

Effects of Using Specific versus General Data in Social Program Research

Janice A. Hauge*
University of North Texas
P.O. Box 311457
Denton, TX 76208
jhauge@unt.edu

Mark A. Jamison
Public Utility Research Center
University of Florida
P.O. Box 117142
Gainesville, FL 32611

Abstract

We present a comparison of results of similar analyses of a particular social program using two sources of data: one representative of the general population, and one representative of the population that actually is eligible for the social program. To do this, we focus on a particular public assistance program as implemented in Florida as our public policy program of choice. We compare the results of an analysis of the program's participation rates using US Census data for the general population of Florida to the results of the exact same model using a dataset that includes only the population of Florida that is actually eligible for the program. We find that while generally signs of effects remain the same, they do not always remain the same. Moreover, significance differs, and the marginal effects of various demographic and socio-economic factors on program participation rates vary greatly. We submit that such differences are important for policymakers to recognize so that they can effectively target programs to those individuals and geographic areas most in need of such programs.

Keywords: welfare programs, empirical estimation, data limitations

***Corresponding Author**

I. Introduction

The use of accurate data is paramount in properly analyzing many of the complex economic questions being asked today. Generally data are vigilantly collected, and most researchers take great care in ensuring the data collected are accurate, relevant to the question(s) posed, and correctly used in estimations. In some cases, the data a researcher most wants are proprietary or unavailable for some other reason. Such is the case in many analyses of public assistance programs in place today; due to confidentiality requirements for program recipients, information on the characteristics of participants is limited. In such cases, researchers often employ datasets that contain the target population as well as other persons. When this is done, researchers may attempt to account for any data shortfalls through various statistical methods and alternative estimations. For example, a paper by Currie and Yelowitz (2000) uses data from the Survey of Income and Program Participation (SIPP) and the Current Population Survey (CPS), and US Census data to study the effects of public housing on children. The paper provides a cursory analysis of take-up rates but finds the databases used undercount the number of recipients and equally importantly, seem to over-represent White recipients and undercount Hispanic recipients. Similarly, in a study examining the effects of welfare programs on the labor supply, Gensler and Walls (1995) limit the dataset used in an attempt to focus solely on welfare recipients; however, they acknowledge that even within their limited parameters, forty-four percent of the population within their sample received no welfare payments. This implies a substantial subset of their dataset examining effects of welfare programs do not participate in such programs. Ideally, researchers might have preferred to know

exactly who is enrolled in each program and the precise demographic and socio-economic characteristics of these participants, but such data were unavailable.

Using as an example a particular public assistance program, the Lifeline Assistance Program (Lifeline) as implemented in Florida, we consider erroneous conclusions that might be drawn when overly broad datasets are used. Lifeline is a nationwide telephone discount program created by the Federal Communications Commission (FCC) in 1984 and implemented on a state-by-state basis to help low-income households afford telephone service. The Telecommunications Act of 1996 made it compulsory for states to implement Lifeline programs.

Recently there has been interest in why so few low-income households who qualify for the Lifeline Program actually take advantage of it. As of April 2004, only one-third of eligible households in the United States actually subscribed to the program, and states varied greatly in their participation rates: California had a 131.9 percent participation rate, whereas West Virginia had a 3.3 percent participation in 2002 (FCC, 2004).¹ While the FCC includes Lifeline statistics in many of its annual telecommunications reports that provide excellent information on the status of the Lifeline program, the reports provide only limited analysis of factors that drive the rate of participation. As with other public assistance programs, the participation rate is of critical interest to policymakers.

Our goal in this paper is not an analysis of the Lifeline program per se, but rather a comparison of results of similar analyses using different sources of data.² We

¹ Due to the manner in which the state of California enrolls Lifeline subscribers, more than 100% of eligible households enrolled in the program, i.e., some enrolled despite not meeting eligibility criteria.

² Our formal analysis of the Lifeline Assistance Program in Florida is currently under review; our analysis of the Lifeline Assistance Program nationwide is forthcoming in *Public Finance Review*.

simply use the Lifeline program as our public policy program of choice to compare the results of an analysis of Florida's Lifeline participation rates using Census data for the general population of Florida to the results of the exact same model using a dataset that includes only the population of Florida that is actually eligible for the Lifeline program. We find that while generally signs of explanatory variables remain the same across models using different datasets, they do not always remain the same; moreover, the significance of effects differs across datasets, and the marginal effects of various demographic and socio-economic factors on program participation rates vary greatly. We submit that such differences are important for policymakers to recognize so that they can effectively target programs to those individuals and geographic areas most in need of those programs.

The paper proceeds as follows. Section II describes the Lifeline program and literature in which more general datasets have been used to answer public program participation questions. Section III provides a theoretical model to explain why key estimates differ when various independent variables differ, and details the empirical models from which our results are drawn. Section IV provides an explanation of our datasets that are used for the comparison, and Sections V and VI provide the results and conclusion, respectively.

II. Background and Literature Review

Background on Lifeline

In Florida, Lifeline provides low-income households with a discount of up to \$13.50 on the price of basic local telephone service.³ Under the FCC guidelines that pertain to all US states, there are four tiers of monthly federal Lifeline support. The first tier of federal support is a credit (\$6.50 at the time of this study) available to all eligible households for the federal subscriber line charge. The second tier of federal support is a \$1.75 credit. The third tier of federal support is one-half the amount of additional state support up to a maximum of \$1.75 in federal support. Because Florida's telecommunications carriers that qualify to participate in Lifeline provide an additional \$3.50 credit to Lifeline customers' bills, Florida Lifeline subscribers during the period of our study (2003-2005) received a total monthly credit of \$13.50. At no time is the consumer's bill for local service less than zero. The fourth tier of federal support is available only to eligible consumers living on Native American tribal lands, but no tribal lands in Florida qualified for this support at the time of our study.

