The Marginal Labor Supply Disincentives of Welfare Reforms*

Robert Moffitt†
Department of Economics, Johns Hopkins University

February 18, 2019

ABSTRACT: Existing research on the static effects of the manipulation of welfare program benefit parameters on supply has allowed only restrictive forms of heterogeneity in preferences. Yet preference heterogeneity implies that the marginal effects of welfare reforms on labor supply may differ in different time periods with different populations and which sweep out different portions of the marginal distributions of preferences. A new examination of the heavily studied AFDC program examines changes in its tax rates in 1967, 1981, and 1996 and estimates the marginal effects on labor supply of each of the reforms in those years. A theory-consistent reduced form model is formulated 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 individuals on the margin at each of the historical reform dates differed because of differences in the composition of who was on the program and who was not, and who the marginal person was, in each period.

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

*The author would like to thank Marc Chan, Kai Liu, Shaiza Qayyum, Kyungmin Kang, and Sue Bahk for research assistance as well as the participants of a large number of conferences and departmental seminars, including formal discussant comments by James Ziliak. Research support from the National Institutes of Health is gratefully acknowledged.

†moffitt@jhu.edu.

The classic form of a welfare program for a low-income population is that represented by a negative income tax, with a guaranteed minimum cash payment for those with no private income and with a positive marginal benefit-reduction rate, or tax rate, applied to increases in earnings. In the U.S., the only major program 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 1970-2015. The fraction of families participating has gradually fallen, first in the late 1970s and then again after 1996. 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 likely to be 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.

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 the tax rate changes in the AFDC program in 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 and provides a formal definition of the labor supply 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.

The second section 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),Bjorklund 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, 2005); 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. 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. 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.

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

[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 not conducted.

reforms in each year. For example, 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.

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

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

where H_i is hours of work for individual i, Y_i is disposable income, P_i is a program participation indicator, θ_i is a vector of labor supply preference parameters, and ϕ_i is a scalar representing fixed costs of participation in utility units. The presence of P_i allows for the presence of fixed costs of participation–in money, time, or utility, with the exact type unspecified and scaled in units of utility. Fixed costs are required to fit the data because many individuals who are eligible for transfer programs do not participate in them, as revealed by the presence in the data of nonparticipating eligibles for virtually all programs. The presence of fixed costs also makes the participation decision separable from the H decision.[2] 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

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

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

[2] 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)).

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), below breakeven but off the program (II), and below breakeven and on the program (I'; I is the pre-program location for this individual).

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

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

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

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

where $G(\theta_i, \phi_i)$ is the joint c.d.f. of the two heterogeneity components. An individual is on the margin of participation if she is indifferent between participating and not participating. There is a locus of combinations of the labor supply and fixed cost parameters that put a person on the margin, with the locus defined as the set of values θ_D and ϕ_D satisfying the condition

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

The labor supply response of the marginal individual (see (6)) therefore differs for those with differing values of ϕ.[3]. Defining $S_{\theta\phi}$ as the set of parameters in regions demarcated by the θ_D and ϕ_D boundary points which generate $P = 1$, the mean labor supply response of those participating in the program is

$$\begin{aligned}\widetilde{\triangle}_{P_i=1} &= E(\triangle_i \mid W_i, N_i, G, t, r, P_i = 1) \\ &= \frac{1}{P} \int\limits_{S_{\theta\phi}} \int \triangle_i(\theta_i \mid W_i, N_i, G, t, r) dG(\theta_i, \phi_i)\end{aligned} \tag{9}$$

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

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

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 $\partial\widetilde{\triangle}/\partial P$. 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.

In this transfer program model, an exogenous increase in participation can be induced by a reduction in ϕ_i, the fixed cost parameter (an observable proxy for these costs will be used in the empirical work). A reduction in this parameter only shifts who participates in the program and does not shift the values or the latent distribution of the population responses in eqn(6). With unrestricted heterogeneity of preferences, the marginal responses so induced can be greater or smaller than the responses of those initially on the program. A reduction in ϕ_i means that individuals with lower values of the non-cost portion of the utility gain of going onto welfare, $dV = V[W_i(1-t), G + N_i(1-r); \theta_i] - V[W_i, N_i]$ will

[3]This focus is a generalization of the U_D parameter of Heckman and Vytlacil (2005)

participate. But there is no determinant relationship between the magnitude of this portion of the utility gain and the magnitude of labor supply reductions because, for example, those with smaller utility gains who enter the program may have relatively greater preferences for consumption goods than for leisure than those initially on the program, leading to a smaller labor supply reduction on the margin. Alternatively, marginal participants could have a smaller relative valuation of consumption relative to leisure, giving the opposite result. For the same reason, if ϕ_i continues to fall, the marginal labor supply responses can fall, rise, or remain the same with each increase in participation. Thus how marginal labor supply reductions change as participation expands or contracts can only be determined empirically.

It is also worth noting that the effect of changes in the program parameters G, t, and r on hours of work do not identify marginal responses because they also have inframarginal responses on those initially on the program.[4] The effect of changes in those parameters is a combination of marginal and inframarginal responses. But the levels of those parameters affect who is on the margin because they affect the values of θ_D and ϕ_D and hence will alter the set $S_{\theta\phi}$, which in turn determines the values of θ of individuals on the margin of participation. Consequently, the marginal response to a reduction in ϕ_i will vary with the level of the program parameters, and those parameters must be conditioned on to properly estimate marginal responses after a reduction in fixed costs.

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,

[4] With unrestricted heterogeneity, those inframarginal responses can also vary arbitrarily with successive changes in those parameters.

