Dynamics of technological development in the energy sector

J. Doyne Farmer and Jessika Trancik
Santa Fe Institute
1399 Hyde Park Road
Santa Fe NM 87501

Abstract

This paper reviews the literature on trends of technological improvement, focusing on the energy sector. We discuss the extent to which past trends can be used to predict the future improvement paths of technologies. The historical trends for certain technologies, such as wind and photovoltaics, have been much more regular than those of other technologies, such as nuclear fission or natural gas. Reasons for this include different degrees of dependency on scarce resources (which is high in the case of natural gas), as well as technology improvement drivers other than cost (such as a push to increase safety in the case of nuclear fission). Data from the United States show that retail electricity prices have fluctuated over the last forty years, but with no clear increasing or decreasing trend. In contrast the cost of several renewable technologies has dropped considerably; for instance, the cost of photovoltaics has dropped by more than two orders of magnitude during that same period. A blind extrapolation of historical trends suggests that the cost to achieve parity is not prohibitive, though we stress that there are large uncertainties involved. In an effort to better understand the reasons for these trends, we review theories for the functional form of technological improvement curves and discuss how this problem can be understood in terms of portfolio theory.

I. Introduction

In order to achieve a global transition to a low-carbon energy infrastructure we will have to make a series of investment choices in both the public and private sectors. Making the right choices depends on our beliefs about the future of each possible technology. As we invest in a technology, we expect its costs to drop, but by how much? Which technologies are most likely to satisfy, at the lowest cost, requirements for low carbon emissions and other environmental criteria?

Currently this debate often occurs in the following way: Experts are queried and present technical arguments for a technology based on their own judgment, sometimes taking into account engineering extrapolations for components of a technology. Since experts usually differ in their opinions, sometimes dramatically, this places decision makers in the difficult position of having to make their own subjective judgments about who is right and who is wrong. This problem is exacerbated because each expert is most often a specialist in only one technology or a related family of technologies, and there is a tendency to champion these technologies. This hampers objective decision making.

An alternative approach uses the past performance trends of technologies to extrapolate into the future. Based on existing data it seems clear that predictions made in this way are better than random. However, it is still unclear how much better, making it difficult to know how much to rely on them and limiting their usefulness in constructing public and private investment portfolios. Nonetheless, the predictions made using this method are good enough that an awareness of the large literature on this subject should be an important part of any decision making process for technological investment.

This approach can be substantially improved by developing a more scientific understanding of the factors that drive technological improvement, and using extrapolations based on past and present performance data to systematically predict future performance. There is considerable evidence that there are patterns to technological improvement. By developing an understanding of the underlying reasons for these patterns in past data it should be possible to predict future technological performance in reliable ways, even if in some cases this simply means more accurately predicting the uncertainty of extrapolations.

In this article we present a brief review of the literature on technology performance curves and offer our opinions about what research needs to be done to have a better understanding of the patterns of technological improvement. We also comment on how this understanding can be used to make better forecasts of future technological performance. We also discuss how insight into this problem can be gained by casting it in terms portfolio theory.

This paper is organized as follows: In Section II we review the literature on technology performance curves in general and in Section III we do this specifically for the energy sector. In Section IV we discuss how performance curves can be used to make investment extrapolations, and how they can be sensitive to small changes in model parameters. In Section V we discuss the factors that influence technology performance and in Section VI we give a qualitative discussion of the differences between renewable vs. non-renewable sources of energy. In Section VII we discuss theories that attempt to explain the functional form of technology improvement curves and in Section VIII we cast the problem of technological evolution in terms of portfolio theory. Section IX presents our conclusions.

II. Background on performance curves

One of the most striking facts about technologies is that they tend to improve with time and experience. The first systematic observation of this was published by T.P. Wright in 1936 [1]. He gathered and analyzed data on the costs of airplane manufacturing in the United States, and showed that they systematically dropped as a function of the cumulative number of units produced. He proposed that the cost per unit c as a function of the cumulative number of units produced n was reasonably well described by a power law of the form:

$$c = kn^{-\alpha} \qquad (1)$$

where α and k are constants. k determines the scale of the changes and α determines the rate at which improvements occur. It is common to re-express α as a *progress ratio R*, which is defined as $R = 2^{\alpha}$. The progress ratio is the fraction by which the cost drops under a doubling of cumulative output. A progress ratio of 80%, for example, implies that if over a given period the cumulative production n doubles, the costs c are 80% of their value at the beginning of the period, i.e. they drop by 20%.

An example is given in Figure 1 for the price of the Ford Model T from 1909 – 1923 from Abernathy and Wayne [2]. The data is fit to a power law of the form of equation (1). To make this relationship clear the data are plotted on double logarithmic scale. Taking the logarithms of both sides of equation (1) gives $\log c = -\alpha \log n + \log k$, i.e. the logarithm of the cost is linear in the logarithm of the cumulative production, so that in double logarithmic scale a power law becomes a straight line. We see that for the Ford Model T there is a fairly good fit to a power law.

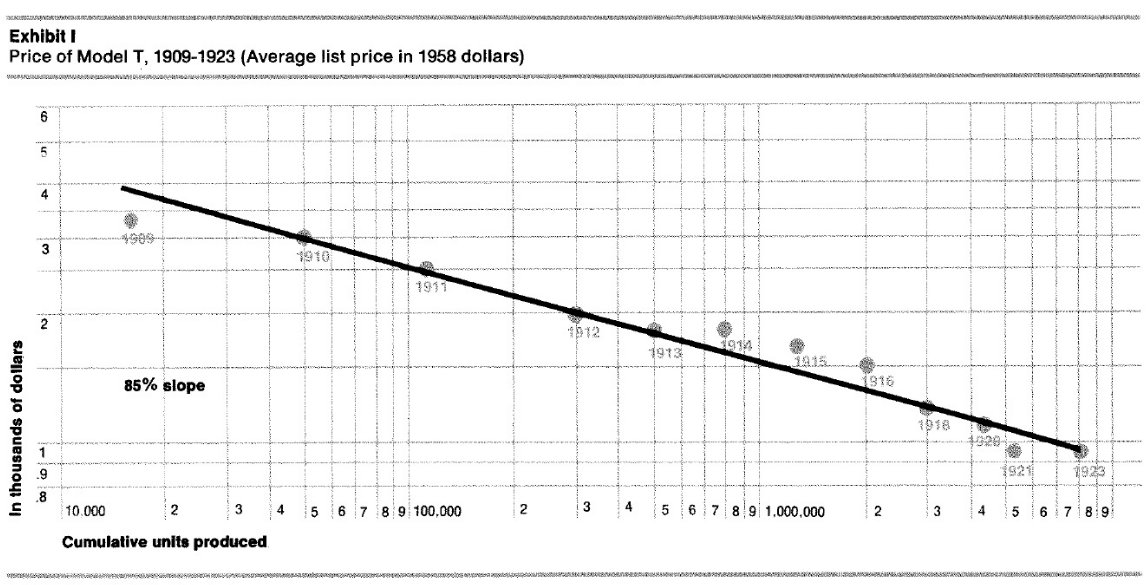

Figure 1. The price of the Ford Model T from 1909-1923[2].

Since Wright's investigation of airplanes there have been many other studies of technological improvement [1, 3-14]. Various metrics have been used for measuring both technological improvement and cumulative production. The most common measure of technological improvement is the cost of producing a single unit, though in some cases "cost" is taken to include only labor and not capital costs. In addition to becoming cheaper to manufacture, technologies get better in a variety of other ways, which complicates the question of what one should take as a measure of cumulative production. For example, if one just counts airplanes, is a small biplane equivalent to a large commercial jet liner? In such an example the right metric of production is not clear.

Studies of technological improvement have taken place on many levels, including individuals, firms, and entire industries. At the individual level, for example, the average time it takes for a lathe operator to produce a given product has been shown to decrease with experience. Many of the early studies were at the level of individual firms or even individual plants, for example, tracking the labor output required to produce a product as a function of cumulative output for that particular plant. Plots of performance vs. cumulative production are known by various names, such as learning curves, progress functions and experience curves. A learning curve or progress function typically refers to the relationship between technical change and learning from production experience at the firm level. The experience curve concept, in contrast, was developed to include all costs to manufacture and market a product at the industry level [5-7]. In this paper we introduce the more general term performance curve to include all of the above, as well as other performance metrics that may not involve cost.

