Preliminary and Incomplete

The Marginal Labor Supply Disincentives of Welfare Reforms*

Robert Moffitt†
Department of Economics, Johns Hopkins University

October 22, 2018

ABSTRACT: Existing research on the static effects of the manipulation of welfare program benefit parameters on supply has allowed only restrictive forms of heterogeneity in preferences. Yet preference heterogeneity implies that the marginal effects of welfare reforms on labor supply may differ in different time periods with different populations and which sweep out different portions of the marginal distributions of preferences. A new examination of the heavily studied AFDC program shows that changes in its tax rates and guarantees in 1967, the 1970s, 1981, 1996, and 2010 had different marginal effects on labor supply because of differing heterogeneity of marginal participants in each case. The models used to estimate these effects allow for a nonparametric specification of how changes in welfare program participation affect labor supply on the margin. Estimates of the model using a form of local instrumental variables show that the effects of each of the historical reforms on labor supply differed because of differences in the composition of who was on the program and who was not, and who the marginal person was, in each period.

KEYWORDS: Welfare, Labor Supply, Marginal Treatment Effects
JEL CLASSIFICATION: I3,J2,C21

---

*The author would like to thank Marc Chan, Kai Liu, Shaiza Qayyum, Lu Xu, and Sue Bahk for research assistance as well as the participants of a large number of conferences and departmental seminars for comments. Joseph Altonji, Richard Blundell, Christopher Flinn, Eric French, Michael Keane, Christopher Taber, and James Ziliak provided specific helpful comments and feedback. Research support from the National Institutes of Health is gratefully acknowledged.

†moffitt@jhu.edu.

The classic form of a welfare program for a low-income population is that represented by a negative income tax, with a guaranteed minimum cash payment for those with no private income and with a positive marginal benefit-reduction rate, or tax rate, applied to increases in earnings. In the U.S., the only major program has taken this classic shape was the Aid to Families with Dependent Children (AFDC) program, which took that shape from its formation in 1935 to the early 1990s, when its structure was changed. The program has undergone significant fluctuations in the size of its caseload, growing from the 1960s to the early 1970s, flattening out in the 1970s, then growing again in the 1980s before declining after 1996. These fluctuations were in large part a result of changes in its guaranteed minimum payments and its tax rateps: the tax rate was reduced from 100 percent to 67 percent in 1967; real guarantees fell over the 1970s; the tax rate was brought back to 100 percent in 1981; and tax rates were then reduced again in 1996. Real guarantees have steadily fallen since that time. These reforms and their effects on labor supply have been evaluated exhaustively in the literature (see literature reviews by Moffitt (1992),Blundell and Macurdy (1999),Moffitt (2003),Ziliak (2016), and Chan and Moffitt (2018)).

This paper revisits this literature, arguing that the empirical models used to evaluate these reforms have been excessively restrictive in the representation of unobserved heterogeneity in the eligible population (i.e., heterogeneity conditional on the observables). By definition, the effect of any reform on labor supply depends on the labor supply responses of inframarginal individuals (i.e., those who remain on the program both before and after the reform) but also on the labor supply responses of marginal individuals who either join or leave the program in response to the reform. With sufficient heterogeneity of preferences, these two responses are not the same, but the existing literature on the effects of AFDC reforms on labor supply has almost entirely assumed they are equivalent. Yet, if there have been changes over time in terms of who is on the margin and who is not for that program, the effects of the reforms on labor supply may very well have varied. This paper provides evidence that, in fact, who was on the margin was quite different in 1967 and in each succeeding year when reforms took place and that, consequently, the marginal effects of the reforms on labor supply varied significantly. This finding should obviously be important for any future reforms as well.

The first section of the paper lays out the familiar static labor supply model in the presence of a classic welfare program but adds a general form of preference heterogeneity to that model. The model is then used to analyze the marginal labor supply response of a expansionary reform and shows how it will vary depending on the initial distribution of preferences of those on the program as well as the distribution of preferences of those brought into the program. The model also shows that that the marginal labor supply responses from an expansionary reform can be greater or smaller than the responses of those initially on the program, which also implies that a program that is continually expanded can have marginal labor supply responses that grow, fall, or remain the same.

The second section presents a reduced form model designed for the estimation of marginal labor supply responses. Those responses can be identified over the range of participation rates supported by the instruments and their support. The model draws directly on the literature on the reduced form estimation of marginal treatment effects (Heckman and Robb (1985),Bjorklund and Moffitt (1987), and many subsequent papers) and applies a modified form of the local instrumental variable (LIV) method of Heckman and Vytlacil (1999, 2005) (see also Heckman et al. (2006)). The parameters of the reduced form model are related directly to those of the structural model.

The third section uses cross-section data from the late 1980s and the early 1990s (the last time the AFDC program had its classic shape) to estimate marginal labor supply responses to variation in participation rates nonparametrically using sieve methods. The instruments are based on variables representing fixed costs of participation which do not affect labor supply reponses conditional on participation. Over the support of the propensity score provided by the instruments, the results show considerable heterogeneity in marginal labor effects, which follow a quadratic pattern with program expansion–growing (i.e., becoming increasingly negative) with initial expansions but eventually declining as the program grows beyond a certain threshold.

The fourth section uses the estimated model to show how the 1967 and later reforms of the program affected labor supply and how the labor supply response of the marginal individual varied over time. Although the model estimated in the previous section was not fully structural, because it was formulated precisely as the reduced from of a structural model, estimates of the marginal labor supply responses to the specific manipulations of minimum guarantees and tax rates that occurred in each historical period can nevertheless be obtained. The estimates show that the marginal labor supply effects of reforms in 1967, the 1970s, 1981, 1996, and 2010 were all different, and that the marginal labor supply response to an expansion of participation in the period closes to the present time (2010) is smaller than in any year since 1967.

The implication of the analysis is that policy forecasts of every incremental reform must be based on the nature of the specific population on the program at the time of the reform and on the nature of the reform (since which parameter is changed affects who is brought in or who leaves). A general model of how responses depend on population heterogeneity is needed to be able to make reliable forecasts for policy.

The methodological approach taken here should be also applicable to models of dynamic labor supply responses to program changes (Blundell et al. (2016)) and to the estimation of more complex reforms than simple manipulations of guarantees and tax rates (e.g., the imposition of time limits as in Chan (2013)). The approach should also be applicable to the estimation of behavioral responses to programs other than the AFDC program, as well as to other studies of policy impacts where heterogeneity is likely to be important.

