand) has determined fertility.

It turns out that we can travel a fair way by asserting nothing more than the sign of the covariance in (7), (8) and (9) when they are substituted into (4). In so doing, our framework economizes on assumed optimizing behaviour, and the centrality of a simple dynamic force, the ITD, is thrown into stark relief.

To track the evolution of inequality in the three eras – which we have designated Mathusian, Industrial Revolution and Modern – we substitute into (4) and let V_t stand for variance $V(Y_t)$ in generation *t*.

[20] There may be a combination of subsidiary factors that caused a negative relation between income and fertility. The expansion of opportunities for education and participation in the workforce increased the opportunity of the time spent on child-bearing and child-rearing for women, and the improvement in health care meant that the need for multiple children to ensure that any reached maturity was much reduced.

[21] In their model the shadow price of quality is the number of children multiplied by the price of child-quality, so one less child makes it cheaper to invest in the quality of children, reinforcing the desire to reduce fertility.

$$
\begin{aligned}
V(Y_{t+1}) \quad &= \phi^2 \left[V(Y_t) + \left(\frac{y}{n}\right)^2 V(N_t) - 2\left(\frac{y}{n}\right) Cov(Y_t, N_t) \right] \\
\Rightarrow V_{t+1} \quad &= \phi^2 \left[1 - \Lambda\right] V_t + \Omega \\
where \quad & \Lambda = \frac{\beta_1 y}{n}\left(2 - \frac{\beta_1 y}{n}\right), \quad \Omega = \left(\frac{y \phi \sigma}{n}\right)^2
\end{aligned} \tag{10}
$$

Using (10) to describe the three eras requires that we declare something about the parameter values in each era.

Based on Clark and Cummins' (2010) database, we assert that income was stagnant during the *Malthusian era*, and that wages were quite low. Together these imply a low value of y/n and a ϕ value of unity. As we have outlined above, the central claim of the ITD mechanism turns on the sign of β_1 which we assume to be positive in that era.

It turns out that since $n=\beta_0+\beta_1.y$, and both n and β_0 are assumed positive, the second term in the expansion of Λ is always positive since $\beta_1\, y/n$ is always less than unity. This implies that overall Λ takes the same sign as β_1. Thus, in the Malthusian era, where β_1 is positive, Λ is positive, and so, for an assumed $\phi = 1$, $\phi^2(1–\Lambda)$ must be less than unity. Thus, (10) has a steady state solution given by:

$$
\begin{aligned}
V \quad &= \left[1 - \Lambda\right] V + \Omega \\
V \quad &= \frac{\Omega}{\Lambda} = \frac{\sigma^2}{\beta_1\left(\frac{2n}{y} - \beta_1\right)}
\end{aligned} \tag{11}
$$

Thus, in the Malthusian era ITD-inequality converged to a steady state aided by no secular economic growth (ϕ=1), low income per capita (n/y being large in the denominator) and driven by innovations to child numbers unrelated to income (σ^2). There would also have been some tendency towards convergence across countries, according to (11) because the steady state does not depend on the starting values of inequality. Crucially, the sign of β_1 guarantees convergence. Economically, at least as far as the ITD was concerned, Smith's *universal opulence which extends itself to the lowest ranks of the people* was a result of the positive covariance between income and surviving children: Looking all the way up the income ladder, children of the ever-richer parents would climb down ever-more rungs in the next generation, some even reaching the lowest ranks of the people.

When we come to the Industrial Revolution, β_1 was a parameter in transition. Without committing to whether a parameter configuration for our model exactly matches what historians dub the timespan for the Industrial Revolution,[22] we will assert a period existed when economic growth was beginning but ϕ was still approximately unity. And, further, we will claim that β_1 was, over that same period, effectively zero implying the same fate for Λ. The solution to (10) now exhibits a trend.

$$
\begin{aligned}
V_{t+1} \quad &= \phi^2 \left[1 - \Lambda\right] V_t + \Omega = V_t + \Omega \\
V_t \quad &= V_0 + \Omega . t = V_0 + \left(\frac{y.\sigma}{n}\right)^2 t
\end{aligned} \tag{12}
$$

Thus the factors in (11) that determined the steady state now determine the growth out of the steady state. However, two important differences are that income per head was starting to grow, and so (12)

[22] One set of dates is 1760 – 1800 in Clark and Cummins (2010).

would, loosely speaking, start to accelerate. Secondly, V_0 in (12) implies that this equation does not share the convergence property of (11). If some countries did not get to a steady state described in (11) – or other non-ITD factors impacted upon their inequality to take them away from the steady state at the time the industrial revolution started – then county differences in inequality no longer had a tendency to be eliminated by the ITD operating through (11). Instead, whatever the state of their inequality, they trended from that point (V_0). The contentious history of global inequality was born.

It should be noted that this inequality was a form of the ITD process that was driven solely from the uncorrelated variance in the number of children, rather than β_1. That is, variation in the numbers of children, uncorrelated with income, created inequality, which transmitted in an undiscounted way onto the next generation.