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 $\phi_i = m(Z_i, \nu)$, where Z_i is an observable correlate of fixed costs. Then mean hours of work in the population conditional on the budget constraint and on Z_i can be expressed as

$$
\begin{aligned}
E(H|W,N,G,t,r,Z) &= E_\theta[H(W,N;\theta) \mid W,N,G,t,r,Z] \\
&+ E_{\theta,\nu}[\triangle \mid P=1, W,N,G,t,r,Z;\theta,\nu] Pr(P^* \geq 0 \mid W,N,G,t,r,Z;\theta,\nu)
\end{aligned} \tag{12}
$$

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 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) and its associated equations(4)-eqrefeq:5 can be written in unrestricted form as

$$H_i = \beta_i + \alpha_i P_i \tag{13}$$

$$P_i^* = m(Z_i, \delta_i) \tag{14}$$

$$P_i = 1(P_i^* \geq 0) \tag{15}$$

where β_i (mean hours off welfare) and α_i (a relabeling of $\triangle$) are scalar random parameters and δ_i is a vector of random parameters. All parameters are allowed to be individual-specific and to have some 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. The function m can likewise be unrestricted and can be saturated if Z_i is assumed to have a multinomial distribution, although we shall discuss restrictions on δ_i below.

The object of interest is the distribution of α_i. Selection in this model can occur either on the intercept (β_i) or the slope coefficient (α_i) because both may be related to δ_i and, in fact, the theoretical model implies that they must be because the participation equation contains the parameters of the labor supply function. Conditional on Z_i, eqn(12) now takes 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 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 for those on the program (i.e., the effect of the treatment on the treated) conditional on Z_i, and depends on the shape of the distribution of α_i and how different fractions of participants are selected from different portions of that distribution. While the first assumption is familiar, the second may be less so. The usual assumption in the literature is that the two potential outcomes, β_i and $\beta_i + \alpha_i$, are fully independent of Z_i, which implies that α_i is as well. Eqn (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 the effect of the participation probability (i.e., the propensity score), and only at specific values of Z_i.[5] If this were not so, different values of Z_i would lead to different conditional means of α_i through some other channel, which

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

would rule it out as a valid exclusion restriction.[6]

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

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

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

where ϵ_i and ν_i are mean zero and orthogonal to the RHS by construction. No other restriction on these error terms need be made, as this is a reduced form of the model. The first equation just states that the population mean of H_i equals a constant plus the mean response of those in the program times the fraction who are in it. The implication of this way of specifying the model is that preference heterogeneity is detectable by a nonlinearity in the response of the population mean of H_i (taken over participants and nonparticipants) to changes in the participation probability. If responses are homogeneous and hence the same for all members of the population, the function g reduces to a constant and therefore a shift in the fraction on the program has a linear effect on the population mean of H_i. If the responses of those on the marginal vary, however, the response of the population mean of H_i will depart from linearity. This 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.[7]

Nonparametric identification of the parameters of the model–β and the function g at every point F–is straightforward and has been extensively discussed in the literature. F is

[6]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 assumption affects the interpretation of the estimated distribution.

[7]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.

identified at every data point Z_i from the second equation from the mean of P_i at each value of Z_i (apart from sampling error). If there is a value of Z_i in the data for which $F(Z_i) = 0$, then β is identified from the mean of H_i at that point and hence g is identified pointwise at every other value of Z_i and hence F. If no such value is in the data, then g can only be identified subject to a normalization or multiple variables of g can be identified. For example, the LATE of Imbens and Angrist (1994) is identified by the discrete difference in H between two points z_i and z_j divided by the difference in F between those two points. A marginal treatment effect is a continuous version of this and requires some smoothing method across discrete values of Z, and is computed by $\partial H/\partial F = g'(F)F + g(F)$. In the empirical analysis below, the g function will be approximated by a nonparametric but continuous function which implicitly means that interpolation between the data points identifies the pointwise derivatives in the MTE.

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:

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

Note that the theory imposes two restrictions on the form of the equation. First, the intercept of the equation must not include the welfare program parameters G, t, and r because the intercept represents labor supply off welfare. 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 the coefficient on changes in participation induced by changes in Z 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.

Fully 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.[8] 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 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).

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

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 an able-bodied spouse or partner, with the practical implication that the caseload was almost entirely composed of single women with children. However, major structural reforms of the program began in 1993 with the introduction of work requirements and time limits, and it has not returned to its classic form since that time. Consequently, the analysis here will use data on disadvantaged single women with children from the late 1980s 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 study reflected in its name. I use all waves of panels interviewed in the Spring of each year 1988-1992 (only Spring to avoid seasonal variation) and pool them into one sample, excluding overlapping observations by including only the first interview when the person appears 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 1. 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). 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 and t) are needed. To address the familiar problem of missing wages for nonworkers, a traditional selection model is estimated, with estimates shown in Appendix Table 2. 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 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 accurate approximations to the parameters actually faced by recipients.[9] The mean effective tax rate on earnings across years is approximately 0.41. 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. 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.

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

The vector Z_i consists of variables that affect the probability of participation but not labor supply conditional on participation, and hence must represent fixed costs of participation. Here we take advantage of the existence prior to 1996 of a federal quality control audit program imposed on state AFDC programs (Hansen and Tepping (1990),of Representatives Committee on Ways and Means (1994)). Because the program was federally regulated in that period and because eligibility conditions were set at the federal level, the federal government audited state records annually to determine whether errors in assessing eligibility were made. 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.[10] 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, as well as 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.

While random error rates would not necessarily affect AFDC participation, there is a large literature in social work journals from the period showing that those rates were not random but were part of systematic agency policies to strategically reduce caseloads, with the most common mechanism simply to deny eligibility as frequently as possible using subjective interpretations of the eligibility rules (Piliavin et al. (1979),Brodkin and Lipsky (1983),Lipsky (1984),Lindsey et al. (1989),Kramer (1990)). This literature shows that caseworkers conducting eligibility assessments on individual applicants were able to interpret the rules for what types of income to count, whether an able-bodied spouse or partner was present, which assets to count, and other factors affecting eligibility broadly and subjectively. Usually, the discretion was exercised at the explicit or implicit direction of welfare department administrators who, in turn, took their guidance from the legislature

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

and governor of the state. Hence, the degree of discretion exercised at the caseworker level reflected the attitudes of the state governance structure toward the AFDC program.