1 Adding Heterogeneity to the Canonical Static Labor Supply Model of Transfers

The canonical static model of the labor supply response to transfers (Moffitt (1983)) assumes utility to be

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

where H_i is hours of work for individual i, Y_i is disposable income, P_i is a program participation indicator, θ_i is a vector of labor supply preference parameters, and ϕ_i is a scalar representing fixed costs of participation in utility units. The presence of P_i allows for the presence of fixed costs of participation–in money, time, or utility, with the exact type unspecified. Fixed costs are required to fit the data because many individuals who are eligible for transfer programs do not participate in them, as revealed by the presence in the data of nonparticipating eligibles for virtually all programs. The presence of fixed costs also makes the participation decision separable from the H decision.[1] The separability of P_i from the U function is for analytic convenience and is not required for any of the following results.

The individual faces an hourly wage rate W_i and has available exogenous non-transfer nonlabor income N_i. The welfare benefit formula is $B_i = G - tW_iH_i - rN_i$ (assuming, for the moment, that the parameters G, t and r do not vary by i) and hence the budget constraint is

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

[1]The existence of a cost function also opens an avenue for instruments that affect fixed costs but not hours of work directly, the same role that cost functions often play in models of schooling and human capital (see, e.g., p.674 of Heckman and Vytlacil (2005)).

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

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

$$P_i^* = V[W_i(1 - t), G + N_i(1 - r); \theta_i] - V[W_i, N_i; \theta_i] - \phi_i \tag{4}$$

$$P_i = 1(P_i^* \geq 0) \tag{5}$$

where V is the indirect utility function and $1(\cdot)$ 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), below breakeven but off the program (II), and below breakeven and on the program (I'; I is the pre-program location for this individual).

The response to the program for individual i is

$$\triangle_i\,(\theta_i \mid W_i, N_i, G, t, r) = H[W_i(1 - t), G + N_i(1 - r); \theta_i] - H[W_i, N_i; \theta_i] \tag{6}$$

which is a heterogeneous response if θ_i varies with i. The response $\triangle_i$ includes both responses from below breakeven and above breakeven. Individual values of $\triangle_i$ will never be identified by the data, but the mean of those values over some populations or subpopulations can be. To define the subpopulation of participants and their distributions of θ_i, first let S_θ and S_ϕ represent the unconditional supports of the two parameters, so that the participation rate in the population is

$$P = E(P_i \mid W_i, N_i, G, t, r)$$
$$= \int_{S_\phi}\int_{S_\theta} 1\{V[W_i(1-t), G + N_i(1-r); \theta_i] - V[W_i, N_i; \theta_i] - \phi_i\} dG(\theta_i, \phi_i) \tag{7}$$

where $G(\theta_i, \phi_i)$ is the joint c.d.f. of the two heterogeneity components. Then define θ_D and ϕ_D as the values of those parameters for the marginal person:

$$0 = V[W_i(1-t), G + N_i(1-r); \theta_D] - V[W_i, N_i; \theta_D] - \phi_D \tag{8}$$

and define the set $S_{\theta\phi}$ as the set of parameters in regions demarcated by the θ_D and ϕ_D boundary points which generate $P = 1$.[2] Then the mean labor supply effect for those participating in the program is

$$\widetilde{\triangle}_{P_i=1} = E(\triangle_i \mid W_i, N_i, G, t, r, P_i = 1)$$
$$= [\int_{S_{\theta\phi}}\int \triangle_i(\theta_i \mid W_i, N_i, G, t, r) dG(\theta_i, \phi_i)]/P \tag{9}$$

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

$$\widetilde{\triangle} = E(\triangle_i P_i \mid W_i, N_i, G, t, r)$$
$$= \int_{S_{\theta\phi}}\int \triangle_i(\theta_i \mid W_i, N_i, G, t, r) dG(\theta_i, \phi_i) \tag{10}$$

The marginal response to a change in program participation (i.e., the marginal treatment effect, or MTE), which is the mean $\triangle$ of those who change participation, is $\partial\widetilde{\triangle}/\partial P$.

The values of the response quantities $\triangle_i$, $\widetilde{\triangle}$, $\widetilde{\triangle}_{P_i=1}$, and $\partial\widetilde{\triangle}/\partial P$ must all be

[2]These boundary parameters are a generalization of the U_D parameter of Heckman and Vytlacil (2005).

nonpositive according to theory, but their magnitudes are affected by who participates and who does not. Participation decisions are a function of the budget constraint parameters and the joint distribution of θ_i and ϕ_i, as is clear from equation (4). The mean labor supply response of participants, $\widetilde{\triangle}_{P_i=1}$ is therefore also a function of these same variables and distribution, as is the marginal labor supply response. A change in a welfare program parameter like G or t will not only change the mean labor supply response of those already on the program but will also change the values of θ_D and ϕ_D and hence will alter the set $S_{\theta\phi}$ which determines the values of θ of individuals on the margin of participation. With no restrictions on the shapes of that joint distribution, those marginal values could be greater or smaller than those of initial participants.

A case of particular interest for the empirical work below is a program expansion induced by a reduction in ϕ_i, the fixed cost parameter (an observable proxy for these costs will be used in the empirical work). A reduction in this parameter only shifts who participates in the program and does not shift the values or the latent distribution of the population responses in eqn(6). A reduction in ϕ_i means that individuals with lower values of the non-cost portion of the utility gain of going onto welfare, $dV = V[W_i(1-t), G + N_i(1-r); \theta_i] - V[W_i, N_i]$ will participate. But there is no determinant relationship between the magnitude of this portion of the utility gain and the magnitude of labor supply reductions because, for example, those with smaller utility gains who enter the program may have relatively greater preferences for consumption goods than for leisure than those initially on the program, leading to a smaller labor supply reduction on the margin. Alternatively, marginal participants could have a smaller relative valuation of consumption relative to leisure, giving the opposite result. Thus how marginal labor supply reductions change as participation expands or declines can only be determined empirically.