In the Modern era (10) cannot be solved to give a steady state, or even a constant growth rate. It is still true that overall Λ takes the same sign as β_1 and so the switch to a negative covariance implies the root of the difference equation is outside the unit circle. As if that weren't enough, we now assume that ϕ exceeds unity, and together with growing y/n this scales up both the impact of the current income distribution on next periods distribution, and the effect of random fluctuations in children unrelated to income (σ^2).[23]

$$V_{t+1} = \phi^2.[1-\Lambda].V_t + \Omega$$
$$\text{where } \ \phi > 1, \ \Lambda < 0 \ \text{ and } \ \Omega = x^2.\sigma^2 \ \text{ and } \ x = \left(\frac{y.\phi.}{n}\right) \text{ is growing} \qquad (13)$$

Economically, all the factors that worked towards equality and a stable steady state in the Malthusian era have disappeared. We saw in the solution of (12) that a steady state inequality would be possible in the modern world only if β_1 returned to its (positive) Malthusian sign reinforcing the key mechanism of the ITD. However, that is a necessary, rather than a sufficient condition. In (12) the general growth in the economy—which impacts the model via ϕ––would need to be offset sufficiently if $\phi^2.(1-\Lambda)$ were to fall below unity. Otherwise, (13) obtains.

We conclude this section with a summary of the key parameter parameter configurations over our three eras, and their implications for ITD inequality.

Table 3: Inequality in Three Eras

Era	ϕ	*y/n*	β_1	$1-\Lambda$	**Converges ?**	**Intergenerational Inequality**
Malthus	1=stagnant	Low	> 0	< 1	yes	Stable steady state
Ind Rev	1^+= emergent	medium	= 0	= 1	no	Positive trend
Modern	>1 growth	high	< 0	> 1	no	Explosive

4. Numerical Estimates

In view of the preceding analysis, and especially since we cannot calculate a steady state in the modern world, it is natural to do some simulations for discrete periods. We begin by creating a stylized numerical equation for each of the Modern and Mathusian eras. For the modern era, we go to Figure 2.

[23] It may be true that the lower numbers of children in the modern world place a cap on the size of the uncorrelated variance *u* in (7), in which case σ^2 could be lower in (13). However, that is only a level issue and does not alter the dynamic explosiveness of (13) leading to higher inequality than the Malthusian era.

Based on the bottom (dark green line) which asymptotes to 1.6 for high incomes, and a reported elasticity of -0.22, we linearized $N=(1.6)Y^{-.22}$ around a (normalized) income of unity:

$$N = 1.952 - .352.Y \qquad 0 < Y < 1. \tag{14}$$

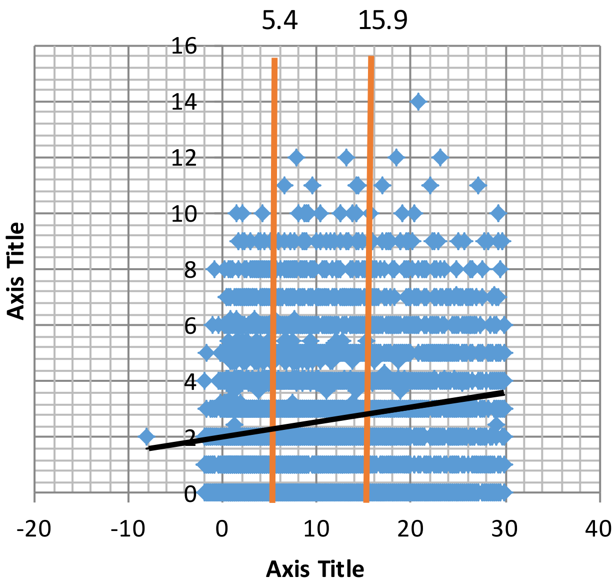
Given the near-linear relationship evident in the figure, equation (14) tracks the actual data quite well even as Y approaches its minimum value (normalized to zero).

For the Malthusian era, we regressed the number of children on income, excluding the outliers for income.[24] Since the maximum income (excluding outliers) was £30, we rescaled income to the interval [0, 1] and multiplied the coefficient by 30 to give the same marginal effect.

$$N = 1.9938 + 1.5503.Y \qquad 0 < Y < 1. \tag{15}$$

Figure 4: A Malthusian Equation (5)

(Vertical axis, surviving children. Horizontal axis, £income)



Interestingly, in both eras, the intercept of the equation is 2, with the effects of normalized income operating more powerfully in the Mathusian era. The R^2 of the Malthusian era is low, and it might be expected that if unit record data were available for Figure 2, it would also display variation in children unrelated to income. We already flagged in the previous section that the one positive development for inequality in the Modern era is the likely fall in this variance.

