The Effect of Transfer Programs on Labor Supply in the Presence of Preference Heterogeneity and Variable Takeup

Robert Moffitt
Johns Hopkins University

November, 2005

Earlier versions of this paper have been presented at Stanford University, University College London, the ESPE Conference in Bergen, Norway in June 2004, and the World Congress of the Econometric Society in August, 2005. The author would like to thank the National Institute of Child Health and Human Development for research support and Jiawei Chen, Berna Demiralp, Steffen Reinhold, and Katie Winder for research assistance.

welfls1_v5.wpd
11/2/05

Abstract

This paper applies recent developments in the treatment effects literature to the traditional static model of the effect of transfer programs on labor supply. Preferences are allowed to be heterogeneous, which generates heterogeneity in the labor supply effect of transfer program participation on the basis of both observables and unobservables. The resulting selection problem implies that those on the margin of participation may have different labor supply responses than inframarginal individuals, and that the mean labor supply response in a population may vary with the level of the participation rate and the distribution of observable characteristics. A simple version of the model is specified which makes minimal identifying assumptions and which imposes minimal parametric restrictions on the unobservables. A flexible parametric specification for the observables is implemented with series approximations. The model is applied to a cross-section of the Survey of Income and Program Participation. Strong heterogeneity in response on the basis of observables is found as well as somewhat weaker evidence of heterogeneous response on the basis of unobservables. Those with greater labor supply disincentives are shown to participate in the program earlier than those with smaller disincentives. Nonlinearities in the effects of the instruments and of the exogenous conditioning variables are found to be important and to affect the estimates of program impact.

There is a large empirical literature estimating the effects of transfer programs on labor supply, starting with the studies of the effect of a negative income tax in the 1960s and continuing through to the present time. There have been studies of a wide range of types of programs both in Europe and the U.S.; see Danziger et al. (1981), Hausman (1985), Blundell and MaCurdy (1999), and Moffitt (1992, 2002, 2003) for reviews. However, many recent developments in the treatment effects literature have not been incorporated into these models. The most important is the recognition of the importance of heterogeneous response, but also important is the recognition of the importance of making minimal identifying assumptions, of using flexible or nonparametric methods, and of allowing for nonlinearities in response; see Imbens and Angrist (1994) and Heckman and Vytlacil (1999, 2000, 2001), as well as Heckman and Robb (1985) and Björklund and Moffitt (1987) for earlier studies of heterogeneous response. Most of the literature on the labor supply disincentives of transfer programs has assumed homogeneous responses and many studies have used highly parametric models which constrain the form in which nonlinearities appear.

This paper incorporates these developments into a simple model of the effect of transfer programs on labor supply. The model is not fully structural because labor supply responses are summarized by the coefficient on a binary participation indicator and are not decomposed into income and substitution elasticities, a topic left for future work. The coefficient is allowed to vary in the population on the basis of both observables and unobservables, implying that the mean labor disincentive of a transfer program can be expected to vary with the level of the participation rate and the distribution of observables in the population. A two-equation treatment effects model is specified in a way designed for relatively easy estimation with nonlinear least squares but in which treatment responses are also identified with minimal assumptions, albeit only over the support of the participation probabilities in the data. Nonlinearities in the participation response, as well as in the effects of the instruments and the other conditional variables in the model, are modeled with series estimation using linear splines.

The model is applied to a cross section of the Survey of Income and Program Participation on a sample of single mothers eligible for the Aid to Families with Dependent Children program in the early 1990s. The results show strong evidence of heterogeneity of labor supply responses, with most estimates implying that those with greater labor supply disincentives are the first to join the program. When the heterogeneity is broken down into observable and unobservable components, it is found that responses vary strongly with observable characteristics and somewhat more weakly with unobservable characteristics. Nonlinearities in the effects of the instruments and of the conditioning variables is also found to be somewhat important, suggesting that the conventional approach of entering those variables in linear index form may be problematic.

The next section of the paper incorporates preference heterogeneity into the traditional static labor supply model of transfer program participation. The econometric model is set up in the following section, and the subsequent section reports the empirical results. The findings are summarized at the end.

I. Adding Heterogeneity to the Canonical Labor Supply Model of Transfers

The canonical model of the labor supply response to transfers with variable (i.e., incomplete) takeup is (Moffitt, 1983):

$$U(H_i, Y_i; \theta_i) - \phi_i P_i \tag{1}$$

where H_i =hours of work for individual i, Y_i =disposable income, P_i is a program participation indicator, θ_i is a vector of labor supply taste parameters, and ϕ_i represents fixed costs of participation in utility units. The separability of P_i from the function for H_i and Y_i is for analytic convenience and is not required for any of the following results. Allowing for fixed costs of participation--in money, time, or utility, with the exact type unspecified--is required because many individuals who are eligible for transfer programs do not participate in them. This is a pervasive empirical phenomenon and an intrinsic part of the model because, without variable takeup, the treatment effects model is not applicable to the problem (see below).

The individual is assumed to face an hourly wage rate W_i and to have available exogenous, non-transfer nonlabor income N_i. The welfare benefit formula is $B_i = G - tW_iH_i - rN_i$ and hence the budget constraint is[1]

$$\begin{aligned} Y_i &= W_i(1-t)H_i + G + (1-r)N_i && \text{if } P_i = 1 \\ Y_i &= W_iH_i + N_i && \text{if } P_i = 0 \end{aligned} \tag{2}$$

[1] The parameters G, t, and r are held fixed for expositional purposes but will vary across individuals in the data, as described below.

The resulting labor supply model is represented by two functions, a labor supply function conditional on participation and a participation function:

$$H_i = H\,[\,W_i(1 - tP_i)\;,\; N_i + P_i(G - rN_i)\;;\;\theta_i] \tag{3}$$

$$\begin{aligned} P_i^* &= V\,[\,W_i\,(1\text{-}t),\; G + N_i(1\text{-}t)\;;\;\theta_i] \;-\; V\,[W_i\,,N_i;\,\theta_i] \;-\; \phi_i \\ P_i &= 1(P_i^* \geq 0) \end{aligned} \tag{4}$$

where V is the indirect utility function and 1(•) is the indicator function. Nonparticipants, those for whom P* is negative, are of two types: low-work individuals for whom a positive benefit is offered and a utility gain (in V) could be obtained but who do not participate because ϕ_i is too high; and high-work individuals for whom the utility gain (in V) is negative and who would not participate even if ϕ_i were zero (these individuals are above the eligibility point, or "above-breakeven" in the terminology of the literature). Figure 1 is the familiar income-leisure diagram showing three different individuals who respond to the transfer program constraint by continuing to work above the breakeven point (III), working below breakeven but off the program (II), and working below breakeven and on the program (I'; I is the pre-program location for this individual).

Our initial main interest centers on the response to the program for individual i, which we denote Δ_i:

$$\Delta_i(\,\theta_i\,) \;=\; H\,[W_i(1\text{-}t)\,,\,G{+}N_i(1\text{-}r);\,\theta_i] \;-\; H\,[W_i,\,N_i;\,\theta_i] \tag{5}$$

which is a heterogeneous response if θ_i varies with i. Individual values of Δ_i will never be identified by the data, but the mean of those values over some populations or subpopulations will be. Letting

$$\begin{aligned} S_\phi &= \text{support of } \phi \\ S_\theta(\phi) &= \text{set of } \theta \text{ s.t. } P_i=1 \text{ conditional on } \phi, \end{aligned} \tag{6}$$

the mean effect of the transfer program over the entire population, participants and non-participants combined, conditional on the budget constraint variables, is

$$\begin{aligned} \tilde{\Delta} &= E(\Delta_i P_i \mid W_i, N_i, G, t, r) \\ &= \int_{S_\phi} \int_{S_\theta(\phi)} \Delta_i(\theta_i \mid W_i, N_i, G, t, r) \; g(\theta_i, \phi_i) \, d\theta_i \, d\phi_i \end{aligned} \tag{7}$$

where $g(\theta_i, \phi_i)$ is the p.d.f. of the two heterogeneity components.[2] Letting S_θ be the unconditional support of θ, the participation rate in the population is

$$\begin{aligned} P &= E(P_i \mid W_i, N_i, G, t, r) \\ &= \int_{S_\phi} \int_{S_\theta} 1\{V[W_i(1-t), G+N_i(1-t); \theta_i] - V[W_i, N_i; \theta_i] - \phi_i \geq 0\} \; g(\theta_i, \phi_i) \, d\theta_i \, d\phi_i \end{aligned} \tag{8}$$

