NBER WORKING PAPER SERIES

THE MARGINAL LABOR SUPPLY DISINCENTIVES OF WELFARE REFORMS

Robert A. Moffitt

Working Paper 26028
http://www.nber.org/papers/w26028

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2019

The author would like to thank Marc Chan, Kai Liu, Shaiza Qayyum, Kyungmin Kang, and Sue Bahk for research assistance as well as the participants of a large number of conferences and departmental seminars and numerous specific individuals for comments, including formal discussant remarks by James Ziliak. Comments at a seminar at the University of Chicago were particularly helpful. Research support from the National Institutes of Health grant NIH R01 HD 27248 is gratefully acknowledged. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

The Marginal Labor Supply Disincentives of Welfare Reforms
Robert A. Moffitt
NBER Working Paper No. 26028
June 2019
JEL No. C21,I3,J2

ABSTRACT

Existing research on the static effects of the manipulation of welfare program benefit parameters on labor 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 distribution of preferences. A new examination of the heavily studied AFDC program examines changes in its tax rates in 1967, 1981, and 1996 and estimates the marginal effects on labor supply of a change in participation in each of those reform years. The estimates are based on a theory-consistent reduced form model which allows 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 marginal treatment effects are quadratic, rising and then falling as participation rates rise. Applying the estimates to the three historical reform periods shows that marginal responses differed in each period because of differences in the composition of who was on the program and who was not.

Robert A. Moffitt
Department of Economics
Johns Hopkins University
3400 North Charles Street
Baltimore, MD 21218
and IZA
and also NBER
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 that 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. Notable reforms in the program took place in 1967, 1981, and 1996, with a decrease in the tax rate in the first year from 100 percent to 67 percent, an increase in the tax rate back to 100 percent in the second year, and a decrease in the tax rate again in the third year to approximately 50 percent (albeit accompanied by many other reforms). The effects of these reforms on labor supply have been heavily studied (see Moffitt (1992), Moffitt (2003), and Ziliak (2016) for reviews).

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.

Prima facie evidence for differences in the composition of the caseload is shown in Figure 1, which shows both the caseload of the program and the participation rate of single-mother families (the primary eligible group for the program) over the period 1967-2015. The caseload rose in the late 1960s, flattened out over the 1970s and the 1980s, rose again in the early 1990s, and fell sharply thereafter. These changes are almost certainly associated with changes in the composition of who participated and who did not. The fraction of single mother families participating does not follow the exact same trend because the number of such families grew strongly in the 1970s and early 1980s, leading the program participation rate (at least measured against all single mother families) to fall. For present purposes, clearly if labor supply responses to program participation are heterogeneous, the response of the marginal mother is likely to have changed over time.

But who participates and who does not is also affected by the level of the program guarantee and the tax rate on earnings, since those parameters affect the gains to participation for women at different levels of labor supply. Real guarantees fell from 1967 to 1981 and rose back again from 1981 to 1996 (Moffitt (2003)), and major changes in tax rates occurred in the aforementioned three years, which are indicated in the Figure. Other demographic features of the eligible and participation populations could also have changed. Estimating the labor supply of the marginal individual over the last 30 years therefore requires an analysis which takes into account changes in the demographic composition of the population, the fraction of the population participating in the program, and the levels of the program parameters in each year.

This paper specifies a model allowing the composition of the caseload to affect who is on the margin and who is not, and uses estimates of the model to simulate marginal labor supply responses of changes in participation in the tax rate reform years of 1967, 1981, and 1996. The first section 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, then uses that model to provide a formal definition of the labor supply response of the marginal individual. 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 over time in an arbitrary and unrestricted fashion.

The second section of the paper presents a reduced form model designed for the estimation of marginal labor supply responses. An example of a structural model which could generate the reduced form is given. Those responses can be nonparametrically identified over the range of participation rates provided 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),Björklund and Moffitt (1987), and many subsequent papers) and sets up a modified form of the local instrumental variable (LIV) estimation method (Heckman and Vytlacil (1999, 2001, 2005, 2007); see also Heckman et al. (2006)). The parameters of the reduced form model are directly related to those of the structural model and are fully theory-consistent.

The third section estimates the reduced form model with cross-sectional data from the late 1980s and the early 1990s–the last time the AFDC program had its classic shape–using sieve methods to nonparametrically estimate the shape of the marginal labor supply effect curve. State-level variables which affect fixed costs of participation but not labor supply conditional on participation are used as instruments. The instruments provide variation in participation rates not over the full range of participation rates but only over a subrange of it, but the range covered includes the range of participation rates in the three historical tax-rate reform years of interest (1967, 1981, and 1996). Consequently, a full range of participation rates is not needed for the goal of the paper.[1] Estimates from the model show that marginal labor supply responses are significantly different over the covered range of participation rates and are U-shaped and non-monotonic, growing in size as participation increases and then declining after a certain level of participation. This pattern is shown to

[1]Methods for extrapolation beyond the support of participation rates in the data, as discussed by, for example, Brinch et al. (2017) and Mogstad et al. (2018), are therefore also not needed.

arise from changes in the full-time/part-time response to participation as participation rises. The marginal response also varies modestly with the levels of the guarantee and the tax rates conditional on the participation rate, and varies with demographic characteristics.[2]

The fourth section uses the estimated model to calculate who was on the margin at the time of the 1967, 1981, and 1996 changes in the program tax rate. The initial participation rate, the guarantee, the tax rate, and demographic characteristics were different in each of those years, leading to differing marginal changes in response to the reforms in each year. Marginal responses were largest in 1967, smallest in 1981, and in between in 1996. The calculations also show that the reduction in the tax rate from 100 percent to 67 percent in 1967 and the increase in the tax rate back to 100 percent in 1981 did not have symmetric effects because the populations on the margin were different in those two years.

The general implication of the analysis is that policy forecasts of incremental reforms must be based on the nature of the specific population that is participating in the program at the time of the reform as well as on the nature of the reform, since which program parameters are 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, and this should apply to future reforms of transfer programs in general. 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.

[2]Estimates assuming homogeneous effects are also presented and are shown to provide a very misleading picture of marginal responses.

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), Chan and Moffitt (2018)) assumes utility to be

$$U(H_i, Y_i; \theta_i) - \phi_i P_i \quad (1)$$

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 whose distribution is in the positive domain. The presence of P_i allows for the presence of fixed costs of participation–in money, time, or utility (stigma), with the exact type unspecified and scaled in units of utility (Moffitt (1983), Daponte et al. (1999), Currie (2006)). 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 partially separable from the H decision.[3] 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

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

[3] 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 H is the labor supply function, 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 2 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).

Eqns(3)-(5) are in the form of a generalized Roy model, but where the outcomes for the two regimes are notationally represented in the single eqn(3) instead of two separate equations. The fixed cost term ϕ_i plays the role of the cost term in the generalized Roy model while the change in V corresponds to the gain in earnings or other outcome in that model. Unlike the Roy model where the earnings gain is typically assumed to be linear in the selection equation (e.g., Heckman and Vytlacil (2005)), here the unobservable θ_i enters nonlinearly through the indirect utility function V. The two unobservables, θ_i and ϕ_i, have a joint distribution function $G(\theta_i, \phi_i)$.

The response to the program for individual i conditional on the budget constraint parameters is

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

where $C_i = [W_i, N_i, G, t, r]$ is the set of budget constraint variables. The response in eqn(6) is a heterogeneous response if θ_i varies with i. There is a latent distribution of these responses for the full population, including those who do not eventually participate.

To define the marginal labor supply response, or marginal treatment effect, first note that eqn(4) implies that individuals on welfare must have increases in V from participation that are greater than their ϕ_i values. A shift downward in the distribution of the ϕ_i will bring onto the program those whose increases in V had put them just on the margin of participation initially. The values of $\triangle_i$ for those individuals are the labor supply responses of those on the margin.

More formally, define θ_D and ϕ_D as the values which make an individual indifferent between participation and non-participation:

$$\begin{aligned} 0 &= V[W_i(1-t), G + N_i(1-r); \theta_D] - V[W_i, N_i; \theta_D] - \phi_D \\ &= dV(\theta_D|C_i) - \phi_D \end{aligned} \tag{7}$$

where the second line just defines $dV(\theta_D|C_i)$. Eqn(7) defines a locus of the two unobservables along which marginal individuals locate. That locus is shifted by the budget constraint parameters. Following the literature (e.g., Heckman and Vytlacil (2005), eqn(4)), the marginal labor supply response can be defined as $\triangle^{MTE}(C_i) = E_{\phi_D} \triangle\left[\theta_D(\phi_D, C_i)|Ci\right]$ where $\theta_D(\phi_D, C_i)$ is the function solving eqn(7) for θ as a function of ϕ_D and C_i.[4]

The question raised in the Introduction is whether the labor supply responses of those on the margin are greater or smaller than those initially on the program, holding constant

[4] As previously noted, in the typical generalized Roy model, the unobservables are linearly related in this indifference locus and hence only the composite error term matters for selection. Here, with the unobservables nonlinearly related, selection depends on the two unobservables separately.

C_i. The answer is that the sign is ambiguous. While those on the margin have, by definition, smaller values of $dV(\theta_i|C_i)$ than those initially on the program, there is no necessary relationship between the magnitude of those utility differences and the magnitudes of the $\triangle_i|C_i$. Intuitively, the utility gain $dV(\theta_i|C_i)$ is achieved by some combination of increase in leisure and increase in goods consumption. The mix depends on relative preferences for those two goods, and those relative preferences can vary arbitrarily over the dV distribution. Consequently, as ϕ_i falls in successive increments and as program participation rises, $\triangle_i$ can rise, fall, or remain the same in any arbitrary pattern.[5] It is the goal of the empirical work in the sections below to identify that pattern.

A hypothetical pattern of a relationship between dV_i and $\triangle_i$ is shown in Figure 3. While dV is a function of θ and not $\triangle$, θ can be defined without loss of generality to be monotonically related to $\triangle$ and hence the horizontal axis can be represented with either parameter. Figure 3, reflecting the just-mentioned result that dV and $\triangle$ can have any arbitrary relationship, assumes that they have an alternating pattern of positive and negative association. For individuals with a value of ϕ_0, three regions are identified where $P = 1$ and each is associated with a range of labor supply responses, $\triangle$ (those ranges are labeled 1, 2, and 3 on the horizontal axis). A fall in the value of ϕ to ϕ_1 increases participation, and the regions of $\triangle$ of participation expand. The mean $\triangle$ of those newly joining the program is the integral over the distribution of $\triangle$ in the new regions of participation. Of course, in actuality there is a joint distribution of $\triangle$ and ϕ, so the actual regions of participation and of $\triangle$ will depend on that joint distribution and must be integrated over both.

Since the locus of indifference is where $dV = \phi$, the indifference locus showing the values of ϕ which make participation marginal for any value of θ or $\triangle$–that is, the locus corresponding to eqn(7)–will have the same pattern as Figure 3. It is shown in Figure 4,

[5] As in the generalized Roy model, there is positive selection on gains to participation unless costs are positively correlated with gains. But positive selection occurs on V, not $\triangle$, and those do not have a determinate relationship.

along with the regions where $P = 0$ and $P = 1$. The joint distribution of the two parameters determines the magnitude of the participation and non-participation rates. The locus is shifted when the budget constraint parameters C_i change or when the parameters of the joint distribution $G(\theta_i, \phi_i)$ shift. For example, if $\phi_i = \overline{\phi}(Z_i) + \nu_i$, where Z_i is an observable proxy for costs and ν_i represents unobserved costs, the line of indifference is the same as in Figure 4 but with the vertical axis measuring ν instead of ϕ, and with the indifference line understood to be conditional on Z_i. A shift in Z_i hence shifts the indifference locus.

It is worth noting that changes in the program parameters G, t, and r do not identify marginal responses because they also have inframarginal responses on those initially on the program. Changes in those parameters shift the line of indifference and hence the participation rate, but they also change the latent population distribution of responses, $\triangle_i$. The literature on the labor supply responses to welfare has mostly consisted of regressions of H on those parameters and hence do not identify marginal effects.

The marginal labor supply response like that illustrated in Figure 3 is typically identified by a change in the mean effect of the treatment on the treated (i.e., the mean labor supply response of participants) as participation expands. That mean in this model is

$$\begin{aligned} \widetilde{\triangle}_{P_i=1} &= E(\triangle_i | C_i, P_i = 1) \\ &= \frac{1}{P} \int\limits_{S_{\theta\phi}} \int \triangle_i(\theta_i | C_i) dG(\theta_i, \phi_i) \end{aligned} \tag{8}$$

where $S_{\theta\phi}$ is the set of parameters in regions demarcated by the θ_D,ϕ_D locus which generates $P = 1$, and where

$$\begin{aligned} P &= E(P_i | C_i) \\ &= \int\limits_{S_\phi} \int\limits_{S_\theta} 1\{V[W_i(1-t), G + N_i(1-r); \theta_i] - V[W_i, N_i; \theta_i] - \phi_i\} dG(\theta_i, \phi_i) \end{aligned} \tag{9}$$

is the participation rate (S_θ and S_ϕ represent the unconditional supports of the two parameters). The mean effect of the transfer program over the entire population, participants and non-participants combined, conditional on the budget constraint, is

$$\begin{aligned}\widetilde{\triangle} &= E(\triangle_i P_i | C_i) \\ &= \int\limits_{S_{\theta\phi}} \int \triangle_i(\theta_i | C_i) dG(\theta_i, \phi_i)\end{aligned} \tag{10}$$

The marginal treatment effect is traditionally defined as the marginal response to an exogenous increase in program participation, which in the notation here is the mean $\triangle$ of those who change participation, or $\partial\widetilde{\triangle}/\partial P$.[6] The values of the response quantities $\triangle_i$, $\widetilde{\triangle}$, $\widetilde{\triangle}_{P_i=1}$, and $\partial\widetilde{\triangle}/\partial P$ must all be nonpositive according to theory.