For simplicity and clarity, our simulations of inequality are an extension of (3) in which $\phi n\ (Y/N)=n\ Y/N$ so that $\phi=1$. Furthermore, we do not add independent noise u to the determination of the numbers of children. The first of these two modelling choices biases downwards the level of inequality, and its transmission across generations, but we want to isolate the key difference between inequality-generation across the three eras, namely the evolution of β_1. As we shall see the transmission is nevertheless substantial.[25]

In the analysis up to this point, we have used variance of family income as a measure of inequality, but this measure depends on the units of income chosen. For this reason, our simulations use the more

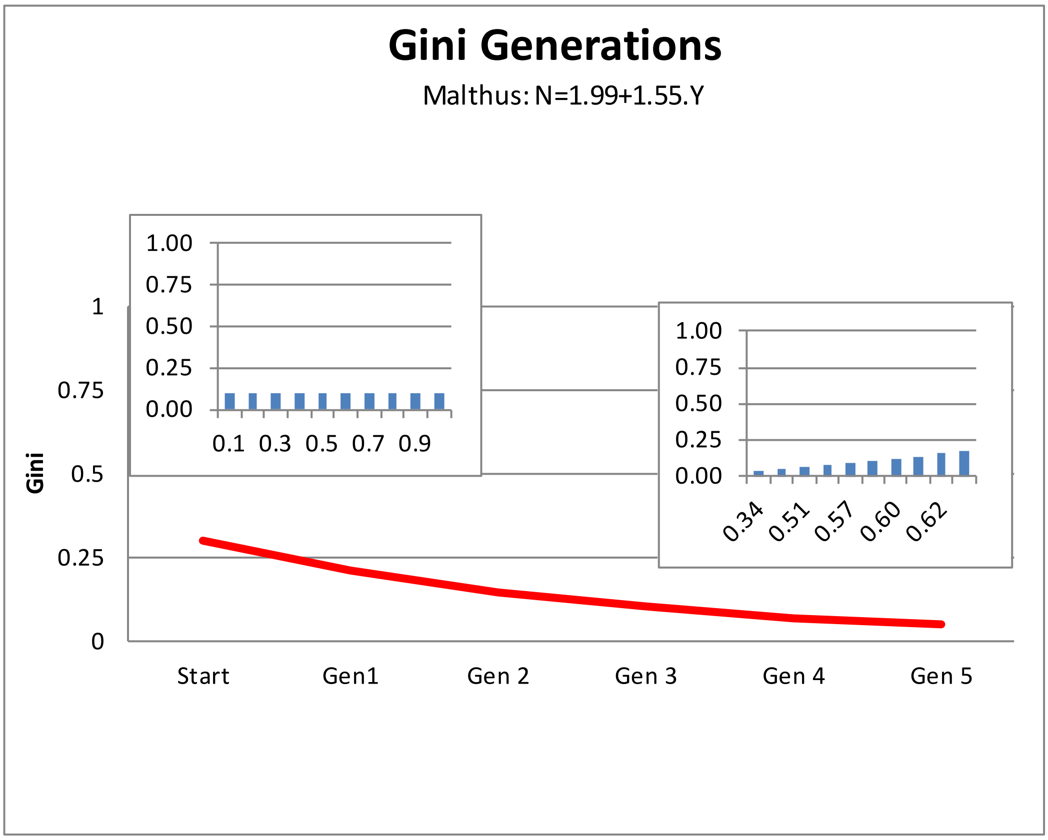
[24] We ran original regressions without excluding outliers, and there was no significant relationship. However, an examination of the scatterplot showed some very high income levels with high leverage. We then excluded the top 15% of incomes from the analysis. In the Mathusian era, these were incomes over £30.

[25] Another more pragmatic reason for not adding noise is that we cannot infer it from Figure 2.

standard measure of inequality, namely the Gini coefficient.[26] We start with an economy with income uniformly distributed over the interval (0, 1). Each generation, a cohort of children grow up with earnings given by (3) and the pre-multiplication by n ensures that the average income of each generation is roughly constant. We do this so that the equations we have estimated (14) and (15) will remain roughly valid, although we do not force income to stick strictly below unity. We then calculate our Gini coefficients for each generation based on the distribution of income (see Appendix 2). The nature of the Gini measure is such that it does not matter if every earner has income scaled up each generation, so we are obtaining a set of estimates for the ITD process, whilst still being able to use (14) and (15) each generation.

We proceed in order of history. The first simulation is for the Malthusian era. We use equation (15) over five generations. The results are shown in Figure 5. Interestingly, the locus of Ginis over five generations appears to be heading towards a low steady state. Reassuringly, the parameter Λ, the rate of decline of the variance in the second line of (10) with $\phi=1$ and $\sigma^2=0$, is about one half and the Gini's in Figure 5 do indeed halve each generation. Theoretically, with no independent noise in child numbers, (11) suggests that this steady state should be zero, a state of perfect equality. We don't mean to suggest that this tendency would dominate all other factors in a real-world empirical setting, but the tendency *toward* convergence is obvious and powerful.

