Randomizing Development: Method or Madness?

Lant Pritchett

Oxford Blavatnik School of Government

June 30, 2019

Abstract. An important argument for the increased use of randomized control trial methods in development is that the evidence from these studies will encourage the uptake of effective programs and projects (both through discouraging ineffective projects and improving design of new projects) and this will lead to reduced poverty and improved human well-being. However, cross-national evidence shows that the four-fold transformation of national development, to higher productivity economies, to more responsive states, the more capable organizations and administration and to more equal social treatment produces gains in poverty and human well-being that are *orders of magnitude* bigger than the best that can be hoped from better programs. Arguments that RCT research is a good (much less "best") investment depend on *both* believing in an implausibly low likelihood that non-RCT research can improve progress national development *and* believing in an implausibly large likelihood that RCT evidence improves outcomes.

Randomizing Development: Method or Madness?

Bill Gates has recently been promoting chicken ownership to address poverty in Africa. In an open letter, Professor Blattman of University of Chicago pointed out that cash transfers may be more cost effective than chickens said: "*It would be straightforward to run a study with a few thousand people in six countries, and eight or 12 variations, to understand which combination works best, where, and with whom. **To me that answer is the best investment we could make to fight world poverty***. *The scholars at Innovations for Poverty Action who ran the livestock trial in Science agree with me. In fact, we've been trying, together, to get just such a comparative study started.*"[1][emphasis added]

I think it is important for the development community to stop and reflect on how we, as a development community, arrived at this two-fold madness. First the madness that Bill Gates, a genius, a humanitarian, an important public intellectual, could be even semi-seriously talking about chickens. Second, the madness about method, that the response of Chris Blattman, also a genius, an academic at a top global university, and also an important public intellectual would respond not "Chickens? Really?" but rather that the "best investment" to "fight world poverty" is using the *right method* to study the competing program and design elements of chickens versus cash transfers[2].

That this *is* madness is, I hope, is obvious. The top 20 most populous developing countries in the world are (in order): China, India, Indonesia, Brazil, Pakistan, Nigeria, Bangladesh, Russia, Mexico, Philippines, Ethiopia, Vietnam, Egypt, Iran, Turkey, DR Congo, Thailand, South Africa, Tanzania and Colombia. Together these countries have 4.6 billion people. Imagine gathering a couple of dozen of the leaders from any one of these countries (where "leadership" could be political, social, economic, intellectual, popular, mass movement, civil society, or any combination) and saying: "We, the experts in the development community, think 'fighting world poverty' is the center of the development agenda and we think that the 'best investment' we can make to promote development/fight poverty in *your country* [fill in the blank: Indonesia, Brazil, Nigeria, DRC, Tanzania, South Africa, Egypt, India] is a set of studies using the right method to resolve the questions of whether anti-poverty programs should promote chicken ownership or distribute cash and, within that, how best to design such chicken or cash transfer programs?"

I imagine two responses from country leaders. One, how could you have come to such trivial and trivializing ideas about our country's goals, aspirations, and challenge? How can we as [Indonesians/Indians/Nigerians/Egyptians/Tanzanians] not take as outright contempt the suggestion that either "chickens" or "studies about chickens" are the top

---

[1] https://www.cgdev.org/blog/getting-kinky-chickens

[2] With dozens on studies on conditional cash transfers, micro-finance, and a sobriquet 'Worm Wars' to describe a massive debate on whether deworming is cost-effective (and a bouquet of RCT studies of boutique anti-poverty and kinky goal interventions) this madness has seeped far more broadly.

priorities for our country? Two, we can easily list for you many pressing, urgent, if not crisis, development issues affecting the current and future well-being of the citizens of our country. These questions are important whether or not your preferred method for producing research papers can address them[3].

I am using "studies of chickens versus cash" not to single out Professor Blattman, but to stand in for the whole *randomista* movement in development. Development economists, rather than finding it hard to think of "anything else" (Lucas 1988) but the big picture issues around national development, are now so committed to a method they are thinking about "anything but" national development. There are now literally thousands of published RCTs, with dozens on studies on conditional cash transfers, on micro-finance, and literally hundreds of studies of boutique interventions in water, sanitation, education, health, business training, etc[4]. I argue this madness about a method in development academia is a symptom, not the disease. The big debate is about the relative importance of "national development" versus "kinky development" and whether "national development" can be accelerated. RCT as a method can only even pretend to any importance if either (a) one interprets the development in a narrow way as achieving specific, low-bar, targets ("kinky" development) or (b) one takes the view that "national development" is completely beyond the influence of ideas or evidence.

National development is a four-fold transformation of an *intrinsically social* grouping (country or region or society) to higher levels of capabilities in four dimensions: an economic transformation from lower productivity to higher productivity; a political transformation to governments more responsive to the broad wishes of the population, an administrative transformation to organizations (including those of the state) with higher levels of functional capability for implementation, and a social transformation to more equal treatment of the citizens of the country (usually with a sense of common identity and, to some extent, shared

---

[3] Four (of many possible) anecdotes to back this assertion up. First, a colleague of mine was in the front office of the prime minister of a large and important country. At the request of prominent randomistas who had done considerable work in that country he managed to set aside two hours for a meeting between these academics and the prime minister. At the end of the meeting the prime minister pulled my friend aside and said: "Never, ever, waste my time like that again." Second, my colleague Arvind Subramanian was a top policy adviser in India, a country that has been a focus of randomistas activity, for three years. In a speech to my students in 2018 he said that *never* in his three years of being involved at many levels (from mid-level to the highest) in discussing the range of economic challenges facing India did he hear the results of any RCT play any role. Third, in my work as a development practitioner I have been in all but two of those top twenty population countries and have lived for years in two of them (Indonesia and India) and never, ever, outside of the narrow confines of development agencies and projects have I heard either chickens or rigorous studies mentioned as priorities. Fourth, when the "livestock trial in Science" study was being promoted in the media a reporter from a US based publication called to ask me my view of this important study. I responded that I had not read it as it wasn't a particularly interesting or important study from my viewpoint as a development scholar/practitioner. She asked me how, in light of the august authors and preeminent publication I could say such a thing. I responded that if she could find any mention of that study in the local press or media in *any* of the seven countries I would change my mind, read the study, and give her comments. Since of course the reporter never called back, I had a research assistant search for media mentions in any of the study countries (canvassing for people who spoke the local languages to help) and we could come up with not a single mention of the study.
[4] There is even a term 'Worm Wars' to describe a hotly contested debate on the questions of whether, when and where, deworming is a cost-effective intervention.

purpose). National development is about countries like Haiti or India or Bolivia or Indonesia achieving the high levels of economic, political, administrative, and social *functional* capabilities that Denmark or Japan or Australia possess. National development is not an end but a means of achieving higher level of human well-being.

"Kinky development" (Pritchett 2014, Kenny and Pritchett 2013) is the view that development is primarily, if not exclusively, about reaching very low-bar levels of specific indicators: "eradicating extreme poverty" or "universal primary school completion" or "access to safe water" are "kinky" goals in that they draw some completely arbitrary line or threshold in some dimension of human well-being and then pretend that "kinking" the distribution of well-being, pushing people to just that threshold, is the goal of development. The distinctive element of kinky development is that gains to human well-being above the low-bar threshold count for *nothing*.

Section I, empirically demonstrates two things.

One, median income/consumption, one of the four elements of national development, is both (a) empirically *necessary* and *sufficient* for reducing headcount consumption poverty and (b) (related) accounts for that *essentially all* of the cross national variation in poverty rates. The effect of anti-poverty programs (and *a fortiori* to the design of such programs and *a fortiori squared,* so to speak, studies about the design of anti-poverty programs) are just are tiny compared to the effects of inclusive growth.

Two, for omnibus measures of human well-being, such as the Social Progress Index, (a) high levels of national development are empirically *necessary* and *sufficient* for achieving high levels of human well-being and (b) this relationship is empirically tight for the Social Progress Index (and other omnibus human well-being measures). Moreover, all (less one) of the dozen of specific measures of human well-being that go into the Social Progress Index (e.g. access to water, personal security, health, education, etc.) are also tightly correlated with national development.

Section II presents a decision-tree framework to evaluate the claim that a specific intellectual activity (such as an RCT study) about targeted programs (like cash versus chickens) could be the "best investment" for "fighting poverty" (or, more generally, any measure of human well-being). I show *all* the links in the chain of reasoning that are needed to arrive at such a conclusion are false.

I)   National Development and Human Well-Being

I propose a rough and ready definition and empirical measures of "national development" and then show its empirical relationship to measures of human well-being, both kinky measures, like low-bar poverty, and broader measures.

II.A)   National Development as a four-fold transformation of countries

The very word "development" implies a change over time in which something becomes a better, more mature, more advanced version of its ontological type. A human develops from

zygote to mature adult, a frog from zygote to tadpole to frog. Rocks neither "develop" to become frogs nor, do rocks through erosion, "develop" to become sand. The first is impossible and the latter not directional. What is it that "develops" with "development"? With "national" development what "develops" is typically a country, but is always and intrinsically a *social* (and socially constructed) aggregate[5]. A country has (at least) four important dimensions along which it "develops" and each is *intrinsically* and *ontologically* social and cannot be meaningfully individuated.

Economic development. This is usually understood as the productive capability of a *place*. This has some elements of the characteristics of the individuals but also a general "total factor productivity"-like element which is place specific and not individuated. A country's labor productivity, as measured by GDP per worker, is one possible indicator of economic development, though there can be many others (e.g. Hidalgo and Hausmann (2009) measures of economic complexity), and GDP can be adjusted in many ways (e.g. green accounting). These measures are *never* intended as direct measures of human well-being but are measures of the economic product and productivity of a place.

Administrative development. This is typically conceived of as some aggregate of the capability of (mostly state) organizations to accomplish public purposes[6]. Countries have an array of organizations to carry achieve purposes: armies, central banks, post offices, police forces, courts, land registries, etc. While there is of course variation within countries in the capability of organizations (Kaufmann et al 2002), an aggregate of the administrative capability of the state is another element of national development. The Fragile Sates Index, as one example of such a measure, ranks countries from 0 (best) to 10 (worst, most fragile) on their "broad based provision of public services" and Denmark scores .9, Indonesia 5.6 and Haiti 9.4.

Political development. This is obviously hugely value laden and, like anything said about politics, is itself political, but *descriptively* when people described the "development" of states they usually had in mind some notion that those in political power and exercising sovereign power in a country: (a) are responsive to the needs, wishes, wants, desires of the citizens of the country and that political processes allowed those to be expressed by citizens and aggregated in fair and legitimate ways and (b) respected at least some set of "negative" rights that preserved liberty and security of the person (and perhaps in addition some "positive" rights) and (c) there is some degree of "rule of law." The Fragile States Index, for instance, has two distinct measures, one for "state legitimacy" (not "democracy") and one for

---

[5] While "nation" or "nation-state" are often used casually as synonyms for "country" this language brings in massive ideological baggage about what a "nation" is and is relationship to sovereign states as "countries." We can talk about the "development" of regions (e.g. Southern versus Northern Italy) or of provinces/states within a country (e.g. Tamil Nadu versus Uttar Pradesh).

[6] In our work *Building State Capability* we distinguish between the *capability* of organizations, which is a feature of an organization, and *capacity* as a feature of individuals and point out that capability or an organization is not the aggregation of the capacity of the individuals. This is to emphasize there are two distinct concepts, but we acknowledge one could just as well used the words exchanged (e.g. capacity as a feature of organizations) and, as long as one were consistent about distinguishing the two concepts, achieve the same goal.