Studies of performance curves at different levels show that technological costs as a function of time do not always follow power law relationships. For example, if we look at the costs to manufacture Fords other than the Model T, the story is quite different, as shown in Figure 2. This figure shows that there was a dramatic drop in cost immediately before Ford began producing the Model T, due to a change in corporate strategy to focus on producing cheaper cars. While there was a fairly steady drop in costs that followed a power law during the Model T era, this ended shortly after the demise of the Model T. Ford changed their strategy back to producing larger, more expensive cars, and manufacturing costs went up for almost the next four decades. This shows the danger of blindly extrapolating a performance curve. This is confirmed by other studies in the literature. While there are many examples of steady improvement along power laws, it is by no means a universal rule. As the Ford example illustrates, it may be valid in some regimes and not in others, and indeed when there is a major regime change one may expect discontinuous behavior. One obvious factor that can be extremely important is the objective function that the producers of a technology optimize: During the Model T era Ford concentrated on making the cheapest possible automobile. When market tastes changed circa 1925, Ford changed their focus to comfort and performance and allowed costs to rise. Thus the subsequent increase in cost probably does not reflect a decrease in manufacturing efficiency, but rather a change in the product itself.

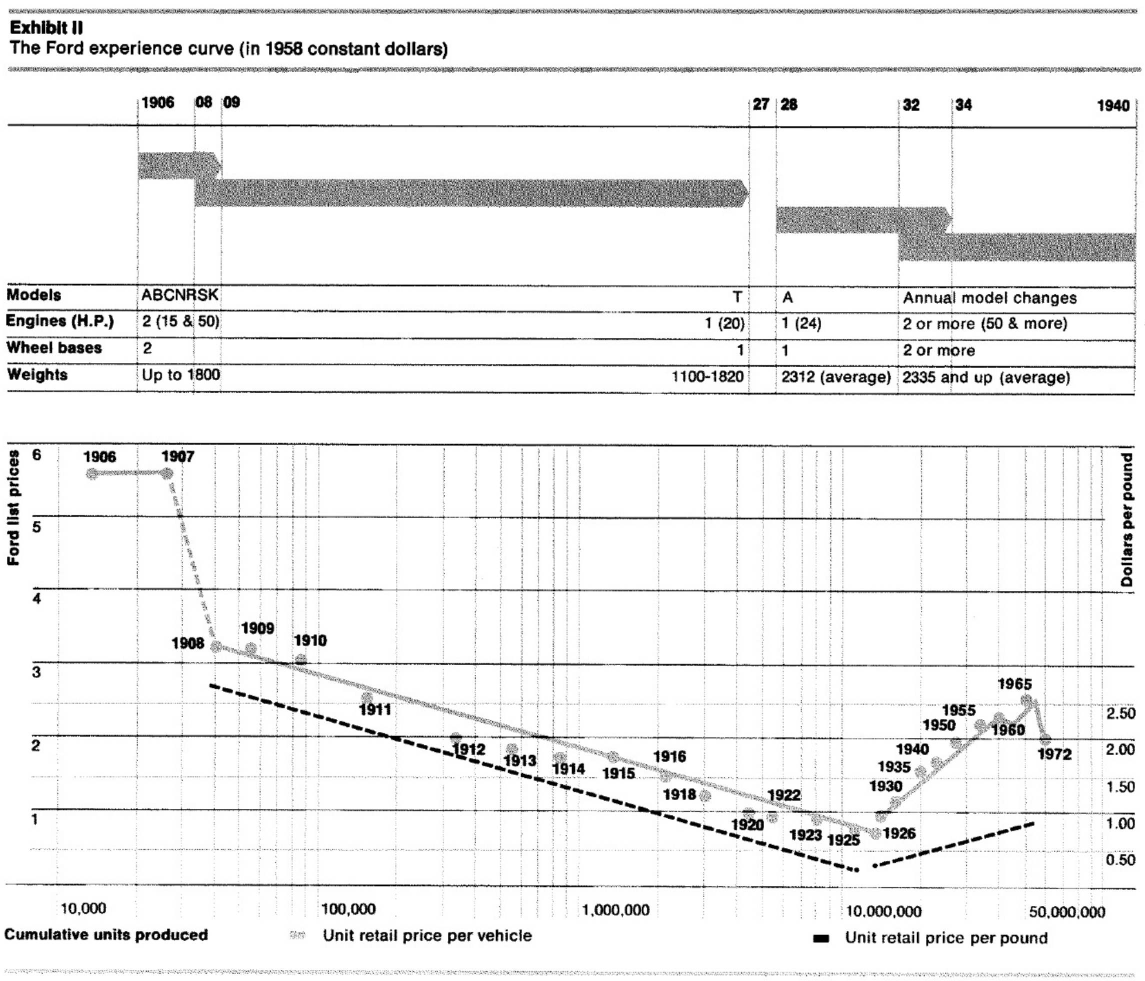

Figure 2. A study of the cost for producing Ford automobiles over a longer period shows that the power law improvement observed during the Model T period ceased to be valid in other periods due to a change in consumer tastes and corporate strategy [2], i.e. a change in the optimization function.

One should naturally ask why performance is taken to be a function of cumulative production rather than some other variable. It is generally assumed that cumulative output is not the direct cause of improvement, but rather is simply an easily measurable quantity that is correlated to other variables such as accumulated knowledge. It would be nice to have a more direct way of measuring know-how, but cumulative performance has the advantage of being straightforward to measure compared with most other factors. Trends in time are generally less consistent, with the notable exception of Moore's Law.

Even if a power law relationship is assumed to hold, the corresponding progress ratios can be highly variable. Figure 3 shows the results of a study done by Dutton and Thomas that tabulated progress ratios from 108 different cases. While the center of mass is near

80%, the approximate value originally observed by Wright, the progress ratios vary from 55% to 108%. Unfortunately, in the absence of a clear null hypothesis for the statistical process of technological improvement and more careful data analysis it is impossible to tell whether these variations are real or simply artifacts of statistical fluctuations due to short samples. However, the Ford example suggests that such variations should not be at all surprising. Even if we assume that learning always drives progress along a power law (which seems unlikely), when cost is not the sole optimization criterion, or when there are factors other than learning influencing costs, we should not expect to see a power law improvement when we look at cost alone.

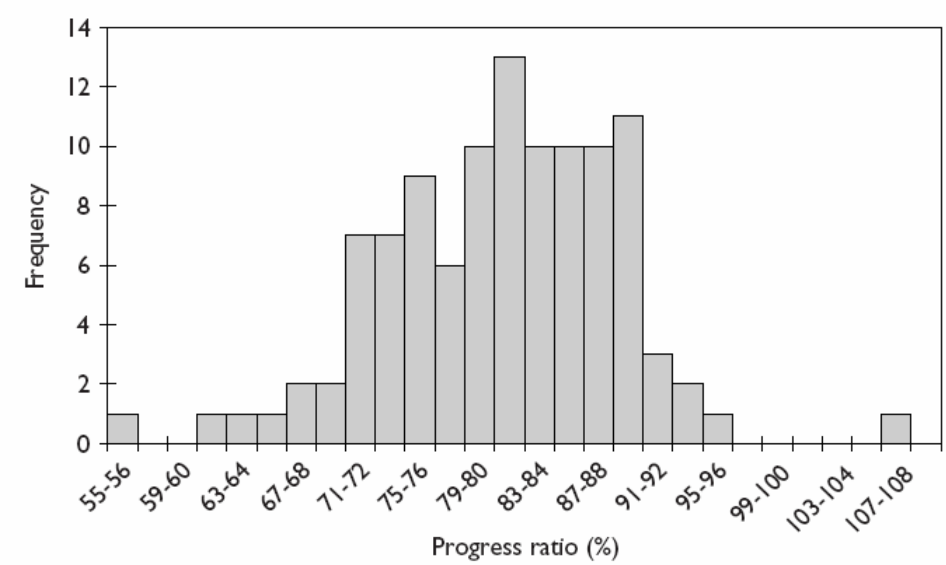

Figure 3. Progress ratios 108 cases, 22 field studies, electronics, machine tools, system components for electronic data processing, papermaking, aircraft, steel, apparel, and automobiles [7].