While the lack of data on positive errors made by states does not permit a direct examination of whether positive and negative errors balance out, the test of the bias hypothesis in the social work literature can nevertheless be made just by determining whether negative error rates in a state affect the participation rate. If the two types of errors balance out, then negative error rates should have no effect on participation. This can be tested with the SIPP data at hand by determining whether the error rate in the state of residence of each individual significant affected that individual's probability of participation in the program.

The data provide information on four measures of state AFDC administrative actions which are potential correlates for non-financial administrative barriers: the percent of applications denied eligibility in error, the percent of applications denied for procedural reasons, the percent of applications denied overall, and the percent of applications dismissed for eligibility reasons. The first is the most direct measure of state error rates, while the second is often considered to be a result of administrative barriers because procedural reasons generally mean failure to complete paperwork properly. The third and fourth are indirect measures of administrative barriers, but the probability that a given individual is on welfare, conditional on their earned and unearned income, should not depend on state denial rates if those rates are only based on earned and unearned income of the applicant.

Table 1 shows the descriptive statistics of the four potential instruments.[11] The mean ineligible error rate is 4.7 percent, the mean percent of applications denied for procedural reasons is 14.1 percent, the mean percent of applications denied is 24.3 percent, and the mean percent of applications dismissed for eligibility reasons is 19.8 percent. Not only are these rates often quite high, they have an enormous range across the states: from 2 to 7

[11] The administrative variables bounce around from year to year for each state, so the state-specific mean values over the 1988-1992 period are used for the analysis.

percent for the first, from 1 percent to 34 percent for the second, from 5 percent to 47 percent for the third, and from 1 percent to 39 percent for the fourth.

The first four columns of Table 2 report estimates of the coefficients on these instruments from a probit equation for program participation in the SIPP analysis sample. Three of the four instruments have negative effects on program participation but with high standard errors. The F-statistics for the four from estimates of a linear probability model are also very low. However, a more detailed analysis of the power of these instruments reveals that the low power exhibited by these estimates hides the existence of differential power in different ranges of the participation probability. Figure 3 shows a histogram of the predicted participation rates using the percent of applications dismissed for eligibility, the strongest instrument in Table 1. The support of the predicted participation probabilities is approximately the same for probabilities below 0.6 but declines for probabilities above that value. But instrument power requires instead an assessment of how the instruments change the distribution of probabilities, and an initial, crude assessment is shown in Figure 4, which shows the incremental explained variance of the instrument at different levels of a baseline predicted probability, where the baseline is a probability predicted by excluding the instrument from the equation. The incremental explained variance is the amount by which $\sum[P - \hat{F}]^2$ is increased when the instrument is added. At low and high values of $\hat{F}$, the instrument has very little contribution to explained variance compared to that variance in the middle ranges.[12]

The bottom rows of Table 2 show pseudo F-statistics which are a more formal way of assessing instrument power in different ranges of the predicted participation probability. The statistics use conventional F-statistic formulas for the instrument but calculated separately for quartiles of the baseline predicted probability. F-statistics as calculated in this way can be negative in some ranges if the addition of the instrument to the participation equation worsens fit in one region of the distribution. The results show that

[12]More generally, it is unlikely that any instrument would have much power in the tails of conventional c.d.f.'s where probabilities are near 0 or 1.

the third and fourth instruments have much stronger power in the second quartile of the distribution (.25-.50) than would be implied by a conventional F-statistic, which is implicitly defined over the entire range.

That power is strongest in the .25-.50 range because the mean participation rate, .37, lies in that range. An additively separable instrument in the probit index function just moves probabilities above and below that mean. The power of instruments can be expanded to a wider interval around .37 by relaxing the additive separability and by allowing the effect of the instrument to vary with other indvidual and state characteristics. An analysis of significant interactions of the instruments with the other covariates in the participation equation revealed that at least two–the AFDC guarantee and the level of nonlabor income–significantly affected the effect of the instruments on participation probabilities. In all cases, higher levels of the AFDC guarantee led to more negative effects of the administrative barriers on participation probabilities of the SIPP single mothers, as did higher levels of their nonlabor income. Possible reasons for these effects are that states with high guarantees may be particularly interested in using non-financial barriers to lower the caseload and hence reduce costs, while women with higher levels of nonlabor income may be easier to deny eligibility with subjective assessments of the level of such income.[13]

Adding interactions with these two covariates improves the pseudo F statistics in the middle two ranges of the participation probabilities considerably, as shown in the last four columns of Table 2. The best performing instrument is the percent of applications denied, which has pseudo F statistics of 6.1 and 4.5 in the middle two quartiles. This instrument will therefore be used in the second-stage estimation of the hours equation in the next section (full estimates are shown in Appendix Table B3). But none of the instruments have power in the bottom or top quartile, and hence the instruments have no power to estimate MTEs in those ranges. Consequently, MTEs in those ranges will not be considered further.[14]

[13]The single mother's actual nonlabor income is included elsewhere in the regression.

[14]As noted in the Introduction, the three historical tax reforms had participation rates within the middle

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. 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.[15]

The full estimates of the hours equation are shown in Appendix Table B4 and the key coefficient estimates are shown in column (1) of Table 3 for the basic model.[16] 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 g function to be identified by the fixed cost instruments. 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 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 B4) 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 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

two quartiles and hence the estimates are capable of simulating their effects.

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

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

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, for participation rates between .25 and .75.[17] 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 about .35 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 middle of the significant range. Further additions of knots only serve to add additional small bumps in the curve without changing its general shape.[18]

Columns (2) and (3) of Table 3 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 no heterogeneity in shape is present. The third column presents estimates of a specification with two

[17] 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.

[18] Generalized cross-validation statistics (not shown) for goodness of fit also stop improving after 5 knots.

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.