"human rights and rule of law" (10 is worst, 0 is best). For State Legitimacy Haiti is 8.7, Indonesia 4.8 and Denmark .9 while for Human Rights and Rule of Law Haiti is 7.4, Indonesia is 7.3 and Denmark is 1.2. The Polity2 measure on a 10 is complete democracy, minus 10 is complete autocracy of the POLITY IV project has been 10 for Denmark since 1915 (with the interregnum of WWII), in Indonesia was -5 in 1998 (last year of Suharto's rule), jumped to 6 in 1999 and was 9 by 2017, in Haiti this was 0 2010 to 2015 and 5 in 2016 and 2017.
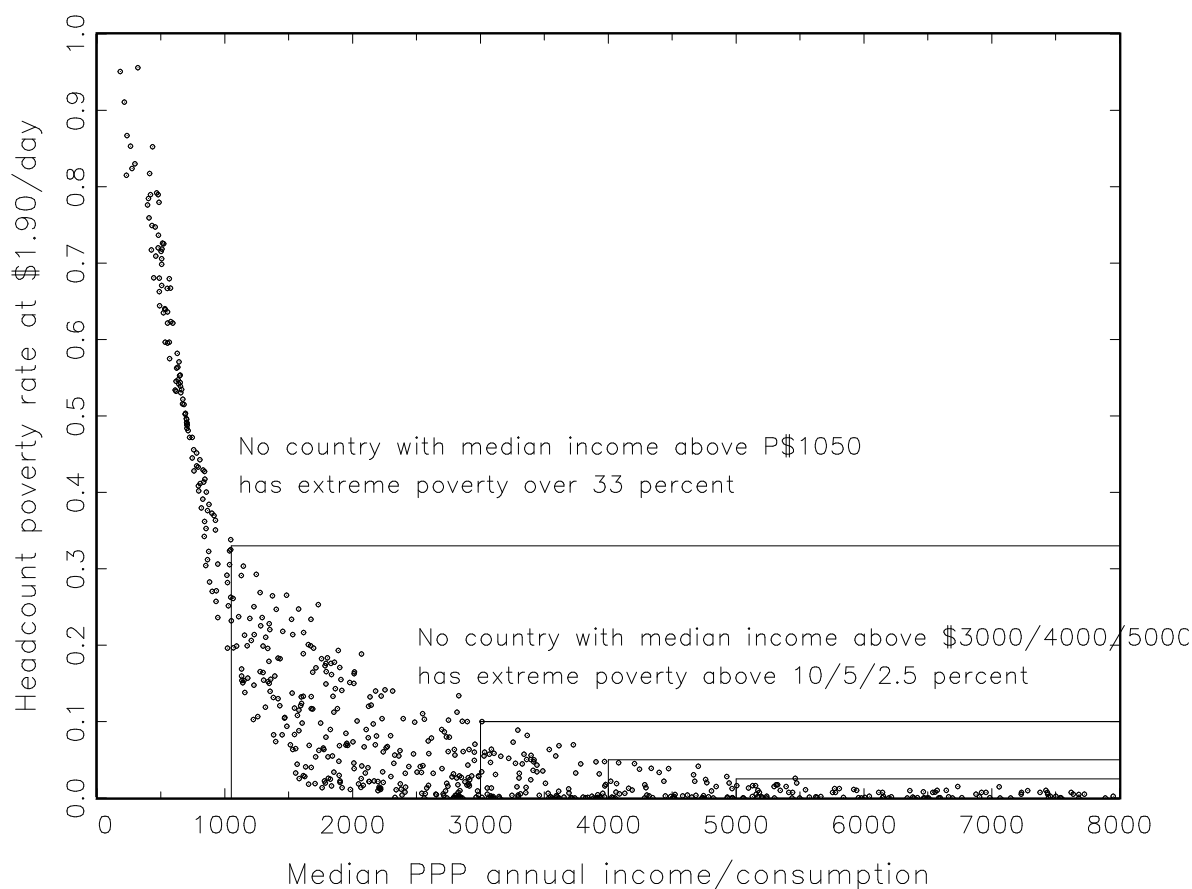
Social Development. Even more value laden and hence, if anything, more political than political development is the notion of how citizens/members of a common society treat each other changes as an intrinsic part of development. While these ideas were flawed in many ways (and in many ways reprehensible projections of social constructs of colonialists and colonialism) there was an important notion that "social equality"--in the sense that people were treated by other people equally independent of their social identities (kin, hereditary class, clan, tribe, ethnicity, race, sex, religion)--was, in and of itself, part of development. One part of the social development was the creation/adoption of a shared identity. These are obviously historically constructed values of the Western experience and do not have universal validity, but I would argue were often bundled into notions of "modernization" and "development" for good or ill. Today of course this is most obvious in the views that development needs to be gendered and that societies that do not treat the sexes fairly are considered less "socially developed" at least in one important sense, than those that do.

The units at which national development happens: a market, an organization, a polity, a society are about processes in which individuals participate and into which they are embedded but are *ontologically* not individuated.

II.B) Levels of median income/consumption completely explain poverty

National development, and in this case, just one measure of one element of national development, the levels of median consumption, is *sufficient* to (essentially) eliminate "low bar" or "dollar a day" (now, with inflation, P$1.90 a day where "P$" means purchasing power adjusted dollars) poverty.   The standard World Bank data, limited to all country/year pairs with actual survey data, one has over 800 country/year observations on measured poverty rates and on median income or consumption. Figure 1 shows that *no country* with median annual income above P$3,000 (about the level of Peru or Mongolia around 2010) has low-bar poverty more than 10 percent.   By P$5,000 (about the level of Costa Rica) essentially no country has low-bar poverty above 2 percent. Also, no country with median income above P$1,000 (about the level of Bangladesh in 2010) has low-bar poverty more than a third of their population. The whitespace in the "northeast" of Figure 1 is important as those are combinations of median income/poverty that *never* happen. There is a level of median income/consumption that is empirically *sufficient* to reduce poverty below any given percent of the population.

Figure 1: Median income/consumption is *sufficient* to eliminate extreme poverty

No country with median income above P$1050
has extreme poverty over 33 percent

No country with median income above $3000/4000/5000
has extreme poverty above 10/5/2.5 percent

Source: Author's calculations with data from [PovcalNet: the on-line tool for poverty measurement developed by the Development Research Group of the World Bank](#)

Figure 2 shows the levels of median income/consumption that are *empirically necessary* to reach various levels of $5.5 per day poverty rates[7]. By "empirically necessary" I am not asserting any logical necessity (like a theorem) but just that is doesn't happen. The whitespace in the "southwest" of Figure 2 are low median/low poverty are never seen. No country has pushed $5.5/day poverty below 75 percent of all households without median income above P$1045. That implies 42 of the 164 countries have a latest observed level of income such that *no country* has *ever* been observed with a poverty rate at P$5.5 less than 75 percent with their level of income. 107 of the 164 countries have a level of income such that (almost) no country has been observed with poverty below 10 percent at their level of income.   No country (but one[8]) has pushed P$5.5/day poverty below 10 percent without

---

[7] This is the highest level the World Bank source provides data but this is a "moderate" not a "high" poverty line. I, and many other people, argue for upper bar poverty definitions of P$10/day or above, which are still far below those actually used in richer countries.

[8] This country/year is Azerbaijan in 2005, whose data show median income of P$5655 in 1995 and poverty headcount 5.5$/day poverty of 5 percent and median income of P$5197 in 2015 and poverty of essentially zero but in 2005 a median income of P$2785 and poverty of 7.7 percent, which is the anomalously low observation, even for this country.

having median income/consumption above P$3535 (roughly the level of "upper middle income" countries like Peru (P$3486 in 2015), Kazakhstan (P$3557 in 2015) or Thailand (P$3549 in 2010).

Figure 2: High levels of median income/consumption are empirically necessary to eliminate poverty (and these levels are higher the higher the poverty line).
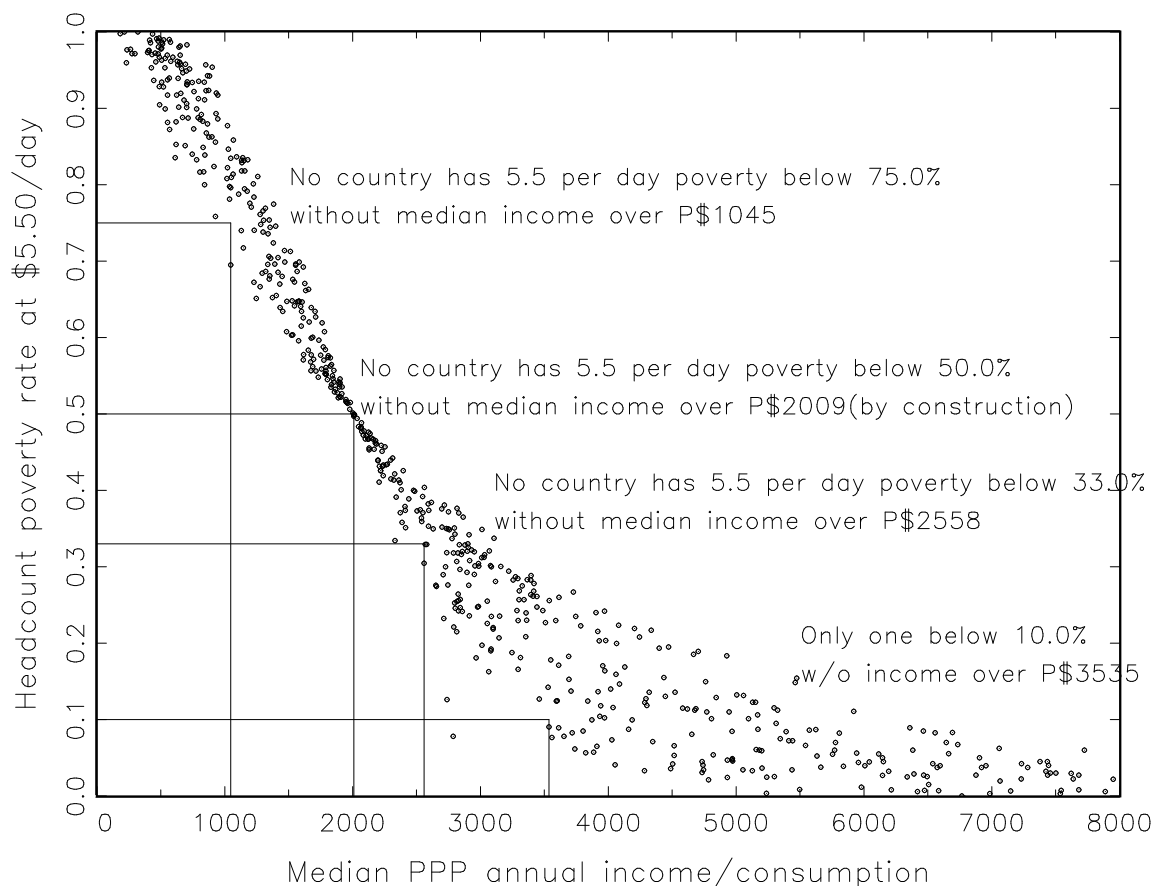
Source: Author's calculations with data from [PovcalNet: the on-line tool for poverty measurement developed by the Development Research Group of the World Bank](#)

So far I have been using 810 observations from the World Bank data whether the data was for income or consumption. But for exploring connections with programs or projects consumption expenditures are a better measure as they more reliably measure post-tax and transfer outcomes and hence reflect consumption expenditures inclusive of any benefits from programs.   Figure 3 shows the relationship between country level poverty rates at the three poverty lines in the World Bank data, P$1.9, P$3.2 and P$5.5, and the median of the distribution of consumption using just the 389 country/year observations using consumption data. Since the poverty rates *must be, by construction*, non-linear in the median, I fit a completely flexible functional form including all powers of the median from -2 to 5.