2. A Reduced Form Econometric Model

The object of the empirical work is to estimate the marginal effect on hours of work of a change in participation induced by a change in fixed costs. Eqn(3) implies that, definitionally,

$$\begin{aligned}H_i &= P_i H[W_i(1-t), G + (1-r)N_i; \theta_i] + (1-P_i)H(W_i, N_i; \theta_i) \\ &= H(W_i, N_i; \theta_i) + P_i \triangle_i\end{aligned} \tag{11}$$

where $\triangle_i$ is defined in eqn(6). Now assume that $\phi_i = m(Z_i, \nu_i)$, where Z_i is an observable correlate of fixed costs and ν_i represents variation in ϕ_i conditional on Z_i. Then mean hours of work in the population conditional on the budget constraint and on Z_i can be

[6] The MTE is more usually defined as the derivative of $E(H_i|P)$ w.r.t. P (ignoring other conditioning covariates) but since $E(H_i|P) = \mathit{constant} + \widetilde{\triangle}$ (see next section), the two are equivalent.

expressed as

$$E(H|W,N,G,t,r,Z) = E_\theta[H(W,N;\theta) \mid W,N] \\ + E_{\theta,\nu}[\triangle \mid P=1,W,N,G,t,r,Z]E_{\theta,\nu}(P|W,N,G,t,r,Z) \quad (12)$$

where individual subscripts have been omitted for simplicity. Both the left hand side and the last term on the RHS are identified in the data so the question is whether the conditional mean of $\triangle$ can be (this is the effect of the treatment on the treated and was expressed in the last section as eqn(8)). This can be most easily seen, and the estimation method also clarified, by letting the conditioning on the budget constraint variables be implicit. Then eqn(11) (which comes from eqn(3)) and its associated equations(4)-(5) can be written in unrestricted form as

$$H_i = \beta_i + \alpha_i P_i \quad (13)$$

$$P_i^* = m(Z_i, \delta_i) \quad (14)$$

$$P_i = 1(P_i^* \geq 0) \quad (15)$$

where β_i is hours worked off welfare and α_i is a relabeling of $\triangle_i$. Eqn(13) is equivalent to eqn(3) and the β and α parameters are functions of θ. The participation equation is a representation of eqns(4)-(5)and the parameters δ_i represent the combined variation from θ and ν (i.e., variation in ϕ conditional on Z). 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 and eqn(13) is in the form of a familiar random coefficients model.[7] The function m can likewise be unrestricted and can be saturated if Z_i is assumed

[7]Setting it up as a random coefficients model goes back to Heckman and Robb (1985) and Björklund and Moffitt (1987).

to have a multinomial distribution, although we shall discuss possible 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) or both because both may be related to δ_i and, in fact, the theoretical model implies that they must be because the participation equation contains θ, the labor supply preferences. Eqn(12), which conditions on Z_i and hence is a reduced form, and the associated participation equation, now take the form

$$E(H_i \mid Z_i = z) = E(\beta_i \mid Z_i = z) + E(\alpha_i \mid P_i = 1, Z_i = z) \Pr(P_i = 1 \mid Z_i = z) \tag{16}$$

$$E(P_i \mid Z_i = z) = \Pr[m(z, \delta_i) \geq 0] \tag{17}$$

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

$$A1. \;\; E(\beta_i \mid Z_i = z) = \beta \tag{18}$$

$$A2. \;\; E(\alpha_i \mid P_i = 1, Z_i = z) = g[E(P_i \mid Z_i = z)] \tag{19}$$

where g is the labor supply effect of the treatment on the treated conditional on Z_i, which 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 A1 is familiar, A2 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 (19) 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 its effect on the participation probability (i.e., the propensity score), and only at specific values of Z_i.[8] If this were not so, different values of Z_i would lead to different conditional means of α_i through some other channel (e.g., Z_i

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

could be correlated with tastes for work), which would mean that α_i is not independent of Z_i.[9]

Inserting the two assumptions into the main model in eqns (16)-(17), and denoting the participation probability as $F(Z_i) = E(P_i \mid Z_i)$, we obtain two estimating equations

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

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

where ϵ_i and ξ_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 implies 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 this way of specifying the model–that is, as a random coefficient 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 margin vary, however, the response of the population mean of H_i to a change in participation will depart from linearity.

This formulation of the heterogeneous-response treatment model has been noted by Heckman and Vytlacil (2005) and Heckman et al. (2006), and eqn(20) follows from their work. However, here it will form the basis of the estimation of the model and the conditional mean function in eqn(20) will be estimated directly by regressing labor supply on the participation probability with a coefficient that is a nonparametric function of that

[9] 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. This condition is satisfied if the distribution of ϕ_i in eqn(4) is in the positive (or at least non-negative) domain.

probability.[10]

When nonparametric identification of the parameters of the model–β, the function g at every point F, and F itself–can be identified has been extensively discussed in the literature and need only be briefly stated. 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). With identification of F, 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 between those two points. With multiple values of z, multiple LATE values can be identified. A marginal treatment effect is a continuous version of this and requires some smoothing method across discrete values of Z, and is computed by $\partial H/\partial F = g'(F)F + g(F)$. However, while the MTE $\partial H/\partial F$ is identified, g and g' are not unless there is a value of Z_i in the data for which $F(Z_i) = 0$. In that case, β is identified from the mean of H_i at that point and hence g is identified pointwise at every other value of z since F is identified. If no such value is in the data, then g can only be identified subject to a normalization of its value at a particular value of z or if the value of g is known at some value.

Reintroducing the budget constraint parameters, the reduced form can be expressed by conditioning on those parameters as well as on Z_i, leading to eqn(12) with the identifying restrictions imposed:

$$\begin{aligned} E(H \mid W_i, N_i, G, t, r, Z_i) &= E_\theta[H(W_i, N_i)] + E_{\theta,\nu}[\triangle_i \mid W_i, N_i, G, t, r, P_i = 1]E_{\theta,\nu}(P_i \mid W_i, N_i, G, t, r, Z_i) \\ &= h_0(W_i, N_i) + g[W_i, N_i, G, t, r, F(W_i, N_i, G, t, r, Z_i)]F(W_i, N_i, G, t, r, Z_i) \end{aligned} \tag{22}$$

Note that the theory imposes two restrictions on the form of the equation. First, the intercept of the equation, denoted by the h_0 function, must not include the welfare program parameters G, t, and r because the intercept represents labor supply off welfare.

[10] Eqn(20) 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(20) by allowing the coefficient on the propensity score to be nonparametric in that score is equivalent because it just factors the propensity score out of the function.

Hence these parameters should not be "controlled for" in the H regression. Second, the function g, which is the mean labor reduction for those participating in the program, must contain the budget constraint parameters because those parameters affect the labor supply of inframarginal participants. They must be included so that changes in the coefficient g induced by changes in F hold the budget constraint fixed, which is required for changes in that coefficient with respect to participation to identify the responses only of marginal participants and not those who are inframarginal. Of course, a fully parametric model which makes use of a specific parametric utility function and assumptions on which parameters of that function are heterogeneous would result in specific functional forms for h_0, g, and F. An illustrative fully parametric model is provided in Appendix A to demonstrate what those functional forms would look like for one such model.

Full nonparametric estimation of the three functions h_0, g, and F would make the estimation subject to the curse of dimensionality. Considerable dimension reduction can be achieved by using traditional linear indices in the observables, with

$$H_i = X_i^\beta \beta + [X_i\lambda + g(F(X_i\eta + \delta Z_i))]F(X_i\eta + \delta Z_i) + \epsilon_i \tag{23}$$

$$P_i = F(X_i\eta + \delta Z_i) + \nu_i \tag{24}$$

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, t, and r. Exogenous characteristics thus linearly affect labor supply off welfare and linearly affect the g and F functions.[11] However, the g function will continue to be nonparametrically estimated, using sieve methods (see below; normality will be assumed for F, however). With these two functions specified, we will employ two-step estimation of the model, with a first-stage probit estimation of eqn(24) and second-stage nonlinear least squares estimation of eqn(23) using fitted values of F from the first stage. Consistency and

[11] Some specifications to be estimated will interact X with g.

asymptotic normality of two-step estimation of nonlinear conditional mean functions with estimated first-stage parameters is demonstrated in Newey and McFadden (1994). Standard errors are obtained by jointly bootstrapping Eqn(23) and (24).

3. Data and Main Results

Data. 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 one parent, 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 to the early 1990s, just before the change in structure occurred.

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 and 1990s. 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 survey 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 to avoid dependent observations.

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 3,381 observations.

The means of the variables used are shown in Appendix Table B1. 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 (the state unemployment rate is also used as a conditioning variable).[12] Thirty-seven percent of the observations were on AFDC. For the budget constraint, variables for the hourly wage rate (W), nonlabor income (N), and the AFDC guarantee and tax rate (G, t, and r) are needed. To address the familiar problem of missing wages for nonworkers, a traditional selection model is estimated, with estimates shown in Appendix Table B2. The OLS estimates are almost identical to selection-adjusted estimates, so the former are used. For N, the weekly value of nontransfer nonlabor income reported in the survey is used. AFDC guarantees and tax rates by year, state, and family size are taken from estimates by Ziliak (2007), who used administrative caseload data to estimate effective guarantees and tax rates. The nominal guarantees and tax rates in the AFDC program are complex and depend on the use of numerous deductions; effective guarantees and tax rates estimated by regression methods are more accurate approximations to the parameters actually faced by recipients.[13] The mean effective tax rate on earnings across years is approximately 0.41 and that on unearned income is

[12] The empirical work will report some estimates separating the extensive margin from the intensive margin of H.

[13] See the references in Ziliak for the prior literature.

approximately 0.30.[14] The analysis also controls for the guaranteed benefit in the Food Stamp program, which was available over this period to both participants and nonparticipants in the AFDC program. The Food Stamp guarantee is set at the national level and hence varies only by family size and year, and consequently has relatively little variation in the sample used here. Those benefits are assumed to be equivalent to cash, as most of the literature suggests.

Fixed Cost Proxies, First-Stage Participation Estimates, and Instrument Power. The instruments Z_i represents fixed costs of participation that affect participation but not labor supply directly, and therefore which meet the mean independence conditions in eqns(18) and (19). Here we shall draw upon a sizable literature well known to students of the AFDC program in the 1970s and 1980s which studied non-financial administrative barriers to program participation imposed by the states. This literature, including widely cited studies by Handler and Hollingsworth (1971),Piliavin et al. (1979),Brodkin and Lipsky (1983),Lipsky (1984),Lindsey et al. (1989), and Kramer (1990), appearing mostly in social work journals, showed that caseworkers conducting eligibility assessments on individual applicants in many states were able to subjectively 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. Caseworkers often also imposed heavy paperwork requirements on recipients and used failure to complete the paperwork properly as a reason for denying applications as well as imposing administrative obstacles in general ("mechanisms to limit services...through imposing costs and inconvenience on clients" (Lipsky, 1984, p. 8)). 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, who would encourage welfare administrators to control caseloads and hence costs (Handler and Hollingsworth, 1971, p. 11). Hence the degree of discretion exercised at the caseworker level reflected the attitudes of the state

[14]The tax rate on unearned income was invariably insignificant in the empirical analysis and hence is not represented in the estimations reported in the next section.

governance structure toward the AFDC program. The use of administrative non-financial barriers was necessary because the financial rules (benefit levels, asset requirements, earnings deductions, etc.) were regulated by the federal government and hence the manipulation of financial eligibility rules to control caseloads and costs was more difficult.[15] [16]

However, the federal government did have an interest in ensuring that states were making eligibility conditions in a fair and proper manner, so they conducted audits of state eligibility decisions, visiting the states annually and drawing random samples of application records and recalculating eligibility (Hansen and Tepping (1990),U.S. House of Representatives (1994)). The federal government then calculated error rates, by state and by year, revealed by their audits. While some of the data on these error rates are published, some are unpublished but exist in the internal files of the Department of Health and Human Services and were obtained for this project.[17] The error rates collected by the auditors concern errors in assessing eligibility and as well errors related to specific mechanisms such as errors in denying applications, the reason for denial, including paperwork errors as well as errors in determining income, assets, and other financial factors affected eligibility. For almost all the categories, only error rates resulting in denial of eligibility are reported; those resulting in incorrect approval of eligibility are not. This is not a problem for the purpose at hand because if negative error rates were exactly balanced by positive error rates, the level of negative errors should have no effect on program participation in the states.

