Graz Economics Papers – GEP

GEP 2012–05

Modelling the Emergence of New Technologies using S-Curve Diffusion Models

Miriam Steurer, Robert J. Hill, Markus Zahrnhofer, and Christian Hartmann

November 2012

Department of Economics
Department of Public Economics
University of Graz

Modelling the Emergence of New Technologies using S-Curve Diffusion Models

Miriam Steurer[a*], Robert J. Hill[b], Markus Zahrnhofer[a] and Christian Hartmann[a]

a: POLICIES, Joanneum Research Forschungsgesellschaft, Graz, Austria (miriam.steurer@joanneum.at)

b: Department of Economics, University of Graz, Austria

*: Corresponding author

21 February 2012

Three theoretical benchmark models of diffusion of new technologies are the substitution, mortality and social-learning models. These models tend to generate symmetric, right-skewed and left-skewed S-curves respectively. The empirical literature has focused primarily on fitting either Logistic or Gompertz functions to real data. Given that Logistic is symmetric and Gompertz is right skewed, the former is typically matched with the substitution model and the latter with the mortality model. Neither function can be used to describe the left-skewed social-learning model. We show here how the Generalized-Extreme-Value (GEV) function – which includes Gompertz as a special case and can be either left or right skewed – is more flexible and can be matched with either the mortality or social-learning model. Using cumulative citations as a proxy for diffusion, we fit Logistic, Gompertz and GEV S-curves to 12 citations data sets. Logistic emerges as the best fit for 6 data sets and GEV for the other 6 (all of which are right skewed). It follows that the social-learning model does not fit with any of our data sets. Truncating our data sets in 1996 or 2001 in all but one case does not change the best fit function. This suggests that our fitted S-curves could be useful for modelling aspects (such as the asymptotic upper limit) of a new technology's future path.

Keywords: Diffusion of New Technologies; S-curve, Innovation; GEV function; Cumulative citations

1. Introduction

New technologies do not receive universal approval and usage overnight – rather they need time to gain acceptance and diffuse through an economy. The pattern and speed with which a new technology receives acceptance depends on a variety of factors such as the characteristics of potential users and the intensity of their interactions. Most of all, though, the diffusion path of a technology will depend on its type. Different types of technologies generate different diffusion paths. In this paper we consider how these diffusion paths can be measured.

The diffusion process of most technologies can be illustrated using S curves (also known as sigmoid functions). This has been done in a variety of different applications. One of the earliest examples is Kuznets [1] who was among the first to recognize that technological change can be represented by an S curve. Other early examples include Ryan and Gross [2] and Griliches [3] who used the S-curve model to estimate the diffusion of hybrid corn in different areas of the United States.

The S-curve literature is characterised by an almost complete dichotomy between theory and practice. On the one hand practitioners use S-curves for forecasting purposes as well as characterization of types of processes (see for example Porter et al. [4] and Twiss [5]). On the other hand theorists have developed the theoretical implications of various diffusion models without actually trying to fit their models to real data (see for example Young [6]).

The theoretical underpinnings of the S curve can be found in different fields, primarily in Sociology, Marketing and Economics. Given the multi-disciplinary nature of the literature and the associated differences in notation, assumptions and perspective it can be hard to disentangle the common themes that emerge. Young [6] provides an excellent analysis of the theoretical similarities and differences of various diffusion models. He considers three broad classes of diffusion models: contagion (a model mostly used in Marketing), social influence (from the field of Sociology) and social learning (from Economics). Young shows that each model "leaves a characteristic 'footprint' on the shape of the adoption curve which provides the basis for discriminating empirically between them". This 'footprint' has been exploited in the empirical literature on emerging new technologies for some time. The contagion and social influence models, as described by Young, are essentially equivalent respectively to the substitution and mortality processes that are widely discussed in the technology diffusion literature (see for example Porter et al. [4]). A key 'footprint' of the social-learning and mortality models is skewness. The social-learning model tends to generate a left-skewed S curve while the mortality model generates a right-skewed S curve. The substitution model by contrast tends to generate a symmetric S curve.

The key parameters of an S-curve are the asymptotic upper bound, the point of inflexion, standard deviation, skewness and kurtosis. Most empirical researchers (e.g., Porter et al. [4]; Bengisu and Nekhili [7]) have focused on two S-curve models – Logistic and Gompertz. These models have fixed skewness and kurtosis coefficients. In particular, Logistic is symmetric while Gompertz is right skewed. By implication, neither can provide a good fit for the left-skewed social-learning model.

For this reason we consider here a more flexible S-curve – the Generalized Extreme Value (GEV) function – that can be either left or right skewed and includes Gompertz as a special case.

The diffusion of new emerging technologies can be measured using a variety of different indicators. Traditionally, direct measures such as sales data or the number of firms adopting the technology have been used. We show here that the S-curve model also applies if the diffusion process is measured indirectly by cumulative citations in academic journals and conference proceedings.

Citations data have two advantages over sales or usage data. First, compared to actual sales data citations are easier to obtain and more complete. Young [6] for example laments the fact that a good empirical analysis of S-curve (or diffusion) models is almost impossible due to the lack of good data. As conditions for good empirical analysis he states that a curve needs to be measured at frequent time intervals and that one needs data that covers the whole life cycle (from the very beginning to end) of the adoption process.

> "[O]ne needs data that cover the whole life cycle of the adoption process, including the very early phases when adoption is just getting started. Many of the datasets in the literature are quite weak in this respect; the reason, of course , is that researchers are not likely to focus their attention on an innovation until it is already well underway." (Young [6])

Using citation (or patent) data overcomes some of these difficulties of incomplete data sets. The second advantage concerns the "time element". Wygant and Markley [8], Choo [9} and Martino [10] all note how the diffusion of ideas starts before the diffusion of the actual technologies themselves. Thus, by using cumulative citations in academic journals we have an ideal "leading indicator" for the performance of emerging technologies.

The main contributions of this paper can be summarized as follows. First, we show how the use of the generalized-extreme-value (GEV) distribution can improve the empirical fit of an S curve to real diffusion data. Second, we identify 12 technologies and fit Logistic, Gompertz and GEV S-curves to cumulative citations data on these technologies. We find that the best fit functional form is either Logistic or GEV. In the latter case, Gompertz (a special case of GEV) is never the best fit. Also, none of our fitted S-curves are left skewed. This suggests that the social-learning model may not be a good benchmark for describing the diffusion of citations data. Third, we find that truncating our data sets in 1996 or 2001 in all but one case does not change the best fit S-curve functional form. Given that the functional form provides some insights into the future path of emerging technologies (such as the asymptotic upper limit), this finding (i.e., that the S-curve functional forms are reasonably stable) could be of use to firms in their research and development planning and to governments in the formation of industrial policy.