For all three measures the data say that *very nearly all* the observed variation (R2 of .983 to .988) across countries and time in poverty rates is associated with variation in the

median (50th percentile) of consumption. An R2 of .988 implies that the correlation of actual poverty rates and the poverty rate predicted from the median is .994 (=√.988).

Figure 3: Median income/consumption is of a country predicts the level of poverty *exactly* for high poverty lines and near exactly even for low poverty lines
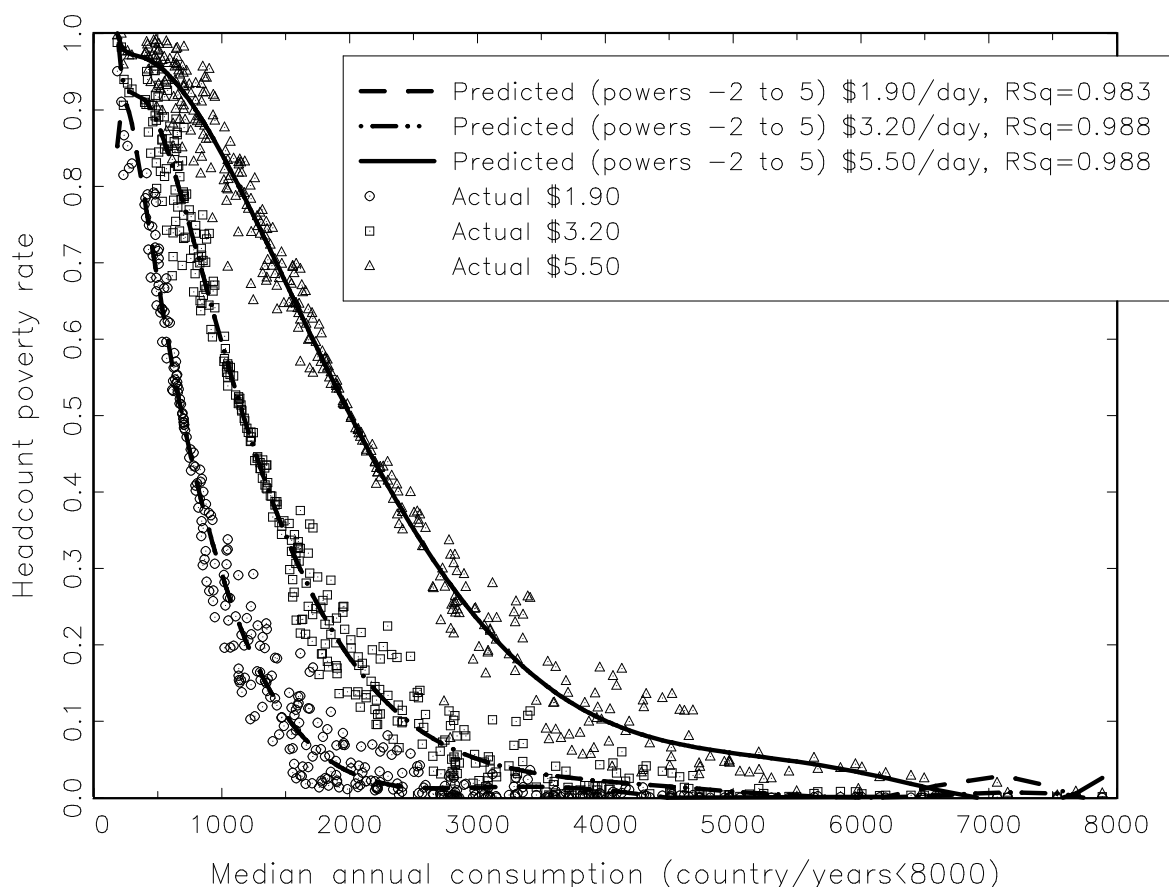
Source: Author's calculations with data from PovcalNet: the on-line tool for poverty measurement developed by the Development Research Group of the World Bank

This of course doesn't mean that other factors like the change in the inequality or the adoption of "poverty" programs *cannot* make a difference or even that they *cannot* in principal make a "substantial" difference, it just says that empirically, relative to the massive changes associated with the change in the median (from poverty of 100 percent to near zero percent), the differences at a given level of consumption are very modest compared to the gains from growth. Table 2 shows calculations of various poverty counter-factuals. For a country in the middle of the bottom quartile the poverty rate is 72.2 percent. If the country moved "due south"—had a lower poverty for the same median consumption—by one standard deviation of the residual the poverty rate would be 68.6 percent. In contrast if that country had the median consumption of having grown by 2 ppa faster over the previous 20 years (roughly a standard deviation of cross-national growth rates) its poverty would have been more than halved, to 35.9 percent. It would take a growth rate only .2 percent higher (e.g. 2.2 ppa vs 2 ppa)—which is only a tenth of a cross-national standard deviation--to

produce the same poverty reduction as improving poverty for a given median by a standard deviation.

Table 1: Even very small improvements in growth produce poverty reduction near the same as substantial (standard deviation of residual) improvements in poverty for a given level of median consumption.

Poverty rate	Quartile I of consumption, $1.90/day poverty line	Quartile II, $5.50/day
At average median consumption in the country quartile	72.2%	74.1%
If poverty is one standard deviation of the residual better for same consumption	68.6%	70.2%
If medium run growth (20 years) were 2.0 ppa higher (one cross-national standard deviation of growth rates)	35.9%	51.8%
If medium run growth (20 years) is better by .2 ppa (one tenth of a cross-national standard deviation of growth rates)	67.8%	72.2%
Source: Author's calculations with regressions shown in Figure 3 above.		

This super-tight correlation of measured poverty rates and median income/consumption also hold in changes over time within countries (Kraay (2006))[9]. Figure 4 shows an R2 of .93 between the change in "dollar a day" (P$1.90) poverty with the change in the predicted poverty based on just the shift in the median and the estimated functional form for the longest observed spell for each country (longer than 10 years).

---

[9] All of the empirical work here relies on the standard World Bank sources on household incomes/consumption, not on estimates of GDP per capita. Pinkovskiy and Sala-i-Martin (2016) argue, based on satellite data of light at night, that GDP per capita is a better, more reliable measure of progress and this shows faster progress and more poverty reduction.

Figure 4: Changes in poverty rates are also tightly associated with changes in median income/consumption
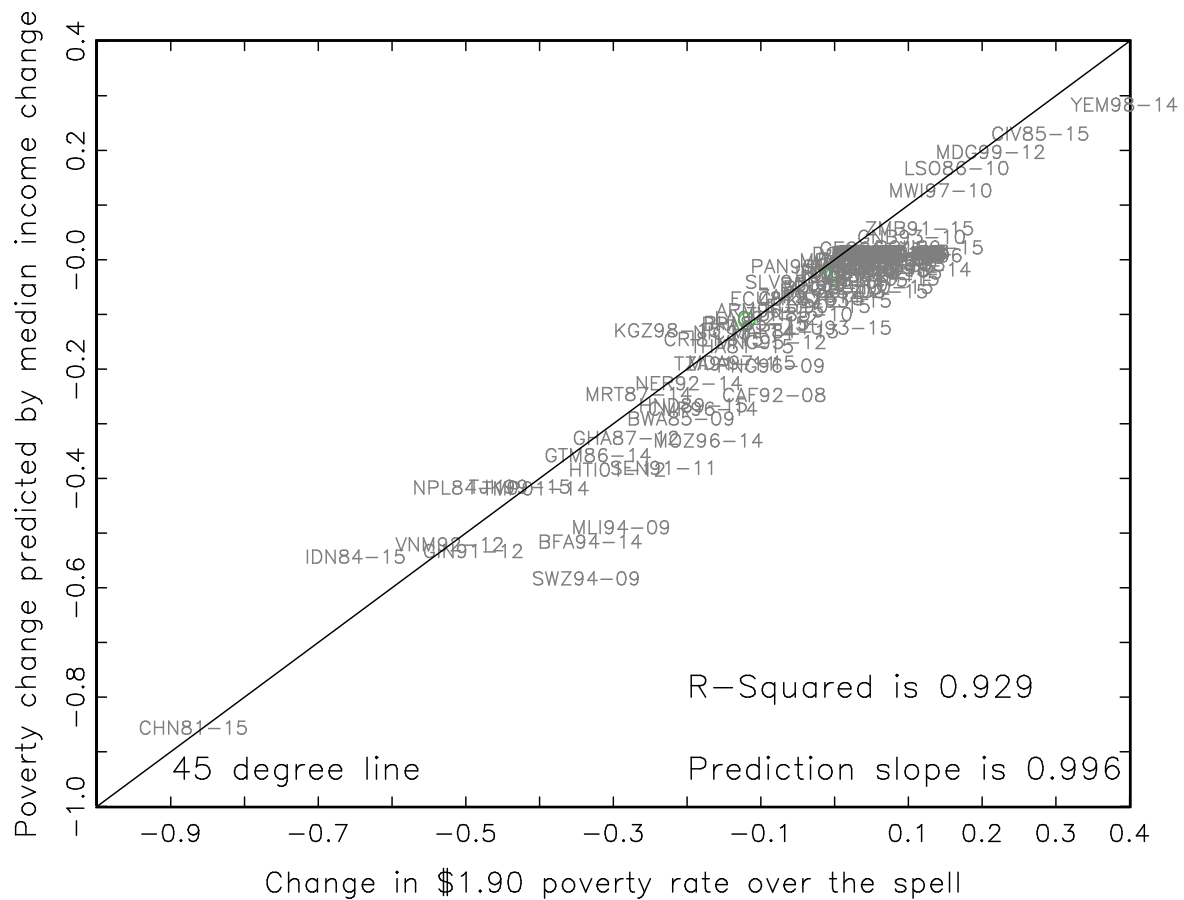
Source: Author's calculations with data from PovcalNet: the on-line tool for poverty measurement developed by the Development Research Group of the World Bank

Figure 5 shows some large countries (China, Indonesia, Vietnam) and to a lesser extent India, that have seen extreme poverty fall rapidly from very high levels to low levels. These poverty reductions happened "right in front of our eyes" and we have reasonably good household surveys tracking poverty over most (or all) of these periods so careful empirical work can be done to decompose the proximate determinants of this fall. How much of this fall in poverty was "accounted for" by changes in the central tendency (mean/median), how much was general change in inequality and how much was due to shifts in the distribution below the poverty line, conditional on mean and overall inequality of the type that "anti-poverty programs could in principle be responsible for). It is not too terrible a caricature of these results to say that "all" or "more than all" of the reduction in poverty in these countries was due to shifts in the mean/median (and "more than all" is that in many cases the inequality got worse (in the case of China much worse) and hence the increase in the central tendency had to offset that poverty worsening increase in inequality to reduce poverty).

