Superexponential Long-Term Trends in Information Technology

Béla Nagy
J. Doyne Farmer
Jessika E. Trancik
John Paul Gonzales

SFI WORKING PAPER: 2010-11-030

Superexponential Long-term Trends in Information Technology

Béla Nagy[1,*], J. Doyne Farmer[1], Jessika E. Trancik[1,2], John Paul Gonzales[1]

[1]Santa Fe Institute
1399 Hyde Park Road
Santa Fe, New Mexico 87501
USA

[2]Engineering Systems Division
Massachusetts Institute of Technology
Cambridge, Massachusetts 02138
USA

November 3, 2010

Abstract

Moore's Law has created a popular perception of exponential progress in information technology. But is the progress of IT really exponential? In this paper we examine long time series of data documenting progress in information technology gathered by Koh and Magee (2006). We analyze six different historical trends of progress for several technologies grouped into the following three functional tasks: information storage, information transportation (bandwidth), and information transformation (speed of computation). Five of the six datasets extend back to the nineteenth century. We perform statistical analyses and show that in all six cases one can reject the exponential hypothesis at statistically significant levels. In contrast, one cannot reject the hypothesis of superexponential growth with decreasing doubling times. This raises questions about whether past trends in the improvement of information technology are sustainable.

1 Introduction

Since Gordon Moore first proposed his famous law, the predicted dramatic improvement in information technology has revolutionized life in most parts of the world. Moore's original prediction

*Contact author: bn@santafe.edu `http://www.santafe.edu/profiles/?pid=355`

was restricted to the statement that transistor count per unit area increases exponentially with a doubling time of one year (Moore 1965), later revised to two years (Moore 1975). Since then his hypothesis, with some variations in doubling times, has been extended to apply to almost every performance metric for information technology hardware. But is the rate of improvement really exponential?

In this paper we show that if one looks over a sufficiently long span of time, all of the relevant performance metrics appear to improve superexponentially. We examine several different hypotheses for superexponential growth, some of which include singularities in finite time, and some of which do not, and show that it is not possible at this stage to distinguish between them. This raises questions about whether or not the historical trends for information technology are sustainable.

Our analysis uses functional performance metrics originally proposed by Koh and Magee (2006)[1]. These include information storage (per unit volume), bandwidth, and calculations per second, as well as their costs, making a total of six different performance metrics, as summarized in Table 1. Koh and Magee (2006) contains an extensive historical database, in many cases going back in time for more than a hundred years. Although there are many missing values (times for which no observations are available), their data are unrivaled in terms of scope and reliability, covering long time periods with high quality[2].

We follow Koh and Magee and analyze a time series based only on those data points that represent "the best performance at a given time", i.e. those that are not dominated by any previous data points. This quantifies the upper envelope of progress consisting of a sequence of world records. Given the incompleteness of the data, this has the important advantage that we do not have to monitor all technologies, but only the best, to have a well-defined series.

Koh and Magee (2006) hypothesized that all six trends are approximately exponential[3] and estimated annual exponential progress rates under the assumption that they are constant (using simple linear regression on the logarithmic scale). We revisit this using a more complex analysis based on generalized nonlinear regression, including a richer statistical model that allows for correlated errors and conclude that the progress rates are increasing.

Without exception the empirical evidence points to shrinking doubling times as time progresses, meaning that all of these trends are in fact not exponential but superexponential. To illustrate that, in the next section we fit hyperbolic trends arising from a flexible family of power functions that can progress faster or slower than an exponential.

Additional functional forms are explored in Section 3, presenting alternatives to power laws and providing different ways to reject the exponential hypothesis. Finally, in Section 4 we summarize our findings, discuss some limitations, and identify some directions for future work. In addition to the empirical analysis, the appendix also provides a possible theoretical explanation for the power function, describing how certain probabilistic mechanisms may give rise to superexponential, constant rate exponential, or subexponential trends over time.

---

[1] Functional performance metrics have also been proposed for energy technologies (Koh and Magee 2008) and for wireless communication (Amaya and Magee 2008).

[2] All the data can be found in Koh and Magee (2006) complete with citations to the original sources. In addition, the numbers used in this paper are also available for free download in CSV format or as HTML tables at `http://pcdb.santafe.edu/` and can be plotted online in a web browser by using this web-enabled Performance Curve Database. Table A3 in Koh and Magee (2006) had some incorrect data for the "Alpha Server SC ES45/1 GHz/3024" in the year 2001, as confirmed by Magee (2009), and because of that this data point was excluded from the analysis.

[3] Kelly (2010) also agrees with an "invariant slope" hypothesis for "a long-run emergent exponential", illustrated by the example of Kryder's Law (Walter 2005) for information storage densities of successive magnetic technologies.

