Stuart Polikoff
Stephen Pastorkovich
OPASTCO
21 Dupont Circle, N.W.
Suite 700
Washington, D.C. 20036

David D. Oxenford
Bruce D. Jacobs
Stephen J. Berman
Fisher Wayland Cooper
Leader & Zaragoza L.L.P.
2001 Pennsylvania Avenue, N.W., #400
Washington, D.C. 20006

Randall B. Lowe
Julie A. Kaminski
Prism Communication Services, Inc.
1667 K Street, N.W.
Suite 200
Washington, D.C. 20006

Ben Wiles
Senior Attorney
Public Utility Law Project
90 State Street
Suite 601
Albany, NY 12207

James D. Ellis
Alfred G. Richter, Jr.
Roger K. Toppins
Mark P. Royer
Counsel for SBC Communications, Inc.
1401 Eye Street, N.W., 11th Floor
Washington, D.C. 20005

Guy Christiansen
Directory, Regulatory Affairs
SkyBridge LP
3 Bethesda Metro Center
Suite 700
Bethesda, MD 20814

Rikke Davis
Jay C. Keithley
Counsel for Sprint Corporation
401 9th Street, N.W.
4th Floor
Washington, D.C. 20004

Charles C. Hunter
Catherine M. Hannan
Hunter Communications Law Group
1620 I Street, N.W.
Suite 706
Washington, D.C. 20006

Lawrence E. Sarjeant
Linda L. Kent
Keith Townsend
John W. Hunter
Julie E. Rones
Counsel for USTA
1401 H Street, N.W., Suite 600
Washington, D.C. 20005

Robert B. McKenna
Blair A. Rosenthal
Counsel for US WEST
1020 19th Street, N.W.
Suite 700
Washington, D.C. 20036

Jeffrey L. Sheldon, Esq.
Brett Kilbourne, Esq.
United Telecom Council
1140 Connecticut Avenue, N.W.
Suite 1140
Washington, D.C. 20036

Paul J. Sinderbrand
Robert D. Primosch
Wilkinson Barker Knauer, LLP
2300 N Street, N.W.
Suite 700
Washington, D.C. 20037

Larry Goldberg, Director
Media Access
WGBH Educational Foundation
125 Western Avenue
Boston, MA 02134

ITS*
445 12th Street, S.W.
Washington, D.C. 20554

Michelle Mundt

Michelle Mundt

DCDOCS:158827.1(3#JV01!.DOC)

ALL STATE® LEGAL 800-222-0510 ED11 RECYCLED

A Response to Olbeter and Robison's "Breaking the Backbone"

EXECUTIVE SUMMARY

The recent paper by Erik Olbeter and Matt Robison, entitled "Breaking the Backbone: The Impact of Regulation on Internet Infrastructure Deployment," claims that differences in regulation governing the RBOCs and the Independent LECs is responsible for the paucity of Internet hubs in rural states. Their underlying theory is based on flimsy logic and a poor understanding of the Internet. But most significantly, Olbeter and Robison's statistical findings are the result of the exclusion of data, incorrect specification and measurement of the variables, careless data collection, and a lack of expertise in regression analysis.

This brief response to the Olbeter-Robison study shows, and simple logic dictates, that the most important determinant of Internet hub deployment in a state is its population. Yet, Olbeter and Robison take no direct account of population in their analysis. Failing to do so leads the authors to conclude that lifting the interLATA restriction on 226,000 access lines in Wyoming will generate as many new Internet hubs as the lifting that restriction on 16.1 million access lines in California. According to Olbeter and Robison, after lifting the interLATA restriction each Internet hub in Wyoming will service only 16,000 persons while in California 157,000 persons will be forced to share each Internet hub. Olbeter and Robison also forecast that if the interLATA restriction is lifted, Delaware, with a population of about 730,000, would have as many Internet hubs as Virginia, with a population of 6.7 million and a thriving Internet community. Obviously, these predictions are implausible.

"Garbage In, Garbage Out" is probably the best explanation for the findings of the Olbeter-Robison study. As described in detail in this document, the Olbeter-Robison can be criticized on the following points (at a minimum):

- Even using the same data and model described by Olbeter and Robison, their statistical results cannot be replicated.

- Four states (plus the District of Columbia) are excluded from the sample used for the analysis for extraneous reasons. Yet, when these states are added to the sample, and all states are included in the statistical analysis, the reverse effect to the one claimed by Olbeter and Robison is proven. Namely, their analysis shows that removing the interLATA restriction on the RBOCs would

decrease the number of Internet hubs and cause the "disconnected dozen" to become the "disconnected thirty-two."

- Olbeter and Robison include only large (Tier One) independents in their analysis, excluding roughly one-third of non-RBOC lines. Inclusion of all independent LECs (even using Olbeter and Robison's carefully chosen sample) undermines their conclusion.

- As discussed above, by ignoring population, the statistical model generates implausible forecast values of hub counts resulting from the lifting of the interLATA restriction.

- Olbeter and Robison's data collection was sloppy at best. For example, the authors contend that there are no Internet hubs in Maine, Montana, North Dakota, and West Virginia. Yet, *Boardwatch* magazine (their source for hub counts) lists PSINet as operating hubs in all of these states.

- Olbeter and Robison's discussion of regression analysis indicates a profound lack of expertise with the statistical technique. For example, it is impossible for Olbeter and Robison's model to forecast a decrease in Internet hubs, yet they forecast Ohio will lose 4 hubs if Ameritech receives interLATA relief. Further, the authors inappropriately employ diagnostic statistics when they do, and fail to perform necessary diagnostics when needed.

Statistical analysis can be very helpful to social scientists and policymakers trying to understand causal relationships in a complex factual setting. The analysis of the costs and benefits of alternative regulatory regimes requires that the magnitude of particular causal relationships be measured. This task is the role of statistical research. But statistical analysis can also be poorly constructed and performed, perhaps even manipulated, to indicate fictitious relationships that simply do not exist. As detailed in this report, the Olbeter and Robison study falls into this latter category of statistical research poorly done and, hopefully, ignored.

George S. Ford, Ph.D.
Senior Economist
MCI Worldcom Inc.
1801 Pennsylvania Avenue NW
Washington, DC 20006
george.ford@wcom.com

A Response to Olbeter and Robison's "Breaking the Backbone"

GEORGE S. FORD*

In a recent paper entitled "Breaking the Backbone: The Impact of Regulation on Internet Infrastructure Deployment," Erik Olbeter and Matt Robison attempt to assess the impact of "regulation" on the deployment of Internet hubs. The authors' central thesis is that differences in the forms of regulation governing the Regional Bell Operating Companies (RBOC) and the Independent Local Exchange Carriers (INLEC) are responsible, at least in part, for the deployment of Internet hub facilities. Regulation of the RBOCs and INLECs is similar, but the authors focus only on the interLATA restrictions faced by the RBOCs. While Olbeter and Robison mention the differences in unbundling/resale requirement not faced by smaller INLECs, these smaller independents are excluded from their analysis.