Figure 5: Transmission of Inequality: Malthusian Era



The inset boxes in Figure 5 show the distribution of income at the start (where it is uniform over (0, 1) by assumption) and after five generations. Here, the range of income shrinks substantially, over an interval of (0.28, 0.64) in generation five and the distribution is skewed towards higher wealth. Any families with low wealth have few surviving children (due to the positive covariance between Y and N),

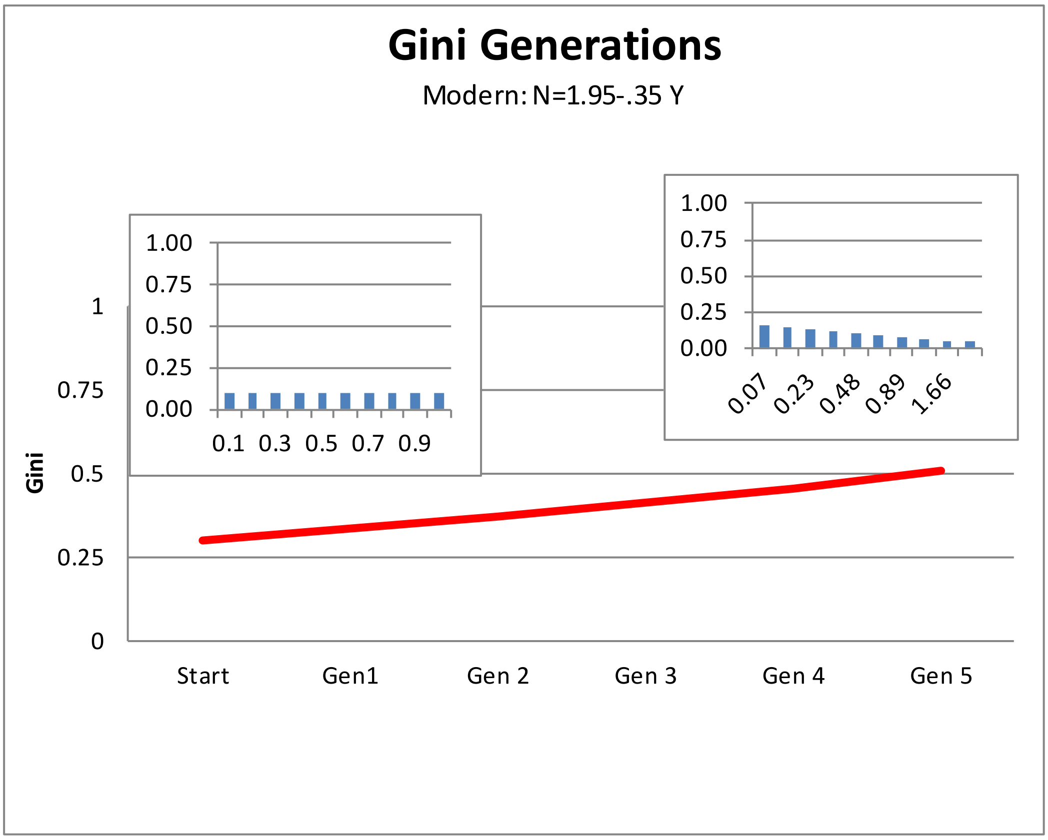
[26] The relevant spreadsheet for this section, gini.xlsx, is in our online appendix. In it, we also calculate the variance corresponding to the Gini coefficients, and find that they tell the same story.

and the relatively high investment in the children improves their relative earning power in the next generation. The Gini calculations go from 0.3 to around 0.05, which is a substantial decline. Using the contemporary Ginis in Corak (2012), this corresponds to moving from a country like New Zealand, Sweden or Australia--all around 0.3--to a mythical country whose degree of inequality would be well below that of the least unequal country (Denmark, with a Gini of around 0.1).

There is no need to draw a graph of the static Industrial Revolution era.[27] With the assumption that no other factors affected inequality and no covariance between Y and N, there is no movement in the Ginis from 0.3, and no change in the income distribution. As we noted earlier, this is a conservative picture. If there were independent noise affecting the numbers of children this would affect inequality. However, with a zero β_1 the effects would not be ITD inequality as described in this section. We are not, after all, trying to *predict* inequality, but rather to *simulate* structural dynamic change in inequality arising from changes in β_1 under the assumptions of our model.

The simulation for the modern era in Figure 6 shows a marked increase in inequality across five generations.

Figure 6: Transmission of Inequality: Modern Era



As before, the inset boxes show the uniform distribution of income at the start, and after five generations. Here, the range of income rises substantially, over an interval of (0.07, 1.78) in generation five and the distribution is skewed towards lower wealth. Any families with low wealth have many children (due to the negative covariance between Y and N), but no less importantly the high income families have few children who have yet higher income in subsequent generations. The Gini calculations go from 0.3 to around 0.5. Again, using the contemporary Ginis in Corak (2012), this