Operation	*Functional performance metric*	*Unit*	*Time period*	*Length*
Storage	Information per unit volume	Megabits per cubic centimeter	1890 – 2004	24
Storage	Information per unit cost	Megabits per dollar	1919 – 2003	22
Transportation	Bandwidth	Kbps	1858 – 2002	19
Transportation	Bandwidth per cable length per unit cost	Kbps per km per dollar	1858 – 2002	21
Transformation	Calculations per second	MIPS	1891 – 2004	27
Transformation	Calculations per second per unit cost	MIPS per dollar	1891 – 2004	31

Table 1: Summary table of six functional performance metrics for measuring progress in IT during six overlapping historical time periods. The lengths of the data sets (i.e. the number of data points in each time series) are in the last column. Kbps is Kilobits per second, dollar refers to real 2004 U.S. dollars (using the GDP deflator as the inflation adjustment), and MIPS is Million Instructions Per Second. Storage technologies are punch card, magnetic tape, magnetic disk, and optical disk. Transportation technologies are single cable, coaxial cable, and optical cable. Transformation technologies are mechanical calculators, vacuum tube based computers, transistor based computers, and integrated circuit based computers ranging from personal computers to supercomputers. Length refers to the number of data points.

2 Analysis

In this section we begin by highlighting the assumptions underlying a linear regression analysis, such as that in Koh and Magee (2006). Then a more general nonlinear regression model is derived by relaxing some of the assumptions in subsection 2.2. Finally, we conclude this section by presenting the results of the generalized nonlinear regression analysis in subsection 2.3.

2.1 Linear Regression Assumptions

The linear regression model used by Koh and Magee is

$$\log y(t) = \alpha + \beta t + \varepsilon(t),$$

where $\log$ is the natural logarithm, $y(t)$ denotes one of the six functional performance metrics as a function of the time variable t, α and β are the intercept and slope parameters for the linear time trend, and $\varepsilon(t)$ is a Gaussian white noise term. There are four key assumptions in this model:

1. Linearity: the $\alpha + \beta t$ trend is linear.
2. Independence: the observations $\log y(t)$ and $\log y(t')$ are independent of each other for any two time points t and t'.
3. Equal variance: the variance of $\log y(t)$ is constant.
4. Normality: the departures of $\log y(t)$ from the $\alpha + \beta t$ trend are normally distributed.

Alternatively, we can express the last three statements by specifying that the $\varepsilon(t)$ random variables are independent, identically distributed Gaussians with zero mean and a constant variance

σ^2. Besides simplicity, the advantages of using such simple linear regressions include familiarity and easy interpretability. But residual analyses of these linear models for the six time series in Table 1 reveal that the first two assumptions of the above four are not satisfied. So we relax the linearity and independence constraints (while keeping equal variance and normality), resulting in a more complex statistical model. This is described in detail in the next subsection.

2.2 Generalized Nonlinear Regression Model

A common trick used by statisticians when some of the assumptions for a linear regression are not satisfied is to search for a transformation that makes the assumptions more reasonable. Perhaps the most popular family of such transformations is the one named after Box and Cox (1964). In our case this means transforming $y(t)$ by applying a power function f, parameterized by a shape parameter λ:

$$f\left(y(t)\right) = \begin{cases} \frac{1}{\lambda}\left(y(t)^{\lambda} - 1\right), & \text{if } \lambda \neq 0; \\ \log y(t), & \text{if } \lambda = 0. \end{cases} \tag{1}$$

This is a continuous family of transformations in λ, meaning that

$$\frac{1}{\lambda}\left(y(t)^{\lambda} - 1\right) \to \log y(t) \quad \text{as} \quad \lambda \to 0.$$

After finding the best λ (that made the assumptions least violated), a traditional linear regression analysis would typically then proceed by fitting a linear trend $\alpha + \beta t$ to the transformed $f\left(y(t)\right)$ response. However, that requires transferring everything over to the new scale (defined by the shape parameter λ). To avoid that step, we instead use nonlinear regression. Substituting $\alpha + \beta t$ in place of $f\left(y(t)\right)$ in equation (1) leads to a nonlinear function for $y(t)$:

$$y(t) = \begin{cases} \left(1 + \lambda(\alpha + \beta t)\right)^{1/\lambda}, & \text{if } \lambda \neq 0; \\ \exp\{\alpha + \beta t\}, & \text{if } \lambda = 0. \end{cases} \tag{2}$$

Taking logarithms on both sides and adding a generalized $\varepsilon(t)$ noise term to the right hand side gives the statistical model that we use:

$$\log y(t) = \begin{cases} \frac{1}{\lambda}\log\left(1 + \lambda(\alpha + \beta t)\right) + \varepsilon(t), & \text{if } \lambda \neq 0; \\ \alpha + \beta t + \varepsilon(t), & \text{if } \lambda = 0. \end{cases} \tag{3}$$

This includes an exponential trend as a special case when $\lambda = 0$. This model is continuous in λ, since the limit of the $\lambda \neq 0$ case as λ goes to zero is the $\lambda = 0$ case. This is a nonlinear regression model because the trend $\frac{1}{\lambda}\log\left(1 + \lambda(\alpha + \beta t)\right)$ is no longer linear. In addition, we also relax the independence assumption for the noise term $\varepsilon(t)$ by allowing its covariance at different times to be nonzero. After testing several hypotheses, we chose the following functional form:

$$\text{Cov}\left(\varepsilon(t),\ \varepsilon(t')\right) \ = \ \sigma^2 \exp\left\{-|t - t'|/\rho\right\}, \tag{4}$$

where σ^2 is the variance and ρ is the characteristic decay time[4].