While performance curves are often thought to reflect only incremental improvements, depending on the level of aggregation of the technologies studied, they may also capture more radical, new discoveries. For example a performance curve for integrated circuits may include radical changes to thin-film deposition techniques, and a performance curve for electronic display screens may include fundamental advances in transparent conducting materials. The degree to which radical changes decrease the reliability of future predictions based on past performance will depend on the level of aggregation and many other factors. The case of integrated circuits (measured in terms of the number of transistors per integrated circuit) provides an excellent example of how progress in a given technology can follow a relatively smooth trajectory even in the midst of many radical innovations.

Despite the fact that there are hundreds of papers studying performance of technologies, based on the current literature it is difficult to form a clear and quantitative picture of their usefulness as a forecasting tool [4, 7, 9, 15]. No single study has systematically collected the data from all these studies and performed the careful statistical analysis that would be required to quantitatively measure their usefulness as a forecasting tool; for the best examples see [9, 15, 16]. Nonetheless, there are a few qualitative conclusions that can be drawn: (1) Technological costs tend to drop with time, and often do so following a power law relationship. (2) The improvement is most clear when costs are the main factor that is being optimized, and when the driving force is innovation rather than other factors[1]. (3) While the quality of forecasts is uncertain, it is clear that there is at least some forecasting power to extrapolating technological trends – the forecasts are much better than random. We will say more about this in the next section.

III. Performance curves in the energy sector

What can performance curves tell us about formulating rational policies and making good investments in energy technologies? At the very least, they can give us an idea of historical trends. For example, in Figure 4 we show the price of a watt of installed capacity as a function of cumulative capacity for five different energy technologies, namely photovoltaics, solar thermal, wind, U.S. nuclear power, and NOx controls. As indicated in the figure, the data span somewhat different periods for each technology, which should be borne in mind in considering the results (e.g. the solar thermal data ends in 1991 and so does not reflect more recent improvements). For energy technologies we have the advantage that, under the assumption that a watt of power is equally useful at any point in time, we can measure cumulative capacity in a consistent way (i.e. we don't have the problem encountered with airplanes of having to compare airplanes of different speeds, reliability, etc.).

A problem occurs when we try to compare technologies such as photovoltaics or wind, where the costs are almost entirely the capital costs of the initial installation, to technologies such as coal or natural gas fired power plants, where there are significant ongoing fuel costs. All the technologies shown in Figure 4 are dominated by capital costs, and so it is fairly reasonable to focus on the capital costs per unit of peak power production capacity. Of course, in reality the amount of energy a working photovoltaic installation will generate over time depends on many other factors, such as the average amount of sunlight for the location in which it is installed. A similar consideration applies to a fossil fuel plant: The cost of the plant itself does not include the cost of the fossil fuel, which will vary over time as the price of the fuel fluctuates, and may be a significant component of the total cost of the power. One must also address the fact that

[1] This is at least the traditional thinking, but not all evidence supports it. For example, photovoltaics are a good example of a technology that has followed a clear power law improvement curve. Nonetheless, Nemet [11] has concluded that the main drivers of improvement were economies of scale, efficiency improvements, and material costs. See our remarks in Section VII, where we suggest that economies of scale can also obey Wright's Law.

the need for electricity varies depending on time of day and weather conditions, i.e. power generated on a sunny day when air conditioning demand is at its peak may be more valuable than power generated at night, and as the mixture of different generating sources changes, the demand for power conditioned on other factors, such as weather, may change. To make a proper comparison one must estimate all costs of actually

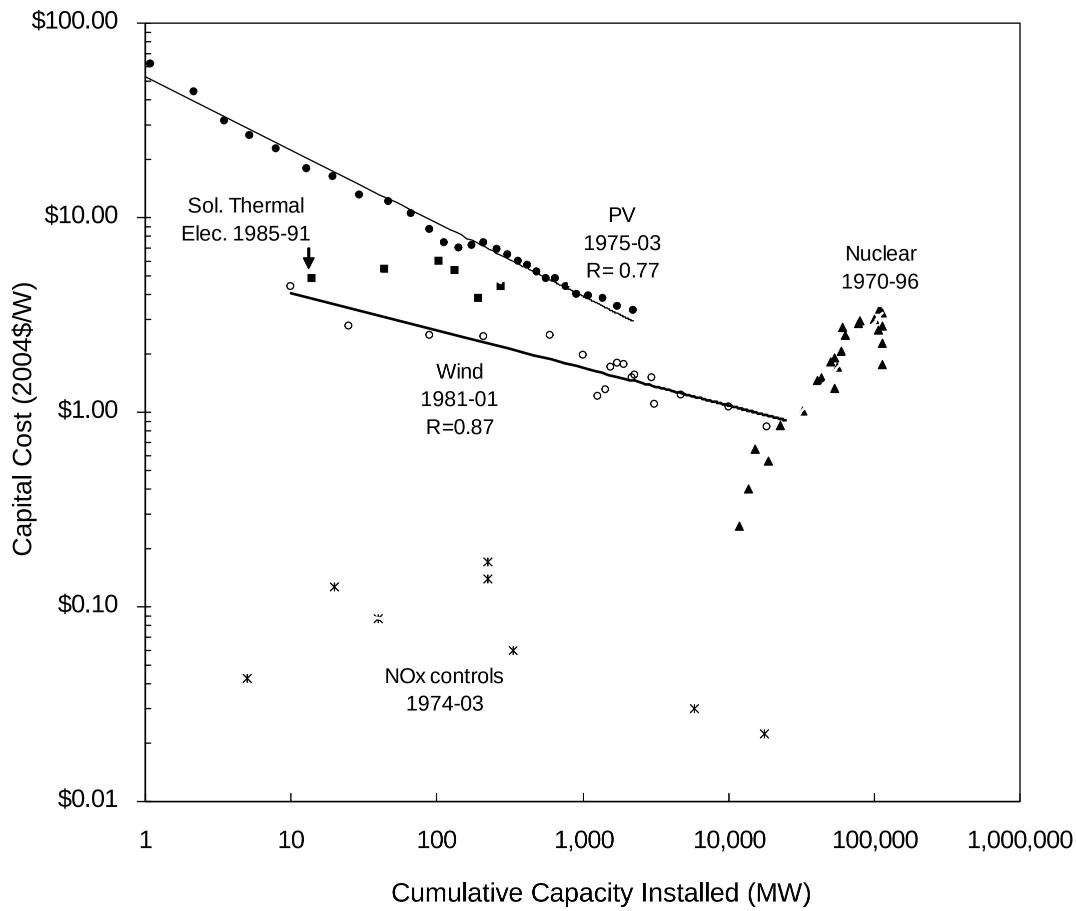

Figure 4. *Cost vs. cumulative capacity for electricity generation technologies*. Cost is measured in terms of price per watt of peak generating capacity, and cumulative capacity is measured in terms of megawatts of installed capacity. Data is given for wind, PV = photovoltaic, solar thermal, and nuclear power. NOx refers to the additional cost of installing pollution controls that reduce nitrogen oxide to a targeted regulatory level. This data is from selective catalytic reduction (SCR) units installed on natural gas-fired power plants. See [17]. The NOx and nuclear power data are from the US, and the other data sets are global. Several of the data sets are fairly well described by a power law, and the progress ratio (R) is shown for these.

generating the power, including maintenance, lifetime, transmission, and ability to deliver during times of peak demand. This can obviously get too complicated to capture with a single average measure. Comparisons between different technologies are typically reduced to a cost per kilowatt-hour of the energy actually generated, but one should keep the complexities discussed above in mind. In any case when making an extrapolation it can be useful to decompose the total costs into their components. This is discussed further in section VI.