2. Fitting S-Curves to Technological Diffusion Data

2.1. Literature review

The literature on the diffusion of emerging technology (measured as market share, cumulative sales, cumulative production, etc.) generally uses S curves to predict the diffusion process. This is because the new technology typically grows slowly at first, then gathers momentum and exhibits a growth rate > 1, followed by a period of slower growth (growth rate < 1) and eventually reaches obsolescence and stops developing . Twiss [5] refers to these three stages of the technological S curve as incubation, rapid growth and maturity. Plotted with cumulative acceptance, market share (or whatever it is one is measuring) on the y-axis against time on the x-axis the process can be described by an S curve.

We focus our attention on cumulative citations in scientific journals as our measure of diffusion. Discussions on new technologies in scientific journals start before market production of these technologies. Citations in the scientific literature are therefore an ideal leading indicator of the success of a new technology before it has an impact in the market (as measured by sales volume or market share). Similarily, Bengisu and Nekhili [7] show that there exists a strong positive correlation between citations data and patenting behaviour.

The empirical S-curve literature, in the technology-diffusion context, has tended to focus on just two functional forms (e.g. Porter et al. [4]; Franses [11]; Bengisu and Nekhili [7]).

These are the Logistic and Gompertz functions.[1] One appealing feature of Logistic and Gompertz is that, for a given upper limit, they can be reformulated into linear form and the other parameters estimated using ordinary least squares (OLS).

We show how this property of the Logistic and Gompertz functions can be exploited to estimate the asymptotic upper bound of the S-curve endogenously rather than imposing a value obtained say from expert opinion (as is often done in the literature).[2]

This is done in two stages. In stage 1, the upper limit on cumulative citations is fixed and the other parameters are then estimated using OLS. This is repeated for a range of possible values for the upper limit. A one-dimensional grid search can then be used to determine the optimal upper limit. By contrast, using nonlinear least squares to estimate all the parameters simultaneously implies undertaking a multi-dimensional grid search which increases the risk that the solution found is not a global optimum. In stage 2, the optimal upper limit is then determined using a grid search.

Logistic and Gompertz, however, are not the only functions that can be linearized for a given upper limit. The cumulative density functions (cdfs) of the Cauchy and the Generalized Extreme Value (GEV) distributions, in particular, also have this property. Neither has been previously considered in this literature.[3] It turns out that the Cauchy distribution is not flexible enough for our purposes and never provides a good fit for our

---

[1] The Fisher-Pry [12] (or Pearl) function has also received some attention (see Porter et al. [4]). On closer inspection, however, it can be seen that it is simply a rescaled version of the Logistic function. The Gompertz function is also known as the "reversed Gumbel" or "Gumbel max" function. Confusingly, some authors refer to Gumbel when they mean reversed Gumbel.

[2] While a strand of the literature (see for example Porter [4]) prefers to rely on expert opinions, these are often biased in addition to being inherently subjective. Schnaars [13] and Tichy [14] find that experts often are over-optimistic in their predictions concerning their area of expertise. The approach presented in this paper could be used to complement expert judgement.

[3] In principle a number of other functional forms could also be considered, such as the cumulative density functions (cdfs) of some well-known unimodal statistical distributions such as Gamma and Beta. For example, Jeuland [15] uses gamma distributions to model the diffusion of innovation. However, the parameters of the Gamma and Beta distribution would need to be estimated using some kind of multidimensional grid search algorithm. While this can easily be done using a standard econometrics program, one cannot be sure that the solution found is a global optimum. OLS on the other hand guarantees a global optimum.

data. We show that GEV, by contrast, performs well and hence could be a useful additional weapon in the empirical researcher's toolkit.

The GEV function is, in fact, a generalization of the Gompertz function. Compared to the Gompertz function the GEV function has an extra parameter, gamma, that can influence the skewness and kurtosis of the S-function. We let gamma take on a number of different values (in the vicinity to zero) and then transform GEV into linear form to allow estimation of the other parameters using OLS. Gompertz is obtained in the limit as gamma tends to zero.

The Logistic model has been used to describe a substitution process (see Porter et al. [4]) in which one technology gets replaced (or is substituted for) by a radically new technology entering the market. The slow growth of the technology in the beginning is due to high initial costs of the technology and uncertainty about whether the technology is going to "make it". However, once it has made it over the acceptance threshold further adaptation is very rapid and there is very strong growth in the dissemination of the technology. Once people or firms start switching towards this new technology almost everyone else does so as well. Statistically the action is concentrated around the midpoint of the distribution (which should be approximately symmetric). The Logistic model has a kurtosis of 1.2, which is consistent with the thin-tailed substitution scenario just described.

The Gompertz function generates a right-skewed S curve in which the period of acceleration is shorter than the period of deceleration. Compared to the Logistic model, the Gompertz model describes a more incremental technological change. In a technological context, the Gompertz function is therefore more appropriate to describe cases in which equipment replacement is driven by equipment deterioration rather than obsolescence. It is therefore often referred to as the "mortality model" (see Porter et al. [4]).[4]

Gompertz has a skewness coefficient of 1.14 and a kurtosis of 2.4. The point of inflexion of the Logistic function occurs at a higher cumulative percentage (50%) than the Gompertz curve (36.8%). Both models are rigid both with respect to skewness and kurtosis. The Gompertz function is therefore no more flexible than the Logistic function.

The GEV function, by contrast, is more flexible and lets the data determine the skewness (and hence cumulative percentage of the point of inflexion) and the level of kurtosis rather than fixing them in advance. The greater flexibility of the GEV function allows us to check for the presence of other classes of diffusion models. Allowing for the possibility of left skewness, in particular, is important since Young [6] shows that the social-learning diffusion model tends to generate a left-skewed S curve. Left skewness also arises in business cycles, where the period of expansion is of longer duration than the subsequent contraction. Empirical evidence of such left-skewness in reference to economic recessions and recoveries is provided by King and Plosser [16], Balke and Wynne [17], and Harding and Pagan [18].

