A Model of the Marginal Labor Supply Response to Transfer Programs, with a Historical Illustration*

Robert A. Moffitt†
Johns Hopkins University

Matthew V. Zahn
Georgetown University

Original: July 2019
Revised: September 2025

Abstract

We extend the textbook model of the labor supply response to a transfer program to incorporate responses of those on the margin of participation, which we term the marginal labor supply response. Government policies which alter program participation can have effects on the labor supply of those on the margin which differ from those of inframarginal individuals. The model shows theoretically that the marginal labor supply responses can grow, decline, or remain the same as participation expands and hence are ambiguous in sign. We provide a historical illustration, estimating marginal labor supply responses of single mothers in the 1980s AFDC program, the last program to assume the textbook form of a cash transfer program. The empirical results show a non-monotonic, U-shaped marginal response curve, with marginal labor supply responses small at low participation rates, growing in magnitude as the program expands, but then falling again as participation expands further. The pattern can be explained by changing movements between full-time work, part-time work, and non-work in response to program expansion. Marginal labor supply responses are quite modest, on average, consistent with the literature, but are larger in certain ranges of participation.

Keywords: Welfare, Labor Supply, Marginal Treatment Effects

JEL Codes: I3, J2, C21

*The authors would like to thank Marc Chan, Kai Liu, Shaiza Qayyum, Kyungmin Kang, and Sue Bahk for research assistance as well as the participants of a large number of conferences and departmental seminars and numerous specific individuals at those presentations, including formal discussant remarks by James Ziliak. Comments from two referees are also appreciated. Research support from the National Institutes of Health is gratefully acknowledged.

†Corresponding author: `moffitt@jhu.edu`.

Reforms in transfer programs can have effects on individual outcomes either because they alter who participates and who does not, with those who change participation having altered outcomes, or because they have effects on those already participating in a program, or both. Most of the literature on transfer program reforms does not attempt to separate these channels and often lacks the identifying variation to do so. However, some reforms affect only program participation, and those have effects only on those who were initially on the margin of participation. The effect of the program is measurable by the change in their outcomes after their participation changes due to the reform.

This paper contributes to the large literature on the effect of transfer programs on labor supply by providing a study of marginal effects for labor supply outcomes. We provide a theoretical model by extending the textbook static labor supply of a classic, negative-income-tax transfer program with a welfare guarantee and a tax rate, to include response heterogeneity that generates marginal labor supply responses. This model is new to the literature and has implications for empirical work as well because it shows that marginal effects can be identified by changes in the costs of program participation. The model provides a precise definition of such marginal effects and how they depend on preference heterogeneity, the budget constraint, and program parameters. The model also shows, theoretically, that the labor supply responses of those on the margin of participation and who are brought into the program by an expansion can be higher, lower, or the same as the responses of those already on the program and can also rise, fall, or remain the same as participation expands, leaving those effects ambiguous in sign. Resolving the sign of these effects is an empirical task and may depend on the specifics of the impacted program and population.

We provide an empirical illustration by estimating marginal labor supply responses to the historical AFDC program. AFDC was the last major program to have the structure of a cash program with a simple benefit system typically used in the classic model of labor supply responses. Using data from the 1980s and early 1990s and an instrumental variable that altered participation in the program in that period, we estimate the marginal labor supply response curve for the program, showing how marginal responses vary with the level of participation. The estimates show a U-shaped curve, with smaller labor supply responses when the fraction participating is low, but growing as participation expands, but then declining again when participation reaches higher levels. We show that this pattern can be explained by movements between full-time work, part-time work, and non-work. The average labor supply effect is quite modest, consistent with the literature, but can be larger in certain ranges of the participation rate and caseload. While the empirical illustration is of historical interest, more contemporary applications would be of greater interest. We conclude our analysis by suggesting several applications of our framework where marginal approaches could be applied.

The marginal labor supply response we estimate is called a marginal treatment effect (MTE) in the causal models literature. The MTE was introduced by Björklund and Moffitt (1987) and extended in a series of papers by James Heckman and coauthors, beginning with Heckman and Vytlacil (1999, 2005). Reviews of the method can be found in Cornelissen et al. (2016) and Mogstad and Torgovitsky (2024). There have been a modest number of empirical applications of the MTE framework, including applications to foster care and child removal (Doyle, 2007; Bald et al., 2019), the Social Security Disability Insurance program (Maestas et al., 2013), education (Carneiro et al., 2011), health insurance (Kowalski, 2016), early child care (Cornelissen et al., 2018), migration (Johnson and Taylor, 2019), incarceration (Bhuller et al., 2020), surgery (Tafti, 2022), misdemeanor prosecution (Agan et al., 2023), and electricity plan choice (Ito et al., 2023), among others. But no applications have been made to the labor supply effects of transfer programs, despite the large existing literature on the topic (for reviews, see Moffitt, 1992, 2002, 2003, 2014).

The paper proceeds as follows. Section 1 is the theoretical section, modifying the textbook model of the labor supply response to transfer programs to allow for marginal responses, and establishes the theoretical ambiguity of the sign of the MTE curve. Section 2 of the paper provides the empirical illustration to the historical AFDC program. A summary with suggestions for future research in Section 3 concludes the paper.

1 Modifying the Textbook Model to Allow Marginal Effects

In this section, we modify the textbook static model of the labor supply effects of a classic transfer program to allow marginal effects. This model assumes that an individual has a utility function $U(H_i, Y_i; \theta_i)$ where H_i denotes hours of work by individual i and Y_i denotes disposable income. We add the parameter θ_i to represent preference heterogeneity for leisure and income. An individual faces an hourly wage rate W_i and has available exogenous non-transfer non-labor income N_i. The welfare benefit formula is $B_i = G - tW_iH_i - rN_i$, where G is the guarantee and t and r are the tax rates on earnings and non-labor income, respectively. An individual not on the program has the budget constraint $Y_i = W_iH_i + N_i$ and utility maximization results in a labor supply function $H_i = H(W_i, N_i; \theta_i)$. An individual on the program has the budget constraint $Y_i = W_i(1-t)H_i + G + (1-r)N_i$ and labor supply function $H_i = H(W_i(1-t), G + (1-r)N_i; \theta_i)$. Figure 1 shows the textbook income-leisure diagram. An individual has the indifference curve O if off welfare and the indifference curve I if on welfare, with a labor supply response to participation in the program indicated by $\triangle_I$.

An additional feature that is typically added to this model is to allow for the possibility of incomplete take-up, i.e., that not all eligibles actually enroll in the program. Incomplete take-up has been studied in the literature for many years, with Currie (2006) conducting the first survey, and with a subsequent ongoing literature that includes recent work on administrative burden (Herd and Moynihan, 2018).[1] The literature has identified

[1]Ko and Moffitt (2022) have followed Currie (2006) with a new survey of the literature, including estimates

a number of reasons for partial take-up, including stigma from welfare participation, time and hassle costs of applying and participating—and administrative burden more generally—and lack of information about eligibility, among other factors. There is evidence in support of all these factors but assigning weight to each has been difficult, so here we will just assume there is some collection of costs to participation that capture these various factors. In our empirical application below, we will use data on one particular type of impediment to welfare take-up (a measure of administrative barriers to participation).

To write the model down, we adopt the form proposed by Moffitt (1983) and Chan and Moffitt (2018), with the utility function:

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

where P_i is a program participation indicator and ϕ_i is a scalar representing some type of participation cost, either from the perspective of the individual or imposed by the government agency. There are two choices—a choice of labor supply, H_i, and a choice of program participation, P_i—each with its own equation. We write them as follows:

$$H_i = H[W_i(1 - tP_i), N_i + P_i(G - rN_i); \theta_i] \tag{2}$$

$$P_i^* = V[W_i(1 - t), G + N_i(1 - r); \theta_i] - V[W_i, N_i; \theta_i] - \phi_i \tag{3}$$

$$P_i = 1(P_i^* \geq 0) \tag{4}$$

where $V[\cdot]$ is the indirect utility function and $1(\cdot)$ is the indicator function. Equation (3) is part of our extension to the textbook model and shows how an individual will participate if the direct utility gain from participation (i.e., the change in $V[\cdot]$) is greater than the cost of participation.[2]

of partial take-up of welfare programs around the world.

[2]By using the indirect utility function, the model assumes only intensive margin responses, but this is only for illustration. The empirical example below is reduced form in nature—regressing H on program participation—and captures the extensive and intensive margins.

Given the model, we can formally define the *marginal* labor supply response, which is the labor supply response to a change in program participation for those just on the margin of participation. These responses will consequently be those resulting from a small increase in participation that brings individuals into the program. First, the labor supply response to the program for individual i, holding fixed the budget constraint variables, is the change in hours worked from participating:

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

where $C_i = [W_i, N_i, G, t, r]$ is shortened notation for the set of budget constraint variables. This response differs across individuals because of heterogeneity in the taste parameter θ_i.[3] The marginal labor supply response is the value of $\triangle_i$ for individuals whose θ_i puts them on the margin of participation. But who is on the margin of participation also depends on ϕ_i. It is the set of joint values of these two variables that determines who is on that margin. The set of values that make participation indifferent are the values of θ_D and ϕ_D that satisfy the equation:

$$0 = V[W_i(1-t), G + N_i(1-r); \theta_D] - V[W_i, N_i; \theta_D] - \phi_D \tag{6}$$

and it is the set of values (θ_i, ϕ_i) that fall on one side of the (θ_D, ϕ_D) locus defined by this equation that generate participation. If we use this locus to define a function $\theta_D(\phi_D|C)$—preference values that put the individual on the margin of participation if their participation cost is ϕ_D—then the marginal labor supply response (again, holding constant the budget constraint variables C) is:

$$\triangle^{MTE}(C) = E_{\phi_D} \triangle [\theta_D(\phi_D|C)|C] \tag{7}$$

[3] We say θ_i represents "tastes" but it can represent anything that causes the labor supply response to differ—availability of child care, availability of jobs, and other factors not captured in the model.

Formally, the marginal response is the average labor response of those on the margin averaged across all values of participation costs. What we will call the "marginal response curve," or MTE curve, is how this response varies as participation changes as a result of variation in participation costs and is the derivative of the function with respect to the participation rate.

We end with a discussion of whether the labor supply responses of those on the margin who are brought into the program by an expansion of participation differ from the responses of those already on the program. In many applications, participation in a program or activity is considered to be a function of "gains" minus "costs" where the gains are some measurable outcome like earnings. This typically generates positive selection, meaning those brought into a program have smaller gains and smaller responses than those initially on the program.[4] But here, there is no implication that those newly brought into a program will have larger or smaller labor supply responses than those already on the program (i.e., selection can be positive, negative, or zero on labor supply). This is because the gains from participation are utility gains (i.e., gains in $V(\cdot)$) and not directly based on the size of labor supply responses. The direct gains from participation can be in the form of additional consumption rather than leisure, and how much utility is gained from each can vary arbitrarily across individuals.