The point estimates for the marginal labor responses are often relatively large, peaking at approximately 30 hours per week for the best-fitting specifications. While these effects are large, it is worth emphasizing that they occur only in a specific part of the participation probability distribution and therefore only at certain caseload levels. An interpretation of the reason for the large sizes can be obtained by examining how the distribution of hours worked changes as participation rises–in particular, by examining how individuals reduce hours from 40 per week or 20 per week to lower levels, including nonwork. That movements between full-time work, part-time work, and nonwork may be important is demonstrated in Table 4, 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 7, which shows the result of estimating the hours worked equation by successfully replacing the dependent variable 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 fell 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 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. 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 reflect a decline in movements out of full-time work.

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 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)).

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.

Table 4 shows the results of using the estimates from the 5-knot model reported in the last section to calculate the marginal labor supply response in each of these years. The calculations ignore differences in the effects of differences in the demographic and other exogenous covariates in the model, which is left for future work, and take into account only the differences in the initial participation rate, in the level of G, and in the initial level of t. In 1967, the initial participation rate was .36, the guarantee was 32 percent above the level in the SIPP sample used for estimation, and the tax rate was the same as in that sample.[19]

[19] The tax rates used in the estimation were effective, rather than nominal, tax rates. We therefore assume that the difference between the two was the same in 1967 as in the SIPP sample in the late 1980s and early

The model estimates a marginal labor supply response of -31 which, although having a wide confidence interval, is 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. Both of these effects work to reduce the marginal labor supply response. The marginal response in 1981, just before the reform, 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 1981 values. Both of these factors pushed the labor supply response of the marginal individual in a negative direction. At the time of the 1996 reform, the marginal response was about 28 hours, and hence had risen most of the way back to its 1967 level.[20]

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 program, this would imply that the hypothetical marginal response to an increase in participation would be insignificantly different from zero.

5. Summary and Conclusions

This paper has had both a substantive and methodological goal. The substantive goal was to estimate the marginal labor supply responses in the years of historical reforms of the AFDC reform which altered the tax rate on earnings in the program either upward or downward. Major reforms took place in 1967, 1981, and 1996. The methodological goal

1990s.

[20]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.

was to demonstrate how a theory-consistent reduced form model could be estimated nonparametrically to determine the size of the marginal labor supply response as a function of the participation rate and of the budget constraint.

The substantive findings are that the marginal labor supply responses in the reform years 1967, 1981, and 1996 varied significantly in magnitude. The largest marginal responses were in 1967 when participation rates were fairly low and guarantees were fairly high. The lowest marginal responses were in 1981, when participation rates were high and guarantees were low and, in that year, the 95 confidence interval includes zero. The marginal responses in 1996 fell in between those in the other two years.

The methods used here are reasonably simple to estimate and should be applicable to cases where more than a binary instrument is available and where the values of the instrument sweep out a reasonably large portion of the participation rates in the program. The instruments used here provide a much greater range of participation rate estimates than would a binary instrument but still provide variation over a much smaller range than the full possible support of participation rates. Looking for instruments which can cover a reasonably wide range of participation rates would therefore seem to be a challenge for future research using this method.

Estimation of structural models using these methods would also be worth exploring. Because the reduced form model developed here is fully theory-consistent, it contains all the budget constraint and other variables that would enter into a fully structural model. Whether the structural model parameters could be identified from the estimates of the reduced form parameters and distributions in the model specified here would also be a worthy topic for future work.

References

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

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

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.

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

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

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

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.

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.

of Representatives Committee on Ways, U. S. H. and Means (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.

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.

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

First-Stage Probit Estimates of the Participation Equation

Log W	-3.008
	(0.337)
Log (N+10)	-0.303
	(0.060)
Log G	1.312
	(0.192)
Log W(1-t)	0.264
	(0.154)
Age	0.010
	(0.005)
Black	0.150
	(0.056)
Family size	-0.054
	(0.033)
Number of Children Less than 6	0.298
	(0.034)
Food Stamp Guarantee	1.972
	(1.281)
State Unemployment Rate	0.043
	(0.022)
Northeast	0.141
	(0.113)
Midwest	0.178
	(0.145)
West	1.080
	(0.285)
Pct. App. Denied	0.011
	(0.006)
Pct. App. Denied*G	-0.176
	(0.060)
Pct. App. Denied*N	-0.001
	(0.000)
Constant	5.513
	(1.114)

Notes:
Standard errors in parentheses.

Appendix Table B4
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)
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 B4 (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 F, the predicted participation probability. The g function is specified as

$$g(Fhat) = g_1 + g_2 F + g_3 N3 + g_4 N4 + g_5 N5 \tag{42}$$

where

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

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

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

where

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

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

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

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

Table 1
Four Fixed Cost Proxies

	Mean	Stnd Dev	Min	Max
Percent Ineligible in Error	4.7	1.1	2.2	7.0
Percent Applications Denied for Procedural Reasons	14.1	8.9	1.3	34.6
Percent Applications Denied	24.3	11.2	5.3	47.8
Percent Applications Dismissed for Eligibility Reasons	19.8	10.6	1.0	39.4

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

First Stage Participation Probit Instrument

Coefficient Estimates and Instrument Power

	Uninteracted				With Interactions			
	Pct. Ineligible	Pct. App. Den.Proc. Reasons	Pct. App. Denied	Pct. App. Den. Elig.	Pct. Ineligible	Pct. App. Den.Proc. Reasons	Pct. App. Denied	Pct. App. Den. Elig.
Intercept	0.017	-0.034	-0.051	-0.057	0.097	0.011	0.011	0.005
	(0.026)	(0.036)	(0.033)	(0.040)	(0.046)	(0.006)	(0.006)	(0.001)
With G	--	--	--	--	-0.734	-0.174	-0.176	-0.082
					(0.468)	(0.080)	(0.060)	(0.090)
With N	--	--	--	--	-0.002	-0.001	-0.001	-0.002
					(0.000)	(0.000)	(0.000)	(0.000)
OLS F-stat	0.39	0.97	1.81	2.53	1.01	2.27	4.80	1.27
Pseudo F-stat by Part. Prob.								
0-.25	-0.18	0.89	0.45	0.39	-0.96	-0.83	-1.22	0.13
.25-.50	0.47	0.46	3.11	2.02	5.21	3.40	6.06	2.49
.50-.75	0.18	-0.12	-0.32	0.21	3.10	2.72	4.45	2.06
.75-1.00	-0.08	-0.04	0.12	0.03	0.59	0.13	0.59	0.01

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: Defining RSS as the residual sum of squares, equal to the sum of [P-Fhat]**2 over all observations, where Fhat is the predicted probability for each individual, the pseudo F-stat is the ratio of the difference in RSS for the restricted model excluding the instruments and the RSS including the instruments divided by the d.o.f. to the residual variance, each numerator computed separately for observations predicted to be in the designated interval without the instruments.