[4] Based on the likelihood this covariance function provided the best fits compared to linear, spherical, and rational quadratic correlations (all under the assumption that the covariance is a function of the time difference $|t - t'|$). The model was fitted independently for each of the six data sets by generalized nonlinear least squares, using the *gnls* function in the *nlme* package in the statistical software *R*.

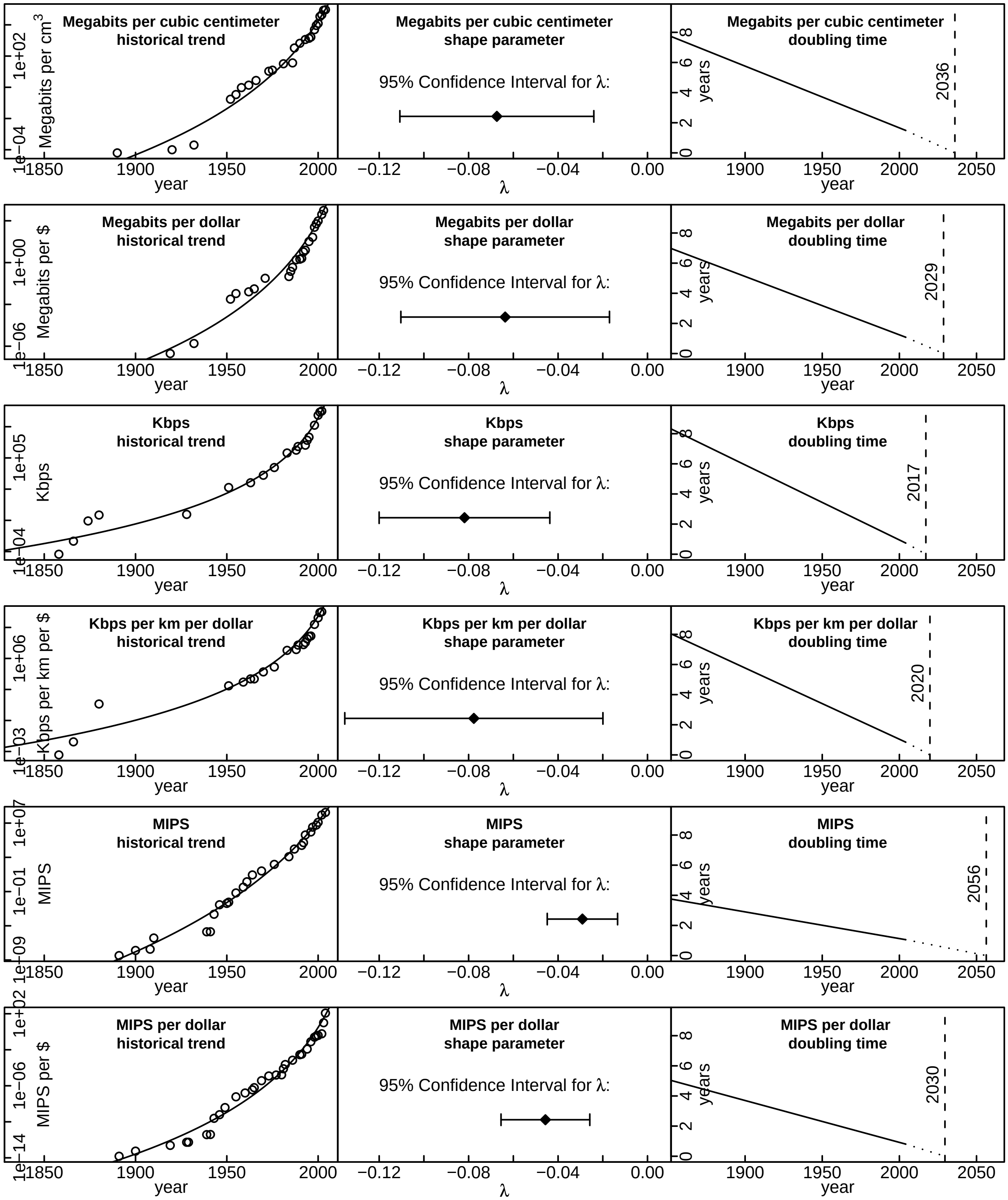

Figure 1: Generalized nonlinear least squares fits for the six performance metrics in Table 1. Left: fitted curves (solid lines) over the historical data points (circles), plotted on semi-log scales; Middle: approximate 95% confidence intervals for the shape parameter λ; Right: estimated doubling times (solid lines) as a function of time and extrapolations (dotted lines). The estimated finite time singularities are indicated by vertical dashed lines (where the extrapolated doubling times hit zero).