[2] We have assumed for convenience that g is independent of W,N,G,t, and r. Note that $S_\theta(\phi)$ is a function of the budget constraint parameters but this is not explicitly represented in its argument.

and the mean labor supply response among those who participate is

$$\tilde{\Delta}_{P_i=1} = \tilde{\Delta} / P \tag{9}$$

The marginal response to a change in a program parameter p (p=G,t, or r)--that is, the mean Δ of those who change participation--is $[(\partial\tilde{\Delta}/\partial p)/(\partial P/\partial p]$. These effects have been discussed extensively in the treatment effects literature, and are defined within an econometric model in the next section.

The distribution of θ_i affects the mean response in the population in two ways: first, by affecting the distribution of Δ_i across the population--that is, the distribution of response if all individuals participate--and, second, by altering which of those individuals participate because θ_i appears in (4). The distribution of ϕ_i affects mean response only through the latter mechanism, by altering the composition of the participant population.

While Δ_i, $\tilde{\Delta}$, and $\tilde{\Delta}_{P_i=1}$ must be negative according to theory, how they change as the participation rate changes is ambiguous in sign. This is because the magnitude of Δ_i has no determinate relationship to the magnitude of $V[W_i(1\text{-}t),\ G+N_i(1\text{-}t)\ ;\ \theta_i] - V\ [W_i,N_i;\ \theta_i]$. If Δ_i is decreasing in an element of θ_i at some point in the budget space, implying a greater labor supply disincentive as θ_i rises, this does not imply a greater utility gain to going onto welfare. Ceteris paribus, a greater reduction in labor supply corresponds to a lesser increase (or greater decrease) in consumption, and therefore whether the utility gain to going onto welfare is increasing or decreasing in θ_i depends on the relative utility to the individual of increased consumption versus increased leisure. Thus whether initial entrants to a program are those with greater or lesser work disincentives than the population as a whole is an empirical question.[3]

Most of the labor supply literature in general, and the literature on labor supply responses to transfer programs in particular, has assumed that heterogeneity resides only in an intercept term:

$$H_i = \theta_{0i} + h\,[W_i(1\text{-}tP_i)\,,\, N_i\text{+}P_i(G\text{-}rN_i)\,;\, \theta_1] \tag{10}$$

In this case, Δ_i is a function only of θ_1 and hence is homogeneous. There is still a selection problem because $\tilde{\Delta}$ is a function of θ_{0i} and hence participants are selected on their values of labor supply, but marginal effects are constant. This model is particularly implausible because it implies that the response is the same whether the individual is a high-work individual who is offered a small benefit at their initial hours of work or a low-work individual who is offered a high benefit at those initial hours. The only significant introduction of heterogeneous response in the nonlinear budget constraint literature appeared in Burtless and Hausman (1978) and Hausman (1981), who assumed that the conditional labor supply function has a heterogenous income coefficient. However, its distribution was restricted and no other response parameters were allowed to vary across individuals.

The Burtless and Hausman (1978) and Hausman (1981) models, like many of those in the literature, assumed full participation and 100 percent welfare takeup, contrary to the data. With

---

[3] An additional consideration is that the lower limit of zero for hours of work limits the magnitude of the labor supply response. Consequently, for example, those whose labor supply in the absence of the program is close to zero necessarily can have only small responses when going onto the program. This feature is not represented in the model presented here but will be present in the estimated responses reported below.

full participation, there is no meaningful distinction between the choice of hours of work and the choice of welfare participation, and hence the treatment effects model in general is not applicable. There is still an estimation problem because nonlinear budget constraints typically induce nonlinear labor supply functions, but this nonlinearity can be addressed by estimating the unconditional labor supply function H(W,N,G,t,r) directly with appropriate nonlinear methods.

II. A Formulation of the Heterogeneous Treatment Effects Model

In this section we formulate an econometric model to capture these heterogeneous effects. However, it will be reduced form in the sense that labor supply will be estimated as a function of program participation alone, and the underlying utility structure of the problem will not be explicitly specified. The work here can be viewed as an initial step toward a model with such a structure. Our specification consists of a two-equation model, one of which is an approximation of (3)--that is, H_i as a function of P_i--but with W_i not explicitly represented (assume that a wage equation has been substituted in) and where the restrictions in that function required by theory are not imposed. The other equation is an approximation of (4), again without theoretical restrictions imposed. The labor supply function in (3) implies that the effect of participation on labor supply should interact with observable and unobservable determinants of wages, and with nonlabor income, and this will be represented in the model.

The model is adapted from those in the treatment effects literature.[4] Let y_i be an outcome

[4] There is a voluminous literature. Papers which emphasize heterogeneity of response include Imbens and Angrist (1994), Heckman and Vytlacil (1999,2000,2001), and others, and older work by Heckman and Robb (1985) and Björklund and Moffitt (1987).

variable for individual i, P_i a dummy variable signifying participation in the program, and Z_i an instrumental variable with a multinomial distribution. The variable Z_i represents observable determinants of ϕ_i which do not appear elsewhere in the model. Ignoring observables other than these for the moment, a completely unrestricted model can be written as

$$y_i = \beta_i + \alpha_i P_i \tag{11}$$

$$P_i^* = k(Z_i, \delta_i) \tag{12}$$

$$P_i = 1(P_i^* \geq 0) \tag{13}$$

where β_i and α_i are scalar random parameters and δ_i is a vector of random parameters. All parameters are allowed to be individual-specific and to have some joint distribution $f(\beta_i, \alpha_i, \delta_i)$; a separate model (11)-(13) exists for each individual. Eqn (11) is to be interpreted as a description of potential outcomes, not just a description of how y_i varies with P_i in any particular population; hence α_i is the object of interest. What can be estimated, however, is only the mean of α_i on some population(s).

If we condition (11) on P_i, we obtain

$$E(y_i \mid P_i) = E(\beta_i \mid P_i) + E(\alpha_i \mid P_i) P_i \tag{14}$$

which illustrates one conditional mean of interest. But to see which of the classes of objects can be identified, we work instead with the reduced form by conditioning (11)-(13) on Z_i:

$$E(y_i \mid Z_i{=}z) = E(\beta_i \mid Z_i{=}z) + E(\alpha_i \mid P_i{=}1, Z_i{=}z)\, \text{Prob}(P_i{=}1 \mid Z_i{=}z) \tag{15}$$

$$E(P_i \mid Z_i{=}z) = \text{Prob}\,[k(z, \delta_i) \geq 0] \tag{16}$$

We make the following minimal identifying assumptions:

A1. $E(\beta_i \mid Z_i{=}z) = \beta$ (17)

A2. $E(\alpha_i \mid P_i{=}1, Z_i{=}z) = g[E(P_i{=}1 \mid Z_i{=}z)]$ (18)

A3. $E(P_i \mid Z_i{=}z) - E(P_i \mid Z_i{=}z')$ is zero or the same sign for all i for any distinct values z and z' (19)

Assumptions A1 and A2 are mean independence assumptions needed for Z_i to be a valid instrument. They state that the mean of the random intercept and the conditional mean of the random coefficient, respectively, are, in the former case, independent of Z_i and, in the latter case, dependent only on the fraction treated and not directly on Z_i. The necessity of (19) was recognized by Imbens and Angrist (1994), who termed it a "monotonicity" assumption, and constitutes a restriction on the distribution of δ_i.