2 A Reduced Form Model

The model estimated is reduced form but completely consistent with the theoretical model just outlined and hence is capable of capturing marginal labor supply responses appropriately. The model is simply that in eqns(3)-(5). For reduced form estimation, let H_i be a function of P_i as in (3) and assume, for the moment, that all budget constraint and other covariates have been conditioned out. Likewise, condition out all observables in (4) except for an observable proxy for fixed costs, Z_i. An unrestricted reduced-form model of (3)-(4) with full individual heterogeneity is

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

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

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

where β_i and α_i are scalar random parameters and δ_i is a vector of random parameters.[3] All parameters are allowed to be individual-specific and to have some unrestricted joint distribution which is generated by the latent heterogeneity in the structural parameters θ_i and ϕ_i. A separate model therefore exists for each individual i. The function m can likewise be unrestricted and can be saturated if Z_i is assumed to have a multinomial distribution, although we shall discuss restrictions on δ_i below. The object of interest is the distribution of α_i. Selection in this model can occur either on the intercept (β_i) or the slope coefficient (α_i) because both may be related to δ_i and, in fact, the theoretical model implies that they must be because the participation equation contains the parameters of the labor supply function. Assuming that the three parameters in (11)-(13) are mean independent of Z_i, we can condition both equations on it to determine what is identified and estimable:

[3]In the empirical analysis below, corner solutions at $H = 0$ will be allowed in the data but (11)-(13) will continue to represent the reduced form of the model. With corner solutions, equation (4) must be modified to represent the choice of direct utility at $H = 0$ versus utility at $H > 0$. The reduced form of that representation is still of the form in(11)-(13).

$$E(H_i \mid Z_i = z) = E(\beta_i \mid Z_i = z) + E(\alpha_i \mid P_i = 1, Z_i = z) \Pr(P_i = 1 \mid Z_i = z) \tag{14}$$

$$E(P_i \mid Z_i = z) = \Pr[m(z, \delta_i) \geq 0] \tag{15}$$

Identification of $E(\alpha_i \mid P_i = 1, Z_i = z)$ requires that Z_i satisfy two mean independence requirements, one for the intercept and one for the slope coefficient:

$$A1. \;\; E(\beta_i \mid Z_i = z) = \beta \tag{16}$$

$$A2. \;\; E(\alpha_i \mid P_i = 1, Z_i = z) = g[E(P_i \mid Z_i = z)] \tag{17}$$

where g is the effect for those on the program (i.e., the effect of the treatment on the treated) conditional on Z_i, and depends on the shape of the distribution of α_i and how different fractions of participants are selected from different portions of that distribution. While the first assumption is familiar, the second may be less so. The usual assumption in the literature is that the two potential outcomes, β_i and $\beta_i + \alpha_i$, are fully independent of Z_i, which implies that α_i is as well. Eqn (17) is a slightly weaker condition which states that all that is required is that the effect of the treatment on the treated be dependent on Z_i only through the effect of the participation probability (i.e., the propensity score).[4] If this were not so, different values of Z_i would lead to different conditional means of α_i through some other channel, which would rule it out as a valid exclusion restriction. [5]

Inserting the two assumptions into the main model in eqns (14)-(15), and denoting the participation probability as $F(Z_i) = E(P_i \mid Z_i)$, we obtain two estimating equations

$$H_i = \beta + g[F(Z_i)]F(Z_i) + \epsilon_i \tag{18}$$

[4] The terms "propensity score" and "participation probability" are used interchangeably throughout.

[5] The monotonicity condition of Imbens and Angrist (1994) constitutes, in this model, a restriction on δ_i, requiring that the difference between propensity scores at distinct values z and z' be zero or the same sign for all i.

$$P_i = F(Z_i) + \upsilon_i \tag{19}$$

where ϵ_i and ν_i are mean zero and orthogonal to the RHS by construction. No other restriction on these error terms need be made, as this is a reduced form of the model. The first equation just states that the population mean of H_i equals a constant plus the mean response of those in the program times the fraction who are in it. The implication of the model is that preference heterogeneity is detectable by a nonlinearity in the response of the population mean of H_i (taken over participants and nonparticipants) to changes in the participation probability. If responses are homogeneous and hence the same for all members of the population, the function g reduces to a constant and therefore a shift in the fraction on the program has a linear effect on the population mean of H_i. If the responses of those on the marginal vary, however, the response of the population mean of H_i will depart from linearity. This feature of the heterogeneous-response treatment model has been noted by Heckman and Vytlacil (2005) and Heckman et al. (2006), and eqn(18) follows from their work. However, here it will form the basis of the estimation of the model and the conditional mean function in eqn(18) will be estimated directly.[6]

Nonparametric identification of the parameters of the model–β and the function g at every point F–is straightforward and has been extensively discussed in the literature. F is identified at every data point Z_i from the second equation from the mean of P_i at each value of Z_i (apart from sampling error). If there is a value of Z_i in the data for which $F(Z_i) = 0$, then β is identified from the mean of H_i at that point and hence g is identified pointwise at every other value of Z_i and hence F. If no such value is in the data, then g can only be identified subject to a normalization or multiple variables of g can be identified. For example, the LATE of Imbens and Angrist (1994) is identified by the discrete difference in H between two points z_i and z_j divided by the difference in F

[6]Eqn(18) appears in the middle of p.690 of Heckman and Vytlacil (2005). Those authors estimate the MTE by a direct nonparametric computation of the slope of the outcome-propensity-score regression line, termed local instrumental variables. Estimation of eqn(18) by allowing the coefficient on the propensity score to be nonparametric in that score is equivalent.

between those two points. A marginal treatment effect is a continuous version of this and requires some smoothing method across discrete values of Z, and is computed by $\partial H/\partial F = g'(F)F + g(F)$. Here, the g function will be approximated by a nonparametric but continuous function which implicitly means that interpolation between the data points identifies the pointwise derivatives in the MTE.

If we now condition on the budget constraint parameters, this reduced form can be written in terms of the structural functions defined in the last section:

$$\begin{aligned} E(H \mid W_i, N_i, G, t, r) &= E(H \mid W_i, N_i, P_i = 0) + E[\triangle_i \mid W_i, N_i, G, t, r, P_i = 1)]E(P_i \mid W_i, N_i, G, t, r) \\ &= h_0(W_i, N_i) + g(W_i, N_i, G, t, r)F(W_i, N_i, G, t, r) \end{aligned} \tag{20}$$

where $h_0(W_i, N_i)$ is mean labor supply off welfare, $g(W_i, N_i, G, t, r)$ is the mean labor supply response of participants as defined above, and $F(W_i, N_i, G, t, r)$ is the function for the probability of participation. All three functions have the budget constraint parameters as arguments and all can be estimated as approximations in those parameters.