Figure 1 illustrates this point. Individual I is already on the program, whose labor supply response is $\triangle_I$. But consider two other individuals, I_A and I_B, whose initial locations were the same as that as individual I (at O) but had greater participation costs and hence were off welfare. If their participation costs were to fall and they were brought into the program, the two individuals would benefit in different ways. I_A gains from participation relatively more in the form of additional consumption (Y), while I_B gains relatively more in the form of additional leisure. I_A has a smaller labor supply response than individual I and I_B has a larger one. How the average response of new participants

[4]The Roy Model is the most common form of this model. Björklund and Moffitt (1987) also framed the MTE in these terms.

differs from that of existing participants depends on whether there are more individuals like I_A or I_B and this can arbitrarily change as participation expands.

2 Empirical Illustration to Historical AFDC program

2.1 Regression Specification

The theoretical model in the previous section directly translates into a familiar model with outcomes as a function of participation. There is a first-stage equation for welfare participation which includes a variable affecting participation but not outcomes directly. Using an observable proxy for participation costs which satisfies the usual validity and relevance conditions for an instrumental variable leads to a standard IV estimation approach. The only needed modification in the regression equation for the outcome is one to allow the effect of participation on labor supply to vary with the level of participation (i.e., preference heterogeneity), rather than being a fixed, constant coefficient as usually estimated.

To make this modification, rewrite Equation (2) as

$$
\begin{aligned}
H_i &= m(W_i, N_i, X_i; \theta_i) + [m(W_i(1-t), N_i + G + (1-r)N_i, X_i; \theta_i) - m(W_i, N_i, X_i; \theta_i)]P_i \qquad (8) \\
&= m(W_i, N_i, X_i) + [\triangle(W_i, N_i, G, t, X_i; \theta_i)]P_i + \epsilon_i \qquad (9)
\end{aligned}
$$

where X_i is a vector of exogenous covariates. Labor supply is equal to the function $m(\cdot)$ if an individual is off welfare ($P_i = 0$). If an individual is on welfare ($P_i = 1$), labor supply is equal to $m(\cdot)$ plus $\triangle$ which denotes the difference in their labor supply when on and off welfare. If the unobservable in $m(\cdot)$ enters linearly, it can broken out and entered as a linear error term. This formulation is a random coefficients model, where the effect of participation on labor supply depends on an error term, θ_i, and differs across individuals.

This equation can be estimated in reduced form as follows. Let the first stage participation equation be represented as the propensity score,

$\Pr(P_i = 1|C_i, X_i, Z_i) = F(C_i, X_i, Z_i)$ where Z_i is an instrument and $F(\cdot)$ is a probability function ranging from 0 to 1. Then it can be shown that the reduced form of Equation (9) can be written as:[5]

$$H_i = m(C_i, X_i) + g[F(C_i, X_i, Z_i), C_i, X_i]F(C_i, X_i, Z_i) + \nu_i \tag{10}$$

Mean hours of work (H), conditional on the observables in the budget constraint and exogenous socioeconomic variables, equals mean labor supply off welfare plus the fraction of the population on welfare (F) times the mean labor supply response of those on, which is denoted by the function $g(\cdot)$. This function not only depends on observables C and X but also on the fraction on welfare (F). As a result, the effect of an increase in the fraction participating in welfare (F) on labor supply depends on this fraction because the composition of who is on the margin can change as that fraction changes. This feature allows a marginal response curve to be non-linear and non-constant. In a 2SLS estimation, a first-stage estimate of F, $\hat{F}$, can be used in the outcome equation in both places it appears. Some functional form for g can be adopted which allows it to be flexibly estimated as a function of F. The marginal labor supply response curve is the partial derivative of this estimated equation with respect to F, capturing both the direct effect on mean labor supply in the population stemming from a simple change in the fraction on welfare as well as the change in the response to that change because of a changing composition of those on the margin.

To implement the model empirically, we take the following linearized outcome equation to data for estimation:

$$H_i = X_i\beta + [X_i\lambda + g(F_i)]F_i + \nu_i \tag{11}$$

[5] This derivation involves the following identity: $E(H|C,X,Z) = m(C,X) + E[\triangle(C,X,\theta|C,X,Z,P = 1)\Pr(P = 1|C,X,Z) = m(C,X) + g(C,X,F)F$.

where, for notational simplicity, we redefine X to include both the budget constraint variables and exogenous characteristics. The term $X_i\lambda$ allows treatment effects to depend on other observables, but this is common in the literature, where subgroup effects are often estimated. We will estimate g with a flexible non-linear form described below allowing the marginal response curve to be any shape.

Data. The Aid to Families with Dependent Children (AFDC) program was an open-ended cash transfer program similar to the one featured in the theoretical model. Starting in 1993 this program was transformed with the introduction of work requirements, time limits, and other features that made it a different type of program than that illustrated above. These changes led to the program being renamed Temporary Assistance for Needy Families (TANF). We therefore provide an analysis using the pre-1993 version of the AFDC program, to illustrate the method and for historical interest. We will remark on the implications of our findings for the modern TANF program in Section 3 with our conclusions.

We use data from 1988–1992, 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 began in 1984 for which a set of rolling, short (12 to 48 month) panels are available throughout the 1980s and 1990s. The SIPP is commonly used for the study of transfer programs because respondents were interviewed three times a year and their hours of work, wage rates, and welfare participation were collected monthly within the year, making them more accurate than the annual retrospective time frames used in most household surveys. The SIPP questionnaire also provided detailed questions on the receipt of transfer programs, a significant focus of the survey reflected in its name. We use all waves of panels interviewed in the Spring of each year from 1988–1992 to avoid seasonal variation and pool them into one sample. We exclude 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 prior AFDC research. To concentrate on the AFDC-eligible population, we restrict the sample to those with 12 years or less of completed education, non-transfer non-labor income less than $1,000 per month, assets less than $1,500, and between the ages of 20 and 55. The resulting data set has 3,381 observations.

The means of the variables used in our sample are shown in Appendix Table A1. The variables include hours worked per week in the month prior to interview (H) (including zeroes), which averages 21 hours for the full sample but only 4 hours for single mothers on AFDC, almost all of whom (84 percent) were not working in this period of the program.[6] About 37 percent of the single mothers were on AFDC at any time in the prior month (P). Covariates for education, age, race, and family structure (several state characteristics are also used as conditioning variables) are also included. For the budget constraint, variables for the hourly wage rate (W), non-labor income (N), and the AFDC guarantee and tax rates (G, t, and r) are needed. To address the familiar problem of missing wages for non-workers, a traditional selection model is estimated. Appendix Table A2 reports estimates of this equation using OLS and a selection-bias adjustment. The OLS coefficient estimates are similar to the selection-adjusted estimates for most of the variables, but not all. We will use the OLS estimates for our main analysis and then estimate the model with the selection-bias adjusted estimates as a sensitivity test.[7]

For N, the weekly value of non-transfer non-labor income reported in the survey is used.[8] AFDC guarantees and tax rates by year, state, and family size are taken from estimates by Ziliak (2007), who used administrative caseload data to estimate "effective" guarantees and tax rates. The effective guarantees and tax rates in the AFDC program

[6]The empirical work will report some estimates separating the extensive and intensive margin of H.
[7]These results are similar to those presented in the text and are available upon request.
[8]Logarithmic specifications of monetary variables typically work better in labor supply regressions, so we enter $\log(N + 10)$ in light of the many values of zero in the data.

differ from the nominal rates because the benefit formula has numerous exclusions and deductions which generate regions of zero tax rates and others with positive values but below the nominal rates because of earnings-related deductions. A long literature has used estimated effective guarantees and tax rates by regression methods, which are more accurate approximations to the parameters actually faced by recipients.[9] The mean effective tax rate on earnings across years is approximately 0.41, considerably below the nominal rate of 1.0, and that on unearned income is approximately 0.30, also far below 1.0.[10] The analysis also controls for the guaranteed benefit in the Food Stamp program, which was available over this period to both participants and non-participants in the AFDC program. The Food Stamp guarantee is set at the national level and hence varies only by family size and year, and consequently has relatively little variation in our sample. Those benefits are assumed to be equivalent to cash, as most of the literature suggests.

Instruments. As shown in our model and discussed in the regression section, we require instruments Z_i that proxy costs of participation that affect participation but not labor supply directly. We use measures of what were called administrative barriers to participation in the 1980s literature on the AFDC program, which were error rates made by the states in the determination of eligibility. Each year, federal auditors visited each state, recalculated eligibility for a sample of applicants, and then computed error rates made by states in that determination. Students of the AFDC program in the 1970s and 1980s know that there is a sizable literature, appearing mostly in social work journals, discussing the non-random and intentional nature of these error rates (Handler and Hollingsworth, 1971; Piliavin et al., 1979; Brodkin and Lipsky, 1983; Lipsky, 1984; Lindsey et al., 1989; Kramer, 1990). More errors were made incorrectly denying eligibility than errors incorrectly approving eligibility. This literature showed that administrative barriers

[9] See the references in (Ziliak, 2007) for the long prior literature.

[10] Both G and t have substantial cross-sectional variation, with the 1988 G for a family of 3 ranging from \$100 per month to \$753 per month, and with the effective tax rate on earnings ranging from 0.12 to 0.66. The tax rate on unearned income also has a wide range, but it was invariably insignificant in the empirical analysis and hence is not represented in the estimates reported in the next section.

were politically driven at the gubernatorial and state legislature levels and were aimed at keeping caseloads in the program low. States were able to subjectively interpret the rules for what types of income to count, whether an able-bodied spouse or partner was present, which assets to count, and other factors affecting eligibility. Heavy paperwork requirements on applicants were imposed and states used failure to complete the paperwork properly as a reason for denying applications ("mechanisms to limit services...through imposing costs and inconvenience on clients" Lipsky, 1984, p.8).

