Returns to Civilization Scale

Our objective is to identify the relationships between civilizational size, interconnectedness, technological advancement, and well-being.

The goals for this project are as follows.

- Build a model of socioeconomic output in terms of an economy's population size and interconnectedness.
- Build a model of technological complexity and relate it to a civilization's size and interconnectedness.
- Conduct a literature review of relevant concepts and present it in a narrative format.

Socioeconomic quantities as a function of civilizational size and interconnectedness

Goal: develop a model of an economy with multiple cities, C_1, C_2, ..., C_m, where C_i has a population N_i. Each city is, internally, fully mixed and satisfies the urban scaling properties as described by **Bettencourt**. Cities i and j and connected by a link L_{ij}, with distance d_{ij} such that the distance d satisfies the properties of a **metric**. The distances may be interpreted as resistance in intercity links.

The Mixing Populations property might have to be relaxed or modified. One possibility would be to add over multiple types of socioeconomic interactions, rather than aggregating them into a single quantity y. Some high value interactions, such as tourism or business travel, might show global mixing, while other interactions, such as daily commuting, would only mix to the intraurban level.

Hypothesis: we can generalize the urban scaling model to get a 7/6 exponent on world socioeconomic output in terms of population, albeit with a smaller coefficient than we would have for a single city, assuming there are economically meaningful interactions that mix globally.

An obvious challenge will be in empirically testing any result. As there is only one global civilization at a given point in time, there is no obvious way to run a regression to test if the scaling relationship holds.

Technological complexity, scale-dependence, and limits to growth

Related to the goal above, we want to quantify the extent to which total factor productivity--or relatedly, technological progress--is embodied in scale-dependent ways. Technical capability should probably not be regarded as an exogenous factor or as disembodied knowledge, but rather as (at least partially) embedded in a civilization in the form of human capital, physical capital, economies of scale, etc.

Civilizational Returns to Scale

Agglomeration economies at the urban level

An agglomeration economy is the socioeconomic benefit of urban concentration. The principle that cities generate wealth, through facilitating easy travel, trade, and communication, via geographic concentration is well-established, though there is some debate in the literature on the precise mechanisms.

Jane Jacobs emphasized the benefit of knowledge spillovers between industries, a point confirmed by **Glaeser et al**. Older research by **Alfred Marshall** emphasizes the economies of scale brought about by the industrial concentration that cities enable. Thus there is some tension in the literature between the relative importance of local diversity versus specialization. **Michael Porter** emphasizes the importance of cities fostering competitive business environments. **Additional mechanisms** that have been identified for agglomeration economies include the sharing of production inputs, including labor; concentration near important natural resources such as harbors; increased employment opportunities for workers; and expanded markets for firms.

Regardless of the precise combination of mechanisms, cities are indeed observed to increase in socioeconomic outputs as their populations grow. Research by **Lobo et al.**, **Bettencourt**, **West et al.**, and **Glaeser** find that measures of economic output, productivity, per-capita GDP, and patents rise by 7-27% for each doubling of a city's population.

Certain beneficial interactions that cannot easily be expressed in economic terms also tend to grow, per capita, as a city grows. Examples include **access to health care**, **access to cultural amenities and entertainment**, and **social interaction**.

City size brings about efficiency as well. **Cottineau et al.** estimate that the doubling of a city size increases population density by about 6%, which is associated with reduction in greenhouse gas emissions and energy consumption. The theoretical work of Bettencourt also demonstrates that more populous cities have higher population density and less per capita infrastructure (such as roads, water pipes, etc.).

Certain negative socioeconomic interactions also scale superlinearly with city size. Several works by **Barthelemy** **and** **collaborators** and **other authors** show that traffic congestion, as measured by lost time, increases with a city size. Crime **tends** **to** **increase** **with** **city** size, with studies reporting increases of 10-33% for a doubling of city size. While an increase of city size tends to reduce driving and overall nitrous oxide generation, the higher population density **tends to increase** individual exposure to NO_2 by 41% for a doubling of city size. **Evidence from China** suggests greater individual exposure to particulate matter in larger cities, though worldwide the evidence is **ambiguous**. Individuals in larger cities tend to be exposed to more **light pollution**. Larger cities increase risk of exposure to **communicable** **diseases** and some **mental** **health** risks.

