Revisiting the Work Incentives of Welfare Programs

Robert Moffitt*

Johns Hopkins University

May, 2017

Abstract

The standard static model of the effect of welfare programs on labor supply is newly estimated with maximum response heterogeneity. Both reduced form and structural models are estimated allowing response heterogeneity whose shape is estimated nonparametrically. Estimated marginal treatment effects representing the effect of program participation on labor supply for the historical AFDC program are found to be quite heterogeneous, varying with the initial level of participation and initial levels of welfare program parameters, all of which affect the composition of those brought into or who leave the program on the margin. Applying the estimated model to major historical changes in benefits and tax rates in the AFDC program in 1967 and 1981 shows that those reforms had marginal effects on labor supply which were quite different because the populations were different.

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

There is a long and extensive literature on estimating the effects of welfare programs on labor supply. Studies in the early 1970s estimating labor supply equations with cross-sectional data to forecast the effects of welfare programs (Cain and Watts (1973)) were followed by results from a series of negative income tax experiments (Burtless (1987),Moffitt and Kehrer (1981)). A literature using more sophisticated econometric methods to handle the selectivity of welfare participation developed in the 1980s (Burtless and Hausman (1978),Hausman (1981),Moffitt (1983)) which was followed by a literature of reduced-form estimates of the effects of specific welfare reforms in the 1980s and 1990s (Gueron and Pauly (1991),Grogger and Karoly (2005)). Reviews of these literatures have been published by Danziger et al. (1981), Moffitt (1992), and Blundell and Macurdy (1999). For the most part, the parts of this literature which aim to estimate the effects of welfare tax rates and guarantees on labor supply have found them to be in the expected direction, negative in both cases. More recently, progress on dynamics and human capital effects of welfare programs have been studied (Blundell et al. (2016)).

Heterogeneity in labor supply response has, however, not been a source of major interest in this literature, yet the importance of heterogeneous response is increasingly recognized in other literatures. The older estimates of welfare effects using the standard static model typically assumed constant coefficients, for example, although this was occasionally relaxed, as in the Burtless and Hausman (1978) study, where the coefficient on nonlabor income was allowed to be heterogeneous but in a restrictive parametric form. The literature on nonlinear budget constraints of which the Burtless-Hausman paper was a part (including, as previously referenced, Hausman (1981) and Moffitt (1983)) did allow heterogeneous preferences in the additive error terms which were assumed to include preference heterogeneity and therefore affected utility on different segments of the constraint, so heterogeneity did have an impact on which segment to locate on but not on the change in hours of work from welfare participation for a given change in tax rate and benefit level. More recent structural models of labor supply responses to welfare programs typically have only a small number of heterogenous components and assume homogeneity for most parameters of interest. In reduced form models, a more direct focus of heterogeneity was the subject of the work of Bitler et al. (2006) and a follow-up paper by Kline and Tartari (2016), which demonstrated that some of the welfare reforms of the 1990s induced positive effects on labor supply and negative effects on others, depending on the initial level of hours of work.

This paper introduces and estimates a model of heterogeneity in the labor supply response to classic cash welfare program, defined as a program with a standard textbook structure offering a lump sum benefit for those who do not work and specify a tax rate at which the benefit is reduced as earnings rise (i.e., a negative income tax). While there are no existing programs of this type in the U.S., they nevertheless are of great historical importance and have important policy lessons for current programs. In the most important historical program in the U.S., the Aid to Families with Dependent Children (AFDC) program–the program which is studied here–the tax rate on earnings was 100 percent prior to 1967, when it was lowered to 66 percent, and the tax rate was then raised back to 100 percent in 1981 (the program was abolished in 1996). The basic benefit grew at different rates in different years and in different states as well. In a heterogeneous response model, all these reforms can have very different effects, depending on the distributions of the heterogeneity of the parameters of the model and on which portions of those distributions are affected by the reforms. The model estimated in this paper is applied to those specific historical reforms and implies that their effects on labor supply varied markly because of differences in the work incentives of the marginal recipients affected by the reforms.

In a structural model with heterogeneous parameters, a conventional selection bias problem arises just as it does in reduced form models, where those who participate in the program have different values of the counterfactual than those who do not. The conventional approach to this problem in reduced form models is to assume the existence of valid exclusion restrictions, and the same method will be applied here. However, the reduced form treatment effects literature has also extensively explored the differences between marginal treatment effects and its closely related cousin, the LATE, and these concepts have typically not been applied in structural models. The model in this paper does so, focusing most heavily on the concept of a marginal treatment effect (MTE), applied here to the effect of program participation on labor supply of participants who are on the margin. The marginal treatment concept was introduced explicitly as a random coefficient model by Bjorklund and Moffitt (1987) following on the suggestion of such a model by Heckman and Robb (1985). The well-known LATE concept introduced by Imbens and Angrist (1994) is a discrete version. The most extensive work analyzing the MTE concept is that of Heckman and Vytlacil (1999), Heckman and Vytlacil (2005), and Heckman et al. (2006), and the idea has now been applied in a number of different reduced-form settings.[1]

The model used here also incorporates the important feature of almost all welfare programs of incomplete, or partial, takeup. The percent of eligible families who participate in U.S. welfare programs is almost always far below 100 percent, which means that the takeup decision has to be modeled simultaneously with the labor supply decision (Moffitt (1983)). Models that assume that all eligible individuals receive the benefit and hence are on the envelope of their budget sets represent a specification error since eligible nonparticipants do not actually face those sets. There is obviously heterogeneity in takeup which may be correlated with hetereogeneity of labor supply responses and hence it must be incorporated. In addition, methodologically, it is incomplete takeup that connects models of welfare effects to the treatment effects literature because, if takeup is complete, there is no participation decision separate from the labor supply decision.

[1] For recent work on extrapolating estimating marginal treatment effects to larger supports of the distribution, see Brinch et al. (2017) and Kowalski (2016). For work on marginal treatment effects in dynamic treatment models, see Heckman et al. (2016).

The application here is to the AFDC program prior to the 1990s reforms, which roughly fit the classic welfare program model. Estimates of a static labor supply model with endogeneous takeup which allows full unobserved heterogeneity in substitution and income effects shows their distributions to have significant dispersion. The estimated model is used to demonstrate the impact of the 1967 reduction in the welfare tax rate, the 1981 increase in that tax rate, the changes in the basic benefit that occurred over the 1970s and 1980s. The model estimates imply that the marginal recipient in 1967 would have had a smaller work reduction from program participation than in the late 1980s, but the effects of increases in the tax rate were greater. Simulations showing the marginal effects of tax rate changes at current participation rates are yet different again. Thus the labor supply effects of changes in tax rates depend on the composition of the initial population which varies over time.

The paper is organized as follows. The first section lays out the simple static labor supply model when takeup is endogeneous and all responses are heterogeneous. The analysis demonstrates that marginal treatment effects–that is, the net labor supply response from increasing or decreasing the caseload and hence drawing individuals onto or off of the program–are ambiguous in sign and can be either increasing or decreasing in the level of the takeup rate. The second section introduces the data and estimates marginal treatment effects in a reduced form model where participation in the program per se has heterogeneous effects. The third section reestimates a structural static model of labor supply and shows how marginal treatment effects work through the distributions of the wage and income elasticities. The fourth section (to be completed) uses the estimated model to simulate the effects of historic reforms of the AFDC program (1967 and 1981 changes in tax rates, changes in guaranteed benefit levels, etc.) as well as hypothetical reforms that involve manipulation of the welfare program parameters. The last section suggests directions for future research, including the extension to dynamic models and to more complex welfare programs that those with the classic NIT form.

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

The canonical static model of the labor supply response to transfers with incomplete) takeup (Moffitt (1983)) posits 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 separability of P_i from the U function is for analytic convenience and is not required for any of the following results. Allowing for fixed costs of participation–in money, time, or utility, with the exact type unspecified–is required because many individuals who are eligible for transfer programs do not participate in them. If this were not the case, then all individuals would locate on the boundary of their budget sets and program participation would be automatically determined by the choice of H, meaning that there would be no separate participation decision.

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

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

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

The response to the program for individual i is

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

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. Letting

$$S_\phi = \sup \quad \phi \tag{7}$$

$$S_{\theta(\phi)} = set \;\; of \;\; \theta \;\; s.t. \;\; P_i = 1 \;\; conditional \;\; on \;\; \phi \tag{8}$$

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

$$\widetilde{\triangle} = E(\triangle_i P_i \mid W_i, N_i, G, t, r) \tag{9}$$

$$= \int_{S_\phi} \int_{S_{\theta(\phi)}} \triangle_i \left(\theta_i \mid W_i, N_i, G, t, r\right) dG(\theta_i, \phi_i) \tag{10}$$

where $G(\theta_i, \phi_i)$ is the joint c.d.f. of the two heterogeneity components. Note that the two sets S are functions of the budget constraint parameters, which is not made explicit. Letting S_θ be the unconditional support of θ, the participation rate in the population is

$$P = E(P_i \mid W_i, N_i, G, t, r) \tag{11}$$

$$= \int_{S_\phi} \int_{S_\theta} 1\{V[W_i(1-t), G + N_i(1-r); \theta_i] - V[W_i, N_i; \theta_i] - \phi_i\} dG(\theta_i, \phi_i) \tag{12}$$

and the mean labor supply response among those who participate is

$$\widetilde{\triangle}_{P_i=1} = \widetilde{\triangle}/P \tag{13}$$

The marginal response to a change in program participation, which is often interpreted as the mean $\triangle$ of those who change participation, is $\partial\widetilde{\triangle}/\partial P$. These effects have been discussed extensively in the treatment effects literature and are defined within the econometric model in the next section.

The distribution of θ_i affects the mean response in the population in two ways: first, by affecting the distribution of $\triangle_i$ across the population–that is, the distribution of response if all individuals participate–and, second, by altering which of those individuals participate because θ_i appears in eqn(4). The distribution of ϕ_i affects mean response only through the latter mechanism, by altering the composition of the participant population; this feature will lead to an exclusion restriction in the econometric model below.

