DEVELOPING THE CHAPIN HALL CHILD CARE SUBSIDY ELIGIBILITY MODEL

Abstract:

Administrative records are readily usable for analyzing the characteristics of program participants. However, they are significantly less directly usable for determining who is eligible or qualified for a program. One possible approach for making this type of imputation is applying known eligibility rules to relevant household characteristic data. Such an approach implicitly assumes that the household characteristic data are reliable and that the relationship of household members to each other is known or unimportant. The model described in this paper does not accept these assumptions. Instead it models family relationships and qualifying income based on recorded values for these in program administrative record data.

Introduction:

The Child Care Subsidy program is a federal-level program that is tailored and administered separately by each state. The program is intended to assist low-income families in paying for childcare so to allow adult family members the ability to seek and retain gainful employment. To evaluate the utilization and the efficacy of the Child Care Subsidy (CCS) program it is useful to be able to distinguish those families that are eligible to participate in the program from those that are not. Being able to make this distinction allows the computation of take-up rates both in aggregate and according to various demographic factors and also allows the comparison of outcomes among eligible families between those participating in the program and those not participating in the program. The Chapin Hall Center for Children in conjunction with the U.S. Bureau Of The Census, University of Chicago, Columbia University National Center for Children in Poverty, University of Texas at Austin Ray Marshall Center for the Study of Human Resources, University of Baltimore Jacob France Institute sponsored a project to establish a database that allowed for the evaluation of the use and performance of the Child Care Subsidy program in several states: Maryland, Illinois, and Texas. Critical to evaluation was the determination of Child Care Subsidy Eligibility for families included in the study database. For didactic simplicity, this paper will focus on the establishment of an eligibility model for Texas residents. Modeling for Maryland and Illinois largely follow the methods developed originally for Texas. Also, the modeling effort describe herein focused on CCS *income* eligibility. Additional work beyond this discussion added qualification by work- or work training-status to the modeling regime.

To operationalize the analysis of the Child Care Subsidy program, it was decided that this analysis would be conducted on the data collected from the 2001 American Community Survey and Supplementary Survey. These surveys use identical data collection instruments and differ only in the sampling methods used to select participants. Joint weights have been established such that the ACS/SS01 response data can be used to produce estimates for the U.S. resident population. Note that ACS/SS 2001 is ideal for the purposes of analysis due to the extensiveness of data elements, the availability of the data for Census researchers, and the size of the sample. Persons captured by the 2001 ACS/SS 2001 form the universe for the application of this model, and population projections can be made from it by applying existing ACS/SS 2001 weights to the results of the model.

One of the key decisions affecting the development of the eligibility modeling was to have certain of the determinations to be rendered probabilistically rather than deterministically. To clarify this point, consider that for a given family we could have made a yes or no determination regarding Child Care Subsidy eligibility. Rather than do this, the model establishes a probability of eligibility for each modeled family. Also, part of the modeling process consists of the identification of family relationships. This too is characterized probabilistically rather than deterministically. So in cases where more than one adult in a household may potentially be responsible for one or more of the children in a household, a probability characterizes the existence of such relationships. This probability replaces a definitive assignment of responsibility for each child to a particular adult household member.

The reason we chose to characterize eligibility and family relationships probabilistically rather that deterministically is our belief that these probabilistic characterization would provide more effective analysis given that

- They conform more to our state of knowledge that a deterministic characterization
- They allow the level of confidence of the determination to be accounted for in the generated analysis, presumably leading to more accurate results
- They do not result in determinations that later can be shown definitively to be untrue.

Our modeling procedure can be understood to comprise three tasks:

- Development of a family relationship model
- Development of an income eligibility model
- Application of models to database to establish eligibility likelihood.

Family Relationship Module:

To understand the approach that we used, it is useful to review how Texas Department of Family and Protective Services consider families in establishing eligibility and enrollment under the Child Care Subsidy program. First, a childcare subsidy can only be provided to a family that needs this service for a child age twelve or under. Any family not having such a person will be considered not eligible for the subsidy. Based on our review of Texas child care subsidy data, we see that for each child enrolled in the subsidy program, there is a single adult who is recorded as being responsible for the child. Presumably this person is someone with legally established responsibility for the child. It is unclear who is selected as this person with primary responsibility in cases where there is more than one adult with legal responsibility for a child, such as a mother and father in a traditional family structure. Nevertheless, our modeling seeks to replicate the person indicated on administrative records as responsible.