By considering only the interLATA restrictions, Olbeter and Robison conclude that the INLECs are "unregulated (p. 11)" while the RBOCs are "regulated."[1] Within this context, testing for a relationship between the percentage of lines served by INLECs and the number of Internet backbone hubs in a state becomes the central theme of the study. If a positive relationship is found, the authors contend, then "regulation" matters and regulators could promote Internet hub deployment by "deregulating" the RBOCs. This positive relationship is found to exist using regression analysis, and the authors forecast a 118% increase in the number of Internet hubs if the RBOCs are "deregulated."

In this brief report, I evaluate the statistical analysis of Olbeter and Robison. My focus on the statistical analysis should not be interpreted as the acceptance of their central thesis. Indeed, the authors' assertion that the entire Internet backbone community is systematically forgoing profitable opportunities is, on its face, a dubious proposition. The extent to which the authors claim the Internet community has failed to recognize profitable market opportunities for hub

---

* Senior Economist, MCI Worldcom Inc, 1801 Pennsylvania Avenue NW, Washington, DC 20006, george.ford@wcom.com.

[1] Obviously, describing local exchange carriers that are subject to price regulation, interLATA restrictions, and a plethora of other regulations as "unregulated" is inappropriate.

deployment - deploying less than one-half the "optimal" number of hubs -- provides the strongest evidence as to the daftness of their thesis. Perhaps more importantly, the authors incorrectly claim that Internet hub cities (listed in *Boardwatch* magazine) are the only place where customers can access Internet backbones. In addition, the hub count data from *Boardwatch* magazine substantially understate the true number of hub facilities. For example, UUNET only provides *Boardwatch* the names of hub cities with multiple OC-12 connections.[2] Even worse, Olbeter and Robison fail to include all the hub cities listed in *Boardwatch* in their analysis.

I. The Olbeter and Robison Approach

The Olbeter-Robison thesis holds that the less regulated are a state's local exchange companies, the larger number of Internet hubs the state will have. Because INLECs are not subject to the interLATA restriction faced by the RBOCs, the authors contend that the percentage of lines served by INLECs is an inverse proxy for the degree of regulation in a state. Given variation across states in the percentage of lines served by INLECs, a statistical test of their thesis may be possible. To this end, the authors posit an ordinary least squares regression (OLS) to measure hub presence in a state as a function of INLEC share of presubscribed access lines.[3] If a greater INLEC presence produces more Internet hubs, Olbeter-Robison claim this finding is evidence that "deregulation" will increase the hub count in a state.

Olbeter and Robison assess their hypothesis - more Internet hubs will be associated with more INLEC lines -- by estimating a OLS regression of the following form:[4]

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \qquad (1)$$

where

$y =$ number of backbone hubs per state;
$x_1 =$ per capita income, per state;

---

[2] Thus, only the highest capacity hub cities are listed for UUNET.

[3] For a basic exposition of least squares regression, see A.H. Studenmund, *Using Econometrics* (1992).

[4] Because the dependent variable is count data, with a substantial number of zeros, the Poisson regression would be preferred to OLS.

x_2 = number of cities with population of 100,000 or more, per state;
x_3 = percent of local access lines controlled by *Tier One* INLECs, per state[5];
x_4 = number of CLEC resold lines, per state;
ε = econometric disturbance term.[6]

The actual data and sources for these variables are provided in Attachment A.

The authors posit that more hubs should be located in states with higher per capita incomes (x_1; $\beta_1 > 0$) and a greater number of large cities (x_2; $\beta_2 > 0$). They also test for a local competition effect by including the number of resold lines in the state as a variable (x_4). The authors posit no explanation for why resale, rather than facilities-based competition, is a better measure of the impact of CLEC presence. Section II.5 of this document considers the impact of a facilities-based measure of competition on Olbeter and Robison's conclusions. The most important variable is the percent of lines served by INLECs in the state (x_3). According to the authors, proof of their hypothesis is supported by a positive and statistically significant coefficient on x_3 ($\beta_3 > 0$).

For those less familiar with regression techniques, the gist of the Olbeter-Robison test is illustrated in Figure 1. According the Olbeter-Robison thesis, if we plot the observed number of hubs and the percent of Tier One INLEC lines per state we would find a relationship like the line labeled AA in the figure, where the relationship between the two variables is positive.[7] Note that the slope of the line AA is β_3. The alternative hypothesis is that more INLEC lines leads to fewer Internet hubs in a state and this relationship is illustrated by the line labeled BB. Of course, there may be no relationship so that a fitted line through the scattered points is flat (i.e., the slope is zero) like line CC in the figure. Note that in

---

[5] While the text of the Olbeter-Robinson reports indicates that the variable x_3 measures all non-RBOC lines, the figures in their report suggest otherwise. On page 26, for example, the percent of INLEC lines is listed as 18.8%, which is the percentage lines served by Tier One INLECs in the state and not all INLECs.

[6] The authors describe the estimated parameters, i.e., the β's, as being the exogenous variables rather than the x's as is typical in econometrics work. Nor do the authors include a disturbance term, ε, in their specification of the model - also customary in econometrics work.

[7] The example is an illustrative simplification of the thesis, and does not contain real data. The influences of other factors are assumed to already be accounted for or zero.

regression analysis, a flat line is implied by a statistically insignificant coefficient.[8]

Olbeter and Robison estimate the OLS regression (equation 1) using a sample of 46 states; Connecticut, Nevada, Hawaii, Alaska, and the District of Columbia were excluded. The results of their statistical analysis are provided in Table 1.[9] As expected, the OLS regression indicates that more hubs are located in states with a higher number of large cities ($\beta_1 > 0$) and higher incomes ($\beta_2 > 0$). Competition, as measured by CLEC resold lines, does not affect the number of hubs in a state ($\beta_4 = 0$; i.e., it is statistically insignificant). According to the results provided by the authors, the OLS coefficient β_3 is positive and statistically significant. Given the positive and statistically significant coefficient (β_3) on the percent INLEC lines, the authors conclude that their thesis is correct – a higher percentage of INLECs lines will produce a higher hub count per state. Thus, the authors conclude that RBOCs should be regulated more like INLECs, i.e., the interLATA restrictions should be lifted.