Figure 4 indicates that four of the five technologies have tended to improve with time throughout the period of the study. Photovoltaics and wind power improved fairly steadily. Both yield reasonable fits to a power law, though from the plot one can see that there is some fluctuation in slope. For solar thermal we have very little data so it is not surprising that the trend is not obvious. The cost of NOx controls, in contrast, has a noisy trajectory, and although there was a tendency for improvement it is difficult to say what kind of model might be appropriate.

A striking counter example is U.S. nuclear power, which increased in price by almost an order of magnitude over the course of roughly two decades. Although we cannot state this with certainty, we believe this is primarily due to the increased focus on power plant safety during this period, which was significantly driven by tighter regulatory control[2].

As in the case of Ford, when cost ceases to be the dominant objective function and optimization occurs in other dimensions, there is no reason to expect that prices will decrease according to a power law, or even to expect that they will improve at all. This message is of course very relevant for the energy sector as we move forward. When we impose penalties for carbon emissions, thereby changing the objective function, we should expect that fossil fuel technologies are likely to increase in price, whereas technologies that do not produce much carbon anyway will be relatively unaffected.

We have not included fossil fuels in Figure 4 because we are still compiling the data to reduce the analysis to a similar form, and because there are problems in measuring cumulative production in a way that allows a reasonable comparison. We will return later to discuss fossil fuels in more detail.

IV. When do alternative technologies become competitive?

If we believe that there is a reliable trend of technological improvement, then it is possible to use this trend to extrapolate the level of investment that would be required to reach break-even at a given price level. This of course assumes that we know in advance

[2] One piece of evidence for this is the suggestion that trends in European nuclear power plants are quite different, and do not show the same increase in cost. However reliable cost data for Europe is difficult to obtain.

what a competitive price level is likely to be – the dominant technologies that set the price level are likely to themselves be evolving.

In fact, trends in retail electricity rates in the U.S. suggest that the evolution of the prices set by the dominant technologies has been fairly slow in recent years. In Figure 5 we show a comparison of the costs of electricity generation by photovoltaics to U.S. retail electricity rates during the same period.

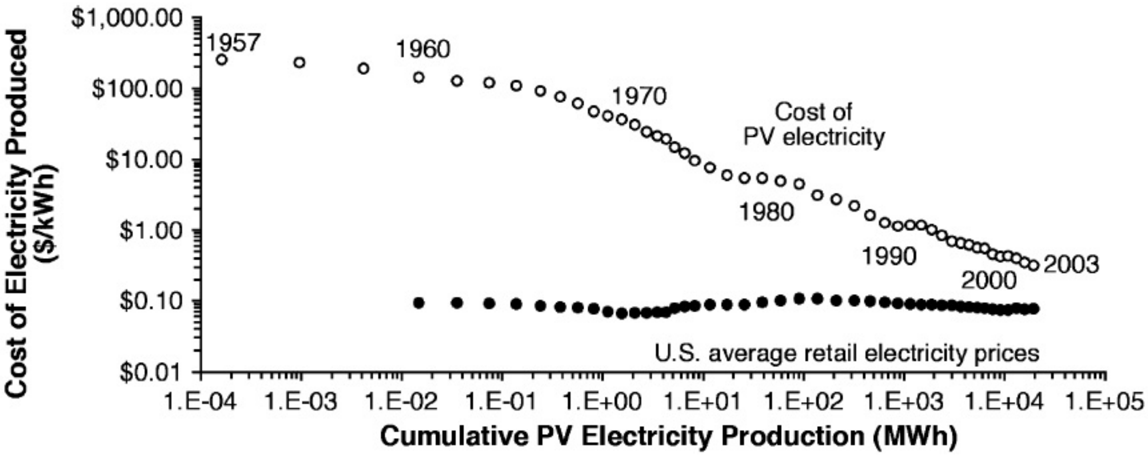

Figure 5. Aggregate cost of electricity in the US, versus cost of PV electricity [11].

We see that retail rates have fluctuated (bear in mind the logarithmic scale), the overall trend has remained relatively constant at about 10 cents per kilowatt-hour, while in the same period while the cost of photovoltaic electricity has dropped by more than two orders of magnitude. This suggests that if we simply keep building more photovoltaics we should shortly reach the cross-over point at which photovoltaic prices become competitive with existing dominant power generation methods. Assuming for the moment that this assumption is true, at what point would that occur?

In Table 1 we present a comparison of the cumulative production levels that would need to be reached and the corresponding cost of reaching this level, assuming a target of 1 $/W. The cost to reach the cross-over point is found by extrapolating the curve to the target and taking the integral up to that point. The answer is very sensitive to the progress ratio. In-sample the progress ratio has tended to be roughly 80%, but since the out of sample value is uncertain, we have included a range of plausible values from 70 – 90%.

Progress ratio	70%	75%	80%	85%	90%
Breakeven cumulative production (GW)	23	48	148	957	39700
Cost of reaching 1.0 $/W ($ billion)	37	74	211	1240	46800

Table 1. The cumulative production of electricity (in gigawatts of peak capacity) needed to break even with a residential cost of 10 cents per kilowatt hour, and the cumulative cost of achieving this level of production, assuming a simple trend extrapolation. These are given as a function of the progress ratio that is assumed to hold during the out of sample period. Costs are given in dollars as of the year 2000 [12].

As is apparent from the table, the results are highly sensitive to the progress ratio. In the most optimistic case where we assume a progress ratio of 70% the investment required is only 37 billion dollars, whereas for a 90% progress ratio it is 46.8 trillion dollars. In the case that the present rate of 80% continues it is about 200 billion. While this is a large number, it is a small fraction of the yearly tax revenue of the United States. Note that this is not really “investment” in the usual sense – we are simply assuming that installations are made to increase the capacity, and that the pattern of improvement follows a power law with the given progress ratio. That is, we assume that if the rate of installations per unit time were to increase, e.g. due to a crash program to install photovoltaics, all the factors that cause improvements would have the same correlations with cumulative capacity that they would otherwise, so that Wright’s Law continues to hold. To consider an analogous situation, production of airplanes and other war machinery under crash programs during World War II form some of the best examples of Wright’s Law [18], are were roughly in line with earlier trends for cumulative production, suggesting that such an assumption is not unreasonable.

To illustrate the sensitivity of historical trends, in Figure 6 we show two different global surveys of PV prices. Fitting a power law to these two data sets separately results in a progress ratio of between 0.74 and 0.83 and an investment (area under the curve) that varies by more than an order of magnitude. The data shown in Figure 4 takes an average of these data sets. This highlights the importance of obtaining data sets that are as comprehensive as possible. Relying on time series data that do not extend over a long enough time period can also cause problems.

Another challenge is that data is most commonly obtainable in terms of price rather than costs. This adds extra difficulty in interpreting trends since profit margins may change [5]. This is likely to be less of a problem for long data sets in competitive markets, where we can average profit margins over time and where we expect the margins to remain fairly constant. For cases where these conditions do not hold, it is important to consider this additional source of uncertainty in projections.