2.3 Generalized Nonlinear Regression Results

Figure 1 gives an overview of our results. In the first column the fitted nonlinear trends are shown together with the empirical data points on semi-log scales (i.e. the performance metrics on the vertical axes are on a logarithmic scale, but the time variable on the horizontal axes is on a linear scale). In this view exponential trends are straight lines. That is not what we see. Instead, all of them are superexponential, as indicated by the fact that in every case $\lambda < 0$ by a statistically significant amount. Non-exponential behavior implies that doubling times are not constant. In the third column of Figure 1 we plot the doubling times as a function of the time t. This way of parameterizing superexponential behavior implies that there is a finite time singularity in the year $-(\alpha + 1/\lambda)/\beta$.

Note that here we are talking about a mathematical singularity (division by zero) of the kind described for example in von Foerster, Mora and Amiot (1960); Meyer and Vallee (1975); Kremer (1993); Johansen and Sornette (2001); Bettencourt, Lobo, Helbing, Kühnert and West (2007). This is different from the "technological singularity" discussed by various authors (Vinge 1993; Kurzweil 2005; Yudkowsky 2007) [5]. The finite time singularity estimates are indicated in the third column in Figure 1 by dashed lines, and listed with approximate standard errors in the last two columns of Table 2. As we can see, the uncertainties are much too large to enable any meaningful timing of these singularities. (This may change in the future when we will have more data points.)

Unit of performance metric	*slope*	*1900*	*1950*	*2000*	*singularity*	*standard error*
Megabits per cubic centimeter	-0.047	6.4	4	1.7	2036	20
Megabits per dollar	-0.044	5.7	3.5	1.3	2029	17
Kbps	-0.057	6.6	3.8	0.97	2017	9
Kbps per km per dollar	-0.054	6.4	3.8	1.1	2020	14
MIPS	-0.02	3.2	2.1	1.1	2056	25
MIPS per dollar	-0.032	4.1	2.5	0.93	2030	12

Table 2: Shrinking doubling times. Estimated slopes (for the doubling time declines shown in the third column of Figure 1) and estimated doubling times for the years 1900, 1950, and 2000. The slopes can be estimated by multiplying the estimates of λ by $\log 2 = 0.693$. The projected years for the singularities are shown with the corresponding approximate standard errors, in years.

3 Alternatives

If the hyperbolic functional form used here is correct, then a regime change is inevitable. This is because of the finite time singularity: it is physically impossible for performance to go to infinity. However, there are many other alternative functional forms, and as we will show, it is not clear whether the evidence supports a finite time singularity or an alternative superexponential.

In the following two subsections, we describe alternatives proposed by others and project the resulting fitted curves several decades into the future to dramatize the importance of identifying the right functional form. Figure 2 includes the power function in black (same as in Figure 1), a piecewise exponential in red, a double exponential in green, and the (simple) exponential in blue (corresponding to the $\lambda = 0$ case in equation (3)).

[5] A recent review article by Sandberg (2010) provides an excellent overview of several different singularity concepts.