The data provide information on seven measures of state AFDC administrative actions which are potential correlates of non-financial administrative barriers: the percent of eligibility denials that were made in error, the error rate from improperly denying

[15]Nichols and Zeckhauser (1982) argued that these "ordeals" may be socially optimal because they screen out individuals with low marginal utility gains from program participation.

[16]As noted by Lurie (2006), caseworker discretion in the program became far greater after 1996 than it had been in the 1970s and 1980s.

[17]The rates which were published appear in annual issues of the publication Quarterly Public Assistance Statistics in the 1980s and 1990s.

requests for hearings and appeals, the percent of cases dismissed for eligibility reasons other than the grant amount, the overall percent of applications denied, the percent of applications denied for procedural reasons (usually interpreted as not complying with paperwork), the percent of cases resulting in an incorrect overpayment or underpayment, and the percent of cases resulting in an underpayment. There are also error rates and percents of actions related to income, assets, or employment, but these are directly or indirectly related to the applicant's labor supply and earnings level and hence are not used.

The means and distributional statistics of the seven administrative variables are shown in Table 1.[18] While the means of two of the variables are less than 1 percent, others range from 2 percent to 24 percent. The cross-state variation is also wide, with some states making underpayment errors in over 10 percent of cases, procedural denial rates of almost 35 percent, and overall denial rates of almost 50 percent.

The question is whether these state administrative barrier variables are correlated with program participation rates in the SIPP sample at hand. This can be examined by matching the state of residence of each SIPP observation to the barrier variables in that state. Column (1) of Table 2 shows probit estimates of a program participation equation using, for example, the Percent Applications Denied variable. The equation conditions on the budget constraint variables (W, N, $W(1-t)$, and G) as well as on demographics for age, race, and family composition. State-level unemployment rates are included as well as four region variables, which makes the estimates of the effect of the barrier variable arise from cross-state within-region variation. Wages and nonlabor income have expected negative effects on participation probabilities and net wages and guarantees have the expected positive effects.[19] The effects of demographic variables are also line with the literature.

[18]The administrative variables bounce around from year to year for each state and have no significant trend. The analysis therefore uses the state-specific mean values over the 1988-1992 period. The lack of meaningful variation over time also precludes the use of state fixed effects.

[19]Exploration of nonlinear effects of the guarantee revealed no significant nonlinearities, implying that positive selection occurs throughout its range.

As the table also shows, the Percent of Applications Denied has a negative effect on program participation, with a t-statistic of 1.54 and thus has the expected sign if the variable is indeed representing administrative barriers, but is weak in significance. Table 3 displays the coefficients on all seven barrier variables estimated from equations with the same specification as column (1) of Table 2. Five of the seven have coefficient point estimates that are negative in sign, consistent with an administrative barrier interpretation. The two that have positive signs are those representing payment errors rather than eligibility-related barriers and hence should, on a priori grounds, be expected to be the weakest of the variables.

While the negative signs on these instruments are consistent with an administrative barrier interpretation, there are obvious threats to this interpretation. As noted previously, the maintained assumption for this interpretation is that the state-specific variation in these variables is mean independent of the unobservables in the labor supply equations of the low income single mother population in the state. However, it should be emphasized that the budget constraint variables are in the conditioning set of the equations, so the negative signs on the state-level variables imply that two women who have the same welfare and nonwelfare budget constraints, and who therefore face the same locus of offered benefits over the hours worked range, have different probabilities of participation that are correlated with the state level variables. While it is possible that the budget constraint is measured with error, for the barrier variables to be invalid requires otherwise that caseworkers in some way directly base decisions on actual or potential hours of work of applicants. There is no qualitative evidence that caseworkers make decisions on that basis in the ground-level social work literature referenced at the beginning of this section.[20]

An indirect test of the validity of the instruments is whether the state administrative barrier variables are correlated with observables that affect labor supply. All the variables

[20]The administrative barrier variables are correlated with the political makeup of the state. In regressions not shown, the party in control of the legislature significantly affects the probability of errors in the state, suggesting political forces are partly responsible for the administrative actions.

in the participation equation will be included in the labor supply equation save the instrument itself, and all are significant correlates of single mother labor supply. Yet tests for interactions between the variables in the participation equation and the administrative barrier variables were all insignificant except those interacting the guarantee level and nonlabor income, as shown in Column (2) of Table 2. Most notably, the interaction with the hourly wage rate, the most direct determinant of labor supply, was statistically insignificant. This provides some additional, albeit indirect, evidence that the portion of caseworker discretion being picked up by the administrative barrier variables is not related to labor supply.[21]

Instrument Power. Turning to the issue of instrument power, the instruments clearly have low overall power. However, this arises because the instruments have high power in some range of participation probabilities and low power in other ranges. Figures 3 and 4 illustrate this issue. Figure 3 shows the conventional histogram of predicted participation rates obtained from the estimation, in this case using the estimates in Column (1) of Table 2 (the other instruments provide similar distributions). The propensity scores have a reasonable distribution and tail off only above participation rates around 0.60. But most of this variation is from the variables in the equation other than the instruments. Figure 4 shows, by illustration, the amount of incremental variance in predicted participation rates contributed by the instrument, shown separately for deciles of the participation probability.[22] The instrument has very little explanatory power at low and high participation probabilities and, instead, has relatively greater power in the middle range of probabilities between 0.30 and 0.80. This is no doubt because the cdf of predicted probabilities is relatively flat in the tails (i.e., at low and high probabilities) and the instruments move the probability of participation very little in those ranges.

[21]The negative coefficient on the interaction with G is consistent in a rough sense with the optimal use of administrative barriers suggested by Nichols and Zeckhauser (1982), whose argued that imposing such barriers and reducing the caseload would allow higher benefits to be paid to those who are on the welfare rolls.

[22]The deciles of the baseline predicted probability on the horizontal axis are obtained from estimates of the equation without the instrument.

A more formal way to express instrument power in different ranges of participation probabilities is to compute F-statistics separately by range. F-statistics computed separately by quartile of the predicted probability distribution for the Percent of Applications Denied instrument are shown at the bottom of Column (1) of Table 2. While the conventional OLS F-statistic is 1.81, the statistic rises to 3.11 in the range of participation rates between .25 and .50.[23] Further, the greater power in the .25-.50 range is primarily because the mean participation rate in the sample is .37, and what the instrument does in this case is to provide its major explanatory power in an interval around that mean. Column (2) shows, instead, the results of tests for interactions between the instrument and the other variables in the equation, which show maximum incremental explanatory power obtained when the instrument is interacted with the welfare guarantee and private nonlabor income. The F-statistics rise to above 4 in the .50-.75 range and above 6 in the .25-.50 range.

Table 4 shows similar results for the other instruments, excluding the last two in Table 3, with each specification representing that which yielded the largest F statistics in the middle ranges of participation probabilities.[24] Compared to the Percent of Applications Denied variable, the first instrument (Percent of Ineligibles made in error) has better fit in the .25-.50 range but slightly worse fit in the .50-.75 range. The other three instruments have lower fit in the .25-.50 range and varying fits in the .50-.75 range. The analysis in the next section will therefore estimate the MTE curve only in the .25-.75 range and will lead with the Percent of Applications Denied instrument, but results using the other four instruments, especially that for Percent of Ineligibles In Error, will be presented as well. As the estimates will show, all five instruments yield the same shape of the MTE curve but with slightly different confidence intervals.[25]

[23]An F-statistic defined in this way can be negative if adding the instrument to the equation reduces the fit in one region while increasing it in others.

[24]A large number of other specifications were tested, including additional interactions with the other regressors, polynomials in the instruments, and entering multiple instruments. None yielded F statistics as high as those shown in the table.

[25]That all five instruments give approximately the same MTE results suggest that they are all picking up

While the instruments therefore do not provide sufficient power over the full range of participation probabilities, the .25-.75 range covers the participation rates which arose in the three historical AFDC reform episodes which are the final object of the paper. Consequently, for the purposes at hand, extrapolation to other ranges and other historical episodes where participation rates lay outside this range is not conducted.

Results of Hours Equation Estimation. Estimation of eqn(23) using the fitted values of the participation probabilities for F yields estimates of β, λ, and the parameters of the g function. The g function is estimated with conventional cubic splines, hence $g(F) = g_0 + \sum_{j=1}^{J} g_j Max(0, F - \pi_j)^3$, where the π_j are preset spline knots. The initial estimates use five knots approximately equally spaced over the range (.25,.75) but estimates using fewer and a greater number of knots are obtained. Generalized cross-validation statistics (GCVs) are used as a measure of goodness of fit. Given the well-known tendency of polynomials to reach implausible values in the tails of the function and beyond the range of the data, natural splines are typically used, which constrain the function to be linear before the first knot and beyond the last knot (Hastie et al. (2009)). Imposing linearity on the function in those two intervals requires modifying the spline functions to accommodate this; the exact spline functions for a five-knot spline are shown in Appendix C.[26]

The full estimates of the hours equation are shown in Appendix Table B3 and the key coefficient estimates are shown in column (1) of Table 5 for the basic model.[27] The first four rows show the λ coefficients on the four budget constraint variables which, as noted in a previous section, must be conditioned on for the coefficient on $\hat{F}$ in the g function identify marginal treatment effects. The standard errors are high for all variables except nonlabor income, which has a positive effect on g, implying that increases in the participation rate have a smaller negative effect on hours of work for those with higher

common procedures across the states. This is also probably why entering two instruments at the same time into the participation equation yielded little additional explanatory power (see previous footnote).

[26]Consistency of sieve methods is discussed by Chen (2007).

[27]Standard errors are calculated by bootstrapping the hours equation, the participation equation, and the wage equation jointly.

levels of such income. Higher levels of nonlabor income result in a lower potential benefit to participation and, ceteris paribus, to a lower reduction in labor supply. But higher nonlabor income also reduces the level of labor supply off welfare (see its negative coefficient in Appendix Table B3) and it is plausible that those with lower levels of initial labor supply choose to take more of their benefit in the form of additional consumption than in more leisure upon participation.

Although these results imply relatively little heterogeneity in response by most of the budget constraint variables, the cubic spline variables in the g function imply non-trivial heterogeneity in labor supply response from increases in participation induced by reductions in fixed costs-that is, the function g is not constant w.r.t. the participation rate. The coefficients on the splines are not easily interpretable, so the third, five-knot panel of Figure 5 shows the marginal labor supply responses corresponding to those estimates and their 95 percent confidence intervals for the region where the instruments have power (participation rates between .25 and .75).[28] The marginal responses are U-shaped and non-monotonic, starting off at F=.25 insignificantly different from 0 but then growing in (negative) size as participation increases. The marginal response peaks at a participation probability of about .35, when it reaches -31 hours per week, and then declines, becoming insignificantly different from 0 at approximately F=.47. The point estimate approaches zero as participation rises further but remains insignificantly different from 0 for all higher participation levels.

The sensitivity of the results to the number of knots chosen is also demonstrated in the Figure with the MTEs for 3, 4, and 6 knots. When the smaller numbers of knots is used, the MTE estimates are monotonic in the participation rate and decline over the range shown. However, this is because an insufficient number of knots is used to pick up the shape of the MTE curve over the early participation ranges. When 6 knots are used, the shape of the MTE curve changes very little, only adding a small new submode in the

[28]The MTE function is, as noted previously, just the derivative of the hours equation w.r.t the participation rate. All MTE curves are evaluated at the means of the other variables in the model.

middle of the significant range. Further additions of knots only serve to add additional small bumps in the curve without changing its general shape.[29]

Columns (2) and (3) of Table 5 test expansions of the specification in two ways. Column (2) tests whether the budget constraint variables affect the shape of the MTE curve and not just its level–the specification in Column (1) just shifts the MTE curve up and down in parallel fashion. But the interactions of the budget constraint variables with the participation probability have large standard errors, indicating that no significant heterogeneity in shape is present. The third column presents estimates of a specification with two additional λ variables, obtaining after testing for the presence of effects of all the other variables in the hours equation. Both age and race (Black) significantly affect the MTE, with older women having smaller labor supply disincentives in response to rising participation and Black women also having smaller disincentives. The estimates in Column (3) will be used for the rest of the analysis, and the MTE curve for that specification is shown in Figure 6 and is essentially identical to that in Figure 5.

Estimation of a homogeneous effects model, equivalent to specifying the g function as a constant, yields a point estimate of -31.5 hours per week (s.e.=5.9). As is well known, linear IV assigns weights to the different MTEs at different points in the propensity score distribution (Heckman and Vytlacil (1999); Angrist et al. (2000); Heckman and Vytlacil (2001, 2005)). In this application, the weights are concentrated around .35. Linear IV would therefore give a wildly distorted picture of how marginal responses vary and would completely miss the U-shaped response function which actually occurs.

Figure 7 shows the MTE estimates for the other four instruments shown in Table 4. For three of them–the Percent of Hearings and Appeals Improperly Denied, the Percent of Cases Denied for non-Grant Reasons, and the Percent of Applications Denied for Procedural Reasons–the MTE estimates are very close to those in Figure 6. Significant MTEs occur in the approximate range of participation probabilities (.26, .47), the MTE

[29] The GCV statistics (not shown) achieve their minimum at 5 knots.

curve is U-shaped, and the maximum labor supply response is approximately -30. For one of the instruments–for the Percent of Applications Denied in Error–the significance band is narrower, in the (.26, .37) range and the maximum response is also lower, about -21 hours. However, the shape of the MTE curve is, like the others, U-shaped and hence also implies growing then declining marginal labor supply responses as the caseload grows.

