FUTURE GROWTH OF CABLE TELEVISION

Rolla Edward Park

December 1970

P-4527

FUTURE GROWTH OF CABLE TELEVISION*

Rolla Edward Park

The RAND Corporation, Santa Monica, California

The growth of cable television raises a real dilemma for public policy. On the one hand, it holds out the promise of more diverse programming made possible, even promoted, by the ability of cable to carry a large number of signals. On the other, it poses a threat to over-the-air television broadcasting. When cable carries distant signals, it fragments the local audience, tending to reduce local station revenue. Some stations might be forced off the air, reducing the amount of service available to cable non-subscribers. The loss to those viewers who are in areas not served by cable, and to those who cannot afford the cable subscription fee, might be considerable.

---

*This paper is based on work supported by a Ford Foundation grant and reported more fully in my RAND R-587, "Potential Impact of Cable Growth on Television Broadcasting," October 1970. It should not, however, be interpreted as the official opinion of either the Ford Foundation or The RAND Corporation.

In 1970, roughly 7.5 percent of U.S. television homes were cable subscribers,[1] and this figure is increasing rapidly. Over the past decade, the number of subscribers has increased at an average annual rate of about 21 percent,[2] while the number of cable systems has grown by 14 percent per year.[3]

The future impact of cable on television broadcasting certainly depends on how far this remarkable growth continues. If 90 percent of all television homes ultimately subscribe to cable service, the impact will obviously be greater than if only 30 percent do. This paper presents estimates of average cable penetration levels ultimately to be expected. These estimates suggest that ultimate penetration may be on the order of 40 to 45 percent nationwide.

LOGISTIC GROWTH CURVE

The logistic growth curve,

$$Y = e^{\alpha-\beta/T} , \tag{1}$$

is frequently used to represent growth processes. The size of the growing entity is denoted by Y, T denotes time since growth began, e is the base of natural logarithms, and α and β are parameters. This curve is sketched in Figure 1. The shape of the curve makes clear its relevance for many growth processes. The entity grows slowly at first, then at an increasing absolute rate as it gets bigger. As it

---

[1] There were 4,500,000 cable subscribers in January 1970 and 59,388,600 television households in September 1969 (ARB figure) according to Television Factbook, 1970-1971 Edition, No. 40, Services Volume, published by Television Digest, Inc., Washington, D.C., 1970, pp. 66-a and 84-a.

[2] From 650,000 in 1960 to 4,500,000 in 1970. Television Factbook, p. 66-a.

[3] From 640 in 1960 to 2,350 in 1970. Television Factbook, p. 66-a.