Figure 5: In several countries the most rapid reductions in extreme poverty in history in several countries had been underway for 20 years by 2000
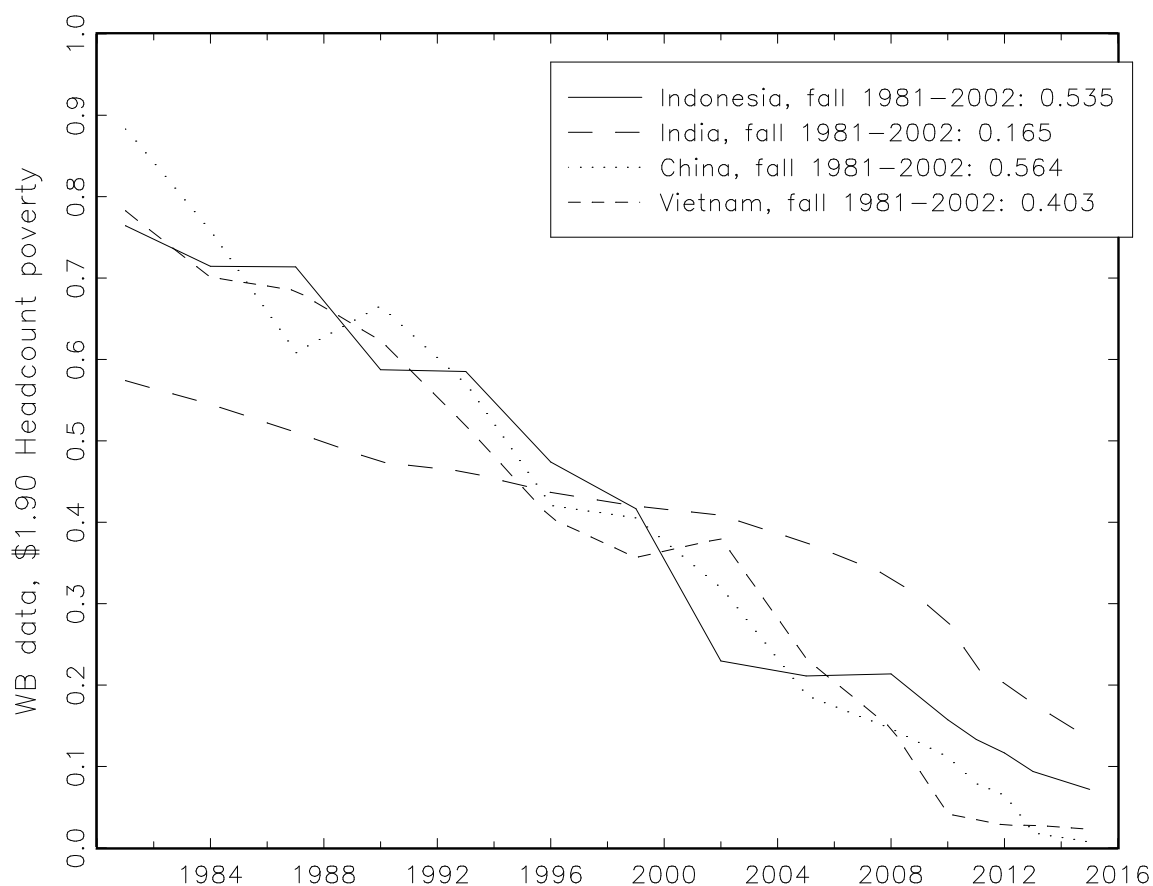
Source: Author's calculations with data from PovcalNet: the on-line tool for poverty measurement developed by the Development Research Group of the World Bank

By "poverty programs" many people seem to mean "interventions" that raise the consumption of "the poor" at a given level of the median. This is what would be expected from a cash transfer (conditional or not), a "graduation" type livestock program (referred to above), micro-finance, chickens, business training or pretty much any other *targeted* anti-poverty program. These are all intended to bringing up the "left tail" of the consumption distribution (benefitting "the poor") while holding its central tendency fixed (or possibly lowered, depending on how it is financed). The simple correlations say that differences across the country/years in the impact of "poverty programs" conditional on the median account for *at the very most* 1.2 percent of the total cross-national variation in poverty rates.[10] This is an upper bound as *everything* besides the median (measurement error, non-

---

[10] The standard poverty measures and medians are just different summary statistics of the *same* distribution. The standard headcount measure is just a partial integral of the distribution below a poverty line (I have published papers on methods for calculating poverty, e.g. Pradhan et al 2001). This doesn't mean a high correlation is "baked in" as it would be possible, in theory, for programs to "kink" the distribution and reduce poverty for a given median.

programmatic differences in left-tail versus median consumption (e.g. different relative prices of goods the poor consume intensively), non-programmatic differences in incomes driven by different relative prices of assets owned by the poor (e.g. unskilled labor), etc.) adds up to 1.2 percent of the observed variance in poverty so poverty programs could account for as little as.1 percent (given the existence of scaled and effective programs in at least some places cannot be zero).

II.C) National Development and Broader Measures of Social Progress

In addition to its impact on a kinky goal like extreme poverty, achieving high levels of national development is also a necessary and sufficient condition for achieving high levels of overall human well-being. The correlation of an omnibus measure of human well-being (Social Protection Index) and national development are extremely high (.967) (Pritchett 2016).

The Social Progress Index is the result of the effort of the Social Progress Imperative to create a new and better ways to compare development performance across countries. They explicitly do *not* use GDP per capita (or other measures of national development), but rather focus on direct measures of human well-being. The Social Progress Index (SPI) has three aggregate components called: 1) basic human needs, 2) foundations of well-being, and 3) opportunity. Each of these three components are built from four sub-indicators, which are each themselves built up from specific measures. For instance, the aggregate "basic human needs" (I) has four sub-components: I.1 "nutrition and basic medical care" I.2 "water and sanitation" I.3 "shelter" I.4 personal safety. Each of these is based on specific indicators, so, for instance, sub-component I.2 "water and sanitation" is based on: I.2.a "access to piped water", I.2.b "rural access to improved water source", and I.2.c "access to improved sanitation." I am not saying the SPI is the best measure of country-level human well-being, but it is a thoughtful and careful attempt to measure of social progress across countries and uses 53 distinct indicators—which include economic, education and health indicators but also non-standard indicators like religious tolerance, freedom from crime, political rights.

I regress the SPI (re-scaled 0 (worst) to 100 (best)) on three indicators of national development: (ln) GDP per capita (proxy for productive economy), the POLITY2 measure of autocracy/democracy (proxy for responsive polity) and World Governance Indicator of Government Effectiveness (proxy for capable administration), also each scaled 0 to 100[11] for 140 countries (excluding high income oil countries and one country (El Salvador) whose GDPPC data seemed wrong). The National Development Index adds the three components using OLS coefficients as weights.

---

[11] I don't think any hinges on using these particular three proxies for the underlying concepts of national development. For instance, the Fund for Peace presents a Fragile States Index that has multiple components. Two of those, "Public Services" and "State Legitimacy" are potential alternative empirical proxies for the concepts of "administrative capability" and "political responsiveness." A regression the overall Social Progress Index on GDP per capita, FSI: Public Services and FSI: State Legitimacy (all scaled to 100) the R-Squared is .947 (even higher), with all three indicators having powerful roles.

Figure 7 shows that national development is empirically necessary and sufficient for achieving high levels of the SPI. No country has achieved an SPI in the top third of countries (above 70.1) without a National Development Index above 68.6 (Argentina's level)[12]. Similarly, no country in the top third of NDI has an SPI less than 61.6.

The SPI and NDI have a correlation of .967 (R2 of the regression was .935). This is an amazingly tight relationship of two conceptually and empirically different measures as different cross national measures of the *same thing* from different sources or methods—like "years of schooling of the adult population" or "child mortality"—often don't have cross-national correlations as high as.96, just due to pure measurement error.

Figure 7: National development is empirically necesary and suffiicent for high levels of the Social Progress Index
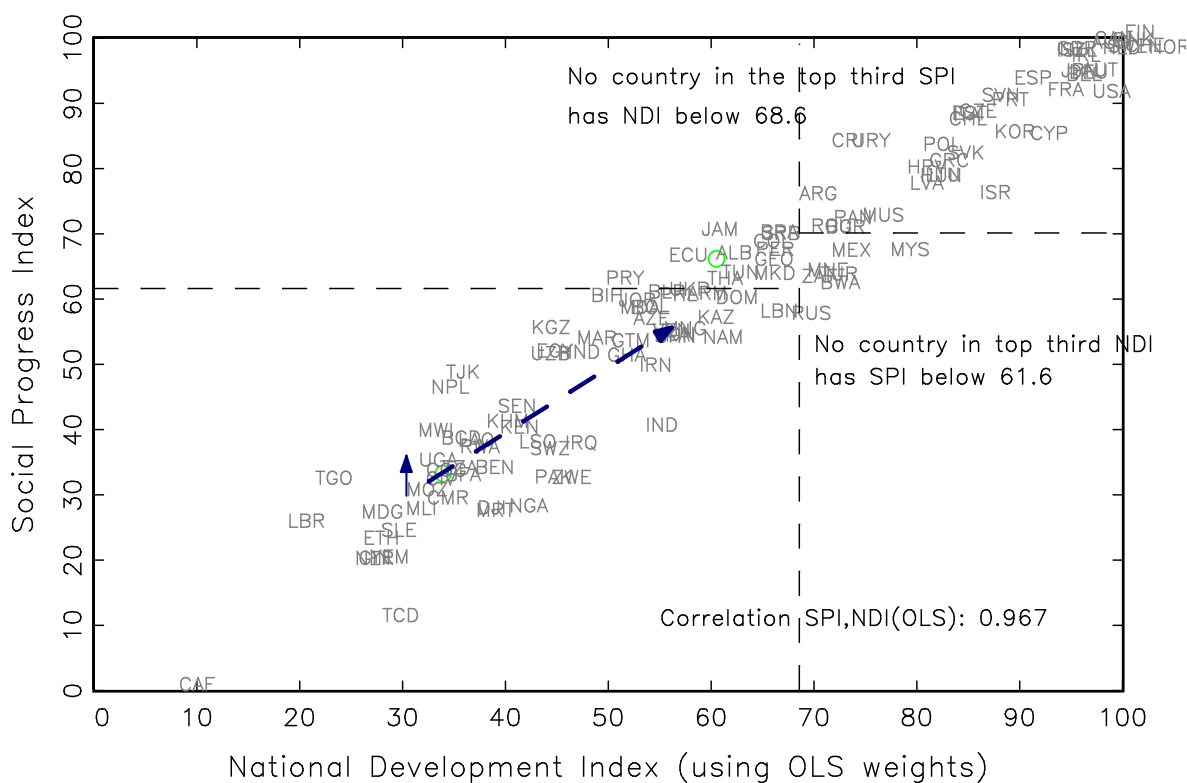
Source: Author's calculations with data and procedures as described in the text.

---

[12] Measures of human well-being are sometimes to point out that GDP per capita is a weak proxy for human well-being (for which of course no economist ever proposes it) by showing "outliers" that achieve high SPI with low(ish) GDP per capita. But "national development" *includes* politics, state capability, and social transformation. With this broader definition countries that are sometimes high performers for their GDPPC like Costa Rica (CRI in the graph, which overlaps URY) does have high SPI and "over-performs" even its NDI, but it is not a massive "outlier" as it has high NDI.

As with poverty, the strong and tight relationship implies the potential gains in social progress for a given level of national development are quite small (relative to the range of SPI). Mozambique (abbreviation MOZ) has roughly the same actual and predicted SPI (hence NDI) of about 30. Suppose somehow Mozambique were a "star performer" on Social Progress for a given level of national development, in the specific sense it has SPI higher by a residual standard deviation (so, on the assumption of a normal distribution was in the 84th percentile of countries with its NDI rather than 50th). Then its SPI would be 36 (illustrated with the vertical arrow in Figure 7). This gain is not nothing, but still would leave Mozambique's SPI below Laos, Bangladesh or Kenya. In contrast, if Mozambique improved by one standard deviation on each of the elements of national development the SPI would reach 56, higher the SPI of upper-middle income countries like Morocco or Indonesia (dashed "northeast" arrow in Figure 7).

