January 2007

Estimating the Effect of Policy Changes on Participation in Medicaid While Allowing for Heterogeneous Treatment Effects Based on Observables and Unobservables*

John C. Ham
University of Southern California, Federal Reserve Bank of San Francisco and IZA

I. Serkan Ozbeklik
The Ohio State University and University of Southern California

Lara Shore- Sheppard
Williams College and NBER

Correspondence Author Contact Information:

John C. Ham
Department of Economics
University of Southern California
3620 S. Vermont Ave. KAP 300
Los Angeles, CA 90089
Tel #: (818) 439-3531
Email: johnham@usc.edu

*We are grateful to Steven Cosslett, Curtis Eberwein, Geert Ridder, Frank Vella and seminar participants at Ohio State and Georgetown for comments. Some of this research was conducted while Ham was a visitor at the Federal Reserve Bank of New York where he was provided with a hospitable and productive environment. This research has been supported by grants from the National Science Foundation (SES-0136928 and SES-0627968) and the National Institute of Child Health and Human Development (R01 HD39369-01A1). Any opinions, findings, and conclusions or recommendations in thispaper are those of the authors and do not necessarily reflect the views of the National Science Foundation, National Institute of Child Health and Human Development , the Federal Reserve Bank of New York, the Federal Reserve Bank of San Francisco or the Federal Reserve System. We are responsible for any errors.

Estimating the Effect of Policy Changes on Participation in Medicaid While Allowing for Heterogeneous Treatment Effects Based on Observables and Unobservables

Abstract

Previous research on participation in Medicaid has estimated a constant treatment effect model where becoming eligible affects the take-up of each child in an identical fashion. This approach has two drawbacks if effects do differ across families. First, it provides no information about which groups have low take-up conditional on eligibility. Second, it may lead to misleading predictions of the effects of expanding eligibility if the newly eligible have different characteristics, and thus different take-up rates, from individuals who were previously eligible. We explore two approaches to allowing the effect of eligibility on participation in Medicaid to differ across demographic groups. The first incorporates interactions between eligibility and demographic variables in the now standard linear probability model approach, which does not require a distributional assumption. The second is based on a switching probit model and has the advantage of allowing newly eligible individuals to differ from those previously eligible in terms of both observed and unobserved characteristics. Using data from the Survey of Income and Program Participation and exploiting the exogenous policy changes in Medicaid eligibility that took place in the late 1980's and early 1990's, we estimate the average take-up rate of many demographic groups using these two models. We then measure how different demographic groups would respond to a hypothetical policy experiment of increasing eligibility in 1995 by raising the 1995 income limits by 10%. We find groups vary dramatically in terms of their average take-up rates and their response to the policy change; moreover we are able to quantify these differences. While our emphasis is on Medicaid participation, our approach is applicable to the crowding out of private insurance by public insurance, as well as to participation in other social programs.

Keywords: Medicaid, Take-up, Heterogeneous Treatment Effect, Switching Probit Model, Linear Probability Model.

1. Introduction

Participation in means-tested social programs such as Medicaid represents a significant research area in public economics, health economics and labor economics. An increasingly common way of modeling such participation is to estimate a linear probability model of participation where a dummy variable for eligibility for the program is an endogenous explanatory variable.[1] An implication of this model is that the estimated effect of becoming eligible on the probability of participation is restricted to be constant across individuals. However, there are several reasons to believe that participation conditional on eligibility is not constant. If this is the case, using this approach can lead to misleading predictions of the effects of expanding eligibility if newly eligible individuals differ from the previously eligible in ways that affect take-up. In addition, this approach does not allow researchers to determine which groups have low rates of participation given eligibility, an important part of accurately targeting policies such as outreach programs.

In this paper we explore two approaches to allowing the effect of eligibility on program participation to differ across groups. The first incorporates interactions between eligibility and demographic variables in the now standard linear probability model approach. This method accounts for observable differences between the newly eligible and the previously eligible, is easy to use and interpret, and does not require a distributional assumption. The second method is based on a switching probit model. It allows newly eligible individuals to differ from those previously eligible in terms of unobserved as well as observed characteristics while addressing some technical problems that arise in the

[1] Papers using such an approach include Cutler and Gruber (1996), Currie and Gruber (1996a, 1996b), LoSasso and Buchmueller (2004), Hudson, Selden, and Banthin (2005), Ham and Shore-Sheppard (2005), and Shore-Sheppard (2005).

simultaneous equation linear probability model, albeit at the cost of making distributional assumptions. We show for both models how one can use the parameter estimates to calculate separate take-up rates across different demographic groups and the take-up rate among those made newly eligible by a policy change. While our emphasis is on Medicaid take-up, our approach is applicable more broadly. For example, it can be used to examine participation in the State Children's Health Insurance Program (SCHIP), the crowding out of private insurance by public insurance, and take-up of other social programs such as Food Stamps and Section 8 housing.

The late 1980's and early 1990's were an era of significant changes in the legislative environment towards Medicaid. For the first time, eligibility for Medicaid among low-income children and pregnant women did not require that the individual also be eligible for Aid to Families with Dependent Children (AFDC). Family structure criteria for eligibility were loosened, and income limits for Medicaid eligibility were set at levels far above AFDC levels. With these changes in eligibility criteria, about 30 percent of children aged 0 to 18 years had become eligible for public health insurance by 1996, and only about half of these children were from welfare-eligible families (Selden, Banthin, and Cohen 1998). These expansions of Medicaid eligibility for children outside the traditional welfare population raise the question of how newly eligible families respond to the policy change in terms of Medicaid participation, and how their coverage by private insurance is affected (crowding out).

In the last decade a number of studies have examined these issues. All of those studies used the form or timing of the expansions, and the fact that some groups were affected by the policy changes while others were not, to disentangle the effects of Medicaid expansions on the health insurance coverage of individuals. Although the identification strategies of the previous papers are similar in this very fundamental aspect, the methods used and the estimates obtained differ considerably. Some of the most important studies in this literature treat Medicaid eligibility as endogenous in an instrumental variable, linear probability model, where the instrument varies only due to the legislative amendments over the sample period. Although this method is simple and the coefficients are easy to interpret, it ignores the fact that different groups are likely to respond differently to a change in eligibility status. In particular, the reaction may differ across demographic groups, since over time the policy changes are making higher income families eligible.