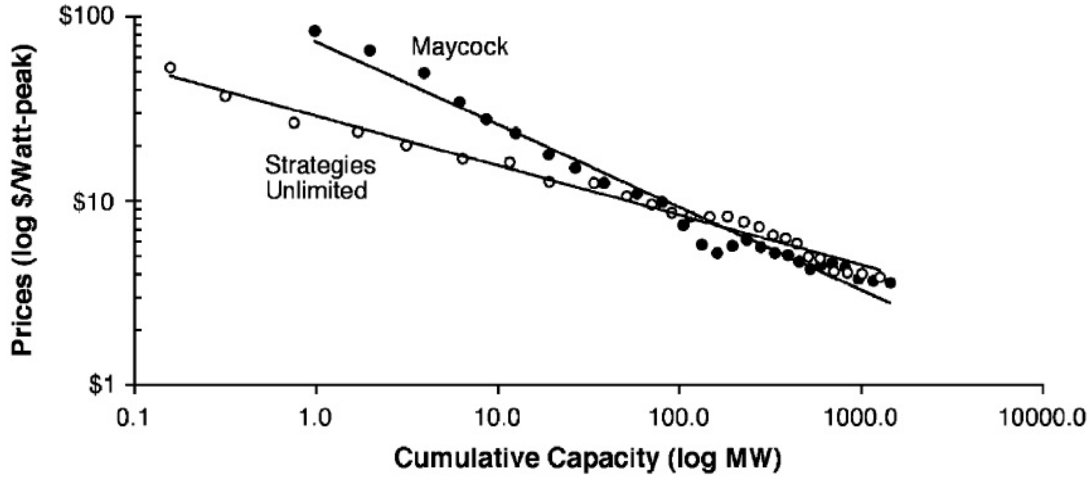

Figure 6. Illustration of two different historical data sets for PV performance curves, with progress ratios that vary from 0.74 (Maycock) to 0.83 (Strategies Unlimited). [11]

The extrapolation exercise discussed above illustrates several important points. First, it shows that the cost of transitioning to a new technology depends critically on the rate at which that technology improves. Some technologies have shown fairly steady trends of improvement in the past, and a multitude of historical examples suggest that it is likely that they will continue these trends in the future. Of course, we can never be sure of this, which illustrates the need to maintain a portfolio of diversified investments, as discussed in Section VIII. It also motivates obtaining a better understanding of the factors that drive technological improvement, a subject that we feel has received inadequate attention.

Performance curve projections can have significant effects on, for example, cost optimization models that attempt to determine the minimum cost to stabilize CO_2 concentrations in the atmosphere at a given level. These cost estimates are in turn used to determine acceptable levels for carbon regulations such as caps on emissions. This highlights the importance of making reliable forecasts. In the next section we outline steps one can take in curve extrapolation and decomposition of technologies to increase the reliability of forecasts.

V. What factors does technology performance depend on?

The reasons that technologies improve have been the subject of considerable debate [7, 19]. One approach to addressing this question is to understand the effect of individual sources of improvement on the technology as a whole. This can be done by decomposing technologies into the processes that drive improvement, which we call process decomposition, or by decomposing them into their physical components, which we call input decomposition.

Process decomposition separates the means by which technologies improve or become more economical into components that are often described as economies of scale, learning by doing and learning by using. *Economies of scale* occur when a process has fixed, scale independent costs that become smaller in relative terms as scale increases. *Learning by doing* commonly refers to improvements that occur as people perform a procedure over and over again [20], and *learning by using* refers to improvements generated by feedback from users or as a result of users changing the way they operate a device [21, 22]. And of course, although much harder to characterize, there are the bursts of creativity and originality that give rise to new technologies and radical transformations [23, 24]. Since all the latter source of improvements can be hard to distinguish it can be useful to lump all performance improvements that include learning or creativity in any of their forms under the general heading of *innovation*, which thus includes all sources of improvement except economies of scale. An example of a measure of innovation is shown in Figure 7, where the efficiency of several thin-film cells is plotted as a function of time. Similarly in Figure 8 the carbon intensity of electricity generated via coal combustion is shown as a decreasing trend, primarily influenced by an increase in the efficiency of coal-fired power plants.

Decomposition of inputs is another tool to study performance improvement. Input decomposition refers to a study of the physical characteristics of a given technology. For example we may want to characterize a technology based on its unit scale (defined as the investment required to produce one example of a technology) and degrees of freedom in a

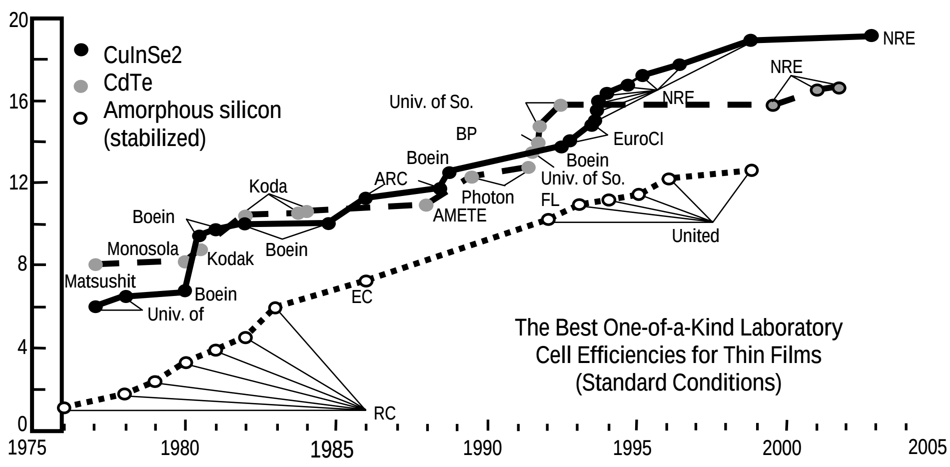

Figure 7. The change over time in laboratory cell efficiency for several thin-film photovoltaic cells [25].

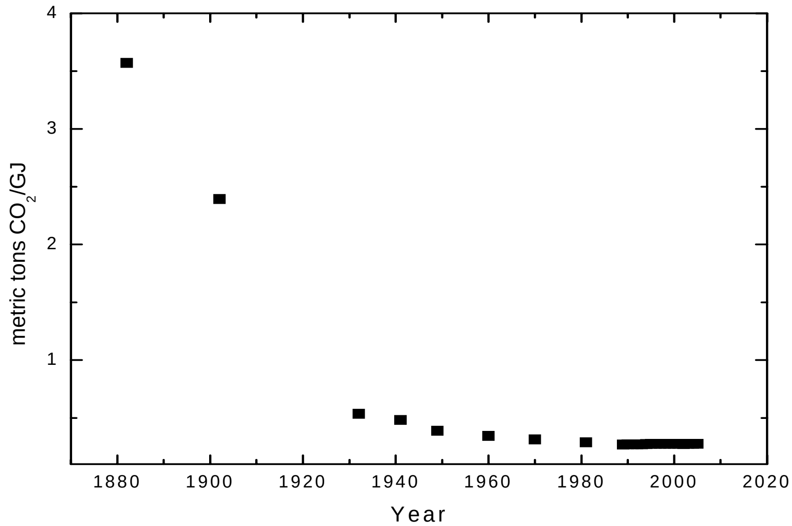

Figure 8. The carbon intensity of coal electricity as a function of time. This is based on US data from the Energy Information Administration.

device (defined as the number of modular parts that make up the whole). Input decomposition can also refer to dividing a technology into its parts and studying each of the parts separately as technologies unto themselves. For example, the inputs for a photovoltaic cell are the light absorbing materials, the absorber partners, the top and bottom contacts, barriers to the environment, circuit elements, a mounting scheme, and all of the processing equipment required to make each of the above [25]. Cells themselves are assembled into arrays and systems of arrays. Thinking in terms of input decomposition takes advantage of the fact that technologies are inherently recursive, i.e. complex technologies are often built out of other simpler technologies [26-29].

By constructing performance curves for individual processes and inputs and using multivariate forecasting it may be possible to improve the reliability of forecasts, or at least gain a more accurate estimate of uncertainty. This approach allows one to ask the following questions: What is the relative importance of economies of scale versus innovation? Are performance improvements based on economies of scale more predictable than those based on innovation, as suggested in [11, 30]? Does improvement depend more on internal factors, such as unit scale (e.g. transistors vs. nuclear power plants), on complexity of design, or on external factors, such as level of R&D investment? Studies of this kind have already begun to yield results. For example a recent study of photovoltaics, which uses a combination of process and input decomposition, suggests that the main factors contributing to the decrease in photovoltaic module costs were economies of scale, efficiency improvements and silicon cost [11].

VI. Performance curves for renewable vs. non-renewable sources of energy

In comparing performance curves for technologies based on renewable and non-renewable resources there are several important differences to take into account. One has already been mentioned in the paragraphs above, namely that because of the different nature of cost contributions (namely one-time capital costs vs. recurring fuel costs) it is important to make comparisons in terms of both peak power costs and average energy generation cost per unit time. An additional consideration is that in the case of a non-renewable resource the fuel costs will follow their own performance curve which will depend on the technology for extracting the fuel as well as price fluctuations driven by variations in supply and demand. The availability of a fuel is in turn influenced by the overall size of the resource deposits and their geographical distribution [31]. As more fuel is extracted the quality of fuel deposits declines, so that in absolute terms the fuel becomes more difficult to extract, and all else being equal fuel extraction costs will increase. For non-renewable resources this resource scarcity curve competes against the technology improvement curve for the technologies used for resource extraction. The point where resource scarcity becomes dominant depends on the total size of deposits; for example, for coal we expect that the point at which scarcity significantly drives costs up is further in the future than it is for natural gas.