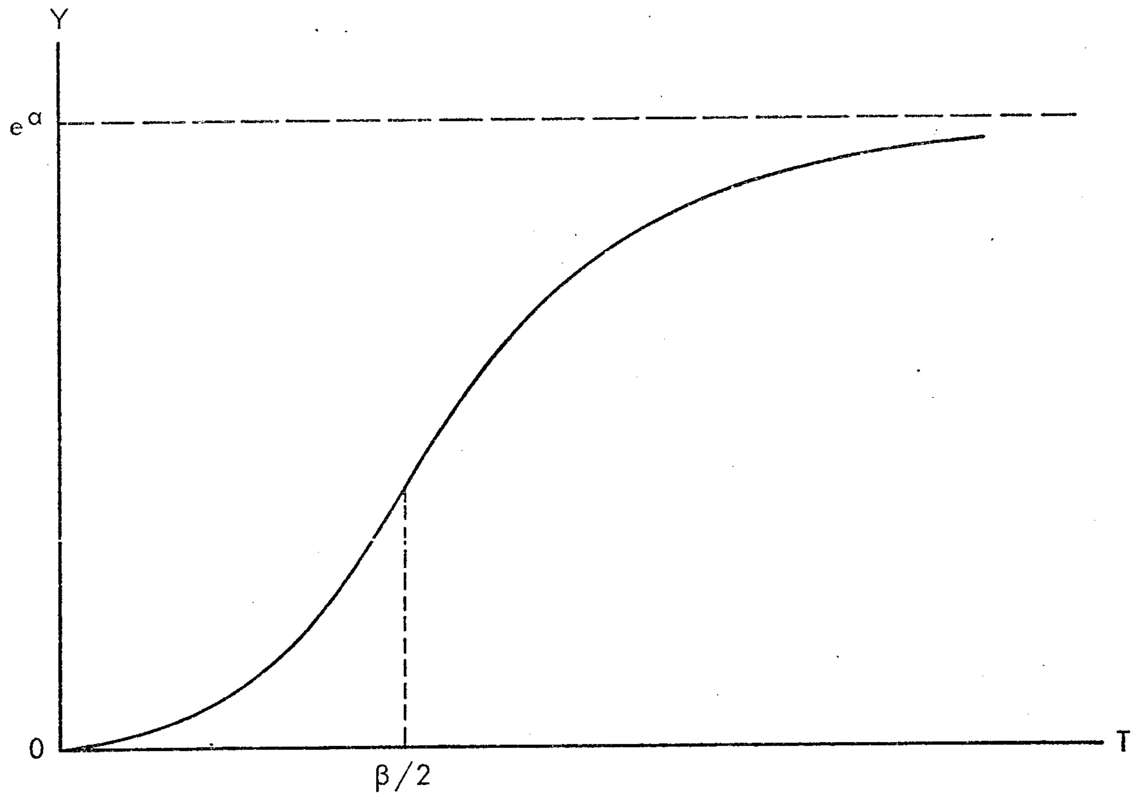

Fig. 1 — Logistic growth curve

approaches its mature size, its growth rate begins to decrease. Finally, it approaches its ultimate size asymptotically.

Qualitatively, at least, the logistic curve is a good descriptor of growth of many entities whose ultimate size is limited, for example, a tree, or a colony of fruit flies in a finite container, or a cable television system with a finite service area. Figure 1 is easily interpreted in terms of a cable system. The new system gets off to a relatively slow start for a variety of reasons: since it is new to the community, many people do not know of its existence or understand the service it provides. Nor is the firm likely to be staffed to sustain a maximal growth rate from the beginning. But as the system grows, the word gets around. More people now have neighbors who subscribe, and so know first hand about the service. The growth rate picks up. Perhaps additional installers and sales personnel are hired. At some point, though, most of the easy sales have been made. The growth rate slows as fewer and fewer potential subscribers remain unsigned. The system slowly approaches its ultimate size, with all households that desire service being served.

THE MODEL

To estimate ultimate cable penetration levels, I fit a set of logistic growth curves to data on actual cable systems. For this purpose, equation (1) needs some embellishment.

When (1) is applied to a cable system, Y denotes the number of subscribers. Obviously, the ultimate number of subscribers, e^{α}, will vary from system to system, depending on the number of households in the system's service area, the type of service offered, and other factors. To account for this, I specify

$$e^{\alpha} = F_i H e^{u} \quad . \tag{2}$$

Here H is the number of households in the service area; F_i is the fraction of all households expected ultimately to subscribe if the system offers service of types i; u is an error term introduced to represent the influence of all other factors — income, availability of alternative entertainment, variations in taste and in cable system management, for example.

Type of service is defined initially in terms of the numbers of local and distant signals carried by the system. If other things were equal, one would expect a system that carried many distant signals to have a higher ultimate penetration than one with few distant signals. Also, other things again equal, one would expect that a system in an area with more local signals would have a lower ultimate penetration than one in an area with fewer. If local signals are abundant, distant signals available only on the cable offer less incentive to subscribe. "Service of type i" is specified more concretely in the following section, as are the other variables in the model.

Referring to Figure 1, we note that the parameter β is a measure of how stretched out the growth curve is. The larger is β, the longer is the time until the inflection point on the growth curve is reached. This parameter, too, may be expected to vary from system to system. I expect it to be larger the more households there are in the system's service area, and so specify

$$\beta = \beta_1 + \beta_2 H + \beta_3 H^2 \quad . \qquad (3)$$

The H^2 term is included to allow for possible curvature in the relation, there being no reason a priori to expect it to be linear. Admittedly, (3) should be a stochastic relation like (2). I omit the error term, making the relation deterministic, for pragmatic reasons: to make it possible to estimate the resulting equation, (4) below, using conventional methods.

Substituting from (2) and (3) in (1) and taking natural logarithms, we get

$$\log Y = \log F_i + \log H + \beta_1(-1/T) + \beta_2(-H/T) + \beta_3(-H^2/T) + u \quad . \qquad (4)$$

With some slight additional manipulation, and assuming that the errors u are distributed independently (of each other and of the independent variables) with zero mean and constant variance, (4) is an appropriate subject for ordinary least squares estimation. The next section discusses the data used to estimate (4), and the section after that presents the estimates themselves.