The point estimates for the marginal labor responses are often relatively large, peaking at approximately 30 hours per week for the best-fitting specifications and most of the instruments. While these effects are large, they occur only in a specific part of the participation probability distribution and therefore only at certain caseload levels. Some insight into the mechanics behind the U-shaped pattern of responses can be gleaned by examining where in the distribution of hours the responses come from over different ranges–in particular, by examining how individuals reduce hours from 40 per week or 20 per week or to lower levels, including nonwork. That movements between full-time work, part-time work, and nonwork may be important is demonstrated in Table 6, which shows the distribution of welfare participants and non-participants across the hours categories. What is striking about the table is that welfare participation is essentially equivalent to not working, with almost no participants working part-time and even fewer working full-time. Among non-recipients, the distribution is the opposite, with almost no one not working and over 80 percent working full-time. While these distributions are not causal, they suggest that being off welfare is generally associated with working full-time and being on welfare is generally associated with not working, and that some of those who go onto welfare may reduce their hours by 40 per week.

Evidence suggesting this is the case is shown in Figure 8, which shows the result of estimating the hours worked equation by successively replacing the dependent variable for H with dummies for not working, working part-time, and working full-time. The Figure shows the MTEs for those regressions. The leftmost panel shows that the probability of nonwork rises sharply as participation goes from .25 to .35, the same range where the MTE for average hours falls the most. The middle panel shows that the MTE for part-time work actually starts off at a positive level, implying an increase in part-time work that can only come from full-time workers reducing labor supply to the part-time level. The part-time MTE becomes less positive as participation increases and eventually becomes negative, implying that some part-timers move at that point to nonwork. But the right panel shows that the MTE for full-time work is large and negative in the .25 to .35 participation rate range implying, when combined with the other panels, that all of the reduction in labor supply over that range is from full-time work to nonwork upon participation, which is where the prior figures show the maximum reduction occurs. Eventually, however, after participation rises high enough, movements out of full-time work fall to zero. Thus the decline in the labor supply reductions in average hours when participation rates rise sufficiently reflects a decline in movements out of full-time work.

Further evidence that it is the high-hours-worked individuals who participate "early" (i.e., when administrative barriers and fixed costs are high and hence participation is low) who are responsible for the large marginal effects in the lower ranges of the participation rate distribution is shown in Table 7, which displays a few labor-supply related variables by quartile of the fitted propensity score distribution. Those who are on the margin at low participation probabilities have higher wage rates, are less likely to be black, are older, and have fewer young children, all of which are correlated with higher levels of work. Nonlabor income is higher for the early participants as well, which is typically correlated with lower levels of labor supply but, for discrete moves from full time work to nonwork, this means that those individuals also have a larger income cushion if they do not work. Those who are on the margin at higher participation rates have lower wages, are more likely to be black, are younger, and have more children, all of which are correlated with lower levels of work and hence lower marginal effects of labor supply upon participation.[30]

[30]It may be worth noting that the theoretical model outlined previously did not imply that initial participants (i.e., those joining when fixed costs are high) could not be those with higher labor supply than later participants, only that initial participants are those with the highest utility gains from participation. It is quite possible that those mothers with higher wage rates and working full-time have the greater marginal

4. Marginal Effects of Major AFDC Reforms

The AFDC program has experienced three major changes in the tax rate on benefits over its history. From its creation in 1935 to 1967, the tax rate was 100 percent. This high tax rate was the subject of well-known criticisms of the program by Friedman (1962), Lampman (1965), and Tobin (1966) for its resulting work disincentives. In 1967, Congress lowered the tax rate to 67 percent in order to provide work incentives to AFDC participants. However, the Reagan Administration in its early days in 1981, based on a prior reform in California when Reagan was governor, concluded that low tax rates just increased the caseload and hence costs without any significant work incentives. At the Administration's recommendation, the tax rate in the program was raised back to 100 percent by Congress. A reversal of this decision took place in 1996, when major welfare program legislation transformed the AFDC program into a more pro-work program with work requirements and time limits. As part of that reform, states were allowed to set their own tax rates rather than have them federally mandated, and most states chose to implement major reductions. On average, the tax rate after the reform was approximately 50 percent.

A simple model of labor supply responses without much heterogeneity would predict that the 1981 tax rate increase would just reverse the labor supply effects of the 1967 tax reduction, and that the 1996 reduction would have effects similar to those of the 1967 reduction, although presumably slightly larger given the larger magnitude of the reduction. However, the participation rate in the program was very different in the three reform years. That rate was modest, around .36 in 1967, but rose in the late 1960s and early 1970s before leveling off (Moffitt (1992)). By 1981, the participation rate was just over .50. In the 1980s, the participation rate began to decline, reaching the .37 level reported above but then rising again in the early 1990s. By 1996, the participation rate had risen back to .40 (Ziliak (2016)). Because marginal labor supply effects differ depending on the participation rate,

utility gains from increased leisure.

marginal effects should have therefore been different at each of these historical periods.

In addition to differences in participation rates, real guarantees were very different in the three years. Guarantees were very high in the 1960s and in 1967 in particular but, over the latter half of the 1970s and early 1980s, they were allowed to fall in real terms as state legislatures failed to raise the nominal amounts sufficiently to offset inflation. By 1981, guarantees were 30 percent lower than they had been in 1967. But over the early 1990s, states began raising guarantee levels again and, by 1996, they had reached a level about halfway between their 1967 high level and their 1981 low level. Thus guarantee levels were also different in the different years, as were the initial tax rates at the time the tax-rate reforms took place. Since the model shows that marginal effects depend on the initial levels and guarantees, since those affect the composition of the recipient population, marginal labor supply effects could also differ across periods for this reason.

To estimate the effects of these factors, Current Population Survey (CPS) files were obtained for 1967, 1981, and 1996. All demographic variables in the estimated participation and hours equations were constructed for each of those years from the CPS data. The levels of G and t in those years were also obtained. Using the estimated participation equation from the SIPP data for 1988-1992 as reported in the last sections, the effects of changes in demographics as well as changes in the guarantees and tax rates on program participation between 1988-1992 and each of those other years on the participation rate could be calculated. Finally, using the fitted model of marginal labor supply effects reported in the last section, those marginal effects could be computed for 1967, 1981, and 1996.

The results are reported in Table 8. In 1967, the participation rate was .36, not very different from the .37 value in 1988-1992. But this was the result of offsetting effects of differences in demographics in 1967 and 1988-1992, which pushed the participation rates down by 5 percentage points, and the higher guarantees in 1967, which pushed the rate up by 4 percentage points. At the participation rate of .36, the marginal individual had a labor supply effect of -31 hours, with a wide confidence interval but bounded away from zero.

But things were quite different in 1981, when the participation rate had risen to .53 and guarantees had fallen to one-third of their 1967 level. The decline in guarantees pushed the participation rate down by 5 percent points but the demographics were about the same as in 1988-1992. The higher participation rate was due to other factors and not to other observables in the data. The marginal response in 1981 was -13 hours per week and insignificantly different from zero. Thus raising the tax rate back to its 1967 level did not have opposite marginal effects because the participation rate and the guarantee level were different in 1981 than in 1967.

By 1996, the participation rate had fallen to .40 and guarantees had also risen by 20 percent from their 1988-1992 values. The increase in the guarantee pushed up the participation rate by 5 percentage points. Changes in demographics added another 2 percentage points. At the .40 participation rate, the marginal response was about -28 hours, and hence had risen most of the way back to its 1967 level.[31]

Simulations for marginal responses in years later than 1996 cannot be conducted with the model estimated in this paper because the program no longer took the simple form which the model represents. However, participation rates in the program (now called TANF) are known to be approximately 10 to 15 percent. Ignoring the other differences in the TANF and AFDC programs, this would imply that the hypothetical marginal response to an increase in participation at the current time would be insignificantly different from zero.

5. Summary

This paper has provided a model and a reduced form estimation method for nonparametrically analyzing the marginal labor supply response in a classic transfer program of the textbook negative income tax type. Applying the model to the Aid to Families with Dependent Children in the late 1980s and early 1990s shows that marginal

[31] This simulation ignores all the structural changes in the program that occurred in 1996 and hence is only a hypothetical marginal response that would have occurred in the absence of those other reform elements.

labor supply respones are non-montonic and quadratic, with the magnitude of the marginal response increasing as participation rates increase but eventually declining after participation rates pass an inflection point. Marginal responses are insignificantly different from zero at low and high participation rates. Using the estimates to estimate marginal responses at three historical years when major reforms of the program took place shows that marginal responses were different in each year, both because the demographic composition of the caseload was different, the level of the program parameters was different, and because a different fraction of the population was participating in the program. The largest marginal response was in 1967 when participation rates were fairly low and guarantees were fairly high. The lowest marginal response was in 1981, when participation rates were high and guarantees were low and, in that year, a 95 percent confidence interval includes zero. The marginal response in 1996 fell in between those in the other two years.

A number of obvious extensions of the analysis would be worthwhile. One is to estimate a structural model which pins down the underlying parameters of a formally defined utility function whose parameters vary in the population. That would allow a better analysis of counterfactuals than can the method used here. Another is to extend the static model to dynamic models where dynamics are introduced through intertemporal elasticities of labor supply, human capital, and preference persistence (Chan and Moffitt (2018)). Yet another avenue for more model development is to add an analysis of inframarginal responses to transfer program reforms to the analysis of marginal responses, since any reform involving alteration of program parameters affects both.

There are also many programs of interest other than the simple negative-income-tax cash program type analyzed here. The 1996 reform of the AFDC program introduced work requirements, time limits, and other features, which have been show to have had effects on average labor supply (Chan (2013)). Their marginal effects are likely to be quite different than those analysed here because those reforms almost surely affected different portions of the labor supply preference distribution. In addition, the participation rate in the program has dropped by 80 percent, which surely affects who is on the margin of participation in the program. The analysis of the marginal responses to in-kind transfers, which requires modeling the consumption of the subsidized good jointly with labor supply, is another obvious extension given the expansion of those types of transfers in the U.S. over the last 30 years.

References

Angrist, J. D., K. Graddy, and G. Imbens (2000, July). The Interpretation of Instrumental Variable Estimators in Simultaneous Equations Models with an Application to the Demand for Fish. *Review of Economic Studies 67*(3), 499–527.

Björklund, 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.

Brinch, C. N., M. Mogstad, and M. Wiswall (2017, August). Beyond LATE with a Discrete Instrument. *Journal of Political Economy 125*(4), 985–1039.

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.

Currie, J. (2006). The Take-Up of Social Benefits. In A. J. Auerbach, D. Card, and J. M. Quigley (Eds.), *Public Policy and the Income Distribution*, pp. 80–148. New York: Russell Sage Foundation.

Daponte, B. O., S. Sanders, and L. Taylor (1999). Why Do Low-Income Households Not Use Food Stamps? Evidence from an Experiment. *Journal of Human Resources 34*(3), 612–28.

Friedman, M. (1962). *Capitalism and Freedom*. Chicago: University of Chicago Press.

Handler, J. F. and E. J. Hollingsworth (1971). *The Deserving Poor: A Study of Welfare Administration*. Chicago: Markham.

Hansen, M. H. and B. J. Tepping (1990). Regression Estimates in Federal Welfare Quality Control Programs. *Journal of the American Statistical Association 85*(411), 856–864.

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 (2001). Local Instrumental Variables. In C. Hsiao, K. Morimune, and J. Powell (Eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*. New York: Cambridge University Press.

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

Heckman, J. J. and E. Vytlacil (2007). Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimates to Evaluate Social Programs and to Forecast Their Effects in New Environments. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6. Amsterdam: Elsevier.

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

Keane, M. and R. Moffitt (1998, August). A Structural Model of Multiple Welfare Program Participation and Labor Supply. *International Economic Review 39*(3), 553–89.

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.

Lampman, R. (1965, May). Approaches to the Reduction of Poverty. *American Economic Review 55*, 521–29.

Lindsey, E. W., S. Colosetti, B. Roach, and J. S. Wodarski (1989). Quality Control and Error Reduction in the AFDC Program: A Review and Synthesis of State Strategies. *Administration in Social Work 13*(2), 29–45.

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

Lurie, I. (2006). *At the Front Lines of the Welfare System.* Albany, NY: Rockfeller Institute Press.

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.

Mogstad, M., A. Santos, and A. Torgovitsky (2018, September). Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica 86*(5), 1589–1619.

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.

Nichols, A. L. and R. J. Zeckhauser (1982). Targeting Transfers Through Restrictions on Recipients. *American Economic Review 72*(2), 372–77.

Piliavin, I., S. Masters, and T. Corbett (1979, August). Administration and Organizational Influences on AFDC Case Decision Errors: An Empirical Analysis. Institute for Research on Poverty Discussion Paper 542-79.

Tobin, J. (1966, Fall). On the Economic Status of the Negro. *Daedalus 94*(4), 878–98.

U.S. House of Representatives (1994). *Background Material and Data on Programs Within the Jurisdiction of the Committee on Ways and Means*. Washigton, D.C.: U.S. Government Printing Office.