Table 2 shows the empirical relationship of the three components and 12 sub-components of the Social Progress Index with proxies for national development. Each of the three components of the SPI has a very strong correlation with NDI (Basic Needs .904, Foundations of Well-Being .925, and Opportunity .932). All of the 12 sub-components (less one[13]) are also strongly associated with national development.

---

[13] The indicator without a strong positive correlation is "environmental quality", which includes greenhouse gas emissions, which are positively associated with GDP per capita.

Table 1: The Social Progress Index—and all of its components and sub-components—are strongly associated with three indicators of national development

Social Progress Indicator, its three components (Basic Human Needs, Foundations of Well Being, Opportunity) and the four sub-components of each component	Economic Productivity ((ln) GDP per capita, PWT8.0, rescaled 0 to 100)		Administrative Capability (World Governance Indicators, Government Effectiveness, rescaled 0 to 100)		Political Responsiveness (Polity IV Project, Polity 2, rescaled 0 to 100)		R-Squared of regression on national development indicators
	OLS coeff.	t-stat.	OLS coeff.	t-stat	OLS coeff.	t-stat	
Social Progress Index	0.53	13.67	0.34	7.38	0.12	5.01	**0.935**
I) **Basic Human Needs**	0.74	12.10	0.18	2.46	-0.02	-0.43	**0.835**
I.1) Nutrition and Basic Medical Care	0.57	8.86	0.34	5.17	0.18	5.06	0.865
I.2) Water and Sanitation	0.31	4.95	0.51	8.15	0.23	7.11	0.873
I.3) Shelter	0.80	9.74	-0.09	-0.95	0.04	0.79	0.672
I.4) Personal Safety	1.17	11.78	0.01	0.06	0.06	1.12	0.784
II) **Foundations of Well-Being**	1.06	13.30	0.04	0.47	-0.01	-0.36	**0.820**
II.1) Access to Basic Knowledge	-0.02	-0.27	0.77	7.86	-0.09	-1.83	0.603
II.2) Access to Info and Comm.	1.00	10.62	-0.11	-1.09	0.04	0.73	0.707
II.3) Health and Wellness	0.53	8.02	0.22	3.25	0.21	6.11	0.816
II.4) Environmental Quality	-0.18	-1.55	0.50	4.34	0.01	0.13	0.242
III) **Opportunity**	0.11	1.33	0.52	6.43	0.18	4.34	**0.709**
III.1) Personal Rights	-0.08	-0.86	0.53	5.68	0.55	11.58	0.765
III.2) Personal Freedom and Choice	0.16	2.06	0.66	8.65	-0.01	-0.37	0.757
III.3) Tolerance and Inclusion	0.19	1.71	0.41	3.70	0.14	2.48	0.517
III.4) Access to Advanced Education	0.93	11.21	0.17	2.04	0.03	0.73	0.824
Source: Author's calculations.							

National indicators of subjectively assessed well-being are also highly correlated with national development. Regressing the Cantril "ladder of life" measure of average subjective well-being on the three national development indicators has an R2 of .66 (correlation .812 with an OLS NDI). The World Happiness Report has developed another index of human well-being based on the empirical relationship of seven factors (like "perceptions of corruption", "Healthy life expectancy", "social support" and measures of affect) to the "ladder of life" measure of subjective well-being. An equally weighted index of the six elements of the happiness index regressed on the three indicators of national development produces an R2 across 120 countries of .788 (correlation with OLS NDI .887). Again, the correlation between this six element "happiness" index and the directly observed "ladder of

life satisfaction" measure is .81. While these are lower than the SPI/NDI correlation, the three indicators of human well-being (SPI, subjective life satisfaction, World Happiness Report) are only about as tightly correlated among themselves as each is with a measure specific national development index.

I.D) National development brings elimination of poverty and high levels of human well-being

With the accumulation of more and better data we can show the relationships of national development with poverty, overall human well-being, or specific indicators of well-being are as anyone ever claimed they would be.

What is odd is that anyone ever doubted this. Four-fold national development is a human well-being machine. Take any objective that contributes to well-being that is strong and widely spread—access to water, better health, improved shelter, more schooling—national development is built to increase the accomplishment of that objective. A more productive economy that produces broad based increases in incomes allows households more income to pursue their objectives so, to the extent these objectives are private goods, it would be very strange indeed if higher private incomes did not lead to higher levels of consumption (and indeed all that empirical matters in the SPI components for "water and sanitation" and "shelter" and "access to basic knowledge" the only significant correlate is GDPPC)[14].

But if the human well-being objectives require "public goods" (non-rival, non-excludable) or the markets for these goods have "market failures" then this is precisely what governments that are responsive and capable can address. Indeed, for the component "environmental quality" the only strong partial correlates were capability and polity, not GDP per capita and for "personal safety" the only partial correlate was state capability. No one, even the most ardent and market-oriented economist, ever made the case income alone would solve all problems. A responsive polity and capable state was always an integral part of the vision of development.

II)   *RCTs in development as a method for improving human well-being*

Back to the madness. How did we get to studies of chickens? How did development economics get to thinking about anything *but* national development? How would one provide argumentation or evidence or warrant for a claim that a study with a particular method of the relative effectiveness of targeted programs of chickens versus cash was the "best investment" for fighting poverty?   There are three multiplicative elements to such a claim: (a) the likelihood a study produces reliable and useable knowledge, the likelihood the knowledge

---

[14] And, one would expect the relationship with national development to be even stronger/tighter for "necessities" as economists definition of "necessity" is something for which marginal utility gets very high as consumption of it falls and, related, something for which the price elasticity (especially at low levels) is expected to be very low. A simple Engel curve—that food share in consumption declines linearly with (log) aggregate income/consumption is the arguably the best documented fact in all of economics.

changes events in the world that improves outcomes and (c) the total gain to HWB (in some normative evaluation) from such changes.

Figure 8: The empirical magnitudes to be resolved to make decisions about the expected relative value of various types of investment in research
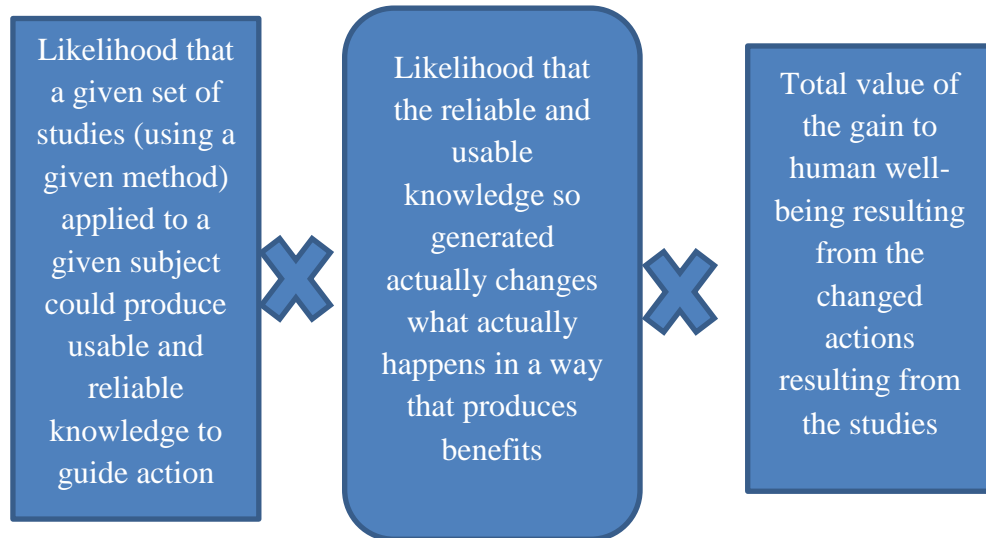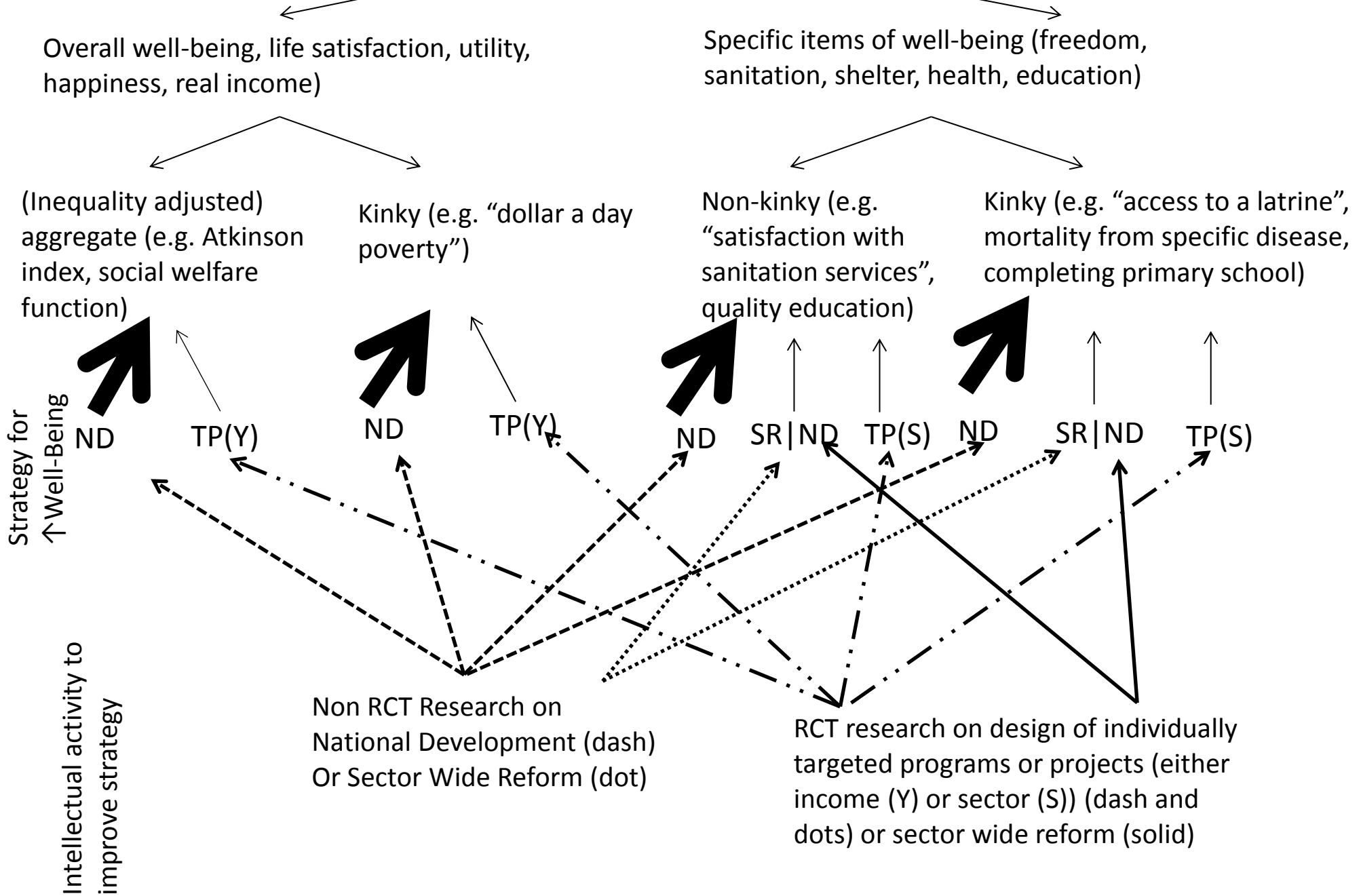

Figure 9 from the top down provides a map of the array of framings of measures of human well-being (omnibus/aggregate and specific indicators or domains) and whether the normative evaluation of those is kinky or not. The essence of the "kinky" measure is *not* that the poorer (those with less sanitation/education/energy) receive more weight in the measure of human well-being and the richer (those with more of a specific thing) less weight. Any of the standard inequality measures of aggregates, like the Atkinson index or a standard SWF with the assumption of declining marginal utility can accommodate that (with parameters giving different intensities of "preference for the poor") and similarly sector measures can give greater weight to specific levels of service or certain groups. The essence of a kinky measure is that the gain to human well-being above some arbitrary threshold (like a poverty line, or "primary school completion" or "access to a latrine") is *exactly zero*.

