BAM Attachment 9 – Gompertz Penetration Rate

Penetration Rate Analysis

In order to determine penetration rates of new broadband deployments in unserved areas, we chose to perform a combination of several statistical and regression analyses. Our primary data source was a table of Home Broadband Adoption metrics from the Pew Internet & American Life Project. Since 2001, the Pew Research Center has conducted extensive, anonymous phone surveys on broadband adoption in the United States, breaking out responses by various demographics. Their surveys reveal positive and negative correlation factors between certain demographic characteristics and broadband adoption[1]. The Pew study notes the most significant factors are, in order of importance:

Positively Correlated	Negatively Correlated
Income greater than $100K	Less than high school education
Income between $75K – $100K	Senior citizen (65+)
College degree or greater education	Rural
	High school degree only

We obtained the results of the Pew study on broadband adoption covering 19 survey periods from October 2001 to November 2009. This data aggregated adoption percentages in each period by the following demographics:

- Overall
- Race
- Income
- Age
- Education Level
- Rural or Non-Rural

Preliminary findings of the data revealed that the trends in broadband adoption matched those of standard technology adoption lifecycles. Our approach to this analysis was to understand the shape and characteristics of the Pew adoption curves in an attempt to model the results into a mathematical model, by which future broadband adoption could then be measured. We began by examining two popular mathematical models used to forecast technology adoption: the Fisher-Pry model and the Gompertz model[2]. The table below highlights the differences between the two.