In urban geography, a "city" is typically defined as a commuter shed around a central area. Such a definition is clearly dependent on the means of commuting. **Gerritse and Arribas-Bel** show that an increased number of roads and highways increase the productivity-to-population elasticity of a city; in other words, agglomeration economies work best when transportation infrastructure facilitates efficient travel. Frick and Rodríguez-Pose **show** that, in Subsaharan Africa, cities with a population greater than 3 million show a loss of economic productivity with population growth, which may be the result of **inadequate infrastructure**.

There is not necessarily a single ideal city size; in addition to infrastructure, the economic role of a city within a broader economy influences its optimum size. As Alain Bertaud **demonstrates**, cities with highly specialized knowledge-based professions, such as New York City, should naturally be larger than agricultural cities, with manufacturing cities occupying an intermediate position.

The observation of agglomeration economies naturally suggests that concentrating a nation's population into larger cities, either through densification or through conurbation via faster transportation technology, may be a strategy for national economic growth. **Ganong and Shoag**, **Glaeser and Gyourko**, **Herkenhoff et al.**, and **Hsieh and Moretti** have argued that the United States could achieve an annual GDP gain of $400 billion to $1.7 trillion per year by loosening zoning restrictions in the largest and most productive cities, thereby allowing them to grow, achieve agglomeration benefits, and shift workers from lower to higher productivity areas. However, it is unclear how much of these gains could be induced through zoning reform, and how much would better be addressed through lessening the negative effects of city size, such as through improving transportation, law enforcement, and health.

Agglomeration economy at the super-urban level

On the scale of a national, global, or some other super-urban scale, we hypothesize that returns to scale are driven by two key processes: interconnection and overall size. The economic gains from trade are well established and are rooted in the principles of specialization and comparative advantage, as described by David Ricardo. Briefly, Ricardo **demonstrated** that, given two individuals, cities, or nations, each will have a comparative advantage over the other in producing a good or service. Both will be better off if each produces in the area of their comparative advantage and trade, rather than not trading, so long as the cost of trading (e.g. shipping cost) is less than the mutual gain from trade.

The empirical evidence that trade is a driver of growth and income is extensive, as presented by, for instance, **Frankel and Romer**, **Dollar and Kraay**, **Dreher**, though **Ortega and Peri** do not find that openness to trade drives growth. Ortega and Peri do find, however, that openness to migration drives growth. The survey of **Drinkwater et al.** finds that immigration generally increases GDP in the host country, though can have mixed effects on wages.

The importance of transportation in sustaining trade and thus economic prosperity is well-studied. For transportation in general, causality is established by **Beyzatler et al**. **Campante**, **Feyrer**, and **Green** have demonstrated growth induced by air travel, especially **business travel**. **Pascali** demonstrates the historical importance of steamships in driving globalization and economic growth in the late 19th century. **Donaldson and Hornbeck** demonstrate the importance of railroads in driving growth in the American West in the 19th century, while **Donaldson** does so in the context of colonial India. **Chen et al.** and **Chong et al.** have demonstrated economic activity and social welfare driven by high speed rail in China. **Duranton and Turner** find that presence of interstate highways drove growth in American cities in the 20th century. **Botasso et al.** establish economic growth driven by ports.

While the matter is complex and in some cases up for considerable debate, there is evidence to support the hypothesis that a larger national population supports economic growth and higher GDP per capita. **Michael Kremer** has developed a model, backed by long-term empirical observation, of increasing incomes in terms of population by treating the rate of technological progress as dependent on overall population, rather than exogenous. This is in contrast with the Malthusian model, that in general technological progress should fail to keep pace with population growth and keep the world in a subsistence state.

Wesley and Peterson argue that high population growth can be an impediment to economic growth in low-income countries, while low population growth is also an impediment in high-income countries. **Garza-Rodriguez et al.** find a positive, bicausal link between GDP and population growth in Mexico from 1960 to 2014. **Abdullah et al.** show a negative correlation between population growth and GDP growth in Bangledesh. One should intuitively expect that population growth would depress GDP in the short term, as societal resources must be invested in the care and education of children who are not economically productive, and enhance GDP in the long term as the new generation enters the workforce, and this is generally confirmed by **Headey and Hodge**.