Here we only consider values of gamma in the vicinity of zero. Over this range the GEV function is right skewed. (It is symmetric at gamma equal to about -0.3 – see Markose and Alentorn [19]). A left-skewed S-curve is obtained by taking the mirror image of the GEV S-curve. For reasons that are explained later, the original GEV curve is referred to here as GEV-Max while its mirror image is referred to as GEV-Min. We try fitting both GEV-Max and GEV-Min functions to our citations data sets (and thus allow for the possibility of left skewness).

[4] The Gompertz function has been used by the demographer Gompertz in order to describe the phenomenon that the mortality rate of a population increases dramatically as the population ages (see Porter et al. [4]).

By classifying a technological diffusion series as Logistic, Gompertz or some variety of GEV we obtain important information about the dynamics of the process involved. In particular, fitting an S-curve provides an indication of whether the underlying dynamics of the diffusion process are best described by the substitution model, the mortality model, or the social learning model. As these three models differ significantly in their predictions about the pace of further development this is important information.

2.2. Parameter estimation

We begin by reviewing the Logistic and Gompertz functions, which are defined as follows:

Logistic:

$$Y_t = \frac{L}{1 + e^{a-bt}}$$

Gompertz (Gumbel Max):

$$Y_t = Le^{-e^{a-bt}}$$

where Y_t denotes the cumulative number of citations at time t, and L is the asymptotic upper limit. L, a and b are positive valued parameters that are estimated endogenously. Time t can take positive or negative values.

For each function we proceed first to estimate the parameters a and b for given values of L. The optimal value of L is then estimated afterwards. For given L, these functions can be transformed into linear form as follows:

Linearization of Logistic:

$$Y_t = \frac{L}{1 + e^{a-bt}} \Rightarrow \ln\left(\frac{L - Y_t}{Y_t}\right) = a - bt.$$

Linearization of Gompertz (Gumbel Max):

$$Y_t = Le^{-e^{a-bt}} \Rightarrow \ln\left[-\ln\left(\frac{Y_t}{L}\right)\right] = a - bt.$$

The parameters a and b can be estimated using OLS once a value of L has been specified and the dependent variable Y_t has been suitably transformed. It should be noted though that in each case we have taken a nonlinear transformation of the dependent variable. Hence our solutions for a and b only approximately minimize the sum of squared errors in the original model.

Having solved first for a and b, for a range of possible L values, we then solve for the optimal L by doing a grid search to find the value that minimizes the sum of squared errors in the original model. There is the slight risk, as with any nonlinear least squares estimation that the solution found is only a local minimum. However, given the search is only one-dimensional this is not very likely. Also, it is important to note that each value of L in this process has its own OLS a and b parameters.

As noted above, Gompertz is a special case of the GEV-Max function. The GEV distribution is the limiting distribution of the extreme value in a sample of data, with GEV-Max describing the maximum extreme value while GEV-Min describes the minimum extreme value (see Fisher and Tippett [20]). Extreme value theory has been applied in many fields ranging from hydrology and climate modelling to finance and insurance (see for example Embrechts, Kluppelberg and Mikosch [21], Reiss and Thomas [22], and Markose and Alentorn [19]).

Generalized Extreme Value (GEV) Max:

$$Y_t = Le^{-[1+\gamma(a-bt)]^{-\frac{1}{\gamma}}}$$

Gompertz (otherwise known as the type I extreme value function, "reversed Gumbel", or "Gumbel Max") is obtained in the limit as γ in the GEV-Max function tends to zero. When γ is greater than zero, GEV-Max is referred to as the Fréchet (or type II extreme value, or

"Fréchet-Max") function, while when γ is less than zero it is referred to as the reversed Weibull (or type III extreme value or "Weibull-Max") function.

For given γ and L, GEV-Max can be transformed into a form where the parameters a and b can also be estimated using OLS:

Linearization of Generalized Extreme Value (GEV) Max:

$$Y_t = Le^{-\left[1+\gamma(a-bt)\right]^{-\frac{1}{\gamma}}} \Rightarrow \frac{1}{\gamma}\left\{\left[\ln\left(\frac{L}{Y_t}\right)\right]^{-\gamma} - 1\right\} = a - bt.$$

Our objective with regard to the GEV-Max function is to see whether it is possible in general to improve on the fit achieved by Gompertz by allowing γ to differ slightly from zero. Hence we do not solve for γ as such, but instead try a few values in the vicinity of zero. We also consider the GEV-Min function, which is the mirror image of the GEV-Max function and hence has the opposite skewness.

Generalized Extreme Value (GEV) Min:

$$Y_t = L\left[1 - e^{-\left[1-\gamma(a-bt)\right]^{-\frac{1}{\gamma}}}\right]$$

An important special case of the GEV-Min function is Gumbel-Min, which is obtained in the limit as γ tends to zero.

Gumbel-Min:

$$Y_t = L\left[1 - e^{-e^{-(a-bt)}}\right]$$

When γ is greater than zero, the GEV-Min function is known as reversed Fréchet (or Fréchet-Min). Conversely, when γ is less than zero, the GEV-Min function is known as Weibull (or Weibull-Min).

In an attempt to minimize unnecessary confusion, we will henceforth refer to the GEV-Max, GEV-Min, Gumbel-Max (=Gompertz), Gumble-Min, Fréchet-Max, Fréchet-Min, Weibull-Max and Weibull-Min functions.

An important implication of switching from the Max to the Min version of any of these functions is that it reverses the direction of skewness. More specifically, the max versions of each of these functions are right skewed, while the min versions are left skewed. The skewness and kurtosis coefficients of the Logistic, GEV-Max and GEV-Min functions (for gamma in the range [-0.03,0.03]) are shown in Figure 1.

2.3. Choosing between functional forms

There are many different model selection criteria (see for example Fox [23]). In our context, however, they are all equivalent to minimizing the sum of squared errors. This is because we have only one explanatory variable (time) and the same number of parameters given that we fix the gamma parameter in advance for each version of the GEV function that we consider. This means that the parameters to be estimated are a, b, and L. In this case even the Akaike information criterion reduces, under standard assumptions, to comparing the sum of squared errors. It is important, however, that the sum of squared errors are compared for the original version of each function (i.e., with Y_t rather than some function of Y_t as the dependent variable). Otherwise, the sum of squared errors will not be comparable across functional forms.