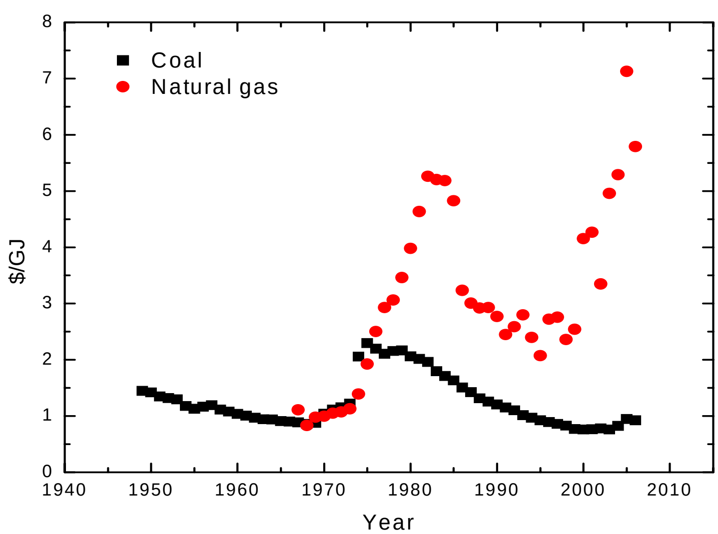

Figure 9. Cost of coal and natural gas in the U.S. in 2006$ per unit of usable energy extracted. These are actually prices rather than costs, i.e. they include profit margins, but we refer to them as ‘costs’ because they are essentially cost components from the point of view of electricity generation.

In Figure 9 one can see that the price of coal in the United States has fluctuated significantly during the time period 1940-2007. This had a significant effect on the cost of coal-based electricity. Figure 9 also shows the cost of natural gas, which has fluctuated and where the cost today is several times higher than the starting costs. The fluctuations themselves are important because they decrease the certainty of performance curve extrapolations. We are currently beginning to investigate whether there are systematic trends in the cost of these fuels by comparing empirical data to a basic null model where the price of oil, natural gas, or coal follows a random walk.

In summary, three factors influencing the cost of fossil fuel based energy are (1) technological improvement in the conversion technology driving the cost down, (2) technological improvement in the fuel extraction technology driving the cost down, and (3) progression from extraction of high quality deposits to low quality deposits driving the cost up. For the case of coal, the historical trend does not clearly show that we have a reached a point in which the cost of the fuel is trending upwards, whereas in the case of natural gas we may already be seeing indications of this. It is clear that for fossil fuels there are significant fluctuations in the cost of the resource that affect the reliability of performance curve projections. In the case of renewable technologies where the resource is essentially free, such as solar and wind, the performance curves will depend primarily on factor (1) above. There are several exceptions to this in cases where rare materials are used in manufacturing (such as the platinum group metals) and where there is a temporary mismatch in supply and demand (such as in the recent case of silicon based solar cells). However, for technologies that are dominated by installation costs, these effects are at least known at the time of manufacturing.

VII. Theories for Wright's Law

Wright originally postulated that technologies should improve as a power law in cumulative production according to equation (1). While this hypothesis is clearly violated in many cases, there are nonetheless enough examples where this seems to be approximately true that it is interesting to ask why this particular functional form might be special.

The literature on this question has focused on the idea that the process of innovation can be viewed as a search through a space of possible technologies; for reviews see [32, 33]. Muth made an extremely simple model by simply assuming a random search, in which new technologies are sampled without any reference to previous technologies, and replace the current best technology only if their cost is lower. Using arguments from extreme value theory he showed that if the space of possible technologies is reasonably behaved this will give rise to Wright's Law. However, the idea that the search is independent of previously known technologies is very unrealistic, and the values of the progress ratio that are obtained under reasonable assumptions about the search space are much too low to be plausible. A more sophisticated theory was given by Auerswald et al., who viewed the production of a technology in terms of a recipe for its manufacture, consisting of N steps. Innovations randomly modify one of the steps in the recipe. They allow for the possibility that each step in the recipe is coupled to M other steps, so that when one step is modified the other steps are automatically affected. Through simulations they show that under certain circumstances it is possible to obtain an approximate power law improvement over a finite range. One of the key assumptions is that M/N has the right intermediate value, i.e. M > 1 but also M/N << 1, so the technology is complex but not *too* complex.

An alternative idea about how one obtains a power law is due to Sahal [34]. His observation is that one automatically obtains a power law from two competing exponentials. If a technology proliferates exponentially in time proportional to $dn/dt \sim \exp(at)$, and its cost is drops at an exponential rate proportional to $c \sim \exp(-bt)$, then $c(n) = kn^{-\alpha}$, where $\alpha = b/a$. Exponential growth in cumulative numbers or exponential decrease in cost will occur whenever both processes can be approximated by first order linear differential equations, i.e. when they follow dynamics of the form dy/dt ~ y. This is true even if neither process is strictly exponential providing there exists some coordinate transformation z that makes each of them exponential, i.e. $z(c) = \exp(-bt)$ and $z(n) = \exp(at)$, there will be a power law relation between c and n. Of course, this only complicates the story, as one must explain why the individual processes $z(c)$ *and* $z(n)$ are exponentials. There are numerous comments in the literature that time trends are much less reliable than those based on cumulative production, which casts some doubt on this idea.

The motivation that search is the driving force behind Wright's Law comes from the observation that there are many examples of learning by doing in which there is little or no capital investment – where the sole source of improvement is alterations in the procedure for making something. A famous example is the Horndal Iron Works [4]. Most examples in the literature are not so simple, and involve both learning by doing and economies of scale. For example, this is true for Ford motor during the era of the Model T: The lowering of cost is likely partially due to more efficient manufacturing procedures, such as improved assembly line methods, and partially due to the fact that Ford sold an ever-increasing number of automobiles and so benefited from economies of scale in their production process.

Economies of scale can also give power laws. For example, consider the simplest case in which the cost for producing a single unit is the same as the cost for producing n units. In this case the cost per unit for n units is $c = K/n$, so the relationship between cumulative production and cost per unit is a power law with exponent $\alpha = 1$, corresponding to a progress ratio of 50%. This number is intriguing since it is also roughly the lower bound on progress ratios observed in Figure 3. In the more general case where production of additional units is cheaper than the first unit, but still not free, it is not obvious why the cost per unit would drop as a power law rather than some other functional form. But it is certainly plausible that this would happen, and in this case we would expect $\alpha < 1$. In any case that power laws can be generated either by learning or by economies of scale.

Another intriguing idea is that the form of technological improvement may come about because of the nature of technologies themselves. Technologies have been shown to be recursively built out of inputs which consist of other technologies [27-29], and improvements in technologies typically come about via recombination of existing technologies. These ideas can be extended to systems of technologies as follows: An ecology of related technologies can be visualized as a graph in which each technology is a node with directed links connecting input and output technologies. If an input improves its outputs also improve. These improvements are then transmitted to other outputs, and so on. If the graph has loops, the effects can cycle back to their origin, amplifying the original improvement. This is analogous to the chemical phenomenon of autocatalysis. An example is the interaction of computers with their inputs: as computers improve, our ability to model semi-conductor devices improves, which improves computer components, which improves computers.

For any fixed graph one can model the performance of a family of coupled technologies in terms of a set of coupled differential equations whose variables represent the information at the nodes and whose nonzero interaction terms represent the links. For a minimal model the variables at each node are the performance and prevalence of the technology. As new nodes are added, representing new discoveries, the set of differential equations changes, which in turn may alter the pattern of addition of nodes and links. This approach to modeling is called *metadynamics*, and was originally used in studies of autocatalytic networks in chemistry [35-41]. In simple examples such dynamics can give rise to nonlinear equations with power law solutions.