While $\triangle_i$, $\widetilde{\triangle}$, $\widetilde{\triangle}_{P_i=1}$, and $\partial\widetilde{\triangle}/\partial P$ must be negative according to theory, how they change as the participation rate changes is less clear and requires making a distinction between different sources of change in participation. How the effect varies with a change in participation induced by a change in ϕ_i, for example, is ambiguous in sign because the magnitude of $\triangle_i$ has no determinate relationship to the magnitude of the non-cost portion of the utility gain of going onto welfare, $dV = V[W_i(1-t), G + N_i(1-r); \theta_i] - V[W_i, N_i]$. For example, those with greater gains dV may be those with greater marginal utilities of consumption and hence those with smaller marginal utilities of leisure; it is the relative marginal utility of consumption and leisure that matters. An increase in participation induced by a reduction in ϕ will draw new individuals onto welfare whose values of dV are smaller than those of initial recipients (for any given value of ϕ, participation is positively selected on dV), but those smaller values of dV could be associated with either greater or smaller labor supply reductions. Thus, one central question of the analysis can only be determined empirically.

Participation rate expansions induced by changes in the budget constraint, on the other hand, have quite different effects because they induce changes in mean labor supply reductions for those initially on welfare as well. An expansion of the generosity of the program, for example, will increase participation and necessarily increase mean labor supply reductions. Thus $\triangle_i$, $\widetilde{\triangle}$, and $\widetilde{\triangle}_{P_i=1}$ will necessarily become more negative as participation rises. However, this gross marginal response, $\partial\widetilde{\triangle}/\partial P$–that is, not holding the budget constraint variables fixed–cannot be interpreted as the mean response of those brought into the program because it will include not only their responses but also the mean increase in labor supply reductions of those initially on the program. But the correlation between $\triangle_i$ and dV will still be at play in this case because it will determine whether the labor supply reductions of the new entrants are greater or smaller than those of the initial recipients after both face the same new budget constraint. Consequently, heterogeneity in response may make the increase in labor supply reductions arising from budget constraint expansions greater or smaller than would be predicted if responses had been assumed to be homogeneous and unchanging as the program expands. These effects will be separately identified in the econometric model in the next section.

II Reduced Form Model

The model to be estimated is that in eqns(3)-(5). It is useful to first estimate that model in reduced form before proceeding to the structural model for a number of reasons. The approach to the estimation of marginal treatment effects to be used in the structural model is based on that used in the reduced form model, where the intuition is more clearly visible and where the role of the instruments is clearest. The main results for the pattern of heterogeneous responses will, in fact, be the same in both models, but the structural model will show how that pattern manifests itself in parameters of a utility function and in wage and income elasticities of labor supply; thus the contribution of the structural model is best gauged against a reduced form model. case.

Therefore let H_i be a function of P_i as in (3) and assume that all budget constraint and other covariates have been conditioned out. Likewise, condition out all observables in (4) except for an observable proxy for fixed costs, Z_i. An unrestricted reduced-form model of (3)-(4) with full individual heterogeneity can be written as follows:

$$H_i = \beta_i + \alpha_i P_i \tag{14}$$

$$P_i^* = m(Z_i, \delta_i) \tag{15}$$

$$P_i = 1(P_i^* \geq 0) \tag{16}$$

where β_i and α_i are scalar random parameters and δ_i is a vector of random parameters.[2]

[2] In the empirical analysis below, corner solutions at $H = 0$ will be allowed in the data but (14)-(16)

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 of this type exists for each individual i. The function m can likewise be unrestricted and can be saturated if Z_i is assumed to have a multinomial distribution, although we shall discuss restrictions on δ_i below. The object of interest is the distribution of α_i. Selection in this model can occur either on the intercept (β_i) or the slope coefficient (α_i) because both may be related to δ_i and, in fact, the theoretical model implies that they must be because the participation equation contains the parameters of the labor supply function. Assuming that the three parameters in (14)-(16) are mean independent of Z_i, we can condition both equations on it to determine what is identified and estimable:

$$E(H_i \mid Z_i = z) = E(\beta_i \mid Z_i = z) + E(\alpha_i \mid P_i = 1, Z_i = z)\,\Pr(P_i = 1 \mid Z_i = z) \tag{17}$$

$$E(P_i \mid Z_i = z) = \Pr[m(z, \delta_i) \geq 0] \tag{18}$$

What we wish to identify is $E(\alpha_i \mid P_i = 1, Z_i = z)$ (if we can identify that, we can also integrate over the support of Z_i to obtain the mean of α_i conditional only on participation). Identification requires that Z_i satisfy two mean independence requirements, one for the intercept and one for the slope coefficient:

$$A1.\;\; E(\beta_i \mid Z_i = z) = \beta \tag{19}$$

$$A2.\;\; E(\alpha_i \mid P_i = 1, Z_i = z) = g[E(P_i \mid Z_i = z)] \tag{20}$$

where g is the effect for those on the program (i.e., the effect of the treatment on the treated) conditional on Z_i, and depends on the shape of the distribution of α_i and how

will continue to represent the reduced form of the model. With corner solutions, equation (4) must be modified to represent the choice of direct utility at $H = 0$ versus utility at $H > 0$. The reduced form of that representation is still of the form in (14)-(16).

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 (20) 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. If this were not so, different values of Z_i would lead to different conditional means of α_i through some other channel, which would rule it out as a valid exclusion restriction.

The "monotonicity" condition of Imbens and Angrist (1994) constitutes, in this model, a restriction on δ_i and can be expressed as

$$P_i(Z_i = z) - P(Z_i = z') \; is \; zero \; or \; the \; same \; sign \; for \; all \; i \; for \; any \; \text{distinct} \; values \; z \; and \; z' \tag{21}$$

Inserting the two assumptions into the main model in eqns (17)-(18), and denoting the participation probability as $F(Z_i) = E(P_i \mid Z_i)$, we obtain two estimating equations

$$H_i = \beta + g[F(Z_i)]F(Z_i) + \epsilon_i \tag{22}$$

$$P_i = F(Z_i) + \upsilon_i \tag{23}$$

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 merely 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. The implication of the model is that preference heterogeneity is detectable by a nonlinearity in the response of the population mean of H_i (taken over participants and nonparticipants) to the participation probability. If responses are homogenous and hence the same for all members of the population, the function g reduces to a constant and therefore a shift in the fraction on the program has a linear effect on the population mean of H_i. If the responses of those on the marginal vary, however, the response of the population mean of H_i will depart from linearity. This feature of the heterogeneous-response treatment model has been noted by Heckman and Vytlacil (2005) and Heckman et al. (2006), and eqn(22) follows from their work. However, here it will form the basis of the estimation of the model and the conditional mean function in eqn(22) will be estimated directly.

Nonparametric identification of the parameters of the model–β and the function g at every point F–is straightforward and has been extensively discussed in the literature. F is identified at every data point Z_i from the second equation from the mean of P_i at each value of Z_i (ignoring 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 $\partial H/\partial F = g'(F)F + g(F)$. Here, the g function will be approximated by a nonparametric but continuous function which implicitly means that interpolation between the data points identifies the pointwise derivatives in the MTE.

Introducing exogenous covariates, let X_i^{β} denote a vector which includes W_i , N_i, and sociodemographic characteristics (age, education, family composition, etc.), all of which affect labor supply when off welfare. Let X_i denote a vector which augments X_i^{β} with the welfare-program variables G and t, which will affect labor supply on welfare; X_i will shift the function g, the effect of welfare on labor supply. The vector X_i will affect the probability of participation as well, thus shifting F. While extensive interaction is in principle possible by estimating the model separately for every set of values of these covariates, a less ambitious and more conventional approach will be taken here, which is to introduce index functions of covariates and to allow these index functions to affect the means of β, g, and F, and to be additively separable with Z. With this formulation, the model specializes to

$$H_i = X_i^{\beta}\beta + [X_i\lambda + g(F(X_i\eta + \delta Z_i))]F(X_i\eta + \delta Z_i) + \epsilon_i \tag{24}$$

$$P_i = F(X_i\eta + \delta Z_i) + \nu_i \tag{25}$$

which leaves only the functions g and F unspecified.[3] For g, we will estimate its shape with sieve methods but will assume normality for F.[4] With these two functions specified, we will employ two-step estimation of the model, with a first-stage probit estimation of eqn(25) and second-stage nonlinear least squares estimation of eqn(24) using fitted values of F from the first stage. Consistency of two-step estimation of conditional mean functions is demonstrated in Newey and McFadden (1994). Standard errors are obtained by bootstrapping.

III Application to AFDC, Data, and Reduced Form Results

The last significant U.S. welfare program to provide NIT-style cash benefits without restrictions imposed from work requirements, time limits, or other nonmonetary rules was the Aid to Families with Dependent Children (AFDC) program prior to 1994, so this is the program we study and which many others have studied in the past literature. Extensions

[3] Forcing the parameters β and λ to be fixed implicitly assumes that any heterogeneity in them does not vary with welfare participation. If they did, then they would have to be treated as α is being treated here, i.e., the coefficients would have to be made a function of F. Such an extension would be of interest for future work. The constancy of the parameters η and δ is innocuous since the conditional mean of P integrates out over the population distribution of those parameters.

[4] Equations (24)-(25) are equivalent to the classic Lee (1979) two-regime switching regression model but without the bivariate normal distribution assumptions in that model.

to other programs are discussed in the Conclusions when future work is discussed. We use data from the Survey of Income and Program Participation (SIPP), which is a set of rolling, short (12 to 48 month) panels which are representative samples of the U.S. population. We use all waves of panels interviewed in the Spring of each year 1988-1993 (only Spring to avoid seasonal variation) and pool them into one sample, excluding overlapping observations by including only the first interview when the person appears in more than one year (to avoid dependent observations, which would complicate the estimation). Eligibility for AFDC in this period required sufficiently low assets and income and, for the most part, required that eligible families be single mothers with at least one child under 18. Our sample is therefore restricted to such families, similar to the practice in past AFDC research. To concentrate on the AFDC-eligible population, we restrict our 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 5,151 observations.

