

Is a singularity just around the corner?

What it takes to get explosive economic growth

(Received 10 April, 1998)

Robin Hanson

School of Public Health

*University of California, Berkeley
Warren Hall, CA 94720-7360*

net: <http://hanson.berkeley.edu/>

email: hanson@econ.berkeley.edu

ABSTRACT

Economic growth is determined by the supply and demand of investment capital; technology determines the demand for capital, while human nature determines the supply. The supply curve has two distinct parts, giving the world economy two distinct modes. In the familiar slow growth mode, rates of return are limited by human discount rates. In the fast growth mode, investment is limited by the world's wealth. Historical trends suggest that we may transition to the fast mode in roughly another century and a half.

Can some new technology switch us to the fast mode more quickly than this? Perhaps, but such a technology must greatly raise the rate of return for the world's expected worst investment project. It must thus be very broadly applicable, improving almost all forms of capital and investment. Furthermore, investment externalities must remain within certain limits.

Introduction

Many technological enthusiasts, impressed by the potential of various envisioned technologies, have speculated that technology may soon become so productive as to induce a "singularity", a period of extremely rapid growth. (A collection of speculations can be found [here](#) and [here](#).)

How rapid? Over the last few centuries, population has doubled roughly every 70 years, per-capita consumption has doubled roughly every 35 years, and scientific progress has doubled roughly every 15 years. In contrast, [computing power](#) has doubled roughly every two years for the last half-century. Many have speculated that perhaps economic growth rates will soon match or even greatly exceed current computing-power growth rates.

What would it take, exactly, for a technology to make the economy grow this fast? This paper offers a simple economic analysis intended to illuminate this question. We will find that while very rapid growth is possible in principle, this requires enabling technologies to meet some strong conditions. It is hard to see how

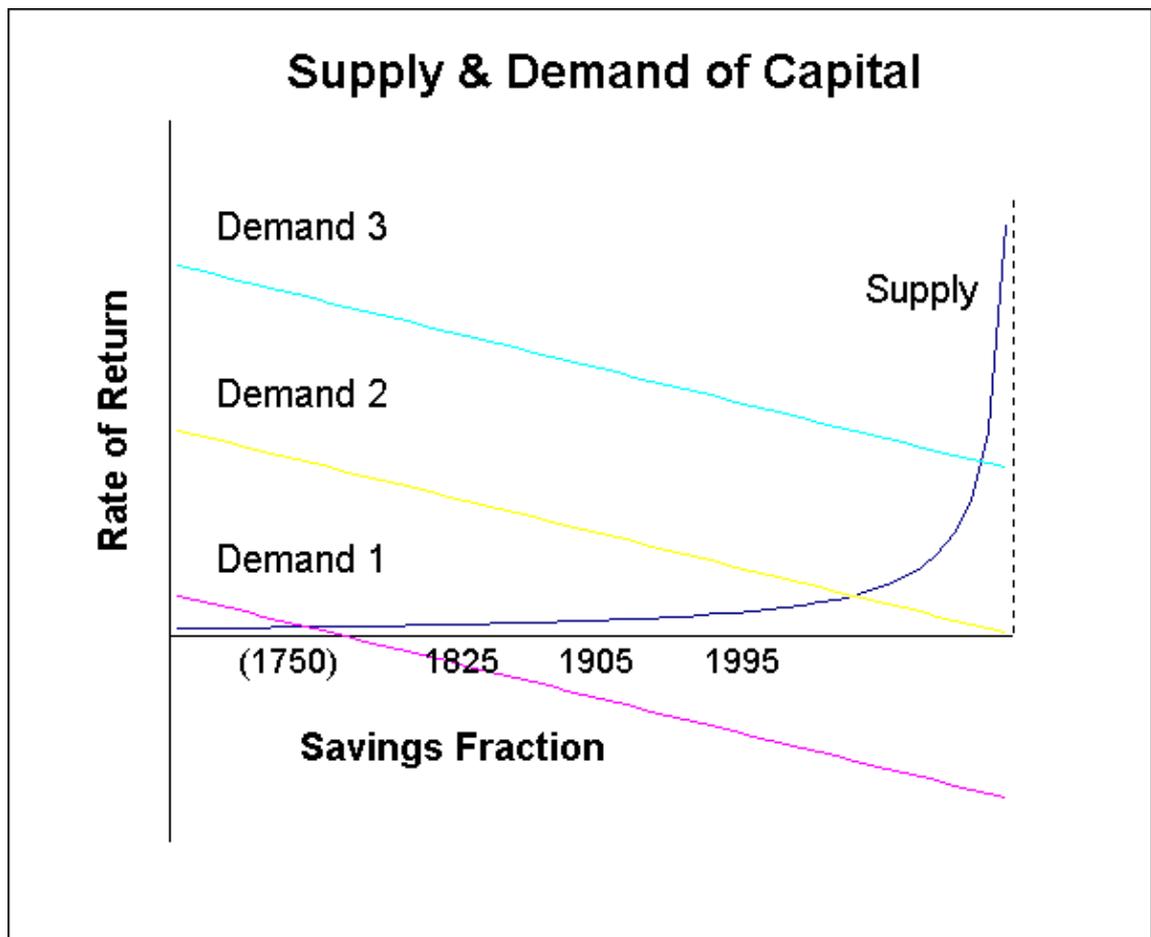
any single new technology could do this, though historical trends suggest that the accumulation of all new technologies over the century and a half might.