With these assumptions, and letting $F(Z_i)=E(P_i{=}1 \mid Z_i)$, (15) and (16) can be rewritten as

$$y_i = \beta + g[F(Z_i)]\, F(Z_i) + \epsilon_i \tag{20}$$

$$P_i = F(Z_i) + \nu_i \tag{21}$$

where F is a proper probability function and where $E(\epsilon_i \mid gF) = E(\nu_i \mid F) = 0$ by construction. No other restriction on the distribution of ϵ_i or ν_i is made. The implication of response heterogeneity can be seen in (20) to be an effect of program participation which varies with the level of participation, hence resulting in an inherent nonlinearity of y in F. A homogeneous-effects model is one in which g is a constant.[5]

Nonparametric identification of the parameters of (20) and (21) is straightforward given that P_i is binary and Z_i has a multinomial distribution. $F(Z_i)$ is identified at each point Z_i=z from the population mean of P_i at that z. The parameter β and the function $g[F(Z_i)]$ at Z_i=z are identified provided a normalization is made to the g function, e.g., $g[F(Z_1)]=1$.[6] The treatment effects discussed in the literature are all defined by g: the average treatment effect is g(1); the effect of the treatment on the treated when a fraction F_j is treated is $g(F_j)$; the local average treatment effect between two participation fractions F_j and $F_{j'}$ is $[F_j g(F_j)-F_{j'} g(F_{j'})]/(F_j-F_{j'})$; and the marginal treatment effect at some point F_j is $\partial y/\partial F=g'(F_j)F_j+g(F_j)$, which must be obtained by some smoothing method given the multinomial assumption on Z_i.[7]

The problem will be approached with a flexible parametric rather than fully nonparametric approach. We will be parametric on F but flexible parametric on g and k, approximating each with a series expansion using regression splines with parameter vectors

---

[5] Heckman et al. (2004) have recently noted that a test of heterogeneity in response can be conducted by testing for nonlinearities of y in F. This suggestion also follows from the formulation here and will be part of the empirical work below.

[6] In the special case that the support of Z_i in the data contains an element s.t. $F(Z_i)=0$, no normalization is necessary.

[7] The effect of the treatment on the treated as defined here is conditional on z; however, by integrating z out, the more conventional effect of the treatment on the treated is obtained.

θ_1,and θ_2, respectively. The series expansion approach is nonparametric in the limit but can also be treated as a flexible-form parametric approach.[8] In addition, it is computationally convenient for a fixed order of the series because splines lead to easily estimable nonlinear functions. Consequently, (20) and (21) can be consistently estimated by joint nonlinear least squares (NLS) with any simple regression package (with robust standard errors). The estimation problem is

$$\min_{\theta_1,\theta_2} \sum_i w_{1i} [y_i - \beta - g[F(k(Z_i,\theta_2)); \theta_1] F(k(Z_i,\theta_2))]2 \quad + \quad \sum_i w_{2i} [D_i - F(k(Z_i,\theta_2))]^2 \tag{22}$$

where w_{1i} and w_{2i} are weights to improve efficiency.[9]

Adding a vector of exogenous observables X_i to the model will also be approached nonparametrically, again building off of the conventional linear regression model. The model is now

$$y_i = \alpha_i P_i + h_i(X_i) \tag{23}$$

$$P_i^* = k(Z_i, X_i, \delta_i) \tag{24}$$

$$P_i = 1(P_i^* \geq 0) \tag{25}$$

---

[8] Heckman et al. (2004) suggest a local linear kernel approach, instead. That approach would be more difficult to implement here.

[9] In the application below, cross-equation weights will be used as well. Two-stage estimation (i.e., two-stage least squares) is an option and corresponds to instrumental variables estimation when k is allowed to be fully nonparametric in Z_i, and yields the local average treatment effect just defined. However, if k is considered instead to be a smoothed function of Z_i or if Z_i is continuous, insertion of a first-stage predicted F into (20) will generate inconsistent parameter estimates because F appears nonlinearly in (20), and hence two-stage estimation should not be used.

We assume

B1. $E[\ h_i(X_i) \mid X_i{=}x, Z_i{=}z) \ = \ h(x)$ (26)

B2. $E(\alpha_i \mid P_i{=}1, X_i{=}x, Z_i{=}z) \ = \ g[E(P_i{=}1 \mid X_i{=}x, Z_i{=}z), x]$ (27)

B3. $E(P_i \mid Z_i{=}z, X_i{=}x) - E(P_i \mid Z_i{=}z', X_i{=}x)$ is zero or the same sign for all i for any distinct values z and z' (28)

Then, conditioning (23)-(25) on X_i and Z_i, we have:

$$y_i \ = \ g[F(Z_i,X_i), X_i]\, F(Z_i,X_i) \ + \ h(X_i) \ + \ \epsilon_i \tag{29}$$

$$P_i \ = \ F(Z_i,X_i) \ + \ \nu_i \tag{30}$$

where, again, the errors are mean-independent of the regressors by construction. The two equations are estimated jointly with nonlinear least squares, and regression splines are used to approximate the unknown functions. In this model, the effect of participation on labor supply varies with observable as well as unobservable characteristics.

It should be noted that the effects of welfare participation on labor supply that are nonparametrically identified in this model are only those induced by the instrument, Z_i. For example, effects of the budget constraint variables on labor supply, which are contained in X_i, are not nonparametrically identified, yet are of obvious interest. However, because the approach below will not be fully nonparametric, but only flexible parametric, a mapping from Z_i to X_i will be possible, and hence the effects of the budget constraint on labor supply will be inferred.[10]

III. Empirical Analysis

We study the labor supply effects of the well-known U.S. cash transfer program, the Aid to Families with Dependent Children (AFDC) program using data from the Survey of Income and Program Participation (SIPP) in the early 1990s. The SIPP is a set of rolling, short (12 to 48 month) panels which are representative samples from the U.S. population. The initial panels began in 1984 and subsequent panels have begun approximately annually, and have between 30,000 and 70,000 families in each. We draw a sample from all the panels that have data in the period 1989-1991, which is just prior to a major restructuring of the program which began in 1992-1993 and which imposed work requirements, time limits, and other features not captured in our model. Prior to 1992-1993, the AFDC program was close to a straight cash transfer program with a conventional benefit formula. The SIPP, unlike other U.S. survey data sets such as the Current Population Survey, the Michigan Panel Study on Income Dynamics, and the National Longitudinal Study, asked welfare participation, labor supply, income, family structure, and other questions all as of the month previous to the survey, not as of the prior calendar year; past studies using the latter data sets have been required to measure one of more of these variables at a different calendar time than the others. To increase sample size, we draw our samples from all

---

[10] Ichimura and Taber (2000) have emphasized the necessity of an exchangeability condition between Z_i and X_i to be able to extrapolate beyond the nonparametrically estimated effects of Z_i.

individuals interviewed in the Spring of 1989, 1990, and 1991 SIPP surveys.[11]

Eligibility for AFDC in this period required sufficiently low assets and income and, for the most part, required that eligible families be single mothers with children under 18. Our subsample is therefore restricted to such families, as most past studies of AFDC have also done. In order to concentrate on the AFDC-eligible population, we also restrict our sample to those with completed education of twelve years or less, with nontransfer nonlabor income less than $1,000 per month, and between the ages of 20 and 55. The resulting data set has 5,722 observations. We define variables for average hours of work per week in the prior month (y), participation in AFDC at any time in the prior month (P), and demographics such as education, age, race, and family structure. The hourly wage rate is omitted because it is only available for workers and is assumed to be proxied by demographic characteristics. Nontransfer nonlabor income in the prior month is included in the model, however. The AFDC guarantee for a family of four in the individual's state of residence is also included. Tax rates in the AFDC program are not included because they were uniform nationwide (both equal to 100 percent over this period). The demographic and budget constraint variables constitute the X vector presented in the prior section. The names, definitions and means of the variables used in the estimation appear in Appendix A. Thirty-one percent of the sample was on AFDC in the prior month.