THE DATA

Data used are for a cross section of cable systems as of February 1969, taken from the 1969-1970 Television Factbook.[1] The sample includes all 46 of the systems listed as having 10,000 or more subscribers, plus every sixth listed operational system with fewer than 10,000 subscribers, read from a randomly chosen starting point.[2] Since the listing is alphabetical by state, the geographical distribution of systems in the sample is the same as that of all listed systems. The total number of systems in the sample is 416.

Entries in the Factbook usually include a list of television stations carried by the cable system. By referring to maps in the CATV Atlas,[3] one can usually determine which of the stations are distant signals (that is, carried by a cable system outside the station's Grade B contour) and which are local signals. The number of systems for each combination of distant and local signals is shown in Figure 2. The systems represented in Figure 2 total 395; for 21 of the systems in the gross sample, stations carried are not listed in the Factbook.

The systems in Figure 2 are divided into six groups, each providing a roughly homogeneous type of service. Assignments to groups are made based on a priori judgment, and on the need to have a minimum of thirty or so systems in each group to get good estimates. Analytically, the most important division is that between systems with two or fewer local signals, that is, systems in areas where a full network lineup is not available over the air, and systems with three or more local signals. In the former case, the cable typically carries the missing network signal or signals, presumably making cable service especially attractive. The other divisions, shown by lines in the figure, are chosen with less a priori justification primarily so that systems are well distributed among different types of services.

---

[1] Television Factbook, pp. 363a-591a.

[2] Alexander City, Alabama.

[3] CATV and Station Coverage Atlas, 1968-1969, published by Television Digest, Inc., Washington, D.C., 1968.

Distant signals \ Local signals	0	1	2	3	4	5	6	7	8	9	10 +
10 +	1	2	4	0	2	0	0	0	0	0	0
9	3	2	1	2	0	0	0	0	0	0	0
8	4	1	6	2	2	2	0	2	0	0	0
7	4	7	3	7	3	2	0	0	0	1	0
6	6	5	7	6	5	5	1	0	0	1	0
5	13	6	7	7	6	7	2	1	0	1	0
4	9	18	9	9	8	6	2	0	3	0	0
3	3	5	23	9	3	4	4	2	2	0	2
2	0	3	2	10	6	5	2	5	1	1	2
1	2	1	4	6	6	0	1	0	3	3	5
0	0	0	0	6	7	7	2	14	4	3	14

Fig. 2 — Number of systems carrying different combinations of local and distant signals

To obtain estimates of the asymptotic penetration levels for each type of service, the F_i in equation (4), dummy variables, D_i, are defined corresponding to the service classifications as follows. $D_i = 1$ if:

i	Type of Service	
1	Local signals ≤ 2	Distant signals ≤ 3
2	Local signals ≤ 2	Distant signals ≥ 4
3	$3 \leq$ Local signals ≤ 6	Distant signals ≤ 3
4	$3 \leq$ Local signals ≤ 6	Distant signals ≥ 4
5	Local signals ≥ 7	Distant signals $= 0$
6	Local signals ≥ 7	Distant signals ≥ 1 .

Otherwise, $D_i = 0$. A system with service of type 1, for example, is represented by $D_1 = 1$, $D_2 = \ldots = D_6 = 0$.

The number of households variable, H, is constructed in the following manner. The Factbook listing usually includes population of the system's service area. This figure is converted to number of households by dividing by the average number of persons per household in the state in which the system is located. Average persons per household, in turn, is calculated from census data[1] by dividing state population by number of occupied dwelling units in the state.

The time variable, T, is calculated from the Factbook listing, which usually includes the date that the system began service. Time in months from begin-service date to February 1969 is the value used for T.

Finally, the number of subscribers, Y, is taken directly from the Factbook listing.

Because subscribers, population, or begin-service date is missing from some listings, the usable sample is further reduced to 352 observations.

[1] County and City Data Book, 1967, U.S. Department of Commerce, Washington, D.C., 1967.

THE ESTIMATES

Making use of the dummy variables defined in the preceding section, (4) can be rewritten in a form suitable of ordinary least squares estimation as

$$\log Y - \log H = \sum_{i=1}^{6} \log F_i(D_i) + \beta_1(-1/T) + \beta_2(-H/T) + \beta_3(-H^2/T) + u \; . \qquad (5)$$

Regression of log Y - log H on the D_i's, $-1/T$, $-H/T$, and $-H^2/T$, with the intercept suppressed, yields estimates for the log F_i's and the β's. The estimated coefficients for this first regression, together with their t values, are shown as line (1) in Table 1.