The variables we use for estimation are average 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 we construct covariates for education, age, race, and family structure. The hourly wage is omitted because it is only available for workers and is assumed to be proxied by demographic characteristics, especially education. However, nontransfer nonlabor income is explicitly included among the covariates. The AFDC guarantee for a family of four in the individual's state of residence is also included. AFDC tax rates on earned and unearned income were nominally 100 percent over this period but, because of deductions and other exclusions, averaged .40 for earned income and .30 for unearned income on average (Ziliak (2007)). Because they are constant, they are not included in the model. The names, definitions, and means of the variables used in the estimation appear in Appendix Table A1. Thirty-one percent of the sample was on AFDC in the month prior to interview.

The exclusion restrictions (Z) are selected to proxy costs of participation in AFDC. There is a large social work literature on the use by states of non-financial and non-monetary mechanisms to keep AFDC caseloads down and to do so strategically, with the most common mechanism simply to deny eligibility as frequently as possible using subjective interpretations of the eligibility rules (Brodkin and Lipsky (1983), Lipsky (1984), Kramer (1990)). A quality control program operated by the federal government sent auditors each year to each state to determine whether states were erroneously assessing eligibility and denying applications. The error rates assessed in the program are available by state and year.The various error rates were pretested in OLS estimations of the welfare participation equation and, from this exercise, two emerged as consistently significant and with the expected sign, namely, negative on welfare participation rates: the error rate made by the state resulting in incorrect denial of eligibility and the percent of applications denied because of a failure on the part of the applicant to comply with all procedural requirements, which is a common indicator of the amount of paperwork and bureaucracy imposed on prospective recipients. Both error rates vary dramatically across states, from 2 percent to 11 percent for the first error rate and from 1 percent to 35 percent for the second error rate. Conditional on other individual characteristics, the F-statistics for these two variables are approximately 11 in the OLS welfare participation equations. Further details are given in Appendix A.

For estimation of marginal treatment effects, the overall power of the instruments is more complex than in the standard linear IV case because the instruments need to move the welfare participation probability (F above) around not just on average but at every value of that probability. This issue is important in the data here, as illustrated in Figures 2 and 3. Figure 2 shows a histogram of predicted welfare participation rates from probit estimation of that function, including both exogenous variates X as well as the instruments Z. The support of the distribution is concentrated at $H = 0$ but is also fairly good at ranges below about .50 but not above that. However, much of this variation is driven by the X vector. Figure 3 provides instead an indication of the explanatory power of the Z vector alone, showing the variance of the residual $[\widehat{F}(X, \overline{Z}) - \widehat{F}(X, Z)]$ plotted against deciles of the distribution of $\widehat{F}(X, \overline{Z})$. The instruments have reasonable incremental explanatory power for participation in the range (0.3, 0.8) but little power outside that range. Therefore it is unlikely that these instruments will have much power in estimating the MTE below 0.3 or above 0.8.

For the initial results we set $\lambda = 0$ (hence no interactions of X with participation) and estimate the hours equation by OLS, regressing hours on X and P. OLS gives a response estimate of -22.7 (s.e.=.50), which is only slightly smaller than the raw mean difference between participants and non-participants of -25.1, implying that conditioning on X has little effect. Next we estimate eqn(25) assuming a constant g, which is equivalent to the homogeneous-effect model and equivalent to 2SLS, though using probit for the first stage instead of the linear probability model. The estimate of the program effect is -38.5 (s.e.=3.6), larger than the OLS estimate, suggesting that OLS is biased downward. Full coefficient estimates are shown in Appendix Table A-2.

To obtain the greater information yielded by marginal treatment effects, the g function is specified as a series function with cubic splines as the bases of the function:

$$g(F) = \gamma_0 + \sum_{j=1}^{J} \gamma_j Max(0, F - \pi_j)^3 \tag{26}$$

where the π_j are preset spline knots. Cubic splines are the most commonly used order in the literature but because they typically are ill-behaved in the tails, they are commonly forced to be constant below the lowest knot and above the highest, which requires a slight modification in (26) (Hastie et al. (2009), pp.145-146). Fit is gauged by the generalized cross-validation (GCV) statistic, which is simply the sum of squared deviations adjusted for degrees of freedom. See Chen (2007) for consistency results for sieve models.

The object of interest in the estimation is the MTE, which is $\partial H / \partial F$. Estimation of

(24) with $\lambda = 0$ for spline specifications with 3, 4, 5, and 6 knots generates the MTEs and 95 percent confidence intervals shown in Figure 4. The GCV for all four are extremely close, differing only at the 4th significant digit, so all fit the data about the same. But all of them also show the same general pattern of non-monotonic MTEs, with labor supply disincentives rising then falling as participation rises. The 95 percent confidence interval bands show that the estimates are significantly different from zero only in the middle range of participation probabilities, which may be because the instruments have low power in those ranges, as noted above. The central range where the estimates re significant differs across the specifications but is approximately from .10 to .15 at the bottom to .60 to .80 at the top. The maximum work disincentive is large and over 40 hours per week, although the confidence bands range down to 35 or so. But most of the disincentive effects are lower than this, down to around 20 hours at low participation rates and down to around 10-15 hours at higher participation rates, all only while in the 95 percent significance band.

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

Estimates with $\lambda \neq 0$ allow the effect of participation on labor supply to differ by characteristics. Estimates are shown in Table 1. Those with higher levels of education, who have larger family sizes, and face higher welfare guarantees have greater work reductions when entering welfare. Those who are older, who have higher nonlabor incomes, and are in states with higher unemployment rates have smaller work reductions. In addition, these results, taken together with the other estimates of the model, allow a rough determination of the consistency of the results with the static labor supply model if education is taken as a proxy for the wage and if its estimated effects as well as those for nonlabor income and the welfare guarantee are considered together. For a classic NIT with a 100 percent tax rate, the effects of changes in wages, nonlabor income, and the welfare guarantee on hours of work are portrayed in Figures 5-7. An increase in the wage raises H off the program (assuming substitution dominate income effects), lowers the probability of being on the program, and increases the work reduction from going onto the program since hours are always reduced to zero from the non-program level. An increase in the level of nonlabor income reduces hours of work off the program, reduces the probability of being on the program, and reduces the work reduction from going onto the program. An increase in the welfare guarantee has no effect on H off the program, increases the probability of participating in the program, and increases the work reduction since those newly brought onto the program have higher H off the program. These effects are summarized in Table 2 by the parameter vector in the model where they appear. Tables 1 and A-2 show that these signs coincide in every case with estimates from the model.

IV Structural Model

Despite the correspondence of signs of the budget-constraint-related coefficients in the reduced form model, a structural model would identify those effects precisely by showing how heterogeneous parameters of the utility function are correlated with program participation as participation expands. Moreover, a structural model of labor supply would allow a superior estimation of the effects of different changes in the budget constraint, such as welfare guarantees and tax rates, for differing changes in those parameters will affect different parts of the preference distribution and hence have different effects. Accordingly, this section reports estimates of a structural model with marginal treatment effects and the next section will use the estimates to show the differing effects of varying welfare policies. However, the labor supply function conditional on participation will be fully specified.

Assume the quadratic utility function (Keane and Moffitt (1998)):

$$U_H = -\alpha_i H - \beta_i H^2 + Y - \delta_i Y^2 - \phi_i P \tag{27}$$

where H=hours of work, Y=income, and P=a welfare participation indicator and where concavity of the function requires $\alpha_i \geq 0$, $\beta_i \geq 0$, $(1/2Y) \geq \delta_i \geq 0$. The three parameters α_i, β_i, and δ_i, which map into the intercept, wage effect, and nonlabor income effect in the labor supply function, are allowed to be fully heterogeneous, implying that the labor supply response to any change in any budget constraint parameter will also be heterogeneous. In line with the approach of most of the literature on similar models, we should group the data on H takes into categories for nonwork, part-time work, and full-time work, and will evaluate the budget constraint options at the three weekly values of 0, 20, and 40. Denote the value of Y at each by Y_H. If the individual is on welfare,Y_H will be higher by the value of the welfare benefit. For purposes that will be clearer below, define a new hypothetical welfare participation dummy P', which is a dummy indicating whether we are calculating income including the welfare benefit or not. It will differ in some instances from whether the individual is actually on welfare, which is still denoted by P. Denote utility conditional on P and P' and at the three hours values as:

$$U_0(P', P) = Y_0(P') - \delta_i(P) Y_0^2(P') - \phi_i P \tag{28}$$

$$U_{20}(P', P) = -\alpha_i(P)20 - \beta_i(P)400 + Y_{20}(P') - \delta_i(P) Y_{20}^2(P') - \phi_i P \tag{29}$$

$$U_{40}(P',P) = -\alpha_i(P)40 - \beta_i(P)1600 + Y_{40}(P') - \delta_i(P)Y_{40}^2(P') - \phi_i P \tag{30}$$

where the values of the a, β, and y are allowed to be different for those on welfare and off. Let $Y_H(P') = WH + N + P'B_H$ with $B_H = Max(0, G - WH - N)$. Now define the three indicator variables =1 if H is observed at each of the three points: $I_0(P) = 1$ if $H = 0$, $I_{20}(P) = 1$ if $H = 20$,and $I_{40}(P) = 1$ if $H = 40$. Then the labor supply equation can be written

$$\begin{aligned} H &= 20[I_{20}(1)P + I_{20}(0)(1-P)] + 40[I_{40}(1)P + I_{40}(0)(1-P)] \\ &= [20I_{20}(0) + 40I_{40}(0)] + \{20[I_{20}(1) - I_{20}(0)] + 40[I_{40}(1) - I_{40}(0)]\}P \end{aligned} \tag{31}$$

and now compute the mean of H conditional on the instruments, Z (i.e.,the reduced form in terms of the structural parameters):

$$\begin{aligned} E(H|Z) = 20E[I_{20}(0)] + 40E[I_{40}(0)] + \{20[E(I_{20}(1)|P=1) - E(I_{20}(0)|P=1] \\ + 40[E(I_{40}(1)|P=1) - E(I_{40}(0)|P=1)]\} \Pr(P=1|Z) \end{aligned} \tag{32}$$

using the identifying assumption that Z only affects labor supply through welfare participation, and the first two terms assume that Z does not affect labor supply of people off welfare. We will plug in the predicted probability from the first stage for $\Pr(P = 1|Z)$ and then estimate this conditional mean equation with nonlinear least squares.