From these regression coefficients, Olbeter and Robison forecast the number of hubs that would exist in a state if the RBOCs were "deregulated." The forecast hub counts are calculated with the following formula:

$$y^* = y + 27.923 \cdot (1 - x_3) \tag{2}$$

where y^* is the forecast number of hubs. The estimated change in hub counts is provided in Table 2. As illustrated in the table, the forecasted increase in hubs in some states is substantial. For example, the authors predict that if Bell Atlantic is deregulated in Delaware, hub count in that state will increase from 2 to 40. Note that this latter figure matches the current hub count for state of Virginia (the Silicon Valley of the East with nearly ten-times the population of Delaware).[10]

The dramatic increase in hub counts for the relatively small, less populated states is a consequence of the specification of the variable measuring INLEC lines. As illustrated in the table, Idaho and California expect to see an identical

---

[8] Statistical significance, at the 5% level, can generally be detected by a t-statistic of 2.00 or greater. Critical t-statistics will vary, however, with sample size (and, of course, with significance level). For the sample sizes dealt with here, the 5% critical value is about 2.01 and the 10% critical value is 1.68.

[9] As described later in the text, the estimated coefficients are not valid even if the same model is estimated using the same data.

[10] See Section II-3 for further discussion of this topic.

increase in the number of hubs. The increase is identical because in percentage terms, the number of INLEC lines is approximately equal for Idaho (17.6%) and California (18.5%). Thus, the formula in equation (2) would predict the same increase in hubs because $(1 - x_3)$ is identical for the two states. In other words, "deregulating" about 226,000 RBOC lines in Wyoming is equivalent to "deregulating" 16.1 million RBOC lines in California.

The chosen specification of the model, and not economic realities, is responsible for the preposterous forecasted hub counts. Furthermore, because the authors measure the percent of INLEC lines using only Tier One INLECs, they understate the percent of INLEC lines thus overstating $(1 - x_3)$ by about 50% on average and artificially driving up the forecast value of hubs.[11] Since the authors claim that they use all non-RBOC lines, and the theory underpinning their forecasts require them to do so, it is clear that their data collection methods were careless.

II. A Review of the Olbeter and Robison

As illustrated above in Table 2, Olbeter and Robison's analysis shows that changing the manner by which the RBOCs are regulated -- making RBOC regulation more like the regulation of the INLECs -- packs a powerful punch. As I will describe in the next few paragraphs, this punch turns into a wet noodle upon closer inspection of Olbeter and Robison's statistical analysis. My review of their work is simple and straightforward. In section II.3, I attempt to replicate the statistical analysis of Olbeter and Robison. The data set used in *their* analysis is re-constructed and *their* model estimated. I find that their results cannot be replicated. After performing this straightforward task (that, while simple, is enlightening), I will evaluate Olbeter and Robison's exclusion of particular states from the sample in Section II.2. Excluding observations can heavily influence the estimated parameters of a model. Thus, excluding data should be done with caution and with full disclosure as to what affects the exclusion has on the estimated parameters.

In Section II.3, I will consider the plausibility of their forecast hub counts by calculating the number of persons per state serviced by the Olbeter-Robison

---

[11] Tier One INLECs represent about 14% of state access lines on average, while all INLECs represent about 26% of state access lines on average (thus, $1 - x_3$ will be smaller if all INLEC lines are included). Note that the coefficient on x_3 is determined from the incorrect measure of INLEC lines so it is not appropriate simply to reduce the forecast increase in hubs using the correct measure of INLEC lines.

forecast number of hubs. Section II.4 provides numerous examples from the Olbeter-Robison report that indicate that Olbeter and Robison's knowledge of regression analysis is questionable and their data collection methods careless. An alternative measure of competition is offered in Section II.5, and the impact of low hub counts on rates for Internet service is discussed is Section II.6. Concluding comments are provided in Section III.

1. CAN THE RESULTS BE DUPLICATED?

Olbeter and Robison list all the sources for their data. These sources, and the data itself, are provided in Attachment A of this paper. With this data, I have re-estimated the *exact* equation of Olbeter and Robison, using the *same* data and the *same* sample (Model 1).[12] In reviewing the data, I discovered that Olbeter and Robison included only Tier One INLECs in their sample. Thus, their data did not include all non-RBOC lines in their measure of INLEC line share, excluding about one-third of all INLEC presubscribed lines from their data.[13] This exclusion is inconsistent with the text of the Olbeter-Robison study, so I also estimate a second regression (Model 2) using an alternative measure of INLEC lines equal to the percent of all non-RBOC lines.

In Table 3, the results presented by Olbeter and Robison are compared to my own. Model 1 provides the results using the same measure of INLEC lines as Olbeter and Robison, while Model 2 replaces the percent of INLEC lines variable with a variable including all non-RBOC lines. Clearly, the econometric estimates of the three regressions are quite different, particularly where it matters. Olbeter and Robison find a statistical significant coefficient on the percentage of INLEC lines (β_3) -- I do not. The t-statistic for the coefficient on INLEC line share is 1.52 (Model 1), which is significant only at the 0.13 level. The estimated parameter on the variable is quite different across models as well and varies by specification of INLEC lines. While the coefficient on INLEC line share is positive (Model 1), it is customary to measure statistical significance at no more than the 0.10 level and often at no more than the 0.05 level. At best, therefore, the Olbeter-Robison model provides only very weak evidence of a positive relationship between hub count and INLEC line share. The alternative specification of INLEC lines (all

---

[12] The regressions were run on two different statistical packages to confirm the estimated parameters were correct.

[13] On average, the exclusion of smaller INLECs understates the INLEC share in a state by 50%.

non-RBOC lines per state), the measure Olbeter and Robison claim to use, is not statistically significant at anywhere near acceptable levels.

The exact source of the difference in parameter estimates and significance levels is not known with certainty, but only three explanations are possible. First, the data between the two studies might be different. My data is gathered directly from the sources listed by Olbeter and Robison and, as such, should be identical to their data.[14] All the data is available in Attachment A and I encourage others to audit the data and use it to assess my estimates. Second, some statistical packages are known to produce slightly different results, but this problem is rare for simple OLS regressions. I have duplicated my results with two different statistical packages and the results are the same. Third, the differences could be attributed to carelessness in the estimation or reporting of the results. For example, observations may have been dropped if the statistical package read zeros as missing data. Or, estimates from a more limited or expanded sample may have inadvertently been presented in the Olbeter-Robison study.[15] Which of these explanations is responsible for the differences is unknown at this time.