The supply and demand of investment capital

Growth in consumption over any extended period requires growth in what economists call "capital", which means any thing which helps make "products" people want. These terms are considered abstractly, so a music concert is a "product", and the skills of a musician are "human capital".

Capital can be used not only to make products, but also to make more capital, and the rate at which this happens is called a "rate of return". The fraction of capital devoted to making more capital is called the "savings rate", and the price people pay to rent capital for this purpose is called a "market rate of return." This is a "real rate of return" if capital is calibrated so that one more unit of capital always produces the same "amount" (or "value" or "utility") of products people like.

What determines how much people save and the interest rate they get for it? Supply and demand, naturally. Technology supplies a pool of investment projects which are physically possible, and investors demand these projects more or less depending on how productive they are. Or, looking at it in terms of investment capital, investment projects *demand* capital to be carried out, and investors *supply* capital to such projects. This view is illustrated in the following figure.



As is usual in supply and demand graphs, the x-axis is a quantity axis, here amount of capital, and the y-axis is a price axis, here a rate of return. The origin where the axes meet has no capital saved and a zero rate of return.

Supply curves slope up, so that higher returns are required to induce investors to save larger fractions of their income. Demand lines slope down, so that the more projects are undertaken, the worse their return. Projects can be thought of as lined up in order along the savings axis, with the best projects at the left and the worst projects at the right. At the point where supply and demand meet, the marginal investor is just barely willing to offer capital to get the market return offered by the marginal project.

While the demand lines shown are fictitious, the shape of the supply curve shown is true to a reasonable understanding of what investment supply looks like. This is based both on observing current human preferences, and on an understanding of what sort of preferences should have been selected for during our evolution. Moreover, the dates on the figure show roughly where the economy has been along this curve. (See the technical appendix for details.)

The supply curve eventually turns up sharply, shooting off to infinity at a particular bound. At least it does this if the population growth rate stays below twice its current value, and if investments on net benefit rather than harm non-investors (again, see the appendix). The supply curve is relatively flat, however, over a large range, at a rate of return determined mainly by the population growth rate and the degree to which our genes have taught us to discount risk and time. Since capital supply should be relatively insensitive to technology, we can hold this curve fixed in our minds as we consider how technology might change the demand for capital, i.e., the supply of investment projects.

When poor technology creates a low demand for investment capital, as with Demand 1 in the figure, the resulting market rate of return is not very sensitive to demand. In this case improving technology mainly just raises the savings rate. Demand 2 in the figure describes an economy nearing the middle of a transition between the two distinct economic modes, which may describe our world economy in roughly a century.

When technology is good enough to create a high demand for investment capital, as with Demand 3, savings becomes insensitive to demand, and the market rate of return can become sensitive to technology. In the high demand mode, growth rates can in principle be very high. Given demand curves with the right shape, growth rates might increase much faster with time in the high growth mode than they did in the low growth mode. Such rapidly increasing growth rates seems most like the imagined singularity scenarios. With other demand shapes, however, fast growth rates might increase more slowly than slow growth rates did.

Burning Up Excess Return in Investment Races

The reader may have noticed that the figure above suggests that while in the low demand mode the market rate of return stays roughly constant, the *average* rate of return of the projects actually tried might depend dramatically on the demand. After all, the top of the demand curve could in principle be very far above the low supply curve. So why couldn't growth rates also be very large in the low demand mode?