Standard errors in parentheses.

Table 3
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)
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*F	--	-8.81	--
		(54.8)	
(N+10)*F	--	-0.30	--
		(0.37)	
Ln G*F	--	-2.55	--
		(18.95)	
Ln W(1-t)*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 4

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 5

Marginal Labor Supply Effects at Three Historical Reforms

1967	
Initial participation rate	0.36
Marginal labor supply effect	-31.5
	(-12.1, -50.8)
1981	
Initial participation rate	0.53
Marginal labor supply effect	-13.2
	(3.5, -30.0)
1996	
Initial participation rate	0.40
	-28.9
Marginal labor supply effect	(--8.7, -49.2)

Notes:
95 percent confidence intervals in parentheses
1967: G = 132 percent of 1990 value, t = 1990 value
1981: G = 1.02 percent of 1990 value, t = 67 percent of 1990 value
1996: G = 1.20 percent of 1990 value, t = 1990 value
Sources for estimates of initial participation rate and guarantee levels: Moffitt (1992), Ziliak (2016).

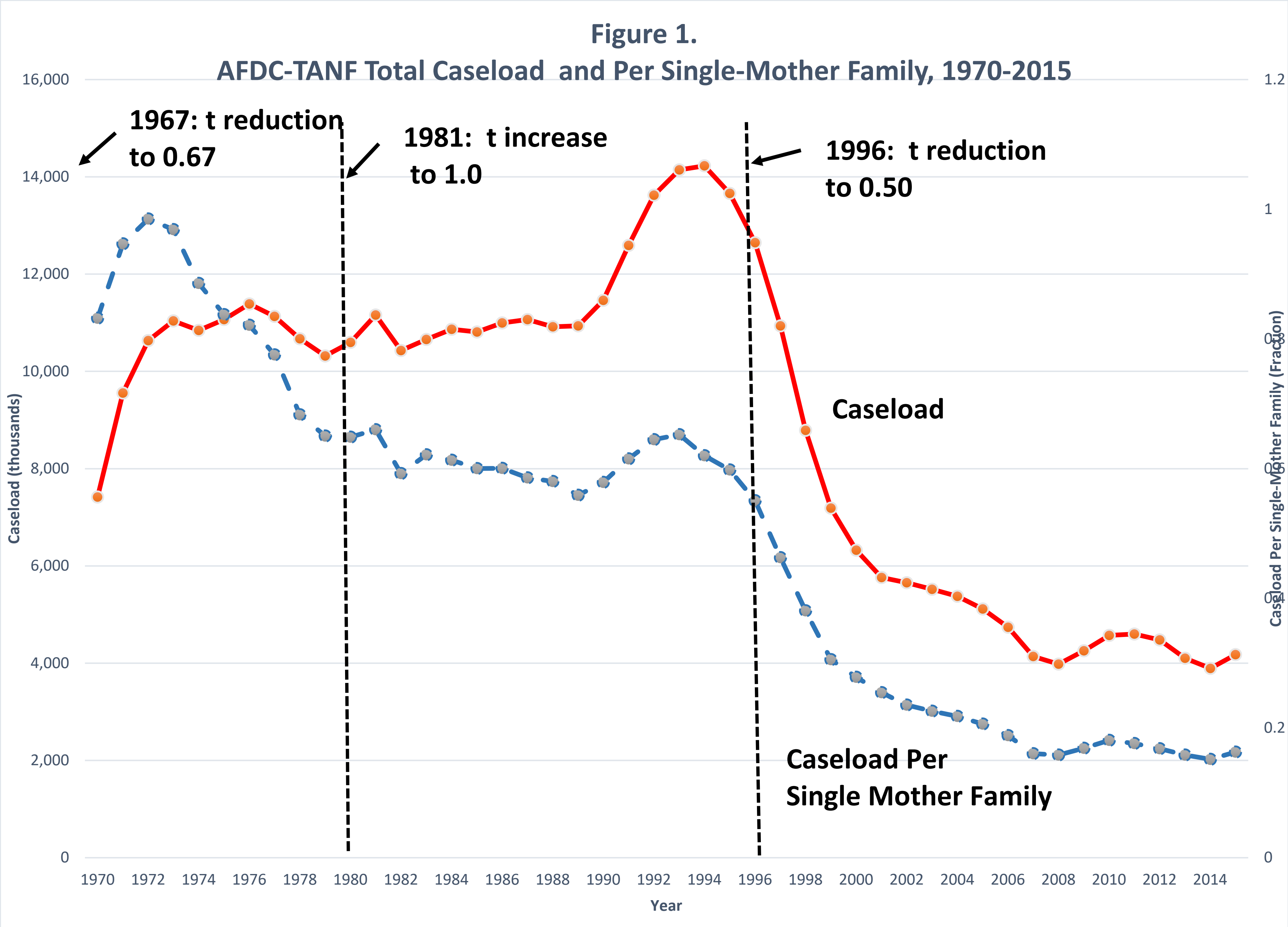


Figure 1.
AFDC-TANF Total Caseload and Per Single-Mother Family, 1970-2015
1967: t reduction to 0.67
1981: t increase to 1.0
1996: t reduction to 0.50
Caseload
Caseload Per Single Mother Family
Caseload (thousands)
Caseload Per Single-Mother Family (Fraction)
Year
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
1.2
1
0.8
0.6
0.4
0.2
0
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014