As already mentioned, an important determinant of the innovation component of technology improvement may be the ratio of unit scale to R&D investment. The scale is defined as the cost of the smallest modular unit that can perform a given function. If the scale is small, then it is possible to make many learning steps with a given level of R&D investment, so progress should tend to occur more rapidly, i.e. the progress ratio may be smaller [42].

VIII. Technological investment viewed from the perspective of portfolio theory

Technology investment choices modeled as portfolios. The problem of investment in technologies can be cast in terms of a problem in dynamic portfolio allocation. The portfolio problem in this context is that of assigning investment weights for a group of competing technologies that perform a common task such as energy generation. For private investment the purpose of forming a portfolio is to optimize returns under risk constraints, but for public investment the goal is to maximize the probability of achieving a socially desirable outcome such as cheap, carbon free energy. The fact that the cost of a given technology tends to drop as more units are produced implies that there is nonlinear feedback between the increasing a portfolio weight and the improving the performance of the technology corresponding to that weight: The unit cost that a technology can achieve depends on how much investment it receives. The resulting nonlinear feedback complicates the portfolio problem and means that a bad choice at the outset can cause lock-in to a suboptimal technology [43]. Because of the feedback between investment and performance, analytic solutions are hard to find and simulation (e.g. dynamic programming) is necessary [44, 45]. While the classic Markowitz portfolio problem is a simple exercise in variational calculus [46-51], in this context it becomes a stochastic nonlinear dynamical system, which is a much less tractable problem. At this point very little is known.

The properties of the optimal portfolio depend on the accuracy of technology performance forecasts and involve a tradeoff between diversification and concentration. At one extreme, the optimal solution is to fully concentrate investment in a single technology. This occurs, for example, when the parameters α_i and k_i for each technology i are known with certainty – there is no reason not to just pick the best technology and invest all resources in it. The more interesting and more realistic case occurs when the parameters are diverse and uncertain. In this case one needs to make a trade-off between diversification and concentration. Too much diversification is bad, diluting individual investments so that no technologies make substantial progress. Too much concentration is also bad, as it is likely to result in lock-in to a poor choice. So far there has been surprisingly little work on this problem (see reference [44] for the current state of the art). There is currently no qualitative theory for understanding this trade-off, and no quantitative simulations of real problems. The optimal portfolio will depend strongly on the distribution of the parameters α and k, their correlations across related technologies, and their predictability through time.

What is clear is that the pressure to diversify in technology portfolios is less obvious than it is in the classic Markowitz portfolio theory currently used in financial markets. In financial markets the differences in future returns that can be reliably forecast are small. This means that pressure to decrease risk will cause well-formed financial portfolios to be very diverse. For technologies, in contrast, since investing tends to cause improvement, there is a strong countervailing pressure to concentrate rather than diversify. A well-formed technology portfolio strikes a balance between diversification and concentration. Where the correct balance lies is not clearly understood even at a qualitative level.

The portfolio problem also depends on risk aversion and time discounting of utility. The proper functional form for time discounting is far from obvious [52]. Exponential discounting in time strongly weights short-term performance and amplifies lock-in. Behavioral studies show that people actually use hyperbolic discounting [53], which strictly speaking means that utility decays as a power law in time. Hyperbolic discounting may be caused by evolutionary selection with uncertain payoffs [54-56]. Recent work has shown that hyperbolic discounting can be rational when discount rates are uncertain [57]. Portfolio optimization over a discounting function that decays as a power law can give dramatically different results than for an exponential.

The problems of technology evolution and portfolio construction are closely related. As a technology evolves, investors are implicitly solving a portfolio problem, even if they do so in a non-optimal manner. Technological evolution is the outcome of a joint process in which public and private investors supply capital for R&D and the manufacture of new technologies, and managers, engineers, and workers create new technologies. Thus, we believe the problem of portfolio construction is intimately interwoven with that of technological evolution, and that it can be very useful to think about technology improvement from this point of view.

Portfolio theory as a way to think about the effectiveness of incentives. Portfolio theory can also be used to study different public policies for carbon reduction. We distinguish (1) reshaping incentives (e.g. carbon taxes or cap and trade), (2) public R&D, and (3) other regulatory market transformation programs, such as renewable portfolio standards. (1) From the point of view of portfolio theory, reshaping incentives amounts to changing the optimization function for investment. In the absence of incentives, cost and carbon generation are decoupled, and there is no reason for private investors to be concerned with carbon. Incentives couple these, reshaping the optimal portfolio for private investors. (2) Public R&D can directly alter a technology's position on a performance curve. Unlike portfolio optimization for a private firm, which is based on increasing marginal profits and tends to foster cost-benefit optimization on a short time scale, public R& D can be directed at achieving targeted long-range goals. This is important since private firms cannot fully appropriate the benefits of R&D investments, and hence tend to under-invest in R&D. (3) From the point of view of portfolio theory, renewable portfolio standards amount to constraining portfolio weights *a priori*. Weight constraints are a commonly used technique in finance to improve out-of-sample performance [47, 51]; in the public policy setting they can also be used to enforce other properties of the solution, such as lowering carbon generation.

IX. Conclusions

In this paper we have highlighted the sizable literature on performance curves and discussed the remaining open questions that need to be answered in order to make predictions based on performance curves realize their full potential. Understanding the uncertainty associated with curve extrapolations is critical because of the high sensitivity of policy implications and investment decisions to assumptions about performance curve functional form and parameters. There are several methods one can adopt to improve forecast reliability, such as retro-casting and ensuring the long length and quality of data sets. We have also outlined more sophisticated methods for making forecasts such as process and input decomposition.

Fossil fuel energy costs follow a complicated trajectory because they are influenced both by trends relating to resource scarcity and those relating to technology improvement. Technology improvement drives resource costs down, but the finite nature of deposits ultimately drives them up. Based on trends from recent years it is not clear which is currently dominating, e.g. Figure 5 shows how even though U.S. electricity prices have fluctuated, the overall trend has remained surprisingly constant for forty years, suggesting that technological improvement has not been sufficient to drive prices down by a significant amount. During the same period costs of photovoltaics and wind decreased dramatically. Extrapolations suggest that if these trends continue as they have in the past, the costs of reaching parity between photovoltaics and current electricity prices are on the order of $200 billion, which is for example comparable to less than a year of the total U.S. expenditure on the Iraq war. We stress, however, that without a deeper understanding of what is driving trends, such forecasts remain highly uncertain.

To stress the value of projections based on historical trends it is useful to consider the case of new technologies, where there is no historical data to extrapolate. While predictions based on the past history of a technology may be uncertain, they provide an important extra piece of evidence. When we consider entirely new energy technologies, such as carbon capture and storage, we should bear in mind that the future of such a technology is likely to be less certain than that of technologies with track records of steady improvement. Obviously this doesn't mean that we should not explore new technologies; it just means that we should bear in mind that all else being equal they have additional risk.

In the final section we have described the unique challenges in formulating an optimal portfolio strategy for new energy technologies, which improve in response to increasing investment. A better understanding of the functional forms of performance curves, the underlying reasons for these curves, and the uncertainty associated with extrapolations will help in the determining optimal portfolio strategies for both the public and private sector. Investing in sensible portfolios is critical for our ability to achieve a near-zero carbon emissions energy infrastructure for the second half of the coming century.

Acknowledgements: This work was supported under National Science Foundation grant SBE-0738187. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Wright, T.P., *Factors Affecting the Costs of Airplanes.* Journal of Aeronautical Sciences, 1936. **10**: p. 302-328.
2. Abernath.Wj and K. Wayne, *Limits Of Learning Curve.* Harvard Business Review, 1974. **52**(5): p. 109-119.
3. Alchian, A., *Reliability of Progress Curves in Airframe Production.* Econometrica, 1963. **31**(4): p. 679-693.
4. Argote, L. and D. Epple, *Learning Curves in Manufacturing.* Science, 1990. **247**(4945): p. 920-924.
5. Boston Consulting Group, *Perspectives on Experience*. 1968.