World history since the Industrial Revolution has been unkind to the Malthusian hypothesis that population growth will reduce the per capita availability of resources and hence living standards. From 1700 to 2012, world population and GDP per capita have both **increased** by about 0.8% annually, or a factor of 12 over that period. Since the 1990s, **the rate of malnutrition** has decreased in absolute terms, as well as a share of the world population. Since 1980, the year the Simon-Ehrlich wager was declared, the cost of a basket of natural resources has **decreased**, in terms of world median wage, by nearly two-thirds, despite growing population (or perhaps because of growing population under the Kremerian hypothesis).

In dealing with environmental challenges, humanity's recent record has been more mixed. Following the Montreal Protocol and the development of refrigerants that do not deplete ozone, the ozone layer has undergone a **partial recovery**. **Steffen et al.** have identified nine "planetary boundaries", the exceeding of which would pose a major threat to human well-being and to the natural environment. Of them, four are in a critical or near-critical state: biodiversity integrity, biogeochemical flows (particularly of phosphorus and nitrogen), climate change, and land-system change.

In principle, it is possible simultaneously to increase significantly human population, per capita GDP, and decrease humanity's contribution to each of the four critical planetary boundaries using technology that is commercialized today or well under development. The energy system, today dominated by fossil fuels, is responsible for about **two-thirds** of anthropogenic global warming. The system could be replaced with a combination of **solar, wind, geothermal**, **ocean**, **nuclear**, and potentially **fusion power**; the demand for liquid fuels could be met with **electrification of transportation**, **hydrogen**, and synthetic hydrocarbons; the demand for **industrial heat** could be met with electricity, hydrogen, **high-temperature gas cooled nuclear reactors**, and **renewable sources**; and residual emissions could be addressed with **carbon capture and sequestration**, **direct air capture**, or geoengineering schemes such as **massive tree planting**. Humanity's spatial footprint, driven primarily by agriculture and pasture, could be drastically cut by producing food with **greenhouses**, **hydroponics**, and using **plant-based** or synthetic meat substitutes, all powered by clean energy. Land use could further be cut through the construction of **dense cities**. While these solutions are technically possible, they pose formidable economic, social, and political challenges that we cannot assume will be easily resolved.

Whether the challenges are easier to solve in the context of a growing human population and economy, or a shrinking population and economy, is unclear. While a growing population and economy requires greater consumption of resources, such a society will also have greater capital and technological know-how to deploy toward solutions. Some environmental NGOs such as the **Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services** call for reductions in population and material wealth in response to environmental challenges such as biodiversity loss, while most are quiet or ambiguous on such questions.

Beyond the risk of resource overextension and environmental degradation, it is worth considering additional potential diseconomies of scale from growing human population and interconnection. Risks, observed harms, and potential harms from global interconnection include psychological tensions arising from **cross-cultural interaction**, the unknown but **potentially catastrophic** risks of **pandemic disease**, a **spike in terrorism** in the 2010s, rising damages from **cyber incidents**, the **drug trade**, and vulnerability to an event such as the

Carrington Event of 1859 and other potentially unforeseen systemic risks. We hypothesize that these risks scale superlinearly with civilizational size and interconnection, analogously to the scaling of harmful socioeconomic interactions observed in cities.

Insofar as an optimal level of interconnectivity and population can be identified for an economy or a civilization, the optimum will be highly dependent on its technological capability and its economy. Any estimation of an optimum level will also require valuations of the intrinsic value of human life, qualities of human civilization, and qualities of non-human nature, all of which may necessarily be highly subjective.

Civilizational Scale and Technology

The growth model of **Robert Solow** characterizes the size of an economy in terms of available labor and capital: $Q = A K^a L^{1-a}$, where Q is output, K is capital, L is labor, $0<a<1$ is a constant to be estimated empirically, and A is an (exogenous in Solow's initial formulation) residual known as Total Factor Productivity, often interpreted as technology. Capital depreciates as a certain percentage per year, and labor saves a fraction of wages to be reinvested in capital. Under these assumptions, it can be shown that Q/L (per capita GDP) will converge to a fixed value. Long term (non-diminishing) growth depends on advancement in technology.