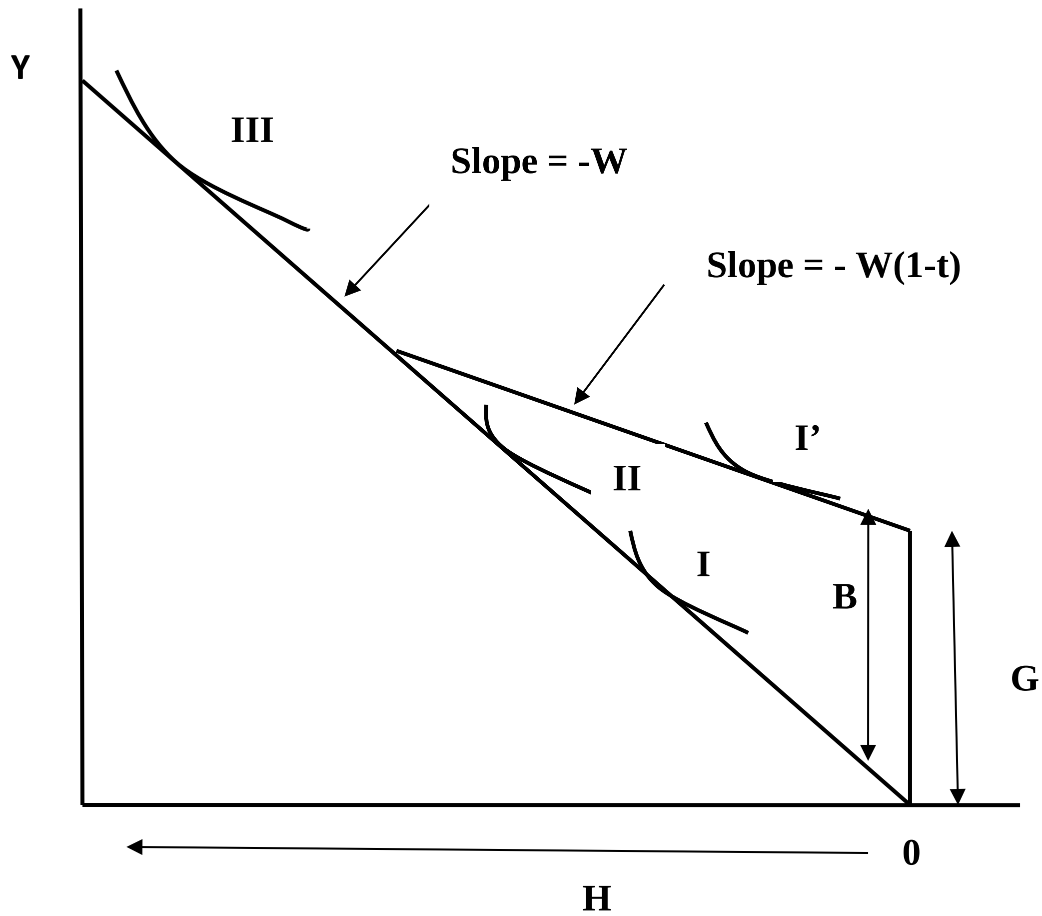


Figure 2: Traditional Income-Leisure Diagram

Figure 3: Histogram of Predicted Participation Rates

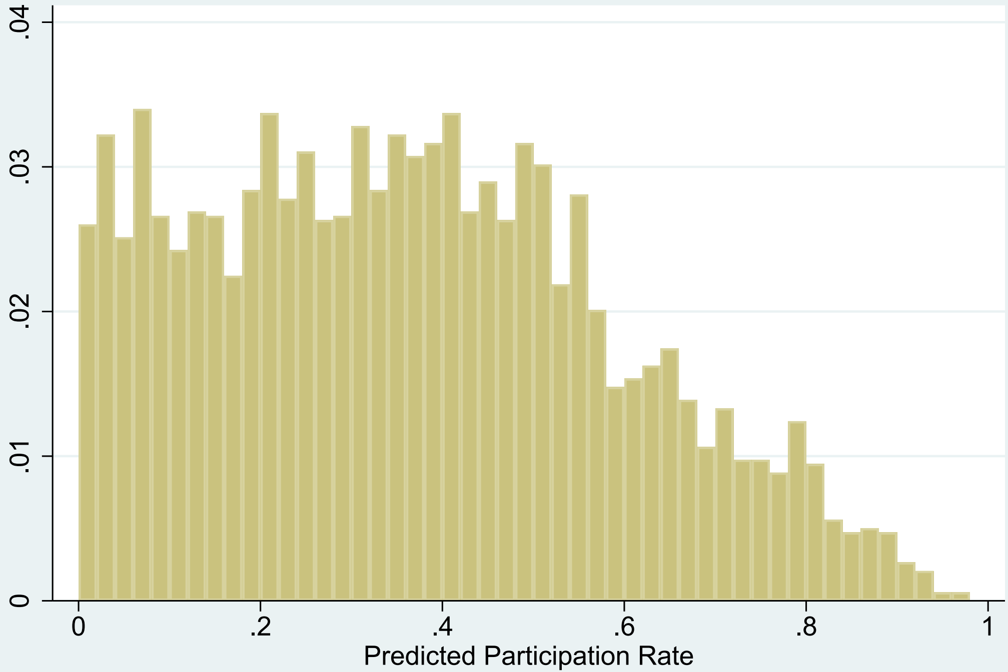


Figure 4: Increment Variance Distribution at Deciles of Baseline F

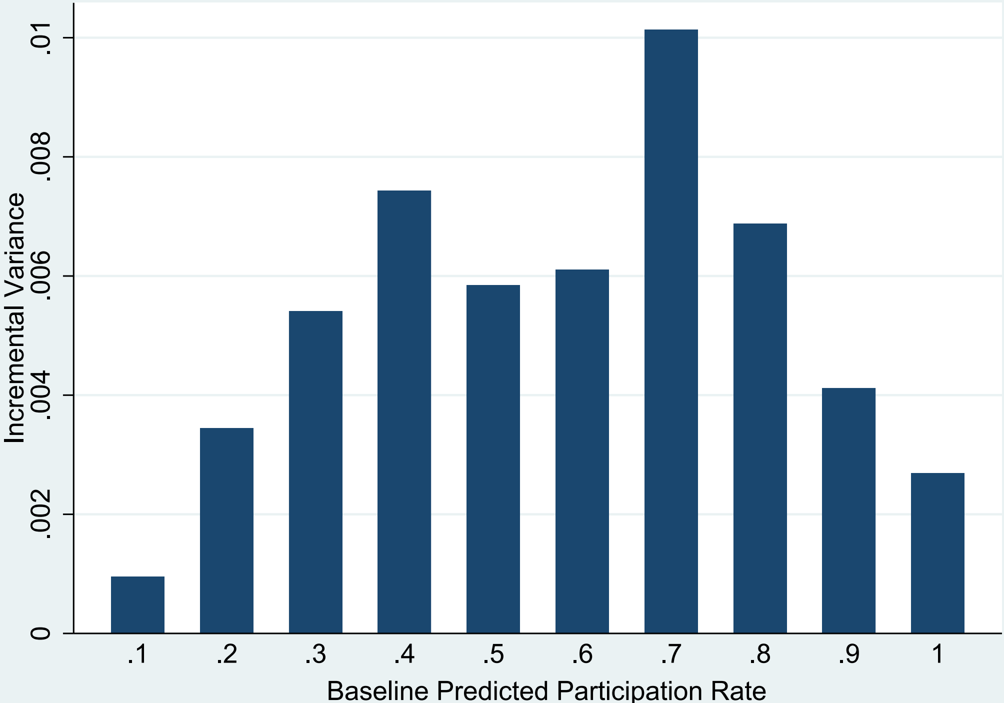


Figure 5. Marginal Labor Supply Curves for Different Natural Cubic Splines
95 percent confidence intervals shown