There are two things to note about this first regression before going on to the definitive form of the relationship. First, the estimated coefficients of D_1 and D_2 are the same; the sample offers no evidence that asymptotic penetration levels for systems with two or fewer local signals depend on the number of distant signals carried. Second, the estimated coefficient of the $-H^2/T$ term is not significantly different than zero at the .95 confidence level; there is no evidence that the β parameter in the logistic growth curve is a non-linear function of number of households.

Consequently, I estimate a revised form of the relationship. Service of types 1 and 2 is lumped together and called type 1. In other words, all systems with two or fewer local signals are classified as offering type 1 service, regardless of the number of distant signals they carry. Also, the H^2/T term is omitted from the equation. The resulting equation to be estimated is

$$\log Y - \log H = \sum_{i=1,3}^{6} \log F_i(D_i) + \beta_1(-1/T) + \beta_2(-H/T) + u \; . \qquad (6)$$

Estimated coefficients and t values are shown as line (2) in Table 1.

Estimated Penetrations

The coefficients of the D_i are estimates of log F_i. By raising e to these powers, one obtains estimates of the F_i themselves, the

Table 1

REGRESSION RESULTS

	Dependent Variable	Estimated Coefficients										
		D_1	D_2	D_3	D_4	D_5	D_6	$-1/T$	$-H/T$	$-H^2/T$	0	R^2
(1)	log Y - log H	-.509 (-3.55)	-.509 (-6.19)	-.986 (-9.45)	-.697 (-7.12)	-1.237 (-6.10)	-.875 (-5.44)	1.194 (2.16)	$.405(10^{-3})$ (6.88)	$-.808(10^{-10})$ (-0.47)		.391
(2)	log Y - log H	-.508 (-7.12)		-.987 (-9.48)	-.696 (-7.13)	-1.223 (-6.10)	-.874 (-5.45)	1.307 (2.63)	$.387(10^{-3})$ (8.73)			.390
(3)	log Y - log H	-.495 (-6.33)		-.976 (-9.09)	-.673 (-6.01)	-1.206 (-5.88)	-.856 (-5.15)	1.300 (2.61)	$.387(10^{-3})$ (8.73)		-.0277 (-0.41)	.391

asymptotic penetration levels. These values, the central results of this paper, are presented in Figure 3. Ninety percent confidence intervals for the estimates are shown within parentheses in the figure.[1]

The relative magnitudes of the estimated asymptotic penetration levels correspond well with a priori expectations. Systems in areas with two or fewer local signals have the highest penetration. Here cable service is especially attractive because it supplies missing network signals and adds greatly to the very limited service available over the air. My estimate indicates that a cable system in such an area can expect, on the average, ultimately to serve 60 percent of all households in its service area. In areas where more local signals are available, estimated asymptotic penetration levels are lower. In an area with between three and six local signals, a cable system that imports three or fewer distant signals can expect an ultimate penetration level of .37; a system importing more than three distant signals will do better, averaging an ultimate penetration of .50. For areas even better endowed with local signals, estimated ultimate penetration decreases still further: .29 for systems that do not import distant signals, .42 for those that do.

Nationwide Average

Estimated ultimate penetration levels in Figure 3 may be used to calculate a rough estimate of expected nationwide average penetration. I make two assumptions, both of which bias the estimate upwards. First, all cable systems will carry four or more distant signals, so the boxes at the top in Figure 3 apply. Second, all television homes are located in areas where cable service can be provided at a reasonable price.

Nearly two-thirds of all television homes are located in areas where three to six signals are received, so the middle column of boxes

[1]Based on a 1.65 standard error band on either side of the estimated log F_i.