Fully nonparametric estimation of all three functions would make the estimation subject to the curse of dimensionality, so considerable dimension reduction will be achieved by using traditional linear indices for in the observables. The equations that will be estimated are therefore:

$$H_i = X_i^{\beta}\beta + [X_i\lambda + g(F(X_i\eta + \delta Z_i))]F(X_i\eta + \delta Z_i) + \epsilon_i \tag{21}$$

$$P_i = F(X_i\eta + \delta Z_i) + \nu_i \tag{22}$$

where X_i^{β} denotes a vector of exogenous socioeconomic characteristics plus W_i and N_i and X_i denotes a vector which augments X_i^{β} with the welfare-program variables G and t and r. Exogenous characteristics thus linearly affect labor supply off welfare and linearly affect the g and F functions. For g, we will estimate its shape with sieve methods but will assume normality for F. With these two functions specified, we will employ two-step estimation of the model, with a first-stage probit estimation of eqn(22) and second-stage nonlinear least squares estimation of eqn(21) using fitted values of F from the first stage. Consistency of two-step estimation of conditional mean functions is demonstrated in Newey and McFadden (1994). Standard errors are obtained by bootstrapping.

3 Data and Preliminary Results

.The Aid to Families with Dependent Children (AFDC) program is the only major cash welfare program the U.S. has had, at least for the nonelderly and nondisabled, with a structure close to that of the classic form outlined above. It was created in 1935 by the U.S. Social Security Act and eligibility required the presence of children and the absence of an able-bodied spouse or partner, with the practical implication that the caseload was almost entirely composed of single women with children. However, major structural reforms of the program began in 1993 with the introduction of work requirements and time limits, and it has not returned to its classic form since that time. Consequently, the analysis here will use data on disadvantaged single women with children from the late 1980s and early 1990s, just before the change in structure began.

Suitable data from that period are available from the Survey of Income and Program Participation(SIPP), a household survey representative of the U.S. population which was begun in 1984 for which a set of rolling, short (12 to 48 month) panels are available throughout the 1980s. The SIPP is commonly used for the study of transfer programs because respondents were interviewed three times a year and their hours of work, wage rates, and welfare participation were collected monthly within the year, making them more accurate than the annual retrospective time frames used in most household surveys. The SIPP questionnaire also provided detailed questions on the receipt of transfer programs, a significant focus of the study reflected in its name. I use all waves of panels interviewed in the Spring of each year 1988-1992 (only Spring to avoid seasonal variation) and pool them into one sample, excluding overlapping observations by including only the first interview when the person appears in more than one year (to avoid dependent observations, which would complicate the estimation).

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 at least one child under 18. The sample is therefore restricted to such families, similar to the practice in past AFDC research. To concentrate on the AFDC-eligible population, I restrict the sample to those with completed education of 12 years or less, nontransfer nonlabor income less than \$1,000 per month, and between the ages of 20 and 55. The resulting data set has 4,115 observations.

The means of the variables used are shown in Appendix Table A1. The variables include hours worked per week in the month prior to interview (H) (including zeroes), whether the mother was on AFDC at any time in the prior month (P), and covariates for education, age, race, and family structure, and nontransfer nonlabor income in the month prior to interview. Thirty-five percent of the observations were on AFDC. The hourly wage is omitted from the reduced form analysis because it is only available for workers but is needed for the structural model. For the latter, a log real hourly wage equation is estimated for workers jointly with a equation for the probability of working, assuming joint normality and including a dummy variable for part-time work to pick up differences in hourly wages for full-time and part-time workers. Estimates are shown in Appendix Table A2 and the mean of the resulting predicted wage for part-time and full-time workers is shown in the table of means.

Over this time period, AFDC guarantees varied by state and hence are assigned by the state of residence of the observation. The guarantees varied by family size as well and hence are assigned on that basis to each observation (family size variables are also included in the list of covariates in the labor supply equation, as noted above, to control for the influence of its variation per se). AFDC tax rates on earned and unearned income were nominally mandated to be 100 percent over this period but, because of deductions and other exclusions, averaged .40 for earned income and .60 for unearned income (Ziliak (2007)). AFDC recipients were also categorically eligible for Food Stamp benefits during the period and hence are assigned to all AFDC recipients, using the federal, nationwide guarantee levels by family size and using the program's tax rate of .24 on earned income and .30 on unearned income (the AFDC benefit is included in unearned income by the Food Stamp program). Food Stamps were also available to single mothers off AFDC and hence are also assigned. Food Stamp benefits are sufficiently low that they are generally considered to be equivalent to cash. Positive tax payments are also calculated for each family including the payroll tax and the federal income income tax, including the Earned Income Tax Credit. The way in which benefits and taxes differs in the reduced form and structural models, as described below.

Fixed Cost Proxies. The empirical work uses variables proxying the fixed cost of AFDC participation, variables which represent state-specific nonfinancial mechanisms used to limit participation in the program. There is a large social work literature on the use by states of non-financial and non-monetary mechanisms to keep AFDC caseloads down over this period and to do so strategically, with the most common mechanism simply to deny eligibility as frequently as possible using subjective interpretations of the eligibility rules (Brodkin and Lipsky (1983),Lipsky (1984),Kramer (1990)). This literature shows that caseworkers conducting eligibility assessments on individual applicants were able to interpret the rules for what types of income to count, whether an able-bodied spouse or partner was present, which assets to count, and other factors affecting eligibility broadly and subjectively. Usually, the discretion was exercised at the explicit or implicit direction of welfare department administrators who, in turn, took their guidance from the legislature and governor of the state. Hence, the degree of discretion exercised at the caseworker level reflected the attitudes of the state governance structure toward the AFDC program.

However, because the program was federally regulated, the federal government conducted a quality control program by sending auditors each year to each state to determine whether states were erroneously assessing eligibility and denying applications. The error rates assessed by the auditors in the 1980s and early 1990s are available from unpublished tables held by the Department of Health and Human Services. For each state in each year, the tables record error rates in applications denied both overall and by reason, in assessment of eligibility, in incorrectly denying requests for hearings to appeal decisions, and related variables. To test the strength of these error rates, OLS regressions were estimated on the SIPP sample, regressing a dummy for AFDC participation on a vector of exogenous covariates and these state-level error rates. From these regressions, two strong instruments emerged, one the average state eligibility error rate resulting in denial of eligibility, and one the average percent of applications incorrectly denied on the basis of procedural requirements (e.g., failure to submit proper paperwork). [7] The state-specific values of these error rates are shown in Appendix Table A3, which shows that first variable ranges widely across the states, from 2.1 to 10.9 percent, as does the second, which ranges from 1 to 35 percent. The two error rate variables have negative effects on welfare participation in the SIPP sample and, combined, yield an F-statistic of a little over 10. The assumption in the analysis below is that these errors are independent of labor supply levels in the states, conditional on the budget constraint and other exogenous covariates used in the empirical model.