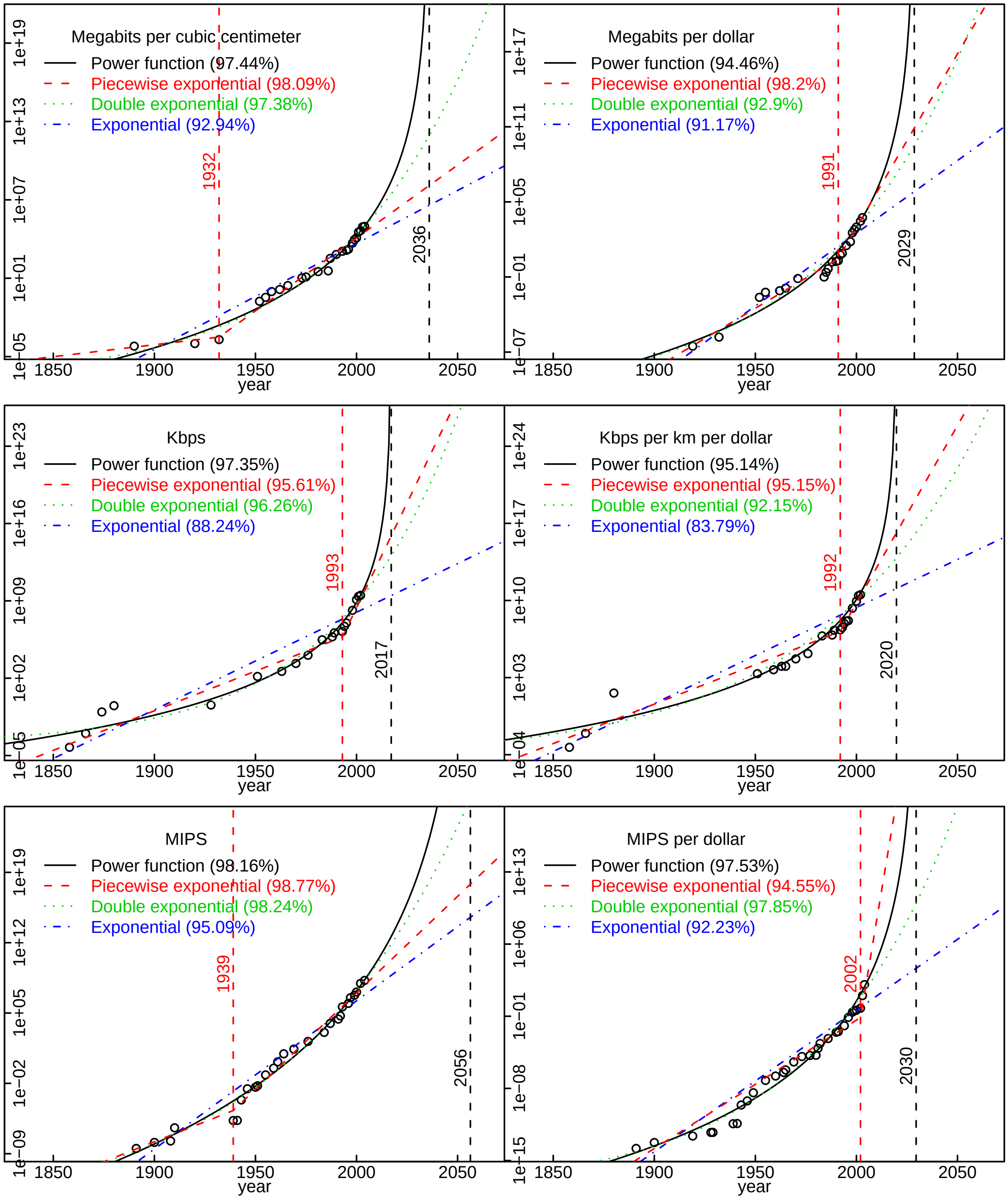

Figure 2: Four different functional forms are fit to each of the six performance metrics described in Table 1. The R^2 value for each fit is shown in brackets. The vertical black dashed lines indicate the estimated singularity times for the power function, and the vertical red dashed lines indicate the estimated regime change τ for the piecewise exponential.

3.1 Piecewise exponential

Based on the data alone we cannot exclude the possibility that abrupt regime shifts have already happened. For instance, this was the interpretation of Amaya and Magee (2008) for a case study of historical wireless throughput evolution. Another example is Nordhaus (2007), who, based on his own measures of computer processing speeds, concluded that "there was a major break in the trend around World War II".

However, one should keep in mind a warning from Jurvetson (2009): "In practice, one can tell any choice of stories from the selection of segments, a post-hoc human judgment. Too tempting a source of bias." To alleviate human bias we do a completely data-driven analysis. We fit two-piece exponentials subject to the constraint that the combined curve is continuous, estimating the parameters by maximum likelihood in the model

$$\log y(t) = \alpha + \beta t + \gamma(t - \tau)I\{t > \tau\} + \varepsilon(t), \tag{5}$$

where τ is a breakpoint parameter (when the regime changes). As before, $\varepsilon(t)$ is a zero mean Gaussian process with a covariance structure defined by equation (4). The breakpoint at τ is implemented by the indicator function

$$I\{t > \tau\} = \begin{cases} 0, & \text{if } t \leq \tau; \\ 1, & \text{if } t > \tau. \end{cases}$$

This model has slope β and intercept α for $t \leq \tau$ and slope $\beta + \gamma$ and intercept $\alpha - \gamma\tau$ for $t > \tau$.

3.2 Double exponential

Yet another alternative is the double exponential proposed by Kurzweil (2001, 2005) "meaning that the rate of exponential growth is itself growing exponentially". So we fitted an intercept ζ plus an exponential time trend $\exp(\eta + \theta t)$ by maximum likelihood, using the additive error term $\varepsilon(t)$ as before, in order to test the double exponential model

$$\log y(t) = \zeta + \exp(\eta + \theta t) + \varepsilon(t).$$

3.3 Model selection

Goodness of fit alone is a dangerous criterion for selecting functional forms. Although the extrapolations diverge wildly, during the period we have data for, the curves are clustered tightly. A crude solution is to use an information criterion that introduces a penalty for the number of free parameters. We tried both the AIC (Akaike 1974) and the BIC (Schwarz 1978) and got similar results. We prefer BIC because it includes a bigger penalty for the number of parameters:

$$\begin{aligned} \text{AIC} &= 2\,(\text{number of parameters}) - 2\,\log \text{likelihood}; \\ \text{BIC} &= \log(\text{number of data points})\,(\text{number of parameters}) - 2\,\log \text{likelihood}. \end{aligned}$$

By looking at the BIC scores in Table 3, we can see that piecewise exponential fits lead to lower (better) scores than those based on power functions with only one exception: the BIC for Kbps. In contrast, scores for the double exponential are higher (meaning worse) than the ones for the power function four times out of six. (Note that the simple one-piece exponential has four parameters, the two-piece exponential has six, and the other two models have five).