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 A. Illustrative Structural Model

Assume the utility function is quadratic and therefore is linear in the taste parameters, and assume that all parameters are potentially heterogeneous (Keane and Moffitt (1998)):

$$U(H, Y, P \mid Z) = -\alpha_i H - \beta_i H^2 + Y - \delta_i Y^2 - (m(Z) + \nu_i)P \tag{25}$$

where, as in the text, Y=income and P=a welfare participation indicator. Define $G(\alpha_i, \beta_i, \delta_i, \nu_i)$ as the joint cdf of the four heterogeneity parameters and assume that G is defined over the region of the parameters which make the utility function quasi-concave. The parameters of this function are the fundamental structural parameters in the model. Assume, in line with most of the structural literature, that individuals have a three-dimensional discrete choice of weekly hours of work, at $H = 0$, $H = 20$, and $H = 40$. The model therefore presumes a six-dimensional discrete choice problem over the six choices defined by H, P combinations.

Following the development in the text, define $I_H(P)$ as the event that hours level H is chosen conditional on a P choice. Then, definitionally,

$$H = 20[I_{20}(1)P + I_{20}(0)(1-P)] + 40[I_{40}(1)P + I_{40}(0)(1-P)] \tag{26}$$

$$= [20I_{20}(0) + 40I_{40}(0)] + \{20[I_{20}(1) - I_{20}(0)] + 40[I_{40}(1) - I_{40}(0)]\}P \tag{27}$$

Computing the mean of H conditional on Z, we have

$$E(H|Z) = 20E[I_{20}(0)] + 40E[I_{40}(0)] + \{20[E(I_{20}(1)|P=1) - E(I_{20}(0)|P=1] + 40[E(I_{40}(1)|P=1) - E(I_{40}(0)|P=1)]\}\Pr(P=1|Z) \tag{28}$$

using the same identifying assumptions discussed in the text. This equation is the structural counterpart to eqn(22).

The dependence of eqn(28) on the structural parameters works through the means and conditional means of the six $I_H(P)$ indicators. Define the function

$$d(\alpha_i, \beta_i, \delta_i, \nu_i \mid \widetilde{H}, \widetilde{P}, H, P, Z_i) = -\alpha_i(\widetilde{H} - H) - \beta_i(\widetilde{H}^2 - H^2) + [Y(\widetilde{H}, \widetilde{P}) - Y(H, P)] - \delta_i[Y(\widetilde{H}, \widetilde{P})^2 - Y(H, P)^2] - m(Z_i)(\widetilde{P} - P) - \nu_i(\widetilde{P} - P) \quad (29)$$

where $Y(H, P)$ is income as defined from the budget constraint for an individual with hours choice H and participation choice P. Then

$$Pr(I_0(1) = 1, P = 1 \mid Z_i) = Pr[d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 0, 1, 0, 0, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 0, 1, 20, 0, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 0, 1, 40, 0, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 0, 1, 20, 1, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 0, 1, 40, 1, Z_i) \geq 0] \quad (30)$$

$$Pr(I_{20}(1) = 1, P = 1 \mid Z_i) = Pr[d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 20, 1, 0, 0, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 20, 1, 20, 0, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 20, 1, 40, 0, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 20, 1, 0, 1, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 20, 1, 40, 1, Z_i) \geq 0] \quad (31)$$

$$Pr(I_{40}(1) = 1, P = 1 \mid Z_i) = Pr[d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 40, 1, 0, 0, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 40, 1, 20, 0, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 40, 1, 40, 0, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 40, 1, 0, 1, Z_i) \geq 0, d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 40, 1, 20, 1, Z_i) \geq 0] \quad (32)$$

and

$$Pr(P = 1 \mid Z_i) = Pr(I_0(1), P = 1) + Pr(I_{20}(1), P = 1) + Pr(I_{40}(1), P = 1) \quad (33)$$

Then

$$E(I_{20}(1) \mid P = 1) = \frac{1}{Pr(P = 1)} Pr(I_{20}(1) = 1, P = 1) \tag{34}$$

$$E(I_{40}(1) \mid P = 1) = \frac{1}{Pr(P = 1)} Pr(I_{40}(1) = 1, P = 1) \tag{35}$$

Define S as the set of $\alpha_i, \beta_i, \delta_i, \nu_i$ for which $P = 1$. Then define

$$\begin{aligned} Pr(I_{20}(0) = 1, P = 1 \mid Z_i) = Pr[&d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 20, 0, 0, 0, Z_i) \geq 0, \\ &d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 20, 0, 40, 0, Z_i) \geq 0, (\alpha_i, \beta_i, \delta_i, \nu_i) \in S) \end{aligned} \tag{36}$$

$$\begin{aligned} Pr(I_{40}(0) = 1, P = 1 \mid Z_i) = Pr[&d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 40, 0, 0, 0, Z_i) \geq 0, \\ &d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 40, 0, 20, 0, Z_i) \geq 0, (\alpha_i, \beta_i, \delta_i, \nu_i) \in S) \end{aligned} \tag{37}$$

Then

$$E(I_{20}(0) \mid P = 1) = \frac{1}{Pr(P = 1)} Pr(I_{20}(0) = 1, P = 1) \tag{38}$$

$$E(I_{40}(0) \mid P = 1) = \frac{1}{Pr(P = 1)} Pr(I_{40}(0) = 1, P = 1) \tag{39}$$

The unconditional means of working $H = 20$ and $H = 40$ in the absence of welfare are

$$\begin{aligned} E[I_{20}(0)] = Pr[&d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 20, 0, 0, 0, Z_i) \geq 0, \\ &d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 20, 0, 40, 0, Z_i) \geq 0)] \end{aligned} \tag{40}$$

$$\begin{aligned} E[I_{40}(0)] = Pr[&d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 40, 0, 0, 0, Z_i) \geq 0, \\ &d(\alpha_i, \beta_i, \delta_i, \nu_i \mid 40, 0, 20, 0, Z_i) \geq 0)] \end{aligned} \tag{41}$$

which do not depend on ν given the additive separability of P in the utility function. This completes the structural expression of all terms in eqn(28).

Appendix B. Additional Tables

Appendix Table B1
Means of the Variables Used in the Analysis

	Full Sample	P=1	P=0
Weekly H	21.4	4.5	31.1
P	0.37	1.0	0.0
Ln W (predicted)	1.79	1.74	1.81
Ln Weekly N	2.97	2.58	3.19
Ln G/100	-2.49	-2.38	-2.55
Ln W(1-t)	1.26	1.22	1.30
Age	32.5	30.3	33.8
Black	0.34	0.41	0.30
Education	10.9	10.5	11.1
Family size	3.1	3.4	3.0
No. Children Less Than 6	0.79	1.14	0.58
Food Stamp Guarantee /100	0.78	0.78	0.78
Unemployment rate	6.4	6.4	6.3
Northeast	0.28	0.28	0.28
Midwest	0.27	0.27	0.26
West	0.22	0.25	0.20
State Percent Services	27.9	28.2	27.8
State Percent Manufacturing	15.2	15.1	15.2
State Percent Urban	76.3	77.5	75.6
Pct. App. Den. Error Rate	0.24	0.12	0.05

Notes:
N = 3381
All dollar-denominated variables are in 1990 PCE dollars.

Appendix Table B2
Log Hourly Wage Equation Estimates

	OLS	Selection-Bias Adjusted
Age	.014 (.001)	.007 (.002)
Education	.046 (.007)	.040 (.007)
Black	-.091 (.027)	.012 (.031)
Northeast	.197 (.048)	0.259 (.051)
Midwest	.087 (.040)	.073 (.042)
West	.116 (.043)	.179 (.046)
State Percent Services	.019 (.008)	.021 (.008)
State Percent Manufacturing	.005 (.004)	.009 (.004)
State Percent Urban	.003 (.001)	.003 (.001)
Constant	-.070 (.228)	.310 (.233)

Notes:
Standard errors in parentheses.

Appendix Table B3
Estimates of Hours Equation with Five-Knot g Spline

	(1)	(2)
λ [1]		
Ln W	-11.99	-20.55
	(14.69)	(15.05)
N	0.22	0.21
	(0.10)	(0.10)
Ln G	-2.43	-0.67
	(4.92)	(5.03)
Ln W(1-t)	-1.38	2.37
	(8.84)	(8.94)
Age	--	0.50
		(0.24)
Black	--	5.64
g		(3.40)
Constant/10	32.2	30.9
	(8.8)	(8.1)
$\hat{F}/100$	-26.9	-26.2
	(6.5)	(6.5)
N3/1000	65.6	64.3
	(16.3)	(16.3)
N4/1000	-90.2	-88.4
	(22.6)	(22.6)
N5/1000	24.9	24.4
	(6.9)	(6.9)
β		
Ln W	23.35	26.41
	(4.87)	(5.52)
Ln (N+10)	-2.83	-2.89
	(1.12)	(1.13)
Age	-0.10	-0.27
	(0.06)	(0.11)
Black	-0.59	-2.75
	(0.80)	(1.57)
Family Size	-0.82	-1.03
	(0.37)	(0.38)
No. Children Less than 6	-1.70	-1.50
	(0.80)	(0.82)
Food Stamp Guarantee	-11.2	-9.9
	(16.3)	(16.3)
Unemployment Rate	-0.64	-0.62
	(0.29)	(0.29)
Northeast	-8.91	-9.54
	(1.96)	(2.00)

Appendix Table B3 (continued)
Estimates of Hours Equation with Five-Knot g Spline

	(1)	(2)
β		
Midwest	-1.86	-2.35
	(1.45)	(1.47)
West	-4.65	-5.20
	(1.83)	(1.85)
Constant	14.67	15.98
	(15.53)	(15.54)

Notes:
For spline variable definitions, see Appendix C.
[1] Variables expressed as deviations from means.
Standard errors in parentheses.

Appendix C. Cubic Spline

The five-knot natural cubic spline is given here, using the same notation as Hastie et al. (2009)[p. 145]. Splines using different numbers of knots are analogous. Let F_1, F_2, F_3, F_4, and F_5 denote the five knot points of $\hat{F}$, the predicted participation probability. The g function is specified as

$$g(\hat{F}) = g_1 + g_2\hat{F} + g_3N3 + g_4N4 + g_5N5 \tag{42}$$

where

$$N3 = d_1 - d_4 \tag{43}$$

$$N4 = d_2 - d_4 \tag{44}$$

$$N5 = d_3 - d_4 \tag{45}$$

where

$$d_1 = \frac{Max(0, \hat{F} - F_1) - Max(0, \hat{F} - F_5)}{F_5 - F_1} \tag{46}$$

$$d_2 = \frac{Max(0, \hat{F} - F_2) - Max(0, \hat{F} - F_5)}{F_5 - F_2} \tag{47}$$

$$d_3 = \frac{Max(0, \hat{F} - F_3) - Max(0, \hat{F} - F_5)}{F_5 - F_3} \tag{48}$$

$$d_4 = \frac{Max(0, \hat{F} - F_4) - Max(0, \hat{F} - F_5)}{F_5 - F_4} \tag{49}$$

Table 1

Seven Administrative Barrier Variables

	Mean	Stnd Dev	Min	Max
Percent ineligible in error	1.7	0.8	0.3	4.7
Percent hearings and appeals improperly denied	2.0	1.5	0.4	5.8
Percent cases elig. denied for non -grant reasons	0.2	0.1	0.0	0.4
Percent applications denied	24.3	11.2	5.3	47.8
Percent applications denied for procedural reasons	14.1	8.9	1.3	34.6
Error rate in payment determination	4.7	1.1	2.2	7.0
Error rate resulting in underpayment	3.5	2.8	1.4	10.6

Notes:

Statistics are taken over all states in the sample and equal the mean percents over all years 1988-1992 for each state.

Source: Unpublished data, U.S. Department of Health and Human Services.

Table 2
Probit Estimates of the Participation Equation
Using Percent Applications Denied

	(1)	(2)
Log W	-2.81	-3.01
	(0.37)	(0.34)
Log (N+10)	-0.44	-0.30
	(0.03)	(0.06)
Log G	0.89	1.31
	(0.14)	(0.19)
Log W(1-t)	0.97	0.26
	(0.28)	(0.15)
Age	0.01	0.01
	(0.01)	(0.01)
Black	0.14	0.15
	(0.05)	(0.06)
Family size	-0.05	-0.05
	(0.03)	(0.03)
Number of Children Less than 6	0.30	0.30
	(0.03)	(0.03)
Food Stamp Guarantee	1.98	1.97
	(1.28)	(1.28)
State Unemployment Rate	0.04	0.04
	(0.02)	(0.02)
Northeast	0.29	0.14
	(0.15)	(0.11)
Midwest	0.14	0.18
	(0.11)	(0.14)
West	0.12	1.08
	(0.15)	(0.29)

Continued on next page.