Using data from the Survey of Income and Program Participation (SIPP) and exploiting the exogenous policy changes that took place in late 1980's and early 1990's, we estimate average take-up rates and the take-up response of different demographic groups according to race, family structure, education and number of earners in the family using the two different methods described above. Our results suggest that groups differ substantially in their take-up rates under both the existing program and a counterfactual expansion of the program. While the fact that there are differences may not seem surprising, we note that we can *quantify* the differences in take-up across groups. The magnitude of such differences is not well known and has heretofore not been estimated in the literature. Further, both of our suggested approaches have the advantage that they can predict different take-up rates for different expansions, since different expansions will affect different parts of the population. The standard linear probability model approach is constrained to predict the same take-up rates for different expansions.

The plan of the paper is as follows. In the next section we review the Medicaid program and the previous studies that have used the two-equation linear probability model approach to estimate the take-up response to the Medicaid expansions. In Section 3, we outline this standard approach to public health insurance take-up and explain how one can extend it using: i) a linear probability model with interactions between eligibility and demographic variables and ii) a switching probit model. We then show how these extensions can be used to estimate average take-up rates for different groups. Next, we show how to use the parameter estimates to predict the response to any expansion in the coverage of the program. After discussing our data briefly in Section 4, we present our results in Section 5. Section 6 concludes the paper.

2. Medicaid Expansions and Previous Literature

Medicaid was first established as a public health insurance program for welfare recipients and low-income aged and disabled individuals. This focus largely remained until the late 1980s, when expansions in eligibility first permitted, and then required, states to cover pregnant women and children with family incomes that made them ineligible for cash welfare. Following the federally-mandated eligibility expansions of 1989 and 1990, states were required to cover children age 6 or younger with family income up to 133 percent of the poverty line and children born after September 30, 1983 with family income up to 100 percent of the poverty line. States were also given the option to increase their eligibility thresholds up to 185 percent of the poverty line. As these eligibility limits were far more generous than the eligibility limits applying to AFDC, the link between Medicaid eligibility and AFDC eligibility greatly diminished for young, low-income children. By 1996, of the approximately 30 percent of children age 19 and younger who were eligible for Medicaid, only about half came from typically welfare-enrolled families (Selden, Banthin, and Cohen 1998). While families who enrolled in cash welfare programs were also automatically enrolled in Medicaid, newly eligible children were not. Consequently the establishment of new Medicaid eligibility raised an important policy question: to what extent did expanded eligibility lead to increased health insurance coverage for the targeted population of children?

There has been a substantial amount of research on this question, and a non-exhaustive list includes Currie and Gruber (1996a, 1996b), Cutler and Gruber (1996), Dubay and Kenney (1996), Thorpe and Florence (1998), Yazici and Kaestner (1999), Shore-Sheppard (2000), Blumberg, Dubay, and Norton (2000), Ham and Shore-Sheppard (2005), and Shore-Sheppard (2005)). Further, there is also research on the related question of how the further public health insurance expansions of the State Children's Health Insurance Program (SCHIP) affected coverage (LoSasso and Buchmueller (2004), Hudson, Selden, and Banthin (2005)). Since our aim in this paper is to suggest an econometrically preferable alternative to the approach used by a particular group of studies, rather than summarizing the literature, we focus on two of the studies that use this now standard approach.

An important study using this approach was the seminal paper of Cutler and Gruber (1996). Cutler and Gruber used the method outlined below in Section 3.1 and data on children from the March Current Population Survey (CPS) from 1988 to 1993 to estimate the effect of imputed Medicaid eligibility on insurance status, controlling for demographics and state and year effects. They used an instrumental variables approach since eligibility is likely to be endogenous. This potential endogeneity arises for several reasons. First, unobservable factors affecting eligibility are likely to be correlated with unobservable individual and family characteristics that determine take-up. Second, eligibility may proxy family income if income, which is also likely to be endogenous, is not included as an independent variable. Finally, parental wages, which in turn determine eligibility, are likely to be correlated with fringe benefits (including private health insurance) of the parents. These benefits are unobserved and must be treated as part of the error term, and will be correlated with eligibility, thus necessitating treating eligibility as endogenous.

As a solution to the problem of the endogeneity of the eligibility variable, Cutler and Gruber (1996) suggested an instrument, $FRACELIG_i$, that is the fraction of a random sample of 300 children of each age imputed to be eligible according to the rules in each state in each year. This instrument, which is essentially an index of the expansiveness of Medicaid eligibility for each age group in each state and year, is correlated with individual eligibility for Medicaid but not otherwise correlated with the demand for insurance.[2] They estimated that the average take-up rate for those affected by the expansions was 23.5 percent.

Ham and Shore-Sheppard (2005) used data from Survey of Income and Program Participation (SIPP) covering the period from October 1985 to August 1995 to replicate Cutler and Gruber's analysis and found a smaller average take-up rate of 11.8 percent for those affected by the expansions. They attributed some of the differences in their results to different samples and recall periods in the data sets used. Ham and Shore-Sheppard slightly modified the Cutler-Gruber instrument $FRACELIG_i$ in their study by using all

[2] Of course, one has to assume that changes in a state's Medicaid provisions are not correlated with changes in the state's availability of private insurance (which are unobservable to the researcher).

sample observations of children of a given age in a SIPP wave except for those from the state for which the instrument is being calculated. Since this instrument is created using a larger sample, it is theoretically superior to the version using a random sample; however in practice it made no difference to the results. We use the data and instrument of Ham and Shore-Sheppard in our empirical application of the approaches described in the next section.

3. Econometric Methodology

3.1. Standard Approach to Estimation

The standard evaluation of eligibility changes in the public health insurance literature involves estimating the following econometric model using a linear probability model. The index function for participation is given by

$$Part_i^* = X_i\beta + \gamma elig_i + u_i, \qquad (1)$$

where X_i is a vector of demographic variables for child *i*, $elig_i$ is a dummy variable coded one if the child is eligible for a program and zero otherwise, and u_i is an error term. A child participates ($part_i = 1$) if $Part_i^* > 0$. The index function for eligibility is given by

$$Elig_i^* = Z_i\delta + e_i, \qquad (2)$$

where $Z_i = (X_i, z_i)$ and z_i is a variable that affects eligibility but not participation conditional on eligibility. An example of such a variable is the Cutler-Gruber instrument $FRACELIG_i$ described above. For our purposes, the crucial issue is that in this model the change in the probability of participating from becoming eligible is a constant equal to γ. It is unlikely that this change in probability is actually constant for all individuals. For example, it is clear from the discussion in Cutler and Gruber (1996) that they have in mind a random coefficients model where the index function for participation is given by

$$
\begin{aligned}
Part_i^* &= X_i\beta + \gamma_i elig_i + u_i \\
&= X_i\beta + \gamma elig_i + (\gamma_i - \gamma) elig_i + u_i \,. \qquad (1)' \\
&= X_i\beta + \gamma elig_i + v_i.
\end{aligned}
$$

As Cutler and Gruber note, the IV estimate of γ will be the average of the γ_i over the variation in their instrument z_i. While this average parameter provides a useful summary of a marginal policy change, it will be less useful for larger, nonmarginal, changes in eligibility since one would need an average of γ_i over those affected by the change.