What is most important about my attempts to duplicate their results is not that I cannot, but that the coefficient on the percent INLEC lines is not statistically significant. Thus, my analysis, based entirely on the proposed methodology of Olbeter and Robison, provides no support for their thesis that a higher percentage of INLEC lines leads to higher hub counts. The rejection of the Olbeter-Robison thesis (as tested by the OLS regression proposed by the authors) is possible without further comment. However, as I will show in the next section, there are other flaws with their approach that are important within the context of their thesis.

---

[14] The hub data was taken directly from the Olbeter and Robison study. Some data was provided on page 26 of their document, and those figures match what is in my data. I have some concern on the measurement of the percent CLEC lines in the Olbeter and Robison study, since their figure for Massachusetts (on page 26) does not match the data in the original source (the *Local Competition Report*). However, excluding this variable from the regression does not remedy the differences in parameter estimates (I compared my results to Model 2, Figure 1, in Olbeter and Robison).

[15] Olbeter and Robison do not include the number of observations in their regression tables, but I suppose that the sample size is 46 as indicated in their description of the sample.

2. SAMPLE SELECTION

Excluding observations from a sample can heavily influence the estimated parameters of a model and, in this section, I will illustrate exactly how heavy this influence can be. Olbeter and Robison exclude four states and DC from their sample. While providing dubious explanations for excluding Alaska, Hawaii, Connecticut, Nevada, and DC, Table 4 suggests a potential alternative explanation for the exclusion of these states.

Recall that the authors' goal is to find support for their hypothesis that in states with high percentages of INLEC access lines, more Internet hubs will exist. As illustrated in Table 4, the states excluded by Olbeter and Robison are all quite "unique" in that they either they have substantially low number of hubs and a high percentage of INLEC lines, or they have a large number of hubs and a low percentage of INLEC lines. Three states -- Alaska, Connecticut and Hawaii - are served entirely by INLECs. Nevada has a very high percentage of INLEC lines (73%).[16] All four states have hub counts well below the national average. In contrast, all lines are served by the RBOC in DC, yet its hub count is well above the national average. While it seems that hub deployment in these states would be most illustrative of the effect of a high percentage of INLEC lines, Olbeter and Robison choose to exclude them. Of course, if one wished to show that the percent of INLEC lines was positively related to hub counts, these four states would certainly not help illustrate the point. Nor would the District of Columbia.

The authors' explanation for excluding Connecticut from the sample is particularly peculiar. The authors, attempting to test for the consequences on hub counts of the inability of RBOCs to provide interLATA services, exclude Connecticut because:

> Connecticut is unique because an integrated carrier (SNET) offers local and long distance service throughout the entire state ... (p. 13).

It would appear that Connecticut is a case study of the Olbeter and Robison thesis, not a peculiar case to be discarded from the sample. The authors also mention that Connecticut's "proximity to New York City (p. 13)" justifies its exclusion. Yet, Pennsylvania and New Jersey do not qualify for exclusion on the same grounds.

If Olbeter and Robison's findings are valid (a higher percentage of INLEC lines leads to a higher hub count) then we clearly have at least five exceptions -

---

[16] The figure is 68% if only Tier One INLECs are included.

five exceptions (about 10% of the sample) that were excluded from the statistical analysis. A figure, similar to Figure 1, may help illustrate the effect of excluding these particular states from the sample. Consider the scatterplot of x-y points in Figure 2. The points indicated by asterisks (*) represent observations like the four states and DC in the Olbeter-Robison data, i.e., these particular points have either a low y and high x value, or a high y and low x value.

Recall that regression analysis, like that used by Olbeter and Robison, fits a line to these data points. If we exclude the asterisk points from the sample, we would fit a line to the data that looks like the dashed line labeled AA. This line has zero slope so that the estimated coefficient (β) of the regression will not be statistically significant.[17] If we fit a line to the data including the asterisk points, however, then the fitted line will have a downward, negative slope like the line labeled BB. Now, the estimated coefficient will be negative and statistically significant.

The hypothetical example illustrated in the figure is not that different from the actual data on hub counts. The impact on the econometric estimates of excluding these four states and DC is presented in Table 5. In Model 1, we provide the OLS regression results from the sample of 46 states used by Olbeter and Robison. As discussed earlier, the coefficient on percent INLEC lines (β_3) is positive but not statistically distinguishable from zero. Thus, the relationship between INLEC lines and hubs per state looks like the line labeled AA in Figure 2. Adding in the five excluded observations produces a sample of 51 observations.[18] Model 2 in the table contains the results from the Olbeter-Robison equation when estimated with the full sample. Just as in Figure 2, the relationship between hub counts and INLEC lines is found to be negative and statistically significant.[19] Thus, when using the full sample, the Olbeter-Robison thesis is not only rejected, but is shown to be contrary to reality. If we embrace the Olbeter-Robison thesis, then we must conclude that more regulation is preferred to less and the INLECs should be forced to adhere to the interLATA restriction just like the RBOCs.

---

[17] The estimated coefficient may be positive or negatively signed, but its t-statistic determines whether or not we can conclude that the slope is not zero.

[18] The negative and statistically significant coefficient on INLEC lines remains even if Alaska and Hawaii are excluded from the sample.

[19] Clearly, these five observations have a substantial influence on the results. While there are legitimate statistical tests for outliers, no such tests were performed by Olbeter and Robison. In any case, the exclusion of particular observations should be done with caution.

Following Olbeter and Robison's forecast method, we can estimate the number of hubs that would exist per state if the RBOCs were "deregulated." The forecast formula is:

$$y^* = y - 14.35 \cdot (1 - x_3) \tag{3}$$

where y^* is the number of forecast hubs. The forecast number of hubs, based on the Olbeter and Robison regression with the full sample, are provided in Table 6. These forecast values indicate that if the Olbeter-Robison prescription of RBOC "deregulation" is taken seriously, the "disconnected dozen" will become the "disconnected thirty-two."[20]

3. Population Per Hub

The first step in evaluating forecasted numbers from regression equations is to consider whether the magnitudes of the forecasts make any sense. The improbability of the forecast hub counts offered by Olbeter and Robison should be obvious to those with only a casual familiarity with the Internet industry. The mere suggestion that the state of Delaware, with a population of about 750,000, would have as many hubs as the state of Virginia does today (population of about 6.7 million and a thriving Internet community) should indicate a problem with the statistical approach. Equally troubling is the contention that roughly 1,165 more hubs would exist if the RBOCs were "deregulated" -- roughly twice as many Internet hubs as exist today (984 according to the authors).[21] In other words, the Internet backbone community has deployed less than half of the appropriate number of hubs and only "deregulating" the RBOCs will eliminate this severe shortage.