Table 2 (continued)

	(1)	(2)
Pct. App. Denied	-0.51	0.01
	(0.33)	(0.01)
Pct. App. Denied*G	--	-0.18
		(0.06)
Pct. App. Denied*N	--	-0.00
		(0.00)
Constant	4.61	5.51
	(1.03)	(1.11)
GCV	323.7	323.3
OLS F-statistic for instruments	1.81	4.80
Pseudo F-statistic by Part. Prob. range		
0-.25	0.45	-1.22
.25-.50	3.11	6.06
.50-.75	-0.32	4.45
.75-1.00	0.12	0.59

Notes:
N=3381
Standard errors in parentheses.
Pseudo F-stat: Define RSS(q) as the residual sum of squares, equal to the sum of [P-$\hat{F}$]**2 taken over all observations in a specific quartile range of $\hat{F}$. The pseudo F-stat is calculated as (1) the difference in RSS(q) for the restricted model excluding the instruments and the unrestricted model RSS(q) including the instruments divided by the d.o.f., divided by (2) the residual variance computed over all observations in the sample, using $\hat{F}$ from the restricted model.

Table 3
Probit Coefficients on Administrative Barrier Variables
In Program Participation Equations

Percent ineligible in error	-.025
	(.037)
Percent hearings and appeals improperly denied	-.011
	(.017)
Percent cases elig. Denied for non-grant reasons	-.571
	(.400)
Percent applications denied	-.051
	(.033)
Percent applications denied for procedural reasons	-.334
	(.356)
Error rate in payment determination	.017
	(.026)
Error rate resulting in underpayment	.001
	(.011)

Notes:
Standard errors in parentheses.
Equations contain all regressors shown in Table 2.

Table 4
First Stage Participation Probit Instrument
Coefficient Estimates and Instrument Power

	Pct. Ineligible In Error	Pct. Hear. & Appeals Denied	Pct. Cases Denied Non-Grant	Pct. Applications Denied	Pct. Apps. Denied Procedural
Main effect	0.020	0.007	-0.028	0.011	0.098
	(0.039)	(0.019)	(0.041)	(0.006)	(0.063)
N interaction	-0.004	-0.002	-.002	-0.001	--
	(0.001)	(0.001)	(.001)	(0.000)	
G interaction	--	--	--	-0.176	-0.189
				(0.060)	(0.075)
Pseudo F-stat by Part. Prob.					
0-.25	1.08	-0.10	0.41	-1.22	-1.12
.25-.50	7.76	3.80	3.45	6.06	3.05
.50-.75	3.95	0.82	2.36	4.45	3.56
.75-1.00	0.38	0.21	0.10	0.59	-0.11

Notes:
All regressions include log hourly wage rate, log nonlabor income, G, log net wage W(1-t), age, black, number of family members, number of children under 18, region dummies, state unemployment rate, and Food Stamp guarantee.
Pseudo F-stat: See footnote to Table 2.
Standard errors in parentheses.

Table 5
Estimates of Hours Equation λ and g Coefficients
with Five-Knot g Spline

	(1)	(2)	(3)
λ[1]			
Ln W	-11.99	-1.47	-20.55
	(14.69)	(45.21)	(15.05)
N	0.22	0.37	0.21
	(0.10)	(0.22)	(0.10)
Ln G	-2.43	-1.17	-0.67
	(4.92)	(15.12)	(5.03)
Ln W(1-t)	-1.38	12.40	2.37
	(8.84)	(26.99)	(8.94)
Age	--	--	0.50
			(0.24)
Black	--	--	5.64
g			(3.40)
Constant/10	32.2	30.0	30.9
	(8.8)	(9.1)	(8.1)
$\hat{F}/100$	-26.9	-25.7	-26.2
	(6.5)	(6.6)	(6.5)
N3/1000	65.6	63.4	64.3
	(16.3)	(16.5)	(16.3)
N4/1000	-90.2	-87.2	-88.4
	(22.6)	(22.8)	(22.6)
N5/1000	24.9	24.1	24.4
	(6.9)	(65.4)	(6.9)
Interactions			
Ln W* $\hat{F}$	--	-8.81	--
		(54.8)	
(N+10)* $\hat{F}$	--	-0.30	--
		(0.37)	
Ln G* $\hat{F}$	--	-2.55	--
		(18.95)	
Ln W(1-t)* $\hat{F}$	--	-23.43	--
		(39.6)	

Notes:
For spline variable definitions, see Appendix C.
Estimated β coefficients shown in Appendix Table B4.
[1] Variables expressed as deviations from means.
Standard errors in parentheses.

Table 6

Percent Working 0, 20 and 40 Hours
by Welfare Participation Status

	H=0	H=20	H=40
All	40.9	12.3	46.7
P=0	17.9	13.4	68.7
P=1	82.9	10.3	6.8

Table 7

Variable Means for Observations in Quartiles
of the $\hat{F}$ Distribution

	2nd Quartile	3rd Quartile	4th Quartile
Hourly wage	\$6.47	\$6.10	\$5.38
Weekly nonlabor income	\$25.2	\$6.8	\$4.1
Black	0.26	0.34	0.43
Age	35.5	32.4	29.6
Child<6	0.39	0.54	0.89

Table 8

Marginal Labor Supply Effects at Three Historical Reforms

1967	
Participation rate	0.36
Difference due to demographics	-.05
Difference due to different G and t	+.04
Marginal labor supply effect	-31.5
	(-12.1, -50.8)
1981	
Participation rate	0.53
Difference due to demographics	0
Difference due to different G and t	-.05
Marginal labor supply effect	-13.2
	(3.5, -30.0)
1996	
Participation rate	0.40
Difference due to demographics	+.02
Difference due to different G and t	+.05
Marginal labor supply effect	-28.9
	(--8.7, -49.2)

Notes:
95 percent confidence intervals in parentheses

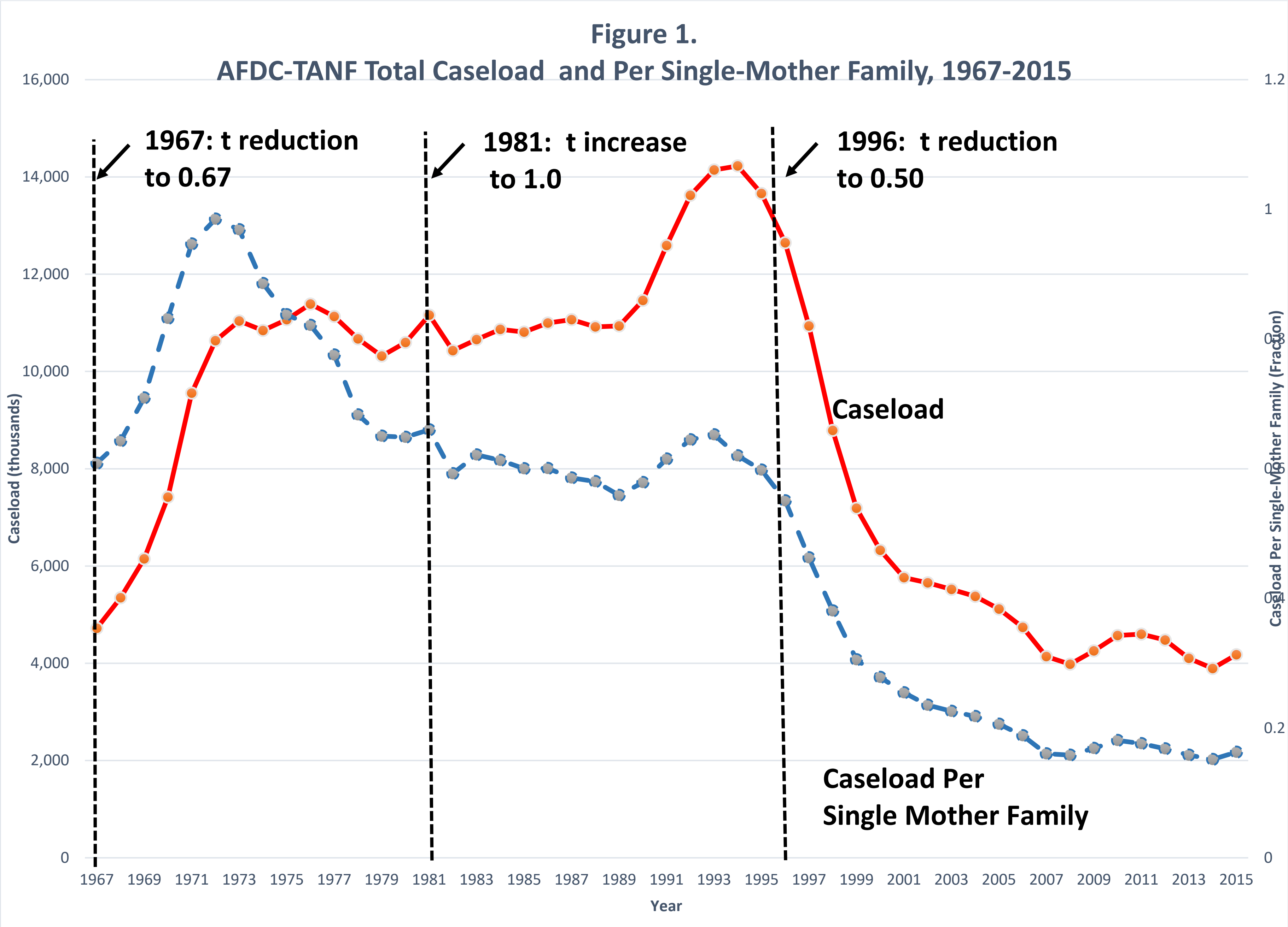


Figure 1.
AFDC-TANF Total Caseload and Per Single-Mother Family, 1967-2015
1967: t reduction to 0.67
1981: t increase to 1.0
1996: t reduction to 0.50
Caseload
Caseload Per Single Mother Family
Caseload (thousands)
Caseload Per Single-Mother Family (Fraction)
Year
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
1.2
1
0.8
0.6
0.4
0.2
0
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015