All that remains is to express the expected value of the six indicator variables in terms of the underlying model parameters: $E[I_{20}(0)]$, $E[I_{40}(0)]$, $E[I_{20}(0)|P = 1]$, $E[I_{40}(0)|P = 1]$, $E[I_{20}(1)|P = 1]$, and $E[I_{40}(1)|P = 1]$. Each of these denotes the probability of choosing one of the three hours points conditional on welfare participation and is determined by the maximum of the three values in (28)-(30), and the maximal value will depend on the values of an individual's preference parameters α_i, β_i, and δ_i. Assume that these parameters have respective untruncated, latent distributions $N(\overline{\alpha}(P), \sigma_\alpha^2)$, $N(\overline{\beta}(P), \sigma_\beta^2)$, and

$N(\overline{\delta}(P), \sigma_\delta^2)$, where

$$\overline{\alpha}(P) = g_0^\alpha + g^\alpha[F(Z)]P + X\theta^\alpha \tag{33}$$

$$\overline{\beta}(P) = g_0^\beta + g^\beta[F(Z)]P + X\theta^\beta \tag{34}$$

$$\overline{\delta}(P) = g_0^\delta + g^\delta[F(Z)]P + X\theta^\delta \tag{35}$$

where again $F(Z) = \Pr(P = 1|Z)$ and where $g^\alpha[F(Z)]$, $g^\beta[F(Z)]$, and $g^\delta[F(Z)]$ are the same spline functions employed in the reduced form model (but without an intercept) and where we let the g coefficients be different for each of the three parameters α, β, and δ.[5] Thus the key feature of the reduced form model, allowing the preference parameters to depend on the fraction of the population participating in the program, is retained in the structural model by allowing the means of the untruncated distributions of those parameters to depend on that fraction.

To maintain concavity of the utility function, the distributions of the three parameters will required to operate only in the permissible ranges $\alpha_i \geq 0$, $\beta_i \geq 0$, $(1/2Y) \geq \delta_i \geq 0$. The actual distributions of the three parameters will therefore follow truncated normal distributions with well-known truncated means and variances involving the parameters of the untruncated distributions. The expression of the six conditional mean functions in terms of the underlying parameters is straightforward and can be computed by integrating over the values of the three parameters over their permissible ranges. The expressions are given in Appendix B. [6]This completes the specification of the model.[7]

The conditional mean function in (32) can again be estimated by nonlinear least squares, using the same natural cubic spline specification with three knots used in the

[5]For notational simplicity, the conditioning of all expectations on an exogenous covariate vector X is left implicit.

[6]Because there are only three discrete outcomes, the variances of all three preference parameters are not identified. The variance of δ is fixed at .0001.

[7]A detail not specified here is that some individuals have high enough wages that they are not eligible for a benefit if $H = 20$ or $H = 40$. In those cases, alternatives are deleted from the choice set and the relevant probabilities are slightly different than those given here.

reduced form model.[8]. For this initial estimation, no exogenous covariates are included and the function therefore just includes the budget constraint variables. The estimates of the parameters of the model are given in Appendix Table A.3. The estimated coefficients themselves are not easily interpretable, as is often the case in models with nonlinear parameters, so the implied results for gross wage elasticities and income elasticities, and how they vary with the probability of participation, are instead shown in Figures 8 and 9. Figure 8 shows that the (uncompensated) wage elasticity is positive at low levels of the participation probability but falls and becauses negative, reaching its trough at $P = .50$, after which it rises as participation rates rise. The income elasticity, on the other hand, grows in size and reaches its largest (negative) value again at approximately $P = .50$. The net effect of the two elasticities is to generate the same concave profile of MTEs as was found in the reduced form model, as can be seen in Figure 10, which shows their values implied by the structural model (the MTE is, again, the derivative of (32) w.r.t. the participation probability). Marginal work reductions are at their lowest values at low levels of participation and grow larger as participation rates rise, peaking at about $P = .38$, after which the work reduction falls as participation increases beyond that point. The range of work reductions is considerably smaller in the structural model than in the reduced form model, possibly because a discrete specification of H is used in the former but a continuous value in the latter.

V Simulations of Alternative Welfare Reforms

This program examined here, the AFDC program, had a 30 percent participation rate among the low income population at the time of our data, the late 1980s and early 1990s. The program had a relatively low caseload in the early 1960s but it grew dramatically from the late 1960s through the 1970s, rising from 1 million in 1965 to 3.7 million in 1980. The

[8] Wage rates for nonworkers are required, and are predicted (for all individuals) from a first-stage equation that uses state-level labor market indicators for industry composition as identifiers.

participation rate in the late 1960s and early 1970s was considerably higher, around 70 percent according to other work. After 1996, the caseload dropped dramatically and today the caseload is about the same as it was in 1970, and the takeup rate is even smaller, around 10 percent.

The nominal tax rates in the program were originally 100 percent. They were reduced to 67 percent in 1967 but the effective tax rates on earned and unearned income by the late 1980s were .40 and .30, respectively. The nominal tax rate was increased by back to 100 percent in 1981. It was then reduced after 1996, to different levels in different states. The real guarantee in the program has also slowly declined over time, by about 20 percent since the 1970s.

Table 3 shows the implied marginal effects of changes in the tax rate and the guarantee at different participation rates, both for the heterogeneous response model estimated here as well as for a model of homogeneous effects where the parameters of the utility function are not allowed to change as the participation rate changes. At the participation rate of 0.3 in the data used here, the MTE is more negative if tax rates were to be increased to 100 percent but changes in the guarantee have very little effect. At a participation rate of 0.70, appropriate to the late 1960s and early 1970s, exogenous increases in participation would have had a smaller work reduction, but the effect of increasing tax rates to 100 percent would have been greater (i.e., increasing work reductions by more). At a participation rate of 0.10, approximately today's participation rate, exogenous increases in participation would have smaller effects of increases in the tax rate to 100 percent would have greater effects than would such an increase in the late 1980s and early 1990s. The constant effects model shown in the second column predicts the same work reduction regardless of the initial participation rate and thus misses the variation arising from heterogeneous response.

VI Summary and Conclusions

In this paper we have modified the traditional static labor supply model of the effect of income transfers to allow for heterogeneous response, implying that changes in welfare generosity have not only direct effects on labor supply of participants but which also change the composition of the caseload and hence change mean work disincentives through compositional effects. Using a modified version of the conventional treatment effects model and estimating the distribution of the unobserved response heterogeneity with series approximation methods, the results show that the marginal treatment effect on hours of work is U-shaped, first negative then positive. This implies that an increase in participation brings into the program individuals who have increasingly larger work disincentives than those initially on the program but eventually brings in those with increasingly smaller work disincentives. A structural model shows that wage and income elasticities vary with the margin of participation and that those with lower and lower labor supply levels are drawn into the program as participation expands, but that the sizes of the reductions increase up to a participation rate of about .70, after which they fall.

These methods can be applied to more complex welfare programs, with time limits and work requirements, with modification, and dynamic models are also possible. Methodologically, the incorporation of hetergeneous parameters into structural models in general should be an area of future work.

Appendix A

Means and Data Sources

The means and standard deviations of the variables used in the analysis are shown in Table A-1. The sources of the state-level variables are as follows. The AFDC guarantee is the monthly maximum amount paid for a family of four in the state, and is obtained from unpublished data provided to the author by the U.S. Department of Health and Human Services for all six years 1988-1993. The state rate of incorrect denial of eligibility was obtained from unpublished tables provided to the author by the U.S. Department of Health and Human Services, compiled from the federal AFDC quality control program. The state percent of applicants denied for failure to comply with procedural requirements was obtained from the issues of <u>Quarterly Public Assistance Statistics</u> published quarterly by the U.S. Department of Health and Human Services, data which were also obtained originally from the federal AFDC quality control program. Fluctuations from year to year in the two quality control variables were smoothed out by computing a weighted mean of each for each state over the 1988-1993 years.

Appendix Table A1

Means and Standard Deviations of the Variables Used in the Analysis

Variable Name	Variable Definition	Total sample	P=1	P=0
Hours	Average hours of work per week in the month prior to survey	21.8 (19.7)	4.5 (11.2)	29.6 (17.5)
P	Dummy variable equal to 1 if individual was on AFDC anytime in the month prior to survey	.31 (.46)	--	--
Age	Age in years at survey date	32.6 (9.0)	30.8 (7.9)	33.4 (9.4)
Education	Years of education at survey date	10.9 (1.9)	10.5 (2.0)	11.1 (1.9)
Family size	Number of individuals in the family at the survey date	3.3 (1.4)	3.4 (1.5)	3.3 (1.4)
No. Childress Lt 6	Number of children less than 6 in the family at the survey date	.70 (.85)	1.1 (1.0)	.51 (.72)
Black	Dummy variable equal to 1 if respondent is black	.34 (.48)	.45 (.50)	.30 (.46)

Appendix Table A1 (continued)

Variable Name	Variable Definition	Total sample	P=1	P=0
Nonlabor income	Nontransfer nonlabor income in the month prior to survey	389.2 (303.0)	75.0 (192.7)	407.1 (411.0)
Welfare G	State monthly AFDC guarantee for a family of four	419.2 (212.5)	440.0 (210.0)	409.0 (213.0)
Eligibility Denial	Percent of applicants incorrectly denied eligibility	4.93 (1.67)	4.77 (1.59)	5.00 (1.71)
Pctdenied	Percent of applications denied for failure to meet procedure requirements in the state	14.7 (8.8)	14.1 (8.3)	15.0 (9.1)
Sample size	--	5,151	1,600	3,551

Notes:

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

Appendix Table A2

Full Estimates for OLS and Basic 2SLS Specifications

	OLS	2SLS
Gamma	-22.7 (0.5)	-38.5 (3.6)
Beta		
Education	1.2 (0.1)	1.0 (0.2)
Age	0.8 (0.2)	0.1 (0.0)
Black	-1.8 (0.4)	-1.1 (0.6)
No. Children Lt 6	-1.1 (0.3)	0.1 (0.5)
Family size	-0.2 (0.2)	0.1 (0.3)
Nonlabor income	-0.2 (0.0)	-0.1 (0.0)
Unemployment Rate	-0.6 (0.1)	-0.4 (0.2)
Constant	18.3 (1.9)	26.9 (3.1)
Nu	--	
Education	--	-0.09 (0.01)
Age	--	-0.01 (0.00)
Black	--	0.25 (0.05)

Appendix Table A2 (continued)

	OLS	2SLS
No. Children Lt 6	--	0.26 (0.03)
Family size	--	0.10 (0.02)
Nonlabor income	--	-0.01 (0.00)
Welfare G	--	0.05 (0.01)
Unemployment Rate	--	-0.06 (0.01)
Constant	--	0.44 (0.21)
<u>Delta</u>		
Elig. Denied	--	-0.06 (0.01)
Pctdenied	--	-0.54 (0.27)

Notes:
Standard errors in parentheses
2SLS corresponds to Table 1, Column (1)

Appendix Table A3
Parameter Estimates of the Structural Model

<u>Alpha</u>	
g_0^{α}	1.2063
g_1^{α}	3.2884
g_2^{α}	-11.0115
$\sigma\alpha$	0.0778
<u>Beta</u>	
g_0^{β} *100	-1.551
g_1^{β} *100	-35.0267
g_2^{β} *100	156.02
$\sigma\beta$ *100	7.0861
<u>Delta</u>[1]	
g_0^{δ} / 10000	31.8451
g_1^{δ} / 10000	30.0628
g_2^{δ} / 10000	-151.0068

Notes:
1. Standard deviation of δ fixed at .0001.