3.2 A Simple but Powerful Alternative: Linear Probability Model with Interactions (LPMI)

As an alternative, we suggest estimating the model

$$Part_i^* = X_i\beta + (X_i elig_i)\theta + u_i \,. \qquad (3)$$

In (3) the natural instruments are X_i and interactions between z_i and X_i. We can use this model to calculate the average take-up rates (i.e. the probability of participation for $elig_i = 1$ minus the probability of participation for $elig_i = 0$) among different demographic groups by simply taking the mean of the following expression for eligible individuals in group *j*

$$\widehat{ATR}_j^{lpmi} = \frac{\sum_{i \in j} X_i \hat{\theta}}{N_j}, \tag{4}$$

where N_j is the number of eligible individuals in group j and $\hat{\theta}$ is the two stage least squares estimate of θ. Since the measure in (4) is a linear sum of regression coefficients, one can calculate the standard error of the measure in a straightforward manner using a standard statistical package, such as Stata.[3] (Details on this calculation are given in the Appendix.)

In addition to calculating the average take-up response by different groups to the existing expansion, it is possible to estimate the take-up response to further policy expansions that make a new group of children eligible. First, one would determine which children become newly eligible under the expansion. Then a natural means of measuring the take-up rate among the newly eligible is

[3] Our Stata programs to implement both approaches suggested in the paper will be available beginning February 1, 2007 at http://lanfiles.williams.edu/~lshore/ and http://www-rcf.usc.edu/~johnham/.

$$\hat{T}_{new}^{lpmi} = \frac{\sum_{i \in new} X_i \hat{\theta}}{N_{new}}, \quad (5)$$

where the sum is over the newly eligible children and N_{new} is the number of newly eligible children. A standard error for the measure in (5) is obtained analogously to the standard error for the measure in (4). This measure takes into account the fact that the newly eligible have different demographic variables than the previously eligible children, but not the fact that they will have nonrandom values of u_i. We can also calculate this average take-up rate for subsets of the newly eligible, e.g. those with zero, one, or more than one earner per family.

Before concluding this section, we note four important problems with the standard approach (and our extension of it). First, it allows for a non-zero probability of participation even if the child is ineligible. Second, while it can account for the fact that in the policy experiment the newly eligible will differ from the previously eligible in terms of observables, it cannot address the fact that the newly eligible will also differ from the previously eligible in terms of unobservables. Third, the estimated take-up rate for an individual or group may be less than zero or greater than one, and we find that both problems occur in our estimates below. Fourth, there is a technical problem with the bivariate probit version of the standard model (1) and (2), and this also occurs in our extended model based on (3). Assuming normality, the log likelihood for the model based on (1) and (2) is

$$L = \sum_{i=1}^{N} \log \Phi_2 (q_{1i}(X_i\beta + \gamma elig_i), q_{2i}Z_i\delta, q_{i1}q_{i2}\tilde{\rho}) \,, \qquad (6)$$

where $\Phi_2(\bullet,\bullet,\tilde{\rho})$ is the cumulative bivariate normal distribution function, $q_{1i} = 2part_i - 1$ and $q_{2i} = 2elig_i - 1$. If there are no misclassification errors, i.e. there is no one classified as ineligible who is seen to participate, this likelihood is maximized by letting the intercept in the participation equation go to a large negative number approaching minus infinity and letting the coefficient on eligibility go to minus this large negative number plus a constant. In other words, the likelihood function for the simultaneous equation probit model is unbounded if there is no misclassification, although one can identify all the parameters besides the intercept. Since this issue would also occur in our extended model (3), some researchers may find this to be a troubling aspect of the standard model. Others may feel that since this last problem arises only in the probit version of the model, and not in the linear probability model itself, it can be ignored.[4] However, even in that case the first three problems remain. We now turn to an approach that addresses all four of these problems, albeit at the cost of making a distributional assumption.

3.3 An Alternative Approach to Estimation: Switching Probit Model (SPM)

To avoid the econometric problems we use a switching model (Quandt 1958, 1960, 1972), which has been applied in the bivariate probit with selection case by van de Ven and

[4] However, note that linear probability estimates of (1) or (1)' can achieve a perfect fit (in the absence of classification error) for the non-eligible by setting $\beta = 0$.

van Praag (1981).[5] The index function for eligibility is still (2), and $elig_i = 1$ if $Elig_i^* > 0$. However, we specify that the probability of participation given non-eligibility is zero. Further, for ***a randomly chosen*** individual, the index function for participation given eligibility is given by

$$Part_i^* = X_i \mu + \varepsilon_i , \qquad (7)$$

where $(\varepsilon_i, e_i) \sim iidN(0, \tilde{V})$ and $\tilde{V} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.[6] The appropriate log likelihood is

$$\begin{aligned} L = &\sum\nolimits_{Elig=1, Part=1} \log \Phi_2 (X_i \mu, Z_i \delta, \rho) + \sum\nolimits_{Elig=1, Part=0} \log \Phi_2 (-X_i \mu, Z_i \delta, -\rho) \\ &+ \sum\nolimits_{Elig=0} \log \Phi_1 (-Z_i \delta). \end{aligned} \qquad (8)$$

As in the case of the linear probability model with interactions, we can look at the predicted take-up rates among the eligible individuals in group j by calculating

$$\widehat{ATR}_j^{spm} = [\sum_{i \in j} 1 - \Phi_1 (-X_i \hat{\mu})] / N_j , \qquad (9)$$

[5] This model is also referred to as a bivariate probit model with sample selection in the econometrics literature.
[6] Following the econometrics literature, it is natural to write the index function for take-up for a randomly chosen (in terms of ε_i) child. However, it is worth emphasizing that unless $\rho = 0$, those eligible will not be a randomly chosen subgroup of the population. Below we take this into account in calculating take-up rates.

where again N_j is the number of eligible individuals in group j and $\hat{\mu}$ is the maximum likelihood estimate of μ. The delta method can be used to calculate the standard errors for these predicted take-up rates, again using a program like Stata. (Details about this calculation are given in the Appendix.) To compute the participation responses of different demographic groups, the process can simply be repeated for each group of interest.