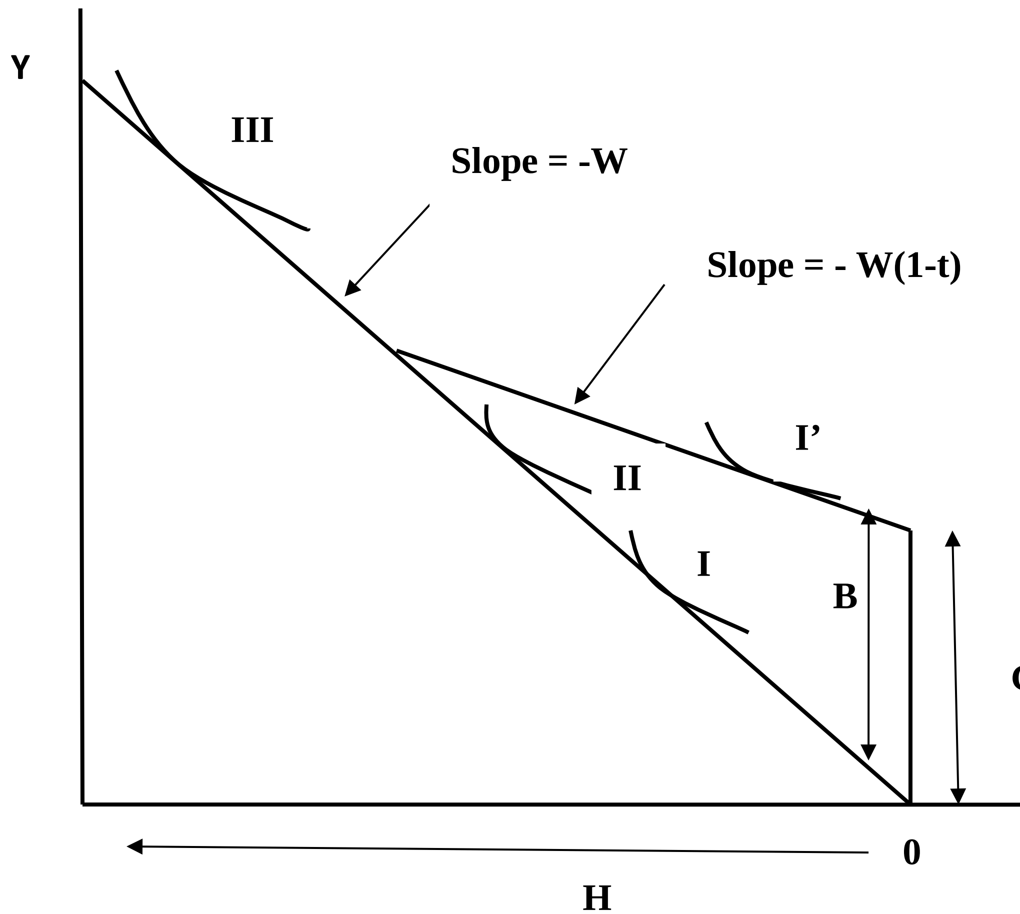


Figure 2: Traditional Income-Leisure Diagram

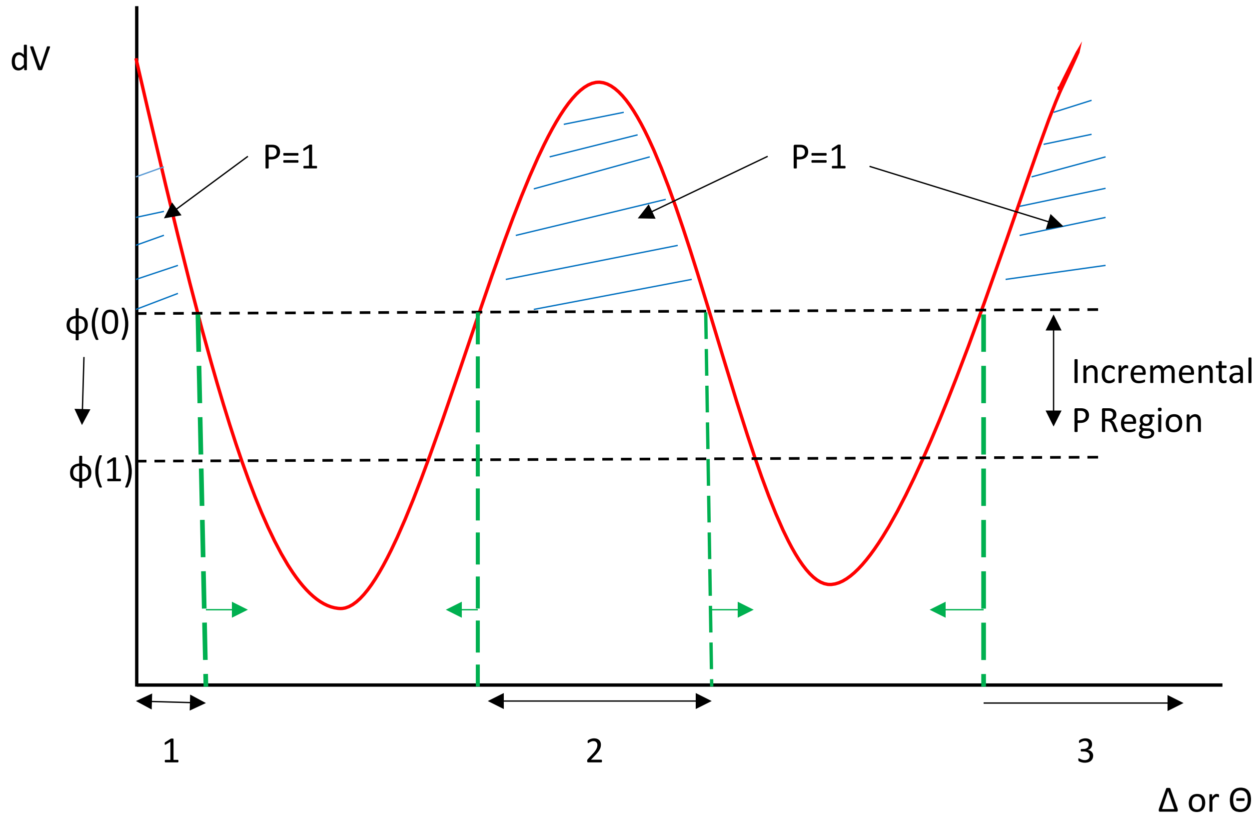


Figure 3. Hypothetical dV curve showing regions of Δ Among Participants (1,2,3)

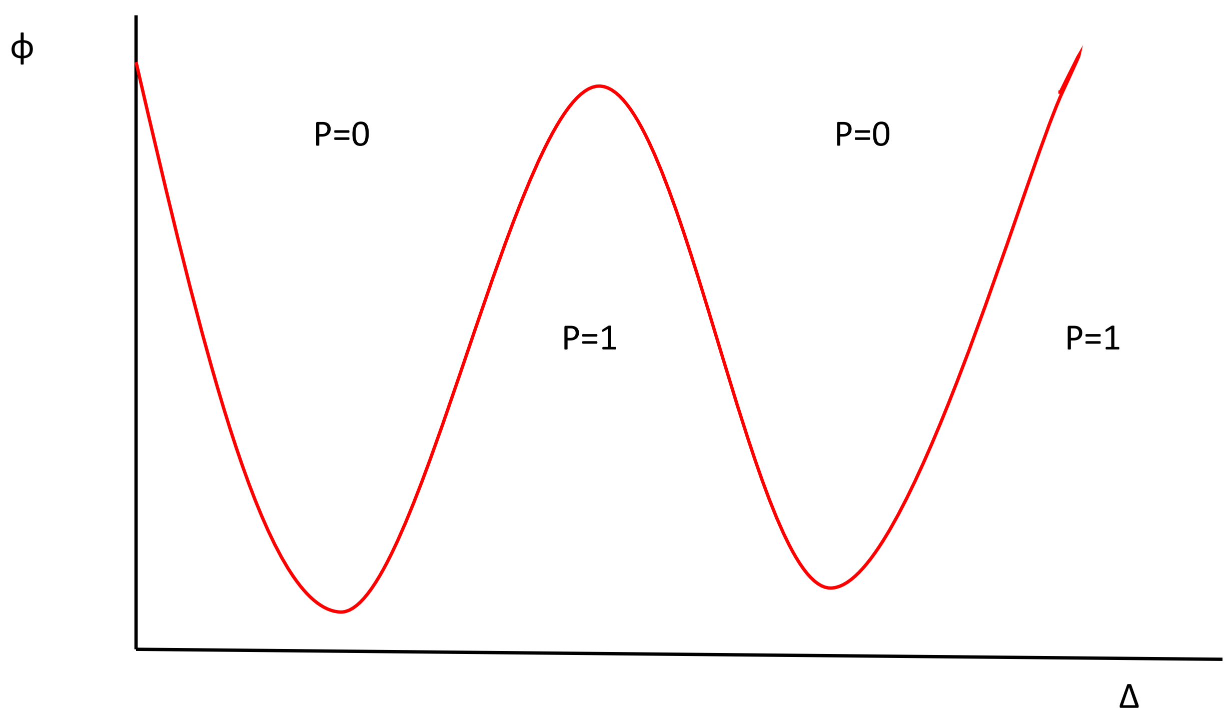
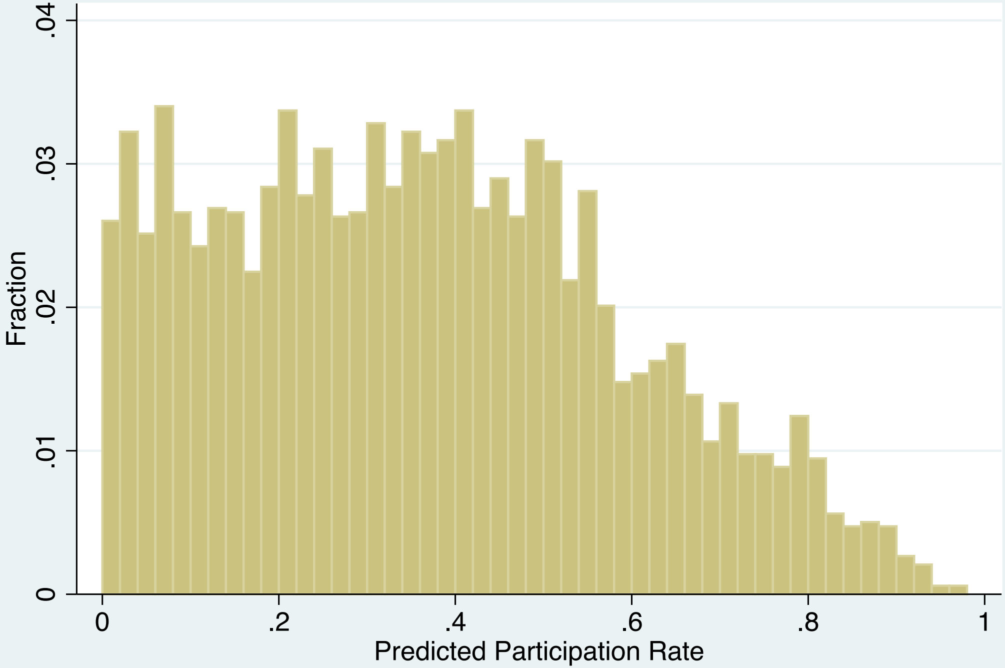


Figure 4. Hypothetical Θ_D, ϕ_D Locus

Figure 5: Histogram of Predicted Participation Rates



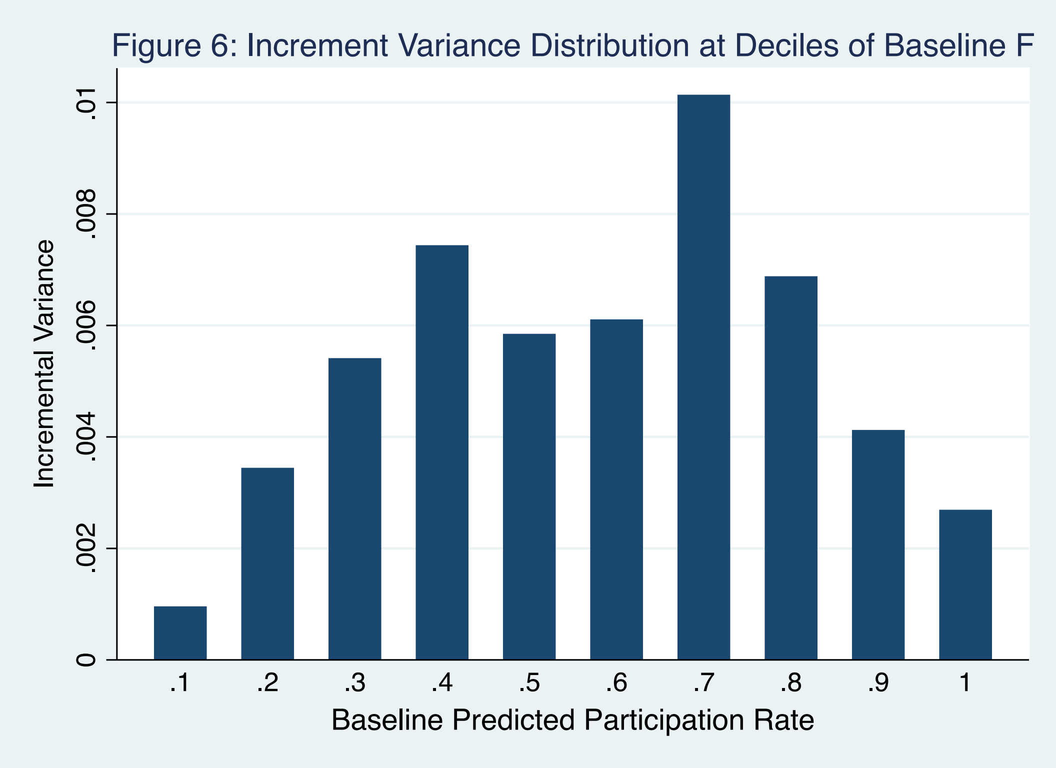


Figure 6: Increment Variance Distribution at Deciles of Baseline F
Incremental Variance
0
.002
.004
.006
.008
.01
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Baseline Predicted Participation Rate

Figure 7. Marginal Labor Supply Curves for Different Natural Cubic Splines
95 percent confidence intervals shown