The eligibility criteria for Florida's Lifeline program evolved moderately during the time of our study. General guidelines, however, held that a household qualified for Lifeline if it was at or below 125 percent of the federal poverty guidelines (this increased to 135 percent in 2005) or participated in one or more of the following programs: National School Lunch Program, Temporary Assistance to Needy Families

³ During the time period for this study (2003-2005), some bills of Florida's Lifeline customers show Lifeline credits to be in excess of \$13.50. As part of the telecommunications price rebalancing in Florida, BellSouth (recently merged with AT&T), Sprint (now called Embarq), and Verizon were restricted by statute from increasing local telephone prices for Lifeline customers. BellSouth implemented this restriction by providing a credit on the bills for its Lifeline customers equal to the increase in local telephone prices that occurred with BellSouth's price rebalancing. Instead of providing a new credit on their customers' bills, Sprint and Verizon increased the Lifeline credit on the bills of their Lifeline customers, so the Lifeline credits to their customers' bills appear to exceed \$13.50.

(TANF), Medicaid, Food Stamps, Supplemental Security Income (SSI), Federal Public Housing Assistance (Section 8), Low Income Home Energy Assistance Program, and Bureau of Indian Affairs Programs. To participate in Lifeline, a household must sign up and verify that it is eligible to participate. The processes for signing up and for verifying eligibility have varied over time but have not changed so dramatically that they might cause significant changes in participation rates in Florida.⁴

Literature Review

There is a wealth of research considering the Lifeline program and associated Universal Service Fund programs.⁵ Using general population data, Burton and Mayo (2005) concluded that restrictions on Lifeline subscribers, such as the inability to subscribe to second lines or to services such as call waiting, reduced the probability of an eligible household signing up for Lifeline. They found similar impacts from higher costs of enrollment in Lifeline and lower local telephone prices. Also using general population datasets, Hauge, Jamison, and Jewell (2007) found that higher local telephone prices, greater Lifeline discounts, higher education levels for the head of household, and higher concentrations of households on public assistance were positively correlated with Lifeline participation.

Furthermore, Garbacz and Thompson (1997, 2002, and 2003) focus on Lifeline and Link-Up and on estimating demand for telecommunications services.⁶ These

⁴ Holt and Jamison (2006) provide a history of Florida's eligibility criteria.

⁵ The Universal Service Fund (USF) is an FCC program amended by the Telecommunications Act of 1996 to provide various levels of support to programs designed to ensure all US citizens have affordable access to basic telecommunications service. The USF covers four programs: high-cost support for incumbents operating in financially uneconomic areas; Lifeline and Link-Up to reduce low-income consumers' basic telephone expenses; schools and libraries to reimburse telecommunications providers for discounts on services provided to schools and libraries; and rural health to provide discounts to rural health care providers for telecommunications and Internet services.

⁶ Link-Up provides a one-time subsidy to low-income households to help initiate telephone service.

papers show that Lifeline and Link-Up have a negligible effect on telephone penetration rates and that the price elasticity of demand for telecommunications service is very low. They do not attempt to explain how governments and telecommunications carriers might more effectively target beneficiaries. Of additional interest for our purposes among these papers is the data employed. In each case, the authors incorporate the characteristics of the general population rather than characteristics of those individuals the policy is designed to assist.

Studies of telecommunications policies are not alone in their use of “best available” data. A number of papers study take-up rates for programs other than universal service. For example, Blundell, Fry, and Walker (1988) analyze take-up rates for a program in the United Kingdom that is analogous to the public housing program in the United States. The authors comment on data integrity issues inherent in UK government surveys that are similar to data integrity issues in some US government surveys; however, they do not consider the population of the dataset to be a potential source of error. Similarly Currie and Yelowitz (2000), and Bitler *et al* (2003) address problems with the datasets they use. However in each case, the authors focus on deficiencies in the datasets; they do not address the issue of including a population that is not directly relevant for the study. Our research concentrates on just this issue. We submit that even if the Census data perfectly represented a geographic population, it would not be appropriate to use complete socio-economic and demographic data to study a public assistance program that applies only to a subset of the population (in this case, the low-income population). Therefore, we examine how closely the results of our policy analysis using Census data match results of the same policy analysis using

alternative data, more specifically, data on the eligible population rather than the total population. The next section will discuss the models employed to study the Lifeline program, followed by the theoretical implications of using general data.

III. Theoretical Models

Lifeline Participation Models

To analyze Lifeline participation rates in our initial research, we used a theoretical model of household utility maximization and an empirical model that includes measures of demographic factors for eligible households at the county level and company-specific measures to ascertain the determinants of the Lifeline participation rate.⁷ The model is described briefly as follows. Assume that household i located in county j maximizes per period utility, given in equation (1), subject to the budget constraint given in equation (2).

$$(1) \quad U_{ij} = U[T_{ij}, Z_{ij}]$$

$$(2) \quad I_{ij} = (P_j \times T_{ij}) + Z_{ij}$$

Utility is a function of telephone services (T) and consumption of a composite good (Z). Income (I) and the price of telephone services (P) are exogenous, and the price of the composite good is normalized to one. Further assume that this household is eligible for the Lifeline program, which allows the household to purchase a fixed amount of telephone services (L) at a discounted price. The price of telephone service when receiving the Lifeline benefit is the difference between (P) and the amount of the discount (S). In addition, it is possible there is some degree of stigma attached to

⁷ This model is from the paper “Participation in Social Programs by Consumers and Companies: A Nationwide Analysis of Participation Rates for Telephone Lifeline Programs” by Hauge, Jamison, and Jewell, forthcoming in *Public Finance Review*, 2007.

Lifeline participation that may be a function of income and other individual characteristics. Assume the Lifeline stigma cost (C) must be subtracted from a household's utility if the Lifeline subsidy is accepted, and assume that stigma cost can vary across households.⁸ For simplicity, assume that program knowledge (K) is a binary variable equal to one if a household knows about the Lifeline program and equal to zero if the household is uninformed.⁹ Under the above assumptions, utility maximization implies the following decision rule for Lifeline participation:

$$(3) \quad \begin{array}{ll} \text{participate} & \text{if } K_{ij} \times \{U[L_{ij}, I_{ij} - (P_j - S) \times L_{ij}] - C_{ij}\} \geq U[T_{ij}, I_{ij} - P_j \times T_{ij}] \text{ and} \\ & \text{do not participate} \quad \text{otherwise.} \end{array}$$

Thus, a household will choose to participate in Lifeline if and only if the utility associated with participating is greater than or equal to the utility associated with not participating and if and only if the household is aware of the program. Label the utility difference in equation (3) y_{ij} , and assume it takes on a linear functional form, so that the decision rule becomes the following:

$$(4) \quad \begin{array}{ll} \text{participate} & \text{if } y_{ij} = \gamma x_{ij} + e_{ij} \text{ and} \\ & \text{do not participate} \quad \text{otherwise.} \end{array}$$

With household-level data, the vector of household-level parameters γ is normally estimated using either probit or logit, depending on the assumed distribution

⁸ We expect stigma costs to be low given studies that have shown stigma decreases with program participation and the primary manner of proving eligibility for Lifeline is through proving participation in another welfare program. Furthermore, a household can enroll from the privacy of their home, avoiding any stigma associated with visiting a welfare office. In addition to stigma, the Lifeline program carries some effort cost of enrolling. In Florida, most enrollment is accomplished through checking a box on a form received upon receipt of any other qualifying welfare service. Additionally, applicants can request Lifeline online through the Department of Children and Families.