The differences in these numbers are much too small to reach any definitive conclusions or to provide a clear ranking between the competing functional forms. The exponential is the only

Unit of performance metric	*Power function*	*Piecewise exponential*	*Double exponential*	*Exponential*
Megabits per cubic centimeter	72.69	69.20	71.72	77.58
Megabits per dollar	61.92	57.70	63.43	64.03
Kbps	72.39	73.08	75.29	78.92
Kbps per km per dollar	80.65	80.06	83.02	81.42
MIPS	107.22	104.82	106.54	112.51
MIPS per dollar	114.98	111.71	115.12	124.18

Table 3: Bayesian information criterion (BIC) for the four different fits for the six performance metrics.

exception, since it can be rejected in favor of a hyperbola, as shown in the previous section. Moreover, we can also reject the exponential in favor of a two-piece exponential by using a likelihood ratio test to justify the extra two parameters (γ and τ). The resulting p-values in Table 4 show that we can reject the exponential hypothesis in favor of a piecewise exponential for all six performance metrics at statistically significant levels.

Unit of performance metric	*Piecewise exponential*	*Exponential*	*p-value*
Megabits per cubic centimeter	-25.0669	-32.4344	0.0006
Megabits per dollar	-19.5769	-25.8325	0.0019
Kbps	-27.7073	-33.5692	0.0028
Kbps per km per dollar	-30.8952	-34.6219	0.0241
MIPS	-42.5214	-49.6623	0.0008
MIPS per dollar	-45.5530	-55.2198	0.0001

Table 4: Log likelihoods for the piecewise exponential and the exponential fits for the six performance metrics and the resulting p-values from the likelihood ratio tests.

4 Discussion

We are not the first to notice accelerating performance trends in information technology. Our contribution is in quantifying this acceleration more rigorously and statistically rejecting the exponential hypothesis. Both the power family and the piecewise exponential family give superior fits to exponentials, though neither the power nor the piecewise exponential can be rejected in favor of the other. These results are inconclusive as to the best functional form.

The double exponential[6] got the third place in a very close race based on BIC, but we cannot rule out that it might result in the best predictions. It will be interesting to watch this race unfold as more data becomes available in the future. Without a more fundamental hypothesis the piecewise exponential is unparsimonious and will likely result in poor predictions: if we allow two trends,

[6] Some of the differential equations for modeling computational power and world knowledge in the appendix of Kurzweil (2005) can also have power law solutions with finite time singularities; however, they are rejected in favor of the double exponential that has no such time limit.

how do we know that new trends will not occur in the future? The possibility of hyperbolic dynamics is intriguing and raises questions about the sustainability of these trends. Obviously it is physically impossible to reach a singularity, indicating that before that hyperbolic growth must necessarily break down. There may be fundamental physical limits (Levitin and Toffoli 2009) to cause the historical trends to be violated. Other limits might be reached even sooner. For example, according to Jurvetson (2004), "another problem is the escalating cost of a semiconductor fab plant, which is doubling every three years, a phenomenon dubbed Moore's Second Law".

The main limitation of this paper is that the fitted curves are merely simplified descriptions of the trends in past performance and may not be predictive of future performance. Even though we have applied parsimony penalties, such in-sample[7] fits are unreliable. Nonetheless, these results provide substantial evidence against simple exponential interpretations of the existing data and suggest that the long-term dynamics are more complicated.

5 Acknowledgments

Partial funding for this work was provided by the National Science Foundation Science of Science and Innovation Policy (SciSIP) program (NSF Award Number: 0738187) and The Boeing Company (Contact: Scott Mathews, TF). We thank Professor Christopher L. Magee for his help with the data, Quan Minh Bui for correcting some mistakes, and Clément Vidal for useful suggestions about how to improve the presentation.

References

AKAIKE, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19** 716–723.

AMAYA, M. A. and MAGEE, C. L. (2008). The Progress in Wireless Data Transport and its Role in the evolving Internet. Working Paper ESD-WP-2008-20, Massachusetts Institute of Technology. URL `http://esd.mit.edu/WPS/2008/esd-wp-2008-20.pdf`.

ARNOLD, B. C., BALAKRISHNAN, N. and NAGARAJA, H. N. (1998). *Records*. John Wiley and Sons.

BETTENCOURT, L. M. A., LOBO, J. A., HELBING, D., KÜHNERT, C. and WEST, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, **104** 7301–7306. URL `http://www.pnas.org/content/104/17/7301`.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, **26** 211–252.

EMBRECHTS, P., MIKOSCH, T. and KLÜPPELBERG, C. (1997). *Modelling extremal events: for insurance and finance*. Springer-Verlag, London, UK.

FRANK, S. A. (2009). The common patterns of nature. *Journal of Evolutionary Biology*, **22** 1563–1585. URL `http://onlinelibrary.wiley.com/doi/10.1111/j.1420-9101.2009.01775.x/pdf`.