The above calculation does not take into account the fact that the eligible children will not be a random sample in terms of the error ε_i in the take-up index function (7). To account for this, we use

$$\widetilde{ATR}_j^{spm} = \frac{1}{N_j}\sum_{i\in j}\Pr(P_i^* \geq 0 \mid E_i^* \geq 0) = \frac{1}{N_j}\sum_{i\in j}\frac{\Pr(P_i^* \geq 0, E_i^* \geq 0)}{\Pr(E_i^* \geq 0)} = \\ = \frac{1}{N_j}\sum_{i\in j}\frac{\Phi_2(X_i\hat{\mu}, Z_i\hat{\delta}, \hat{\rho})}{\Phi_1(Z_i\hat{\delta})}, \qquad (10)$$

where $\hat{\delta}$ and $\hat{\rho}$ are the maximum likelihood estimators of δ and ρ respectively.

To calculate the take-up rate among the newly eligible when we do not account for the fact that they will not be a random sample (in terms of ε_i) of the population, we use

$$\hat{T}_{new}^{spm} = [\sum_{i\in New} 1 - \Phi_1(-X_i\hat{\mu})] / N_{new}, \qquad (11)$$

where again N_{new} is the number of newly eligible individuals. However, one can improve the prediction of take-up among the newly eligible by also taking into account the fact that they will not be a random sample in terms of the unobservable ε_i. To do this, we calculate

$$\tilde{T}_{new}^{spm} = \frac{1}{N_{new}} \sum_{i \in New} \Pr(P_i^* \geq 0 \mid E_i^* \geq 0) = \frac{1}{N_{new}} \sum_{i \in New} \frac{\Pr(P_i^* \geq 0, E_i^* \geq 0)}{\Pr(E_i^* \geq 0)} = \\ = \frac{1}{N_{new}} \sum_{i \in New} \frac{\Phi_2(X_i\hat{\mu}, Z_i\hat{\delta}, \hat{\rho})}{\Phi_1(Z_i\hat{\delta})} \quad . \tag{12}$$

While the estimation of the parameters in (8), and the calculation of the policy effects and the respective standard errors in (9) through (12), may seem somewhat daunting, we emphasize that it can all be carried out in a program such as Stata, and contained in the Appendix.

4. Data

We use data from the Survey of Income and Program Participation (SIPP). SIPP is a nationally representative longitudinal household survey, which is specifically designed to collect detailed income and program participation information. It has a tri-annual feature such that the recall period between each interview is four months for every individual. For the time period covered in this paper (from October 1985 to August 1995), the length of the panels varies from 24 months for the 1988 panel to 40 months for the 1992 panel.[7] Although the sample universe is the entire U.S., the Census Bureau does not separately

[7] In total we used 7 panels from 1986 to 1993. The 1989 panel is not used because it was ended after only three waves.

identify state of residence for residents of nine low population states. Since state of residence information is critical for us to impute Medicaid eligibility, we drop all individuals whose state of residence information is not identified. We also restrict our sample to children living in original-sample households who are younger than 16 years old at the first time they are observed. Furthermore, we drop children who are observed only once, children who leave the sample and then return and children who move between states during the sample period for comparability with earlier studies. (In total, these observations constitute less than 8 percent of the sample.)

Although the four-month period increases the probability of accurate reporting, particularly relative to the fifteen-month recall period of the March Current Population Survey,[8] the SIPP suffers from the problem of "seam bias." Census Bureau researchers have shown that there are a disproportionate number of transitions between the last month of the wave and the first month of the next wave (see, e.g., Young 1989, Marquis and Moore 1990). Because of this seam bias problem, we estimate our models using only the fourth month of the each wave, dropping the first three months. While this approach has the disadvantage that information on the months other than the fourth month is lost, the advantage is that the data in the fourth month of each wave may be the most likely to be accurate since it is closest to the time of interview.

We impute eligibility in four steps. First, we construct the family unit relevant for Medicaid program participation and determine family income. Second, we assign family-specific poverty thresholds based on the size of the family and the year. Since Medicaid

[8] Bennefield (1996) finds that health insurance coverage in the early 1990s is measured more accurately in the SIPP than in the CPS, due in part to the shorter recall period.

eligibility results from AFDC eligibility, we then use information on the family income and family structure, along with the AFDC parameters in effect in the state and year, to impute eligibility for AFDC.[9] Finally, we assign Medicaid eligibility if any of the following conditions hold: the child is in an AFDC-eligible family; the child is income eligible for AFDC and either lives in a state with a "Ribicoff program" or lives in a state with an AFDC-UP program and has an unemployed parent; or the child's family income as a percent of the relevant poverty line is below the Medicaid expansion income eligibility cutoff in effect for that age child in his or her state of residence at that time.

In Table 1 we present the sample means for the variables used in our regressions.[10] Consistent with national trends, Medicaid coverage is higher in our sample in later panels. Furthermore, starting with the 1990 panel, which covers the period 1990-1992, there is a significant increase in our policy instrument $FRACELIG_i$ (discussed in Section 2) capturing eligibility expansions at the federal level occurring during this time period. In addition to the variables in Table 1, we also use state, year and age dummies for each child in the X_i vector in (3) and in our index functions (2) and (7) to control for the possible state-specific, age-specific and year-specific unobservables that could influence participating in Medicaid. Finally, since we use longitudinal data, we cluster the standard errors to account for dependency between person-specific observations.

[9] Families must pass two income tests to receive AFDC, the "gross test," which requires that a family's gross income be less than 1.85 times the state's need standard, and the "net test," which requires that a family's income after disregards be less than the state's payment standard. In determining AFDC eligibility, families are permitted to disregard actual childcare expenses up to a maximum. Since we do not know actual child care expenses, we assume that families deduct the full disregard for all children under age 6, but nothing for older children. These assumptions were also made by Cutler and Gruber (1996).

[10] These sample means have not been weighted, so they should not be considered to be representative of the nation.