The instrumental variables (the Z vector) are selected to proxy fixed costs of participation

[11] We thus draw data from multiple panels: the 1989 panel, which began in February 1989 and ended in January 1990; the 1990 panel, which began in February, 1990 and ended in September, 1992; and the 1991 panel, which began in February 1991 and ended in September, 1993. We draw the sample from all families interviewed from any of these three panels in the four-month period February-May of each year 1989, 1990, and 1991. Families are interviewed in a rolling 4-month cycle and are asked questions about the previous four months; we use only the data from the month prior to the survey

in AFDC, as implied by the theory. Institutional descriptions of the AFDC program have revealed that non-financial barriers have always been present in the program and have hindered participation, perhaps intentionally on the part of the states to keep caseloads down. Data are available on a number of proxies for these barriers. Information on several were gathered from official documents and were pretested in OLS estimations of the welfare participation equation. From this exercise, three emerged as consistently significant and with the expected sign: the error rate made by the state resulting incorrectly in denial of eligibility (collected by the federal government as part of its regulation of the program), the percent of applications denied because of a failure on the part of the applicant to comply with all procedural requirements (an indication of the amount of paperwork and bureaucracy imposed on prospective recipients), and administrative expenses per case in the state (to be interpreted as an indicator of the level of bureaucracy in the program). All three affect welfare participation negatively. Their means and sources are given in Appendix A.

Initial Results. The initial model specification is the following:

$$y_i = X_i\beta + [X_i\lambda + g(F(X_i\delta + Z_i\eta))] F(X_i\delta + Z_i\eta) + \epsilon_i \quad (31)$$

$$P_i = F(X_i\delta + Z_i\eta) + \nu_i \quad (32)$$

and thus we initially assume linearity in the index functions for X_i and Z_i but allow nonlinearity in g, approximating it with a spline function, $g(F) = \gamma_0 + \sum_j \gamma_j Max(0,F-\pi_j)$ where j indexes the splines and π_j is the j^{th} knot.[12] F is taken as the normal c.d.f., but its argument will be expanded flexibly below. Denote u_i as a 2x1 column vector of residuals for individual i, the first

[12] The vector $X_i\lambda$ excludes a constant term.

being the residual in (31) and the second being the residual in (32). Then the unknown parameters in both equations, which we denote θ, are estimated by the criterion

$$\theta = \underset{\theta}{\text{Arg Min}} \sum_{i=1}^{n} u_i' \hat{\Omega}^{-1} u_i \qquad (33)$$

where $\hat{\Omega}$ is a 2x2 matrix of variances of the residuals obtained from an unweighted first-stage estimation. The covariance matrix of the estimated θ is calculated as

$$\text{Var}(\hat{\theta}) = \left(\sum_{i=1}^{n} M_i' \hat{\Omega}^{-1} M_i\right)^{-1} \left(\sum_{i=1}^{n} M_i' \hat{\Omega}^{-1} u_i u_i' \hat{\Omega}^{-1} M_i\right) \left(\sum_{i=1}^{n} M_i' \hat{\Omega}^{-1} M_i\right)^{-1} \qquad (34)$$

where M_i is the 2xK matrix of derivatives of the two conditional mean functions w.r.t. the K parameters in the model.

OLS estimation of the y equation, regressing y on X and P, yields a coefficient on the participation dummy of -24.4 with a standard error of .46, obviously indicating that welfare recipients work less than nonrecipients, even conditional on X; unconditional on X, the gap in hours worked for the two groups is -26.3 hours, as shown in Appendix A, so the X variables have little effect on reducing the gap.

Estimates of several specifications of the model in (31)-(32) are shown in Table 1. Column (1) estimates a model with a constant treatment effect and with λ=0, and shows a disincentive to work of about 30 hours per week, a sizable disincentive, a bit larger than the OLS estimate. However, this coefficient disguises significant variation in the effect and hence is misleading, as will be seen momentarily. Figure 2 shows a histogram of predicted welfare participation rates in the data, obtained from the fitted participation equation in this model (see Table 1). There are a significant number of near-zero probabilities, thereby allowing us to estimate the effect of the treatment on the treated at other participation points; however, this ability is very much a function of the linearity imposed on the participation equation. A rather large spread of participation rates is estimated, with the distribution thinning out only at participation rates above .90. The three instruments in the participation equation are individually weak except for the eligibility error rate, but the three variables are significantly correlated and the F-statistic for their joint significance is 7.25 (p=.00007).

Column (2) allows g to be linear in the probability of participation, again with $\lambda=0$. Labor supply disincentives are implied to be very large at low participation rates and to fall as participation rises, implying that those with the greatest labor supply responses enter first. The marginal treatment effect (MTE), which is the derivative of y w.r.t. F, is shown in Figure 3; those brought into the program have very large responses while those brought in at the end hardly have any labor supply disincentives at all.

These results change in a major way when a covariate vector X is allowed to affect the labor supply response. The coefficients on the response to unobservable participation variation fall in magnitude and significance, implying that most of the response heterogeneity is a result of variation in observable dimensions. The variation in the MTE arising from observables ranges from -7.1 at F=0 to 1.9 at F=1, whereas the value of the observable component of the MTE, $X\lambda$, is -26.9 at mean X.[13]

[13] The MTE is the derivative of y w.r.t P, which is equal to $(\gamma_0 + 2F\gamma_1) + X\lambda$, where γ_0=-7.1 and γ_1=4.5 and λ is set equal to the values in column (3) of Table 1.

This is an important finding because most studies, apart from those in the matching literature, do not interact their treatment variables with observables. These results imply not only that a misleading conclusion about selection on unobservables can result from such a specification, but also that the average effect so obtained is not very meaningful. Figure 4, for example, shows the value of the MTE at the mean of the X vector, and shows that labor supply effects vary from approximately -27 at low participation rates to -18 at high participation rates. Neither the constant-effect estimate of -30.9 in column (1) nor the MTE values from the column (2) specification (see Figure 3) are centered in this interval. The high values of response in those specifications arise from the effects of the distribution of X, which distorts the average response, at least as defined in Figure 4.

It should also be noted that there is no necessary relationship between the magnitude of labor supply disincentives and welfare participation rates across demographic groups (as opposed to expansion or contraction of the group for a given demographic composition), as can be seen by comparing the signs of the coefficients on the same variables in the λ and δ vectors. For example, those with low levels of nonlabor income are more likely to participate in welfare and also have greater work disincentives; but those with low levels of education are also more likely to participate but more likely to have smaller work disincentives.

Columns (4) and (5) in Table 1 show estimates obtained by adding splines to g at the 25th, 50th, and 75th percentile of the predicted participation rate distribution. These added terms are insignificant, indicating that, at least at this crude level, there are no major nonlinearities beyond that implied by the linear probability term.

Nonlinearities in X and Z. The strong evidence for heterogeneity by observables found in the results thus far should raise concerns about the linear-index assumptions in $X\beta$, $X\lambda$, and $X\delta$ made in the model above and in most other research. If labor supply responses to welfare differ by X in quite different ways, as demonstrated above, there is no reason to expect that heterogeneity to be scaled by a summary index of X, at least in the absence of a specific theory in which that index proxies some concrete conceptual quantity. Just as the average labor supply disincentives have been seen to be distorted by inadequate controls for the distribution of X, nonlinearities and interactions which may exist may distort the results yielded by a linear index model. Thus it is important to determine how relaxing these linearity assumptions would affect the estimation and interpretation of the labor supply disincentives of welfare participation.

Linearity in the $Z\eta$ index concerns a more conventional issue of whether increasing the number of instruments by introducing nonlinearities in Z may affect estimated treatment effects. Here it is the distribution of unobservables that affect treatment responses that matters, and whether a linear index in Z adequately captures that distribution. A possibly more important issue is the separability of the X and Z indices in the participation equation, implying that how the Z index sweeps out the distribution of labor supply response unobservables is the same across values of X. Once again, the above finding of heterogeneity in response by X should lead to suspicions of this assumption.