2.4 Inflexion points

The inflexion point in an S-curve occurs when the second derivative of the function with respect to *t* equals zero. The inflexion points t* for our various functions are listed below.

Logistic:

$$t^* = \frac{a}{b} \qquad a, b > 0$$

Gumbel Max (Gompertz):

$$t^* = \frac{a}{b} \qquad a, b > 0$$

GEV Max:

$$t^* = \frac{1}{b}\left[a - \frac{(1+\gamma)^{-\gamma} - 1}{\gamma}\right] \qquad a, b > 0$$

Gumble Min:

$$t^* = \frac{a}{b} \qquad a, b > 0$$

GEV Min:

$$t^* = \frac{1}{b}\left[a + \frac{(1+\gamma)^{-\gamma} - 1}{\gamma}\right] \qquad a, b > 0$$

While the formulas for the inflexion points of the Logistic and Gompertz functions are the same, each function for any given data set will have its own values of the estimated a and b parameters and hence different values of t*. Furthermore, in the limit as γ tends to zero, the point of inflexion t^* of the GEV-Max and GEV-Min functions also converge to a/b.

For the Logistic function, t^* always coincides with the midpoint of the data series. For the Gumbel-Max (Gompertz) function t^* always occurs just after the first third of all data points considered (at 36.8% of the data set). As the range for gamma that we allow in this paper is quite narrow, the inflexion points of the GEV-Max function will also lie in the vicinity of percentile 36.8 while the inflexion points of the GEV-Min function will lie in the vicinity of percentile 63.2. Some examples of inflexion points for our cumulative citations data sets are provided in Table 1.

2.5 Some implications

The choice of functional form for a particular technology provides important insights. It helps us to characterize the dynamics of the trend. Certain technologies are described best by one functional form and other technologies by another. Thus the main benefit of fitting these distributions to data series is to establish the general "character" of the series.

In addition, our methodology allows us to determine the projected upper limit of total citations endogenously. However, establishing upper limits through the data set itself has to be done with caution as even small errors will cause the confidence bounds on the upper limit L to become quite large. The further the limit L lies in the future, the more severe this problem will be. However, in a study covering 35,000 S-curves from simulated data, DeBecker and Modis [24] found that forecasts from fitted Logistic curves are more reliable than one might expect. A rule-of-thumb result provided by DeBecker and Modis is that, given at least half of the S-curve range and an error level of less than 10% on each historical point, the uncertainty on the upper bound L will be less than 20% with a 90% confidence level. This is good news for the accuracy of predictions of emerging trends using citation data. As we deal with citation data the precision of our individual data points is very good and definitely below an error level of 10 percent per data point. Also, DeBecker and Modis get their result by fitting only Logistic S-curves. By providing alternative functional forms our "fit-statistic" should be further improved.

3. Results

Our data set consists of citations in academic papers on particular topics found in ISI Thomson's Web of Knowledge or JSTOR. We specifically focus on older topics that give us complete (or almost complete) S-curves. We then take these data and attempt to find the best fit curve to match these S-curves.

Our data were obtained from keyword searches. We focused on two types of technologies for which we were able to identify keywords. The first was household/computer technologies. The second was concepts from the area of Economics that were popular in the 1950s, 1960s and 1970s. In total we considered five household/computer technologies (ATM, CD Rom, Fax, Mainframe computer, and VCR) and seven economics concepts (Beveridge curve, Hysteresis, Laffer curve, Monetarism, Okun's law, Ricardian equivalence, Solow residual).

Our findings are shown in Table 1. Seven models are considered. The Gumbel-Max, Frechet-Max and Weibull-Max functions are all special cases of GEV-Max. Gumbel-Max, which is the same as Gompertz, is obtained in the limit as the parameter gamma tends to zero, Frechet-Max is obtained when gamma is positive and Weibull-Max when gamma is negative. The GEV-Max function in all cases here is right skewed. Gumbel-Min, Frechet-Min and Weibull-Min are all special cases of GEV-Min. Again Frechet-Min is obtained when gamma is positive, Weibull-Min when gamma is negative, and Gumbel-Min in the limit as gamma tends to zero. GEV-Min is the mirror image of GEV-Max, and hence in all cases here is left skewed. The final function considered here is Logistic which is symmetric.

The fit of each model is measured by the sum of squared errors (SSE) term. A smaller SSE implies a better fit. Also shown in Table 1 are the estimated alpha, beta, L, inflexion points, and gamma parameters. L is the estimated asymptotic upper bound on total citations. We only consider values of gamma in the range from -0.03 to +0.03 at discrete intervals of 0.003. This is because our aim is to focus on the GEV function in the vicinity of Gompertz (i.e., where gamma equals zero) to see how this can improve the fit of the S-curve.

Since it is symmetric, the point of inflexion for Logistic occurs at $L/2$. The points of inflexion for the other functions are derived from the inflexion times given in section 2.4. The level of cumulative citations at the inflexion point is obtained by substituting the inflexion time t^* into the S-curve function. Cumulative citations at the inflexion point are less than $L/2$ for GEV-Max and more than $L/2$ for GEV-Min. The upper limits L are systematically larger for GEV-Max than for GEV-Min, with the Logistic upper limits lying somewhere in between.

We find that the Logistic function fits best for all five household/computer technologies. Given that these technologies are all subject to obsolescence it is intuitively plausible that they should be described by the substitution model.

For the Economics concepts we find that GEV-Max fits best in all cases except for hysteresis, which like the household/computer technologies is best described by the Logistic function. For four of the Economic concepts (Okun's law, Solow residual, Laffer curve and Monetarism) Frechet-Max fits best, implying a positive value of gamma. Weibull Max is the best fit for the other two (Beveridge curve and Ricardian equivalence).

In all six cases where GEV max fits best, the selected value of gamma is either 0.3 or -0.3 which are the boundaries of the range over which we search. It follows that even better S-curve fits could be found by increasing the allowable range for gamma, and thus moving the best-fit function further still from Gompertz. For cases though where Logistic fits best in Table 1, the gap between the sum of squared errors (SSE) of Logistic and the best GEV function is quite large. Increasing slightly the allowable range of gamma would therefore be unlikely to overturn a finding in favour of Logistic.