The problem with this argument is that this figure has neglected to include the time dimension, and a lack of long-term property rights in most investment projects means that returns above the market rate are burned up in a race to be first.

There are, in fact, very few long-term property rights regarding the right to undertake investment projects. Think of developing a new kind of car, colonizing the moon, developing specialized CAD software, or making a movie with a certain kind of gimmick. Each such project requires various forms of capital, such as machines or skilled labor. While in the short term only one investor may have the rights to tackle a given project, in the long term many competing investors could have positioned themselves to have this short-term opportunity.

For example, Microsoft's dominant position in PC operating systems now gives it the right to many very attractive investments. But there was once an open race to become the dominant operating system, and competitors then tried harder because of the prospect of later high returns. And when deciding whether to enter this earlier race and how hard to try, investors mainly wondered if they could get a competitive rate of

return. Similarly, while one group now has the right to make the next Batman movie sequel, there have long been open contests to create popular movie series, and popular comic strips.

Consider a typical as-yet-untried investment project, becoming more and more attractive with time as technology improves and the world market grows larger. If there wasn't much point in attempting such a project very long after other teams tried, then a race to be first should make sure the project is attempted near when investors first expect such attempts to produce a competitive rate of return. This should happen even if the project returns would be much greater if everyone waited longer. The extra value the project would have if everyone waited is burned up in the race to do it first.

Thus most of the return above the market return in our supply and demand figure above should be burned up, leaving the average return at about the market return. Thus in the low demand mode the height of the demand curve is relatively unimportant. If anticipated, a technology which makes a moderate number of investment projects much more productive may have no effect on growth rates or on rates of return.

The width of the demand curve, however, does matter in low demand mode. If technology creates many new attractive investment projects, growth rates can rise due to a rise in the savings rate. This happens by changing the rate of return offered on the marginal investment project, the project actually tried which is expected to offer the worst return. To improve the marginal project, a technology needs to have a very broad range of productive applications.

Investment Bottlenecks

A big problem with having a very broad range of applications, however, is the possibility of bottleneck resources. Most production, for consumption or investment, requires a wide variety of forms of capital. These include land, raw materials, energy, various sorts specialized labor and specialized machines, information sources, places to dispose of waste, access to channels of distribution and advertising, legal adjudication, and regulatory approval. A technology that dramatically improves the productivity of one of these typically gives only a small improvement to the final output. This is the concept of diminishing returns in production. If the rate of improvement in any one input consistently lags behind the rest, that input then becomes a bottleneck limiting the growth rate for the entire process.

For example, computer hardware may be getting cheap very fast, but one cannot now as quickly reduce the costs of training professionals who understand both computers and application areas well. Thus such training becomes a bottleneck limiting the contribution of computers to the larger economy. Similarly, even having the cost of producing electricity drop to near zero would only have a minor effect on the economy, at least in the short term. Most of the price consumers pay for electricity is to transport it, not to produce it, and new ways of organizing the transportation and production of electricity would have to evolve to take more advantage of the new situation.

The notion of bottlenecks applies to timescales in particular. For example, I think the rapid growth in computing speed and communication bandwidth, together with loose notions that most of the economy must be at root computation and communication, have suggested to many that most all economic timescales, including economic doubling times, must soon rise to meet these faster rates. But this ignores the prospect of other physical processes becoming time bottlenecks, and the diminishing returns to raw computation. Note also that in the low growth mode, even changing the total investment time scale by a large factor, which implies changing the capital demand function by a large factor, need not change the market return or growth rate by very much.

It seems hard to escape the conclusion that it just takes a lot of time for the world economy to absorb even very broadly applicable technologies like the computer, especially if the criteria of interest is raising the rate of return on the marginal investment project worldwide. Thus it seems unlikely that a single new technology could quickly knock the economy into a high demand mode with very high growth rates. (For similar conclusions, see [here](#) and [here](#).)

Even if no single technology can create fast growth, the cumulative effect of all new technologies over many decades, might, still slowly push the economy up the supply curve into the high demand mode. This requires that such technologies keep creating enough productive investment opportunities to keep up with a very rapidly growing economy, that investments on average benefit non-investors, and that we don't double population growth rates. This doesn't seem easy, but it may not be impossible either.

Conclusion

If we assume that the risk and time preferences of future investors will not vary much from those which evolution has given us, then the supply curve for investment capital should stay nearly fixed while technology changes the demand curve. And since a lack of long-term property rights in investment projects creates a race to be first, the two parts of the supply curve create two distinct modes for the economy: a low growth rate mode and a high growth rate mode. Historical trends on growth rates suggest that we are slowly moving up the capital supply curve toward the high growth mode, and may reach it in roughly a century and a half.