VII Appendix B

Rather than integrating out over all three preference parameters, it is easier to make use of univariate normal distribution functions for one of the parameters conditional on the other two and then integrate out those other two. To this end, each of the six probabilities will be written as probabilities over α_i conditional on β_i and δ_i and the latter two will be integrated out. Define the following cutoff values for α_i conditional on β_i and δ_i that will define the sign of the three requisite utility differences:

$$A_{20,0}(P', P) = \{-400\beta_i(P) + [Y_{20}(P') - Y_0(P')] - [Y_{20}^2(P') - Y_0^2(P')]\delta_i(P)\}/20 \tag{36}$$

$$A_{40,20}(P', P) = \{-1200\beta_i(P) + [Y_{40}(P') - Y_{20}(P')] - [Y_{40}^2(P') - Y_{20}^2(P')]\delta_i(P)\}/20 \tag{37}$$

$$A_{40,0}(P', P) = \{-1600\beta_i(P) + [Y_{40}(P') - Y_0(P')] - [Y_{40}^2(P') - Y_0^2(P')]\delta_i(P)\}/40 \tag{38}$$

Then

$$U_{20}(P', P) - U_0(P', P) \gtreqless 0 \tag{39}$$

$iff\ \alpha_i \lesseqgtr A_{20,0}(P', P)$

$$U_{40}(P', P) - U_{20}(P', P) \gtreqless 0 \tag{40}$$

$iff\ \alpha_i \lesseqgtr A_{40,20}(P', P)$ and

$$U_{40}(P', P) - U_0(P', P) \gtreqless 0 \tag{41}$$

$iff\ \alpha_i \lesseqgtr A_{40,0}(P', P)$. Defining $\delta^U = 1/2Y$ and using the maximal value of Y in the data to define this upper limit, the six probabilities can be written as follows:

$$
\begin{aligned}
E[I_{20}(0)] &= \Pr(I_{20}(0) = 1) \\
&= \Pr[U_{20}(0,0) > U_0(0,0), U_{20}(0,0) > U_{40}(0,0)] \\
&= \Pr(\alpha_i < A_{20,0}(0,0), \alpha_i > A_{40,20}(0,0) | \beta_i, \delta_i, \alpha_i > 0, \beta_i > 0, \delta^U > \delta_i > 0) \\
&= \frac{1}{\Pr(\alpha_i(0) > 0)\Pr(\beta_i(0) > 0)\Pr(\delta^U > \delta_i(0) > 0)} \\
&\int_0^{\delta^U} \int_0^{\infty} Max\{0, F[(Max[0, (A_{20,0}(0,0)] - \overline{\alpha}(0))/\sigma_\alpha] - F[(Max[0, A_{40,20}(0,0)] - \overline{\alpha}(0))/\sigma_\alpha]\} \\
&f(\beta_i) f(\delta_i) d\beta_i d\delta_i
\end{aligned}
\tag{42}
$$

$$
\begin{aligned}
E[I_{40}(0)] &= \Pr(I_{40}(0) = 1) \\
&= \Pr[U_{40}(0,0) > U_0(0,0), U_{40}(0,0) > U_{20}(0,0)] \\
&= \Pr(0 < \alpha_i < Max[0, A_{40,0}(0,0)], 0 < \alpha_i < Max[0, A_{40,20}(0,0)] | \beta_i, \delta_i, \alpha_i > 0, \beta_i > 0, \delta^U > \delta_i > 0) \\
&= \frac{1}{\Pr(\alpha_i(0) > 0)\Pr(\beta_i(0) > 0)\Pr(\delta^U > \delta_i(0) > 0)} \\
&\int_0^{\delta^U} \int_0^{\infty} \{F[(Min[Max(0, A_{40,0}(0,0)) - \overline{\alpha}(0))/\sigma_\alpha], [Max(0, A_{40,20}(0,0)) - \overline{\alpha}(0))/\sigma_\alpha]] \\
&- F[-\overline{\alpha}(0)/\sigma_\alpha]\} f(\beta_i) f(\delta_i) d\beta_i d\delta_i
\end{aligned}
\tag{43}
$$

$$E[I_{20}(1)|P=1] = \Pr(I_{20}(1)=1|P=1)$$

$$= \Pr[U_{20}(1,1) > U_0(1,1), U_{20}(1,1) > U_{40}(1,1)]$$

$$= \Pr(\alpha_i < Max[0, A_{20,0}(1,1)], \alpha_i > Max[0, A_{40,20}(1,1)] | \beta_i, \delta_i, \alpha_i > 0, \beta_i > 0, \delta^U > \delta_i > 0)$$

$$= \frac{1}{\Pr(\alpha_i(1) > 0)\Pr(\beta_i(1) > 0)\Pr(\delta_i(1) > 0)}$$

$$\int_0^{\delta^U} \int_0^{\infty} Max\{0, F[(Max[0, A_{20,0}(1,1)] - \overline{\alpha}(1))/\sigma_\alpha] - F[(Max[0, A_{40,20}(1,1)] - \overline{\alpha}(1))/\sigma_\alpha]\}$$

$$f(\beta_i) f(\delta_i) d\beta_i d\delta_i \qquad (44)$$

$$E[I_{40}(1)|P=1] = \Pr(I_{40}(1)=1|P=1)$$

$$= \Pr[U_{40}(1,1) > U_0(1,1), U_{40}(1,1) > U_{20}(1,1)]$$

$$= \Pr(0 < \alpha_i < Max[0, A_{40,0}(1,1)], 0 < \alpha_i < Max[0, A_{40,20}(1,1)] | \beta_i, \delta_i, \alpha_i > 0, \beta_i > 0,$$

$$\delta^U > \delta_i > 0)$$

$$= \frac{1}{\Pr(\alpha_i(1) > 0)\Pr(\beta_i(1) > 0)\Pr(\delta^U > \delta_i(1) > 0)}$$

$$\int_0^{\delta^U} \int_0^{\infty} \{F[(Min[Max(0, A_{40,0}(1,1)), Max(0, A_{40,20}(1,1))] - \overline{\alpha}(1))/\sigma_\alpha] - F[-\overline{\alpha}(1)/\sigma_\alpha]\}$$

$$f(\beta_i) f(\delta_i) d\beta_i d\delta_i \qquad (45)$$

$$E[I_{20}(0)|P=1] = \Pr(I_{20}(0)=1|P=1)$$

$$= \Pr[U_{20}(0,1) > U_0(0,1), U_{20}(0,1) > U_{40}(0,1)]$$

$$= \Pr(\alpha_i < Max[0, A_{20,0}(0,1)], \alpha_i > [0, A_{40,20}(0,1)] | \beta_i, \delta_i, \alpha_i > 0, \beta_i > 0, \delta^U > \delta_i > 0)$$

$$= \frac{1}{\Pr(\alpha_i(1) > 0)\Pr(\beta_i(1) > 0)\Pr(\delta^U > \delta_i(1) > 0)}$$

$$\int_0^{\delta^U} \int_0^{\infty} Max\{0, F[((Max[0, A_{20,0}(0,1)] - \overline{\alpha}(1))/\sigma_\alpha] - F[(Max[0, A_{40,20}(0,1)] - \overline{\alpha}(1))/\sigma_\alpha]\}$$

$$f(\beta_i) f(\delta_i) d\beta_i d\delta_i \qquad (46)$$

$$E[I_{40}(0)|P=1] = \Pr(I_{40}(0)=1|P=1)$$

$$= \Pr[U_{40}(0,1) > U_0(0,1), U_{40}(0,1) > U_{20}(0,1)]$$

$$= \Pr(0 < \alpha_i < Max[0, A_{40,0}(0,1)], 0 < \alpha_i < Max[0, A_{40,20}(0,1)]|\beta_i, \delta_i, \alpha_i > 0, \beta_i > 0,$$

$$\delta^U > \delta_i > 0)$$

$$= \frac{1}{\Pr(\alpha_i(1) > 0)\Pr(\beta_i(1) > 0)\Pr(\delta^U > \delta_i(1) > 0)}$$

$$\int_0^{\delta^U} \int_0^{\infty} \{F[(Min(Max[0, A_{40,0}(0,1)], Max[0, A_{40,20}(0,1)] - \overline{\alpha}(1))/\sigma_\alpha] - F[-\overline{\alpha}(1)/\sigma_\alpha]\}$$

$$f(\beta_i)f(\delta_i)d\beta_i d\delta_i \tag{47}$$

These probabilities can be computed by numerical integration.[9]

References

Bitler, M., J. Gelbach, and H. Hoynes (2006). What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments. *American Economic Review 96*(4), 988–1012.

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

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

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

Brinch, C. N., M. Mogstad, and M. Wiswall (2017). Beyond LATE with a Discrete Instrument. *Journal of Political Economy*.

[9] The probabilities for those on welfare sometimes have zero benefits at either $H = 20$ or $H = 40$ or both, which can result if the guarantee is low and/or the wage is high. In that case, the probabilities of locating at those points are set equal to zero.