Preliminary Results. Probit parameters estimates for the propensity score estimation are shown in Appendix Table A4. Figure 2 shows a histogram of predicted welfare participation rates, including both exogenous variates X as well as the instruments Z in the prediction. The support of the distribution is concentrated at $H = 0$ but is also fairly good at ranges below about .50 but not above that. However, much of this variation is driven by the X vector. Figure 3 provides instead an indication of the explanatory power of the Z vector alone, showing the variance of the residual $[\widehat{F}(X,Z) - \widehat{F}(X,\overline{Z})]$ plotted

[7] The averages were computed by averaging the rates for each state over the years of the SIPP sample, 1988-1992. Averaging was necessary to smooth out random fluctuations from year to year.

against deciles of the distribution of $\widehat{F}(X,\overline{Z})$. The instruments have reasonable incremental explanatory power for participation in the range (0.3, 0.8) but little power outside that range. Therefore it is unlikely that these instruments will have much power in estimating the MTE below 0.3 or above 0.8.

For the initial results we set $\lambda = 0$ (hence no interactions of X with participation) and estimate the hours equation by OLS, regressing hours on X and P. OLS gives a response estimate of -22.7 (s.e.=.50), which is only slightly smaller than the raw mean difference between participants and non-participants of -25.1, implying that conditioning on X has little effect. The main object of estimation is to determine whether program participation has a linear effect on hours of work and for that purpose we specify g as a series function with natural cubic splines as the bases of the function, cubic splines being the default polynomial order in most applications.[8] Natural cubic splines tie down the ends of the spline curve to a flat region, which is usually helpful because of the well-known tendency for polynomials to display poor behavior at the end points (Hastie et al. (2009),pp.145-146). Knot points are chosen at percentile points of the propensity score distribution. Fit is gauged by the generalized cross-validation (GCV) statistic, which is simply the sum of squared deviations adjusted for degrees of freedom. See Chen (2007) for consistency results for sieve models.

The MTE is $\partial H/\partial F$. Estimation of (21) with $\lambda = 0$ for spline specifications with 3, 4, 5, and 6 knots generates the MTEs and 95 percent confidence intervals shown in Figure 4.[9] The GCV for all four are extremely close, differing only at the 4th significant digit, so all fit the data about the same. But all of them also show the same general pattern of non-monotonic MTEs, with labor supply disincentives rising then falling as participation rises. The 95 percent confidence interval bands show that the estimates are significantly different from zero only in the middle range of participation probabilities, which may be

[8] That is, $g(F) = \gamma_0 + \sum_{j=1}^{J} \gamma_j Max(0, F - \pi_j)^3$, where where the π_j are preset spline knots.

[9] Estimates of the hours equation for the 3-knot specification are shown in Appendix Table A4.

because the instruments have low power in those ranges, as noted above. The central range where the estimates re significant differs across the specifications but is approximately from .10 to .15 at the bottom to .60 to .80 at the top. The maximum work disincentive is large and over 40 hours per week, although the confidence bands range down to 35 or so. But most of the disincentive effects are lower than this, down to around 20 hours at low participation rates and down to around 10-15 hours at higher participation rates, all only while in the 95 percent significance band.

Why the MTE should follow this particular pattern cannot be determined without some structure imposed on the model (see the next section) but a plausible hypothesis is that when the fixed costs of participation are high and hence participation rates are low, those who participate at those who would have been at $H = 0$ in the absence of the program and who will not reduce hours after joining the program yet who have a high marginal utility of consumption relative to leisure. As participation costs fall, those with higher hours of work and more initial take-home income, and who therefore have less need for additional consumption, could be those who put more marginal value on leisure instead. It is less clear why work reductions would begin to fall as costs fall even further. One possibility is that such families have both low H but larger sources of income not measurable in the data and hence omitted, who have less need to participate in the program in the first place and who, when joining, have both small marginal utilities of leisure and consumption and obtain little extra total utility from participation.

Estimates with $\lambda \neq 0$ allow the effect of participation on labor supply to differ by characteristics. Estimates are shown in Table 1. Those with higher levels of education, who have larger family sizes, and face higher welfare guarantees have greater work reductions when entering welfare. Those who are older, who have higher nonlabor incomes, and are in states with higher unemployment rates have smaller work reductions. The three budget-constraint-related effects are roughly consistent with a static labor supply model for a program with a high tax rate (and hence $H = 0$ if on welfare): an increase in the wage increases hours of work off welfare and implies a larger reduction conditional on participation; a high guarantee means a larger reduction conditional on participation; and an increase in nonlabor income reduces H if off welfare, leading to a smaller labor supply reduction conditional on participation.

4 Marginal Effects of Major AFDC Reforms

Table 2 shows the marginal labor supply effects in 1967, 1981, 1996, and 2010 using the estimates from the 3-knot cubic spline model. In 1967, the participation rate was 36 percent G was 32 percent higher than in the 1988-1992 period from which the data used for the estimation were drawn. At that point, the marginal person had a -47 hour response. In 1981, the guarantee had fallen to a level about 5 percent greater than in 1988-1992 and the participation rate had risen to 53 percent. At that point, the marginal labor supply response was considerably lower than in 1967. In 1996, at the time of the structural change in the program, G had risen slightly but the participation rate had fallen to 40 percent, which is the point where the largest marginal effects occur. The marginal labor supply response rose to -49 at that year.

But the major outlier in the Table is for 2010, when the participation rate had fallen to 13 percent as a result of 1996 reforms that imposed time limits and work requirements in the program and greatly reduced the caseload. The marginal labor supply responses calculated from the model for that year are necessarily hypothetical because the program no longer had the voluntary participation structure it had during the 1988-1992 period used in the estimation. Nevertheless, a marginal expansion of participation in 2010 would have had, if the program had retained its classical form, a labor supply response considerably smaller than at any other of the reform periods back to 1967.

5 Summary and Conclusions

This paper presents a modified version of the traditional static labor supply model of the effect of income transfers to allow for heterogeneous response to estimate marginal responses. The results show some differences in marginal labor supply responses at different points in the history of the AFDC program, with the biggest differences occurring in the most recent period, where a marginal expansion of participation is predicted to have considerably smaller labor supply responses than in earlier periods, assuming the model under classical forms applies. The results imply different marginal labor supply responses to past reforms.

References

Bjorklund, A. and R. Moffitt (1987, February). The estimation of wage and welfare gains in self-selection models. *The Review of Economics and Statistics 69*(1), 42–49.