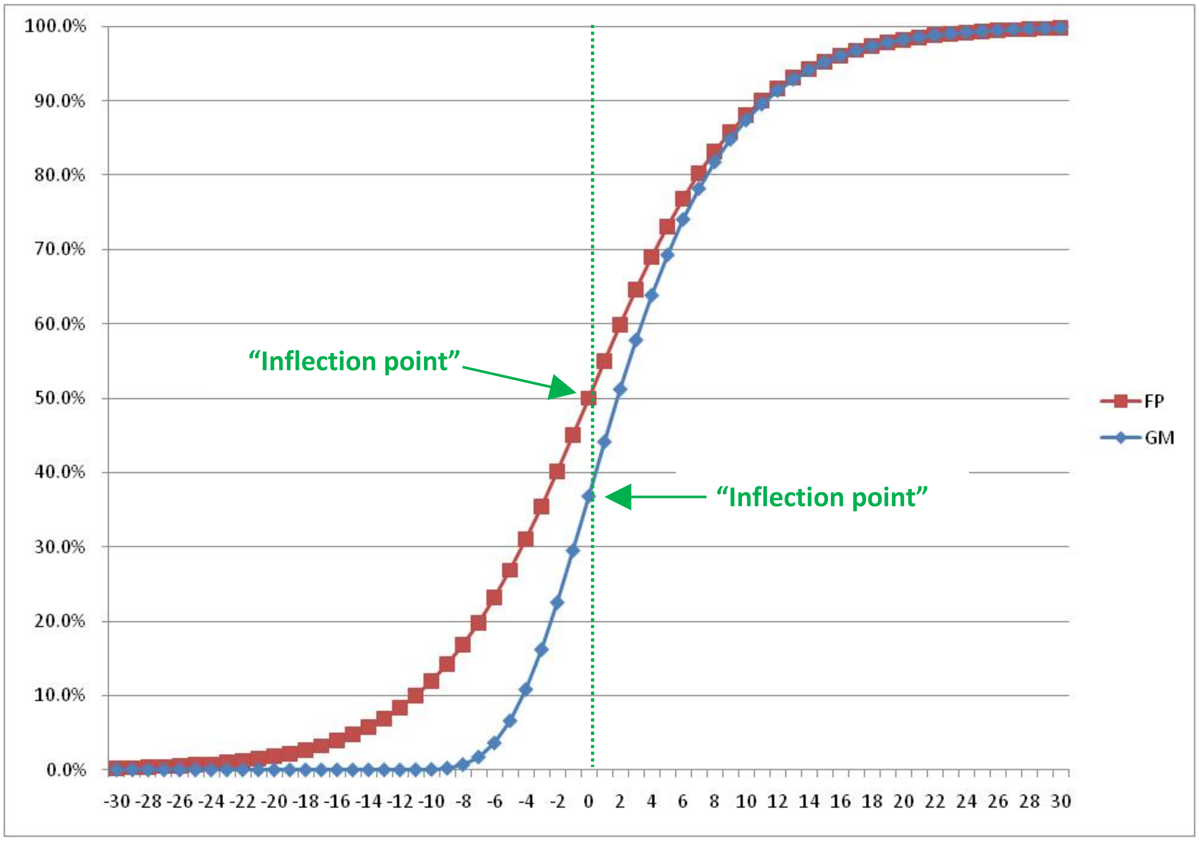
---

[1] Horrigan, John. Home Broadband Adoption 2009. Pew Internet & American Life Project: June 2009. http://pewinternet.org/Reports/2009/10-Home-Broadband-Adoption-2009.aspx

[2] Vanston, Lawrence K. and Vanston, John H. Introduction to Technology Market Forecasting. Austin, TX: Technology Futures, Inc, 1996.

Model	Equation	When Used	Examples
Fisher-Pry	$y = \frac{1}{1+e^{-b(t-a)}}$	When substitution is driven by superior technology. The new product or service presents some technological advantage over the old one.	Internal combustion engine, telegraph, light bulbs
Gompertz	$y = e^{-e^{-b(t-a)}}$	When substitution is driven by superior technology, but purchase depends on consumer choice.	Digital television, mobile phones

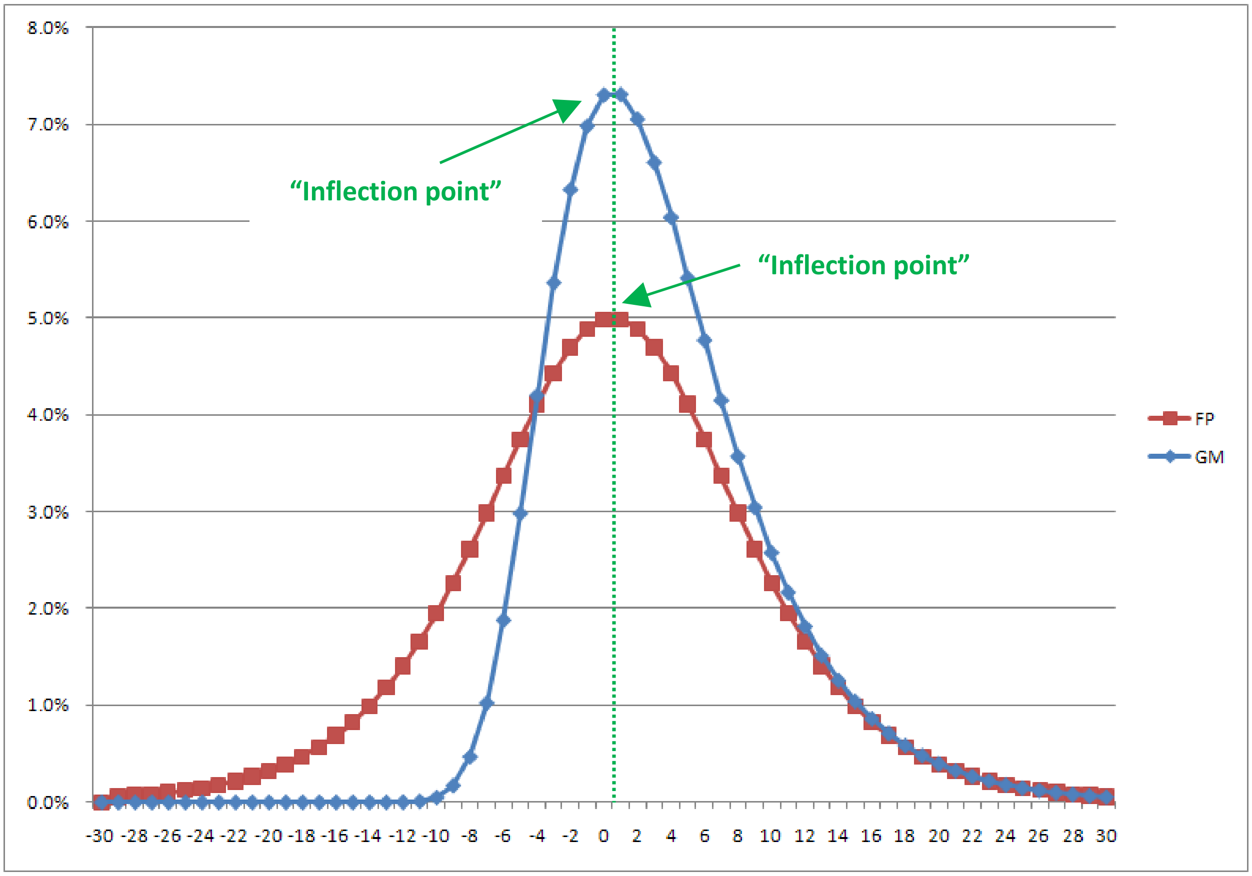
The table below illustrates the <u>cumulative</u> characteristics of the two models as a percentage of the installed base:



"Inflection point"

"Inflection point"

From an <u>incremental</u> standpoint, the period-to-period technology adoption unfolds as follows:



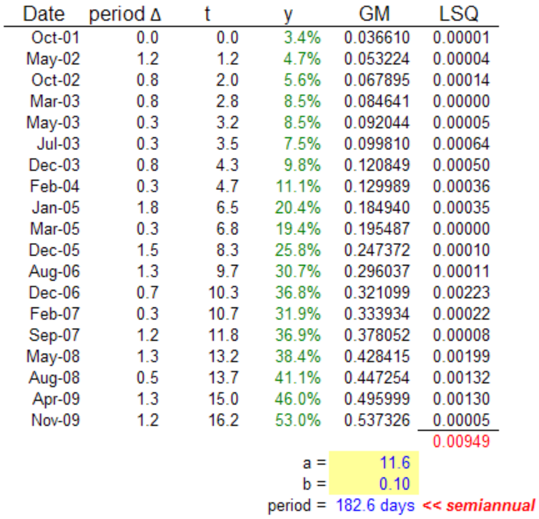
Both models have a characteristic "inflection point" – that is, the point at which the incremental curve is maximized and the cumulative curve flips over[3]. The inflection point should be considered the point where technology adoption reaches its maximum growth rate. The characteristics of the Fisher-Pry and Gompertz models are indicative of the assumptions underlying each model; that is, the Fisher-Pry model reflects revolutionary displacing technology, while the Gompertz curve supports incremental technology change[4]. Consequently, the Fisher-Pry model will inflect at a higher cumulative percentage (50%) than the Gompertz model (36.8%). Also, the adoption will scale much earlier within the chosen timeframe in the Fisher-Pry model than the Gompertz model. Our decision on which model to use for broadband adoption is predicated on how broadband is perceived in the market: is it a revolutionary technology or an iterative one? As innovative as broadband technology has become, it is at its core a replacement for dialup technology (albeit one that is faster and allows access to a larger subset of the Internet). Ultimately, the choice of whether or not to adopt broadband is a consumer choice: broadband, dialup, or nothing? Only the consumer can decide what fits his or her particular needs. Contrast this decision to the light bulb, the antithesis of consumer choice. All things considered, we chose to forecast broadband adoption with the Gompertz model for purposes of this analysis.

---

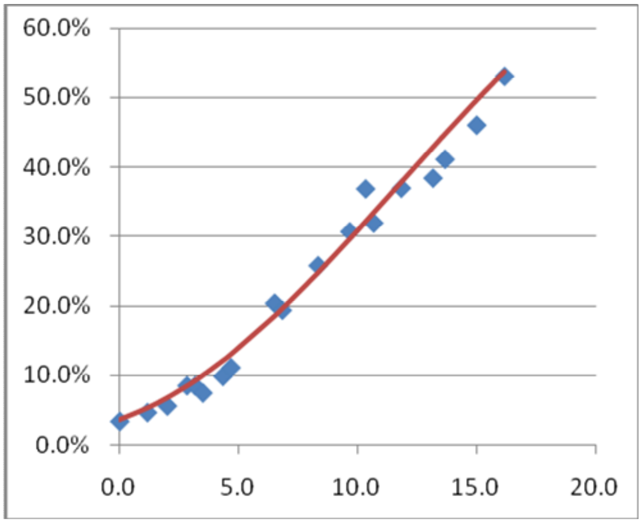
[3] Geometrically speaking, the inflection point on the cumulative curve is the point at which the curve moves from convex to concave. For both the Fisher-Pry and Gompertz models, the slope of the tangential line along the cumulative curve is highest at the inflection point, indicating maximum acceleration of adoption.