We have collected those annual, state-specific error rates from 1980 to 1992 from published and unpublished sources. They varied widely across states. We use them as instruments for AFDC participation in our SIPP data in three different ways. First, we use cross-state variation in the error rates and show that the level of the error rate in a woman's state of residence in the SIPP data is negatively correlated with her probability of participating in the program, holding constant her characteristics as measurable in our data. There are obvious and well-known threats to the validity of any purely cross-sectional state-level government policy instrument, even after conditioning on state-level characteristics. States differ in many demographic and economic characteristics that are difficult to measure and which could be correlated with these error rates, either because both are correlated with some underlying labor supply-related state characteristic or because there might be direct reverse causality running from labor supply levels in a state to administrative barriers. These challenges motivate our use of instruments which use other sources of identifying variation in the administrative barriers from error rates. The second instrument is based on a differences-in-differences (DD) design. Although there was limited state-level policy variation over this period, making a general DD approach infeasible, there was one piece of federal legislation in 1989 that altered the federal monitoring process. We find that this law had differential effects on welfare participation across states and we therefore estimate our model using that variation as an instrument. However, we do not have an explanation for why the federal policy affected different states differently, making it difficult to assess the *a priori* validity of this source of variation. This motivates our third identification strategy. We draw on the literature noted above arguing that political differences across the states were responsible for the differences in error rates. We adopt a traditional close election regression discontinuity (RD) as an instrument, using narrowly elected governors combined with a legislature of the opposite party, which we find to have resulted both in increases in administrative barriers and reductions in AFDC participation.

All three instruments have arguable weaknesses in their *a priori* validity. However, we show that all three, each using a different source of variation, nevertheless yield a marginal response curve with a similar shape, which need not occur since the source of the variation used in each instrument is different. Taken together, this consistency across identification strategies increases confidence in the results of the estimated shape of the marginal response curve.

We should also note that who is screened in and out by this type of administrative barrier to participation need not be the same as those who are screened in and out for other reasons (e.g., stigma, as one example). Our results should strictly be interpreted as showing the MTE effects for our particular type of instrument. In our Conclusions below, we recommend work on the MTE effects of other reasons for non-takeup of welfare programs.

2.2 Cross-State Variation

Our instruments use information on seven measures of state AFDC error rates: the percent of eligibility denials that were made in error, the error rate from improperly denying requests for hearings and appeals, the percent of cases dismissed for eligibility reasons other than the grant amount, the overall percent of applications denied, the percent of applications denied for procedural reasons (usually interpreted as not complying with paperwork), the percent of cases resulting in an incorrect overpayment or underpayment, and the percent of cases resulting in an underpayment. There are also error rates and percents of actions related to income, assets, or employment, but these are directly or indirectly related to the applicant's labor supply and earnings level and hence are not used.

The means and distributional statistics of the seven administrative barrier variables are shown in Table 1.[11] While the means of one of the variables is less than 1 percent, others range from 2 percent to 24 percent. The cross-state variation is also wide, with some states making underpayment errors in over 10 percent of cases, procedural denial rates of almost 35 percent, and overall denial rates of almost 50 percent.

Initial analyses of the seven barrier variables revealed them to be highly correlated, with correlations generally in the range of 0.80. This feature makes it difficult to identify their separate effects and signals that they likely represent packages of similar behaviors by states. We consequently interpret the seven as noisy indicators of a single underlying index and construct the textbook inverse variance weighted average of the seven, which is the lowest variance estimate of a true single variable in the presence of measures with independent mean-zero measurement errors. The summary statistics of this barrier index are reported in the last row of Table 1.[12]

To generate first-stage estimates of the AFDC participation propensity score, we match the state of residence of each observation in our SIPP data to the state administrative barrier index and, for our first, cross-state instrument, estimate probit models for the probability of AFDC participation as a function of the index and other control variables. These controls include the four budget constraint variables which must be included for consistency with the theoretical model. Table 2 reports the estimated coefficients on the barrier index.[13] The first specification features the barrier index alone, while the second specification includes an interaction with non-labor income, which we

[11] The administrative variables vary from year to year for each state because the federal government only took a random sample of records each year. To reduce noise, we compute the average of each barrier for each state over the 1980–1992 period.

[12] The logs of the barrier variables performed better in our analysis than the absolute values. We report the inverse variance weighted mean of the logarithms of the seven barrier variables in the last row of the table.

[13] Appendix Table A3 reports estimates for all coefficients.

found to be highly predictive of AFDC participation.[14] A higher level of administrative barrier in a woman's state of residence reduces the likelihood that she is on AFDC, and the effect is larger for women with higher non-labor income.

The F-statistics reported in the bottom of Table 2 provide information on the strength of the instrument. The uninteracted specification has an OLS-estimated F statistic of over 11, meeting the conventional Stock and Yogo (2005) rule of 10 aimed at keeping bias and coverage at reasonable levels, while the specification with the interaction has a lower OLS F-statistic.[15] However, these statistics do not indicate instrument strength in different regions of the propensity score curve, which is likely to be nonlinear (e.g., of the S-shape typical of cdf's) but are instead indicators of strength of the instrument for some weighted average over all regions. Theoretically, what is needed for a marginal response model with a continuous instrument for a nonlinear propensity score curve is an indicator of instrument strength at each point on the curve (i.e., completely "local"), which would then indicate where in the curve the instrument might be weak and where it might be strong. Such a statistic has not been developed in the weak IV literature, which has focused on models with a single, constant treatment effect (as have Anderson-Rubin and related alternative models). Instead, as an approximation to local strength, we estimate what we term "pseudo-F statistics" for different discrete ranges of the propensity score, first for quartiles and then for terciles of the score distribution.[16] As the results at the bottom of Table 2 show, the instruments are stronger in the central range of the propensity score distribution and very weak in the upper and lower ranges. This pattern is what one should expect for a standard S-shaped cdf curve, where the slope is greatest in the middle and flattest in the tails. We recommend future research on methods for testing for weak instruments in continuous S-shape propensity score curves, but for present purposes we will

[14] For this first-stage equation, a linear specification of N yielded higher F-statistics than $\log(N+10)$, but the latter fits better in the second-stage labor supply equation (see below).

[15] The Stock-Yogo F-statistics for two instruments range from 12 to 20.

[16] The calculation of these statistics is described in the footnote to the table. We divide the observations into ranges based on the propensity score (i.e., the predicted AFDC participation rate), and then compute conventional F-statistics for our instrument within each range of that predicted participation rate.

simply restrict our estimates of the marginal response curve to the approximate region 0.25 to 0.66 (the union of the second quartile and the middle tercile) where the instruments are the strongest. The F statistics for that range are 10 and 17 for the two specifications.[17]

Results. Estimating equation (11) using the fitted values of the participation probabilities for F yields estimates of β, λ, and the parameters of the g function.[18] The g function is estimated with cubic splines, one of the most commonly used flexible forms for estimating the shape of a curve (Hastie et al., 2009). We fit the function $g(F) = g_0 + \sum_{j=1}^{J} g_j \max(0, F - \pi_j)^3$, where the π_j are one of J preset spline knots. For a given J, we use knots chosen to be regularly spaced within the (0.25, 0.66) range where we will focus our attention. We start the estimation with $J = 3$ and then increase the number until a fit measure is optimized. Fit is assessed with a generalized cross-validation statistic (GCV). 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 feature; the exact spline functions for a five-knot spline are shown in Appendix B.[19]

To illustrate our choice of the number of knots, Figure 2 shows the estimated marginal response (MTE) curves in the (0.25, 0.66) propensity score range with 90 percent confidence intervals for three-to-six knots.[20] The 3-knot and 4-knot specifications show

[17] Much of the variation in the propensity score between 0.25 and 0.66 is driven by the covariates. The instruments themselves move the propensity score by less than this—by a maximum of -0.43 but 90 percent of the movements are between -0.12 and +0.09 relative to the mean.

[18] We use the second specification in Table 2 to generate fitted participation probabilities but the results are very similar for the first specification.

[19] Although we use polynomials in the propensity score, this approach does not create "forbidden regression" problems. Our expression for H in Equation (10) is derived from a theory consistent model and the conditional mean of H depends on the variables C_i, X_i, Z_i, and the unobservable ν_i which is mean zero by construction. The H equation is also parameterized to depend on the propensity score $F(C_i, X_i, Z_i; \vartheta)$, which is also a function of these covariates as well as a set of estimated parameters ϑ. Given a consistent estimate for ϑ, one can generate a consistent estimate for $\hat{F}$ and use it to estimate the H equation using standard methods.

[20] The marginal response function is, as noted previously, the derivative of the hours equation with respect to the propensity score. Confidence intervals are constructed by jointly bootstrapping the estimating

monotonic curves but they turn nonmonotonic with 5-knots and stay nonmonotonic at 6-knots. The minimal GCV for all four is at 5-knots, although the specification for 6-knots is only slightly higher. We use the specification with 5-knots for the rest of the analysis.

Table 3 reports the full set of parameter estimates for three versions of the hours equation for the 5-knot specification. The natural spline coefficients are not easily interpretable and instead are shown graphically in MTE form in Figure 2. Column (1) has only the budget constraint variables in the λ vector, which are generally low in statistical significance, implying that we do not detect strong interactions of participation with those variables. The wage itself does have strong positive effects on hours, however, as indicated by its β coefficient. Column (2) tests a set of additional interactions of the participation probability with the budget constraint variables, but none are significant. Column (3) adds Age and Black to the λ vector, which are partially significant and improved the GCV measure. In unreported results, we tested additional X variables in the λ vector but these were either insignificant or had no impact on the GCV metric. The spline coefficients in column (3) are those used in Figure 2. The estimates in columns (1) and (2) yield similar curves.

We return to Figure 2 for substantive interpretation. The marginal labor supply responses are nonmonotonic and U-shaped, starting off at $F = 0.25$ at a negative value but whose confidence interval is slightly excludes zero. The response then grows in (negative) size as participation increases with confidence intervals excluding zero. The marginal response peaks at a participation probability near 0.36, when it reaches a labor supply disincentive of those on the margin of approximately -31 hours per week. It then declines, becoming insignificantly different from 0 at approximately $F = 0.49$. The point estimate approaches zero as participation rises further but remains insignificantly different from 0 for all higher participation levels. Thus the marginal labor supply disincentive of policies which increased participation in the AFDC program in the late 1980s and early 1990s was

equations using weights randomly drawn at the state-level to allow for state-specific clustering. All curves are evaluated at the means of the other variables in the equation.

zero at many margins but non-trivially negative at other margins, depending on the participation rate and caseload level at the reference point.