The built model is a logistic regression estimating a probability that a household member is responsible for a specific 12-year-old or younger household member based on three main effects:

Actual Responsible Adult Status (0 or 1)

Modeled from

- Sex of Other Household Member
- Age Difference Between Other Household Member and Qualifying Child
- Joint Relationship to Reference Person

To estimate our model coefficients, we use data for enrolled CCS families who are also present in ACS/SS 2001. For each child enrolled in the Texas Child Care Subsidy program in calendar year 2001, we look to see if that same child was a reported on person to ACS/SS during one of the months in which he or she was enrolled. If so, then that child, and the other ACS/SS identified family members for that child form a unit of analysis in developing the family relationship model. For each such child, we look to see which other household members may potentially be the person responsible for this child (we are looking to see if this household member could have been considered, potentially, to be the responsible person, in the absence of the knowledge of who actually is). We then characterize this household member's relationship to the child. The characterization is made by the age difference between the other member and the child, the sex of the other member, and the joint relationship to the household reference person. The age difference is categorized as follows:

- 17 Years Difference or Less
- 18 – 24 Years Difference
- 25 – 34 Years Difference
- 35 – 44 Years Difference
- 45 Years Difference or More

The household reference person is the person designated by the respondent as being the primary owner or renter of the household residence. The joint relationship is characterized by one of these categories:

Joint Relationship to Reference Person			
	Potentially Responsible Adult's Relationship to Reference Person		**Qualifying Child's Relationship to Reference Person**
1.	Same or Spouse	←→	Child
2.	Same	←→	Grandchild
3.	Child	←→	Grandchild
4.	*All Other Joint Relationships*		

The joint relationship to the household reference person is used rather than the direct relationship between the other member and the child because this direct relationship status is not collected in the ACS/SS survey interview; rather it must be inferred by the joint relationships to the reference person.

Next, for each household with an enrolled child, we use Texas Child Care Subsidy data to determine which household member was actually the responsible person for the enrolled child. If this person cannot be identified on the ACS/SS roster for this household, then this household is removed from the analysis set. Otherwise, we now can organize these data such that for each enrolled child, there are one or more records each for a potentially responsible adult, depending on the number of children 12 years old or younger in the household. On each of these records there is a Boolean variable indicating whether this adult is or is not the actually responsible person for the enrolled child as determined by reference to Texas CCS administrative records. With the data in this format, we are able to run a standard logistic regression. For this regression, we used no interaction terms.

The model just described can be considered the first stage sub-model of the family relationship model. The second stage seeks to summarize to a single value the potentially multiple probabilities for potentially responsible adults from the first stage sub-model: one for each child equal or under 12-years old. The summarized probability generated from the second stage model represents the probability that the person is the responsible party for *at least* one child in the household. The second-stage sub-model is built on the results (that is, the predicted probabilities of responsibility for each child) of the first-stage sub-model. For the second-stage model, the dependent variable is whether a household member is recorded as the responsible party for at least one child enrolled in the Texas Child Care Subsidy program. All of the independent variables are built on the the complement, $P'_{nr} = 1 - P_{nr}$, of the calculated probability that a household member is not responsible for any child enrolled in the subsidy program, which is computed as $P_{nr} = 1 - \prod_i (1 - p(r_i))$,

Where:

P_{nr}	~	Probability of not being responsible for any enrolled child in household
$p(r_i)$	~	Calculated (from first-stage sub-model) probability of being responsible for the i^{th} enrolled child in the household.

This value, P'_{nr} and its square and cube comprise the independent variables for this logistic regression model, which includes no interaction terms:

$$P(\text{responsible for at least one child}) = 1 \Big/ \left(1 + \exp(-\beta_0 - \beta_1 \cdot P'_{nr} - \beta_2 \cdot {P'_{nr}}^2 - \beta_3 \cdot {P'_{nr}}^3)\right)$$

Based on the first-stage and second-stage model, it is possible to assign a probability of being the responsible party for a child (twelve years-old or younger) should that family apply and be accepted for the subsidy. However, the results of this modeling, will, in general, not yield a total family probability of responsibility of 1, as is eminently desirable. To overcome this deficiency, we use a normalization procedure that forces the sum of potential responsible persons to equal 1, by dividing each calculated probability from the second-stage sub-model by the sum of these. Then, for developing imputations, we assume that only one person is responsible for all qualifying children in the family. While in some instances this assumption may not be correct, we think that due to the overall structure of the eligibility modeling this does not cause any significant biases in derived estimates of family eligibility.