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

Burtless, G. (1987). The Work Response to a Guaranteed Income: A Survey of Experimental Evidence. In A. Munnell (Ed.), *Lessons from the Income Maintenance Experiments*. Boston: Federal Reserve Bank of Boston and Brookings Institution.

Burtless, G. and J. Hausman (1978). The Effect of Taxation on Labor Supply: The Gary Income Maintenance Experiment. *Journal of Political Economy 86*, 1103–1130.

Cain, G. G. and H. W. Watts (1973). *Income Maintenance and Labor Supply*. Boston,MA: Rand McNally.

Chen, X. (2007). Large Sample Sieve Estimation of Semi-Nonparametric Models. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B.

Danziger, S., R. Haveman, and R. Plotnick (1981). How Income Transfers Affect Work, Savings, and the Income Distribution: A Critical Review. *Journal of Economic Literature 19*(3), 975–1028.

Grogger, J. and L. A. Karoly (2005). *Welfare Reform: A Decade of Change*. Cambridge: Harvard University Press.

Gueron, J. M. and E. Pauly (1991). *From Welfare to Work*. New York: Russell Sage Foundation.

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

Hausman, J. (1981). Labor Supply. In H. Aaron and J. Pechman (Eds.), *How Taxes Affect Economic Behavior*. Washington: Brookings Institution.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016, March). Dynamic Treatment Effects. *Journal of Econometrics 191*(1), 164–75.

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). 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), 4730–34.

Heckman, J. J. and E. Vytlacil (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica 73*(3), 669–738.

Imbens, G. and J. Angrist (1994). 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.

Kline, P. and M. Tartari (2016, April). Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach. *American Economic Review 106*(4), 972–1014.

Kowalski, A. E. (2016). Doing More When You're Running LATE: Appying Marginal Treatment Effect Methods to Examine Treatment Effect Heterogeneity in Experiments. National Bureau of Economic Research WP 22363.

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

Lee, L. (1979). Identification and Estimation in Binary Choice Model with Limited (Censored) Dependent Variables. *Econometrica 47*(4), 977–996.

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

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

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

Moffitt, R. and K. Kehrer (1981). The Effect of Tax and Transfer Programs on Labor Supply: The Evidence from the Income Maintenance Experiments. In R. Ehrenberg (Ed.), *Research in Labor Economics*, Volume 4, pp. 103–50. Greenwich, Conn.: JAI Press.

Newey, W. and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume IV. Amsterdam: Elsevier.

Ziliak, J. (2007). Making Work Pay: Changes in Effective Tax Rates and Guarantees in U.S. Transfer Programs, 1983-2002. *Journal of Human Resources 42*(3), 619–642.

Table 1

Lambda Parameter Estimates

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

Notes:

Standard errors in parentheses.

Table 2

Expected Coefficient Signs Under Budget Constraint Interpretation

	λ	β	η
Wage (Education)	<0	>0	<0
Nonlabor income	>0	<0	<0
Welfare Guarantee	<0	0	>0

Table 3

MTEs for Different Scenarios

	Heterogeneous Effects Model	Constant Effects Model
P=0.3		
Current (G=actual, t=.40, r=.30)	-25.2	-15.13
t=r=1.0	-27.9	-26.93
G down by 20%	-26.0	-17.86
P=0.7		
Current	-19.5	-15.13
t=r=1.0	-26.5	-26.93
G down by 20%	-17.3	-17.86
P=0.1		
Current	-22.8	-15.13
t=r=1.0	-30.1	-26.93
G down by 20%	-22.0	-17.86

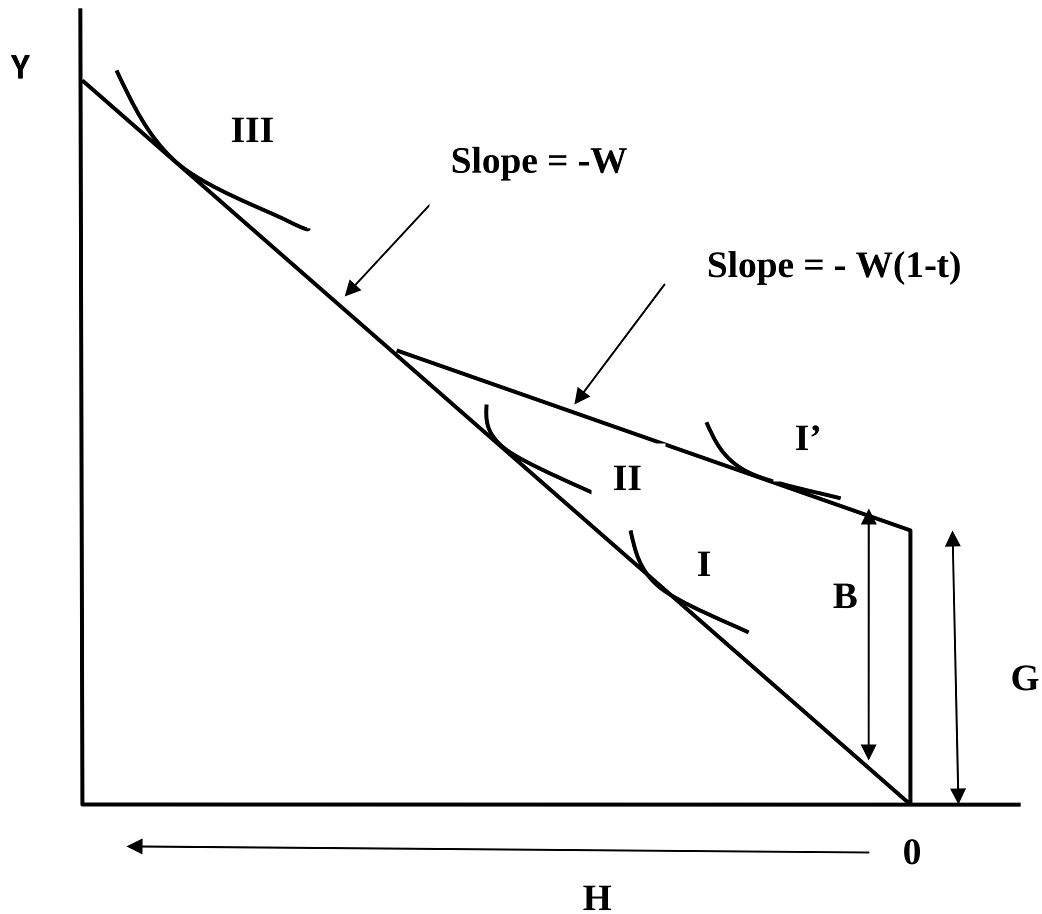
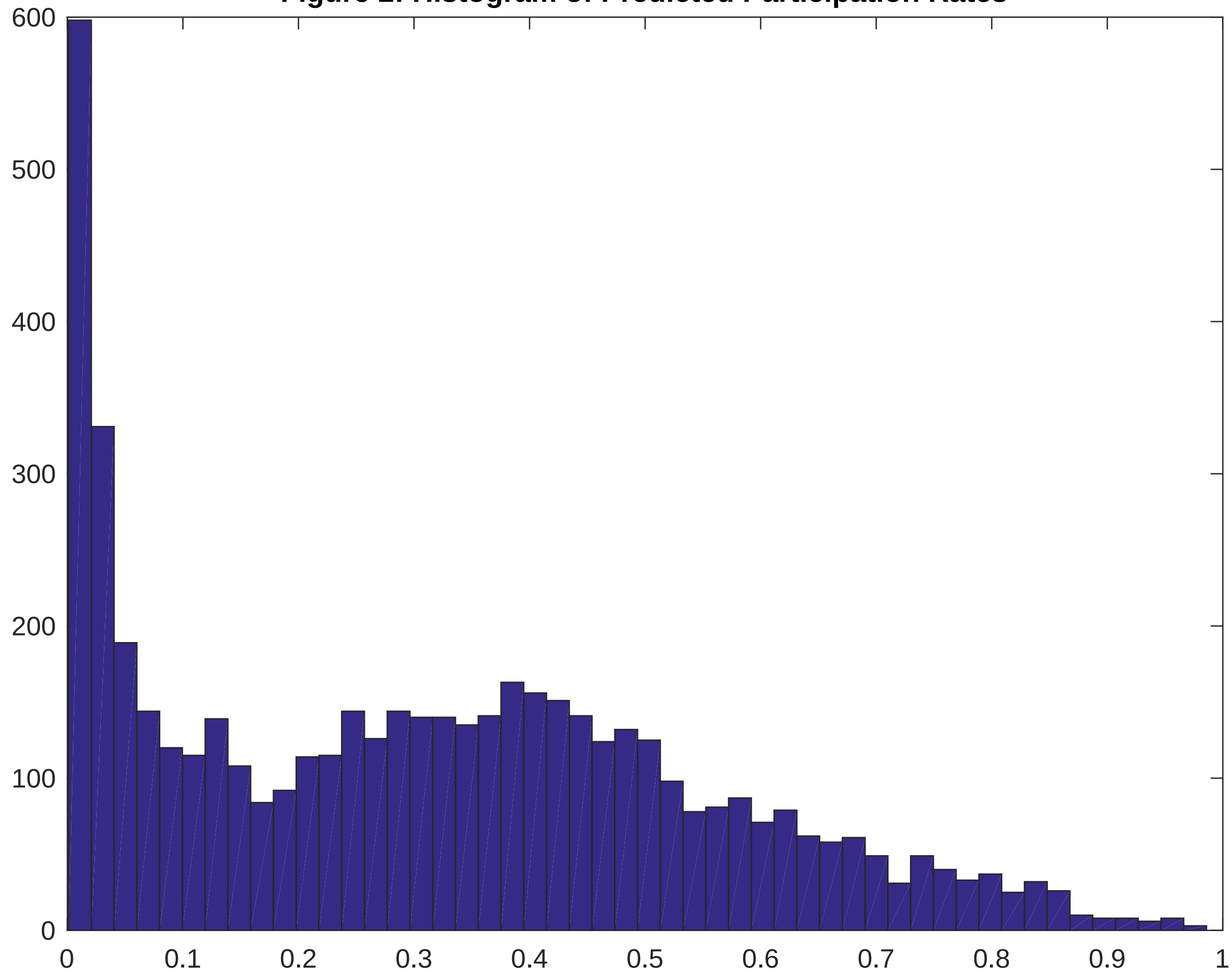