[4] Vanston, Lawrence K. and Vanston, John H. Introduction to Technology Market Forecasting. Austin, TX: Technology Futures, Inc, 1996.

Our analysis of the Pew data consisted of fitting each demographic data breakout (Overall, Race, Income, Age, Education Level, Rural/Non-Rural) into a Gompertz curve using a least squares approach[5]. With a semiannual time period adjustment, the results indicated the Pew data segments could be fit on a corresponding Gompertz cumulative curve with very reasonable least squares accuracy. One such curve fit for a particular demographic (rural populations) is shown below.



Date	period Δ	t	y	GM	LSQ
Oct-01	0.0	0.0	3.4%	0.036610	0.00001
May-02	1.2	1.2	4.7%	0.053224	0.00004
Oct-02	0.8	2.0	5.6%	0.067895	0.00014
Mar-03	0.8	2.8	8.5%	0.084641	0.00000
May-03	0.3	3.2	8.5%	0.092044	0.00005
Jul-03	0.3	3.5	7.5%	0.099810	0.00064
Dec-03	0.8	4.3	9.8%	0.120849	0.00050
Feb-04	0.3	4.7	11.1%	0.129989	0.00036
Jan-05	1.8	6.5	20.4%	0.184940	0.00035
Mar-05	0.3	6.8	19.4%	0.195487	0.00000
Dec-05	1.5	8.3	25.8%	0.247372	0.00010
Aug-06	1.3	9.7	30.7%	0.296037	0.00011
Dec-06	0.7	10.3	36.8%	0.321099	0.00223
Feb-07	0.3	10.7	31.9%	0.333934	0.00022
Sep-07	1.2	11.8	36.9%	0.378052	0.00008
May-08	1.3	13.2	38.4%	0.428415	0.00199
Aug-08	0.5	13.7	41.1%	0.447254	0.00132
Apr-09	1.3	15.0	46.0%	0.495999	0.00130
Nov-09	1.2	16.2	53.0%	0.537326	0.00005
					0.00949
			a =	11.6	
			b =	0.10	
			period =	182.6 days	<< *semiannual*