⁹ Surveys conducted by the Public Utility Research Center at the University of Florida reveal a lack of information about the program to be a main reason for non-participation. Surveys are available at <http://www.purc.ufl.edu>. The surveys include interviews of Floridians in person and over the telephone, as well as written surveys of households that qualify but do not participate and those that qualify and had disconnected their telephone service.

of e . However, we do not observe household Lifeline choices in our data. Instead, we observe the number of Lifeline households that participate out of the number eligible within each Florida county.¹⁰ Although we do not observe household decisions, we can use the utility maximization model discussed above to motivate our county-level empirical analysis. If the data are generated in the manner given in equation (4), then the determinants of Lifeline participation at the county level will be the determinants at the household level (i.e., components of the matrix x_{ij}) aggregated up to the county level. Note that although the determinants at the county level are assumed to be the same as the determinants at the household-level, the coefficients from our county-level analysis cannot be interpreted as household-level effects due to aggregation issues. Specifically, we cannot recover the vector γ of household-level parameters.

We observe the number of positive outcomes (Lifeline subscribers) based on a total number of potential positive outcomes (eligible households); thus, our outcome variable is grouped in percentage terms: the number of households participating in Lifeline divided by the total number of eligible households. Models with grouped data are normally estimated with weighted least squares, as weights are needed to account for the heteroskedasticity associated with observations being clustered by county. We employ a minimum logit chi-square specification in which the dependent variable is the logit of the Lifeline participation rate, i.e., the natural log of the Lifeline participation rate divided by one minus the Lifeline participation rate (Maddala, 1983, p. 30; Papke and Wooldridge, 1996, p. 620; Greene, 2003, p. 687).¹¹

¹⁰ The number of Lifeline participants is provided by each company at the city level. We aggregate to the county level to correspond to the level of the eligible household data.

¹¹ The weights are $1/[n_i p_i (1-p_i)]$, where n_i is the total eligible households of county i , and p_i is the logit probability of the Lifeline participation rate in county i . As discussed by Greene (2003), p_i must be

We construct our model using a complete panel of observations on all 67 Florida counties for the years 2003 through 2005.¹² The panel nature of the data allows us to use a random effects estimator.¹³ In the random effects model, county-specific effects measuring unobservable county characteristics are modeled and estimated as being randomly distributed across counties. Our random effects specification is given in equation (5), where j indicates county, t indicates year, and ρ is the Lifeline participation rate.

$$(5) \quad \ln(\rho_{jt}/(1 - \rho_{jt})) = \alpha + X_{jt}\beta + (u_j + \varepsilon_{jt})$$

The time-invariant, county-specific effect (u_j) is modeled and estimated as part of a well-behaved, normally distributed error term ($u_j + \varepsilon_{jt}$). The matrix X_{jt} contains household and county measures for county j in year t , and the vector β and the constant term α represent parameters to be estimated. The equation is estimated using weighted feasible generalized least squares (FGLS), allowing the variance of u to vary by counties to control for heterogeneity across panels (Greene, 2003).

estimated since it is a function of unknown parameters. Following the prescribed procedure, we estimate p_i using ordinary least squares (OLS) in a first stage because all that is needed is a consistent estimate of p_i . The weights are then computed and used in a second-stage feasible generalized least squares estimation. Several specifications of the dependent variable are available. We choose the minimum logit chi-square estimator due to its similarity to the well-known logit dichotomous choice model. To test robustness, we estimated the model using another commonly employed specification in which the dependent variable is simply the log of the participation rate, finding no major differences between the two specifications. There are 13 year/county observations with zero subscribers. Rather than delete these observations, the number of subscribers is arbitrarily assigned to one in order to compute the dependent variable. Although clearly ad hoc, this solution is the most viable in our situation. We also include a dummy variable in our estimation to account for the 13 observations; the dummy variable (not reported for brevity) is significantly negative.

¹² Some county-level Lifeline participation data are available back to 2000; however, prior to 2003 data is incomplete. We choose to analyze the most-recent three years since they represent complete panels. Summary statistics for 2000-2005 are available from the authors.

¹³ A Hausman test indicates that the assumption of the random effects model concerning the orthogonality of the random effects and the regressors is appropriate. The chi-square statistic (15 d.f.) is 20.88, which is insignificant at any conventional level. Thus, we cannot reject the null of no correlation between the random effects and the regressors. Complete results of this test are available from the authors. We choose the random effects model because it allows the inclusion of time-invariant regressors and is more efficient than a fixed-effects model.

In pursuing this research, we might have used data representative of the general population of each county. Alternatively, we could use data representative of Lifeline eligible households. It is this choice that we analyze in this paper. We develop the theoretical model of why this choice is important next.

Bias in Using Population Characteristics

We introduce a framework to understand the biases introduced when a participation model is estimated using population data (in our case Census data) rather than data on the subgroup of the population eligible for the program (in our case Lifeline).¹⁴ Suppose the true model is given by:

$$(6) \quad t_c^e = \beta_0 + \beta_1 x_c^e + \varepsilon_c.$$

where c indexes the county and t^e represents the take-up by eligible subscribers, x^e some characteristic (for example race, ethnicity, or years of education) among eligible subscribers, and ε is a random error term. Dropping the county subscripts for convenience, note that:

$$(7) \quad x = px^e + (1 - p)x^n$$

where x is a characteristic among the whole county, p is the proportion of eligible subscribers in the county, and x^n is a characteristic among non-eligible people in that county. Assuming that p is constant (i.e., the fraction of eligible subscribers across all counties is the same), we can substitute (7) into (6) to get:

$$(8) \quad t^e = \beta_0 + \beta_1 \left[\left(\frac{1}{p} \right) x - \left(\frac{1-p}{p} \right) x^n \right] + \varepsilon$$

$$= \alpha_0 + \alpha_1 x + v$$

¹⁴ Notes on this framework are attributed to Damon Clark, Assistant Professor of Economics at the University of Florida.

where α_0 is equal to β_0 , α_1 is equal to $\frac{\beta_1}{p}$, and $v = -\beta_1 \left(\frac{1-p}{p} \right) x^n + \varepsilon$. It can be shown that significant biases may be induced when estimating equation (8)¹⁵.