5. Results

5.1 Parameter Estimates

Tables 2 and 3 show our parameter estimates for the linear probability model with interactions (LPMI) and the switching probit model (SPM). The two sets of estimates generally tell the same story. (The base case is a non-white female child in a female-headed family with more than two earners.) However, it is interesting to note that the effect of being white on take-up is not statistically significant in the LPMI while it is very significant and negative in the SPM, and that the normal statistics are substantially higher for the SPM.[11] In terms of the other coefficients, for both sets of estimates, the probability of take-up is significantly smaller than the base case when there is a male head, two parents present, the head is older and the head has more education, while it is significantly larger than the base case when there are fewer earners. The sex of the child does not affect the probability of take-up. In the LPMI the interaction terms for the demographics are jointly significant, and in the SPM the demographic variables in the participation or take-up equation (7) are jointly significant. It is worth stressing that even with very rich models in terms of having state, year and child age dummies interactions with eligibility in the LPMI (3), and having state, year and child age dummies in the take-up equation (7) for the SPM, the effects of demographics on take-up are generally well identified and precisely estimated. Given that all estimation and calculation of the standard errors are carried out in Stata, using these richer models gives important information about take-up without causing

[11] The higher normal statistics will reflect the normality assumption, and the fact that the LPMI will only use data on marginal individuals (in a local instrumental variables sense), while the SPM basically uses all of the data. Since the normality assumption cannot readily be justified on economic grounds, care must be exercised in considering these higher normal statistics as a plus for the SPM.

undue computational difficulties or a loss of statistical precision (see column (1) of Tables 4 and 5 below). As noted above, researchers will not be surprised by the sign of the coefficients for the LPMI model. *However, our analysis provides a quantitative measure of the differences in take-up rates across groups, and this is not currently available in the literature.*

Column 1 of Table 4 gives the predicted average take-up rates using the LPMI for different groups. For each group, the estimates give the take-up rate for eligible individuals within the category. For example, for each white child, we calculate the difference between the predicted probability when $\text{Elig}_i = 1$ and $\text{Elig}_i = 0$, and then take the average of the difference across white children.[12] All of the individual demographic groups have take-up rates that are estimated precisely and follow the pattern in Table 2. However, note that the comparisons in Table 4 answer a slightly different question than the parameter estimates in Table 2. The estimates in Table 2 show the effect on take-up of having a certain characteristic holding other characteristics constant, while those in Table 4 show the average probability of take-up for individuals with a certain characteristic given the other characteristics of those children. The estimate for families with more than two earners demonstrates that the LPMI can run into the problem of a negative estimated probability if a group has a very low probability of take-up.[13] In column 2 of Table 4 we show the estimates of the average take-up rate from the SPM for each demographic group treating each child as having a random draw from the error term distribution. Note that the predicted take-up rates follow the same rankings across groups as in the LPMI, but they are much

[12] Note that the individual error term in the take-up equation differences out of this equation.
[13] In other cases, some individuals have a predicted probability of participation greater than 1.

higher than the LPMI. Although we do not know of a formal treatment of local instrumental variables for the LPMI, we conjecture that the LPMI will estimate take-up rates for marginal children for periods when there was variation in coverage across states and age groups conditional on the other regressors, while the SPM takes the average across children over the whole sample period. Since our sample contains data from years before Medicaid eligibility was expanded when take-up rates were very high, it is not surprising that the SPM predicts much higher take-up rates than LMPI. In this situation the LPMI predictions are likely to be more appropriate for predicting expansions in 1995.[14] Column 3 shows the average take-up rates for the SPM when we take into account the fact that the eligible children will not represent a random draw of the population (because of the correlation between the error in the eligibility index function and the take-up index function). Not surprisingly the take-up rates in column 3 are larger than those in column 2, although the probabilities in column 3 are only larger by .04 to .06.

Table 5 shows the predicted effect on participation in Medicaid using our parameter estimates of those made newly eligible by a policy experiment: increasing the 1995 income limits by 10% for data from 1995. Again we do this using the LPMI, the SPM where we do not take into account the fact that the children are not a random sample in terms of their error term in the take-up index function, and the SPM where we do take into account the distribution of the error term among the newly eligible. For each case the probability of take-up in Table 5 is lower than the respective estimate in Table 4, reflecting the fact that the newly eligible have demographic variables that make take-up less likely than those who

[14] One means of adjusting the SPM to account for the change in the Medicaid eligibility regimes would be to use data only from the early 1990s and on when estimating the SPM.

were eligible under the actual 1995 rules.[15] The comparison of Tables 4 and 5 shows one of the significant advantages of the two approaches we explore here over the standard approach, since using the standard approach all one can do is use the average take-up rate estimated for the whole sample to predict the effect of the expansion. *The approaches we investigate here both adjust for the fact that different expansions will affect children from different families and thus have different take-up rates.* Note again that the predicted take-up rates are much higher for each group using the SPM, and that the LPMI model again runs into trouble in terms of a negative predicted probability for families with more than two earners.[16]

6. Conclusions

In this paper we suggest two approaches to estimating (i) the average take-up rate across demographic groups and (ii) the take-up response to an expansion of a social program attributable to a policy change in eligibility. Both of our approaches, the linear probability model with interactions (LPMI) and the switching probit model (SPM), allow demographics to play a significant role in take-up rates. The two approaches have some costs and benefits relative to each other. The linear probability model with interactions does not require a distributional assumption on the error terms, while the switching probit model does not suffer from criticisms directed at the linear probability model such as positive participation probabilities for ineligible individuals and the possibility that participation probabilities of individuals lie outside the unit interval. The SPM also allows

[15] The standard approach produces a take-up rate of 0.12. In principle this is closest to our estimate of the overall effect of the expansion using the LPMI, and we see that they are quite close also in practice.
[16] These are not completely independent, since the average take-up for the LPMI will be an average over probabilities of which some are estimated to be negative and some estimated to be greater than 1.

us to take into account the fact that distribution of the error term among the newly eligible under expansions of these rules are different from those previously eligible. On the other hand, the LPMI exploits variation in the data from a time which is more comparable to the expansion we consider, and thus is likely to provide more appropriate predictions of take-up under the expansion. Both approaches are easily implemented with Stata, produce precisely estimated effects, and will automatically adjust for the fact that different expansions will affect different children.

Using data from Medicaid and exploiting the exogenous policy changes that took place in the late 1980's and early 1990's, we estimate average take-up rates and the take-up response of different demographic groups according to race, family structure, education and number of earners in the family. Our results indicate that groups differ substantially in their take-up rates under both existing rules and an expansion of eligibility, and more importantly, we are able to quantify these differences, thus providing policy makers with important information that they can use to target groups with low participation rates in outreach campaigns to encourage participation.