Some economic interpretation behind the U-shaped pattern of responses can be gained by examining marginal responses between full-time work, part-time work, and non-work. Figure 3 shows the results of estimating the hours worked equation by successively replacing the dependent variable for H with binary variables for full-time work, part-time work, and non-work. The top left panel shows that, in the range of participation rates from about 0.25 to 0.46, increases in program participation (i.e., through reductions in administrative barriers) draw in women who are reducing their probability of working full-time. There is an insignificant change in the fraction working part-time, though the point estimate is positive, implying that some women on the margin who join the program may be moving from full-time to part-time work. But a much stronger positive effect in this participation-rate range is apparent in the non-work figure, implying that a significant number of marginal participants who are full-time workers move to non-work (although these results cannot distinguish that from exits from part-time work as well). However, as participation rates rise further, the full-time effect falls and eventually becomes statistically insignificantly different from zero, the part-time effect remains essentially zero, and the non-work effect also falls to insignificance. This implies that those who enter welfare at those higher participation rate margins are not reducing their hours at all, but staying in the same hours worked range as they were in off welfare. Since about 84 percent of single mothers on AFDC are not working, on average (Appendix Table A1), this implies that further increases in program participation must come primarily from women who were not working even off the program.[21]

Additional evidence on this interpretation is given in Table 4, which displays labor-supply related variables by quintile of the fitted propensity score distribution within

[21] Similarly, the figures imply that an increase in administrative barriers which make high participation rates fall, initially have no effect on labor supply but, as barriers rise further, some of those not working leave the program to work full-time, and possibly some part-timers leave to work full time as well. But eventually all those who leave the program have the same hours worked off the program as on.

the (0.25, 0.66) range. Those who are on the margin at low participation probabilities in that range have higher wage rates, are less likely to be black, are older, and have fewer young children, all of which are correlated with higher levels of work. Non-labor income is higher for these "early" marginal participants as well, which is typically correlated with lower levels of labor supply. But, for discrete moves from full time work to non-work, this means that those individuals also have a larger income cushion if they do not work. Those who are on the margin at higher, or "late," participation rates have lower wages, are more likely to be black, are younger, and have more children, all of which are correlated with lower levels of work and hence lower marginal effects of labor supply upon participation.[22]

2.3 Difference in Difference Approach

As noted previously, there were no significant legislative changes at the state or federal level regarding state error rates or federal monitoring of those rates over most of our observation period. However, an exception occurred in 1989, when Congress passed new legislation, the Omnibus Budget Reconciliation Act, which modified the quality control inspection program that the federal government used to assess state error rates (U.S. House of Representatives, Committee on Ways and Means, 1994, Section 10). The legislation was motivated by a concern that states were continuing to make errors in their program eligibility assessments, and tightened up the federal monitoring system imposed on the states. The full implementation of the Act started in 1991 and was completed in 1992. We use this legislation in a difference-in-difference (DD) exercise which examines whether error rates in the states changed significantly in the 1991–1992 period compared to previous levels, and whether it did so differentially across states. We then use that cross-state differential change in error rates as the instrument for estimating our marginal response curve. The disadvantage of this method is that the legislation was national in scope and there is no available evidence for why error rates changed differently across states. The

[22] See Chernozhukov et al. (2019) for a related method of using observables at different percentile points to assess heterogeneity of treatment effects.

advantage of this method is that it uses within-state variation in error rates over time rather than the cross-state variation used in the last section, and these are different sources of variation which need not have any relationship to each other.

We implement this method by computing the mean barrier index for each state over the 1988–1990 period and then computing the residual of the actual 1991 and 1992 barrier indices for each at that mean. The residuals have a wide range across the states, with a standard deviation almost equal to its mean. Table 5 shows the first-stage estimates for a standard DD specification, including variables for the state mean barrier index, a binary indicator for the post 1991–1992 period, and an interaction term for the post variable and the state residual barrier index, with the latter constituting the instrument.[23] The estimated coefficient is negative in sign, indicating that those states with above-average residuals had larger declines (or smaller increases) in AFDC participation in the 1991–1992 period, and states with below-average residuals had smaller declines (or larger increases) in participation. Experiments with pseudo-F statistics in different ranges of the propensity score again showed that the range from 0.25 to 0.66 had the largest statistics, which are slightly above 8 for this instrument and thus are on the weak IV borderline.[24]

The estimated response curve from the hours worked equation using this instrument in the hours equation is shown in Figure 4, using a 5-knot natural spline and the specification in column (3) of Table 3. The shape of the curve is remarkably similar to that using the cross-state instrument: U-shaped with confidence intervals bounded away from zero in the 0.26 to 0.63 range, and with a peak (negative) work disincentive of -39 hours per week at approximately a 0.37 participation rate, which is slightly larger than the peak negative for the cross-state instrument. Despite the very different source of variation used with this instrument, the substantive result is the same as for the first instrument, that

[23] Estimates for all parameters are available in Appendix Table A4.

[24] We note that Angrist and Kolesar (2024) have shown that, in just identified models, a better measure of weak IV than the Stock-Yogo F statistic is the correlation coefficient between the endogenous variable and the error term and that little bias in standard errors occurs if that coefficient is less than 0.50. We examined this alternative criterion and found very low coefficients, far below 0.50. These results are available upon request.

marginal work disincentives are small or insignificantly different from zero at many margins of participation but substantial at other margins.

2.4 Close Election RD Approach

For our third instrument, we note that our initial discussion of the literature on state error rates in the 1980s and early 1990s argued that those error rates were a result of political differences across states. But, as is widely recognized, political differences themselves may not be valid instruments because they are likely correlated with state demographics and therefore possibly with the labor market engagement levels of low income families in the state. We draw on the literature on regression discontinuity designs in political economy research using close elections as a plausibly exogenous source of political party governance (see e.g., Lee et al., 2004; Lee, 2008, and the large subsequent literature). The argument in this approach is that states where a party is elected only narrowly are close in unobserved ways to states where parties lose narrowly, and therefore a comparison of the impact of which party is elected in a close election has a better chance of being exogenous than merely political party control itself, which could easily be correlated with state demographics and labor market variables.

We supplement our data set with state-level political variables we collected for the time period from 1988–1992. First, we gathered data on the party affiliation of the governor of each state, and we determine whether that governor was a Democrat or Republican elected in a close election, which we define as having been elected with at most 60% of the vote for Democrats and 55% for Republicans.[25] We control for the governor's share of the vote as the running variable. We also gather information on the political makeup of the state legislature, which we hypothesize could affect the ability of narrowly elected governors to enact policies of their liking and, in particular, to enact policies concerning error rates in their state's welfare programs. We collect data on whether the

[25]For sample size reasons we were unable to go below 60% for Democrats.

legislature is entirely Republican or Democratic (both chambers) or whether it is split, with one chamber controlled by Democrats and one controlled by Republicans (a "split" legislature).[26] We test whether the impact of a governor who has been elected in a close election varies with these legislative party control variables. We implement this analysis separately for closely elected Democratic and Republican governors.

The first two columns of Table 6 show the results of the relevant first stage estimation for narrowly-elected Democratic governors.[27] The first column tests whether a narrowly elected Democratic governor affects the level of the barrier index in the state and the second column tests whether that affects AFDC program participation. While a narrowly-elected Democratic governor results in a reduction in the administrative barrier (coefficient = -0.066) it does not significantly affect AFDC participation. But when interacted with whether the legislature is fully controlled by Republicans, a narrowly elected Democrat governor results in positive effects on barriers and negative effects on AFDC participation. We suspect that in states where the legislature was fully controlled by Republicans and a Democratic governor was elected only narrowly—hence was weak politically—the legislature was able to enact legislation of their liking over the veto threat of the governor. We use the interacted variable as the instrument because of its *a priori* plausibility. The F-statistic in the (0.25, 0.66) region of the propensity score is slightly above 9, marginally greater than the one from the DD analysis presented in the prior section. The instrument is again weak in higher and lower ranges of the propensity score.[28]

The final two columns show the results for an analogous analysis using narrowly-elected Republican governors, alone and interacted with whether the legislature was fully controlled by Democrats. The coefficient on a narrowly-elected Republican governor is positive but not significant and the effect of being combined with a Democrat legislature is further positive but also insignificant on the level of the barrier index.

[26]Nebraska has a unicameral state legislature, so splits are only possible if that chamber is equally divided.
[27]Appendix Table A5 contains coefficient estimates for all variables in the equations.
[28]See the prior footnote and the alternative approach of Angrist and Kolesar (2024). We again found very low correlation coefficients. Results available upon request.

However, narrowly elected Republican governors combined with Democrat legislatures significantly lowers the rate of AFDC participation. The F-statistics in the (0.25, 0.66) is also low, however we note that in the (0.33, 0.66) region the statistic is approximately the same for the Democratic governor specification. The coefficients do not align with conventional assumptions about the preferences of Republicans vs Democrats, and therefore we do not have a good interpretation of the mechanisms underlying them. Despite these concerns, we proceed with MTE estimation to ascertain the implied shapes of those curves.

Figure 5 shows the estimated marginal response curves using these close election variables as an instrument with a 5-knot natural spline and the hours equation specification in column (3) of Table 3, again only for the (0.25, 0.66) range of the propensity score. The curve has the same shape as obtained with the prior two instruments: U-shaped with increasingly negative marginal work incentives as the participation rate rises above 0.25 but peaking at a participation rate of 0.37 where the marginal disincentive is -35 hours per week, and then falling in absolute value as the caseload expands. The confidence interval includes zero at a participation rate of 0.50 or greater.

This third, close election instrument is the weakest of our three instruments, both in first-stage strength and in interpretation of the mechanisms at work. Consequently, the results deserve the smallest weight of the three. Nevertheless, again, this source of variation in the instrument, despite using a very different source than the first two, yields the same U-shape MTE curve as the first two.

3 Summary and Conclusions

This paper has modified the textbook model of the labor supply response to a transfer program to incorporate marginal responses and has shown that that response can be greater, smaller, or the same as responses of inframarginal individuals, and that marginal responses can grow, decline, or remain the same as participation in a program expands, leaving it to empirical work in specific applications to determine the pattern. We have provided an application to the historical AFDC program, the last program in the U.S. to take the classic, open-ended cash negative-income-tax form. Using three different sources of variation in administrative barriers to participation, we estimate a U-shaped marginal response curve, with responses becoming more negative as the program expands from low participation rates to modest participation rates, then becoming less negative as the program expands further, and with the pattern similar for all three instruments.

The approach outline here should be applicable to many more modern policies than AFDC, and to other sources of variation in takeup. In general, any policy that alters eligibility for a program through some measure that does not affect benefits for current participants should only affect program responses (labor supply or others) of those who enter or exit the program, with no impact on those already on the program or who stay on it. The modern TANF program, for example, has many eligibility requirements that do not affect benefits or the behavior of existing recipients (e.g., two-parent eligibility rules). In fact, although the historic AFDC program studied here is dramatically different than the current TANF program, our finding that marginal labor supply responses are near-zero would, if applied to TANF with its currently very ow participation rates, imply that expanding eligibility would have little work disincentives. Our framework could also be applied to other policies that expand eligibility by only increasing the income eligibility level—such as the Medicaid expansion in the Affordable Care Act and Broad-Based Categorical eligibility in the Supplemental Nutrition Assistance Program—as these reforms brought new participants onto these programs and only influenced their behavior. These are both programs of substantive policy interest. Policies which increase or decrease administrative burden and result in changes in participation rates and have no effect on those remaining on the program would also be applicable. Also, the types of marginal individuals who join or exit a program may differ depending on the forces and factors which induce those participation decisions, and it may be that variation in administrative barriers (our application) may not induce the same MTE results as decisions resulting from other forces.