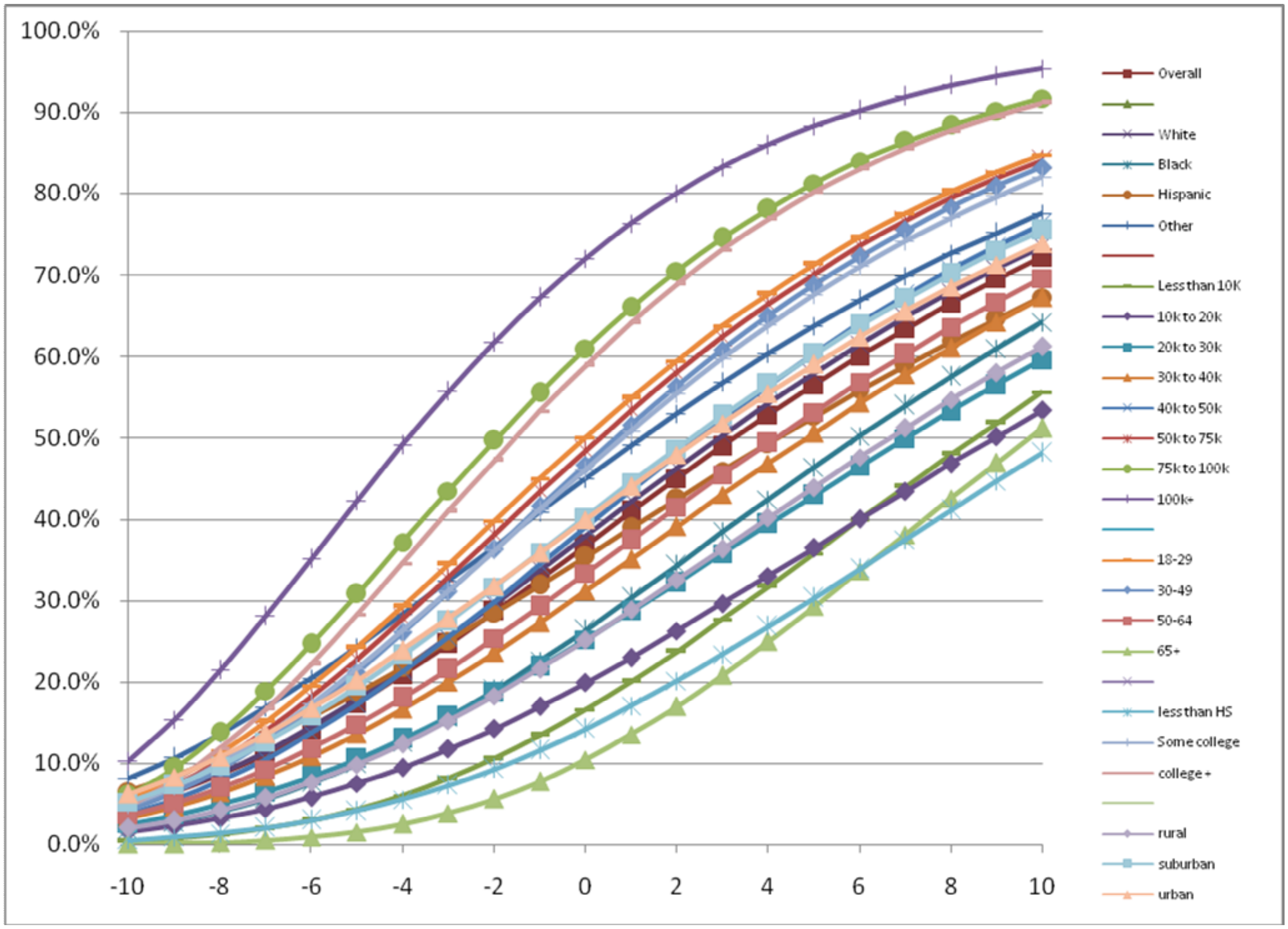


To the right is a table with the survey date, period delta, time (t) in semiannual periods, Pew results for rural populations (y), the assumed Gompertz value (GM) given the ***a*** and ***b*** values shown in yellow, and the squared difference value between (y) and (GM). The process begins by performing iterative analysis to randomly select ***a*** and ***b*** coefficients simultaneously. The process stops after all possible value ranges for ***a*** and ***b*** have been exhausted. The combination of ***a*** and ***b*** values that minimize the sum of the least squares represents the best fit of a Gompertz curve for rural populations. This particular demographic analysis yields a least squares sum of 0.00949. The graph to the right shows the actual Pew numbers (y) represented by the blue points, while the Gompertz curve fit (GM) is represented by the fitted red line. The fit to the observed data can be visually inspected here.

Our analysis provided us with Gompertz curves by demographic. However, consider that the Pew research starts with an arbitrary date of October 2001. This date does not presume the "start" of broadband; it only represents the date at which surveys began. Therefore we must provide a time-based adjustment for every demographic curve. The solution we determined as most appropriate was to develop a series of demographic adoption curves *relative* to the overall adoption curve. The graph below illustrates the relative Gompertz curve fits for every demographic segment. Here, the overall adoption curve inflects at zero on an adjusted time scale.

---

[5] The best fit, between modeled data (Gompertz) and observed data (Pew), in its least-squares sense, is an instance of the model for which the sum of squared residuals has its least value, where a residual is the difference between an observed value and the value provided by the model.



Reinforcing the conclusions of the Pew study, the Income over $75K and College or Greater Education curves are farthest to the left (representing more rapid adoption relative to the mean), while the High School or Less, Rural, and 65+ curves are farthest to the right (representing slower adoption relative to the mean).

It is noteworthy to mention these curves, because they are based upon a nationwide introductory broadband rollout where services have not yet been deployed, represent a Greenfield build. In Brownfield deployments, however, builders are leveraging previous deployments to capture consumers who have already been educated on the benefits of broadband. We therefore allow for an additional time adjustment where Brownfield builds are taking place. This adjustment is discussed later.