In this section, we consider how flexible parametric estimation of the X and Z indices in the model affects the estimates of labor supply responses. We return to the general model in (29)-(30) and consider nonparametric estimation of the functions $g(F,X)$, $h(X)$, and $F(Z,X)$. We shall represent each function with a series approximation and will use linear splines to form the bases of the series.[14] There is a large literature on varieties of spline approximations (de Boor, 2001) but all have the advantage of fitting conveniently into the regression framework used here and permit an easy parametric interpretation as well. Linear splines also allow a relatively straightforward means to control the degree of approximation, which is necessary given the usual limitations imposed by the curse of dimensionality for our sample size. Splines also allow zero regions, unlike polynomial approximations, and are more well-behaved than polynomials. We shall use the general approximation form for an arbitrary function f(x):

$$f(x) = \sum_{k=1}^{K} \theta_{kK} s_{kK} \tag{35}$$

where x is a vector of variables, K is the number of functions in the approximation, s_{kK} is the k^{th} spline function in the series of order K, and θ_{kK} is its coefficient.

To implement the approximation, we need a method of choosing the s_{kK}, a method of adding new s_{kK} as K expands, and a criterion for choosing K. For the first two, we adopt the method of multivariate adaptive splines proposed by Friedman (1991); see also Hastie et al., 2001, for a discussion. In this method, one begins by prespecifying knots for each of the variables in the vector x, defining the s_{kK} as splines corresponding to each variable or as a product of splines for different variables, and setting K equal to the sum of these terms. The core

[14] See Andrews (1991) and Newey (1997) for consistency proofs and convergence rates for series and spline estimators of conditional mean functions. Consistency generally requires, roughly speaking, that n go to infinity faster than the number of approximating basis functions in the series. With special exceptions, convergence is slower than root-n, as is typical for nonparametric estimators.

set of basis functions we use for the X, Z, and F functions are given in Table 2.[15] We specify an especially fine set of splines for Z, given the importance of these variables for identification of the response to treatment; a somewhat coarse set for the continuous variables in X; and a still coarser set for F, both to increase stability and identification of its effects and in light of the results above indicating little higher order nonlinearities in its effect.

In the method proposed by Friedman, one begins by estimating the model with this set of splines or some subset of them. The order of the series is then increased by then multiplying each of the functions in the model by each of the basis functions and adding, in sequence, each of these new product variables to the model; the model is picked which meets some model selection criterion. The added variable, which is a product of one of the splines in the basic set and one of those initially in the model, is then added to the basis functions, now of order K+1. The process is then repeated by multiplying each of the splines in the basic set by each of the functions in the new model, again choosing the best model according to a model selection criterion, and adding that new spline product to the basis function. The process proceeds until an optimality criterion is met.[16]

The advantages of this method of approximation are threefold. First, one can set the initial set of variables according to priors and theory, and one can force any variable or function into the initial set. Thus, for example, we will always begin with the linear specification we have

---

[15] In the method proposed by Friedman (1991), a separate knot is specified for every unique value of every variable x_1 in the data. We use the much simpler procedure of specifying a sufficiently large number of percentile points for each variable to capture the degree of nonlinearity we expect to find in the data, on a priori grounds.

[16] Friedman proposes, and we follow him, in excluding products of the same initial basis function to avoid literal polynomial approximation.

already estimated (continuous variables are just splines with a knot at zero, and discrete variables will be separated completely in any case) because this is the conventional regression specification. Second, the method expands outward rather than inward, permitting one to begin with a modest estimation problem which grows to a larger one only as necessary. Third, it is computationally straightforward although computationally intensive if the order of approximation grows large.

For a model selection and optimality criterion, we use the generalized cross validation function (Craven and Wahba, 1979; Hastie et al., 2001) defined by:

$$GCV = \sum_i \frac{[y_i - \hat{f}(x_i)]^2}{1 - K/n} \qquad (36)$$

where n is the sample size and y_i is the dependent variable (either y or P, in our case). For linear models with simple splines, the GCV is equivalent to the traditional leave-one-out cross-validation (CV) measure. For our model, it is only an approximation to the CV but has the same penalty for increasing K. Preliminary Monte Carlos on the model here indicated that the GCV attains its minimum at the same order K as the CV in most cases. We therefore use the GCV because of its computational ease.

Table 3 shows several results for preliminary models estimated to date. The first column shows the estimates of the g function in the y equation after the above method was applied to the Z variables in the P equation alone, leaving the X vector in that equation in linear form and separable from Z. The GCV was minimized after adding 35 terms, leading to 38 functions of Z including the initial set of three continuous instruments. The F-statistic for the 38 functions falls slightly to 6.90. While the coefficients in the λ vector change slightly, there is a major change in the coefficients on F, with both rising back to levels significant in magnitude and statistical significance. The influence of unobservables in affecting labor supply responses is now closer to that of the observables, and again implying that those with greater labor supply responses enter the program first. Thus nonlinearity of the Z function matters, indirectly implying that different ranges of Z sweep out different types of labor supply unobservables, which is entirely possible given the institutional nature of the non-financial barriers to participation used by the states.[17]

The second two sets of results in Table 3 are generated by applying the spline expansion method jointly to the X and Z variables in the P equation alone. At the time of this writing, 63 terms had been added, leading to a total of 73 variables in the P equation given that 10 were in the initial specification, but the GCV was still falling. Table 3 shows the results of estimating this specification jointly with the y equation.[18] In the middle column, $X\beta$ and $X\lambda$ in the y equation are kept in linear index form. This specification has the drawback of allowing the interactions between different X basis functions in the P equation to identify the effect of welfare participation, which is not desirable, but the results are shown in any case for their possible interest. While the constant term is still negative and significant, the coefficient on linear F disappears. In the last, more interesting, column, the X specification estimated in the P equation is applied exactly to the y equation in the place of both $X\beta$ and $X\lambda$. No further iterations were

---

[17] The traditional CV measure, as well as those related to it such as the GCV, tends to lead to overfitting of the model (Shao, 1993). In future work we will test the sensitivity of the results to reductions in the order of the approximation to levels below that at which the GCV is minimized.

[18] The F-statistic for the Z and X-Z interactions rises to 18.0.

allowed which would fit those functions better in the y equation. The coefficients on the many X interactions are shown in an Appendix available upon request, but the table shows that the coefficients on the F terms return to their signs and significances that were in the first column. Thus all indications are that X interactions do not matter a great deal to the estimation of unobservable components of the labor supply response, given that nonlinearities in the instruments are properly accounted for.

IV. Conclusions

The results in this paper suggest that heterogeneity in the labor supply response to transfer programs is heterogeneous and that those on the margin of participation generally have lower labor supply responses than inframarginal individuals who have entered the program at an earlier stage. This finding significantly modifies the findings from the homogeneous effects models estimated in the literature.

However, the clear majority of the heterogeneity lies in observables, which, after conditioning, leave much less heterogeneity in the unobservables that vary with participation. This suggests that applications of these models should more aggressively consider interactions of the conditioning variables with the participation indicator. Not doing so does not only lead to a failure to capture subgroup effects and a misinterpretation of heterogeneity responses as resulting from unobservables, it also results in a misestimation of the average response, at least as indicated by the results here. The importance of interaction with observables in treatment effects is not dissimilar those captured by matching techniques, but the method of interaction here is quite different, not requiring, for example, either the deletion of cases and consequent inefficiencies in matching methods or the balancing property.

Nonlinearities in the instruments, captured here by a flexible parametric specification with regression splines, appear to be important, but this could vary with the application. Nonlinearities in the conditioning variables, on the other hand, appear to be less important than nonlinearities in the instruments. However, this, again, could vary with the application and the sensitivity of the results to some specifications of those nonlinearities suggests that this should be an issue which should be examined in other applications.