Application of the methods outlined here do, however, require, to be interesting, a multi-valued instrument and not a binary instrument. And, just as for any IV application, the broader the range of participation rates induced by the instrument, the more the results can be generalized to other policies.

References

Agan, A., Doleac, J. L., and Harvey, A. (2023). Misdemeanor prosecution. *Quarterly Journal of Economics*, 138(3):1453–1505.

Angrist, J. D. and Kolesar, M. (2024). One Instrument to Rule Them All: The Bias and Coverage of Just-ID IV. *Journal of Econometrics*, 240(2):1–17.

Bald, A., Chyn, E., Hastings, J. S., and Machelett, M. (2019). The Causal Impact of Removing Children from Abusive and Neglectful Homes. Working Paper 25419, National Bureau of Economic Research.

Bhuller, M., Dahl, G. B., Loken, K. V., and Mogstad, M. (2020). Incarceration, Recidivism, and Employment. *Journal of Political Economy*, 128(4):1269–1324.

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

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

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–81.

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

Chernozhukov, V., Fernandez-Val, I., and Luo, Y. (2019). The Sorted Effects Method: Discovering Heterogeneous Effects Beyond Their Averages. *Econometrica*, 86:1911–38.

Cornelissen, T., Dustmann, C., Raute, A., and Schoenberg, U. (2016). From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions. *Labour Economics*, 41:47–60.

Cornelissen, T., Dustmann, C., Raute, A., and Schoenberg, U. (2018). Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. *Journal of Political Economy*, 126(6):2356–2409.

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

Doyle, J. J. (2007). Child Protection and Child Outcomes: Measuring the Effects of Foster Care. *American Economic Review*, 97(5):1583–1610.

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

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

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

Herd, P. and Moynihan, D. P. (2018). *Administrative Burden: Policy-Making by Other Means*. Russell Sage Foundation, New York, NY.

Ito, K., Ida, T., and Tanaka, M. (2023). Selection on Welfare Gains: Experimental Evidence from Electricity Plan Choice. *American Economic Review*, 113(11):2937–73.

Johnson, J. E. and Taylor, E. J. (2019). The Long Run Health Consequences of Rural-Urban Migration. *Quantitative Economics*, 10:565–606.

Ko, W. and Moffitt, R. A. (2022). Take-up of social benefits. In Zimmermann, K. F., editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–42. Springer International Publishing.

Kowalski, A. E. (2016). Doing More When You Are Running LATE: Applying Marginal Treatment Effect Methods to Examine Treatment Effect Heterogeneity in Experiments. Working Paper 22363, National Bureau of Economic Research.

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

Lee, D. S. (2008). Randomized Experiments from Non-random Selection in U.S. House Elections. *Journal of Econometrics*, 142(2):675–697.

Lee, D. S., Moretti, E., and Butler, M. J. (2004). Do Voters Affect or Elect Policies? Evidence from the U.S. House. *Quarterly Journal of Econometrics*, 119(3):807–859.

Lindsey, E. W., Colosetti, S., Roach, B., and Wodarski, J. S. (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.

Maestas, N., Mullin, K. J., and Strand, A. (2013). Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. *American Economic Review*, 103(5):1797–1829.

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

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

Moffitt, R. A. (2002). Welfare Programs and Labor Supply. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, volume 4, pages 291–363. University of Chicago Press.

Moffitt, R. A., editor (2003). *Means-Tested Transfer Programs in the United States.* University of Chicago Press.

Moffitt, R. A., editor (2014). *Economics of Means-Tested Transfer Programs in the United States*. University of Chicago Press.

Mogstad, M. and Torgovitsky, A. (2024). Instrumental Variables with Unobserved Heterogeneity in Treatment Effects. In Dustmann, C. and Lemieux, T., editors, *Handbook of Labor Economics*, volume 5. Amsterdam North-Holland.

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

Stock, J. H. and Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In Andrews, D. W. and Stock, J. H., editors, *Identification and Estimation for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge University Press, Cambridge, UK.

Tafti, E. A. (2022). Technology, Skills and Performance: The Case of Robots in Surgery. Mimeo.

U.S. House of Representatives, Committee on Ways and Means (1994). Overview of Entitlement Programs: 1994 Green Book. Washington, DC.

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.

Tables and Figures

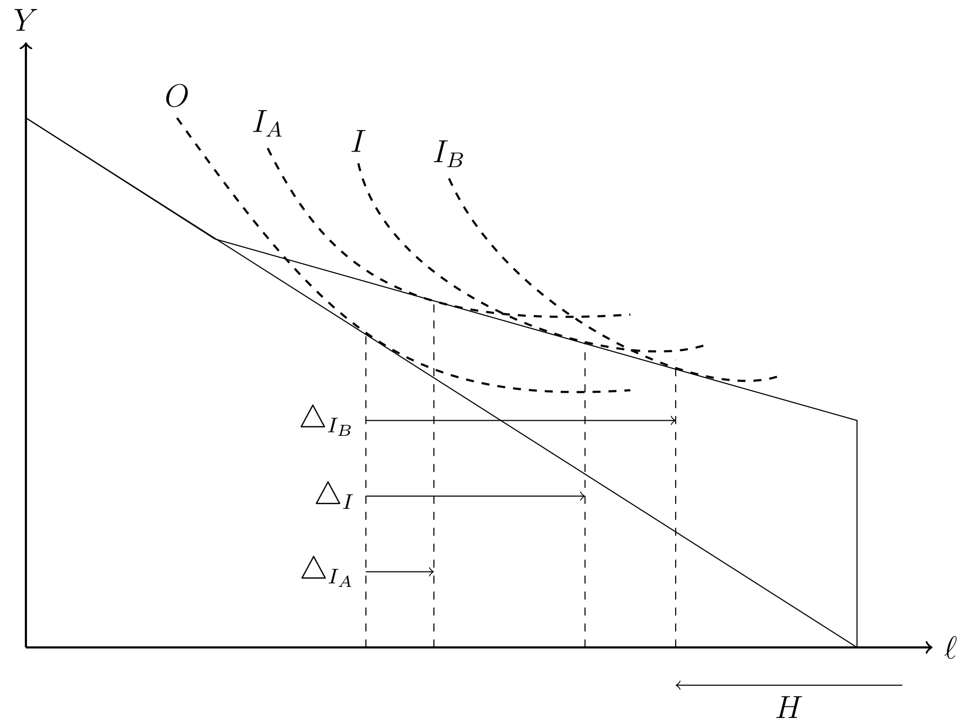

Figure 1: Textbook income-leisure diagram

Notes: This figure presents the textbook representation of income (Y) as a function of leisure (ℓ). Each line denotes the budget constraint. The straight line denotes the budget constraint of an individual not on the transfer program. The line with the kink denotes the budget constraint of an individual on the program. The slope of kinked portion denotes the tax rate on benefits as an individual works more and earns more labor-income. It shows the labor supply responses of three individuals I, I_A, and I_B as they shift from the same off of welfare position (O) to taking up benefits. The labor supply responses are denoted by $\triangle$.

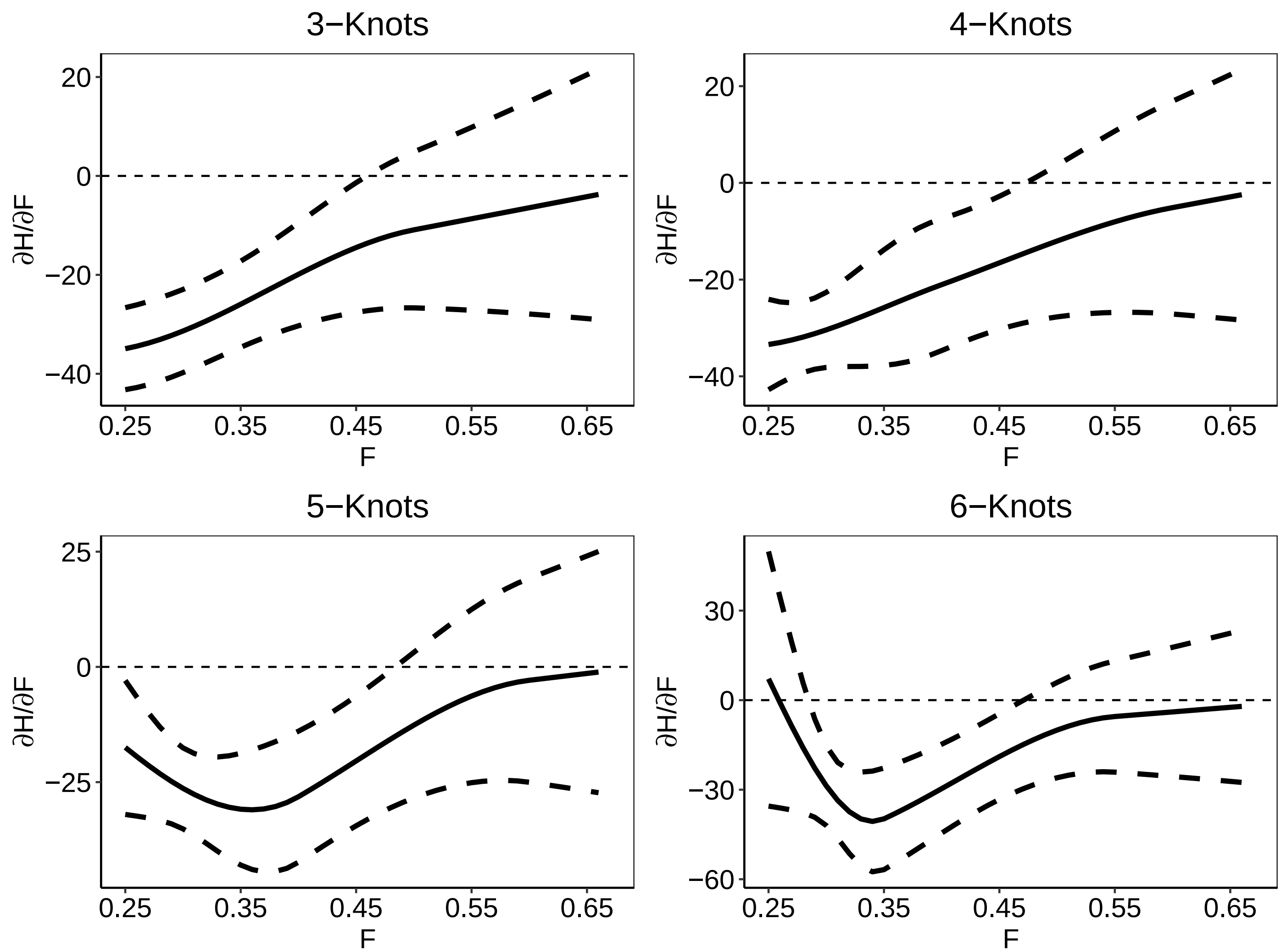

Figure 2: Marginal Labor Supply Curves for Different Natural Cubic Splines

Notes: This figure plots the marginal treatment effect curves using different cubic spline specifications. All specifications use a first stage probit model with the inverse variance weighted log of the AFDC administrative barriers and interactions with N. The dashed lines denote 90 percent confidence intervals that are generated from a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level.