References

Blumberg, Linda J., Lisa Dubay, and Stephen A. Norton. 2000. Did the Medicaid Expansions for Children Displace Private Insurance? An Analysis Using the SIPP. *Journal of Health Economics* 19(1): 33-60.

Currie, Janet and Jonathan Gruber. 1996a. Saving Babies: The Efficacy and Cost of Recent Expansions of Medicaid Eligibility for Pregnant Women. *Journal of Political Economy* 104(6): 1263-1296.

______________. 1996b. Health Insurance Eligibility, Utilization of Medical Care, and Child Health. *Quarterly Journal of Economics* 111(2): 431-466.

Cutler, David M. and Jonathan Gruber. 1996. Does Public Insurance Crowd Out Private Insurance? *Quarterly Journal of Economics* 111(2): 391-430.

Dubay, Lisa C. and Genevieve M. Kenney. 1996. The Effects of Medicaid Expansions on Insurance Coverage of Children. *The Future of Children* 6(1): 152-161.

Ham, John C. and Lara D. Shore-Sheppard. 2005. The Effect of Medicaid Expansions for Low-Income Children on Medicaid Participation and Private Insurance Coverage: Evidence from the SIPP. *Journal of Public Economics* 89(1): 57-83.

Hudson, Julie L., Thomas M. Selden, and Jessica S. Banthin. 2005. The Impact of SCHIP on Insurance Coverage of Children. *Inquiry* 42(3): 232-54.

LoSasso, Anthony T. and Thomas C. Buchmueller. 2004. The Effect of the State Children's Health Insurance Program on Health Insurance Coverage. *Journal of Health Economics* 23 (5): 1059–1082.

Marquis, Kent H. and Jeffrey C. Moore. 1990. Measurement Errors in the Survey of Income and Program Participation Program Reports. *1990 Annual Research Conference Proceedings,* Washington, D.C.: U. S. Bureau of the Census.

Quandt, Richard E. 1958. The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association* 53(284): 873-880.

_____________ . 1960. Tests of the Hypothesis That a Linear Regressions System Obeys Two Different Regimes. *Journal of the American Statistical Association* 55(290): 324-330.

_____________ . 1972. A New Approach to Estimating Switching Regression. *Journal of the American Statistical Association* 67(338): 306-310.

Selden, Thomas M., Jessica S. Banthin, and Joel W. Cohen. 1998. Medicaid's Problem Children: Eligible but not Enrolled. *Health Affairs* 17(3): 192-200.

Shore-Sheppard, Lara D. 2000. The Effect of Expanding Medicaid Eligibility on the Distribution of Children's Health Insurance Coverage. *Industrial and Labor Relations Review* 54(1): 59-71.

———————. 2005. Stemming the Tide? The Effect of Expanding Medicaid Eligibility on Health Insurance Coverage. NBER Working Paper #11091.

Thorpe, Kenneth E. and Curtis S. Florence. 1998. Health Insurance among Children: The Role of Expanded Medicaid Coverage. *Inquiry* 35(4): 369-379.

Van De Ven, W. P. M. M. and B. M. S. van Praag. 1981. The Demand for Deductibles in Private Health Insurance. *Journal of Econometrics* 17: 229-252.

Young, Nathan. 1989. Wave-seam Effects in the SIPP. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Alexandria, VA: American Statistical Association, pp. 393-398.

Table 1: Means of Variables Used in Estimation

	1986	1987	1988	1990	1991	1992	1993
Medicaid	0.1195	0.1158	0.1138	0.1619	0.1669	0.1810	0.2001
Imputed Eligibility	0.1871	0.1755	0.1792	0.2757	0.2953	0.3125	0.3341
Size of HIU	4.2196	4.1603	4.1827	4.1586	4.2161	4.1722	4.2207
White	0.8280	0.8301	0.8248	0.7817	0.8152	0.8047	0.8109
Male	0.5084	0.5154	0.5090	0.5120	0.5129	0.5201	0.5148
Two Parents	0.7585	0.7678	0.7677	0.7104	0.7439	0.7277	0.7335
Male Head Only	0.0220	0.0234	0.0186	0.0267	0.0289	0.0262	0.0212
No Earners	0.1402	0.1310	0.1235	0.1593	0.1539	0.1561	0.1619
One Earner	0.4109	0.4163	0.4188	0.4234	0.4204	0.4132	0.4065
Two Earners	0.3819	0.3901	0.3991	0.3677	0.3774	0.3816	0.3844
Age of Highest Earner	36.56	36.56	36.58	36.75	36.98	36.97	37.18
Education of Highest Earner	12.68	12.69	12.85	12.74	12.92	12.95	12.95
FRACELIG	0.1959	0.1861	0.1921	0.2823	0.3022	0.3208	0.3375
Years Covered	86-88	87-89	88-89	90-92	91-93	92-95	93-95
Number of Observations	44016	45691	40895	99446	66991	108572	101967

Notes: Unweighted means from SIPP panels noted above. See text for description of sample construction.

Table 2- Estimates for Medicaid Participation from IV Linear Probability Model with Interactions

	LPMI
Eligibility	0.0662
	(0.1086)
Eligibility*Size of Household	0.0337*
	(0.0056)
Eligibility*White	0.0143
	(0.0150)
Eligibility*Male	-0.0073
	(0.0105)
Eligibility*Two parents	-0.1360*
	(0.0170)
Eligibility*Male head only	-0.0962*
	(0.0278)
Eligibility*No earners	0.3287*
	(0.0445)
Eligibility*One earner	0.2497*
	(0.0347)
Eligibility*Two earners	0.1325*
	(0.0331)
Eligibility*Higher earner’s age	-0.0027*
	(0.0010)
Eligibility*Higher earner’s education	-0.0184*
	(0.0025)

Notes: All regressions include demographic main effects, year, age, state dummies, eligibility-year, eligibility-age and eligibility-state interactions. Standard errors have been corrected for repeated observations within individuals.
* denotes significantly different from zero at the 1% level of significance.