Appendix A

Means and Data Sources

The means and standard deviations of the variables used in the analysis are shown in Table A-1. The sources of the state-level variables are as follows. The AFDC guarantee is the monthly maximum amount paid for a family of four in the state, and is obtained from unpublished data provided to the author by the U.S. Department of Health and Human Services for all three years 1989-1991. The state "negative case action error rate," the rate of error per applicant resulting in an incorrect denial of eligibility, is taken from the federal government's quality control program for AFDC and was obtained for 1991 from U.S. House of Representatives (1994, Table 10-39). The state percent of applicants denied for failure to comply with procedural requirements was obtained from the 1989, 1990, and 1991 issues of Quarterly Public Assistance Statistics published quarterly by the U.S. Department of Health and Human Services. The data on state administrative expenditures per case was obtained for 1989, 1990, and 1991 from http://www.acf.dhhs.gov/programs/opre/timetren/index.htm.

Table A-1

Means and Standard Deviations of the Variables Used in the Analysis

Variable Name	Variable Definition	Total sample	Welfare=1	Welfare=0
Hours	Average hours of week in the month prior to survey	21.9 (19.4)	3.8 (.79)	30.1 (17.0)
Welfare	Dummy variable equal to 1 if individual was on AFDC anytime in the month prior to survey	.31 (.46)	--	--
Age	Age in years at survey date divided by 10	3.2 (.88)	3.1 (.79)	3.3 (.91)
Education	Years of education at survey date	10.9 (2.0)	10.4 (2.1)	11.1 (1.9)
Family	Number of individuals in the family at the survey date	3.3 (1.4)	3.4 (1.4)	3.2 (1.4)
Family6	Number of children less than 6 in the family at the survey date	.72 (.89)	1.1 (1.0)	.55 (.76)
Black	Dummy variable equal to 1 if respondent is black	.33 (.47)	.44 (.50)	.28 (.45)

Table A-1 (continued)

Variable Name	Variable Definition	Total sample	Welfare=1	Welfare=0
Nonlabor	Nontransfer nonlabor income in the month prior to survey divided by 100	1.12 (2.07)	.40 (1.16)	1.45 (2.29)
G	State monthly AFDC guarantee for a family of four divided by 100	4.69 (1.97)	4.91 (2.01)	4.59 (1.94)
Admin	AFDC Administrative expenditures per case in the state averaged over 1989, 1990, and 1991, divided by 1000	.044 (.021)	.045 (.022)	.043 (.021)
Pctdenied	Fraction of applications denied for failure to meet procedure requirements in the state averaged over 1989, 1990, and 1991	.59 (.17)	.58 (.17)	.59 (.17)
Eligerror	Federally-audited percent error rate made by the state in 1991 in calculating eligibility	2.25 (2.26)	2.05 (1.87)	2.34 (2.40)
Sample size	--	5,722	1,783	3,939

Notes:

Standard deviations in parentheses
All dollar-valued variables are deflated by a 1990 price index using the GDP-based personal consumption expenditure deflator.

V. References

Andrews, D. W. K. 1991. "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models." Econometrica 59 (March): 307-345.

Björklund, A. and R. Moffitt. 1987. "The Estimation of Wage and Welfare Gains in Self-Selection Models." Review of Economics and Statistics 69: 42-49.

Blundell, R. and T. MaCurdy. 1999. "Labor Supply: A Review of Alternative Approaches" " In Handbook of Labor Economics, Vol. 3A, eds. O. Ashenfelter and D. Card. Amsterdam and New York: Elsevier.

Burtless, G. and J. Hausman. 1978. "The Effect of Taxaton on Labor Supply: The Gary Income Maintenance Experiment." Journal of Political Economy 86: 1103-1130.

Carneiro, Pedro; James J. Heckman; and Edward Vytlacil. 2003. "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education." Mimeo.

Craven, P. and G. Wahba. 1979. "Smoothing Noisy Data with Spline Functions." Numerische Mathematik 31: 377-403.

Danziger, S.; R. Haveman; and R. Plotnick. 1981. "How Income Transfers Affect Work, Savings, and the Income Distribution: A Critical Review " Journal of Economic Literature 19: 975-1028.

de Boor, C. 2001. A Practical Guide to Splines. New York: Springer.

Friedman, J. 1991. "Multivariate Adaptive Regression Splines." Annals of Statistics 19 (March): 1-67.

Hastie, T.; R. Tibshirani; and J. Friedman. 2001. The Elements of Statistical Learning. New York: Springer.

Hausman, J. 1981. "Labor Supply." In How Taxes Affect Economic Behavior, eds. H. Aaron and J. Pechman. Washington: Brookings Institution.

Hausman, J. 1985. "Taxes and Labor Supply." In Handbook of Public Economics Vol.I, eds. A. Auerbach and M. Feldstein. Amsterdam: North-Holland.

Heckman, J. and R. Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions." In Longitudinal Analysis of Labor Market Data, eds J. Heckman and B. Singer. Cambridge University Press..

Heckman, J. and E. Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." Proceedings of the National Academy of Sciences 96 (April): 4730-4734.

Heckman, J. and E. Vytlacil. 2000. "Causal Parameters, Structural Equations, Treatment Effects and Randomized Evaluations of Social Programs." Mimeo. UC and Stanford.

Heckman, J. and E. Vytlacil. 2001. "Policy-Relevant Treatment Effects." American Economic Review 91 (May): 107-111.

Heckman, J.; S. Urzua; and E. Vytlacil. 2004. "Understanding Instrumental Variables in Models with Essential Heterogeneity." Mimeo.

Ichimura, H. and C. Taber. 2000. "Direct Estimation of Policy Impacts." Mimeo.

Imbens, G. and J. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." Econometrica 62: 467-76.

Moffitt, R. 1983. "An Economic Model of Welfare Stigma." American Economic Review 73 (December: 1023-1035.

Moffitt, R. 1992. "Incentive Effects of the U.S. Welfare System: A Review." Journal of Economic Literature 30 (March): 1-61.

Moffitt, R. 2002. "Welfare Programs and Labor Supply." In Handbook of Public Economics, Vol.4, eds. A. Auerbach and M. Friedman. Chicago: University of Chicago Press.

Moffitt, R. 2003. "The Temporary Assistance for Needy Families Program." In Means-Tested Transfer Programs in the United States, ed. R. Moffitt. Chicago: University of Chicago Press.

Shao, J. 1993. "Linear Model Selection by Cross-Validation." Journal of the American Statistical Association. 88 (June): 486-494.

Newey, W. 1997. "Convergence Rates and Asymptotic Normality for Series Estimators." Journal of Econometrics 79(1997): 147-168.

U.S. House of Representatives. 1994. Background Material and Data on Programs Within the Jurisdiction of the Committee on Ways and Means. Washington: Government Printing Office.

Table 1

Results of the Initial Estimation

	(1)	(2)	(3)	(4)	(5)
γ					
Constant	-30.9* (4.0)	-77.6* (8.0)	-7.1 (21.5)	-41.4 (37.0)	6.38 (78.7)
F	--	43.0 (6.0)	4.5 (19.2)	67.0 (68.3)	-94.9 (255.4)
Max(0,F-.25F)	--	--	--	--	113.4 (197.6)
Max(0,F-.50F)	--	--	--	-30.6 (51.9)	-11.9 (53.0)
Max(0,F-.75F)	--	--	--	--	18.0 (28.2)
λ					
Education	--	--	-3.16* (.91)	-1.56* (.94)	-1.86 (1.03)
Age	--	--	-1.91 (1.60)	2.41 (1.64)	2.42 (1.64)
Black	--	--	6.98* (3.28)	2.38 (3.48)	3.50 (3.72)
Family6	--	--	.74 (2.42)	-2.75 (2.63)	-1.86 (3.02)
Family	--	--	.98 (.92)	1.06 (1.12)	1.03 (.94)
Nonlabor	--	--	4.16* (1.62)	6.80* (1.93)	5.82* (2.26)
G			-.53 (.62)	-1.14* (.66)	-.98 (.69)