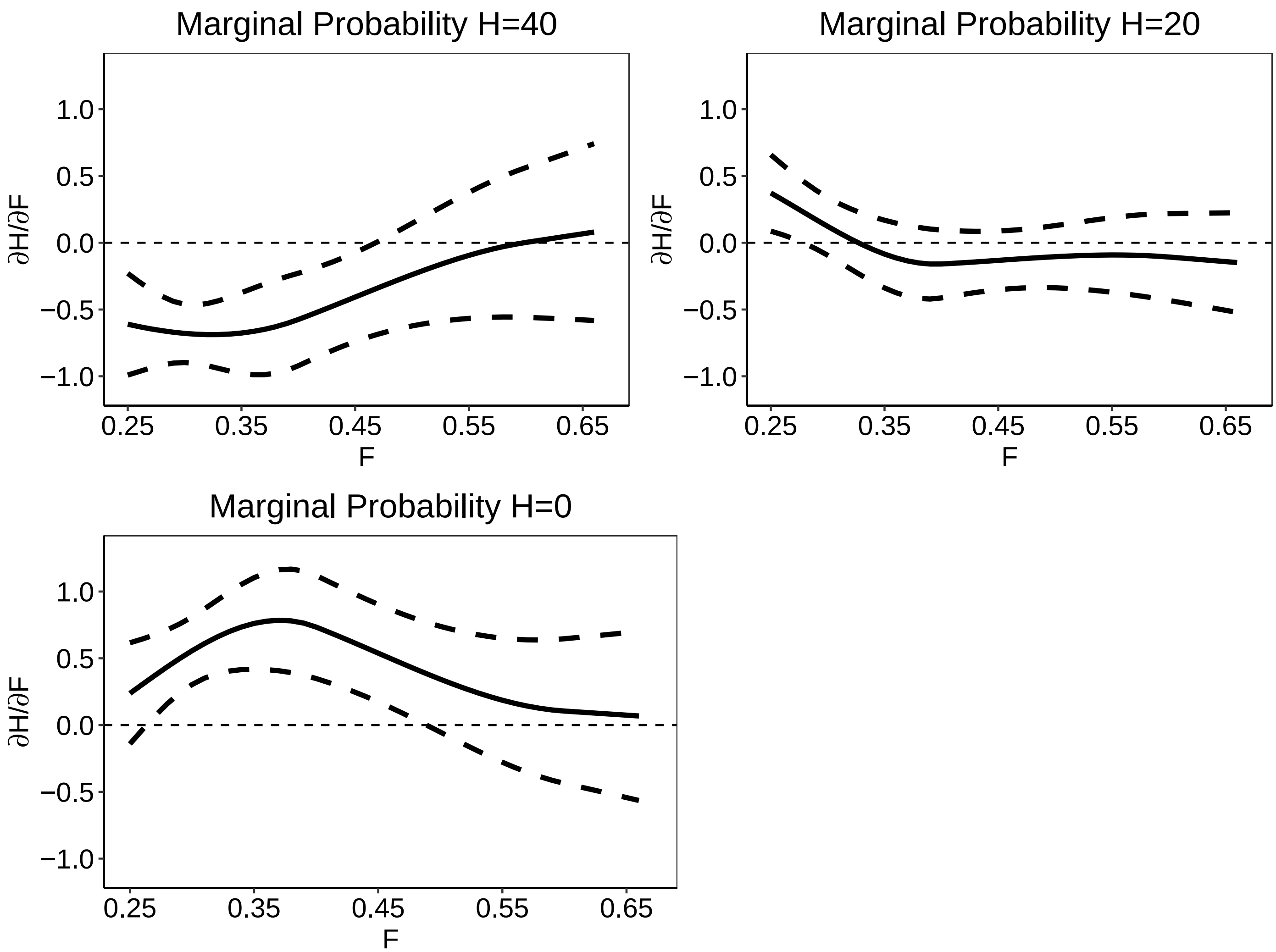

Figure 3: Marginal Labor Supply Curves for Different Types of Workers

Notes: This figure plots the marginal treatment effect curves for non-workers, part-time, and full-time workers using a 5-knot cubic spline specification. The first stage probit model uses the inverse variance weighted log of the AFDC administrative barriers index and interactions with N. The dashed lines denote 90 percent confidence intervals that are generated from a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level.

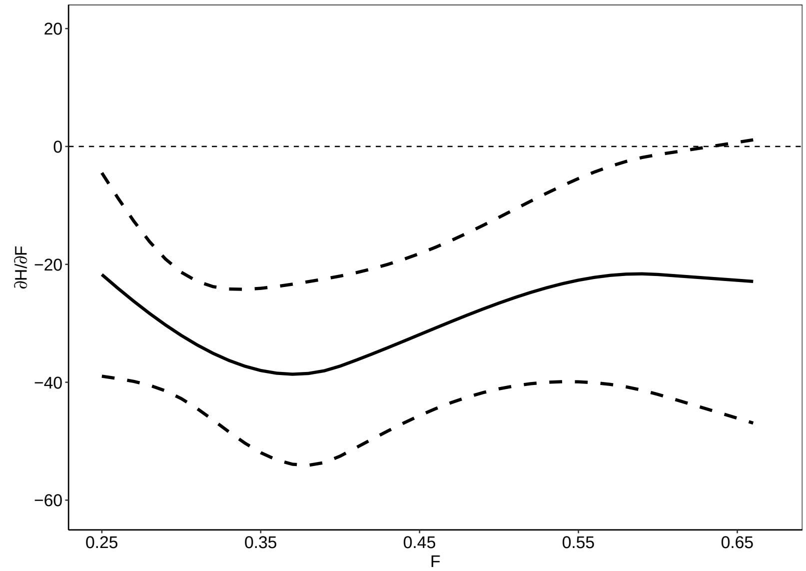

Figure 4: Marginal Labor Supply Curves Using 1989 Law Change Instrument

Notes: This figure plots the marginal treatment effect curves using a 5-knot cubic spline specification. The first stage probit uses the 1989 law change as the instrument. The dashed lines denote 90 percent confidence intervals that are generated from a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level.

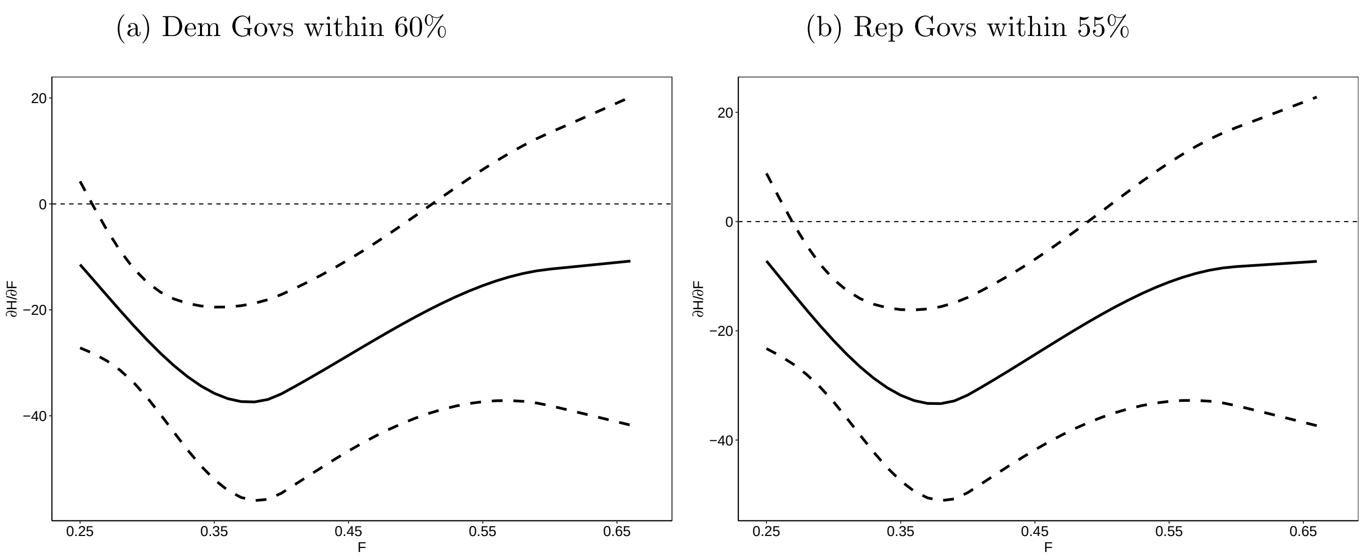

Figure 5: Marginal Labor Supply Curves Using Close Election RD

Notes: This figure plots the marginal treatment effect curves using a 5-knot cubic spline specification. The first stage probit uses the close election regression discontinuity design. The dashed lines denote 90 percent confidence intervals that are generated from a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level.

Table 1: Administrative Barrier Variables

	Mean	Std. Dev.	Min	Max
Individual Barrier Variables				
Pct. ineligible in error	1.7	0.9	0.3	4.7
Pct. hearings and appeals improperly denied	1.8	1.2	0.4	5.8
Pct. cases elig. denied for non-grant reasons	0.2	0.1	0.0	0.4
Pct. applications denied	26.2	9.6	5.3	47.8
Pct. applications denied for procedural reasons	13.8	7.8	1.3	34.6
Error rate in payment determination	4.2	1.3	2.2	7.3
Error rate resulting in underpayment	2.8	1.5	1.6	10.2
Weighted Average Barrier Index				
Inverse Variance Weighted Average	0.3	0.1	0.1	0.5
Inverse Variance Log Weighted Average	1.3	0.2	0.8	1.7

Notes: This table summarizes different administrative barriers for enrollment into the AFDC program from 1980–1992. These measures are obtained from Quarterly Public Assistance Statistics and unpublished data from the U.S. Department of Health and Human Services. Values are averages over all years for each state. "Inverse Variance Weighted Average" is the inverse variance weighted average of the individual barrier variables in levels. "Inverse Variance Log Weighted Average" is the inverse variance weighted average of the individual barrier variables in logs.