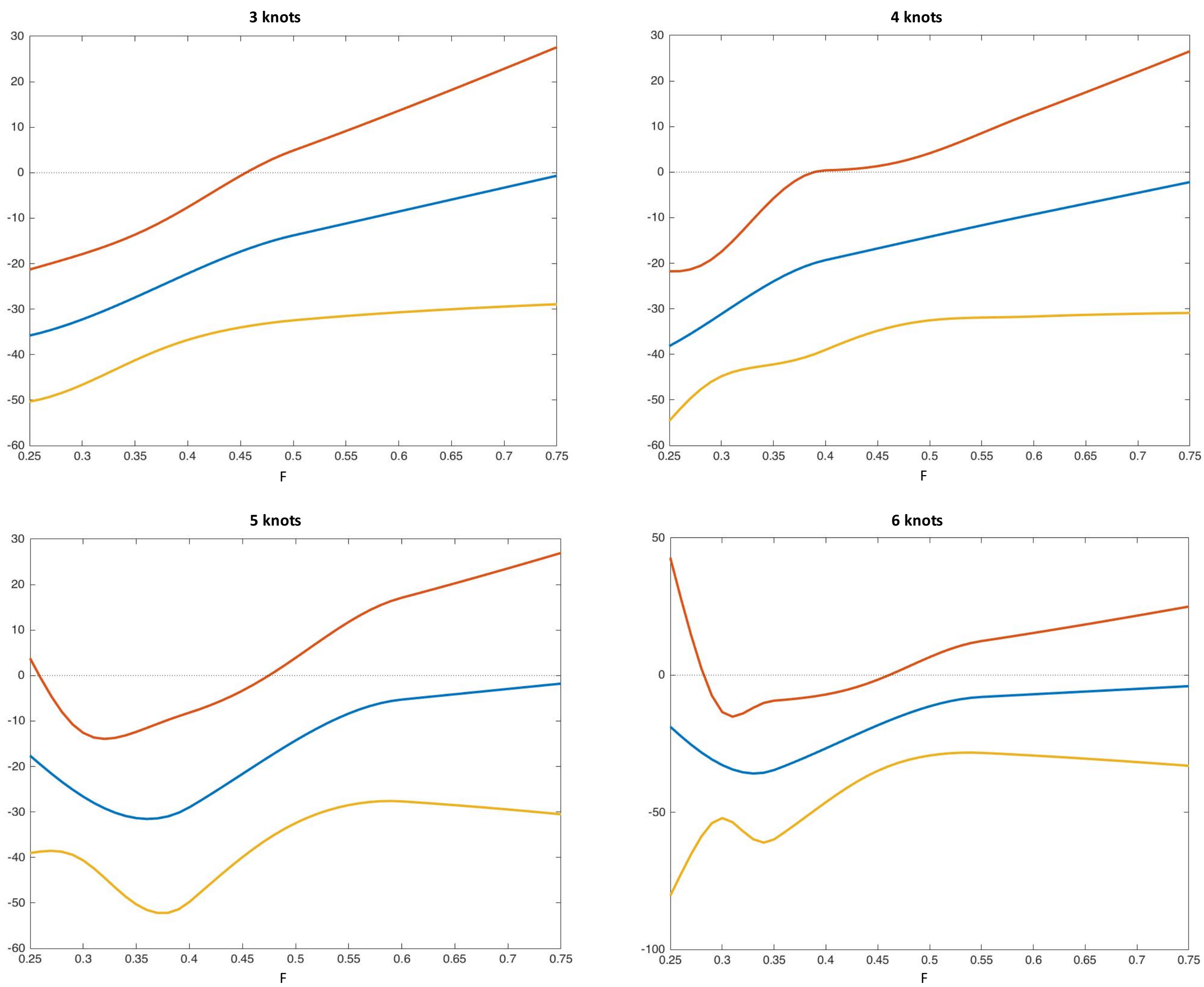


Figure 6. Marginal Labor Supply Curve for Five-Knot Model with Interactions

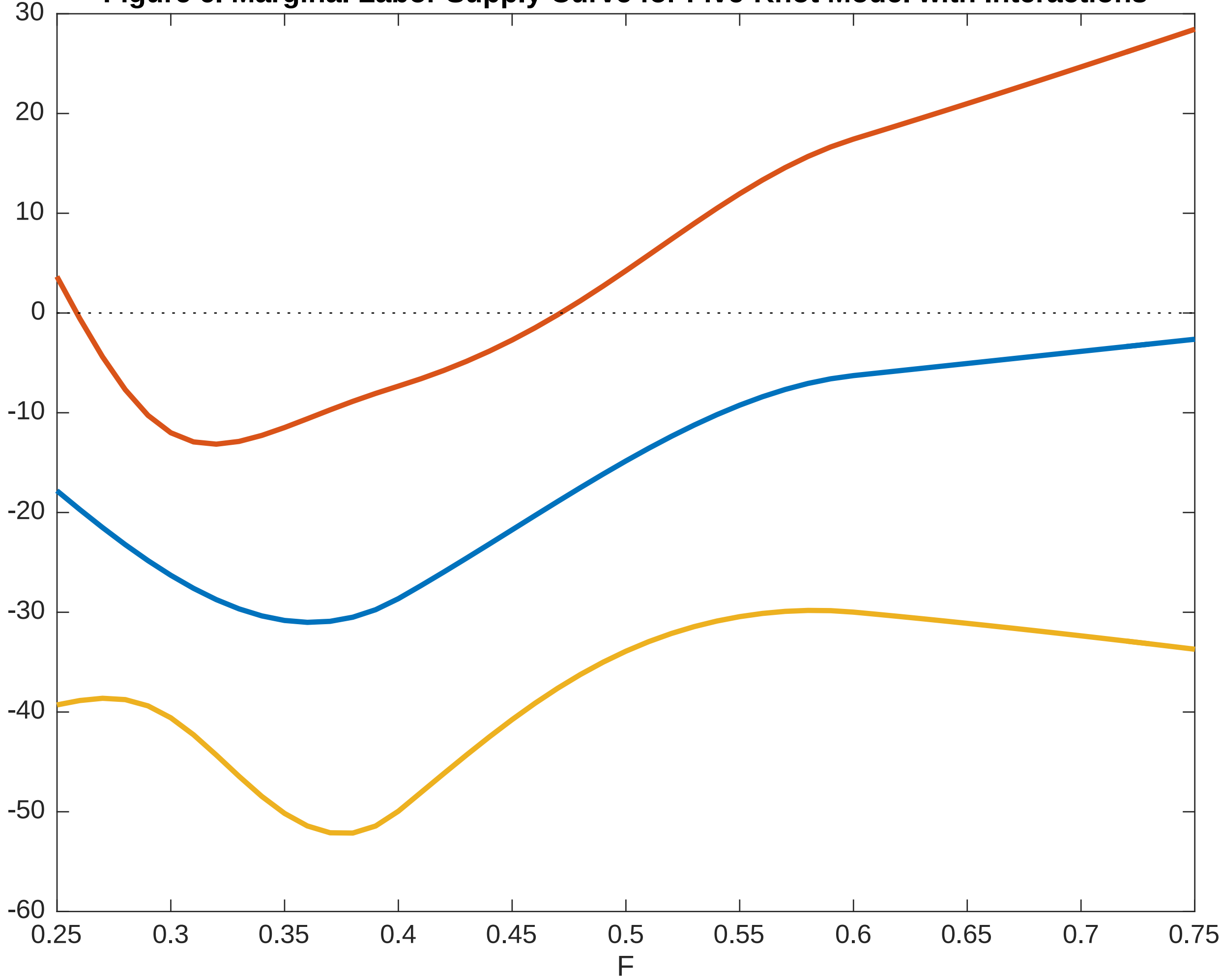