[7]We are currently performing a more extensive analysis using out-of-sample testing for a wide variety of different technologies in reference Nagy, Farmer, Trancik and Bui (2010).

GULATI, S. and PADGETT, W. J. (2003). *Parametric and Nonparametric Inference from Record-Breaking Data*. Springer-Verlag.

JOHANSEN, A. and SORNETTE, D. (2001). Finite-time singularity in the dynamics of the world population, economic and financial indices. *Physica A: Statistical Mechanics and its Applications*, **294** 465 – 502.

JURVETSON, S. (2004). Transcending Moore's Law with Molecular Electronics and Nanotechnology. *Nanotechnology Law and Business*, **1** 70–90. URL http://www.dfj.com/files/TranscendingMoore.pdf.

JURVETSON, S. (2009). Personal communication.

KELLY, K. (2010). *What Technology Wants*. Viking.

KOH, H. and MAGEE, C. L. (2006). A functional approach for studying technological progress: Application to information technology. *Technological Forecasting and Social Change*, **73** 1061–1083.

KOH, H. and MAGEE, C. L. (2008). A functional approach for studying technological progress: Extension to energy technology. *Technological Forecasting and Social Change*, **75** 735–758.

KREMER, M. (1993). Population Growth and Technological Change: One Million B.C. to 1990. *The Quarterly Journal of Economics*, **108** 681–716.

KURZWEIL, R. (2001). The Law of Accelerating Returns. URL http://www.kurzweilai.net/articles/art0134.html.

KURZWEIL, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. Viking Penguin. URL http://singularity.com/.

LEVITIN, L. B. and TOFFOLI, T. (2009). Fundamental limit on the rate of quantum dynamics: The unified bound is tight. *Physical Review Letters*, **103** 160502.

MAGEE, C. L. (2009). Personal communication.

MARTINO, J. P. (1987). Using precursors as leading indicators of technological change. *Technological Forecasting and Social Change*, **32** 341 – 360.

MARTINO, J. P. (1992). Probabilistic technological forecasts using precursor events. *Technological Forecasting and Social Change*, **42** 121 – 131.

MARTINO, J. P. (1993a). Bayesian updates using precursor events. *Technological Forecasting and Social Change*, **43** 169 – 176.

MARTINO, J. P. (1993b). *Technological Forecasting for Decision Making*. McGraw-Hill, Inc., New York.

MCNERNEY, J., FARMER, J. D., REDNER, S. and TRANCIK, J. E. (2009). The Role of Design Complexity in Technology Improvement. URL http://arxiv.org/abs/0907.0036.

MEYER, F. and VALLEE, J. (1975). The dynamics of long-term growth. *Technological Forecasting and Social Change*, **7** 285 – 300.

MOORE, G. E. (1965). Cramming more components onto integrated circuits. *Electronics Magazine*, **38**. URL http://download.intel.com/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore_1965_Article.pdf.

MOORE, G. E. (1975). Progress in digital integrated electronics. *IEEE International Electron Devices Meeting* 11–13.

MUTH, J. F. (1986). Search theory and the manufacturing progress function. *Management Science*, **32** 948–962.

NAGY, B., FARMER, J. D., TRANCIK, J. E. and BUI, Q. M. (2010). Testing Laws of Technological Progress. Unpublished manuscript.

NORDHAUS, W. D. (2007). Two Centuries of Productivity Growth in Computing. *The Journal of Economic History*, **67** 128–159. URL http://www.econ.yale.edu/~nordhaus/homepage/nordhaus_computers_jeh_2007.pdf.

PESSEMIER, E. A. (1977). *Product Management, Strategy and Organization*. John Wiley and Sons.

SANDBERG, A. (2010). An overview of models of technological singularity. *The Third Conference on Artificial General Intelligence (AGI-10)*. URL http://agi-conf.org/2010/wp-content/uploads/2009/06/agi10singmodels2.pdf.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6** 461–464.

SHANNON, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, **27**.

SHARIF, M. N. and ISLAM, M. N. (1980). The Weibull distribution as a general model for forecasting technological change. *Technological Forecasting and Social Change*, **18** 247–256.

VINGE, V. (1993). The Coming Technological Singularity: How to Survive in the Post-Human Era. URL http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html.

VON FOERSTER, H., MORA, P. M. and AMIOT, L. W. (1960). Doomsday: Friday, 13 November, A.D. 2026. *Science*, **132** 1291–1295.

WALTER, C. (2005). Kryder's Law. *Scientific American*, **293** 32–33. URL http://www.scientificamerican.com/article.cfm?id=kryders-law.

YUDKOWSKY, E. (2007). Introducing the Singularity: Three Major Schools of Thought. Singularity Summit 2007. URL http://www.acceleratingfuture.com/people-blog/2007/introducing-the-singularity-three-major-schools-of-thought/.