To further illustrate the improbability of the Olbeter-Robison forecasts, Table 7 illustrates the average number of persons in each state serviced by the forecast number of hubs. For all states in the sample but one, the number of hubs deployed increases; Ohio would see its hub count fall from 56 to 52 (see Table 2). Setting aside for the moment that it is *impossible* for the Olbeter-Robison forecast method to produce a reduction in hub counts for any state (i.e., equation 2 must

---

[20] Thirty-two states will have hub counts of 7 or less. See Olbeter and Robison at p. 3.

[21] At no fault of the authors, the hub counts in *Boardwatch* magazine are substantially below what actually exists. UUNET, for example, provides Boardwatch Magazine only with the names of cities where its hubs have multiple OC-12 connections. Olbeter and Robison also fail to count all hubs in the magazine (see Section II.4).

be positive because $x_3 \leq 1$), Olbeter and Robison conclude that there are too many hubs in Ohio necessary to support its population. From Table 7 we can see that Ohio has 200,000 people per hub. After "deregulating" the RBOCs, Ohio hubs will support 215,000 people per hub. Strangely, Olbeter and Robison claim that while 200,000 persons are too few per hub in Ohio, 16,000 in Wyoming and 24,000 per hub in Vermont is just right. As mentioned earlier, these implausible results are a consequence of the model's specification of INLEC lines as a percentage and, to some extent, the exclusion of roughly half the INLEC lines per state (on average).[22]

Even the most cursory examination of hub deployment (except that of Olbeter and Robison) would lead one to recognize the relationship between population and hub count. Conspicuously absent from the regression specified by Olbeter and Robison is a measure of population, either persons or access lines. Not surprisingly, the number of hubs and population are positively related. Population alone explains about 90% of the total variation in the number of hubs across states (See Table 8, Model 1). It is also true that the states with the highest populations and the highest hub counts just so happen to have the largest number of INLEC lines. For example, the states Olbeter and Robison list as having the most hubs are California, Texas, Florida, New York, and Ohio. All five states are in the top ten most populous states. If the states are ranked by INLEC lines, all five are also in the top ten of this ranking. In fact, population alone explains as much variation in hub counts across states as the four variables in the Olbeter-Robison equation. Note that if population is included as an explanatory variable in equation (1), replacing the number of large cities variable due to high correlation, INLEC line share (Tier One LECs only) is nowhere near statistical significance (Model 2).[23] In fact, population is the only variable in the equation that explains the variation in hub counts across states.

4. Other Problems with the Olbeter-Robison Study

The Olbeter-Robison statistical analysis, for reasons discussed above, contains many flaws – some inadvertent, others perhaps intentional. Most of the

---

[22] If the percent of INLEC lines is replaced with the number of INLEC lines and total lines, then total lines is statistically significant but INLEC lines is not.

[23] The same result holds if the large city variable is included in the regression along with population. The high correlation figure between population and large cities reduces the t-statistics of both variables, but both variables are positively signed and statistically significant. None of the other variables in the model are statistically significant.

problems, I believe, reflect an apparent lack of expertise by Olbeter and Robison in the area of statistical analysis that serves as the foundation for their study.

For example, the authors conclude that their estimated parameters are "unbiased and efficient (p. 18)" and that their "model has no statistical problems (p. 14)" by appealing to the Durbin-Watson statistic. The DW statistic indicates whether or not (first-order) autocorrelation is present in a statistical model. In one particular case, where the DW statistic falls below its acceptable range, Olbeter and Robison state:

> More troubling, the model shows autocorrelation via the low Durbin Watson figure, signifying that the model's estimates are unbiased, but inefficient (p. 19).

While the authors are correct that the DW statistic does detect autocorrelation, and that autocorrelation leads to unbiased but inefficient parameter estimates, they err in the application of the test to cross sectional data. Practitioners of the statistical analysis know that the problem of autocorrelation *is not relevant* when using cross sectional data.[24]

To illustrate, we can estimate the Olbeter-Robison equation and then sort the data by the disturbance term. This sorting process that has absolutely no effect on the statistical properties of the regression results, but will alter the DW statistic. Upon re-estimation, the DW statistic changes from 1.73 to 0.027, the latter indicating severe autocorrelation. If we then sort the data by total lines and re-run the regression, the DW statistic increases to 2.26. If we sort by the number of large cities and re-estimate, the DW statistic falls to 1.88. For all of these different regressions, the estimated coefficients, standard errors, and R^2's are all identical (see Table 9). As this exercise makes apparent, the DW statistic says nothing about the efficiency (the t-statistics are identical) or bias (the β's are identical) of the OLS parameters when using cross sectional data.

The authors are also incorrect in suggesting that the DW test can detect multicollinearity (p. 14, ft. 8). The DW test is neither intended to nor capable of detecting multicollinearity.

---

[24] See, for example, Adrian C. Darnell, *A Dictionary of Econometrics*, 1994, p. 11. The Durbin-Watson statistic can be written as $d = 2(1 - \rho)$, where ρ is the coefficient of a regression of the residuals as a function of their values lagged one time period (Studenmond 1992, p. 343). Clearly, in cross-sectional analysis, there can be no "lagged" values of the residuals. Cross-sectional data, with very few exceptions, has no inherent order.

As opposed to autocorrelation, one real problem with model specification, whether using cross sectional or time series data, is simultaneity bias. If present, simultaneity bias makes the estimated coefficients biased, inefficient, and inconsistent. Olbeter and Robison indicate that simultaneity bias may be a problem with their model, stating:

> There may be a symbiotic relationship at work, meaning that greater Internet hubs helps propel income which in turn propel [sic] more Internet hub deployment (p. 14).

In statistical terms, this "symbiosis" is called "simultaneity." Despite the severe consequences of simultaneity bias, Olbeter and Robison perform no diagnostic tests to detect it or attempt to remedy the problem. If their statement above is correct and simultaneity does exists, then the Olbeter-Robison estimates have none of the desirable properties of OLS coefficients.

Another suggestive example is found on page 20 at Figure 2. The authors present some alternative specifications of their model to test for correct specification. The third model presented is different from the first only in that population density is added as an explanatory variable. Whenever an additional explanatory variable is added to a model, the R^2 will remain the same or rise (the adjusted R^2 may fall however). Yet, Olbeter and Robison indicate that by adding an additional variable, the R^2 falls from 0.91 to 0.28. This is simply not possible.