Many have suggested that some special new technology will induce rapid growth rates much more quickly than this. This is possible in principle, but we have identified a number of conditions which such a technology must meet. Investments must on average benefit, rather than harm, non-investors. And most difficult, the special technology must have such broad and attractive enough applications across the world economy that it dramatically raises the returns on the worst investment anyone undertakes. To do this, such a technology must dramatically improve the productivity of almost all forms of capital, not just a few. And this must be true even though, in a race to be first, each investment project is started as soon as it seemed marginally attractive.

These conditions do not appear to have been met by any of the very diverse range of technologies that have yet appeared. It is possible that such conditions may be met by some future technology, but a persuasive argument for this case should explain, in standard economic terminology, why we should expect this technology to meet these conditions.

References

Barro, R., Sala-I-Martin, X., *Economic Growth*, McGraw Hill, 1995.

Blume, L., Easley, D., "Evolution and Market Behavior", *Journal of Economic Theory*, 58:9-40, 1992.

Hansson, I., Stuart, C., "Malthusian Selection of Preferences," *American Economic Review*, 80(3):529-544, June 1990.

Maddison, A., *Dynamic Forces in Capitalist Development*, Oxford Univ. Press, 1991.

Price, D., *Little science, big science... and beyond*, Columbia Univ. Press, 1986.

See also [Econ Growth Web site](#).

Technical Appendix

There are many more complex models of economic growth out there, but the following simple model is complex enough to embody many results relevant to singularity speculations.

This model is one of the simplest models of *endogenous growth*, allowing for real long-term growth within the model and implicitly including all forms of capital improvement, including research and development. The model's main simplification is that it does not explicitly distinguish between various forms of capital and consumption, aggregating them instead to a single generic form.

First some notation, including standard values where applicable.

n = growth rate of world population, now $\sim 1.4\%/yr$.

g = growth rate of world per-capita consumption, now $\sim 2\%/yr$.

p = typical discount rate, $\sim 3\%/yr$ (= factor of 2 in 23 yrs).

a = typical risk-aversion, ~ 1 ($a = -c*u''/u'$ for utility $u(c)$ of consumption rate c).

r = real market rate of return (not the risk-less interest rate).

s = savings, fraction of world capital making more capital.

A = growth rate of invested capital (corrected for depreciation).

i = internality, fraction of investment return the investor gets.

We have five primary equations using these parameters.

Supply of Capital: $r = p + a * g$

Demand for Capital: $A = A(s)$

Capital Rental: $r = i * A$

Accounting: $g + n = s * A$

SUPPLY: The supply of capital for investment depends on how fast consumption is growing relative to the typical rate at which people discount the future. If people were risk-neutral, the market return would equal the typical discount rate. But for risk-averse people who plan to consume more tomorrow than today, stuff today is worth more.

DEMAND: The demand function $A(s)$ describes the investment projects that current technology makes possible. It says what the expected total (private and external) return on the worst project would be if a fraction s of total income were invested.

RENTAL: The rental price of capital depends on how fast a unit of capital can produce more capital, corrected for the fact that an investor may not internalize all the costs or benefits of a project. A project may, for example, create improvements in technology that others can copy.

ACCOUNTING: The total growth rate in consumption and capital depends on how productive capital is, and on what fraction of income is saved.

Some comments:

1. Typical utility parameter values are predicted from evolutionary selection. When trading resources for a parent now vs. for a child a generation from now, it matters that a child shares only half a parent's genes. And unit risk-aversion, which is log utility $u(c) = \log(c)$, is selected for, at least regarding shocks to all copies of a gene, such as the total market risk in the CAPM.
2. These equations allow for different people to own different amounts of capital, and for different investment projects to have different returns and internality. This is because the supply equation is a very general implication of maximizing expected discounted utility, and the internality parameter i can be considered an average over the projects tried.
3. We have neglected the probably-weak dependence of internality i on the savings rate s . Having $i < 1$ says that investment projects on net benefit non-investors, while $i > 1$ says that non-investors are on net harmed.
4. We've set the market return equal to the average return, and so are assuming no long-term property rights in projects. A project happens at the first time its return becomes competitive. Relaxing this assumption is equivalent to reducing the internality i , which then becomes more dependent on savings s .
5. $A(s)$ summarizes all technical change which creates economic growth. It does not, however, describe

changes in the growth rate. While the return to any one not-yet-started project typically rises with time as technology improves, the best projects will be done first, so the $A(s)$ function may rise or fall with time. $A(s)$ also falls if the number of attractive investments did not grow as fast as the economy did.