These biases may affect estimation results to differing degrees depending on the parameters involved; however, some generalizations can be made. First, as the proportion of eligible subscribers approaches one, using the population characteristic will give a consistent estimate of the true effect; this is intuitively logical. Second, for the proportion of eligible subscribers less than one, since the variance of x^e (some characteristic of the eligible population) is less than the variance of x , the bias will generally underestimate the effect. Furthermore, in cases in which the proportion of eligible subscribers is small, since the covariance of x^e and x^n is likely to be positive (for example common factors drive education rates among the eligible and non-eligible populations within a county), using the population characteristics may not only underestimate the effect, but may give the wrong sign. We assert that such is the case in our analysis of the Lifeline program as it applies to Florida.

IV. Data

We include in our estimates explanatory variables in three categories: measures of the telecommunications environment, characteristics of the eligible population, and county-level socioeconomic characteristics. We do not here reiterate reasons for choosing the independent variables in our model; those details can be found in our 2007 paper.¹⁶ Note that prior welfare participation studies guide our empirical models

¹⁵ The formal proof is in the Appendix.

¹⁶ Hauge, J. A., Jamison, M. A. and Jewell, R. T. (2007) Discounting telephone service: an examination of participation in the Lifeline assistance program using panel data, Working Paper No. 06-06, University of Florida, Public Utility Research Center, under review.

and serve as useful references for predicting our results. Here we take our initial model as given and focus solely on the results of the different datasets.

In comparing the estimates from the Census data to those from the data on eligible households, note that two of the categories (measures of the telecommunications environment, and county-level socio-economic characteristics not attributed to characteristics of subscribers) are the same in both estimates. Half of our explanatory variables – *Percent BellSouth* through *Percent Welfare* as listed in Table One – are constant across specifications because they either were not specific to the eligible population or we lacked measurements specific to the eligible population; the other half differ solely in whether they are specific to the eligible population or relate to the general population. Table One provides brief descriptions of the data used in the analyses. The variables representing the percentages of the population owning their own homes, having a high school degree or less than a high school degree, of white race, of female gender, and of various ages differ as they are from different sources (the US Census and our eligible households datasets). Variables from the two different sources are distinguished in Table One by an asterisk included in the variable description.

The primary difference then in the data is due to the inclusion of specific demographic information with respect to households eligible to receive Lifeline. It is clear from Table One that the use of characteristics of eligible households adds a unique dimension to the original Lifeline study, allowing the results to contribute to the literature by focusing on characteristics of households the program is specifically designed to help. In other words, a key factor in our initial Lifeline research was the

inclusion of the number of households and the characteristics of those households qualifying for Lifeline; those measures reflect the distribution of characteristics within the eligible population, not necessarily within the county as a whole.

Because the difference in data is critical to our current hypothesis, we briefly outline the data sources used. The Census data are taken directly from the statistical files available at www.census.gov, Summary File 3. Summary statistics of the Census data appear in columns three and four in Table One. The data for eligible households was attained in a more complex manner. Specifically, we employ data from Williamson's (2006) study of Lifeline-eligible households in Florida. The study describes how the demographic data on eligible households was determined through a combination of statistical techniques and public program data.¹⁷

Compiling accurate data of eligible participants was essential to the Lifeline study and was a complex task. One method by which prospective participants could prove eligibility for the Lifeline program was through verification of low income. Florida's income eligibility criterion changed during the years of the study (2000 – 2005) from a requirement that household income be at or below 125 percent of the federal poverty guideline (FPG) to at or below 135 percent of the FPG in 2005. Florida also allowed (and still allows) prospective participants to prove eligibility through receipt of certain other welfare programs regardless of the income criterion. Those programs include Bureau of Indian Affairs Programs, Federal Public Housing Assistance (Section 8), Food Stamps, Low Income Home Energy Assistance Program,

¹⁷ Holt, L. and Jamison, M. A. (2006) Making telephone service affordable for low-income households: an analysis of Lifeline and Link-up telephone programs in Florida, Working Paper No. 06-05, University of Florida, Public Utility Research Center. The full paper is available at <http://www.purc.ufl.edu>.

Medicaid, National School Lunch Program, Supplemental Security Income (SSI), and Temporary Assistance to Needy Families (TANF).

There were a number of difficulties to overcome in determining the number of eligible households and in defining characteristics of such households. One such difficulty was the possibility of double-counting a particular household. For example, a household receiving TANF initially would be counted as eligible for Lifeline; that same household also receiving SSI would be counted as eligible under that program's numbers, therefore being counted twice for Lifeline eligibility. Another difficulty was that eligibility criteria differed among the other welfare programs so that it was necessary to account for differences in qualifications across programs in addition to accounting for households who may have been receiving benefits from multiple programs. Careful analysis was undertaken to ensure eligibility estimates based on the various social programs resulted in the number of unduplicated households meeting the criteria to receive Lifeline assistance.¹⁸ Summary statistics of the eligible population data appear in columns five and six in Table One.

It is immediately evident that the characteristics of the county populations differ from the characteristics of the eligible population. To ascertain whether the apparent differences in the data are statistically significant, we ran a series of two sample mean comparison tests assuming unequal variances between Census and eligible data. We found that in all cases the means are statistically different from each other at any

¹⁸ Williamson worked with various agencies as well as undertaking her own statistical calculations. For example, the Department of Children and Families supplied participation data for the TANF, Medicaid, and Food Stamp programs for the years 2003 through 2005. These data were given by individual recipients; therefore, a case study number (typically applied by household) and further calculations were used to estimate eligible households. Additionally, data for recipients of Section 8 Housing assistance from 2001 through 2005 were compiled using primary data provided by Housing and Urban Development (HUD). Details of the methodology for estimating the total number of households and various characteristics of heads of household are provided in Williamson (2006).

conventional level.¹⁹ Our subsequent analysis considers the degree to which these differences affect the results of our participation rate model.