6 Appendix

In this appendix we present a derivation of equation (2) based on assumptions about the time distribution of record-breaking innovations and the additivity of costs for independent technologies. By definition the functional performance metric $y(t)$ quantifies the upper envelope of progress, where each data point represents a new innovation breaking the previous world record. We assume that the time t of each innovation $y(t)$ can be viewed as a sample from a cumulative distribution function $F(t) = P(T \leq t)$, stating the probability that T is less than or equal than a given time t. We assume that $y(t)$ can be written as a function of $F(t)$, meaning that $y(t)$ depends on the variable t only through the function $F(t)$. The system progresses from $F(\inf(t)) = 0$ to $F(\sup(t)) = 1$.

For modeling purposes, it is conceptually easier to think in terms of the unit cost as a function of the cumulative probability, defined as the inverse of the performance metric:

$$c(F(t)) \;=\; \frac{1}{y(t)}.$$

For example, if the functional performance metric y in question is measured in megabits per dollar, then the unit cost c is measured in dollars per megabit. Likewise, if performance is in megabits per cubic centimeter, then the unit cost (in terms of space) is cubic centimeters per megabit.

We now derive the relationship between the unit cost c and $F(t)$ based on three assumptions:

1. c is a monotone decreasing function of $F(t)$.
2. c is non-negative.
3. The costs of independent technologies are additive. That is, if the underlying $T^{(i)}$ random variables are independent for a set of k technologies,
$$P(T^{(1)} \le t, \ldots, T^{(k)} \le t) \;=\; P(T^{(1)} \le t)\; \ldots \; P(T^{(k)} \le t),$$
then the cost function c is *additive*:
$$c\left(P(T^{(1)} \le t, \ldots, T^{(k)} \le t)\right) \;=\; \sum_{i=1}^{k} c\left(P(T^{(i)} \le t)\right).$$

As originally shown by Shannon (1948), up to a multiplicative constant the only function that satisfies these conditions is the logarithm:

$$c(P(T \le t)) \;=\; \log \frac{1}{P(T \le t)} \;=\; -\log P(T \le t) \;=\; -\log F(t). \tag{6}$$

For convenience of notation we fix the multiplicative constant by choosing the natural logarithm. This cost function suggests the information theoretic interpretation that the size of an innovation is proportional to the surprise in its timing. In the beginning, when t is small, the probability $P(T \le t)$ is also small, providing substantial information and surprise in the early stages of development when the cost of the technology tends to be high. (Indeed, this model specifies infinite cost as long as the probability of observing $T \le t$ is zero.) At the other extreme, when t is nearing the end of the support of T, then $P(T \le t)$ goes to 1, and the information content of the innovation process (as well as the cost) in this terminal stage goes to 0.

We now make an assumption about the functional form of $F(t)$ that closes the connection with the Box-Cox transformation. Assume that each innovation consists of a set of requirements, all of which are necessary for the final innovation at time T. If each requirement j is completed at time R_j, then the last step is completed at time $T \equiv \max\{\, R_1,\ R_2,\ \ldots,\ R_n\}$. If the R_j random variables are sufficiently independent and if n is sufficiently large, then in most cases[8] $F(t)$ is well-approximated by a generalized extreme value distribution[9] with a cumulative distribution function

$$F(t) = \exp\left\{ -\left(1 + \lambda(\alpha + \beta t)\right)^{-1/\lambda} \right\}. \tag{7}$$

[8] There are some distributions, such as the Poisson, that do not converge to any of the three max-stable distributions under the operation of taking the maximum, but most common distributions do converge (Embrechts, Mikosch and Klüppelberg 1997).

[9] Extreme value distributions can also be justified based on the principle of maximum entropy by constraints on average location and average tail weighting (Frank 2009). Time lags between innovations have also been modeled with maximum entropy distributions (Martino 1987, 1992, 1993a,b). Other possible ways of modeling record-breaking processes are summarized by Arnold, Balakrishnan and Nagaraja (1998) and Gulati and Padgett (2003).

In the limit as $n \to \infty$ the distribution of T converges to one of the three possible extreme value distributions: reversed Weibull if $\lambda < 0$, Gumbel if $\lambda \to 0$, or Fréchet if $\lambda > 0$. Recalling that

$$y(t) \;=\; \frac{1}{c(F(t))} \;=\; -\frac{1}{\log F(t)} \tag{8}$$

gives equation (2). For the functional performance metrics we analyzed here, we find that $\lambda < 0$. This would suggest that $F(t)$ is a reversed Weibull distribution with an upper limit at the singularity, resulting in a superexponential trend for the performance metric $y(t)$.

Gumbel would cause exponential growth of $y(t)$ for $\lambda = 0$ and Fréchet would lead to a subexponential power law for $\lambda > 0$ (both of these cases have infinite support at $+\infty$ and do not have a finite time singularity). Weibull distributions have also been proposed for modeling the diffusion of technologies (Pessemier 1977; Sharif and Islam 1980) and extreme value distributions have been used in many other ways for modeling cost innovation, e.g. see Muth (1986) or McNerney, Farmer, Redner and Trancik (2009) and the references therein.