6. The growth rate g in the accounting equation is really the growth rate of capital, while the rate g in the demand equation is the growth rate of consumption. When savings s is constant, such as with demand $A(s)$ unchanging with time, these are the same thing. We are thus assuming that the growth rate of savings s is much less than the growth rate of consumption g .
7. We are using depreciation-corrected parameters A and s in the above equations. To have a depreciation d appear explicitly, change to raw terms $A0, s0$, where $A0 = A + d$ and $s0 = (sA+d)/A0$. The accounting and rental equations become $n+g = s0*A0 - d$ and $r = i*(A0-d)$. For non-human capital, depreciation is typically $\sim 5\%/yr$.

The first four columns of the following table shows historical estimates for the annual per-capita and population growth rates g, n for four different dates from 1750 to 1995. (Warning: the per-capita growth estimate for 1750 is very crude.) Using the above model and our standard values for preference parameters a, p , we then infer the next two columns, an annual rate of return r and a savings fraction s/i for each date. The return estimates seem roughly within reason.

Year	# nations in data	Per-capita growth rate	Popul. growth rate	Rate of Return	Savings fraction	Total return	Savings	Raw total return	Raw
		g	n	r	s/i	A	s	$A0$	$s0$
1995	100+	2.0%	1.4%	5.0%	68%	10%	34%	15%	56%
1905	29	1.3%	0.9%	4.3%	50%	9%	25%	14%	53%
1825	7	0.8%	0.5%	3.8%	34%	8%	17%	13%	50%
1750	2	0.0%	0.4%	3.0%	13%	6%	7%	11%	49%

Somewhat arbitrarily assuming internality $i=50\%$, and using the standard value of depreciation $d=5\%$, the rest of the table gives estimates for total return and savings A, s , and their raw values $A0, s0$. A standard estimate of 56% for current raw total savings roughly fits with the standard values that $\sim 20\%$ of raw savings is invested in non-human capital, which gets $\sim 1/3$ of income. Note that the savings growth rate seems to be growing at $\sim 0.3\%/yr$, which is much less than the per-capita growth rate, as we had assumed.

What fast growth requires

Assuming that preference parameters a, p don't change much, what does this model say about the possibility of an economic "singularity," that is, very large growth rates g in per-capita consumption? Such growth rates, if they persisted long, would imply vast changes in per-capita consumption in a single human generation.

The demand equation says that per-capita consumption growth g can't get very large unless the interest rate r does, and the rental equation says that the interest rate r can't get very large unless the total return A does. The accounting equation also says that total consumption growth $n+g$ can't get very large unless total return A does.

Using our equations to eliminate r and g , we get a market equation, with capital demand on the left and capital supply on the right,

$$A(s) = \frac{p - a * n}{i - a * s} .$$

For $p < a n$ and $a s < i$, this gives the functional form shown in the figure in the body of the paper. Thus the relevant savings fraction is actually s/i , savings relative to internality, rather than savings relative to total income.

The limits to fast growth appear more directly in

$$g = s * \frac{p - n}{i - a * s} .$$

Thus the only way to allow very fast growth $g \gg p$ is for $s = \sim i/a$ without $n = \sim p$. Thus since $s < 1$, we require $i < a$.

Thus an economic singularity, $g \gg p$, requires:

$$i < a, \quad s = \sim i/a, \quad \text{and} \quad A(i/a) \gg p .$$

That is, for unit (i.e., log) risk–aversion, an economic singularity requires that

1. Investment projects on net benefit, rather than hurt, non–investors.
2. Savings must be carefully balanced near the internality parameter.
3. The return expected for the worst invested–in project becomes very large.

The bottom line is that this model does allow for an economic singularity under certain circumstances. The historical increase in the savings fraction has been roughly constant with time for the last two centuries, suggesting a near 100% savings fraction near the year 2150. Our ignorance about the internality parameter is cause for concern. Thus it is possible, though not obvious, that a continuation of historical trends will result in an economic singularity of $g \gg p$ in roughly a century and a half.

[*Open peer commentary*](#)

[*Journal of Transhumanism*](#)