V. Results

The results of the weighted FGLS estimations are presented in Table Two. Due to the non-linear construction of the dependent variable, the coefficients cannot be interpreted as marginal effects. We include marginal effects calculated as elasticities for ease of comparison.²⁰ Interestingly, only one of the eight independent variables from the different datasets maintained its sign and significance across both models (*Percent Female*). Three for which there was no effect in the original (eligible) model gained significant signs using Census data (*Percent White, Percent Age 25-54, and Percent Age 55-74*). One variable became insignificant (*Percent High School*) when using the Census data. More disturbingly though are the three variables that changed signs and were significant in both models: *Percent Own Home, Percent No High School, and Percent Age 75+*. These results are disturbing due to the implications for policy. For example, consider conducting a study of the Lifeline program in Florida without the benefit of the dataset developed by Williamson (2006). Results from an estimation using Census data might lead us to suggest to those responsible for improving enrollment or notification procedures that there is no specific age group that they should focus on reaching. Our analysis of the eligible data, however, suggests differently: those over age 75 are less likely to participate than others. For practical purposes, this difference might be significant. For example, one of the changes Florida

¹⁹ Results are available from the authors upon request.

²⁰ The marginal effects are simulated by changing the relevant independent variable, recomputing the predicted rate for each individual, and comparing this new prediction to the rate predicted from the original sample.

has made in an effort to increase Lifeline participation is to provide the ability to enroll via the Internet through the Florida Department of Children and Families (DCF) website. While this ability is likely to increase total enrollment, it is less likely to do so for the population age 75 and over who typically are less computer savvy and might have less occasion to visit a DCF office. If being elderly is negatively correlated with participation, policymakers and telephone companies might re-direct marketing efforts to increase the presence of Lifeline Program information in senior centers throughout the state. Consider another example using the results for cellular telephones.

Policymakers' concern over low Lifeline participation may require greater policy changes than altering enrollment criteria. Currently cellular telephones are not eligible for Lifeline, and we find the number of cellular telephones to be negatively correlated with participation. As the percentage of cellular usage rises, policymakers might consider expanding the Lifeline program to include discounts on cellular telephone bills as well.²¹ Finally, the degree to which policy changes might affect different groups is significant. In terms of absolute values, our analysis of the eligible data consistently provides larger estimates of marginal effects than does our analysis of the Census data, so the lessons for policymakers are more meaningful.

Results involving those characteristics comparable in both estimates are somewhat less troubling than the home ownership, age, and education variables. Three coefficients maintain the same sign and significance across estimates (*Percent BellSouth, Percent Verizon, and Percent Welfare*), while three lose their significance (*Local Phone Rate, Cell Phones, and Percent Rural*) with the Census model. Only

²¹ Garbacz and Thompson (2005) suggest that in developing countries, universal service might be promoted more effectively with subsidies for cellular telephones.

Percent Sprint displays a new (negative and significant) effect when considered in the Census model. While these differences are not as dramatic, which might be expected given that the data across models is the same, conclusions that we draw from these estimates may still be incorrect. For example, we might assert that the local telephone rate has very little effect on participation in Lifeline, perhaps due to highly inelastic demand for phone service and already high penetration rates. Certainly that idea has been promoted in a number of other papers (for example the 2002 and 2003 studies by Garbacz and Thompson). Our analysis of eligible households, however, does indicate that the local telephone rate is important in eligible households' participation decisions. While current research typically cites highly inelastic demand for telecom services, no formal research has examined whether these elasticities differ at different income levels. In other words, while middle-income households may have inelastic demand for telecommunications services, it is possible low-income households have a significantly higher demand elasticity. Such information might cause policymakers to reevaluate the methods by which they intend to assist those most in need.²²

Rather than simply asserting that the results of one model must be “better” due to perhaps more accurate data, we examine our results in light of evidence from prior research on participation in welfare programs. If our subsequent examination finds that our results using the eligible population data are more aligned with prior research than

²² The manner by which the local telephone rate might impact household participation decisions remains an unresolved question. We suggest that if a poorer household is more likely to sign up for any welfare program, and an increase in the local telephone rate makes the household poorer, the reported relationship would hold. This would require an income effect off-setting the substitution effect. It is beyond the scope of this paper to examine these effects and the associated demand elasticities of households of different income levels; however, we believe it an important question for further research.

our results using the Census data, we would then have greater confidence in the accuracy of the results obtained using the eligible data.

In light of our goal to examine data differences, we now compare the results from our models with existing literature by considering the variables presented in Tables One and Two in the order in which they appear. With respect to the effects of the incumbent telecommunications providers (BellSouth, Sprint, and Verizon), we see a difference only in *Percent Sprint*, which initially is insignificant and becomes negative and significant with Census data. It is beyond the scope of this paper to examine different incumbents' policies; however, we believe that part of the effect shown is due to marketing efforts by the incumbents. We expect that greater marketing will increase participation in the Lifeline program in Florida. While it has been asserted that non-participation may simply be a choice (Andrade *et al.*, 2002), we expect that due to program characteristics not fully described here, eligible Lifeline beneficiaries tend to participate if they know about the program. This view is supported by evidence from the previously referenced surveys conducted by the Public Utility Research Center at the University of Florida, analyzing various aspects of participation in Lifeline. The belief that enrollment efforts of incumbents are relevant is supported by results in both models.

Studies of welfare participation overwhelmingly support the assertion that as people perceive the value of an entitlement to rise, the probability of participation in the entitlement program increases. (Andrade *et al.*, 2002; Pan *et. al.*, 2006). We submit that households that are relatively poorer (by virtue of facing higher local telephone rates) are more likely to participate in the program. This result as reflected in the *Local*

Phone Rate variable is found when using the eligible data, however not when using Census data. Additionally, because the Lifeline subsidy was applicable only to landline telephones during the period of our study, we expect a negative relationship between cell phones and Lifeline participation. This is borne out in the eligible data model, however not in the Census data model. This finding challenges researchers to examine substitution of wireless for wireline by population segment.