It is almost certainly inaccurate to treat technology, and hence TFP, an an exogenous variable. **Kuznets** and **Julian** **Simon** argue that the rate of technological progress depends on population, as a greater population means more minds can be put to work on developing new ideas. **Grossman and Helpman** treat innovation as a deliberate process.

Several economists, such as **Westfield**, have relaxed Solow's assumption of constant returns to scale; we may have $Q = A K^a L^b$, where $a+b > 1$. Then some of observed TFP growth would be accounted for in accelerating returns to scale. Relatedly, **Sato** and other economists introduce the concept of a holothetic technology, for which technical progress and economies of scale are indistinguishable. Several economists, such as **Hulten**, have attempted to estimate the portion of TFP that is "embodied" in capital.

A society's technological capability exists in scale-dependent ways. As **Adam Smith** observes in the context of pin making, a seemingly simple task, efficiency improves with division of labor. For a society to be capable of a certain technological proficiency, it must have developed supply chains and a workforce that is trained at each point of the supply chain. Modern globalized manufacturing supply chains are **generally understood** to be more complex today than historically. An indicator of increasing scales of production is trade, as a **share of world GDP,** which has grown from 25% in the 1960s to over 55% in the 2010s.

As demonstrated by **Nagy et al.**, Wright's Law is generally the best predictor of how the price of a given technology will evolve. Wright's Law holds that the cost of a technology tends to decrease as a power of cumulative production: $y_t = Bx_t t^{-w}$, where t is time, y is price, x is cumulative production, and B and w are parameters to be estimated empirically. Solar photovoltaics, for example, have followed a cost reduction curve approximated by Wright's Law. Moore's Law, observed in transistors among many other technologies, is the pattern that costs tend to fall by a fixed percentage every year. When production grows exponentially, Moore's Law and Wright's Law are equivalent. These cost reduction patterns suggest that the market size for a given technology must be sufficiently large to allow either economies of scale or learning-by-doing to drive prices down to an economically viable level.

There are signs of a secular decline in technological advancement and economic growth in rich countries. **Bloom et al.** document that productivity of research in the United States has fallen by a factor of 41 since the 1930s, or 5% per year, with declines in semiconductor performance, agricultural yields, and medical research. **Tyler Cowen** has documented a lack of major technological advancement since the 1970s, relative to previous decades. Productivity growth in the United States **has fallen** from 3.2% from 1995 to 2004 to 1.3% from 2005-2014. Projecting these trends forward, **Robert Gordon** expects no significant increase in standard of living for most Americans over the next few decades.

Some technological capabilities show actual retrogression. In the United States, productivity of construction has fallen by **nearly half** since 1968 and is at the lowest level since the 1940s. The United States is facing rising construction costs of **highways and other infrastructure**. **Bent Flyvbjerg** has documented a growing share of the economy invested in megaprojects ($1 billion or more), which typically suffer delays and cost overruns.

Deep learning, a technology that has been subject to considerable excitement in the past few years and is the basis for some hopes (or **fears**) for a near-term acceleration in productivity, is a case study in severe diminishing returns to scale. Artificial intelligence researcher **Filip Piekniewski** illustrates the central problem: requisite computing power and data grow exponentially in the quality of desired results. This is not sustainable with the **slowing down** of Moore's Law for semiconductors. **Melanie Mitchell** observes that deep learning models tend to be brittle, in that they perform poorly with even slight deviations from the context in which they are trained. This condition, unless resolved, forecloses hope that deep learning systems can perform general purpose tasks in an out-oft-the-box fashion. Deep learning systems **require labor**, which is often hidden from the user, in training and data collection. It has become evident that the self-driving car, a flagship application of deep learning, is **not close** to being ready for commercialization.

The slowdown in technological capability, such as in the case of deep learning, raises the possibility that contemporary technological paradigms are facing diminishing returns to scale. If so, the problem will only become more severe in coming decades, as much of the world faces shrinking and aging populations. This raises the question of whether it is possible to pursue more scale-independent technological paradigms, and if so, what they might look like.