Table 1 (continued)

	(1)	(2)	(3)	(4)	(5)
<u>β</u>					
Constant	21.6* (3.2)	38.8* (5.2)	5.45 (6.89)	7.25 (7.53)	2.78 (10.4)
Education	.91* (.19)	.16* (.31)	2.60 (.43)	2.45* (.42)	2.60* (.48)
Age	1.68* (.31)	1.25* (.41)	1.20* (.53)	1.29* (.53)	1.30* (.54)
Black	-1.86* (.62)	-.14 (.84)	-3.93* (1.28)	-3.72* (1.31)	-4.23* (1.54)
Family6	.06 (.61)	.41 (.70)	-1.29 (1.14)	-1.61 (1.15)	-2.07 (1.41)
Family	-.71* (.19)	-.71* (.23)	-1.06* (.34)	-.98* (.34)	-.98* (.34)
Nonlabor	-.92* (.22)	-1.92* (.27)	-.96* (.31)	-1.02* (.44)	-.67 (.66)
G	-.27* (.14)	-.06 (.17)	-.18 (.25)	-.19 (.25)	-.27 (.27)
<u>δ</u>					
Constant	.62* (.19)	1.06* (.18)	.80* (.17)	.72* (.17)	71* (.17)
Education	-.11* (.01)	-.16* (.01)	-.12* (.01)	-.12* (.01)	-.12* (.01)
Age	-.06* (.03)	-.06* (.03)	-.06* (.02)	-.05* (.02)	-.05* (.02)
Black	.29* (.05)	.31* (.04)	.34* (.04)	.36* (.04)	.36* (.04)

Table 1 (continued)

	(1)	(2)	(3)	(4)	(5)
Family6	.41*	.34*	.38*	.38*	.39*
	(.03)	(.03)	(.02)	(.02)	(.02)
Family	-.04*	-.01	-.03*	-.02*	-.02
	(.02)	(.02)	(.01)	(.01)	(.01)
Nonlabor	-.25*	-.21*	-.20*	-.21*	-.21*
	(.02)	(.02)	(.02)	(.02)	(.02)
G	.05*	.05*	.06*	.07*	.07*
	(.01)	(.01)	(.01)	(.01)	(.02)
$\underline{\eta}$					
Admin	.90	.22	-.41	-2.16*	-1.85*
	(1.19)	(.92)	(1.08)	(1.04)	(1.04)
Pctdenied	-.14	-.08	-.24*	-.31*	-.32*
	(.14)	(.10)	(.12)	(.11)	(.11)
Eligerror	-.03*	-.02*	-.02	-.01	-.01
	(.01)	(.01)	(.01)	(.01)	(.01)

Notes:
Standard errors in parentheses
*: significant at the 10 percent level

Table 2

Core Set of Basis Function for Spline Approximations

X	Z	F
1 (Constant)**	1 (Constant)*	1 (Constant)
Age*	Admin*	F*
Max(0, Age-Age_{25})	Max(0, Admin-$Admin_p$)	Max(0, F-F_{25})
Max(0, Age-Age_{50})	[p=10,20,30,40,50,6,70,80,90]	Max(0, F-F_{50})
Max(0, Age-Age_{75})	Pctdenied*	Max(0, F-F_{75})
Education12*	Max(0, Pctdenied-$Pctdenied_p$)	
Family>3	[p=10,20,30,40,50,60,70,80,90]	
Family6*	Eligerror*	
Black	Max(0, Eligerror-$Eligerror_p$)	
Nonlabor*	[p=10,20,30,40,50,60,70,80,90]	
Max(0, Nonlabor-$Nonlabor_{25}$)		
Max(0, Nonlabor-$Nonlabor_{50}$)		
Max(0, Nonlabor-$Nonlabor_{75}$)		
G*		
Max(0, G-G_{25})		
Max(0, G-G_{50})		
Max(0, G-G_{75})		

Notes:

*: Forced into the initial specification
**: Not included in the initial specification for Xλ but allowed for create new spline interactions
Subscripts denote percentile point of the variable in question
Education12 = dummy =1 if Education=12, 0 otherwise
Family>3 = dummy=1 if family is greater than 3, 0 otherwise

Table 3

Parameter Estimates of g Function for Spline Specifications

	Z splines (P eqn only)	X-Z splines	
		P eqn only	y and P eqns[a]
Constant	-27.5*	-20.7*	-20.4*
	(12.4)	(5.2)	(9.37)
F	25.6*	-.41	9.46*
	(8.7)	(4.03)	(5.70)
$\underline{\lambda}$			
Education	-2.24*	--	--
	(.58)		
Education12	--	-7.60*	--
		(1.86)	
Age	.79	-2.29*	--
	(1.39)	(1.07)	
Black	4.59*	6.58*	--
	(2.29)	(1.64)	
Family6	-1.72	.06	--
	(1.42)	(.87)	
Family	1.69*	--	--
	(.81)		
Family>3	--	5.29*	--
		(1.57)	
Nonlabor	5.05*	2.08*	--
	(1.16)	(.66)	
G	-1.09*	-.14	--
	(.51)	(.39)	

Notes:
Standard errors in parentheses
*: significant at the 10% level
[a] Coefficients on X splines shown in the Appendix