From the bottom up the arrows illustrate claims about the strength/magnitude of causal impacts on human well-being of national development (ND), targeted programs (in income (TP(Y)) or specific sector indicators (TP(S), or sector wide reforms (SR|ND).

Figure 9: What is the best investment in research activity in development for promoting human well-being?

Overall well-being, life satisfaction, utility, happiness, real income)

Specific items of well-being (freedom, sanitation, shelter, health, education)

(Inequality adjusted) aggregate (e.g. Atkinson index, social welfare function)

Kinky (e.g. "dollar a day poverty")

Non-kinky (e.g. "satisfaction with sanitation services", quality education)

Kinky (e.g. "access to a latrine", mortality from specific disease, completing primary school)

Strategy for ↑Well-Being

ND   TP(Y)   ND   TP(Y)   ND   SR|ND   TP(S)   ND   SR|ND   TP(S)

Intellectual activity to improve strategy

Non RCT Research on National Development (dash) Or Sector Wide Reform (dot)

RCT research on design of individually targeted programs or projects (either income (Y) or sector (S)) (dash and dots) or sector wide reform (solid)

The "best investment" claim is that link RCT study to improved targeted program to raise incomes (TP(Y) (dash and dots arrow) times the gains from TP(Y) on Kinky Development ("extreme poverty") (smaller arrow) is larger in benefit-cost ratio than any other. The opposing claim, that research on national development is superior, is *either* thazt (a) a claim that the impact of non-RCT research on national development on national development outcomes (dashed arrow) times the impact of ND on a kinky aggregate development (e.g. "extreme poverty) (big arrow) is bigger *or* (b) that the impact non-RCT research on national development outcomes (dashed arrow) times the impact of national development outcomes on (inequality adjusted) aggregate human well-being is bigger in valuation term (given any reasonable valuation) than that of RCT on TP(Y) on extreme poverty.

There are two elements of figures 8 and 9 that nearly all economists agree on.

First, the magnitude in dollar terms of gain from national development and sector wide reforms are orders and orders of magnitude larger than possible with targeted programs. The *randomistas* do not typically argue that the gains to poverty from growth would not be large as, given the figures above, this is obviously false, but rather the impact of research on growth is small/weak/zero.

Second, the impact of RCT research on national development or sector wide reforms is almost certainly limited. A reason I stressed that the processes of national development operate at a ontological level higher than the individual is that RCTs are typically only possible (and certainly only possible to "power up") when a large number of units can be assigned to "treatment" and "control" status. This is impossible for economy wide or national politics wide or organization wide phenomena.

Therefore, the most common claims by the sophisticated advocates of RCTs are some sets of the following:

- While the impact of national development on all four types of well-being indicators is large, national development is sufficient for achieving kinky goals, and necessary for high goals, the impact of research on national development is very, very near zero (dashed arrows from non-RCT research to national development essentially don't exist) therefore even if the impact of RCT on actual targeted programs (for income Y or specific indicators S) is small, and only on the kinky, the valuation of the research is cost-effective if only because it is effective at all whereas the other types of research have (near) zero effectiveness.

  Or,

- A different line of argument is that the valuation of human well-being is exclusively kinky so gains above the threshold don't matter therefore the national development impact on the non-kinky has very low value.

With either of the above claims*,* one has to add, *and:*

- o RCTs have to be able to generate reliable and useful knowledge about targeted programs for income or specific goals.
- o The reliable and useful knowledge generated by RCTs has to actually change the course of events, that is, the knowledge generated by RCTs has to be a (key) binding constraint to the scale of use of better targeted programs.

II.A) Widely Accepted Claim I: The magnitudes of gains from national development are orders of magnitude larger than from targeted programs

Kenny and Pritchett (2010) show that, on basically any measure of human well-being progress in national development (called 'drive') or gains in sector wide efficacy (called 'shift') dominate, by order of magnitude the gains from targeted programs (called 'kink').

Pritchett, Sen, Kar and Raihan (2016) estimate the net present value of GDP added (or lost) relative to a "business as usual" counter-factual from various episodes of growth or contraction. Our technical method of giving dates and sizes to growth episodes suggests that the growth accelerations in China in 1977 and 1991 produced NPV gains of 2.65 trillion and 11.8 trillion (over 14 trillion total). The growth accelerations in India in 1993 and 2002 produced gains of 1.1 trillion and 2.5 trillion (total of 3.6 trillion). Indonesia's growth acceleration in 1967 produced a NPV gain over BAU of 1.1 trillion. The absolute gains from Vietnam's acceleration in 1989 were smaller, $455 billion, but this was an NPV gain of $6,911 per capita. These growth episodes were also associated with a rapid reduction in "extreme poverty" to very low levels (Figure 5). The losses from decelerations relative to the BAU growth rate are also similarly massive. Brazil's loss from the 1980 deceleration episode was 7.5 trillion dollars, the loss to Indonesia from the 1996 East Asia crisis was near a trillion dollars and the combined losses from the Mexico decelerations of 1981 and 1989 were 1.5 trillion. Many African countries, though small in absolute terms, had massive losses of NPV per capita from growth decelerations: Malawi 1978 P$9,600; Kenya 1967 P$13,300, Cote d Ivoire 1978 P$15,200.

The "livestock" trial published in science showed a complex, multi-faceted "graduation" approach to the ultra-poor raised year 3 incomes in 5 of 6 study sites. The magnitudes, averaged across the five sites, were that $4545 per household in costs in year 1 and 2 produced *$344 per household gain* or, on the assumption of a typical household size of 4, *$86 per person*. On the *assumption* this year 3 amount persists forever, this implies, at a 5 percent discount rate, an average household NPV gross gain of $8472 in gains per household which was about a 7 percent rate of return. Assuming crudely four people per HH that implies an $1136 investment per person produces a once off level gain in year 4 of $86. Suppose we wanted to use the knowledge from this "gold standard" evaluation of an anti-poverty program to raise income in Vietnam by an NPV of $6,911. That would cost $333 billion dollars in

program investments—more than Vietnam's *current* (post growth) total GDP or about three times *total* global development assistance.

The gains from well-functioning financial systems—especially avoiding a large crisis are huge. Estimates of the losses in 2014 to OECD GDP from the 2008 financial crisis was about 3.5% or 1.9 *trillion* dollars, if that is a "permanent" loss relative to a no-crisis counter-factual the Net Present Value of that (at 5 percent) is 38.2 trillion dollars. The US Federal Reserve estimates the NPV of the loss to the USA at US$70,000 for each citizen. The *total* stock of micro-finance assets in 2016 was about 102 billion dollars. Suppose, at the wildest possible positive view, the annual gain to borrowers was 10 percent of the stock and this implies a gain to borrowers of 10.2 billion dollars. Suppose, again at the far reaches of optimism, rigorous research could somehow *double* that gain (relative to a counter-factual) then the gain would be an additional 10.2 billion dollars globally to micro-finance borrowers. The losses from a single (large) global financial crisis were on the order of 200 times larger than the gains from doubling the total benefits from microfinance.

Raising the learning levels in basic education of children to prepare them for their 21st century lives is hugely important. If one takes a view of the challenge how important is research on the enrollment impacts of conditional cash transfers? Using a recent assessment of learning in Zambia, the PISA-D, I estimated that, of the 360,000 children aged 15 in Zambia only 36 percent were in school and assessed and of those only an estimated *5 children total* (not five percent, five children, like the five fingers on your hand) who could read at globally proficient levels (PISA levels 4 or above). Moreover, even if, through whatever heroic efforts, including say, conditional cash transfers, enrollment of 15 year olds increased to 100 percent, at current levels of learning this would add only *14 children* who could read at globally proficient levels. But, Vietnam has learning performance that is massively better than Zambia's in ways that are not accounted for by targeted programs but rather appear to be superior operation of a sector wide education system.

II.B) Widely Accepted Claim II: RCT studies do not address national development

Pritchett (2014) draws on the Vivalt (2015) review of RCT results to compare the topics on which enough RCTs have been done to compare results with some simple questions about whether topic X is even plausibly a major cause of growth. None of the common domains of RCTs (conditional cash transfers, microfinance, improved cook stoves, deworming) are plausibly important determinants of the level of income or of growth. Nor do their advocates make that claim. The reason I emphasized the social nature of the four-fold national development transformation is that what the RCT needs to be successful as a research strategy is (a) (reasonably) clean assignment of units to "treatment" and "control" and (b) enough units for adequate statistical power. This is why, almost necessarily, the method lends itself well to *individualizable* (or small unit, like clinic or school or police station) interventions and not to studying the impact of policy on market performance or the evolution of the governance of a polity or the social transformation. Even if an RCT were to address these topics (like a study on information and voter behavior) they would do so in a way that, if and when the results were extrapolated to the scale of the relevant they would have no

more "rigor" or warrant as evidence than any other method as, in order to use the method precisely the "general equilibrium" effects at the system scale had to be bracketed.

II.C) Needed But False Claim I: The impact of any research (RCT or otherwise) on national development (or sector wide reforms) is vanishingly small

Given the relative magnitudes of the gains to human well-being from national development and that the RCT method is not well applied to promoting national development or sector wide changes, the argument has to be that *national development, including economic growth, is roughly impervious to any sort of research.*

This argument is at odds with commonly accepted interpretations of events in a number of countries. One, there are a number of countries (e.g. China, India, Vietnam, Indonesia) that said (1) "Based on our reading of the existing evidence (including from economists) we are going to shift from policy stance X to policy stance Y in order to accelerate growth", (2) these countries did in fact shift from policy stance X to Y and (3) the countries did in fact have a large (to massive) accelerations of growth relative to BAU as measured by standard methods (Pritchett et al 2016). One had to be particularly stubborn and clever to make the argument: "Politicians changed policies to promote growth based on evidence and then there was growth but (a) this was just dumb luck, the policy shift did not actually cause the shift in growth something else did or (b) (more subtly) the adopted policies did work but that was just dumb luck as there was not enough evidence the policies would work for this to count as a win for 'evidence' changing policy."

There are also a fairly large number of countries that did the opposite. Economists (from their country and others) have said to the leadership of countries: (1) "If you persist in policy stance X you are going to experience large (to massive) negative consequences for economic growth," (2) the leaders have not listened, and (3) there have been precisely the predicted negative consequences. The Venezuelan economy is not in 2018 spiraling into hyperinflation and in the midst of a tragic economic depression because "economists have little useful to say about economic growth" in the sense the advice, if followed, would be useful. If the argument is that research can learn reliable advice but this doesn't mean it will change the course of events, then the question is whether it *never* changes the course of events. There are also cases in which governments have said "based on what economists say we are going to switch paths to avoid massive downturns/hyperinflation", have done so, and it has worked (in the sense at least that a crisis did not happen). While the "growth accelerations" might have been hard to predict with standard policies (Hausmann, Pritchett, Rodrik 2005) there is empirical evidence that "growth collapses" are rather more predictable (Breuer and McDermott 2011).