Figure 1

Figure 2: Histogram of Predicted Participation Rates



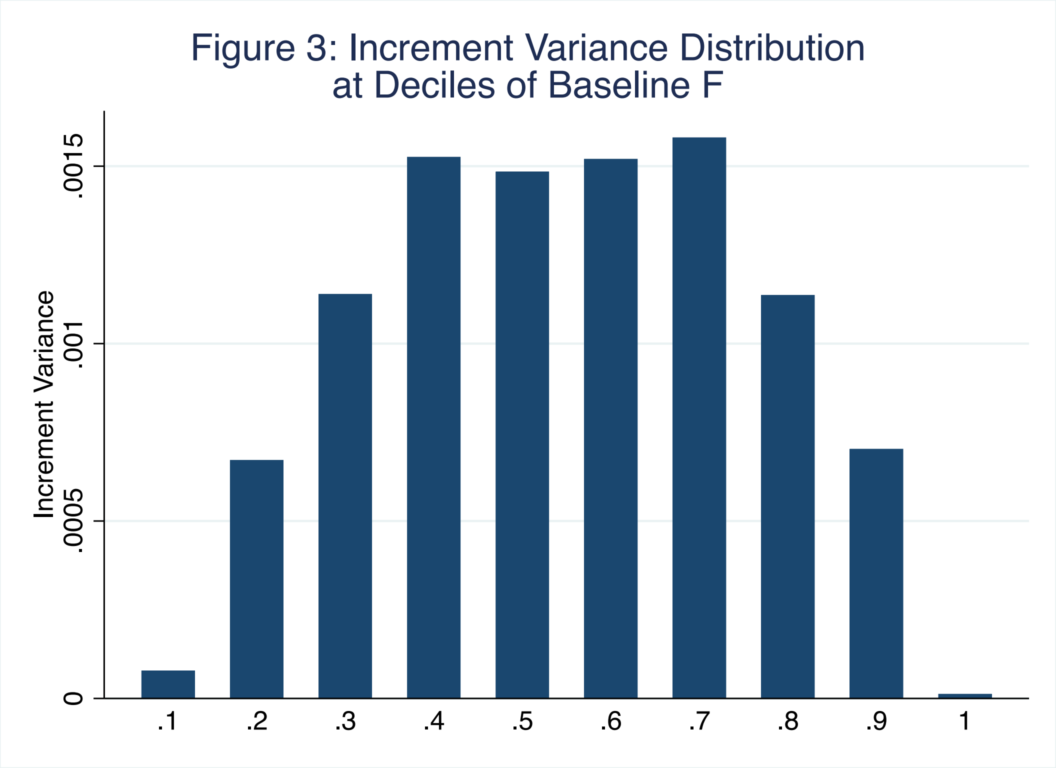
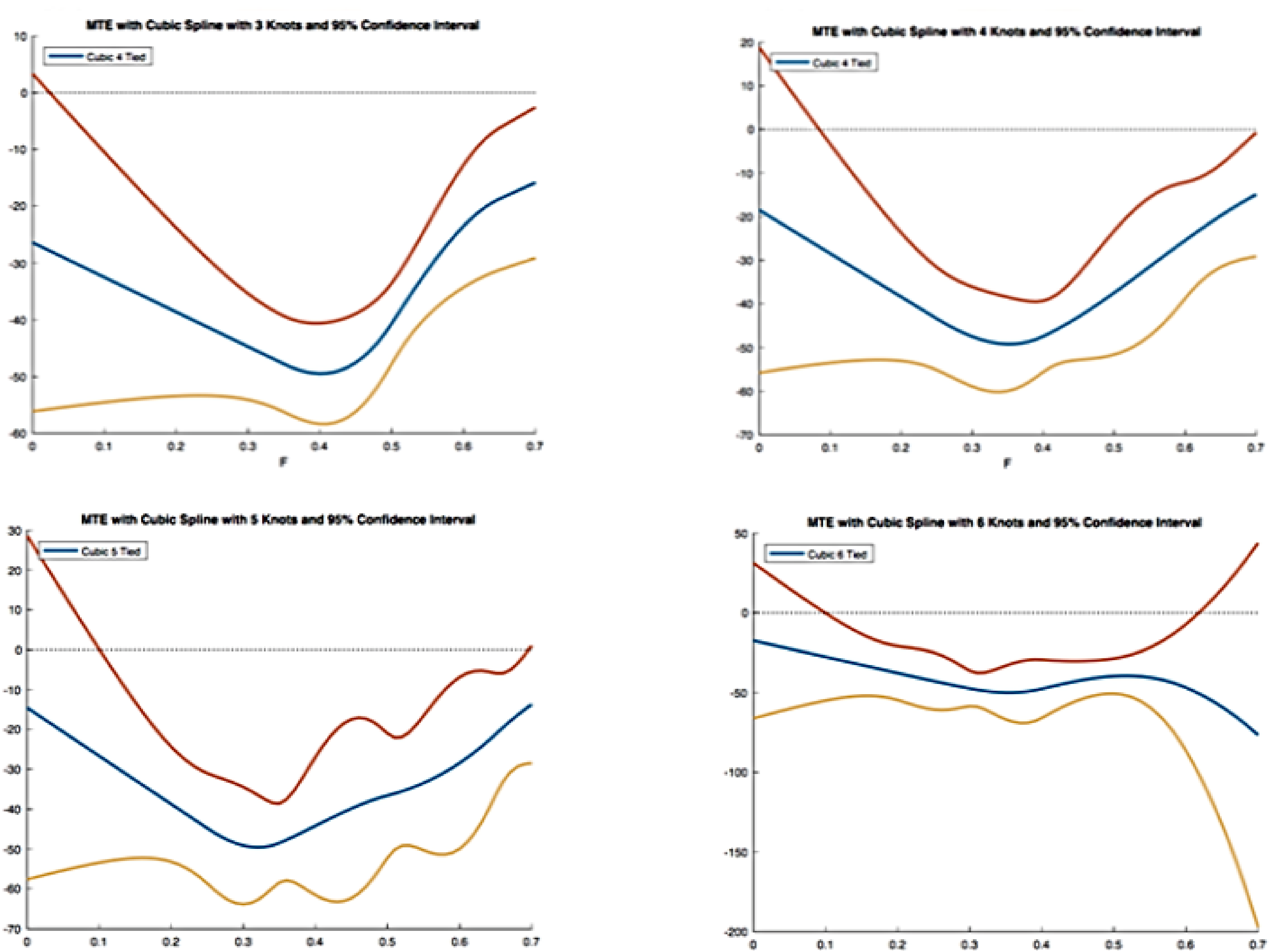


Figure 3: Increment Variance Distribution
at Deciles of Baseline F
Increment Variance
0
.0005
.001
.0015
.1
.2
.3
.4
.5
.6
.7
.8
.9
1