Blundell, R., M. C. Dias, C. Meghir, and J. Shaw (2016, September). Female Labor Supply, Human Capital, and Welfare Reform. *Econometrica 84*(5), 1705–53.

Blundell, R. and T. Macurdy (1999). Labor Supply: A Review of Alternative Approaches. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3A. Amsterdam North-Holland.

Brodkin, E. and M. Lipsky (1983). Quality Control in AFDC as an Administrative Strategy. *Social Service Review 57*(1), 1–34.

Chan, M. K. (2013, May). A Dynamic Model of Welfare Reform. *Econometrica 81*(3), 941–1001.

Chan, M. K. and R. Moffitt (2018). Welfare Reform and the Labor Market. *Annual Review of Economics 10*, 347–81.

Chen, X. (2007). Large Sample Sieve Estimation of Semi-Nonparametric Models. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 5549–5632. Amsterdam: Elsevier North-Holland.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning*. Springer.

Heckman, J. J. and R. Robb (1985). Alternative Methods for Evaluating the Impact of Interventions. In J. J. Heckman and B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*. Cambridge: Cambridge University Press.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006, August). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

Heckman, J. 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*(8), 4730–4734.

Heckman, J. J. and E. Vytlacil (2005, May). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Imbens, G. and J. Angrist (1994, March). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Kramer, F. (1990). Statisticis and Policy in Welfare Quality Control: A Basis for Understanding and Assessing Competing Views. *Journa of the American Statistical Association 85*(411), 850–55.

Lipsky, M. (1984). Bureaucratic Disentitlement in Social Welfare Programs. *Social Service Review 58*(1), 3–27.

Moffitt, R. (1983, December). An Economic Model of Welfare Stigma. *American Economic Review 73*(5), 1023–35.

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

Moffitt, R. (2003). The Temporary Assistance for Needy Families Program. In R. Moffitt (Ed.), *Means-Tested Transfer Programs in the United States*, pp. 291–363. Chicago, Ill.: University of Chicago Press.

Newey, W. and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing. In

R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume IV. Amsterdam: Elsevier.

Ziliak, J. (2007). Making Work Pay: Changes in Effective Tax Rates and Guarantees in U.S. Transfer Programs, 1983-2002. *Journal of Human Resources 42*(3), 619–42.

Ziliak, J. (2016). Temporary Assistance for Needy Families. In R. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States*, Volume 1, pp. 3030–393. Chicago, Ill.: University of Chicago Press.

Appendix

Supplementary Tables

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 six years 1988-1992. The state rate of incorrect denial of eligibility was obtained from unpublished tables provided to the author by the U.S. Department of Health and Human Services, compiled from the federal AFDC quality control program. The state percent of applicants denied for failure to comply with procedural requirements was obtained from the issues of Quarterly Public Assistance Statistics published quarterly by the U.S. Department of Health and Human Services, data which were also obtained originally from the federal AFDC quality control program. Fluctuations from year to year in the two quality control variables were smoothed out by computing a weighted mean of each for each state over the 1988-1992 years.

Appendix Table A1
Means and Standard Deviations of the Variables Used in the Analysis

Variable Name	Variable Definition	Total sample	P=1	P=0
Hours	Average hours of work per week in the month prior to survey	21.8 (19.7)	4.5 (11.2)	29.6 (17.5)
Nonwork	0-5 hours	.41	.18	.83
Parttime work	6-35 hours	.12	.13	.10
Fulltime Work	Over 35 hours	.47	.69	.07
P	Dummy variable equal to 1 if individual was on AFDC anytime in the month prior to survey	.35	--	--
Hourly Wage Rate				
Observed (workers only)				
Predicted Parttime		$4.54 (0.66)		
Predicted Fulltime		$7.28 (1.06)		
Age	Age in years at survey date	32.6 (9.0)	30.8 (7.9)	33.4 (9.4)
Education	Years of education at survey date	10.9 (1.9)	10.5 (2.0)	11.1 (1.9)

Appendix Table A1 (continued)

Variable Name	Variable Definition	Total sample	P=1	P=0
Family size	Number of individuals in the family at the survey date	3.3 (1.4)	3.4 (1.5)	3.3 (1.4)
No. Children Lt 6	Number of children less than 6 in the family at the survey date	.70 (.85)	1.1 (1.0)	.51 (.72)
Black	Dummy variable equal to 1 if respondent is black	.34 (.48)	.45 (.50)	.30 (.46)
State Unemployment Rate		6.3 (1.5)		
Nonlabor income	Nontransfer nonlabor income in the month prior to survey	$389.2 (303.0)	$75.0 (192.7)	$407.1 (411.0)
Welfare G	State monthly AFDC guarantee for a family of four	$419.2 (212.5)	$440.0 (210.0)	$409.0 (213.0)
Food Stamp G		$66.1 (14.5)		
Eligibility Denial	Percent of applicants incorrectly denied eligibility	4.93 (1.67)	4.77 (1.59)	5.00 (1.71)
Pctdenied	Percent of applications denied for failure to meet procedural requirements in the state	14.7 (8.8)	14.1 (8.3)	15.0 (9.1)

Sample size	--	4,115	1,440	2,675

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

Appendix Table A2

Wage Equation (to be included)

Appendix Table A3: Fixed Cost Variable, by State

State	Eligibility Denial	Percent Denied
Alabama	4.393452	0.2856738
Alaska,Idaho,Montana,W	3.048238	0.2494895
Arizona	6.912738	0.4058077
Arkansas	3.791667	0.3690263
California	3.879762	0.2608052
Colorado	2.683333	0.3768443
Connecticut	3.55119	0.3484459
Delaware	6.304762	0.2310895
District of Columbia	6.950357	0.3547997
Florida	7.672381	0.3797023
Georgia	4.110952	0.4041264
Hawaii	3.453691	0.2340734
Illinois	4.570833	0.2765072
Indiana	4.302381	0.2043593
Iowa,North Dakota,South	3.562348	0.2373189
Kansas	3.937381	0.228917
Kentucky	2.735119	0.2338173
Louisiana	4.904524	0.2798858
Maine,Vermont	3.614402	0.2765111
Maryland	6.225714	0.2677891
Massachusetts	3.557143	0.2036663
Michigan	4.016071	0.2991185
Minnesota	2.249167	0.0983536
Mississippi	4.519286	0.2668127
Missouri	4.060595	0.2318035
Nebraska	5.465384	0.1855049
Nevada	3.015	0.477916
New Hampshire	5.025476	0.3242095
New Jersey	4.8975	0.1213643
New Mexico	5.449643	0.3995799
New York	5.545714	0.0530247
North Carolina	2.360476	0.113235
Ohio	6.795714	0.2488872
Oklahoma	3.117024	0.2432857
Oregon	3.547857	0.166573
Pennsylvania	4.727024	0.1318218
Rhode Island	3.555238	0.1751217
South Carolina	4.813452	0.3339927
Tennessee	5.533453	0.238083
Texas	6.00631	0.4670539
Utah	2.789167	0.1749215