Table 3- Estimates for Medicaid Eligibility and Participation from Switching Probit Model

	Participation (1)	Eligibility (2)
Size of Household	0.1508*	0.2739*
	(0.0063)	(0.0049)
White	-0.3378*	-0.2837*
	(0.0185)	(0.0133)
Male	0.0055	-0.0145*
	(0.0153)	(0.0104)
Two parents	-0.8578*	-0.8488*
	(0.0212)	(0.0156)
Male head only	-0.7628*	-0.2355*
	(0.0443)	(0.0293)
No earners	1.9092*	2.4210*
	(0.0689)	(0.0321)
One earner	0.7761*	0.9698*
	(0.0650)	(0.0289)
Two earners	0.4208*	0.1173*
	(0.0639)	(0.0285)
Higher earner's age	-0.0176*	-0.0320*
	(0.0011)	(0.0008)
Higher earner's education	-0.0739*	-0.1474*
	(0.0033)	(0.0020)
FRACELIG	-	0.4661*
		(0.0060)

Notes: All regressions include year, age, and state dummies. Standard errors have been corrected for repeated observations within individuals.
* denotes significantly different from zero at the 1% level of significance.

Table 4- Average Take-up Rates from Linear Probability with Interactions and Switching Probit Models, 1995 Data

	Linear Probability with Interactions (1)	Switching Probit Random Error Draw (2)	Switching Probit Non-Random Error Draw (3)
All Population	0.2022	0.4771	0.5108
	(0.0183)	(0.0044)	(0.0029)
Race			
White	0.1864	0.4044	0.4409
	(0.0164)	(0.0048)	(0.0033)
Non-White	0.2393	0.6474	0.6744
	(0.0267)	(0.0049)	(0.0041)
Education			
High School Drop-out	0.2768	0.6009	0.6232
	(0.0242)	(0.0040)	(0.0036)
High School Graduate	0.1909	0.4444	0.4816
	(0.0171)	(0.0050)	(0.0031)
Some College	0.1418	0.3993	0.4413
	(0.0173)	(0.0056)	(0.0036)
Family Structure			
Female Head	0.2815	0.6957	0.7201
	(0. 0272)	(0.0044)	(0.0035)
Male Head	0.1315	0.2873	0.3230
	(0.0292)	(0.0125)	(0.0131)
Two parents	0.1254	0.2652	0.3082
	(0.0136)	(0.0055)	(0.0035)
Number of Earners			
No earner	0.2886	0.7795	0.7928
	(0.0347)	(0.0035)	(0.0032)
One earner	0.1714	0.2800	0.3271
	(0.0108)	(0.0063)	(0.0040)
Two earners	0.0153	0.1277	0.1860
	(0.0220)	(0.0068)	(0.0053)
More than two earners	-0.1569	0.0757	0.1120
	(0.0354)	(0.0086)	(0.0110)

Note: Standard errors in parentheses. All take-up rates are significantly different from zero at the 1% level.

Table 5- Average Take-up Rates for Newly Eligible Population after Raising the Income Limits by 10%, 1995 Data

	Linear Probability With Interactions (1)	Switching Probit Random Error Draw (2)	Switching Probit Non-Random Error Draw (3)
All Newly Eligible Population	0.0877	0.2179	0.2834
(985 Observations)	(0.0315)	(0.0076)	(0.0057)
Race			
White	0.0848	0.1916	0.2556
	(0.0315)	(0.0074)	(0.0057)
Non-White	0.0995	0.3238	0.3951
	(0.0347)	(0.0096)	(0.0077)
Education			
High School Drop-out	0.1315	0.2886	0.3484
	(0.0332)	(0.0077)	(0.0066)
High School Graduate	0.1049	0.2144	0.2798
	(0.0314)	(0.0081)	(0.0060)
Some College	0.0810	0.2171	0.2867
	(0.0321)	(0.0084)	(0.0063)
Family Structure			
Female Head	0.1677	0.3979	0.4785
	(0.0344)	(0.0108)	(0.0082)
Male Head	0.0540	0.1691	0.2305
	(0.0429)	(0.0123)	(0.0137)
Two parents	0.0598	0.1516	0.2114
	(0.0318)	(0.0068)	(0.0056)
Number of Earners			
No earner	0.1457	0.6280	0.6609
	(0.0489)	(0.0088)	(0.0080)
One earner	0.1264	0.2309	0.2997
	(0.0302)	(0.0084)	(0.0063)
Two earners	0.0146	0.1133	0.1813
	(0.0374)	(0.0075)	(0.0067)
More than two earners	-0.1764	0.0489	0.0908
	(0.0488)	(0.0073)	(0.0104)

Note: Standard errors in parentheses. All take-up rates are significantly different from zero at the 1% level.

Appendix - Calculating the Standard Errors

In this appendix, we describe how we calculate standard errors for each of our parameters of interest in the linear probability model with interactions (LPMI) and switching probit model (SPM) using the delta method.

A1. Linear Probability Model with Interactions (LPMI)

`Calculating the variance of the average take-up rate in the LPMI is straightforward using the delta method since our parameter of interest is only a linear sum of regression coefficients multiplied by the corresponding explanatory variables

$$\widehat{ATR}_j^{lpmi} = \frac{\sum_{i \in j} X_i \hat{\theta}}{N_j}. \tag{A1}$$

So the variance of the average take-up rate among the eligible can be written as

$$var(\widehat{ATR}_j^{lpmi}) = \left(\frac{\partial \widehat{ATR}_j^{lpmi}}{\partial \hat{\theta}} \right)' var(\hat{\theta}) \left(\frac{\partial \widehat{ATR}_j^{lpmi}}{\partial \hat{\theta}} \right) \tag{A2}$$

where

$$\frac{\partial \widehat{ATR}_j^{lpmi}}{\partial \hat{\theta}} = \frac{\sum_{i \in j} X_i}{N_j}. \tag{A3}$$

The variance of the take-up response among the newly eligible families after the policy expansion, $\hat{T}_{new}^{lpmi}$, can be calculated in an identical fashion using the observable characteristics of the sample of the newly eligible families.