Income Qualification Module:

Income qualification forms the other probabilistic component of the eligibility model. This modeling compares modeled income to a deterministically referenced income-eligibility threshold. These thresholds conform to those actually applicable in the various workforce development districts in Texas in-place in 2001. Because we have an error model for modeled income, we are able to calculate a probability that income falls below the determined threshold.

To model income, we were able to use the administratively recorded income for enrolled families at the time of the survey. Based on our understanding of program administration, it seems that this recorded income will be most accurate near the time of original recordation. For this reason, when there is a series of recorded incomes for an enrolled family, we include this family in the model only if the first occurrence of any value of recorded income appeared in the month the same family was interviewed for ACS/SS 2001. To model this income, we made use of the ACS/SS reported categorical income: wage, self-employment, disability, etc. In addition to these data elements, for many of the households used to build the model, we were able to locate the quarterly wage income as reported to the Texas unemployment insurance program, as is required for most employers. These quarterly wage values seemed likely to be in aggregate more accurate that self-reported wage income, a supposition that appeared to bear-out in the modeling process.

Note that our decision to use a model to infer income, and to model to subsidy-program-recorded income rather than make a one-to-one determination of this from ACS/SS and Unemployment Insurance reported values derives from our unwillingness *a-priori* to accept these reported values as unbiased and insubstantially inaccurate. Statistics provided later in this paper seem to confirm that this unwillingness is not unreasonable. Because of the sensitivity of income data and database elements holding these data as well as the complexity of the modeling, we will not present a fully detailed description of the structure of the income model and will instead adumbrate its most significant features.

First, the model was configured differently for married responsible family members that for those non-married. For this purpose, we determined marital status based on the recordation of this status for CCS. For married persons, we included the income of the spouse, as best could be determined by us, as explanatory information in the model, yet included it distinctly from income of the primary. Also, the spouse was included in the count of household members. Otherwise only the responsible person and children of the responsible person, 22 year-old and under were counted as family members. The number of family members is relevant because it determines the applicable family income threshold. In fact we used a direct lookup to then existing income threshold tables, which depend wholly on family size for a given location of residence within Texas. Texas is unusual among states by having thresholds that vary by location of residence, all at a given moment.

Second, different models were used depending on the availability of UI quarterly wage data. So the model used depended on both marital status and UI wage income data availability. Where UI wage data was available, we used data from multiple periods, those nearest to the time of interview as this allowed us to model subsidy-program reported income better than using just a single value.

Third, all versions of the models used had a positive intercept and aggregate income coefficients (that is the sum of all the income coefficients in a given model) of less that one. We have not developed a theory of why this is so. Also, predicted income tends to be less than that computed additively from reported (on ACS/SS and UI) income. This may reflect applicants trying to game the system or may instead be an artifact of our modeling procedure. However, comparing the direct additive computation to appropriate threshold values for participants shows that a substantial number of these are enrolled even though that additive value exceeds the threshold. This provides some credence to our modeling strategy.

Fourth, income modeling did in no way depend on demographic factors other than family relationship status (marital partnership and responsibility for children) such as sex, race, and ethnicity. However, in evaluating model performance and generated take-up rates, we did take these factors into account.

Originally, we attempted to model the error of the income model as Normal with the variance estimated by the regression mean square error statistic. However, further analysis showed that in fact error was not normally distributed and instead had more spread than is found there. By performing an analysis that compared program-reported income to modeled income, we identified the best fitting distribution as a Student-T with six degrees of freedom and a standard deviation equal to the applicable regression root mean square error. Using the point estimate of program-reported income in conjunction with this error model allowed us to determine the probability that predicted income fell below the threshold.

Application of Model:

To determine program eligibility, we used a two-stage process that was build correspondingly on the family relationship and income eligibility modeling. First, we selected all ACS/SS 2001 families that had at least one child 12-year-old or younger, and for these children, we used the family relationship model to assign probabilities for each potentially responsible other household member. Second for each potential household member, we computed a probability of income eligibility. For each household then, we computed the probability of being income eligible for CCS as $P_{ie}^{TH} = \sum_j P_{R,j} \cdot P_{ie,j}$,

Where:

P_{ie}^{TH}	~	Total household probability of being income eligible for CCS.
$P_{R,j}$	~	Probability of j^{th} household member being responsible for children in household.
$P_{ie,j}$	~	Probability of j^{th} household member (and their fellow program-application case members together) being income eligible for CCS.

Note, that there are requirements other than income-eligibility to qualify for CCS. Typically, these relate to the requirement for the responsible person (and their spouse, if married) to be either currently employed or in an approved job-training program. Also, these activities must occur at a certain minimum level specified by hours engaged per week. In addition, there are some families that are categorically eligible for CCS because they are enrolled in the Temporary Aid for Needy Families (TANF) program.