Washington	4.189762	0.2941487
West Virginia	5.890952	0.1824447
Wisconsin	3.825238	0.2375345

Appendix Table A4

First-stage and Second-Stage Reduced Form Coefficients,
Natural Cubic Spline with Three Knots, λ=0

	Participation Equation (Probit Coefficients)	H equation (OLS)
Gammaa		
S1	--	-35.7 (9.5)
S2	--	-.043 (13.0)
S3	--	53.3 (33.1)
Beta		
Education	--	1.08 (0.17)
Age	--	0.13 (0.03)
Black	--	-0.74 (0.60)
No. Children Lt 6	--	0.13 (0.64)

Appendix Table A4 (continued)

	Participation Equation (Probit Coefficients)	H equation (OLS)
Family size	--	-0.96 (0.20)
Nonlabor income	--	-4.38 (0.64)
Unemployment Rate	--	-0.50 (0.16)
AFDC G	--	-0.006 (0.01)
Constant	--	25.3 (3.8)
Eta		
Education	-0.02 (0.01)	--
Age	-0.018 (0.003)	--
Black	0.30 (.05)	--
No. Children Lt 6	0.31 (.03)	--

Appendix Table A4 (continued)

	Participation Equation (Probit Coefficients)	H equation (OLS)
Family size	0.05 (0.02)	--
Nonlabor income	-1.04 (0.08)	--
Unemployment Rate	0.05 (0.02)	--
AFDC G	0.003 (.0006)	--
Constant	-0.18 (0.84)	--
Delta		
Ineligible percent	-0.069 (0.018)	--
Percent denied	-0.432 (.223)	--

Notes:

Bootstrapped standard errors in parentheses.

aThree-Knot natural cubic splines in the propensity score with knots at .15, .33, and .54.

Table 1

Lambda Parameter Estimates

Education	-1.7 (0.8)
Age	0.5 (0.2)
Black	3.9 (2.9)
No. Children Lt 6	0.3 (2.3)
Family Size	-2.9 (1.0)
Nonlabor Income	4.1 (1.4)
Welfare Guarantee	-0.01 (0.00)
Unemployment Rate	1.4 (0.7)

Notes:

Standard errors in parentheses.

Table 2

Marginal Labor Supply Responses at Historical
and Hypothetical AFDC Reforms

1967[a]	-47.2
1981[b]	-31.0
1996[c]	-49.3
2010 Hypothetical Average[d]	-28.0

Notes:
a 1967 G = 1.32 1990 value, Participation rate = .36.
b 1981 G=1.05 1990 value. Participation rate = .53.
c 1996 G=1.20 1990 value. Participation rate = .40.
d 2010 G = .79 1990 value. Participation rate .13.