This is not to say that all research based claims about policies for growth have been right. The "lost decades" in Latin America and the "transition depression" in some (not all) former Soviet dominated countries are both examples of adopting policies for growth based on recommendations that seemed not to work. However, as a paper in this volume points out, among the top ten most prescribed medicines many work on only a third of the patients. So

because a recommendation is not universally successful does not mean it is not a good recommendation. If I can give you a tip that increases your odds of winning a million dollar lottery by 10 percent, it is massively worthwhile. More recent reviews suggest the "pox on all the houses of growth research" stance and a view recommendations had been worthless are too extreme (e.g. Easterly 2018 on the "Washington Consensus", Irwin 2019 on trade).

Keep in mind from Table 1 just how small the expected effect of research on growth has to be to be as poverty reducing as what can be expected from improved poverty programs. Suppose that growth advice was given to 10 countries and in 9 of 10 it either was not adopted or was adopted and did not work but in one of 10 accelerated growth by 2 ppa for 20 years. Then even at this lack of efficacy it is still, for the poorest countries, as poverty reducing. (And obviously if those countries that happen to adopt are large countries (China, India, Vietnam, Indonesia (1960s)) then the total well-being gain is massive even if it is mostly ineffective).

Moreover, the weak performance of growth recommendations in the 1980s and 1990s could just as easily lead to recommendations for much *more* research on how to promote national development rather than less, given the value of getting good rather than bad advice on these hugely consequential issues. It is not as if economics was complacent and either ignoring the negative growth experiences from many episodes of policy reform (e.g. World Bank 2005) or sticking to "mindless growth regressions." An approach taking into account the episodic nature of developing country growth (e.g. Ben-David and Pappell 1998, Pritchett 2000, Jones and Olken 2008 , Berg, Ostry and Zettlemeyer 2012) married with a diagnostic approach (e.g. Hausmann, Rodrik, Velasco 2008, Hausmann, Bailey, Warner 2008, Rodrik 2008) was maturing even as the *randomista* movement was taking off[15].

II.B)   *Needed but False Claim II: Valuation of human well-being is "kinky"*

The other path in Figure 8 and Figures 9 into a priority within development field intellectual activity for RCTs is to adopt exclusively kinky measures of human well-being. This can make the fact that national development is a necessary condition for moderate to high levels of well-being and the massive gains from national development less compelling. I have written extensively elsewhere about why kinky goals generally, and low-bar poverty specifically, are illegitimate in every way: economically (Pritchett 2006, Pritchett 2013), morally (Pritchett 2014c), politically (Gelbach and Pritchett 2002, Pritchett 2005, Pritchett 2014a, Pritchett 2014b) or as goals for development (Pritchett 2015) or development

---

[15] And siphoning off from growth research even funding intended to be channeled to growth research. For instance, the Crépon, Devoto, Duflo, and Pariente (2015) paper re-reviewed by by Florent Bédécarrats, Isabelle Guérin, Solène Morvant-Roux, and François Roubaud (2019) and discussed in this volume was funded and promoted by the *International Growth Centre*, which was originally funded by DFID to improve "growth analytics" in order to lead to more prioritized and pragmatic recommendations to countries for policies to promote growth. Whatever the paper's (de)merits substantively it is a paper about a targeted program and no one pretends it is a paper about promoting national development or even growth.

organizations (Pritchett 2013a) and so can be brief. The simple, but compelling, argument against kinky goals in either income or specific indicators is "introspection plus the Golden Rule."

Introspection. The essence of "kinky" is that gains are *exactly* zero above a low threshold. Ask yourself about yourself: did your personal valuation of income fall to exactly zero when your income passed some low threshold? Did your willingness to pay for higher quality sanitation facilities drop to zero at an outdoor latrine? Did your personal valuation of education drop to zero when you finished primary school? The only honest answer is no.

Golden Rule. A widespread (if not universal) principle of "moral realism" is something like the "golden rule"[16] (do unto others) or the Kantian categorical imperative ("Act only according to that maxim whereby you can, at the same time, will that it should become a universal law" (Kant 1785 (1998)). By the Golden Rule/Kantian Moral Imperative—and frankly common sense--adopting for the general assessment of the well-being of other people by a standard you would never accept for yourself is morally wrong.

Any attempt to "solve" this by claiming the objective function is a "combination" of kinky and non-kinky goals means the overall goals are non-kinky and it is just a question of weights but the massive gains above the threshold are relevant. The replacement of the kinky MDGs with the broad and expansive SDGs should has ended the relevance of the kinky as legitimate expression of development (Pritchett 2015).

II.C)   *Needed but False Claim III: RCTs can reliably generate evidence that improves targeted programs aimed at kinky (aggregate or specific) development goals*

Another path to claiming RCT studies as the "best investment" is to claim that impact evaluation of programs/projects using RCTs are likely to produce rigorous, reliable, and usable evidence that can lead to the design of more effective programs. As I, and many others, including many authors in this volume, have argued: (a) these claims never had any solid evidence but was just asserted on faith, (b) claims that RCTs would "resolve debates" about impacts based on heterogeneity in observational studies was *ex ante* not just empirically unlikely but logically impossible (Pritchett and Sandefur 2013), (c) empirically the reviews of empirical studies fail to show sufficient consistency to be reliable (Vivalt 2015), even within specific topics like improving learning in basic education (Evans and Popova 2016) or deworming (e.g. the "Worm Wars"), and the variability across "rigorous" studies is sufficient that, at least in some instances, relying on the "rigorous" evidence would not reduce the prediction error about program impact in a given context relative to simple methods (Pritchett and Sandefur 2015)—which is exactly what everyone except the *randomistas* expected (Pritchett 2018), (d) the "construct validity" (the robustness of results across variations in the design space) of RCTs is low (Nadel and Pritchett 2016, Kerwin and

---

[16] Parfit (2011) argues that three common approaches to moral questions the Kantian deontological, consequentialism, and contractualism ultimately converge to the same answers and that these are "correct" answers.

Thornton 2018, Kaffenberger 2018), and (e) one cannot use results "proven" with one implementer to extrapolate to impact when implemented by another organization, particularly from an NGO "proof of concept" to scaling with government (Bold et al 2018, Vivalt 2015).

The "livestock study" (Banerjee et al 2015) mentioned by Professor Blattman is sometimes taken as the "proof" than one can create "gold standard" evidence that could guide effective anti-poverty programming. In that context, there are seven points worth nothing. First, the IRR is 7 percent, which is not particularly impressive, it would not pass the 10 percent rate of return traditionally used World Bank project cost benefit analysis. Second, in my calculations above, I was being generous and not including in these calculations one of the six countries, Honduras, in which the livestock (chickens!) died and hence the program had pretty substantial negative impacts on households so the average given does not reflect all experiences. Third, it is not clear the program beats a cash transfer as the costs to produce the gains are very high. Fourth, there is not (yet) "rigorous" evidence that the gains of the program will be sustained. Their calculations suggesting this program produces positive NPV requires the *assumption* that the year three gains are sustained into the distance future, an assumption not supported by their data. If one uses the observed fall in measured annual durables consumption from year 2 to year 3 and extrapolate future income streams using that decay the NPV is *negative* for all but two of the six countries. Fifth, Bauchet, Mordouch and Ravi (2015) did an impact evaluation of a very similar program in South India and they find no impact on income or assets, they argue because the local economy was growing robustly so the livestock option was not attractive so we know *for sure* these results lack external validity at least across some external conditions. Sixth, one suspects there is a lack of "construct validity" in the sense that this "multi-faceted" program was complex and had many elements in part because the design was the result of a long period of more informal "trial and error" and "experiential learning" (Hammer et al 2012) by BRAC and hence even minor variants in the design or the fidelity of its implementation might not produce the positive results. Seventh, while the study was done across multiple sites, responsibility for implementation was the responsibility of the same organization in all sites, so the robustness of these results to any other organization is not at all assured.

The relevance for this paper is that if one wants to claim that the "best investment" is research into a topic that has very, very, limited upside gains (e.g. design of sector specific targeted programs) compared to other research that has massive upside gains (e.g. promotion of national development) the offsetting gains in likelihood of producing reliable, usable results have to be very large. If research into national development has a one in a thousand chance of producing usable results and RCTs a 100 percent chance this is a powerful argument in favor of (some) RCT research. However, there is no compelling or persuasive evidence or argument that the likelihood of producing reliable and useable results from a given magnitude of effort into RCTs is *higher at all*, much less that it is orders of magnitude higher.

II.D)   *Needed but (Probably) False Claim IV: Knowledge of the type RCTs can generate is a binding constraint to the adoption and implementation of better targeted programs*

In order for a proposed public policy/program/project to have (sustained) impact is had to meet a "trinity": it has to be "technically correct" (if implemented it has to be based on a correct set of causal claims about links from inputs to activities to outputs to outcomes), *and* "politically supportable" (one has to be able to generate and sustain a political coalition with sufficient power to authorize the needed actions and resources), *and* "administratively feasible" (one has to be able, with available administrative capability (or the capability that can be mobilized or created) to implement the program with sufficient fidelity to achieve the outcomes). Claims about improvements in human well-being from knowledge gained from RCTs depend on claims that knowledge about program design of the type RCTs can generate is "the" (or at least "a") binding constraint versus other constraints on effective action (Pritchett 2018). But it is not obvious policy design matters for outcomes. Chong et al (2012) show that, for a very specific policy outcome measure, return of misaddressed foreign mail, (a) the *de jure* policy is exactly the same in all countries and (b) the outcome, percent of mail return in compliance with the *de jure* policy, the outcome varies by as much as it possible can (zero percent to 100 percent) and hence (c) all of the variation is due to implementation, none policy.

The "design space" for a project/program aimed at any objective (e.g. women's empowerment, reducing farmer income variability, increasing savings, reducing morbidity from water-borne diseases, etc.) is likely to be large and complex (and unknown) in that there are many choices (e.g. who is responsible for what actions, how frequent should visits be, what is the content of informational messages transmitted, what is the magnitude of a loan, etc.) and many possibilities for each choice and some elements crucial for success might not even be known at the design stage. Doing an RCT establishes an estimate of "impact" which is a point (or set of points, one for each treatment arm) on the "response surface" of outputs or outcomes over a particular design. The previous section (II.C) was about how useful this inference about a point or set of points is when the response surface could vary across contexts or be very rugged (non-robust) with respect to design. But there is an additional concern that knowing the response surface over a project/program design that is administratively or politically impossible has limited or zero value[17] (Gass and Pritchett 2017, Pritchett 2018).

Knowing that projects/programs would have impact X or Y or Z if adopted in contexts where, even when X or Y or Z are fully known and agreed, these projects/programs have zero probability of political adoption may contribute to disciplinary knowledge but cannot be claimed to have benefits for human well-being. Pritchett (2010), drawing on Filmer and Pritchett (Filmer and Pritchett 1999), argues that much of the advocacy around the usefulness of RCTs for "policy making" presumes a "normative as positive" model of politics, even in domains in which that had been shown to be demonstrably false. One doesn't have to buy wholesale into public choice theory to accept that one cannot take

---

[17] This is just the obvious Kuhn-Tucker point from optimization subject to (potentially) many constraints; the "Lagrangian" or "shadow price" on slack constraints is zero.

seriously as a positive model the idea that actors in the public sector (politicians, policy makers, senior technocrats) are optimizing a social welfare function and constrained only by their knowledge of "what works."