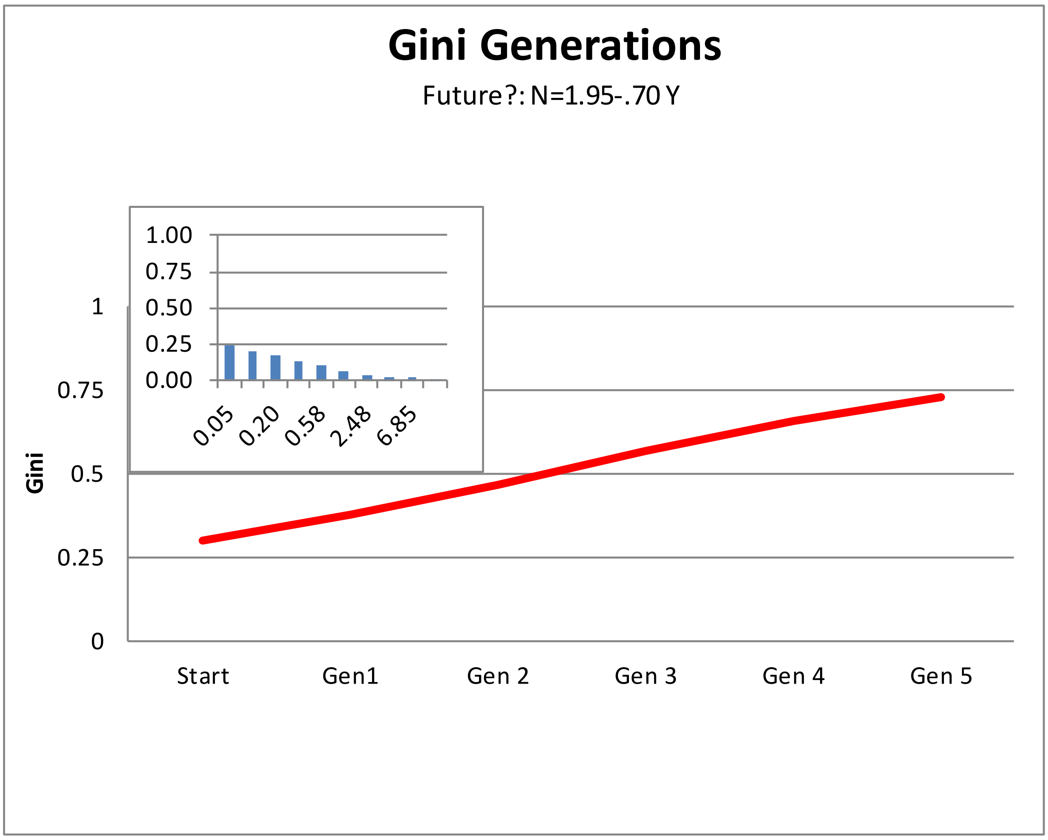
[27] In any case, the Industrial Revolution did not last 5 generations.

corresponds to moving from New Zealand, Sweden or Australia to countries like the US, the UK, Argentina or Pakistan.

The results of our simulations thus far confirm the circumspect stance we adopted at the start of this paper. We wholeheartedly affirm that other factors might prove to be more important for explaining inequality in a particular time and place. Nevertheless, unlike earlier eras, the ITD process now constitutes a 'drift' tendency in our economic system which, if left unchecked, will have a substantial impact over time. And, as we noted earlier, our lack of independent noise for N_t creates a downward bias.[28]

However, five generations is also a long time to assume that the current covariance will last. If we may be permitted a sensitivity analysis that gives pause, suppose that the future negative covariance between income and children strengthens due to, say, an increased prevalence of assortative mating (OECD (2011)). To be extreme, let us suppose the slope coefficient in (5) doubles. Then, the rise in the Gini is much stronger over the generations, as shown in Figure 7.

Figure 7: Transmission of Inequality: A Pessimistic Future



In this case all the qualitative features are as in Fig 6, but the Gini coefficients go from 0.3 to around 0.75. The latter corresponds to some of the most unequal nations on earth, such as Peru and South Africa. Or, to put the result another way, under this scenario it takes just two generations to effect a change in the Gini coefficient that took five generations under the scenario in Figure 5.

[28] There is, however, a complicating factor in our assessment of bias. If n is low, as it has been in the modern era, there will be a number of families with no children when a noise term is added. The treatment of the bequests of the parents in this case will be important, and if their assets are distributed across all income groups via taxation, the noise term will be a force *for* equality. .

5. Discussion and Conclusion

In this paper we have set out a simple mechanism governing the intergenerational transfer of income dispersion, based on the simplest of observations about families. We began, using equation (3), by asserting that the level of intra-family transfers to children is a ratio of the parents' ability to give, which is positively related to parental income, over the number of children in their care. We went on to spell out how this ratio of variables will determine the level of inequality in the subsequent generation.

The mechanism we describe does not depend on any genetic factors or claims about superiority of family environment as such. There is no suggestion that what is inherited from parents is some form of *intrinsically* superior income-earning capacity. That is, we are here making the Smithian assumption that genetic quality is uniformly distributed across the income distribution. This may not be true, of course, but this is the strongest assumption one can make in favour of disposing the process toward equality. If inequality results in this case, it will be even more striking if genetic endowments are sharply divergent.

Instead, inequality in our model is driven by the size of transfers of material goods from parents. And indeed, since all the offspring of any couple share equally and totally in the genetic composition of their parents and the family environment, the critical aspect of automatic "sibling rivalry" associated with the division of family resources would be absent in any process that made appeal to genetic or family environment factors. Nor do we include "assortative mating" factors. Clearly assortative mating would magnify the effects that we discuss: and equally obviously there is considerable evidence that assortative mating occurs (OECD (2011)). But it is not necessary for the ITD process and would for our purposes simply add a gratuitous level of complication.