Figure 4: MARGINAL TREATMENT EFFECTS



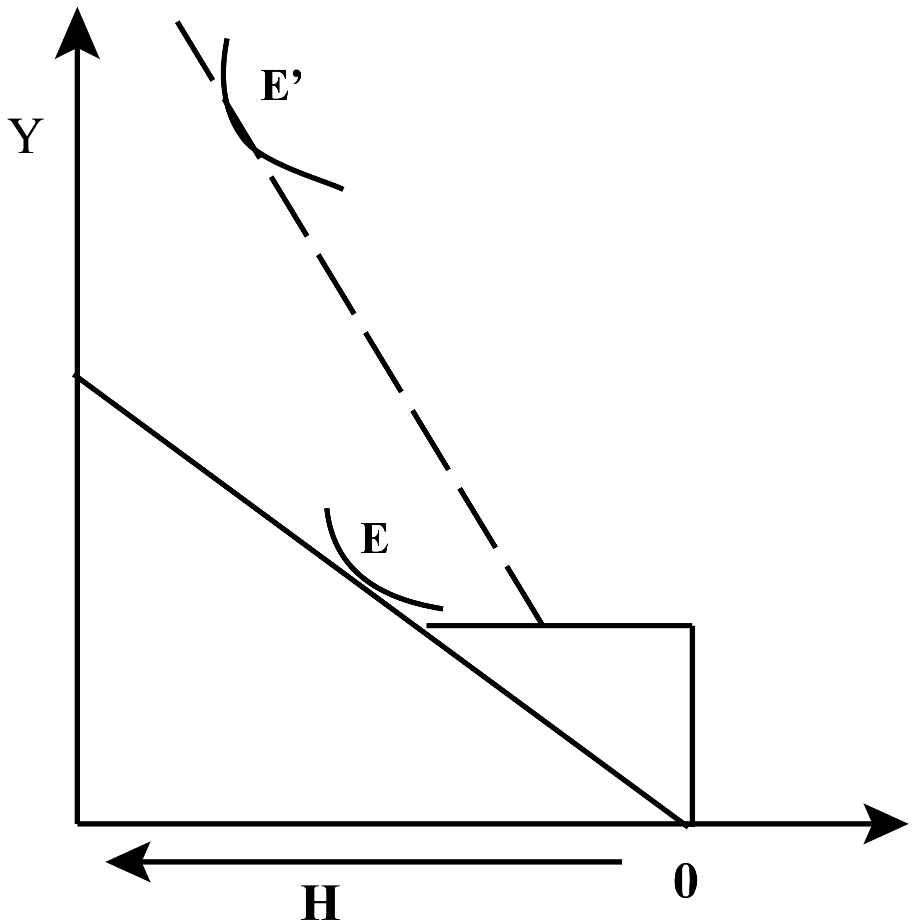


Figure 5. Effect of Increase in the Wage on Labor Supply assuming t=1

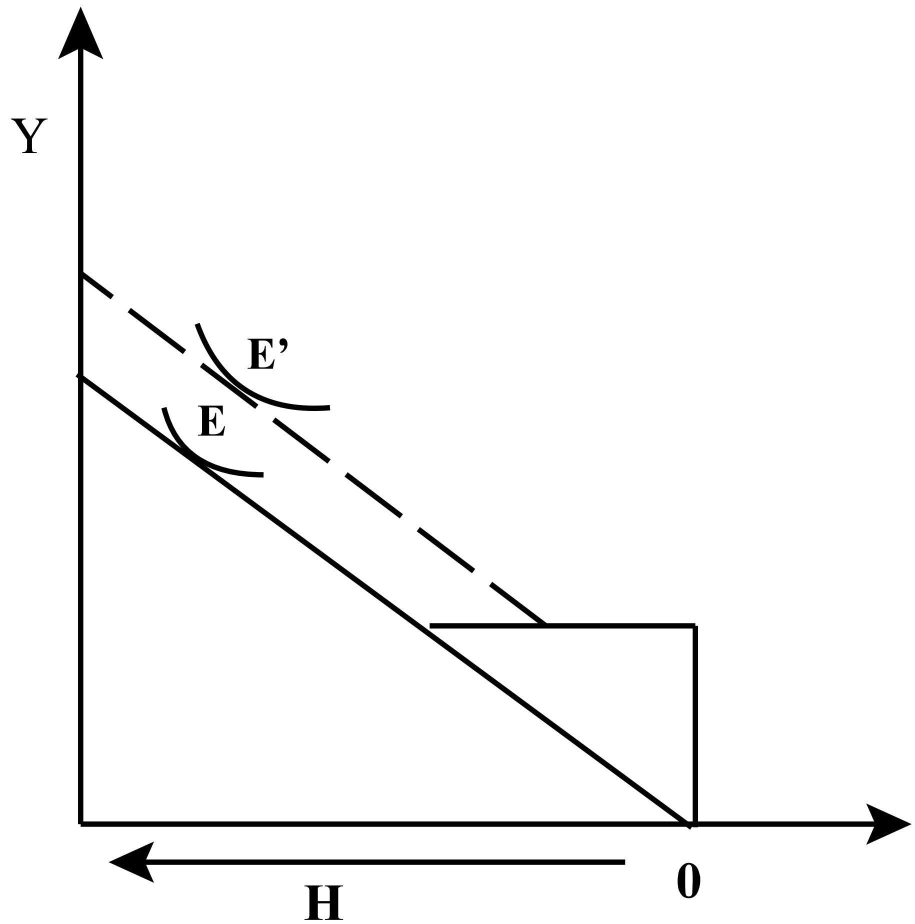


Figure 6. Effect of an Increase in Nonlabor Income on Labor Supply, assuming t=1

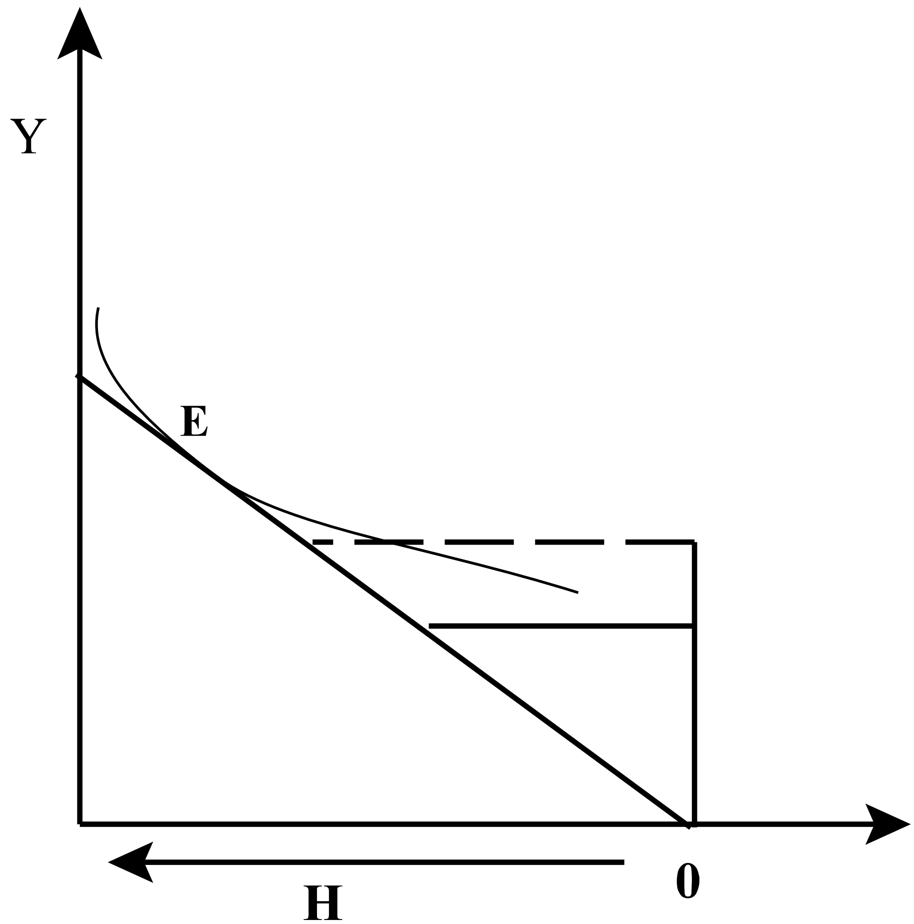


Figure 7. Effect of an Increase in G on Labor Supply, assuming t=1.

Figure 8: Wage Elasticity for Structural Model

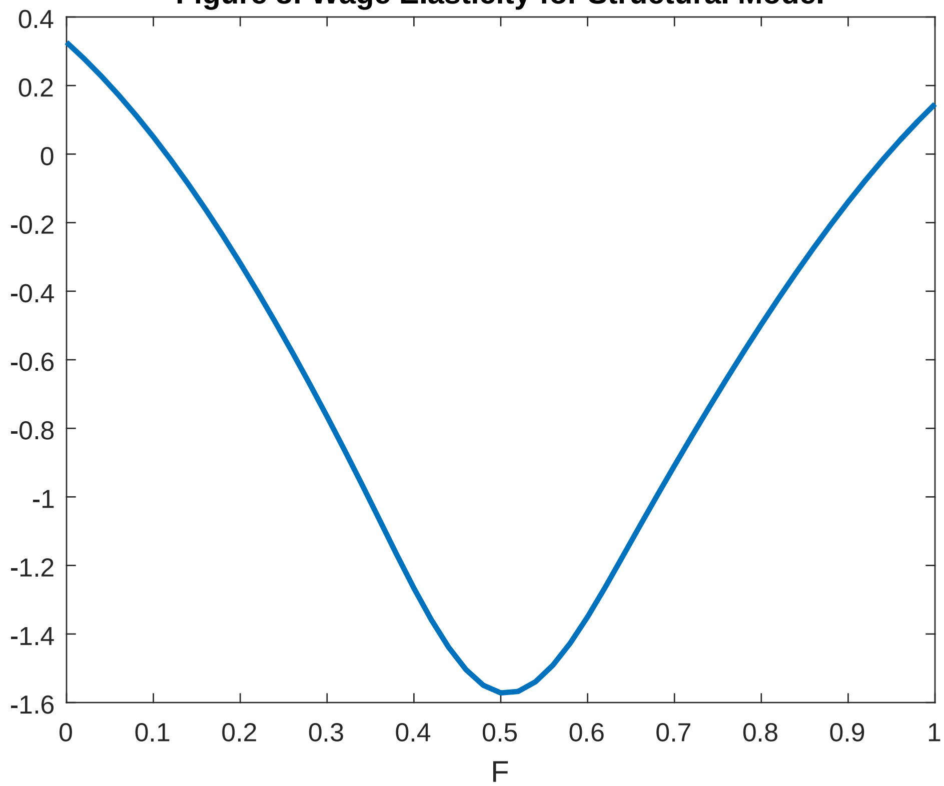


Figure 9: Total Income Elasticity for Structural Model

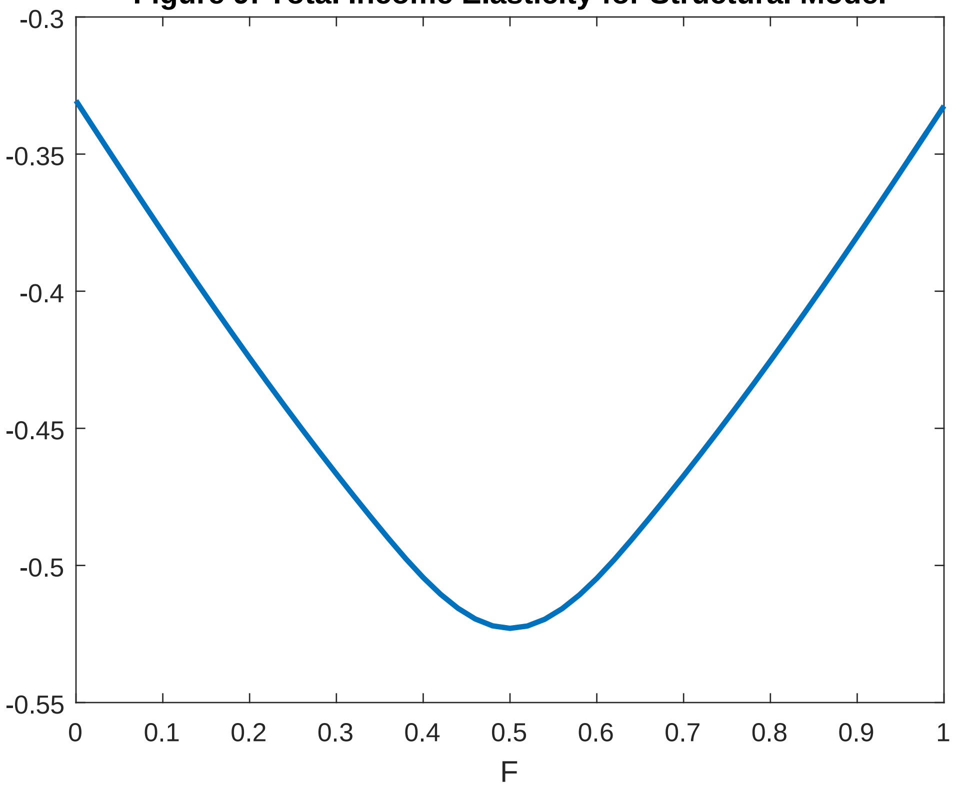


Figure 10: Marginal Treatment Effect at Individual Level