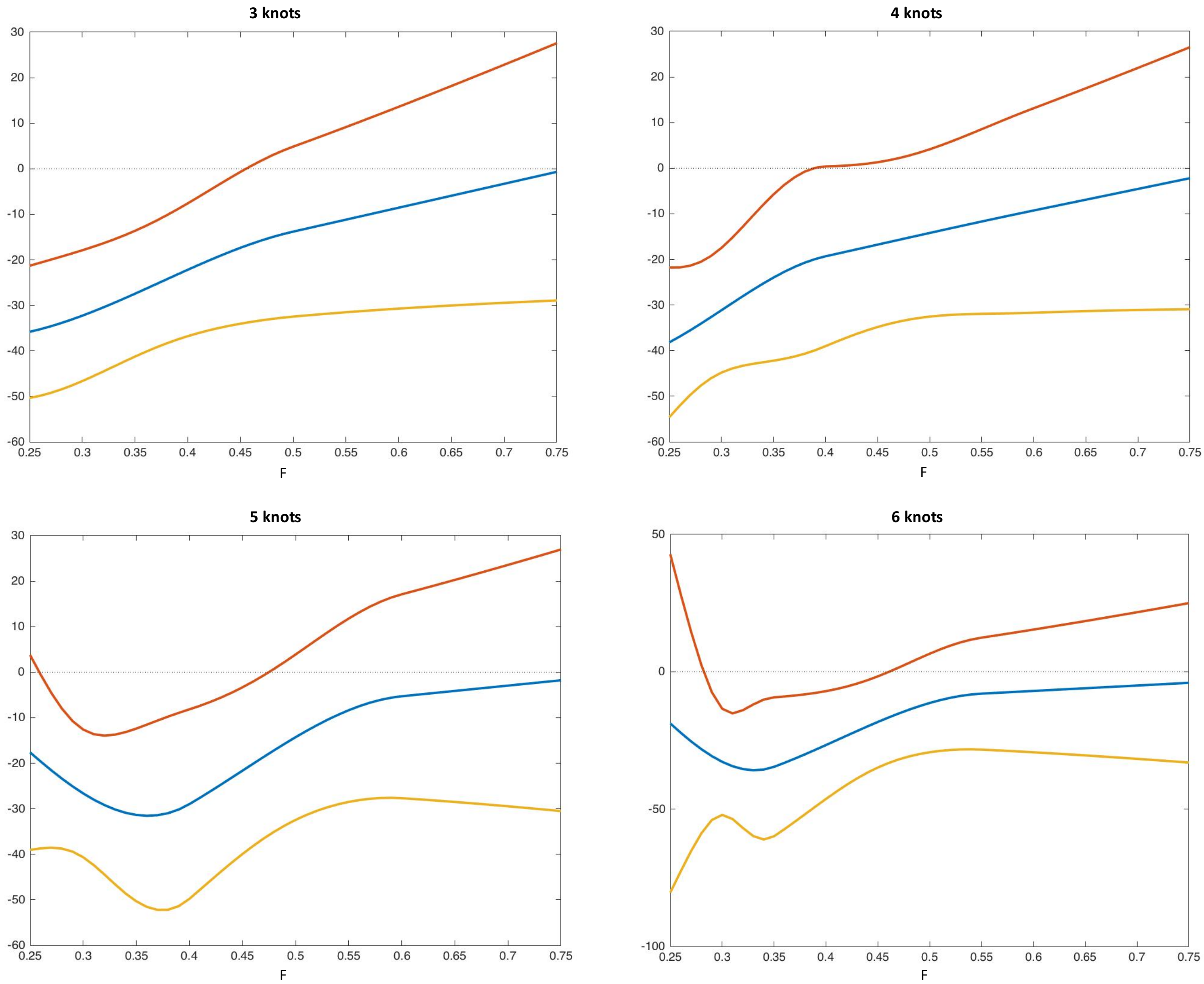


Figure 8. Marginal Labor Supply Curve for Five-Knot Model with Interactions

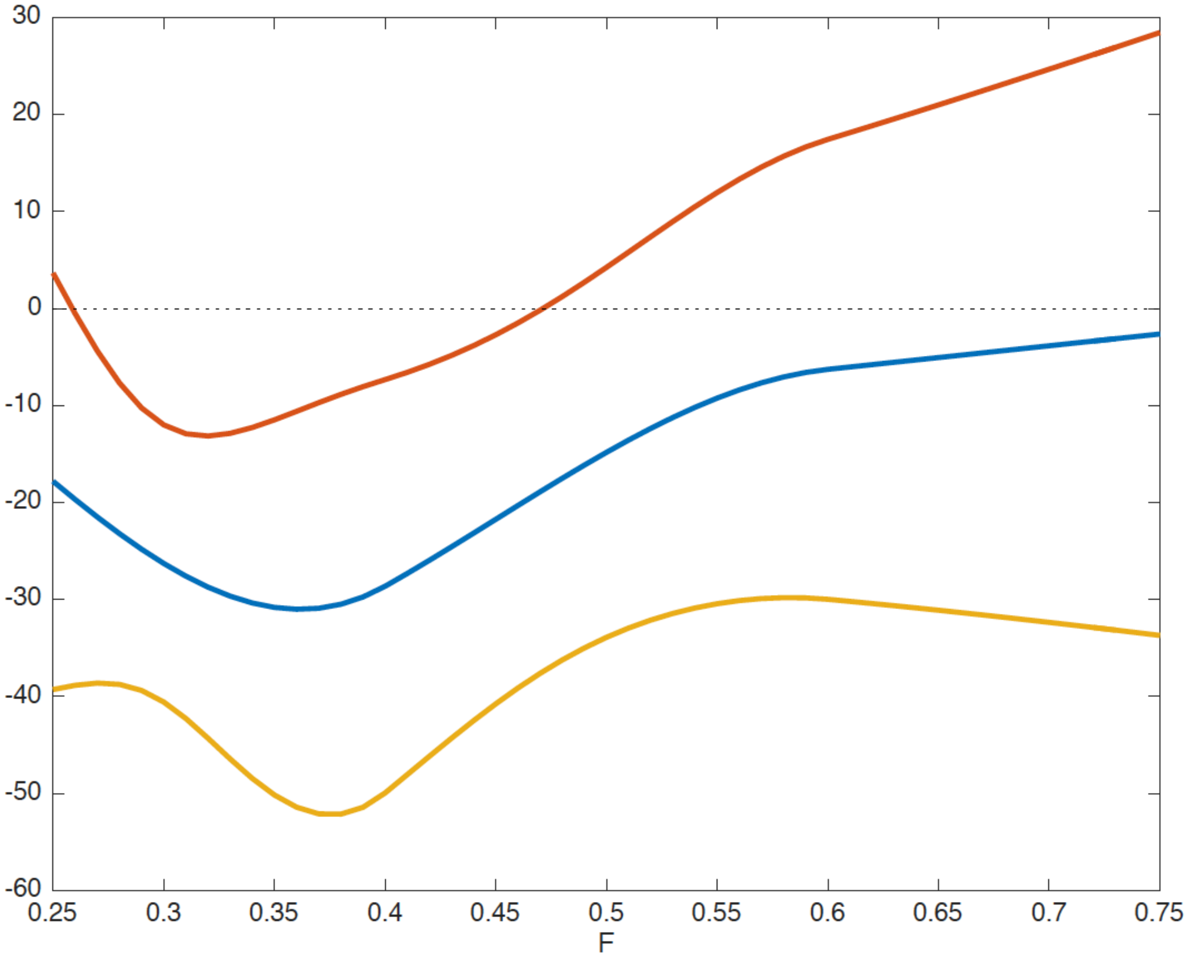


Figure 9. Marginal Labor Supply Curves for Different Instruments
95 percent confidence intervals shown

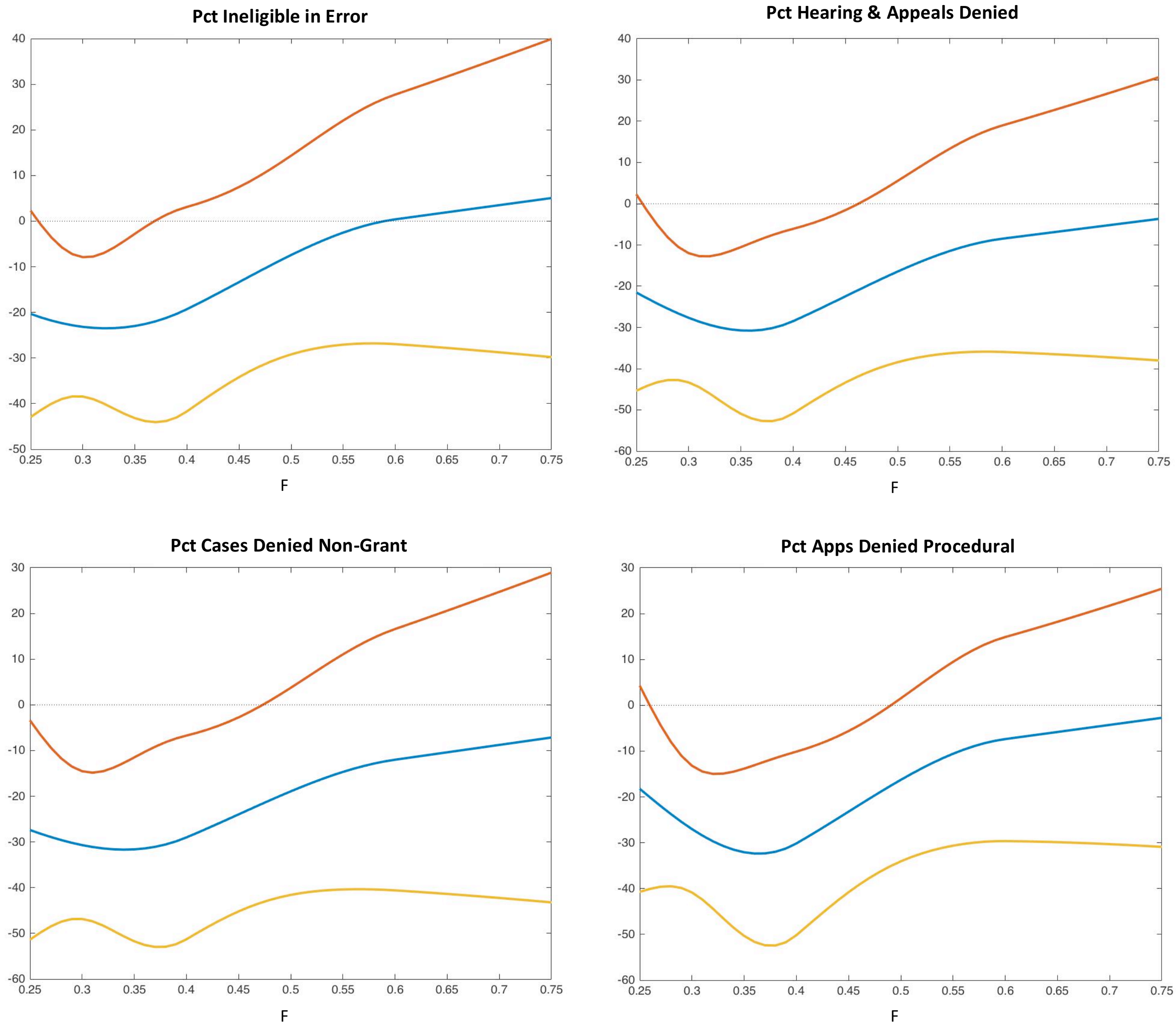
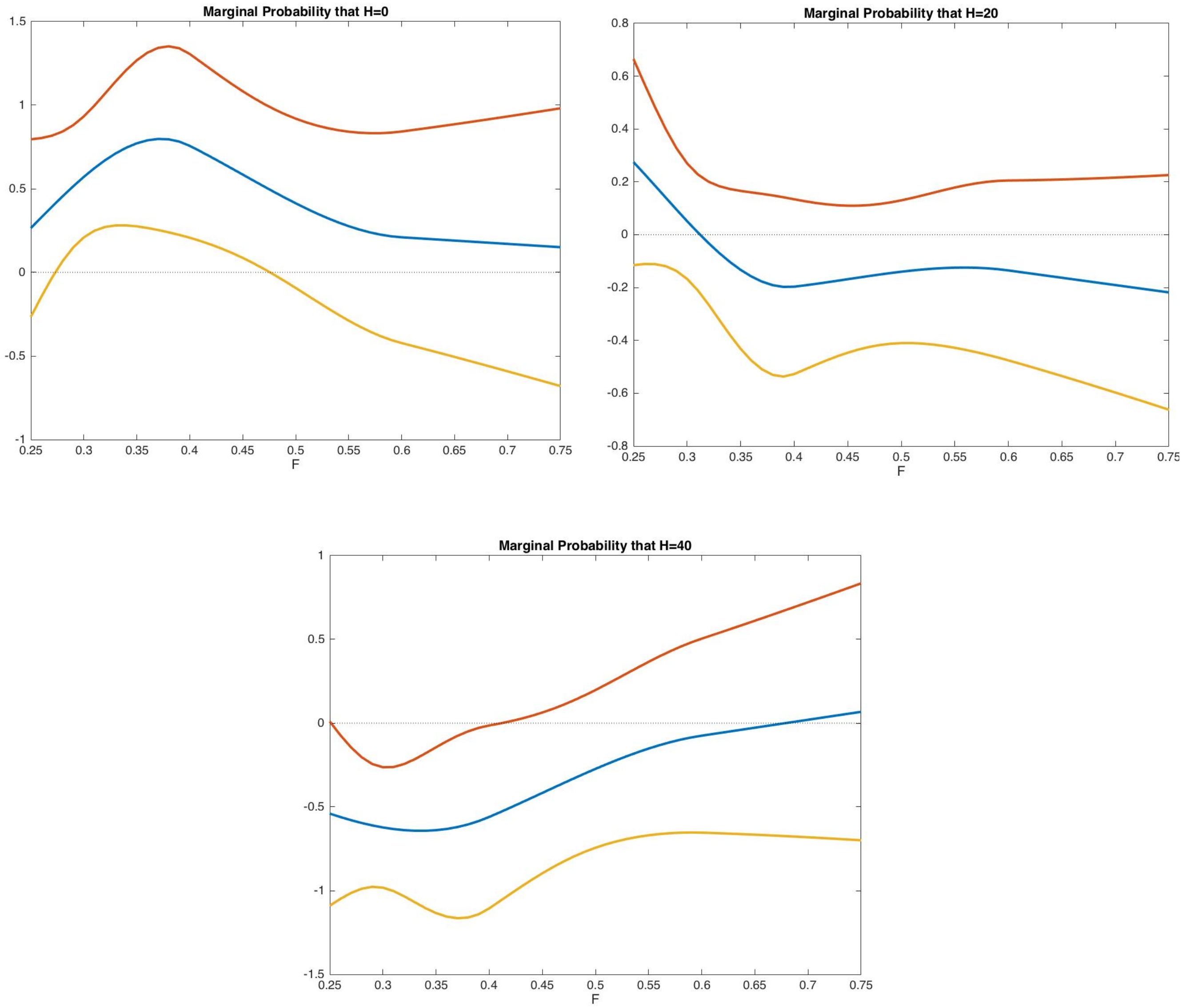


Figure 10



Marginal Probability that H=0
1.5
1
0.5
0
-0.5
-1
0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75
F
Marginal Probability that H=20
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75
F
Marginal Probability that H=40
1
0.5
0
-0.5
-1
-1.5
0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75
F