As we have emphasized the mechanism we describe is simple and, once seen, obvious. And it is only a part of the processes that determine the income distribution at any point in time. But it is a systematic and potentially powerful aspect of that determination process – and it deserves independent exposition and analysis. Predictably, ITD has not totally escaped attention. Many writers on the connection between economics and demography mention the possibility in passing or hint at its operation. But as far as we have been able to detect none has focused on it specifically or drawn attention to its implications for future equality[29].

Moreover, recent discussions[30] of the relation between income dispersion and social mobility make no mention of ITD at all. But as we demonstrate, the 'Iron Law' nature of the mechanism makes ITD absolutely central, and its operation ought to be a focus of current discussions of inequality and of the ongoing policy response. Policies that focus on current transfer programs and/or the need for more extensive public transfer in the light of increasing inequality seem to us to focus on symptoms rather than causes. The challenge, if the ITD process is as relentless as we take it to be, is demographic. And the policy object ought to be at least as much to *increase fertility among the rich* as to reduce it among the poor.

For any commentator concerned with inequality, the long-term future looks bleak in almost all countries. Even those who find current levels of inequality of income roughly acceptable may be deeply troubled by the changes the future appears almost certain to hold. There may well be levels of income inequality that a liberal democratic polity cannot tolerate. The fact that we are on a trajectory that will eventually test such limits is more than a little disturbing.

References

Alexander, K. and A. Cherlin (1990), 'Are Few Better than Many?' review of Family Size and Achievement by Judith Blake, *Contemporary Sociology*, **19**, 3, May, 343-346.

Becker, G. (1981), *A Treatise on the Family*, Chicago Univeristy Press.

[29] Clark and his coauthors seem to us to have come closest to examining its implications for the past.
[30] We have in mind Kreuger (2012) in particular.

Becker, G. and N. Tomes, (1976), ‘Child Endowments and the Quantity and Quality of Children’, *Journal of Political Economy*, 84, 4, Part 2: Essays in Labor Economics in honour of H. Gregg Lewis (August), 143-162

Becker, G. and N. Tomes, (1979), ‘An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility’, *Journal of Political Economy*, 87, 6, 1153-1189.

Blake, J. (1989), *Family Size and Achievement*, Berkeley: University of California Press.

Clark, G. and N. Cummins (2010), Malthus to Modernity: England's First Fertility Transition, 1760-1800, *MPRA Paper* No. 25465, posted 28. September 2010.

Corak, M. (2012), Inequality from Generation to Generation: the United States in Comparison, *University of Ottawa mimeo*.

Dang, H. (2013), ‘The Decision to Invest in Child Quality over Quantity: Household Size and Household Investment in Education in Vietnam’, World Bank e-Library http://econ.worldbank.org/external/default/main?pagePK=64165259&piPK=64165421&theSitePK=469382&menuPK=64216926&entityID=000158349_20131023130441 Policy Research Working Paper.

Goethe, J. W., (1783), published by F. H. Jacobi in *Über die Lehren des Spinoza* in 1785, and in 1789 by Goethe in *Goethes Schriften*.

ICF International (2012), *MEASURE DHS STATcompiler*, at http://www.statcompiler.com , accessed August 30 2013.

Jones, L., Schoonbroodt, A. and M. Tertilt (2008), Fertility Theories: Can they Explain the Negative Fertility-Income Relationship?, *NBER working paper* 14266, August 2008.

Jones, L. and M. Tertilt (2008), An Economic History of Fertility in the U.S.: 1826-1960, in *Frontiers of Family Economics*, ed. by P. Rupert, vol. 1. Emerald Press, forthcoming.

Krueger, A. (2012),The Rise and Consequences of Inequality in the United States, *Center for American Progress*, Speech Jan 12 2012. See http://www.americanprogress.org/wp-content/uploads/events/2012/01/pdf/krueger.pdf

Malthus T.R. (1798), An Essay on the Principle of Population, 1st edition published anonymously, (St. Paul's Churchyard, London: J. Johnson, 1798).

Marx, K. (1875), Critique of the Gotha Programme, *footnote 1 in Marx/Engels Selected Works, 3, 13-30*, first published Abridged in *Die Neue Zeit*, (1890-91), Bd. 1, No. 18,

OECD (2011), An Overview of Growing Income Inequalities in the OECD Countries: Main Findings, in *Divided we Stand: Why Inequality Keeps Rising*, www.oecd.org/els/social/inequality

Pareto, V. (1935), *The Mind and Society*, Vol. 3, New York: Harcourt Brace.

Smith, A. (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, 1st ed., W. Strahan, London.

Appendix 1: Scaling Distributions by Family Size

We begin with the observation of Blake (1989) that in a random sample of people, those from large families will be over-represented. For example, suppose that the pdf of family size is uniform over the integer support N=0 to 4. In generation 2 the distribution of adults classified by family-of-origin size will be 0.0, 0.1, 0.2, 0.3, 0.4 corresponding to N=0 to 4. (This is obtained by multiplying each N by 1/5 and then dividing by the mean of the original pdf, namely 2 children, to ensure the probabilities sum to unity.)

The question is to what extent that distorts the relationship between $V(Y_1)$ worked out with a generation *1* distribution of income and $V(n\ Y_1 / N_1)$ worked out with a generation *2* distribution, over-represented by children from large families. It turns out that the multiplication of Y_1 by n/N_1 exactly compensates for the scaling of the distribution.