6. Conley, P., *Experience Curves as a Planning Tool.* IEEE Spectrum, 1970. **7**(6): p. 63-68.
7. Dutton, J.M. and A. Thomas, *Treating Progress Functions as a Managerial Opportunity.* Academy of Management Review, 1984. **9**(2): p. 235-247.
8. IEA, *Experience Curves for Energy Technology Policy*. 2000: OECD/IEA.
9. McDonald, A. and L. Schrattenholzer, *Learning rates for energy technologies.* Energy Policy, 2001. **29**: p. 255-261.
10. Neij, L., *Use of experience curves to analyse the prospects for diffusion and adoption of renewable energy technology.* Energy Policy, 1997. **25**(13): p. 1099-1107.
11. Nemet, G.F., *Beyond the learning curve: factors influencing cost reductions in photovoltaics.* Energy Policy, 2006. **34**: p. 3218-3232.
12. van der Zwaan, B. and A. Rabl, *The learning potential of photovoltaics: implications for energy policy.* Energy Policy, 2004. **32**: p. 1545-1554.
13. Williams, R.H. and G. Terzian, *A Benefit-Cost Analysis of Accelerated Development of Photovoltaic Technology*. 1993, The Center for Energy and Environmental Studies - Princeton University.
14. Zimmerman, M.B., *Learning Effects and the Commercialization of New Energy Technologies: The Case of Nuclear Power.* The Bell Journal of Economics, 1982. **13**(2): p. 297-310.
15. Gritsevskyi, A. and N. Nakicenovic, *Modeling uncertainty of induced technological change.* Energy Policy, 2000. **28**: p. 907-921.
16. Martino, J.P., *Technological Forecasting for Decision Making*. 1993, New York: McGraw-Hill, Inc.
17. Nemet, G.F., *Policy and innovation in low-carbon energy technologies*, in *Energy and Resources Group*. 2007, University of California: Berkeley, CA.
18. Dutton, J.M., A. Thomas, and J.E. Butler, *The history of progress functions as a managerial technology.* Business History Review, 1984: p. 204-233.
19. Adler, P.S. and K.B. Clark, *Behind the Learning Curve: A Sketch of the Learning Process.* Management Science, 1991. **37**(3): p. 267-281.
20. Arrow, K.J., *The Economic Implications of Learning by Doing.* The Review of Economic Studies, 1962. **29**: p. 155-173.
21. Rosenberg, N., *Inside the Black Box*. 1982, Cambridge: Cambridge University Press.
22. von Hippel, E., *The Sources of Innovation.* 1988.
23. Kleinberg, J., *Bursty and Hierarchical Structure in Streams.* Proc. 8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2002.
24. Basalla, G., *The Evolution of Technology*. Cambridge History of Science. 1988, Cambridge: Cambridge University Press.
25. Trancik, J.E. and K. Zweibel, *Technology Choice and the Cost Reduction Potential of Photovoltaics.* 2006 IEEE 4th World Conference on Photovoltaic Energy Conversion (IEEE WCPEC-4), 2006. **1&2**: p. 2490-2493.
26. Arthur, W.B., *The Logic of Invention*. 2005, Santa Fe Institute.
27. Arthur, W.B., *The Structure of Invention.* Research Policy, 2007. **36**: p. 274-287.

28. Arthur, W.B. and W. Polak, *The Evolution of Technology Within a Simple Computer Model.* Complexity, 2006. **11**(5).
29. Arthur, W.B., *The Nature of Technology: What It Is and How It Evolves*. 2008, New York, NY: Free Press.
30. Lyon, T.P., *Does Dual Sourcing Lower Procurement Costs?* The Journal of Industrial Economiccs, 2006. **LIV**(2): p. 223-252.
31. Cook, E., *Limits to exploitation of nonrenewable resources.* Science, 1976. **191**(4228): p. 677-682.
32. Auerswald, P., et al., *The Production Recipes Approach to Modeling Technological Innovation: An Application to Learning by Doing.* Journal of Economic Dynamics & Control, 2000. **24**: p. 389-450.
33. Muth, J.F., *Search Theory and the Manufacturing Progress Function.* Management Science, 1986. **32**: p. 948-962.
34. Sahal, D., *A Theory of Progress Functions.* IIE Transactions, 1979. **11**(1): p. 23 - 29.
35. Bagley, R.J. and J.D. Farmer, *Spontaneous Emergence of a Metabolism.* 1991: p. 93-140.
36. Bagley, R.J., J.D. Farmer, and W. Fontana. *Evolution of a Metabolism*. in *Artificial Life II*. 1991: Addison Wesley.
37. Bagley, R.J., et al., *Modeling Adaptive Biological Systems.* Biosystems, 1989. **23**: p. 113-138.
38. Farmer, J.D., *A Rosetta Stone for Connectionism.* Physica D, 1990. **42**: p. 153-187.
39. Farmer, J.D., S.A. Kauffman, and N.H. Packard, *Autocatalyic Replication of Polymers.* Physica D, 1986. **22**(1-3): p. 187-204.
40. Farmer, J.D., et al., *Evolution, Games, and Learning: Models for Adaptation in Machines and Nature*. 1986, Amsterdam: North Holland Physics Publishing.
41. Farmer, J.D., N.H. Packard, and A.S. Perelson, *The Immune System, Adaptation and Machine Learning.* Physica D, 1986. **22**(1-3): p. 187-204.
42. Trancik, J.E., *Scale and innovation in the energy sector: a focus on photovoltaics and nuclear fission.* Environmental Research Letters, 2006. **1**: p. 014009 (7 pp).
43. Arthur, W.B., *Increasing Returns and Path Dependence in the Economy*. 1994, Ann Arbor: University of Michigan Press.
44. Blanford, G.J. and J.P. Weyant. *A Global Portfolio Strategy for Climate Change Technology Development*. in *International Energy Workshop 2005*. 2005. Kyoto, Japan.
45. Cowan, R., *Tortoises and Hares: Choice Among Technologies of Unknown Merit.* The Economic Journal, 1991. **101**(407): p. 801-814.
46. Best, M.J. and R.R. Grauer, *On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results.* The Review of Financial Studies, 1991. **4**(2): p. 315-342.
47. Best, M.J. and R.R. Grauer, *Positively weighted minimum-variance portfolios and the structure of assest expected returns.* Journal of Financial and Quantitative Analysis, 1992. **27**(4): p. 513-537.

48. Britten-Jones, M., *The sampling error in estimates of mean-variance efficient portfolio weights.* The Journal of Finance, 1999. **54**(2): p. 655-671.
49. Chopra, V.K. and Z.W. T., *The effect of errors in means, variances, and covariances on optimal portfolio choice.* The Journal of Portfolio Management, 1993. **Winter**: p. 6-11.
50. Jobson, J.D. and B. Korkie, *Estimation for Markowitz efficient portfolios.* Journal of the American Statistical Association, 1980. **75**(371): p. 544-554.
51. Merton, R.C., *On estimating the expected return on the market.* Journal of Financial Economics, 1980. **8**: p. 323-361.
52. Dasgupta, P., *Human Well-Being and the Natural Environment.* Oxford University Press, 2004: p. 1-305.
53. Frederick, S., G. Loewenstein, and T. O'Donoghue, *Time Discounting and Time Preference: A Critical Review.* Journal of Economic Literature, 2002. **40**(2): p. 351-401.
54. Dasgupta, P. and E.S. Maskin, *Uncertainty and hyperbolic discounting.* American Economic Review, 2005. **95**(4): p. 1290-1299.
55. Newell, R.G. and W.A. Pizer, *Discounting the distant future: how much do uncertain rates increase valuations?* Journal of Environmental Economics and Management, 2003. **46**(1): p. 52.
56. Newell, R.G. and W.A. Pizer, *Uncertain discount rates in climate policy analysis.* Energy Policy, 2004. **32**(4): p. 519.
57. Farmer, J.D. and J. Geanakoplos, *A Rational Explanation for Hyperbolic Discounting*. 2008, Santa Fe Institute (in progress).