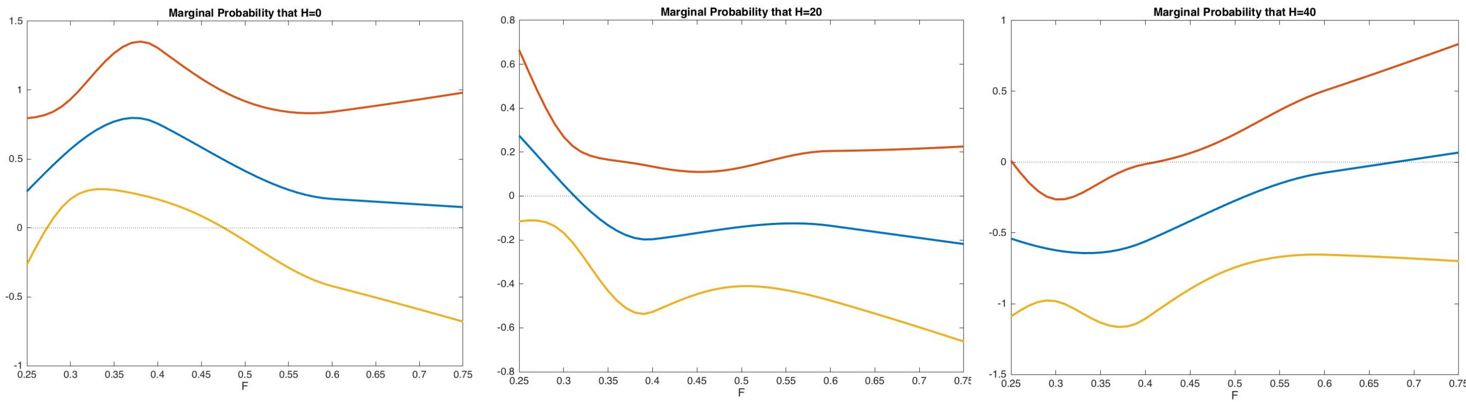


Figure 7



Marginal Probability that H=0
Marginal Probability that H=20
Marginal Probability that H=40
F