Table 2: Estimated Impact of Instruments on AFDC Participation

	(1)	(2)
Barrier Index	-0.593***	-0.482**
	(0.208)	(0.213)
Barrier Index×N		-0.007***
		(0.002)
Budget Constraint	✓	✓
Demographic Controls	✓	✓
State Controls	✓	✓
Region FEs	✓	✓
OLS F-Stat for Instruments	11.13	6.85
Pseudo F-Statistic by Part. Prob. Range		
0.00–0.25	3.16	1.94
0.25–0.50	10.96	15.94
0.50–0.75	0.38	3.13
0.75–1.00	0.91	0.88
0.00–0.33	4.24	4.94
0.33–0.66	9.09	14.62
0.66–1.00	2.07	2.34
0.25–0.66	10.18	17.62
Observations	3,381	3,381

Notes: $***p < 0.01$; $**\,p < 0.05$; $*p < 0.1$. This table reports estimates of a probit model for AFDC participation onto a series of individual and state characteristics. The "Barrier Index" is the inverse variance weighted average of the log of the individual administrative barrier variables in Table 1. "Budget Constraint" variables include $\log \hat{W}$, $\log N + 10$, $\log G$, and $\log \hat{W}(1-t)$. "Demographic Controls" include age, black, family size, the number of children under 6, and the food stamp guarantee. "State Controls" include the unemployment rate, share of the state that is urban, share of the state population that is black, and the per-capita income in the state. Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level. All parameter estimates are available in Appendix Table A3. The second panel reports F-statistics from an OLS version of the probit model and within different participation probability ranges based on the probit estimates. To calculate the F-statistic within a specific range of $\hat{F}$, define $RSS(q)$ as the residual sum of squares, equal to the sum of $[P - \hat{F}]^2$ taken over all observations in the range. The F-statistic is calculated as (1) the difference in $RSS(q)$ for the restricted model excluding the instruments and the unrestricted model $RSS(q)$ including the instruments divided by the degrees of freedom, divided by (2) the residual variance computed over all observations in the sample, using $\hat{F}$ from the restricted model.

Table 3: Estimates of Hours Equation with Five-Knot Spline

	(1)	(2)	(3)
$\boldsymbol{g}$			
Constant	223.049***	196.248***	200.228***
	(67.067)	(68.437)	(68.832)
$\hat{F}$	-1,809.817***	-1,663.513***	-1,683.505***
	(503.843)	(501.764)	(504.806)
S3	43,517.230***	40,590.026***	40,783.838***
	(13,758.907)	(13,557.588)	(13,788.434)
S4	-59,893.273***	-55,797.266***	-56,092.206***
	(19,317.577)	(19,012.526)	(19,376.718)
S5	16,711.270***	15,433.648***	15,597.193***
	(5,759.986)	(5,650.773)	(5,793.138)
$\boldsymbol{\lambda}$			
$\log \hat{W}$	-17.383	19.224	-26.406
	(19.397)	(45.782)	(20.052)
$\log(N+10)$	2.739	8.391	3.591
	(3.196)	(9.689)	(3.621)
$\log G$	-1.059	-10.137	-1.599
	(7.393)	(14.154)	(8.170)
$\log \hat{W}(1-t)$	-7.283	1.053	-11.807
	(9.044)	(26.643)	(10.380)
Age			0.797***
			(0.298)
Black			0.627
			(3.819)
Interactions			
$\log \hat{W} \times \hat{F}$		-37.479	
		(51.298)	
$\log(N+10) \times \hat{F}$		-7.994	
	(12.728)		
$\log G \times \hat{F}$		9.591	
		(14.690)	
$\log \hat{W}(1-t) \times \hat{F}$		-20.125	
		(37.917)	

Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$. This table reports estimates for the hours equation. The first stage is the probit model for AFDC participation in column (2) of Table 2. Variables under the $\boldsymbol{\lambda}$ heading are expressed as deviations from their respective means. Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level. *Table continues onto the next page.*

Table 3: Estimates of Hours Equation with Five-Knot Spline (continued)

	(1)	(2)	(3)
$\boldsymbol{\beta}$			
$\log \hat{W}$	39.198***	36.691***	48.009***
	(7.379)	(8.351)	(10.357)
$\log N + 10$	-1.327	-1.611	-1.442
	(1.073)	(1.517)	(1.127)
Age	-0.237**	-0.261**	-0.542***
	(0.099)	(0.107)	(0.193)
Black	0.372	0.299	0.250
	(0.745)	(0.842)	(1.936)
Family Size	-0.681	-0.515	-0.892
	(0.594)	(0.617)	(0.581)
Number of Children < 6	-1.772**	-2.069***	-1.880**
	(0.737)	(0.741)	(0.800)
Food Stamp Guarantee	-21.382	-25.896	-22.206
	(15.592)	(16.813)	(16.613)
State Unemployment Rate	-0.336	-0.361	-0.351
	(0.216)	(0.223)	(0.223)
State Pct. Urban	-0.277***	-0.302***	-0.297***
	(0.077)	(0.082)	(0.085)
State Pct. Black	-4.114	-4.892	-3.929
	(4.748)	(4.654)	(5.047)
State Per-Capita Income	0.372	0.493	0.363
	(0.322)	(0.337)	(0.333)
Northeast	-13.669***	-14.347***	-15.363***
	(2.643)	(2.749)	(3.123)
Midwest	-4.518**	-4.689**	-5.477**
	(2.099)	(2.226)	(2.356)
West	-5.824*	-5.702*	-6.941**
	(2.974)	(3.072)	(3.386)
Constant	7.558	17.843	7.444
	(20.308)	(21.472)	(21.786)
GCV	318.62	319.06	317.87
Observations	3,381	3,381	3,381

Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$. This table reports estimates for the hours equation. The first stage is the probit model for AFDC participation in column (2) of Table 2. Variables under the $\boldsymbol{\lambda}$ heading are expressed as deviations from their respective means. Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level.

Table 4: Variable Means for Observations in Quintiles of $\hat{F}$ Distribution

	1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
Hourly wage	6.27	6.02	5.91	5.76	5.59
Weekly non-labor inc	11.79	6.99	4.61	2.93	4.03
Black	0.34	0.34	0.43	0.44	0.49
Age	34.50	31.94	30.43	28.98	29.24
Children < 6	0.42	0.58	0.69	0.90	1.31

Notes: This table reports variable means within quintiles of the center of the $\hat{F}$ distribution (0.25 to 0.66). The cutoffs for these quintiles are approximately 0.32, 0.39, 0.46, and 0.53. $\hat{F}$ is generated from a probit model using the inverse variance weighted index of the log of the AFDC administrative barrier variables and interactions with N.

Table 5: First Stage Estimates Using 1989 Law Change

	(1)
State Mean Barrier Index	-0.164
	(0.214)
1991–1992	0.093
	(0.072)
1991–1992 × State Barrier Index Residual	-0.596***
	(0.201)
Budget Constraint	✓
Demographic Controls	✓
Region FEs	✓
OLS F-Stat for Instruments	6.28
Pseudo F-Stat by Part. Prob. Range	
0.00–0.25	-0.95
0.25–0.66	8.19
0.66–1.00	-0.44
Observations	3,381

Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$. This table reports estimates of a probit model for AFDC participation onto a series of individual and state characteristics. The "Barrier Index" is the inverse variance weighted average of the log of the individual administrative barrier variables in Table 1. "State Mean Barrier Index" is the average of the value of the barrier index within a state from 1988–1992. "State Barrier Index Residual" is defined as the difference between the average of the state barrier index from 1988–1990 and the value of the barrier index for the state in 1991 and 1992. "Budget Constraint" variables include $\log \hat{W}$, $\log N + 10$, $\log G$, and $\log \hat{W}(1-t)$. "Demographic Controls" include age, black, family size, the number of children under 6, and the food stamp guarantee. "State Controls" include the unemployment rate, share of the state that is urban, share of the state population that is black, and the per-capita income in the state. Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level. All parameter estimates are available in Appendix Table A4. The second panel reports F-statistics from an OLS version of the probit model and within different participation probability ranges based on the probit estimates. To calculate the F-statistic within a specific range of $\hat{F}$, define $RSS(q)$ as the residual sum of squares, equal to the sum of $[P - \hat{F}]^2$ taken over all observations in the range. The F-statistic is calculated as (1) the difference in $RSS(q)$ for the restricted model excluding the instruments and the unrestricted model $RSS(q)$ including the instruments divided by the degrees of freedom, divided by (2) the residual variance computed over all observations in the sample, using $\hat{F}$ from the restricted model.

Table 6: First Stage Estimates Using Close Election RD

	Dem Govs within 60%		Rep Govs within 55%	
	Barrier Index OLS	AFDC Probit	Barrier Index OLS	AFDC Probit
Elections:				
Gov Vote Share	-0.212	0.517*	0.168	-0.841***
	(0.162)	(0.302)	(0.256)	(0.322)
Close Election	-0.066*	0.035	0.038	0.369***
	(0.035)	(0.079)	(0.073)	(0.131)
State Leg:				
Party Opposite Gov	0.056	0.100	-0.120*	0.368**
	(0.064)	(0.148)	(0.073)	(0.181)
Split	0.236***	-0.136	0.138**	0.140
	(0.044)	(0.094)	(0.063)	(0.157)
Interactions:				
Party Opposite Gov × Close Election	0.180**	-0.814***	0.048	-0.397***
	(0.076)	(0.171)	(0.071)	(0.127)
Budget Constraint	✓	✓	✓	✓
Demographic Controls	✓	✓	✓	✓
Region FEs	✓	✓	✓	✓
OLS F-Stat for Instruments		13.40		6.41
Pseudo F-Stat by Part. Prob. Range				
0.00–0.25		4.60		3.10
0.25–0.66		9.10		4.21
0.66–1.00		1.70		-0.39
0.00–0.33		7.12		0.94
0.33–0.66		6.58		6.36
Observations	3,152	3,152	3,152	3,152

Notes: $***p < 0.01$; $**\,p < 0.05$; $*p < 0.1$. This table reports estimates of an OLS regression of the barrier index onto a series of individual and state characteristics and a probit model for AFDC participation using those same variables. "Gov Vote Share" measures the share of the vote the Democratic or Republican candidate for governor received in the last election. "Close Election" is an indicator for whether the winning Democratic or Republican candidate's vote share was under 60% or 55%. "State Legislature" variables are indicators for the partisan control of the state legislature. Observations for states that aggregated by the SIPP or had outlier values for the political variables are omitted (i.e., Washington DC, Colorado, Maine, Vermont, Iowa, North Dakota, South Dakota, Alaska, Idaho, Montana, and Wyoming). "Budget Constraint" variables include $\log \hat{W}$, $\log N + 10$, $\log G$, and $\log \hat{W}(1-t)$. "Demographic Controls" include age, black, family size, the number of children under 6, and the food stamp guarantee. "State Controls" include the unemployment rate, share of the state that is urban, share of the state population that is black, and the per-capita income in the state. Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level. All parameter estimates are available in Appendix Table A5. The second panel reports F-statistics from an OLS version of the probit model and within different participation probability ranges based on the probit estimates. To calculate the F-statistic within a specific range of $\hat{F}$, define $RSS(q)$ as the residual sum of squares, equal to the sum of $[P - \hat{F}]^2$ taken over all observations in the range. The F-statistic is calculated as (1) the difference in $RSS(q)$ for the restricted model excluding the instruments and the unrestricted model $RSS(q)$ including the instruments divided by the degrees of freedom, divided by (2) the residual variance computed over all observations in the sample, using $\hat{F}$ from the restricted model.