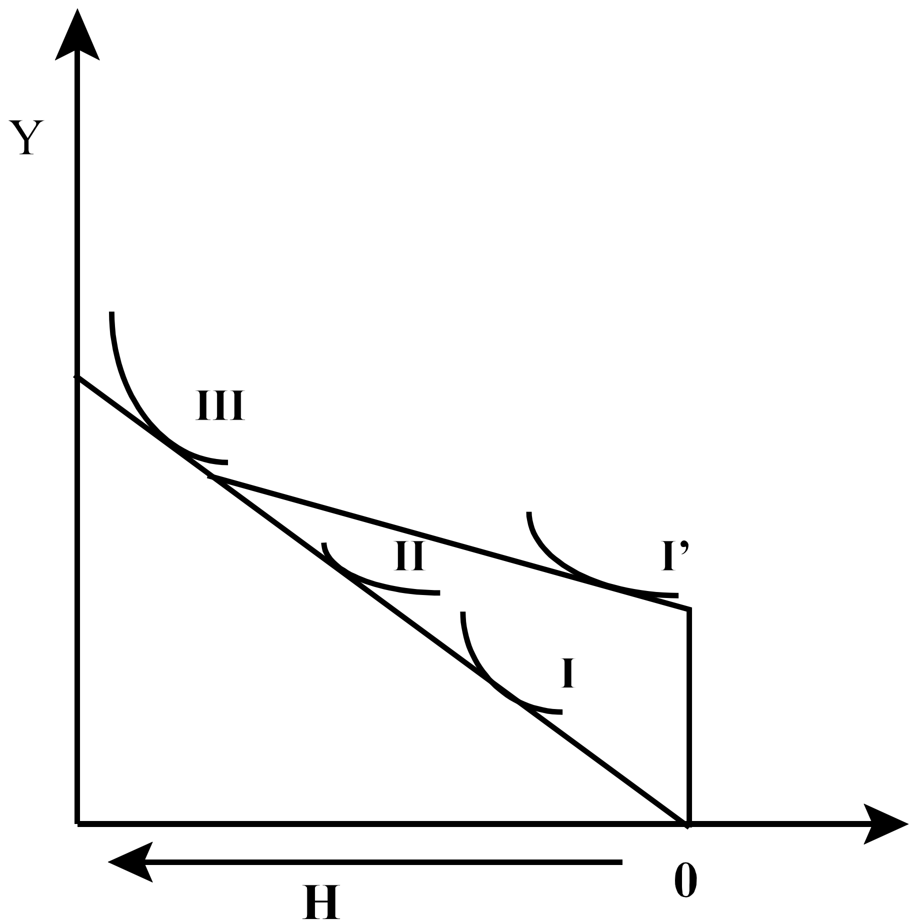

Figure 1

Figure 2: Histogram of Predicted Participation Rates

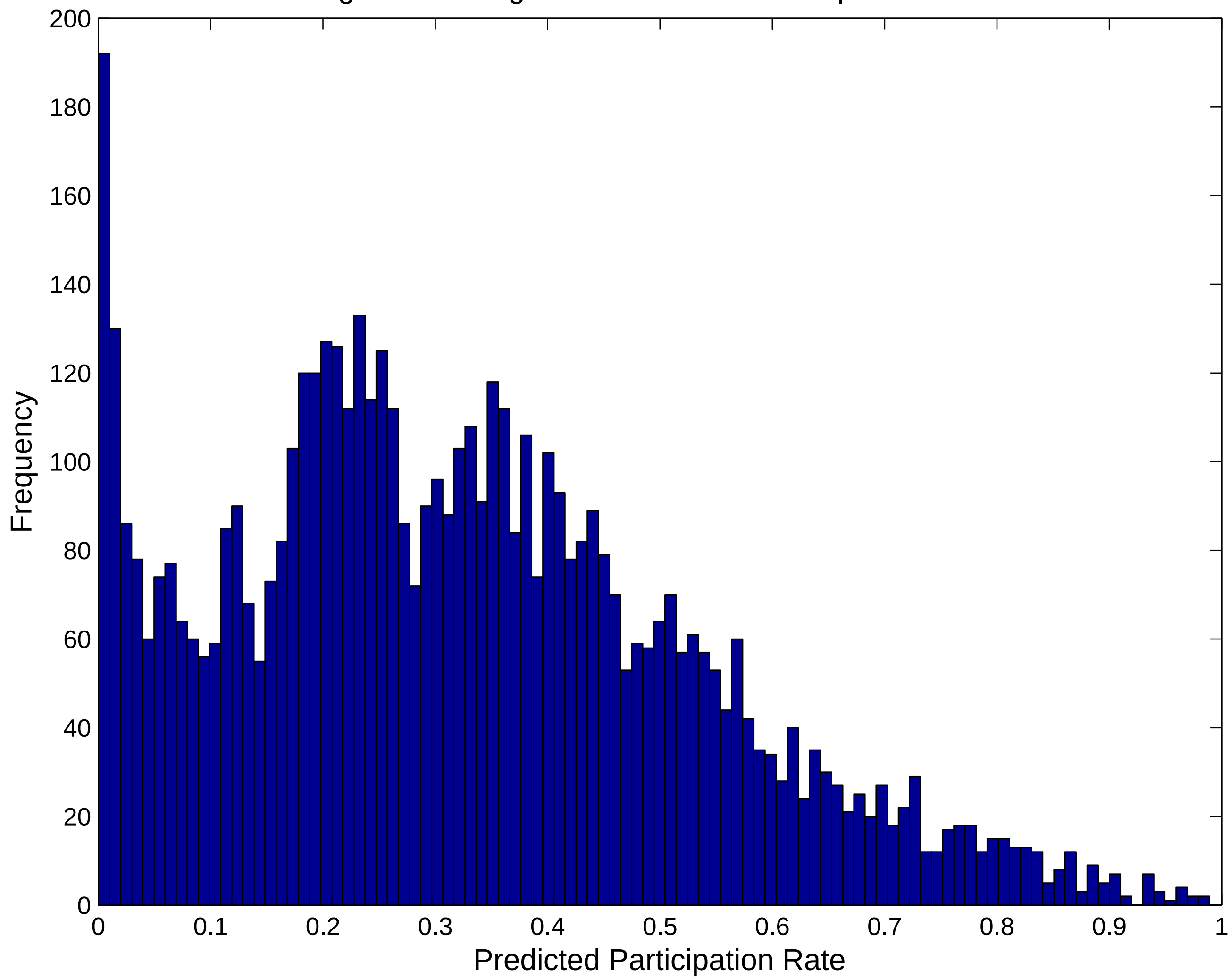

Figure 3. Marginal Treatment Effect (MTE) and 95 Percent Confidence Interval.

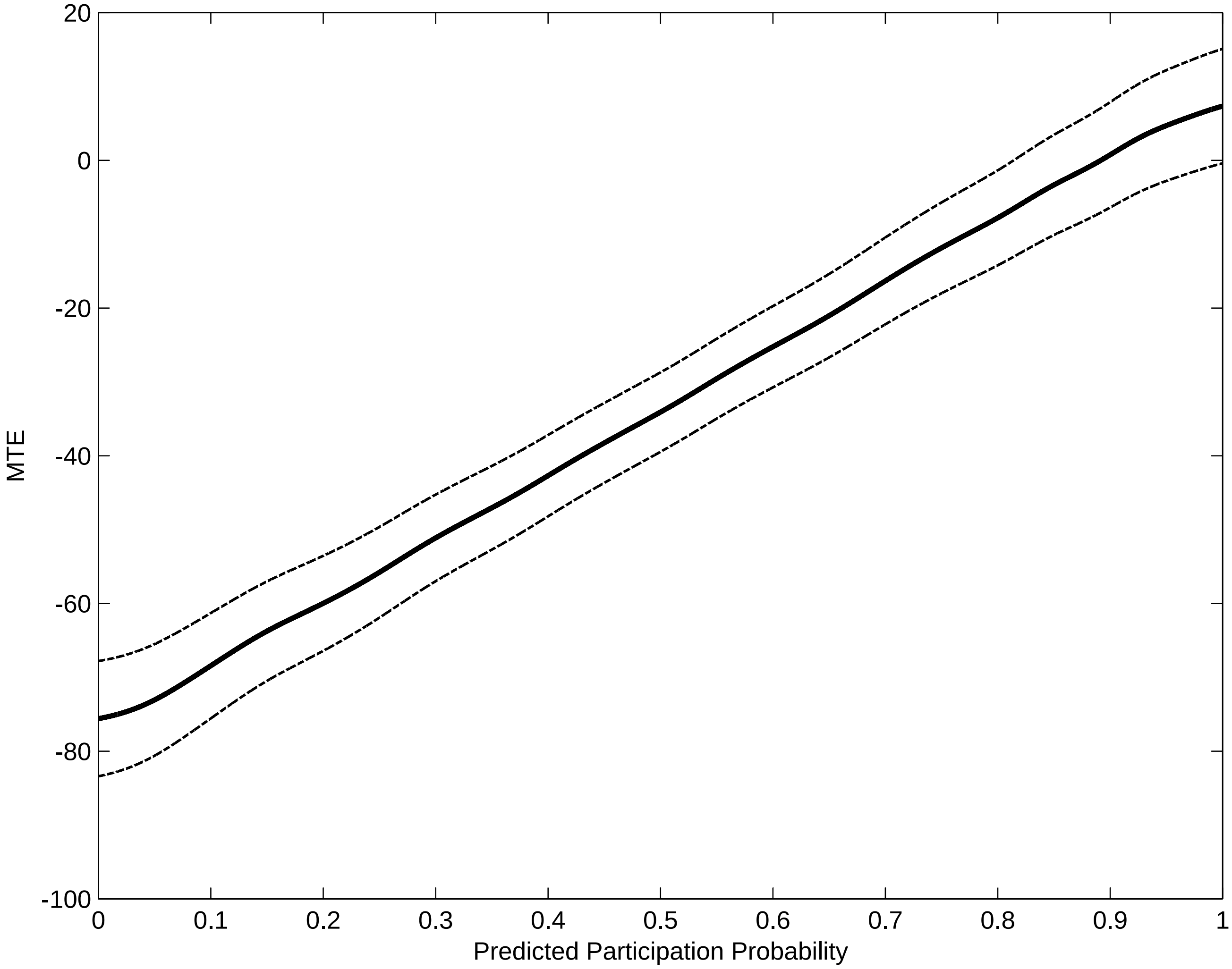

Figure 4. Marginal Treatment Effect (MTE) and 95 Percent Confidence Interval, Excluding X Variation