With respect to our measure of rural environment, we again find evidence to be stronger in the eligible model. If lack of information is a primary reason for lack of participation, and it is more costly for telecommunications companies to market the Lifeline program in rural areas than in urban areas, we would expect a negative correlation between *Percent Rural* and participation. Alternatively, we may assert that rural households, even when aware of Lifeline, do not participate to the same degree as those in urban areas. This possibility is supported by other research finding that rural areas have a higher level of non-participation in welfare programs than urban areas (for example see Pan *et al.*, 2006) Lower participation in rural areas is reflected in the eligible data model but not in the Census data model.

With respect to participation in welfare programs, past research has found that such participation is higher in states with relatively more recipients of any government assistance program (McGarry, 1995; Yelowitz, 2000). Because of the manner in which Lifeline participants frequently prove eligibility (by proving receipt of another welfare program) and because of the wealth of other studies reporting such correlation,

participation in Lifeline should be positively correlated with *Percent Welfare*. We find this to be the case in both models.²³

With respect to the variables for which data differs between the models, we first address owning one's home. While prior studies are ambiguous as to the effects of home ownership on welfare participation, we submit that owning one's home should be positively correlated with participation. Enrolling in Lifeline occurs generally when a householder first connects his or her telephone service. Enrolling, while not overly burdensome, adds to the effort expended while connecting, disconnecting, and reconnecting telephone service. A homeowner may not wish to exert such additional effort for a temporary subsidy. If, however, a homeowner plans to remain in his own home for an extended period of time, that effort to connect the telephone service and sign up for Lifeline would result in a benefit for months if not years into the future. This assertion is supported by the results of our eligible data model, however not by the results of the Census data model. We do find some tangential evidence of this effect in Pan *et al.* (2006), in which the authors find the probability of moving to be positively correlated with leaving a particular welfare program in Iowa.

The next variables measure education. Studies evaluating participation in other welfare programs have concluded that participation falls with education (Andrade *et al.*, 2002). Again, this result is borne out in the eligible data model, however not in the Census data model. There is a greater negative correlation with higher education than with less education.

²³ Interestingly, a 2004 paper by Kang, Huffman and Jensen analyzing TANF found that an increase in non-labor income decreases welfare participation. For our study, this would suggest that participation in TANF, Food Stamps or SSI, for example, would decrease participation in Lifeline. Further research into this possibility seems warranted.

In keeping with prior studies, we expect that race and gender may affect costs and preferences (Blundell, Fry, and Walker, 1988; Blank and Ruggles, 1996; Hoynes, 1996). While different studies find different effects of race, we cannot predict greater or less participation. We find in our models that both reflect positive coefficients on *Percent White*; the Census data model finds the result to be significant. We therefore can support the Census data model with respect to race. With respect to gender, some studies of welfare program participation focus on female heads of households (e.g., Moffitt, 1983; Blank, 1985; Fraker and Moffitt, 1988; Gensler and Walls, 1995; Blank and Ruggles, 1996). In general, these studies find that factors other than gender more frequently determine participation; among such factors is single-parenthood. Since over 80 percent of single parents are women, we expect Lifeline participation to be positively correlated with the percentage of the population that is female.²⁴ This result is found in both models.

Finally, results of prior studies generally find that participation in welfare programs declines with age (Blundell, Fry, and Walker, 1988; Stuber and Kronebusch, 2004). This result is marginally reflected in our eligible data model; it is completely opposed in the Census data model.

Because of the manner in which we estimate the models, there is no standard measure to determine “goodness of fit.” To attempt to determine the strength of the models, we estimated a random effects model using a generalized least squares estimator. While we believe the feasible generalized least squares method to be the most appropriate for our data, the former model provides an R^2 , which can be used as

²⁴ In 2003, there were 12,402,000 single parents in the United States. 81.78 percent (10,142,000) were single mothers. (US Census 2004).

an indicator of goodness of fit. In the model using the Census data, the R^2 is 0.3316; in the model using the eligible population, the R^2 is 0.4324. While not a definitive result, this does provide us with additional confidence in our assertion that using the data specific to the eligible population provides a more accurate result than using data reflecting characteristics of the entire population.

VI. Conclusion

In summary, we reiterate the importance of choosing the most applicable data, and stress that it is imperative that researchers take great care in more than just construction of their empirical models. Using one representative public policy program, we have shown how different conclusions can be drawn given different sources of data; one source is arguably more applicable than the other in that it represents the subset of the population relevant to the study. We suggest that the most readily available or most widely used data does not necessarily imply the best data for any particular study. Further studies might focus on re-examining other public assistance programs with respect to their effects on the low-income population, or the specific population a particular policy is designed to assist.

Table One
Low-Income Households
County-Level Summary Statistics
N = 201

Variable	Variable Description	Census		Eligible	
		Mean	Std.Dev.	Mean	Std.Dev.
Lifeline Subscribers	Total Number of Lifeline Subscribers	2,186	6,039	2,186	6,039
Eligible Households	Total Number of Eligible Households	15,335	27,565	15,335	27,565
$\ln(\rho_{it}/1 - \rho_{it})$	Logit of the Lifeline Participation Rate	-2.60	1.58	-2.60	1.58
<i>Percent BellSouth</i>	Percent of Telephone Service Provided by BellSouth	32.66	41.46	32.66	41.46
<i>Percent Sprint</i>	Percent of Telephone Service Provided by Sprint	39.97	42.61	39.97	42.61
<i>Percent Verizon</i>	Percent of Telephone Service Provided by Verizon	6.76	20.09	6.76	20.09
<i>Local Phone Rate</i>	Average Monthly Charge for Single Residential Line	9.85	1.22	9.85	1.22
<i>Cell Phones</i>	Average Number of Cell Phones per Household	1.14	0.62	1.14	0.62
<i>Percent Rural</i>	Percent Rural Households	41.19	33.30	41.16	33.64
<i>Percent Welfare</i>	Percent Households Receiving Government Assistance	2.60	1.10	2.60	1.10
<i>Percent Own Home</i>	*Percent of Households Owning Home	76.28	7.34	58.23	11.26
<i>Percent No High School</i>	*Percent Heads of Households Not Finishing High School	10.57	27.02	43.16	7.92
<i>Percent High School</i>	*Percent Heads of Households with High School Degree Only	65.03	10.19	31.44	3.33
<i>Percent White</i>	*Percent Heads of Households who are White	80.25	10.20	66.53	15.19
<i>Percent Female</i>	*Percent of Heads of Households who are Female	49.11	3.50	54.39	4.57
<i>Percent Age 25-54</i>	*Percent of Heads of Households Age 25 to 54	51.59	8.36	44.57	6.07
<i>Percent Age 55-74</i>	*Percent of Heads of Households Age 55 to 74	31.15	5.77	29.17	4.86
<i>Percent Age 75+</i>	*Percent of Heads of Households Age 75 and Up	12.88	4.96	17.32	4.93
<i>Year</i>	Year	2004	0.82	2004	0.82