For this paper, we terminate our discussion of CCS eligibility modeling with income eligibility. However, the use of this model requires additional computation or modeling of other eligibility requirements, specifically work activity. Researchers using the modeling described in this paper have taken various approaches to modeling work-qualification (participation in approved work or work-training activities). Work qualification can be most readily ascertained by the existence of substantial UI wages for the quarter during which the ACS/SS interview took place. In addition, ACS/SS-reported employment status and hours-per-week at work may be used.

Results:

We conclude this paper by presenting a computation of 2001 subsidy-eligible families and weighted counts of recipients. From this, putative take-up rates can be computed. For this purpose we considered only work-qualification, not training-based qualification, and excluded families categorically eligible due to TANF enrollment. For modeling purposes, we established work-qualification by the ACS/SS field ESR (Employment Status Recode) indicating anything but ‘Unemployed’ or ‘Not in the Labor Force’ for the responsible person and their spouse, if married.

Estimates of the 2001 Mean Monthly Texas CCS Eligible Families
Hhlds. with Ratio of ACS Reported Income: Income Eligibility Threshold <= 1
Ranked into Quintiles By That Ratio
Weighted by ACS/SS 2001 Final Weight

Rank	Mean Ratio of Modeled Income to CCS Income Threshold	Mean Computed Probability of Income Eligibility	Mean Ratio of ACS/SS Reported Income to CCS Income Threshold	Households in Category Eligible Based on ACS/SS Reported Income	Income-Eligibility Model Derived Estimate of Eligible Households	Count of CCS Enrolled Households from Admin. Records
1	0.37	0.97	0.11	110,725	107,723	4,279
2	0.47	0.97	0.34	97,980	94,733	6,239
3	0.57	0.94	0.54	98,034	92,032	5,190
4	0.68	0.88	0.73	93,979	83,025	5,009
5	0.79	0.78	0.91	83,227	65,218	1,092
All	*0.56*	*0.91*	*0.50*	*483,946*	*442,731*	*21,810*

Estimates of the 2001 Mean Monthly Texas CCS Eligible Families
Hhlds. with Ratio of ACS Reported Income: Income Eligibility Threshold > 1
Ranked into Quintiles By That Ratio
Weighted by ACS/SS 2001 Final Weight

Rank	Mean Ratio of Modeled Income to CCS Income Threshold	Mean Computed Probability of Income Eligibility	Mean Ratio of ACS/SS Reported Income to CCS Income Threshold	Households in Category Eligible Based on ACS/SS Reported Income	Income-Eligibility Model Derived Estimate of Eligible Households	Count of CCS Enrolled Households from Admin. Records
1	0.93	0.60	1.24	0	130,224	4,116
2	1.18	0.28	1.75	0	53,775	852
3	1.49	0.10	2.40	0	18,375	78
4	1.91	0.05	3.33	0	7,545	142
5	3.40	0.04	7.35	0	5,937	0
All	*1.64*	*0.25*	*2.87*	*0*	*215,856*	*5,189*

One noteworthy element of these tables is that nearly one-fifth (comparing last-column “All” row values for the two tables) of eligible families have an ACS/SS calculated income greater than the identified threshold value. Derived take-up rates appear somewhat more uniform using the income-eligible model than the straight ACS/SS computation. Also, for households and families with low income, the probability of being income eligible is very high, near 1, as should be expected.

Conclusion:

The modeling approach taken here takes a hybrid approach of using statistical modeling and probabilistic characterization for some elements of eligibility, family relationship and income, but does not do so for others: age, sex, marital status, number of family members. Of these, we would expect all but marital status to be fairly well reported, and not likely to greatly bias estimates. For marital status, our survey-CCS administrative record matching shows significant levels of disagreement. Were we to attempt to improve this modeling effort, we would consider whether a marital status sub-model would be useful. For the elements that were probabilistically determined, we applied them to computing eligibility by multiplication, which assumes their independence. This does not seem unreasonable, because bias in family relationship determination can be supposed unrelated to bias in income determination. Nevertheless, if there is a covariance between these, they would act to bias our estimates.

While we believe our approach was reasonable given the nature of the data available to us, it is not especially easy to validate this belief based on available data. Tables such as those presented above give us some level of comfort with the results. Additional work is being done by project researchers to evidence the efficacy of our modeling approach. Of these, regressions relating take-up to various social and economic factors seem to produce reasonable associations. We welcome any suggestions that would further this validation or that would allow us to improve this modeling approach in additional iterations.