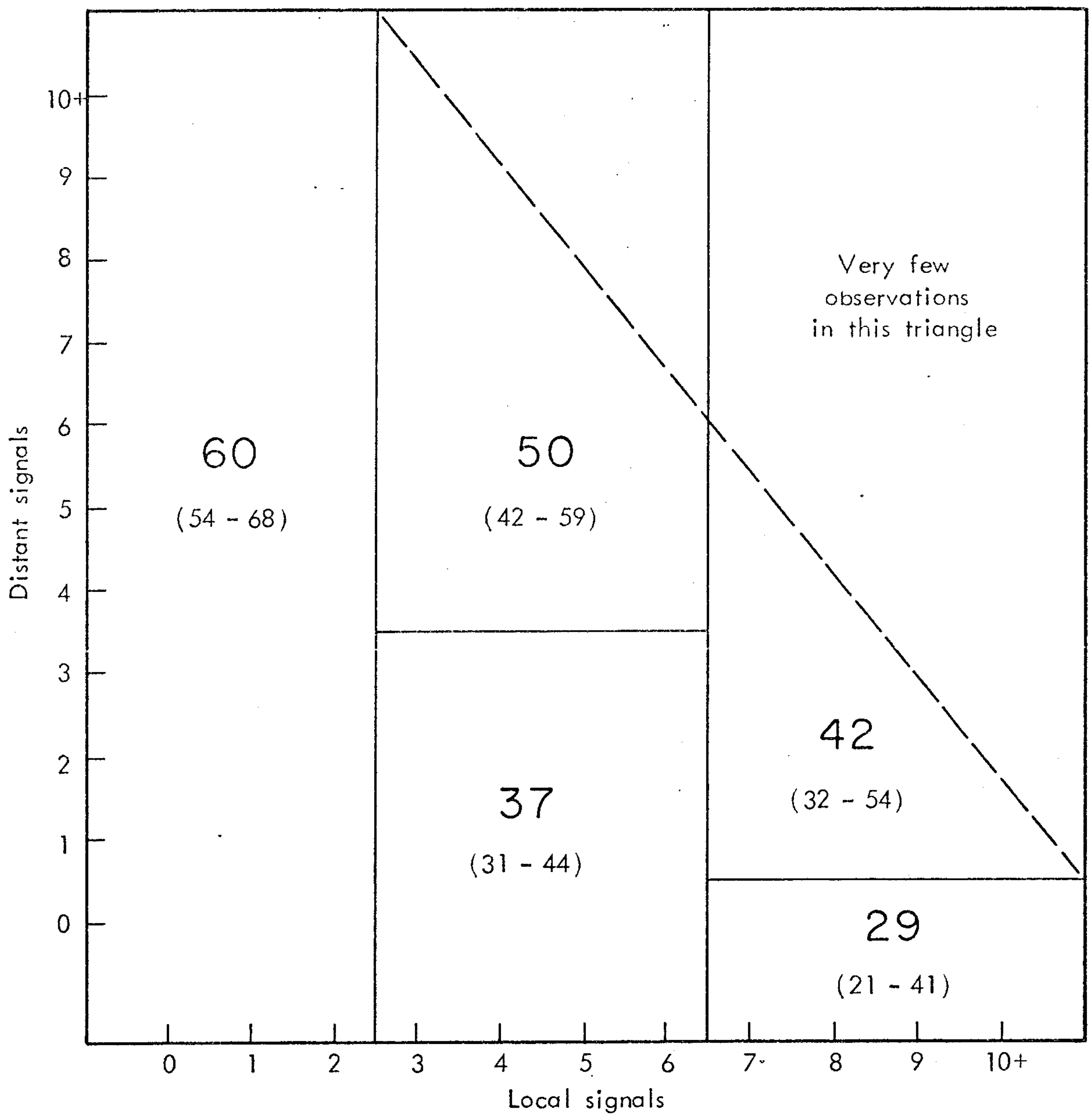

Fig. 3 — Estimated ultimate penetration percentages

in Figure 3 is numerically the most important one. Most of the remainder can receive seven or more signals, so the column to the right is also important. Using the distribution of homes by signals received, and penetration estimates from Figure 3, expected nationwide average cable penetration is calculated in Table 2. Making some allowance for the upward bias introduced by my assumptions, the result can be stated as follows: Expected nationwide average cable penetration is on the order of 40 to 45 percent.

Estimated β Coefficients

It may be of some interest, if only as a check on the plausibility of the model, to discuss the estimated β coefficients, as well. From line (2) of Table 1, the estimated expression for the parameter β in the logistic growth curve is $\beta = 1.307 + .000387H$. In the sample, the mean number of households is 12,929. Recall that the inflection point of the logistic curve comes at $T = \beta/2$. For the average system, then, the estimated inflection point is reached only three months after service begins. For a system with a large service area, say 250,000 households, the estimated inflection point comes 49 months after service begins. These figures strike me as being somewhat on the low side, but not altogether unreasonable.

ASSESSING THE OVERALL SIGNIFICANCE OF THE MODEL

R-squared for the second regression in Table 1 is .39. All it takes is a glance at the t values to assure one that this is statistically a highly significant portion of the total variance in the dependent variable. But the variance of log Y - log H does not have much intuitive meaning.