Table Two
FGLS Results
Dependent Variable = Logit of Lifeline Participation Rate
N = 201

Variable	Census			Eligible		
	Coefficient	S.E.	Marginal Effect	Coefficient	S.E.	Marginal Effect
<i>Percent BellSouth</i>	0.0015***	(0.0004)	0.001	0.0022***	(0.0007)	0.062
<i>Percent Sprint</i>	-0.0012**	(0.0005)	-0.001	0.0010	(0.0006)	0.033
<i>Percent Verizon</i>	0.0034***	(0.0006)	0.001	0.0050***	(0.0008)	0.029
<i>Local Phone Rate</i>	-0.0002	(0.0002)	-0.021	0.0358**	(0.0150)	0.306
<i>Cell Phones</i>	0.0003	(0.0006)	0.003	-0.1215***	(0.0370)	-0.120
<i>Percent Rural</i>	-0.0013	(0.0013)	-0.005	-0.0188***	(0.0021)	-0.666
<i>Percent Welfare</i>	0.8686***	(0.0666)	0.091	0.5488***	(0.0317)	1.219
<i>Percent Own Home</i>	-0.0112*	(0.0061)	-0.074	0.0513***	(0.0063)	2.613
<i>Percent No High School</i>	0.0114*	(0.0058)	0.024	-0.0102*	(0.0054)	-0.381
<i>Percent High School</i>	-0.0022	(0.0063)	-0.007	-0.0272***	(0.0093)	-0.741
<i>Percent White</i>	0.0152***	(0.0029)	0.001	0.0034	(0.0025)	0.193
<i>Percent Female</i>	0.1117***	(0.0081)	0.006	0.0303***	(0.0056)	1.438
<i>Percent Age 25-54</i>	0.0705***	(0.0097)	0.004	-0.0018	(0.0076)	-0.068
<i>Percent Age 55-74</i>	0.0534***	(0.0094)	0.002	0.0083	(0.0091)	0.211
<i>Percent Age 75+</i>	0.0285***	(0.0106)	0.001	-0.0437***	(0.0109)	-0.654
Constant	-0.1468	(0.0089)		-5.976***	(0.6183)	

*** Significant at 1%

** Significant at 5%

* Significant at 10%

References

- Andrade, C. and Faculdade de Ciencias Economicas e Empresariais. (2002) The economics of welfare participation and welfare stigma: a review, *Public Finance and Management*, **2**, 294-333.
- Bitler, M. P., Currie, J. and Scholz, J. K. (2003) WIC eligibility and participation, *Journal of Human Resources*, **38**, 1139-79.
- Blank, R. M. (1985) The impact of state economic differentials on household welfare and labor force behavior, *Journal of Public Economics*, **28**, 25-28.
- Blank, R. M. and Ruggles, P. (1996) When do women use aid to families with dependent children and food stamps? the dynamics of eligibility versus participation, *Journal of Human Resources*, **31**, 57-89.
- Blundell, R., Fry, V. and Walker, I. (1988) Modeling the take-up of means-tested benefits: the case of housing benefits in the United Kingdom, *The Economic Journal*, **98**, 58-74.
- Brown, J. and Jamison, M. A. (2006) Motivations behind low-income households bypass of support for universal service, Working Paper No. 06-07, University of Florida, Public Utility Research Center.
- Brown, J. (2006) Perspectives on communications services and Lifeline: results of a telephone survey of Florida households, Working Paper No. 06-11, University of Florida, Public Utility Research Center.
- Brown, J. (2006) Understanding participation in telecommunications Lifeline programs: a survey of low-income households in Florida, Working Paper No. 06-10, University of Florida, Public Utility Research Center.
- Brown, J. (2006) Disconnecting from communications: a survey of Floridians who qualify for Lifeline and dropped their telephone service, Working Paper No. 06-12, University of Florida, Public Utility Research Center.
- Burton, M. and Mayo, J. W. (2005) Understanding participation in social programs: why don't households pick up the Lifeline? Working Paper, Department of Economics, Georgetown University.
- Currie, J. (2004) The take up of social benefits, Working Paper 10488, National Bureau of Economic Research.
- Currie, J. and Yelowitz, A. S. (2000) Are public housing projects good for kids? *Journal of Public Economics*, **75**, 99-124.