What is perhaps most striking is that for none of our 12 data sets is GEV-Min the best fit, thus suggesting that none of them are left skewed and hence that the social-learning diffusion model is not applicable. Intuitively it perhaps makes sense that the right-skewed mortality model (with the exception of hysteresis) fits the Economics concepts best since as topics become less fashionable in an area, researchers who have already worked on the topic before tend to continue to publish in the area but fewer new researchers are attracted to publish on that topic. Thus the number of articles attributed to the topic per year will decrease but more gradually than in a substitution process.

Our actual citation data series and corresponding best-fit S-curves are graphed in Figure 2. In most cases the fit seems quite good. In practice one important consideration is whether the best-fit function tends to change as more data points become available. To investigate this we try truncating our data sets in 2001 and 1996. We find that in both cases all five of our household/computer technologies and Hysteresis remain Logistic. Similarly five of the six Economics concepts with GEV functional forms remain as before. The exception is Solow residual which switches from Frechet-Max to Logistic. This may be because it exhibits late right-tail activity. Truncation of the data set before this point can lead to it mistakenly being identified as Logistic.

In general though, our best-fit models seem to be reasonably robust to truncation of the data set. From a policy perspective it is useful to be able to identify the best-fit function type before the S-curve is complete, since the functional form provides important predictions about the S-curves right tail (such as the asymptotic upper limit).

4. Conclusion

Three well-known models of diffusion of new technologies are the substitution model which is symmetric, the mortality model which is right skewed and the social-learning model which is usually left skewed. The Logistic function is often used to describe the substitution model and the Gompertz function to describe the mortality model. Both functions, however, lack flexibility in that their skewness and kurtosis coefficients are fixed. We have shown here how the Generalized-Extreme-Value (GEV) function, which includes Gompertz as a special case, can be used to improve the flexibility of fitted S-curves. We find that Logistic fits best for 5 of our 12 data sets on cumulative citations (all of which are household/computer technologies), while for the other 7 (all of which are concepts drawn from the Economics literature) GEV-Max outperforms both Logistic and Gompertz for all but one. None of the fitted S-curves are right skewed. Hence each of our data sets seems to be consistent with either the substitution or mortality models, but not with the social-learning model. The selected functional forms seem to be reasonably stable to truncation of the data sets. Once identified, the functional form can be useful for modelling aspects (such as the asymptotic upper limit) of the future path of an emerging technology.

References

[1] S. Kuznets, Secular Movements in Production and Prices, Houghton Mills, Boston MA, 1930.

[2] B. Ryan and N. C. Gross, The Diffusion of Hybrid Corn in Two Iowa Communities, *Rural Sociology* 8(1) 1943 15-24.

[3] Z. Griliches, Hybrid Corn: An Exploration of the Economics of Technological Change, *Econometrica* 25 1957 501-522.

[4] A. L. Porter, A. T. Roper, T. W. Mason, A. F. Rossini and J. Banks, Forecasting and Management of Technology, Wiley Series in Engineering and Technology Management, John Wiley and Sons, USA, 1991.

[5] B. C. Twiss, Forecasting for Technologists and Engineers, *IEE Management of Technology Series* 15. Peter Peregrinus: London, 1992.

[6] P. Young, Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning, *American Economic Review* 99(5) 2009 1899-1924.

[7] M. Bengisu and R. Nekhili, Forecasting Emerging Technologies with the Aid of Science and Technology Databases, *Technological Forecasting and Social Change* 73 2006 835-844.

[8] A. C. Wygant and O. W. Markley, Information and the Future: A Handbook of Sources and Strategies, Greenwood Press: New York, 1988.

[9] C. W. Choo, Information Management for the Intelligent Organization: the Art of Scanning the Environment. Published for the American Society for Information Science, Assist Monograph Series, 2002.

[10] J. P. Martino, A Review of Selected Recent Advances in Technological Forecasting, *Technological Forecasting and Social Change* 70 2003 719-733.

[11] P. H. Franses, A Method to Select Between Gompertz and Logistic Trend Curves, *Technological Forecasting and Social Change* 46 1994 45-49.

[12] J. C. Fisher and R. H. Pry, A Simple Substitution Model of Technological Change, *Technological Forecasting and Social Change* 3 1971 75-88.

[13] S. P. Schnaars, Megamistakes: Forecasting and the Myth of Rapid Technological Change, Free Press: New York and London, 1989.

[14] G. Tichy, Over-optimism Among Experts in Assessment and Foresight, ITA manuscript, ITA-02-05, Oktober, 2002.

[15] A. P. Jeuland, Incorporating Heterogeneity into Parsimonious Models of Diffusion of Innovation, *University of Chicago Center for Research on Marketing Working Paper* 45, 1981.

[16] R. G. King and C. I. Plosser, Real Business Cycles and the Test of the Adelmans, *Journal of Monetary Economics* 33 1994 405-438.

[17] N. S. Balke and M. A. Wynne (1995), Recessions and Recoveries in Real Business Cycle Models, *Economic Inquiry* 33 1995 640-663.

[18] D. Harding and A. Pagan, Dissecting the Cycle: a Methodological Investigation, *Journal of Monetary Economics* 49 2002 365-381.

[19] S. Markose and A. Alentorn, The Generalized Extreme Value (GEV) Distribution, Implied Tail Index and Option Pricing, *Journal of Derivatives* 18(3) 2011 35-60.

[20] R. A. Fisher and L. H. C. Tippett, Limiting Forms of the Frequency Distributions of the Largest or Smallest Member of a Sample, *Proceedings of the Cambridge Philosophical Society* 24 1928 180-190.

[21] P. Embrechts, C. Kluppelberg and T. Mikosch, *Modelling Extreme Events for Insurance and Finance*, Springer, Berlin, 1997.

[22] R. D. Reiss and M. Thomas, Statistical Analysis of Extreme Values, Birkhauser (Second Edition), Basel, 2001.

[23] K. J. Fox, Information Rich Expressions for Model Selection Criteria, *Applied Economics Letters* 7 2000 59-62.

[24] A. DeBecker and T. Modis, Determination of the Uncertainties in S-Curve Logistic Fits, *Technological Forecasting and Social Change* 46(2) 1994 153-173.