It is also impossible for the forecast method of Olbeter and Robison to predict the hub count in a state will decline. Recall that the forecast number of hubs is calculated using the following formula:

$$y^* = y + 27.923 \cdot (1 - x_3) \tag{4}$$

where y^* is the forecast number of hubs and x_3 is the percent of access lines in the state not operated by RBOCs (i.e., INLEC line share). Obviously, the lowest value x_3 can have is zero (no INLEC lines) and greatest value one (all lines are INLEC lines). The same is true for $(1 - x_3)$. Thus, y^* must be equal to or greater than y, i.e., the change in y $(y^* - y)$ is positive.[25] Despite this fact, Olbeter and Robison

[25] Defining $(1 - x_3)$ as z, it is obvious that $dy^*/dz \geq 0$.

claim that Ohio will experience a decline in hub counts if the RBOCs are "deregulated."[26]

Finally, it is worth mentioning that the data collection process of Olbeter and Robison shows signs of carelessness. Specifically, while Olbeter and Robison list Maine, Montana, North Dakota, and West Virginia as having no hubs, *Boardwatch Magazine* (Spring 1999) shows PSINet as operating hubs in all these states.[27] My guess is that the Olbeter-Robison hub count data is incomplete and fails to include any PSINet hub cities and potentially those of other backbone providers.[28] The finding

5. THE COMPETITION VARIABLE

Olbeter and Robison measure the extent of competition in a state by the number of lines resold by CLECs. While resale can, over time, lead to the increased deployment of CLEC facilities, it is probably too early in the competitive evolution in local exchange markets to observe that effect. In any case, a competition variable that accounts for a more facilities-based presence of CLECs is probably more appropriate.[29] For example, the *Local Competition* report that provided Olbeter and Robison's resale data also has data on the percent of lines in a state addressable by LEC switching centers where at least one CLEC has a co-location arrangement for voice grade and high capacity lines (Table 3.6, *Local Competition Report*).

Table 8 summarizes the regression results when the percent of resale lines is replaced with this alternative, facilities-based measure of competition. The negative sign on INLEC line share is still present, but significant only at the 10.6% level. A positive and statistically significant relationship between hub counts and the facilities-based measure of competition is found. Thus, according to the Olbeter-Robison thesis, regulators that desire to increase hub counts

---

[26] My attempts to duplicate the forecast values, based on Olbeter and Robison's description of their method (summarized by equation 4), failed. The forecast values for a number of states were substantially different than the values provided in the Olbeter-Robison report.

[27] Unlike most of the other Backbone Operators listed in *Boardwatch*, PSINet does not provide a list of hub cities in the text. Rather, its hubs are illustrated on a map.

[28] I did not have adequate time to reconstruct the hub count data from the original source material before the release of this report.

[29] At least within the context of the Olbeter-Robison thesis.

should impose the interLATA restriction on all incumbent LECs and encourage the development of a competitive market in local exchange services.[30]

6. THE CONSEQUENCES OF "DISCONNECTION"

Olbeter and Robison claim that the potential threat of the under deployment of Internet hubs is that "Companies in rural areas without nearby Internet hub access will spend a significant amount more money to conduct crucial business functions... (p. 5)." If this were true, then we would expect ISP rates in states without many hubs to be greater than those with many hubs. If the ISPs must pay more for access, then their customer should be paying more as well. When we compare a sample of rates for a 56k internet connection (Spring 1999 *Boardwatch*) from the five lowest and five highest hub count states (according to Olbeter and Robison), we find no difference in rates. As shown in Table 9, the average rates are $19.39 and $19.40 in low-hub and high-hub states, respectively. There is no statistical difference between the two average rates.[31]

III. Conclusion

The Internet has and continues to fundamentally change social and economic behavior in the United States by substantially reducing the transactions cost of exchange of both a personal and business nature. At this critical time of Internet development, social policy regarding the Internet is extremely important - far too important to be influenced by careless and inept analysis such as that recently conducted by Olbeter and Robison. The problems with their thesis and statistical analysis are numerous - too numerous, in fact, to cover in this brief report. I have only scratched the surface.

The first problem with the analysis of Olbeter and Robison is that it cannot be replicated. Second, the exclusion of particular states from the sample skews the results in a direction desirable by the authors. Third, careless data collection leads to substantial understatement of ILEC lines and hub counts. Fourth, the

---

[30] Again, we caution the reader against accepting the Olbeter-Robison thesis or statistical model for any purposes. These results are simply illustrative of how easy it is to come up with facile policy prescriptions using simplistic and half-baked regression analysis. Using the more limited sample of 46 observations, neither the INLEC share or competition variable is statistically significant.

[31] There are 90 observations of rates from these 10 states. The t-statistic on the means difference test is 0.01.

specification of the model leads to the implausible forecast of hub counts. Fifth, the authors indicate that simultaneity bias may be present but make no attempt to test or correct for the bias. Sixth, the discussion and presentation of the statistical results suggest a lack of expertise by the authors regarding regression analysis. In sum, the recent study by Olbeter and Robison contributes nothing to the debate over Internet or telecommunications policy. Their statistical analysis, from conception to conclusion, is severely defective.

FIGURE 1.

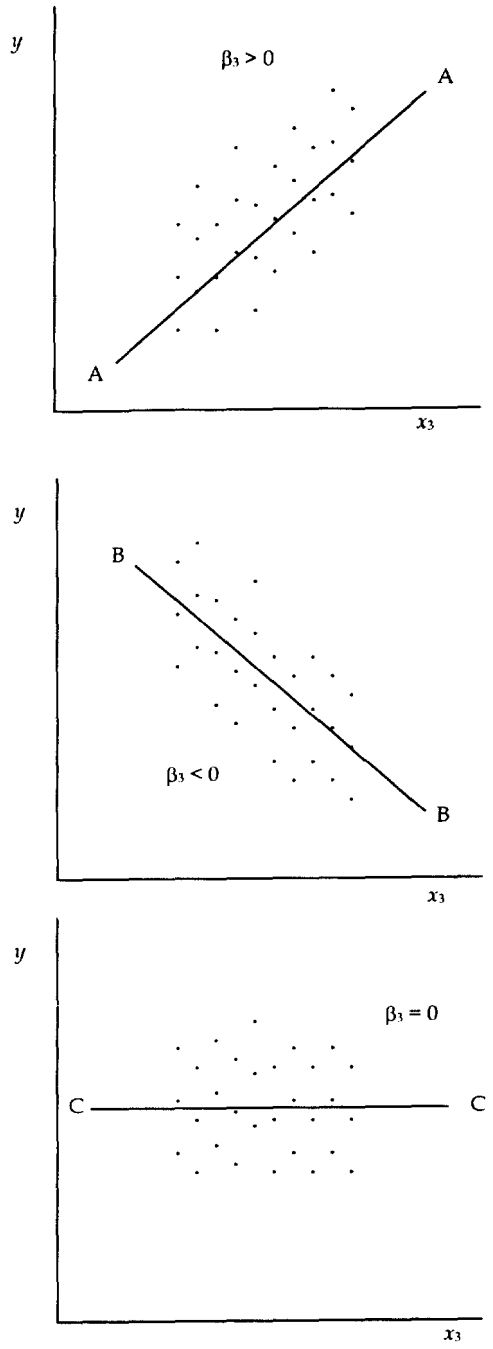

Table 1. Olbeter and Robison's Results

(t-statistics in parentheses)

	Olbeter & Robison
β_0	-34.23 (3.34)*
β_1	3.32 (17.43)*
β_2	0.002 (3.86)*
β_3	27.923 (2.027)*
β_4	-50.29 (0.58)
R^2	0.91
Observations	46

*Statistically Significant at the 5% level or better.