These results provide relative Gompertz curves by every demographic measured in the Pew study; however for a number of reasons, we chose to limit the prediction model to only the demographic factors with the largest positive and negative correlation to broadband adoption. While it would technically be possible to measure adoption changes across all the available demographics on the Pew study, it is not cost effective to do so – either the remaining demographics had minimal influence on broadband adoption, or the demographic data in question was not readily available at the census block level.

The census block demographic variables we chose as to predict broadband adoption are the following:

- Income greater than $100K
- Income between $75K – $100K

- College degree or greater education
- Senior citizen (65+)
- Less than high school education
- Rural
- High school degree only

Using the Gompertz coefficients for each demographic, combined with demographic data at the Census block level, we can build Gompertz curves for every Census block in the nation. To build these custom curves, we weight the demographic Gompertz coefficients (***a*** and ***b***) by the incremental demographics prevalent in the area. For example, if the demographics within the overall curve show 18.5% of households have incomes above $100K, but a particular census block contains 20% households with over $100K income, each “Over $100K” Gompertz coefficient would be weighted by the incremental difference (20% – 18.5% = 1.5%) and added to the overall Gompertz coefficient. By summing up the weightings off each significant variable, our Gompertz equation for each census block would take shape.

The additional step in forecasting broadband penetration rate is to determine how to factor in a Brownfield effect, if any, into the census block time coefficient (***a***). If the Census block was revealed to have a prior broadband deployment, the census block curve would be shifted left a designated number of periods. The number of periods to shift is held constant across all Brownfield deployments.

The final step of developing the census block curve was to determine where to set the inflection point. The zeroed scale is intended to represent the point at which the overall curve inflects, but the time at which the scale hits zero must be determined. The FCC initially chose this scale to be 2 years from the start of deployment; essentially, the overall broadband adoption would reach its maximum growth rate in 24 months. To account for the initial mass influx of customers in the first 24 months, we chose to start with zero subscribers at initial deployment, then trend towards the number of subscribers at 24 months by dividing them into four equal 6-month periods of subscriber adoption. After 24 months, the penetration rates reflected in the Gompertz curve would be in effect. The selection of an inflection point, while initially set at 24 months, is one that can potentially be re-examined and adjusted as needed.

Additional Factors

The resulting census block penetration rate determines the standard broadband adoption rate for that census block. It does not, however, factor in the subscribers of related services (voice, video), the effect of bundled services, or the stratification of tiering (basic vs. premium). To account for each of these we developed factors from which we could adjust the baseline number of expected broadband adopters in every census block. Each factor is discussed below.

Scaling Factor

A scaling factor, in this instance, refers to a multiplying factor developed to predict voice and video subscribers by technology (DOCSIS, FTTp, FTTn, FTTd, and Fixed & Mobile Wireless) based on the number of broadband subscribers. The presumption is that each technology exhibits a constant and unique relationship between broadband subscribers and subscribers to other services like voice and video. In other words, if you know the number of broadband subscribers for a particular technology, you can predict the number of voice or video subscribers as well. Our analysis of industry data affirmed that this relationship is constant and unique for each technology. In this sense, broadband leads the way for technology adoption of ancillary services.

Bundling Percentages

Customers who subscribe to broadband services belong to one of two groups: those that purchase a la carte, or those that purchase as a bundle. Industry analysis confirmed that the relationship between the two subscriber bases is relatively constant for each technology. Using this data we developed a "bundling" percentage based on the broadband subscribers to arrive at the number of bundled subscribers. The number of bundled customers could then be subtracted from the total number of voice and video subscribers to arrive at the number of a la carte subscribers for each.

Tiering Percentages

Tiering, in this case, refers to the tiered services offered by carriers. To limit unneeded complexity we limit the number of tiers in the model to two levels: an introductory, basic level of service; and a "top-shelf" premium service. These low/high tiers are applicable to video (for example, basic vs. premium cable), data (entry-level vs. top-speed), even bundles. Using industry data we were able to develop percentages by technology which break out the respective service subscribers into low-end and high-end tiers. These "tiering" percentages were then applied to the number of broadband, video, and bundled subscribers to arrive at low-tier subscribers and high-tier subscribers for each.