The same logic is true of the capability of organizations expected to implement programs. Knowing that program would have impact X *if* it could be implemented with fidelity doesn't mean the existing organizations in the country, public or private, have that capability. A fair number of existing RCTs have not been able to demonstrate the causal link between the design of the intervention, outputs of the implementing organization, and outcomes. Rather what the experiment learned was that, even in the limited context of an experiment, the "treatment" (whether it was pay for performance, citizen information or top down instructions) could not alter the relevant behavior of the implementing agents to produce "outputs" (e.g. Banerjee, Duflo, Glennerster 2008 (ANMs in Rajasthan), Banerjee et al 2010 (public school headmasters, teachers in Uttar Pradesh), Banerjee et al (2012) (police in Rajasthan). And examples where an experiment worked to produce outcomes when implemented with an NGO did not work when scaled by the government (e.g. compare Duflo, Hanna, Ryan (2012) on cameras in classrooms in NGO schools to Dhaliwal and Hanna (2013) on biometrics in public sector health clinics in Karnataka). How much of observed variation in poverty or sector programs across countries is due to the large differences across countries (which for country indicators of state capability and human well-being outcomes across countries is large in Table 2).

There are pretty good arguments that the "technical" or "codifiable" knowledge that RCTs are best placed to produce are, at best, a minor constraint on the adoption and effective implementation of targeted programs (Pritchett 2018) versus political constraints on the "want to" and the capability of "can do" and neither of these are affected by the results of RCTs. In contrast, a good argument can be made that the use of existing knowledge in a given country is endogenous to politics and capability, rather than an exogenous factor, as the "codifiable" part of knowledge is a public good that, being non-rival and non-excludable should diffuse quickly and easily.

Conclusion

An impact evaluation with an RCT seems to be not really a tool for countries and their governments or for agencies interested in promoting development at all. Rather, it mostly seems a tool to guide that small part of the development process that is "charity" or "philanthropic" that is (a) going to give relative small amounts of money, (b) will not or cannot work though national (or state or local) governments, (c) has relatively "kinky" valuations (perhaps in part because they are rationing tiny resources) and (d) care about the ability of being able to attribute the gain in well-being causally to their specific intervention (rather than about indirect effects). Charity work is a good thing and if charity work can be done better guided by evidence from RCTs that is a good thing. A focus on charity work is likely how Bill Gates and Chris Blattman get to talking about chickens and their impact.

However, to confuse this tiny little segment of the world with the broader process of development is madness. South Korea today is not the South Korea of the early 1960s because its government did a better job promoting ownership of chickens. The world today is night and day better on nearly all objective measures of human well-being because of broad based national development and improved sector wide performance of the kind development was meant to promote (Pritchett 2017). To imagine that the same tools that international NGOs want to use to identify effective humanitarian interventions for the poorest of the poor that are directly attributable to the NGO's actions (and attribution is essential only for the NGO, not the recipient) are the "best investment" in poverty reduction, much less the best investment in development, is the legitimate use of a rigorous method, it is madness.

References

BANERJEE, A., DUFLO, E., GOLDBERG, N., KARLAN, D., OSEI, R., PARIENTÉ, W., SHAPIRO, J., THUYSBAERT, B. & UDRY, C. 2015. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science,* 348.

BANERJEE ABHIJIT; CHATTOPADHYAY, R. D., ESTHER; KENISTON, DANIEL; SINGH NINA 2012. CAN INSTITUTIONS BE REFORMED FROM WITHIN? EVIDENCE FROM A RANDOMIZED EXPERIMENT WITH THE RAJASTHAN POLICE. *NBER Working Papers*.

BANERJEE, A. V., DUFLO, E. & GLENNERSTER, R. 2008. PUTTING A BAND-AID ON A CORPSE: INCENTIVES FOR NURSES IN THE INDIAN PUBLIC HEALTH CARE SYSTEM. *Journal of the European Economic Association,* 6**,** 487-500.

BANERJEE, A. V. B., RUKMINI; DUFLO, ESTHER; GLENNERSTER, RACHEL; KHEMANI, STUTI; 2010. Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India *American Economic Journal: Economic Policy,* vol. 2**,** 1-30.

BEN-DAVID, D. A. D. H. P. 1998. Slowdowns And Meltdowns: Postwar Growth Evidence From 74 Countries *The Review of Economics and Statistics,* 80**,** 561-571.

BERG, A., OSTRY, J. D. & ZETTELMEYER, J. 2012. What makes growth sustained? *Journal of Development Economics,* 98**,** 149-166.

BOLD, T., KIMENYI, M., MWABU, G., NG'ANG'A, A. & SANDEFUR, J. 2018. Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education. *Journal of Public Economics,* 168**,** 1-20.

BREUER, J. B. & MCDERMOTT, J. 2013. Economic Depression in the World. *Journal of Macroeconomics,* 38**,** 227-242.

DHALIWAL, I. & HANNA, R. 2013. Deal with the Devil: The Successes and Limitations of Bureaucratic Reform in India.

DUFLO, E., HANNA, R. & RYAN, S. 2012. Incentives Work: Getting Teachers to Come to School. *American Economic Review,* 102**,** 1241-78.

EVANS, D. & POPOVA, A. 2016. What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews. *World Bank Policy Research Working Paper No. 7203*.

FILMER, D. & PRITCHETT, L. 1999. What Education Production Functions Really Show: A Positive Theory of Education Expenditures. *Economics of Education Review,* 18**,** 223-239.

GASS, J. & PRITCHETT, L. 2017. Returns on Scholarship (versus Organizational Learning) in Development Using (mostly) Education as an Example. *Presentation at University of Washington.*

GELBACH JONAH B; PRITCHETT, L. 2002. Is More for the Poor Less for the Poor? The Politics of Means-Tested Targeting. *The B.E. Journal of Economic Analysis & Policy,* 2**,** 1-28.

HAUSMAN, R., KLINGER, B. & WAGNER, R. 2008. Doing Growth Diagnostics in Practice: A 'Mindbook'. *CID Working Paper No. 177*.

HAUSMANN, R. & HIDALGO, C. 2009. The building blocks of economic complexity. *PNAS,* 106**,** 10570–10575.

HAUSMANN, R., PRITCHETT, L. & RODRIK, D. 2005. Growth Accelerations. *Journal of Economic Growth,* 10**,** 303-329.

HAUSMANN, R. R., DANI; VELASCO, ANDREWS 2008. Growth Diagnostics. *In:* SERRA, N. S., JOSPEH (ed.) *The Washington Consensus Reconsidered: Towards a New Global Governance.* Oxford: Oxford University Press.

JONES, B. & OLKEN, B. 2008. The Anatomy of Stop-Start Growth. *Review of Economics and Statistics,* 90**,** 582-587.

KAFFENBERGER, M. 2018. CONSIDERING CONSTRUCT VALIDITY: SEEMINGLY MINOR DESIGN CHANGES WITHIN THE "SAME" PROJECT IN UGANDA MAKE IT EITHER THE BEST OR WORST OF ALL GLOBAL LITERACY INTERVENTIONS. *RISE* [Online]. Available from: https://riseprogramme.org/blog/considering_construct_validity.

KANT, I. 1998 (1785). *Groundwork of the Metaphysics of Morals,* Cambridge UK, Cambridge University Press.

KAUFMANN, D. M., GIL; GURGUR, TUGRUL 2002. Voice or Public Sector Management? An Empirical Investigation of Determinants of Public Sector Performance based on a Survey of Public Officials *World Bank Research Working Paper*.

KENNY, C. P., LANT 2013. Promoting Millennium Development Ideals: The Risks of Defining Development Down *Center for Global Development, Working Paper*.

KERWIN, J. T. & THORNTON, R. L. 2018. Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures.

KRAAY, A. 2006. When is growth pro-poor? *Journal of development economics,* 80**,** 198-227.

NADEL, S. & PRITCHETT, L. 2016. Searching for the Devil in the Details: Learning about Development Program Design. *Center for Global Development Working Papers,* 434.

PARFIT, D. 2011. *On What Matters,* Oxford UK, Oxford University Press.

PINKOVSKIY, M. & SALA-I-MARTIN, X. 2016. Lights, Camera … Income! Illuminating the National Accounts-Household Surveys Debate. *The Quarterly Journal of Economics,* 131**,** 579-631.

PRITCHETT, L. 2000. Understanding Patterns of Economic Growth: Searching for Hills Among Mountains, Plateaus, and Plains. *World Bank Economic Review,* 14**,** 221-250.

PRITCHETT, L. 2006. Who is Not Poor? Dreaming of a World Truly Free of Poverty. *The World Bank Research Observer,* 21**,** 1-23.

PRITCHETT, L. (ed.) 2010. *The Policy Irrelevance of the Economics of Education: Is "Normative as Positive" Useless or Worse,* Washington DC: Brookings Press.

PRITCHETT, L. 2013a. The Dangerous Seduction of the Kinky Cambridge MA: Center for International Development.

PRITCHETT, L. 2014. Is Your Impact Evaluation Asking Questions that Matter? *Center for Global Devleopment* [Online]. Available from: https://www.cgdev.org/blog/your-impact-evaluation-asking-questions-matter-four-part-smell-test.

PRITCHETT, L. 2014a. A Development Agenda without Developing Countries? The Politics of Penurious Poverty Lines (Part I). *Center for Global Development* [Online]. Available from: https://www.cgdev.org/blog/development-agenda-without-developing-countries-politics-penurious-poverty-lines-part-i.

PRITCHETT, L. 2015. Can Rich Countries Be Reliable Partners for National Development? *Horizons: Journal of International Relations and Sustainable Development,* 2.

PRITCHETT, L. 2016. Turns out Development Does Bring Development. *Center for Global Development* [Online]. Available from: https://www.cgdev.org/blog/turns-out-development-does-bring-development.

PRITCHETT, L. 2017. The Perils of Partial Attribution: Let's All Play for Team Development. Available from: https://www.cgdev.org/publication/perils-partial-attribution.

PRITCHETT, L. 2018a. The Debate about RCTs in Development is Over: We Won. They Lost. New York: DRI.

PRITCHETT, L. 2018b. Knowledge or its Adoption? *Center for Global Development* [Online]. Available from: https://www.cgdev.org/publication/knowledge-or-its-adoption.

PRITCHETT, L. 2018c. We Knew Fire Was Hot. *RISE* [Online]. Available from: https://www.riseprogramme.org/publications/we-knew-fire-was-hot.

PRITCHETT, L., SAMJI, S. & HAMMER, J. 2012. It's All About MeE: Using Structured Experiential Learning ("e") to Crawl the Design Space. *HKS Faculty Research Working Paper Series*.

PRITCHETT, L. & SANDEFUR, J. 2013. Claims to External Validity and Development Practice Don't Mix: Theory, Simulation, and Empirics. *Journal of Globalization and Development,* 4**,** 161-197.

PRITCHETT, L. & SANDEFUR, J. 2015. Learning from Experiments When Context Matters. *American Economic Review,* 105**,** 471-75.

PRITCHETT, L. & SUMMERS, L. H. 2014. Asiaphoria meets regression to the mean. *NBER Working Papers,* 20573.

RODRIK, D. 2008. The New Development Economics: We Shall Experiment, but How Shall We Learn? *HKS Faculty Research Working Paper Series*.

VIVALT, E. 2015. How much can we generalize from impact evaluations? *http://evavivalt.com/wp-content/uploads/How-Much-Can-We-Generalize.pdf*.