Table 2. Olbeter & Robison's Forecast Hub Counts

STATE	Current Hub Count	Forecast Hub Count	STATE	Current Hub Count	Forecast Hub Count
AL	6	NA	MT	0	26
AK	3	0	NE	4	39
AZ	27	53	NV	6	NA
AR	2	28	NH	3	42
CA	177	205	NJ	24	60
CO	28	54	NM	7	29
CT	2	NA	NY	57	62
DE	2	40	NC	25	47
DC	27	NA	ND	0	24
FL	58	72	OH	56	52
GA	36	45	OK	7	33
HI	3	NA	OR	21	42
ID	2	30	PA	32	49
IL	44	58	RI	4	38
IN	12	48	SC	7	29
IA	3	33	SD	0	30
KS	1	44	TN	9	44
KY	8	33	TX	90	106
LA	8	39	UT	11	30
ME	0	29	VT	0	32
MD	25	44	VA	39	66
MA	29	58	WA	31	46
MI	27	60	WV	0	24
MN	11	42	WI	9	43
MS	3	26	WY	1	30
MO	38	45			

NA indicates the state was excluded from the sample.

Table 3. Comparison of OLS Results

(t-statistics in parentheses)

	Olbeter & Robison	Model 1	Model 2
β_0	-34.23 (3.34)*	-34.61 (3.27)*	-39.29 (3.18)*
β_1	3.32 (17.43)*	3.36 (17.48)*	3.43 (18.33)*
β_2	0.002 (3.86)*	0.002 (3.77)*	0.002 (3.83)*
β_3	27.923 (2.027)*	20.04 (1.52)	12.20 (1.05)
β_4	-50.29 (0.58)	-62.57 (-0.53)	-88.29 (0.73)
R^2	0.91	0.91	0.91
Observations	46 (?)	46	46

* Statistically Significant at the 5% level or better.

Table 4. Sample Exclusions

STATE	HUBS	Percent INLEC Lines
AK	3	100%
CT	2	100%
DC	27	0%
HI	3	100%
NV	6	73%
National Average	20.10	26%

FIGURE 2.

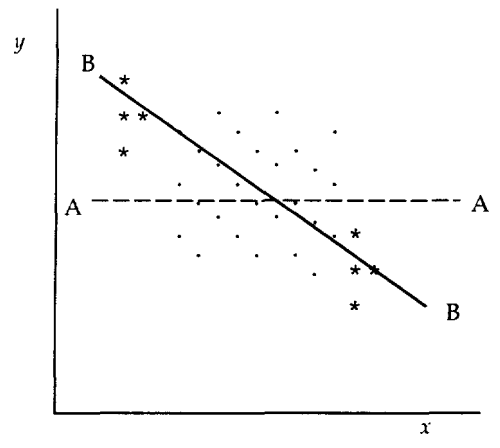

Table 5. OLS with Full Sample

(t-statistics in parenthesis)

Coefficient	Model 1	Model 2
β_0	-34.61 (3.27)*	-15.13 (1.60)
β_1	3.36 (17.48)*	3.50 (17.84)*
β_2	0.002 (3.77)*	0.001 (2.64)*
β_3	20.04 (1.52)	-14.35 (2.30)*
β_4	-62.57 (-0.53)	-49.66 (0.40)
R^2	0.91	0.89
Observations	46	51

*Statistically Significant at the 5% level or better.

Table 6. Forecast Hub Counts with RBOC "Deregulation"

STATE	Current Hub Count	Forecast Hub Count	STATE	Current Hub Count	Forecast Hub Count
AL	6	0	MT	0	0
AK	3	3	NE	4	0
AZ	27	14	NV	6	2
AR	2	0	NH	3	0
CA	177	166	NJ	24	10
CO	28	14	NM	7	0
CT	2	2	NY	57	44
DE	2	0	NC	25	18
DC	27	13	ND	0	0
FL	58	50	OH	56	47
GA	36	24	OK	7	0
HI	3	3	OR	21	11
ID	2	0	PA	32	21
IL	44	32	RI	4	0
IN	12	3	SC	7	0
IA	3	0	SD	0	0
KS	1	0	TN	9	0
KY	8	0	TX	90	79
LA	8	0	UT	11	0
ME	0	0	VT	0	0
MD	25	11	VA	39	28
MA	29	15	WA	31	21
MI	27	15	WV	0	0
MN	11	0	WI	9	0
MS	3	0	WY	1	0
MO	38	27			

Negative values are rounded to zero.

Table 7. Population per Hub based on Olbeter & Robison's "Estimates"

(In thousands)

STATE	Current Population per Hub	Forecast Population per Hub	STATE	Current Population per Hub	Forecast Population per Hub
AL	720	108	NE	414	42
AZ	169	86	NH	391	28
AR	1,262	90	NJ	336	134
CA	182	157	NM	247	60
CO	139	72	NY	318	293
DE	366	18	NC	297	158
FL	253	204	ND	...	27
GA	208	166	OH	200	215
ID	605	40	OK	474	101
IL	270	205	OR	154	77
IN	489	122	PA	376	245
IA	951	86	RI	247	26
KS	2,595	59	SC	537	130
KY	489	118	SD	...	25
LA	544	112	TN	596	122
ME	...	43	TX	216	183
MD	204	116	UT	187	69
MA	211	105	VT	...	18
MI	362	163	VA	173	102
MN	426	112	WA	181	122
MS	910	105	WV	...	76
MO	142	120	WI	574	120
MT	...	34	WY	480	16

Population in Thousands (1998 Statistical Abstract of the US, Table 26, 1997 data).
Hub data is from Olbeter and Robison, pp. 25-5.