Notation:

- f probability density function (pdf)
- $f_{x,\,2}$ pdf of the random variable 'x' in generation 2 (f_x is the pdf in generaton 1)
- Y Income (mean=n)
- N children (mean=y)
- u i.i.d. error with zero mean
- $N=\beta_0+\beta_1 Y+u$ The number of children; Y and u are independent.
- E_i the expectation based on the pdf in generation i ($i=1, 2$)
- V_i the variance based on the pdf in generation i ($i=1,2$)

We seek the mean and variance of $n\ Y_1 / N_1$ using the second generation pdf. We first must establish the joint pdf. Noting that Y and u in generation 1 are independent we can use the conditional probability formula.

$$f_{N|Yu} = \frac{f_{NYu}}{f_Y\ f_u} \quad \Rightarrow \qquad f_{NYu} = f_{N|Yu}\ f_Y\ f_u \qquad (A1.1)$$

We intuit that $f_{N|Yu,\,2}$ relates to the generation 1 density as follows:

$$f_{N|Yu,\,2} = \frac{N\ f_{N|Yu}}{n}. \qquad (A1.2)$$

That is, whatever value of N is drawn for a family in generation 1 the pdf will be scaled up by that value. Then, division by n is required so that the integral of the pdf is unity. This was the intuition above for Blake's observation that you are more likely to meet someone from a large family than a small one. The joint pdf follows when *(A1.2)* is substituted into *(A1.1)*.

$$f_{NYu} = \frac{N\ f_{N|Yu}}{n}\ f_Y\ \ f_u$$

To simplify the algebra, we ignore ϕ in the text by setting it to unity. We are interested in $V_2(nY/N)$ – the variance of children's incomes calculated using the second generation distribution.

$$V_2\left(\frac{n\ Y}{N}\right) = E_2\left(\frac{n\ Y}{N}\right)^2 - \left(E_2\left(\frac{n\ Y}{N}\right)\right)^2 \qquad (A1.3)$$

The expectation of *n Y / N* using the second period distribution is just the original expectation of *Y*.

$$E_2\left(\frac{nY}{N}\right) = \int_N\int_Y\int_u \frac{n\,Y}{N}\frac{N\,f_{N|Yu}}{n}\; f_Y\; f_e\; du\,dY\,dN$$

$$= \int_N\int_Y\int_u Y\, f_{N|Yu}\; f_Y\; f_e\; du\,dY\,dN = \int_N\int_Y\int_u Y\, f_{NYu}\; du\,dY\,dN = y \qquad (A1.4)$$

Equation *(A1.4)* shows how the neatly the multiplication by *n/N* addresses the problem of the scaled distribution. When evaluating $E_2\left(\frac{nY}{N}\right)^2$ in (A1.3) we first place it into the expectations integral, to find how it relates to an expectation based on a generation-1 distribution.

$$E_2\left(\frac{n\,Y}{N}\right)^2 = \int_N\int_Y\int_u \frac{n^2\;Y^2}{N^2}\frac{N\,f_{N|Yu}}{n}\; f_Y\; f_u\; du\,dY\,dN$$

$$= \int_N\int_Y\int_u \frac{n\;\;Y^2}{N} f_{N|Yu}\; f_Y\; f_u\; du\,dY\,dN = E_1\left(\frac{nY^2}{N}\right) \qquad (A1.5)$$

To evaluate the latter, we first do a Taylor series expansion of the term in the expectations.

$$\frac{nY^2}{N} \approx y^2 + 2y(Y-y) - \frac{y^2}{n}(N-n)$$

$$+\frac{1}{2!}\left\{2(Y-y)^2 + \frac{2y^2}{n^2}(N-n)^2 - \frac{4y}{n}(Y-y)(N-n)\right\} \qquad (A1.6)$$

Upon taking generation-1 expectations, the second and third terms fall out.

$$E_1\left(\frac{nY^2}{N}\right) = y^2 + \frac{1}{2!}\left\{2E_1(Y-y)^2 + \frac{2y^2}{n^2}E_1(N-n)^2 - \frac{4y}{n}E_1(Y-y)(N-n)\right\} \quad (A1.7)$$

When (A1.4) and (A1.7) are substituted into (A1.3) we obtain the equivalent expression to equation (4) in the text.

$$V_2\left(\frac{n\,Y}{N}\right) = E_2\left(\frac{n\,Y}{N}\right)^2 - \left(E_2\left(\frac{n\,Y}{N}\right)\right)^2$$