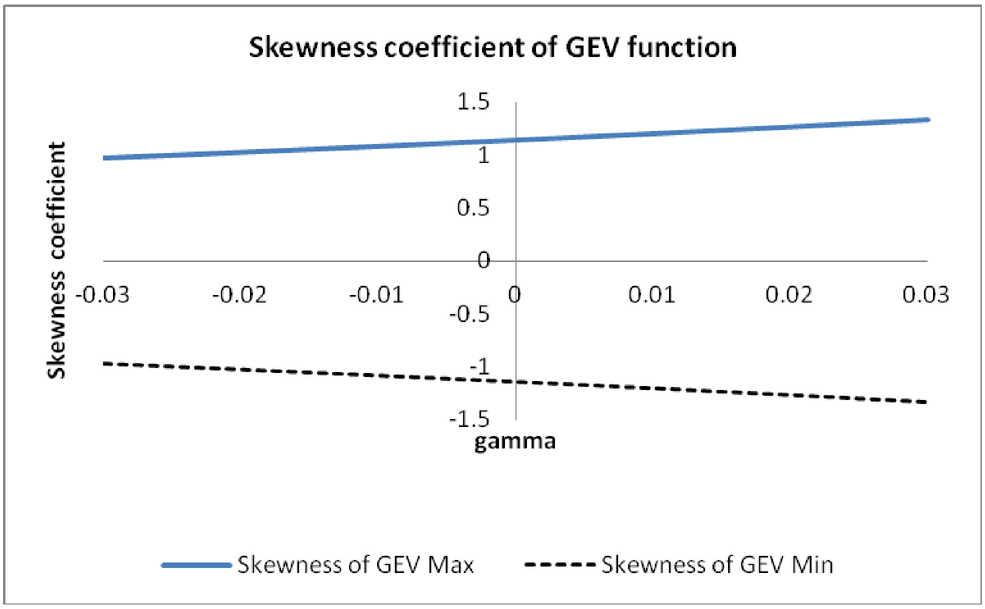
Table 1. S-Curve Sum of Squared Errors and Parameter Estimates

	Gumble Max	Frechet Max	Weibull Max	Logistic	Gumble Min	Frechet Min	Weibull Min
ATM							
SSE	1401575	1385966	1102931	78816	829494	840496	23295731
Alpha	4.07	4.07	4.30	9.99	7.31	7.32	7.15
Beta	0.18	0.18	0.19	0.44	0.30	0.30	0.29
L	6612.01	6612.95	6393.30	5519.34	5468.59	5467.69	5477.56
Inflexion	2432.42	2425.48	2423.59	2759.67	3456.81	3462.27	3401.11
gamma		0.003	-0.03			0.003	-0.03
CDRom							
SSE	14994	15138	14726	11581	70433	96578	1560771
Alpha	2.15	2.15	2.14	4.03	3.17	3.15	3.11
Beta	0.26	0.26	0.26	0.42	0.27	0.26	0.27
L	1395.16	1395.46	1394.56	1351.54	1347.25	1356.57	1349.01
Inflexion	513.25	511.83	516.12	675.77	851.63	872.26	837.62
gamma		0.003	-0.006			0.03	-0.03
Fax							
SSE	2705	2682	2510	904	3421	3460	92479
Alpha	2.42	2.42	2.48	5.57	4.25	4.25	4.17
Beta	0.12	0.12	0.12	0.26	0.18	0.18	0.17
L	335.09	335.12	326.87	274.42	271.00	270.95	271.52
Inflexion	123.27	122.91	123.91	137.21	171.30	171.57	168.59
gamma		0.003	-0.03			0.003	-0.03
Mainframe computer							
SSE	24728	24603	23121	19867	73773	74506	1619611
Alpha	2.54	2.55	2.59	5.28	4.05	4.05	3.96
Beta	0.12	0.12	0.13	0.24	0.16	0.16	0.15
L	1188.02	1188.21	1166.94	1058.59	1051.88	1051.64	1054.25
Inflexion	437.05	435.81	442.37	529.29	664.91	665.92	654.60
gamma		0.003	-0.03			0.003	-0.03
VCR							
SSE	1442	1437	1360	695	3520	3587	220616
Alpha	1.75	1.75	1.77	3.85	3.21	3.21	3.15
Beta	0.10	0.10	0.10	0.20	0.15	0.15	0.15
L	520.40	520.44	508.36	418.35	399.69	399.62	400.40
Inflexion	191.45	190.89	192.71	209.18	252.65	253.05	248.62
gamma		0.003	-0.03			0.003	-0.03
Beveridge curve							
SSE	149.26	148.38	134.36	160.74	511.17	522.35	37714.34
Alpha	1.96	1.96	2.01	4.36	3.71	3.71	3.64
Beta	0.11	0.11	0.1	0.23	0.18	0.18	0.18
L	254.50	254.52	245.46	197.68	182.64	182.61	182.95
Inflexion	93.63	93.35	93.05	98.84	115.45	115.64	113.60
gamma		0.003	-0.03			0.003	-0.03

Table 1. S-Curve Sum of Squared Errors and Parameter Estimates (continued)

	Gumble Max	Frechet Max	Weibull Max	Logistic	Gumble Min	Frechet Min	Weibull Min
Hysteresis							
SSE	106420	105239	94652	16521	84585	85692	2210993
Alpha	2.99	2.99	3.11	7.78	5.81	5.82	5.70
Beta	0.10	0.10	0.11	0.28	0.19	0.19	0.18
L	1848.44	1848.56	1779.78	1379.37	1358.56	1358.30	1361.12
Inflexion	680.00	678.01	674.68	689.68	858.77	860.11	845.14
gamma		0.003	-0.03			0.003	-0.03
Laffer curve							
SSE	7439	7426	7445	7611	7778	8354	260916
Alpha	1.67	1.64	1.67	3.72	3.40	3.44	3.33
Beta	0.07	0.07	0.07	0.16	0.15	0.15	0.15
L	805.16	839.54	800.88	574.42	449.59	438.80	453.91
Inflexion	296.20	299.72	295.51	287.21	284.19	277.85	281.83
gamma		0.03	-0.003			0.003	-0.03
Monetarism							
SSE	6391	5798	6481	40243	131606	132877	2312891
Alpha	1.93	1.92	1.95	3.75	2.96	2.96	2.91
Beta	0.12	0.12	0.12	0.21	0.13	0.13	0.13
L	1259.16	1271.26	1248.56	1177.10	1172.66	1172.50	1174.33
Inflexion	463.21	453.84	463.47	588.54	741.26	742.45	729.16
gamma		0.03	-0.009			0.003	-0.03
Okun's law							
SSE	2166	1995	2182	13074	36973	37428	949041
Alpha	1.83	1.81	1.83	3.77	3.07	3.08	3.03
Beta	0.07	0.07	0.07	0.14	0.09	0.09	0.09
L	771.53	784.84	771.46	677.06	661.94	661.84	662.96
Inflexion	283.83	280.19	284.65	338.52	418.42	419.09	411.64
gamma		0.03	-0.003			0.003	-0.03
Ricardian equivalence							
SSE	3391	3622	3381	7560	22691	22976	580146
Alpha	2.22	2.20	2.25	4.63	3.69	3.70	3.63
Beta	0.12	0.12	0.13	0.24	0.17	0.17	0.16
L	790.29	796.71	776.85	682.18	666.69	666.57	667.85
Inflexion	290.73	292.21	294.48	341.09	421.42	422.08	414.68
gamma		0.003	-0.03			0.003	-0.03
Solow residual							
SSE	562.14	544.38	561.24	1367.29	2748.59	2778.89	73052.34
Alpha	1.79	1.76	1.79	3.67	3.07	3.07	3.02
Beta	0.15	0.14	0.15	0.27	0.20	0.20	0.20
L	317.17	325.00	317.14	276.82	264.34	264.31	264.69
Inflexion	116.68	116.02	117.02	138.41	167.09	167.36	164.34
gamma		0.03	-0.003			0.003	-0.03