Table 8. OLS Results with Population Variable (in thousands)

(t-statistics in parentheses)

	Model 1	Model 2
β_0	-6.10 (3.06)	-2.95 (0.26)
β_1	...	...
β_2	...	0.0001 (0.10)
β_3	...	-6.71 (0.48)
β_4	...	-141.2 (1.45)
Population (β_5)	0.005 (20.04)*	0.005 (16.88)*
R^2	0.90	0.90
Observations	46	46

*Statistically Significant at the 5% level or better.

Table 9. OLS Regressions with Sorted Data

(t-statistics in parenthesis)

	Model 1	Model 2	Model 3
β_0	-39.29 (3.18)*	-39.29 (3.18)*	-39.29 (3.18)*
β_1	3.43 (18.33)*	3.43 (18.33)*	3.43 (18.33)*
β_2	0.002 (3.83)*	0.002 (3.83)*	0.002 (3.83)*
β_3	12.20 (1.05)	12.20 (1.05)	12.20 (1.05)
β_4	-88.29 (0.73)	-88.29 (0.73)	-88.29 (0.73)
R^2	0.91	0.91	0.91
Observations	46	46	46
DW Statistic	0.027	2.26	1.88

*Statistically Significant at the 5% level or better.

Table 10. OLS with Full Sample and Alternative CLEC Measure

(t-statistics in parenthesis)

Coefficient	Estimate
β_0	-23.41 (2.51)*
β_1	2.43 (17.50)*
β_2	0.001 (2.60)*
β_3	-11.40 (1.65)
β_4	10.91 (1.99)*
R^2	0.89
Observations	51

*Statistically Significant at the 5% level or better.

Table 11. Average Monthly Rates for 56K Internet Connection

State	Hubs	Average Monthly Rate (56k Baud)	State	Hubs	Average Monthly Rate (56k Baud)
CA	177	$19.46	ME	0	$18.87
TX	90	$19.86	MT	0	$20.16
FL	58	$18.47	ND	0	$19.95
NY	57	$21.02	SD	0	$19.40
OH	56	$18.17	VT	0	$17.96
Mean		$19.40	Mean		$19.39

Table A. The Data

STATE	INCOME	Cities with more than 100,000 persons	Internet Hubs	Tier One INLEC Lines	All INLEC Lines	Total Lines	CLEC Resold Lines	Percent Lines Served by Colocation Wire Centers
	(a)	(b)	(c)	(d)	(d)	(d)	(e)	(e)
AL	18,155	4	6	237,884	428,440	2,233,362	25,000	45%
AK	22,266	1	3[f]	-	355,185	355,185	-	0%
AZ	19,313	7	27	7,114	158,113	2,414,612	4,000	47%
AR	17,169	1	2	97,401	401,011	1,288,457	15,000	36%
CA	22,964	51	177	3,674,509	3,980,034	19,805,310	290,000	78%
CO	23,300	5	28	-	103,044	2,381,182	16,000	28%
CT	30,935	5	2	2,015,389	2,035,573	2,035,573	31,000	
DE	25,149	0	2	-	-	466,474	7,000	94%
DC	30,933	1	27[f]	-	-	771,630	7,000	79%
FL	21,905	12	58	3,759,046	3,908,462	9,571,502	138,000	58%
GA	20,845	4	36	53,104	702,777	4,275,408	89,000	73%
HI	23,012	1	3[f]	615,288	615,288	615,288	-	0%
ID	17,982	1	2	107,810	163,130	612,755	-	67%
IL	24,310	6	44	1,018,553	1,208,596	7,442,595	201,000	71%
IN	20,488	5	12	1,030,625	1,169,114	3,122,167	5,000	61%
IA	20,214	2	3	261,053	505,906	1,495,268	99,000	21%
KS	20,941	4	1	5,952	241,408	1,486,306	50,000	30%
KY	17,899	2	8	356,384	784,038	1,897,582	21,000	55%
LA	17,841	4	8	-	163,478	2,265,803	44,000	37%
ME	19,088	0	0	-	121,284	754,878	2,000	25%
MD	25,053	1	25	-	5,829	3,052,067	11,000	44%
MA	26,983	4	29	-	3,795	4,151,814	85,000	65%
MI	22,258	8	27	668,455	859,637	5,703,053	155,000	72%
MN	23,263	2	11	3,301	674,569	2,729,586	55,000	65%
MS	15,897	1	3	-	77,964	1,244,747	27,000	35%
MO	20,806	4	38	525,965	738,418	3,064,182	23,000	68%
MT	17,451	0	0	-	147,161	481,698	1,000	0%
NE	20,798	2	4	308,678	426,531	927,923	1,000	66%
NV	23,589	3	6[f]	731,627	779,527	1,074,104	2,000	86%
NH	24,235	1	3	-	45,729	752,763	9,000	68%
NJ	28,302	4	24	180,704	189,400	5,776,498	27,000	40%
NM	17,031	1	7	39,465	120,317	814,166	-	56%
NY	26,452	6	57	820,486	1,189,184	11,562,379	199,000	70%
NC	20,136	5	25	1,522,653	2,099,727	4,166,616	32,000	75%
ND	18,538	0	0	-	146,549	354,244	10,000	0%
OH	21,266	6	56	1,638,953	2,494,138	6,227,640	76,000	79%
OK	17,719	2	7	100,403	311,930	1,822,825	21,000	79%
OR	20,921	3	21	467,317	616,668	1,847,314	45,000	52%

PA	22,496	4	32	1,241,083	1,619,132	7,119,669	71,000	28%
RI	22,280	1	4	-	-	602,318	4,000	76%
SC	18,012	1	7	148,515	652,762	1,962,005	30,000	58%
SD	18,782	1	0	-	127,409	385,081	12,000	0%
TN	19,944	4	9	228,404	606,789	3,071,812	24,000	81%
TX	20,208	23	90	1,681,396	2,301,598	10,678,438	300,000	78%
UT	17,547	1	11	-	42,703	984,594	6,000	74%
VT	20,408	0	0	-	57,939	365,472	1,000	65%
VA	22,861	9	39	825,402	925,961	3,765,373	9,000	32%
WA	22,881	3	31	718,268	1,011,525	3,270,199	46,000	66%
WV	16,498	0	0	-	142,781	846,340	-	0%
WI	21,173	3	9	442,215	1,009,906	3,057,769	30,000	71%
WY	19,541	0	1	-	48,214	274,309	1,000	0%

[a] *Statistical Abstract of the United States* (1998), Table 727.
[b] *Statistical Abstract of the United States* (1998), Table 48.
[c] *Breaking the Backbone*, pp. 25-6.
[d] *Statistics of Communications Common Carriers* (1996), Table 2.3.
[e] *FCC Local Competition Report* (December 1998), Table 3.2 and Table 3.6.
[f] *Boardwatch Magazine* (Spring 1999).