To make possible a more informative assessment of the overall in-sample performance of the model, I rewrite equation (6) as a predictor of number of subscribers:

$$\hat{Y} = \hat{F}_i H e^{-(\hat{\beta}_1 + \hat{\beta}_2 H)/T} \qquad (7)$$

Table 2

EXPECTED NATIONWIDE AVERAGE PENETRATION

Number of Stations Received	Percent of TV Households[a]	Estimated Penetration	Percent on Cable[b]
2 or fewer	3.4	.60	2.0
3 to 6	64.0	.50	32.0
7 or more	32.6	.42	13.7
Total	100.0		47.7

Notes:

[a]From Nielson national sample in September 1967 cited in "A Study of Distribution Methods for Telecommunications (Complan Associates)," A Survey of Telecommunications Technology Part 2, President's Task Force on Communications Policy, June 1969, PB 184 413.

[b]Column 2 times column 3.

As shown in the third line of Table 3, equation (7) explains 31 percent of the total variance in subscribers. Obviously, the "other factors" represented by the error term in the model are important, resulting in 69 percent of the total variance. But it is also true that the factors included explicitly in the model have a highly significant influence. The F statistic for equation (7) is 25.9, and $F_{6,\ 345,\ .01}$ is only about 2.70. Thus the equation is significant far beyond the .01 level. If the other factors remain reasonably constant, then my estimates should be reasonably good predictions. If not, then the other factors should be taken explicitly into account in the model, if possible.

OTHER SERVICES

Obviously, the estimates of asymptotic cable penetration levels presented in this paper (Figure 3) apply only to cable systems whose primary service is delivery of some combination of local and distant television signals. Future cable systems may additionally offer a number of other, essentially different services, such as opinion polling, automatic meter reading, and unique kinds of cable originated programming. If such services should come to motivate an important part of cable demand, my estimates will no longer be relevant. (To the extent that cable originations resemble broadcast programming, however, origination channels could be counted as distant signals, and my estimates could still be used.)

Of course, empirical estimates of the importance of dramatic new services are impossible as long as the services are non-existent. In an attempt to get some feel for the importance of extra services offered by cable systems, I estimate a model that allows for an effect of two unique extra services currently available on some systems: mechanical origination, such as time and weather, and local live origination. The model is

$$Y = F_i H e^{\gamma_0 - (\beta_1 + \beta_2 H)/T} + u \quad . \tag{9}$$

Table 3

ANALYSIS OF VARIANCE OF SUBSCRIBERS ABOUT MEAN

Source of Variance	R-Squared	Sum of Squares (10^6)	Degrees of Freedom	Mean Square	F Statistic
Explained by (2.7)	.311	2440	6	406.7	25.9[a]
Unexplained residual	.689	5400	345	15.7	
Total	1.000	7840	351		

Notes:

[a] $F_{6,\ 345,\ .01} = 2.70$.

The new variable O is a crude index of a system's origination activity. It can take on values of zero, one, or two, with one point assigned for each type of origination offered by the system. In the model, origination increases expected subscribers at any point in time by the factor $e^{\gamma O}$, where γ is a parameter to be estimated.

Model (9) results in a regression equation identical to (6) except that it includes a γO term on the right hand side. Parameter estimates for this model are shown as line (3) back in Table 1. The estimated origination coefficient has the wrong sign, but is not significantly different than zero, with a t value of only -.41. Other specifications of the origination index O -- including canned as well as mechanical and live origination, and using a zero/one origination dummy instead of the additive index described above -- perform even worse. Estimated γ coefficients for these other specifications are larger negatively, but in no case does the t value exceed 0.8 in absolute value, so none are significant even at the .25 level.

Based on these results, one cannot reject the hypothesis that current cable originations have no effect on asymptotic penetration levels. However, other services available in the future may result in penetrations greater than those estimated here.

SUMMARY

Expected ultimate cable penetration levels are estimated by fitting a set of logistic growth curves to 1969 data on cable systems. Highest penetration, 60 percent on average, is to be expected in areas with two or fewer local signals. In such cases, the number of distant signals carried has little or no effect on expected penetration. Lowest penetration is estimated for cable systems that carry no distant signals and operate in areas with many local signals; such a system can expect ultimately to serve 29 percent of all homes in its service area on average. Expected penetration for systems with other combinations of local and distant signals ranges from 37 to 50 percent, as shown in Figure 3. Rough calculations based on these estimates suggest an ultimate nationwide average cable penetration on the order of 40 to 45 percent.

Ultimate penetration may be higher than estimated if radically new cable services are offered in the future. The kind of origination now offered, however, does not significantly affect penetration.