Figure 1. Skewness and Kurtosis Coefficients of GEV Max and Min



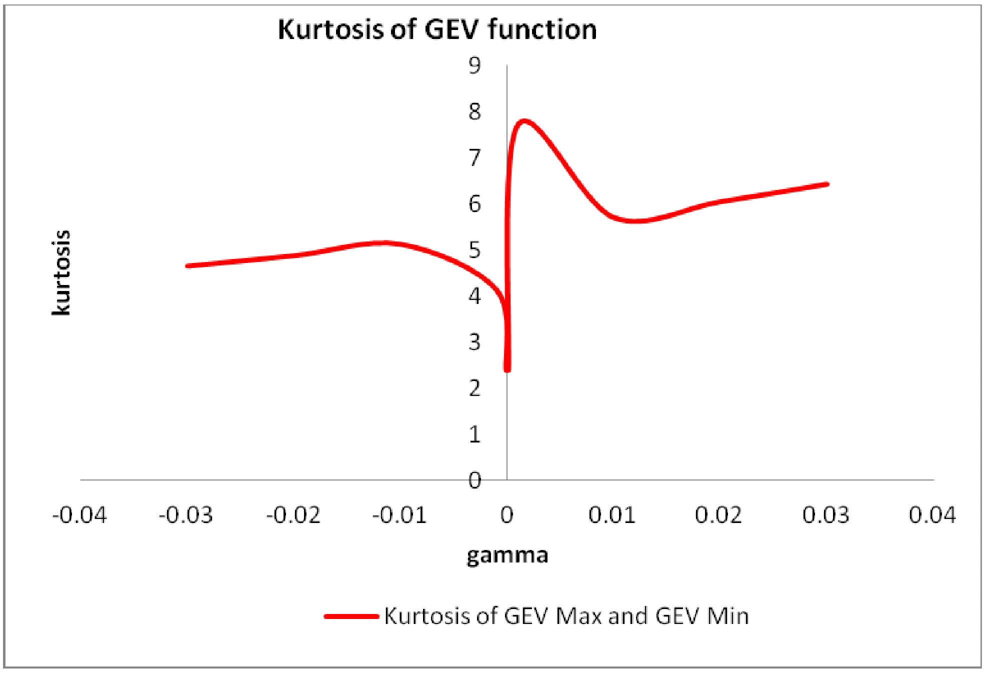


Figure 2. Cumulative Citations and the Best Fit S-Curve

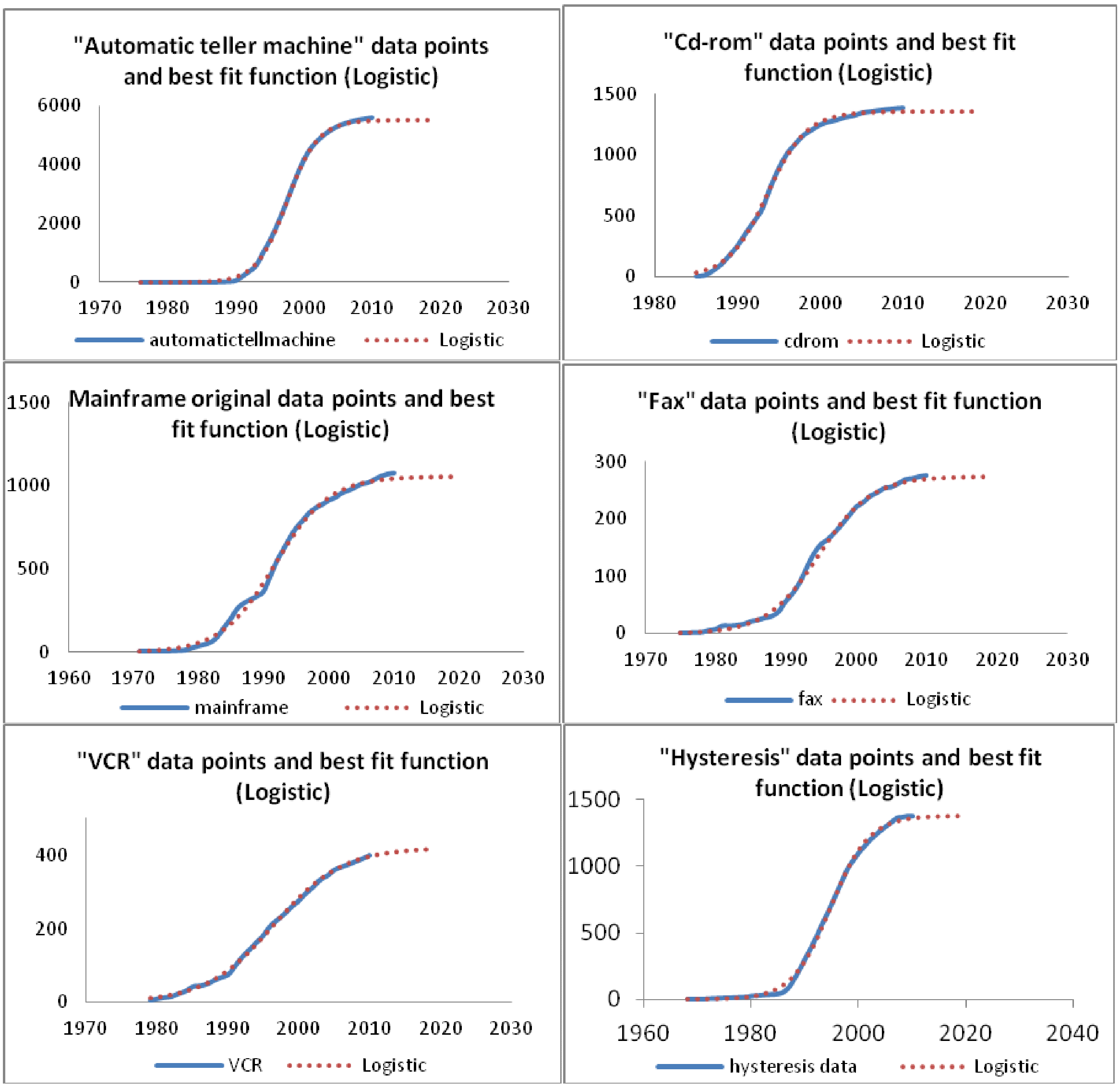
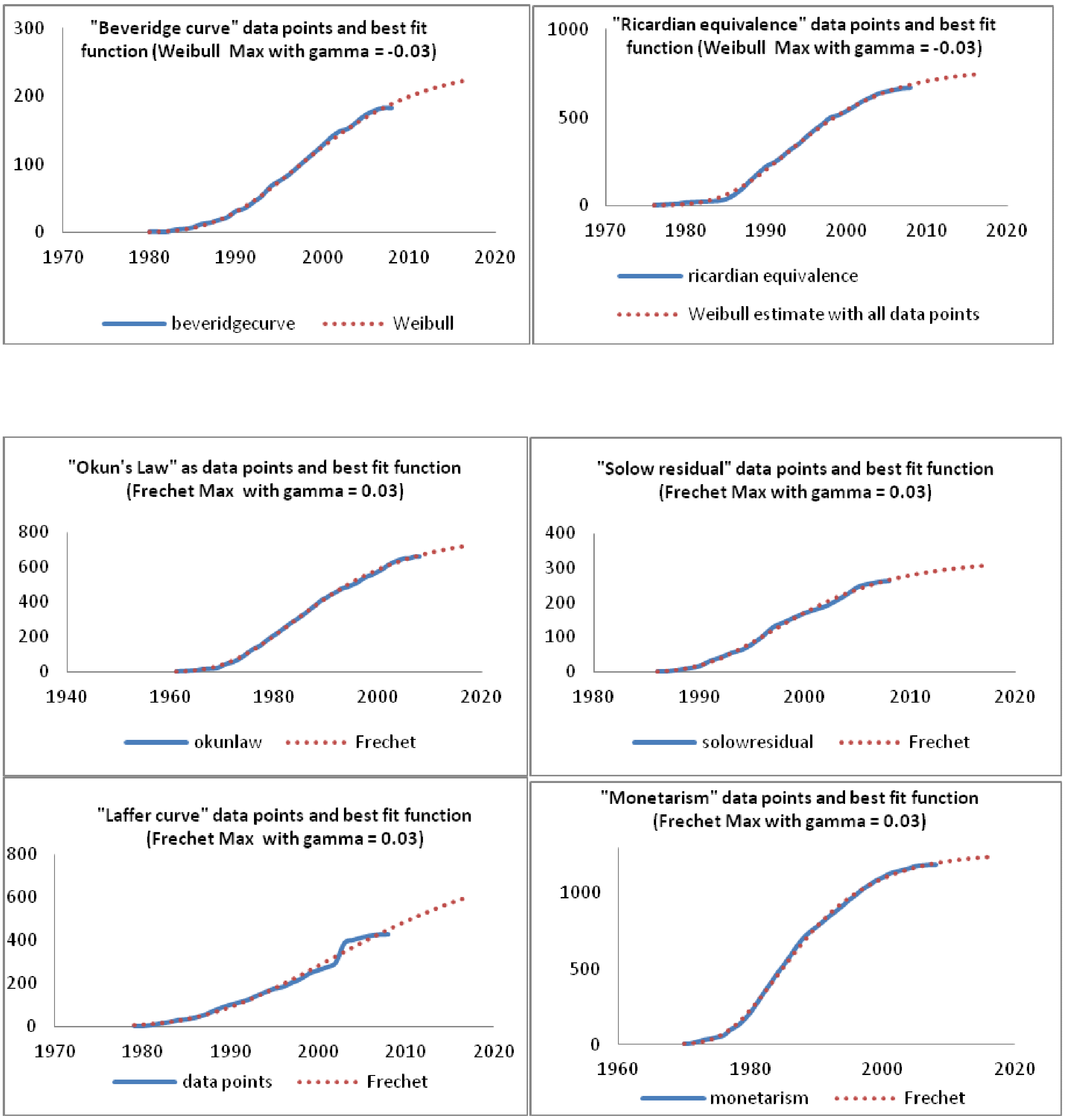


Figure 2. Cumulative Citations and the Best Fit S-Curve (Continued)



Graz Economics Papers

For full list see:

http://www.uni-graz.at/vwlwww/vwlwww_forschung/vwlwww_gep.htm

Address: Department of Economics, University of Graz,
Universitätsstraße 15/F4, A-8010 Graz

05–2012 **Miriam Steurer, Robert J. Hill, Markus Zahrnhofer, Christian Hartmann**: Modelling the Emergence of New Technologies using S-Curve Diffusion Models

04–2012 **Christian Groth, Ronald Wendner**: Embodied learning by investing and speed of convergence

03–2012 **Bettina Brggemann, Joern Kleinert, Esteban Prieto**: The Ideal Loan and the Patterns of Cross-Border Bank Lending

02–2012 **Michael Scholz, Jens Perch Nielsen, Stefan Sperlich**: Nonparametric prediction of stock returns guided by prior knowledge

01–2012 **Ronald Wendner**: Ramsey, Pigou, heterogenous agents, and non-atmospheric consumption externalities