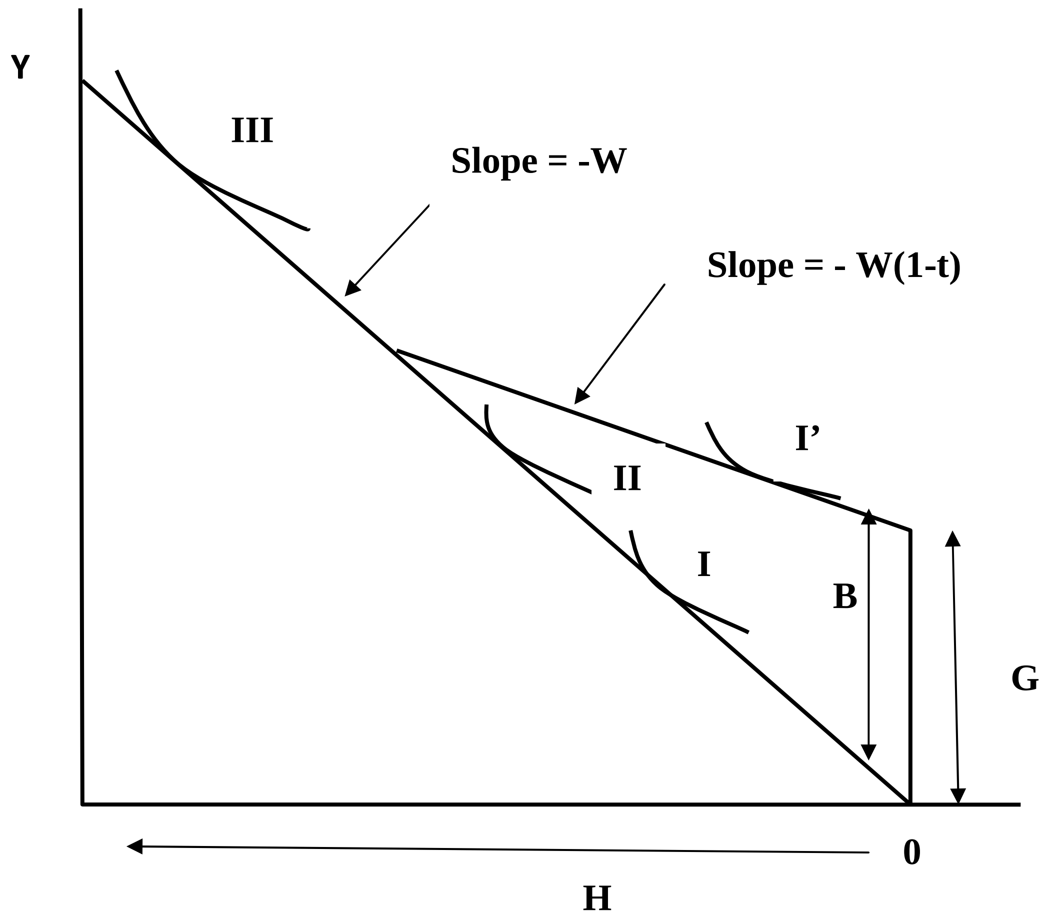
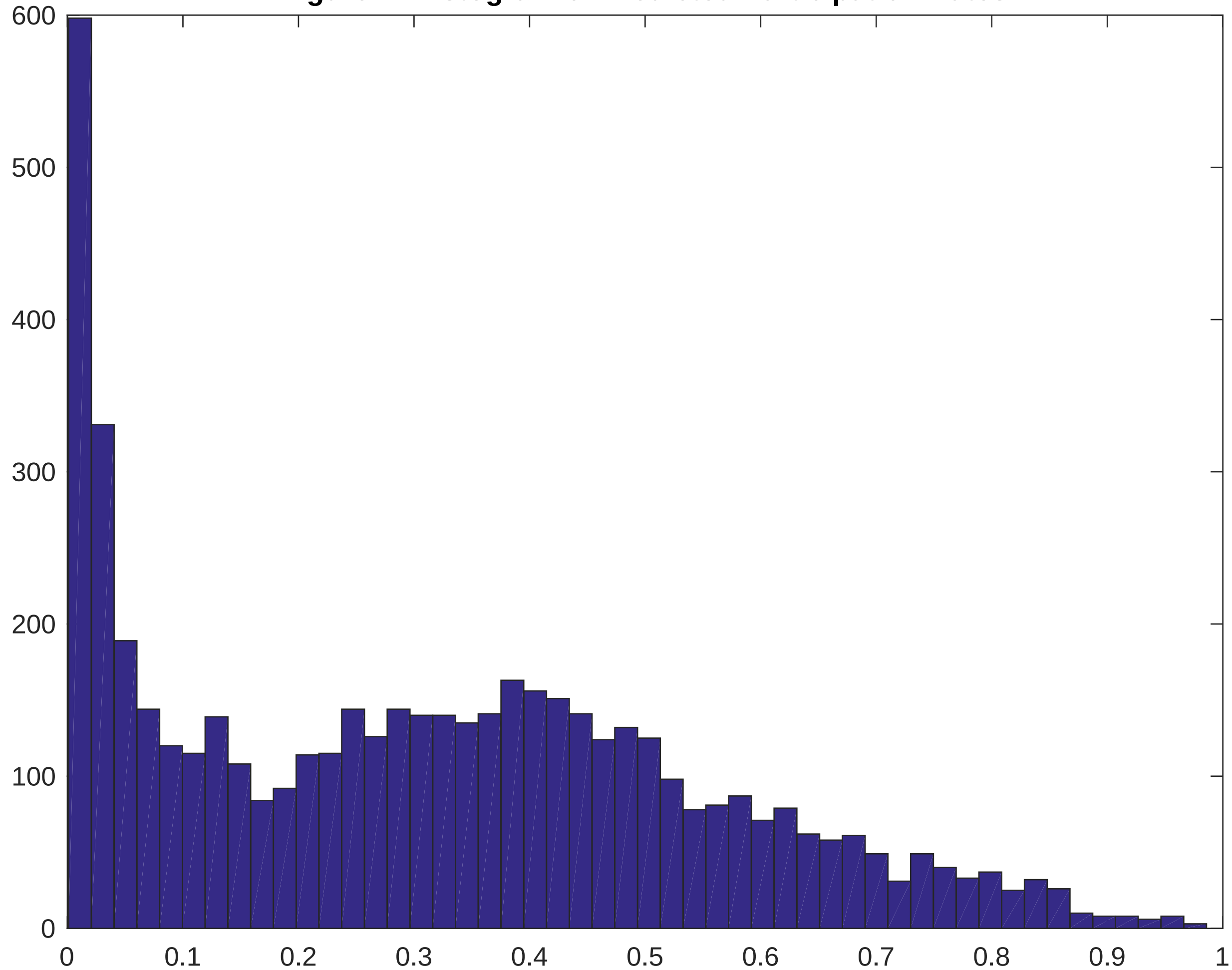


Figure 1

Figure 2: Histogram of Predicted Participation Rates



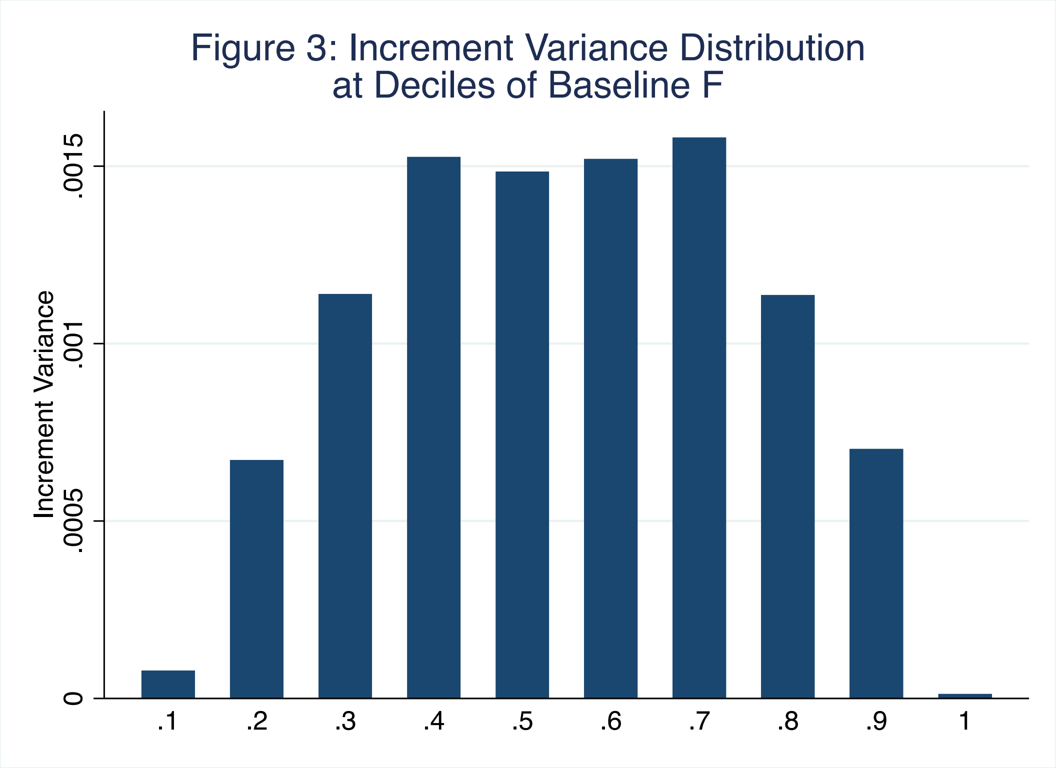


Figure 3: Increment Variance Distribution
at Deciles of Baseline F
Increment Variance
0
.0005
.001
.0015
.1
.2
.3
.4
.5
.6
.7
.8
.9
1

Figure 4: Marginal Treatment Effects