A2. Switching Probit Model (SPM)

A2.1. Marginal Probability of Participation and Take-up Response without Accounting for Unobservables

The marginal probability of participation, which is our average take-up rate as well as the take-up response for the newly eligible population in our policy experiment assuming that the error term in the index function for eligibility is random, is given by

$$\Phi_1[X_i\hat{\mu}] = \int_{-\infty}^{X_i\mu} \phi(\varepsilon_i)d\varepsilon_i \,. \qquad \text{(A4)}$$

Our parameter of interest is

$$\widehat{ATR}_j^{spm} = [\sum_{i\in j} 1 - \Phi_1(-X_i\hat{\mu})]/N_j = [\sum_{i\in j} \Phi_1(X_i\hat{\mu})]/N_j \,. \qquad \text{(A5)}$$

Then, the variance of $\widehat{ATR}_j^{spm}$ is

$$\text{var}(\widehat{ATR}_j^{spm}) = \left[\frac{1}{N_j}\sum_{i\in j}\frac{\partial\Phi_1[X_i\hat{\mu}]}{\partial\hat{\mu}}\right]' \left[var(\hat{\mu})\right] \left[\frac{1}{N_j}\sum_{i\in j}\frac{\partial\Phi_1[X_i\hat{\mu}]}{\partial\hat{\mu}}\right]. \qquad \text{(A6)}$$

The derivative in (A6) can be written explicitly as

$$\frac{1}{N_j}\sum_{i\in j}\frac{\partial\Phi[X_i\hat{\mu}]}{\partial\hat{\mu}} = \frac{1}{N_j}\sum_{i\in j} X_i\phi(X_i\hat{\mu}) \,. \qquad \text{(A7)}$$

The variance of the take-up rate for the expansion is calculated in an analogous fashion.

A2.2. Take-up Response Accounting for Unobservables

Our parameter of interest is

$$\widetilde{ATR}_j^{spm} = \frac{1}{N_j} \sum_{i \in j} \frac{\Phi_2(X_i \hat{\mu}, Z_i \hat{\delta}, \hat{\rho})}{\Phi_1(Z_i \hat{\delta})}, \tag{A8}$$

where

$$\begin{aligned} &\Phi_2(X_i \hat{\mu}, Z_i \hat{\delta}, \rho) = \int_{-\infty}^{X_i \hat{\mu}} \int_{-\infty}^{Z_i \hat{\delta}} \phi_2(\varepsilon_i, e_i, \hat{\rho}) de_i d\varepsilon_i \quad \text{and} \\ &\Phi_1(Z_i \hat{\delta}) = \int_{-\infty}^{Z_i \hat{\delta}} \phi(e_i) de_i. \end{aligned} \tag{A9}$$

In A9, $\phi_2(.,.,\hat{\rho})$ is the bivariate normal density, $\phi(.)$ is the normal density, X_i is the vector of variables in the participation equation, Z_i is the vector of variables in the eligibility equation, and $\hat{\mu}$ and $\hat{\delta}$ are the vectors of coefficients in the participation and eligibility equations respectively.

The variance of $\widetilde{ATR}_j^{spm}$ can be written as follows

$$\text{var}(\widetilde{ATR}_j^{spm}) = \begin{bmatrix} \frac{\partial \widetilde{ATR}_j^{spm}}{\partial \hat{\mu}} & \frac{\partial \widetilde{ATR}_j^{spm}}{\partial \hat{\delta}} & \frac{\partial \widetilde{ATR}_j^{spm}}{\partial \hat{\rho}} \end{bmatrix} \begin{bmatrix} \text{var}(\hat{\mu}) & \text{cov}(\hat{\mu}, \hat{\delta}) & \text{cov}(\hat{\mu}, \hat{\rho}) \\ \text{cov}(\hat{\mu}, \hat{\delta}) & \text{var}(\hat{\delta}) & \text{cov}(\hat{\delta}, \hat{\rho}) \\ \text{cov}(\hat{\mu}, \hat{\rho}) & \text{cov}(\hat{\delta}, \hat{\rho}) & \text{var}(\hat{\rho}) \end{bmatrix} \begin{bmatrix} \frac{\partial \widetilde{ATR}_j^{spm}}{\partial \hat{\mu}} \\ \frac{\partial \widetilde{ATR}_j^{spm}}{\partial \hat{\delta}} \\ \frac{\partial \widetilde{ATR}_j^{spm}}{\partial \hat{\rho}} \end{bmatrix}. \tag{A10}$$

The derivatives can explicitly be written as

$$\frac{\partial \widetilde{ATR}_j^{spm}}{\partial \hat{\mu}} = \frac{1}{N_j} \sum_{i \in j} \frac{\frac{\partial \Phi_2(X_i \hat{\mu}, Z_i \hat{\delta}, \hat{\rho})}{\partial \hat{\mu}}}{\Phi_1(Z_i \hat{\delta})} \ , \tag{A11}$$

$$\frac{\partial \widetilde{ATR}_j^{spm}}{\partial \hat{\delta}} = \frac{1}{N_j} \sum_{i \in j} \frac{\left(\frac{\partial \Phi_2(X_i \hat{\mu}, Z_i \hat{\delta}, \hat{\rho})}{\partial \hat{\delta}} \right) \Phi_1(Z_i \hat{\delta}) - \left(\frac{\partial \Phi_1(Z_i \hat{\delta})}{\partial \hat{\delta}} \right) \Phi_2(X_i \hat{\mu}, Z_i \hat{\delta}, \hat{\rho})}{\left\{ \Phi_1(Z_i \hat{\delta}) \right\}^2} , \tag{A12}$$

$$\frac{\partial \widetilde{ATR}_j^{spm}}{\partial \hat{\rho}} = \frac{1}{N_j} \sum_{i \in j} \frac{\frac{\partial \Phi_2(X_i \hat{\mu}, Z_i \hat{\delta}, \hat{\rho})}{\partial \hat{\rho}}}{\Phi_1(Z_i \hat{\delta})} \tag{A13}$$

where

$$\frac{\partial \Phi_2(X_i \hat{\mu}, Z_i \hat{\delta}, \hat{\rho})}{\partial \hat{\mu}} = X_i \phi(X_i \hat{\mu}) \Phi[(Z_i \hat{\delta} - \hat{\rho} X_i \hat{\mu}) / \sqrt{1 - \hat{\rho}^2}] , \tag{A14}$$

$$\frac{\partial \Phi_2(X_i \hat{\mu}, Z_i \hat{\delta}, \hat{\rho})}{\partial \hat{\delta}} = Z_i \phi(Z_i \hat{\delta}) \Phi[(X_i \hat{\mu} - \hat{\rho} Z_i \hat{\delta}) / \sqrt{1 - \hat{\rho}^2}] , \tag{A15}$$

$$\frac{\partial \Phi_2(X_i \hat{\mu}, Z_i \hat{\delta}, \hat{\rho})}{\partial \hat{\rho}} = \phi_2(X_i \hat{\mu}, Z_i \hat{\delta}, \hat{\rho}) \quad \text{and} \tag{A16}$$

$$\frac{\partial \Phi_1(Z_i \hat{\delta})}{\partial \hat{\delta}} = Z_i \phi(Z_i \hat{\delta}) \ . \tag{A17}$$