$$= E_1\left(\frac{nY^2}{N}\right) - y^2$$

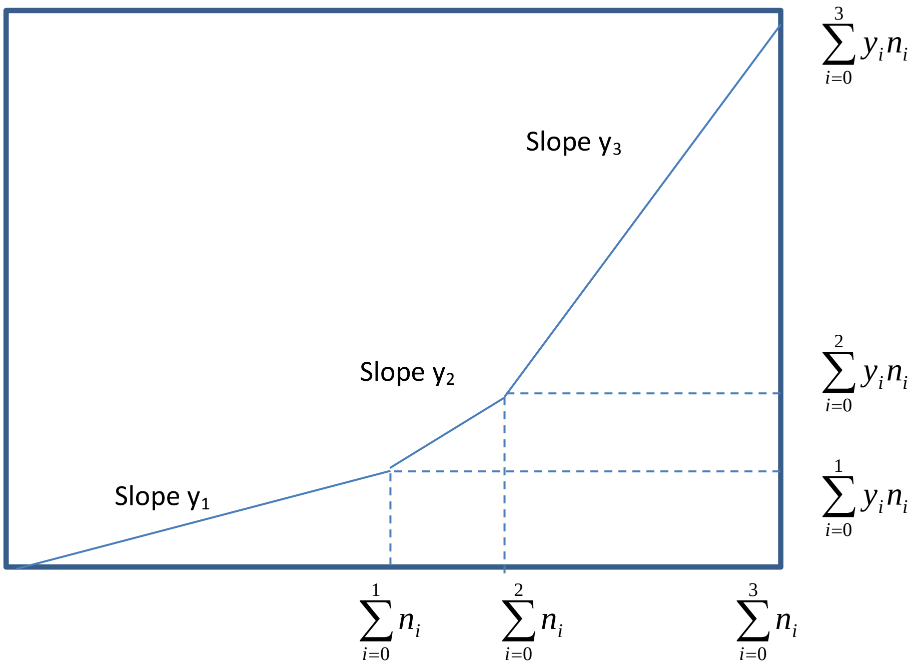
$$= y^2 + \frac{1}{2!}\left\{2E_1(Y-y)^2 + \frac{2y^2}{n^2}E_1(N-n)^2 - \frac{4y}{n}E_1(Y-y)(N-n)\right\} - y^2$$

$$= V_1(Y) + \left(\frac{y}{n}\right)^2 V_1(N) - 2\left(\frac{y}{n}\right)C_1(Y,N)$$

This confirms that equation (4) can be used as the basis for a difference equation, linking inequality across generations. Naturally, it is only as good as the approximation (A1.6) on which it is based, but our numerical simulations give us some confidence in this regard.

Appendix 2: Calculating the Gini Coefficient

Let y_i be individual income and n_i be population per cohort i. Every person in the cohort has the same income. The unscaled Gini box consists of segments each with slope y_i for the ith segment. That is, each extra person adds their y_i to cumulative income. It will be algebraically useful to run i from 0, with $n_0=0$



The horizontal axis has persons, and the vertical axis has cumulative income.

The number of segments runs i=1 to k. In the above k=3. The slopes increase because the Gini methodology ranks income from least (slope) to greatest (slope).

The scaled Gini box shows divides the horizontal axis by total persons, and the vertical axis by total income. We have:

sth cohort runs along the horizontal axis from: $\dfrac{\sum_{i=0}^{s-1} n_i}{\sum_{i=0}^{k} n_i}$ to $\dfrac{\sum_{i=0}^{s} n_i}{\sum_{i=0}^{k} n_i}$

with vertical coordinates: $\dfrac{\sum_{i=0}^{s-1} y_i n_i}{\sum_{i=0}^{k} y_i n_i}$ and $\dfrac{\sum_{i=0}^{s} y_i n_i}{\sum_{i=0}^{k} y_i n_i}$.

The slope of the sth cohort is $\dfrac{\dfrac{y_s n_s}{\sum_{i=0}^{k} y_i n_i}}{\dfrac{n_s}{\sum_{i=0}^{k} n_i}} = \dfrac{y_s \sum_{i=0}^{k} n_i}{\sum_{i=0}^{k} y_i n_i}$.

Using this slope, and the points at s and s-1 we can construct the segments of the (scaled) box above and integrate to obtain the area L under the (Lorenz) curve. The horizontal axis is deemed to be x, and all unmarked summation is from i=0 to i=k.

$$L = \sum_{s=1}^{k} \left\langle \int_{\frac{\sum_{i=0}^{s-1} n_i}{\sum n_i}}^{\frac{\sum_{i=0}^{s} n_i}{\sum n_i}} \left(\left\{ \frac{y_s \sum n_i}{\sum y_i n_i} \left[x - \frac{\sum_{i=0}^{s-1} n_i}{\sum n_i} \right] \right\} + \frac{\sum_{i=0}^{s-1} y_i n_i}{\sum y_i n_i} \right) dx \right\rangle$$

The Gini Coefficient is (0.5-L)/0.5, or, 1-2.L.

Upon integration, and defining p_i be the proportion of agents in each cohort $p_i = n_i / \sum n_i$ we obtain, after a lot of algebra:

$$Gini = 1 - \sum_{s=1}^{k} (p_s \left[\frac{y_s p_s + 2.\sum_{i=0}^{s-1} y_i p_i}{\sum y_i p_i} \right])$$

As a cross-check, note that if all the y's were identical, then the Gini would become

$$Gini = 1 - \sum_{s=1}^{k} (p_s \left[\frac{p_s + 2.\sum_{i=0}^{s-1} p_i}{\sum p_i} \right]) = 1 - \frac{\left(\sum p_i\right)^2}{\sum p_i} = 0 \quad because \qquad \sum p_i = 1$$