- Federal Communications Commission (FCC) (2000, 2001, 2002, 2003, 2004) Universal Service Monitoring Report prepared for the Federal-State Joint Board on Universal Service. Monitoring Report CC Docket No. 98-202, 2003. Prepared by Federal and State Staff for the Federal-State Joint Board on Universal Service in CC Docket No. 96-45.
- Fraker, T. and Moffitt, R. (1988) The effect of food stamps on labour supply: a bivariate selection model, *Journal of Public Economics*, **35**, 25-56.
- Garbacz, C. and Thompson, H. G., Jr. (1997) Assessing the impact of FCC Lifeline and Link-up programs on telephone penetration, *Journal of Regulatory Economics*, **11**, 67-78.
- Garbacz, C. and Thompson, H. G., Jr. (2002) Estimation telephone demand with state decennial Census data from 1970 – 1990, *Journal of Regulatory Economics*, **21**, 317-29.
- Garbacz, C. and Thompson, H. G., Jr. (2003) Estimating telephone demand with state decennial Census data from 1970 -1990: update with 2000 data, *Journal of Regulatory Economics*, **24**, 373-78.
- Garbacz, C. and Thompson, H. G., Jr. (2005) Universal telecommunication service: A world perspective, *Information Economics and Policy*, **17**, 495-512.
- Gensler, H.J., and Walls, W. D. (1995) Labour-Supply Impacts of Effective Welfare Programme Parameters: A Reanalysis, *Applied Economics*, **27**, 461-467,
- Greene, W. H. (2003) *Econometric Analysis*, Fifth Edition, Prentice Hall, Upper Saddle Ridge, NJ.
- Hauge, J. A., Jamison, M. A. and Jewell, R. T. (2007) Discounting telephone service: an examination of participation in the Lifeline assistance program using panel data, Working Paper No. 06-06, University of Florida, Public Utility Research Center, under review.
- Hauge, J. A., Jamison, M. A. and Jewell, R. T. (2007) Participation in social programs by consumers and companies: a nationwide analysis of participation rates for telephone Lifeline programs, forthcoming in *Public Finance Review*.
- Holt, L. and Jamison, M. A. (2006) Making telephone service affordable for low-income households: an analysis of Lifeline and Link-up telephone programs in Florida, Working Paper No. 06-05, University of Florida, Public Utility Research Center.
- Hoynes, H. W. (1996) Welfare transfers in two-parent families: labor supply and welfare participation under AFDC-UP, *Econometrica*, **64**, 295-332.

Kang, G., Huffman, S., and Jensen, H. (2004) An Empirical Analysis of Joint Decisions on Labour Supply and Welfare Participation, *Applied Economics Letters*, **11**, 869-872.

Maddala, G.S. (1983) Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press, Cambridge, UK.

McGarry, K. (1995) Factors determining participation of the elderly in SSI, Working Paper 5250, National Bureau of Economic Research.

Moffitt, R. (1983) An economic model of welfare stigma, *American Economic Review*, **73**, 1023-35.

Pan, S, Jensen, H.H., Fuller, W.A., and Mohanty, S. (2006) The Effects of Local Labour Market Conditions on Welfare Program Participation, *Applied Economics*, **38**, 649-659.

Papke, L. E., and Wooldridge, J. M. (1996) Econometric methods for fractional response variables, *Journal of Applied Econometrics*, **11**, 619-32.

Progress and Freedom Foundation “Progress on Point” papers 1.3 (June 2005), 10.15 (Sept 2003), 11.18 (Oct 2004), and 12.1 (Jan 2005), Released at <http://www.pff.org>.

Stuber, J. and Kronebusch, K. (2004) Stigma and other determinants of participation in TANF and Medicaid, *Journal of Policy Analysis and Management*, **23**, 509-30.

United States Census Bureau (2005) State level demographic data, downloaded August 2005, <http://factfinder.census.gov>.

Williamson, A. R. (2006) A count of the number of households in Florida that qualify for Lifeline telephone discounts 2000 to 2005, Working Paper No. 06-13, University of Florida, Public Utility Research Center.

Yelowitz, A. S. (2000) Public policy and health insurance choices of the elderly: evidence from the Medicare buy-in program, *Journal of Public Economics*, **78**, 301-24.

Appendix²⁵

Accounting for biases when estimating equations in which the true data is a subset of the total dataset can be illustrated as follows.

$$\begin{aligned}\hat{\alpha}_1^{OLS} &= \alpha_1 + \frac{cov(x, v)}{var(x)} \\ &= \frac{\beta_1}{p} - \beta_1 \left(\frac{1-p}{p} \right) \frac{cov(x, x^n)}{var(x)}\end{aligned}$$

where:

$$\begin{aligned}cov(x, x^n) &= E(xx^n) - E(x)E(x^n) \\ &= E\left[\left\{px^e + (1-p)x^n\right\}x^n\right] - E\left[px^e + (1-p)x^n\right]E(x^n) \\ &= p\left\{E(x^e x^n) - E(x^e)E(x^n)\right\} + (1-p)\left\{E(x^n x^n) - E(x^n)E(x^n)\right\} \\ &= pcov(x^e, x^n) + (1-p)var(x^n)\end{aligned}$$

This implies:

$$\begin{aligned}\hat{\alpha}_1^{OLS} &= \frac{\beta_1}{p} \left[1 - (1-p) \frac{pcov(x^e, x^n) + (1-p)var(x^n)}{p^2 var(x^e) + (1-p)^2 var(x^n)} \right] \\ &= \frac{\beta_1}{p} \left[1 - \frac{p(1-p)cov(x^e, x^n) + (1-p)^2 var(x^n)}{p^2 var(x^e) + (1-p)^2 var(x^n)} \right] \\ &= \frac{\beta_1}{p} \left[1 - \frac{p^2 var(x^e) - p(1-p)cov(x^e, x^n)}{p^2 var(x^e) + (1-p)^2 var(x^n)} \right] \\ &= \beta_1 \frac{pvar(x^e) - (1-p)cov(x^e, x^n)}{var(x)}\end{aligned}$$

²⁵ This proof is attributed to Damon Clark, University of Florida Assistant Professor of Economics.

$$= \gamma \beta_1 \text{ where } \gamma = \frac{p \text{var}(x^e) - (1-p) \text{cov}(x^e, x^n)}{\text{var}(x)}$$

With respect to the above equation, note the following:

1. As $p \rightarrow 1$, the numerator approaches $\text{var}(x^e)$ and the denominator approaches $\text{var}(x^e)$. Therefore $\hat{\alpha}_1^{OLS} \rightarrow \beta_1$. This is what we would expect. As the proportion of eligible beneficiaries approaches 1, using the population characteristic will give a consistent estimate of the true effect.
2. For $p < 1$, since $\text{var}(x^e) < \text{var}(x)$, $\gamma < 1$. Hence in general, this approach will underestimate the effect.
3. For p small, since $\text{cov}(x^e, x^n)$ is likely positive, $\gamma < 0$. Hence, when eligibility rates are small, using the population characteristics may give the wrong sign.

This proof involves the simple case in which there is one variable in question and no omitted variables, and eligibility is assumed constant across counties. These assumptions are made for simplicity, however the results hold given multiple variables and varying rates of eligibility.²⁶ When eligibility rates are high, using population characteristics gives estimates that are close to consistent. However, as eligibility decreases, the population approach is likely to underestimate the effects and may even reverse the signs.

²⁶ Formal proof is available upon request.