Appendix A Additional Tables and Figures

Table A1: Means of Variables Used in the Analysis

	Full Sample	P=1	P=0
Weekly H	21.38	4.48	31.14
Share not working (H=0)	0.43	0.84	0.19
Share working part time (H=20)	0.12	0.10	0.13
Share working full time (H=40)	0.45	0.07	0.68
P	0.37	1.00	0.00
Log $\hat{W}$	1.78	1.74	1.81
N	25.03	7.81	34.98
Log (N+10)	2.97	2.58	3.19
Log G	-2.49	-2.38	-2.55
Log $\hat{W}$(1-t)	1.27	1.22	1.30
Age	32.48	30.27	33.75
Black	0.34	0.41	0.30
Education	10.89	10.49	11.13
Family size	3.09	3.37	2.94
No. Children < 6	0.79	1.14	0.58
Food Stamp Guarantee	0.78	0.78	0.78
Unemployment rate	6.35	6.44	6.30
Northeast	0.28	0.28	0.28
Midwest	0.27	0.27	0.26
West	0.22	0.25	0.20
State Percent Services	27.67	27.94	27.52
State Percent Manufacturing	15.39	15.31	15.43
State Percent Urban	76.26	77.49	75.55
Obs	3,381	1,238	2,143

Notes: This table reports the means of variables used in our analysis. To focus on a sample that has a relatively high probability of AFDC, we construct a sample composed of single mothers aged 20–55 with a high school education or less with total assets less than $1,500 and non-transferable non-labor income less than $1,000 per-month drawn from 1988–1992 SIPP interviews. We specify N as log(N+10) to allow a logarithmic specification with values of zero for N. See the text for other variable definitions. All dollar-denominated variables are in 1990 PCE dollars.

Table A2: Log Hourly Wage Equation Estimates

	(1) OLS	(2) Selection-Bias Adjusted
Age	0.014***	0.007***
	(0.001)	(0.002)
Education	0.047***	0.041***
	(0.008)	(0.008)
Black	-0.092***	0.011***
	(0.034)	(0.038)
Northeast	0.206***	0.268***
	(0.051)	(0.048)
Midwest	0.087**	0.074***
	(0.042)	(0.042)
West	0.120*	0.184***
	(0.064)	(0.047)
State Pct. Services	0.016**	0.018***
	(0.007)	(0.006)
State Pct. Manufacturing	0.004	0.008***
	(0.004)	(0.004)
State Pct. Urban	0.003	0.004***
	(0.002)	(0.001)
Observations	1,818	3,258

Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$. This table reports estimates of the log hourly wage equation. The first column uses OLS, while the second uses the selection bias adjustment using a Heckman lambda based on a first stage probit which includes all variables listed in the table plus family size, the number of children under 6, the food stamp guarantee, the state unemployment rate, N, G, and t. Education is measured as the highest grade completed. Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level.

Table A3: Estimated Impact of Instruments on AFDC Participation—Detailed Estimates

	(1)	(2)
$\log \hat{W}$	-3.703***	-3.686***
	(0.703)	(0.712)
$\log N + 10$	-0.436***	-0.071
	(0.031)	(0.090)
$\log G$	1.240***	1.244***
	(0.206)	(0.203)
$\log \hat{W}(1-t)$	1.602***	1.590***
	(0.396)	(0.393)
Age	0.012	0.012
	(0.009)	(0.009)
Black	0.125	0.135
	(0.094)	(0.093)
Family Size	-0.114***	-0.109**
	(0.043)	(0.043)
Number of Children < 6	0.303***	0.303***
	(0.039)	(0.040)
Food Stamp Guarantee	3.868***	3.753***
	(1.367)	(1.368)
State Unemployment Rate	0.022	0.026
	(0.018)	(0.019)

Notes: $***p < 0.01$; $**\,p < 0.05$; $*p < 0.1$. This table reports estimates of a probit model for AFDC participation onto a series of individual and state characteristics. The "Barrier Index" is the inverse variance weighted average of the log of the individual administrative barrier variables in Table 1. Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level. The second panel reports F-statistics from an OLS version of the probit model and within different participation probability ranges based on the probit estimates. To calculate the F-statistic within a specific range of $\hat{F}$, define $RSS(q)$ as the residual sum of squares, equal to the sum of $[P - \hat{F}]^2$ taken over all observations in the range. The F-statistic is calculated as (1) the difference in $RSS(q)$ for the restricted model excluding the instruments and the unrestricted model $RSS(q)$ including the instruments divided by the degrees of freedom, divided by (2) the residual variance computed over all observations in the sample, using $\hat{F}$ from the restricted model. *Table continues onto next page.*

Table A3: Estimated Impact of Instruments on AFDC Participation—Detailed Estimates (continued)

	(1)	(2)
State Pct. Urban	0.023***	0.023***
	(0.005)	(0.005)
State Pct. Black	0.190	0.147
	(0.400)	(0.396)
State Per-Capita Income	-0.106***	-0.099***
	(0.029)	(0.030)
Northeast	0.634***	0.602***
	(0.234)	(0.233)
Midwest	0.184	0.169
	(0.203)	(0.199)
West	-0.024	-0.036
	(0.242)	(0.240)
Barrier Index	-0.593***	-0.482**
	(0.208)	(0.213)
Barrier Index $\times N$		-0.007***
		(0.002)
OLS F-Stat for Instruments	11.13	6.85
Pseudo F-Statistic by Part. Prob. Range		
0.00–0.25	3.16	1.94
0.25–0.50	10.96	15.94
0.50–0.75	0.38	3.13
0.75–1.00	0.91	0.88
0.00–0.33	4.24	4.94
0.33–0.66	9.09	14.62
0.66–1.00	2.07	2.34
0.25–0.66	10.18	17.62
Observations	3,381	3,381

Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$. This table reports estimates of a probit model for AFDC participation onto a series of individual and state characteristics. The "Barrier Index" is the inverse variance weighted average of the log of the individual administrative barrier variables in Table 1. Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level. The second panel reports F-statistics from an OLS version of the probit model and within different participation probability ranges based on the probit estimates. To calculate the F-statistic within a specific range of $\hat{F}$, define $RSS(q)$ as the residual sum of squares, equal to the sum of $[P - \hat{F}]^2$ taken over all observations in the range. The F-statistic is calculated as (1) the difference in $RSS(q)$ for the restricted model excluding the instruments and the unrestricted model $RSS(q)$ including the instruments divided by the degrees of freedom, divided by (2) the residual variance computed over all observations in the sample, using $\hat{F}$ from the restricted model.

.

Table A4: First Stage Estimates Using 1989 Law Change—Detailed Estimates

	(1)
$\log \hat{W}$	-2.958***
	(0.634)
$\log N + 10$	-0.441***
	(0.031)
$\log G$	0.928***
	(0.190)
$\log \hat{W}(1-t)$	1.177***
	(0.427)
Age	0.009
	(0.007)
Black	0.147*
	(0.080)
Family Size	-0.057
	(0.039)
Number of Children < 6	0.302***
	(0.039)
Food Stamp Guarantee	4.003***
	(1.399)

Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$. This table reports estimates of a probit model for AFDC participation onto a series of individual and state characteristics. The "Barrier Index" is the inverse variance weighted average of the log of the individual administrative barrier variables in Table 1. "State Mean Barrier Index" is the average of the value of the barrier index within a state from 1988–1992. "State Barrier Index Residual" is defined as the difference between the average of the state barrier index from 1988–1990 and the value of the barrier index for the state in 1991 and 1992. Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level. The second panel reports F-statistics from an OLS version of the probit model and within different participation probability ranges based on the probit estimates. To calculate the F-statistic within a specific range of $\hat{F}$, define $RSS(q)$ as the residual sum of squares, equal to the sum of $[P - \hat{F}]^2$ taken over all observations in the range. The F-statistic is calculated as (1) the difference in $RSS(q)$ for the restricted model excluding the instruments and the unrestricted model $RSS(q)$ including the instruments divided by the degrees of freedom, divided by (2) the residual variance computed over all observations in the sample, using $\hat{F}$ from the restricted model. *Table continues onto next page.*

Table A4: First Stage Estimates Using 1989 Law Change—Detailed Estimates (continued)

	(1)
Northeast	0.303
	(0.214)
Midwest	0.111
	(0.189)
West	0.094
	(0.217)
State Mean Barrier Index	-0.164
	(0.214)
1991–1992	0.093
	(0.072)
1991–1992 × State Barrier Index Residual	-0.596***
	(0.201)
OLS F-Stat for Instruments	6.28
Pseudo F-Stat by Part. Prob. Range	
0.00–0.25	-0.95
0.25–0.66	8.19
0.66–1.00	-0.44
Observations	3,381

Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$. This table reports estimates of a probit model for AFDC participation onto a series of individual and state characteristics. The "Barrier Index" is the inverse variance weighted average of the log of the individual administrative barrier variables in Table 1. "State Mean Barrier Index" is the average of the value of the barrier index within a state from 1988–1992. "State Barrier Index Residual" is defined as the difference between the average of the state barrier index from 1988–1990 and the value of the barrier index for the state in 1991 and 1992. Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level. The second panel reports F-statistics from an OLS version of the probit model and within different participation probability ranges based on the probit estimates. To calculate the F-statistic within a specific range of $\hat{F}$, define $RSS(q)$ as the residual sum of squares, equal to the sum of $[P - \hat{F}]^2$ taken over all observations in the range. The F-statistic is calculated as (1) the difference in $RSS(q)$ for the restricted model excluding the instruments and the unrestricted model $RSS(q)$ including the instruments divided by the degrees of freedom, divided by (2) the residual variance computed over all observations in the sample, using $\hat{F}$ from the restricted model.

Table A5: First Stage Estimates Using Close Election RD—Detailed Estimates

	Dem Govs within 60%		Rep Govs within 55%	
	Barrier Index OLS	AFDC Probit	Barrier Index OLS	AFDC Probit
$\log \hat{W}$	-0.228	-2.564***	-0.260	-2.629***
	(0.296)	(0.646)	(0.312)	(0.665)
$\log N + 10$	-0.006**	-0.445***	-0.006**	-0.445***
	(0.003)	(0.031)	(0.003)	(0.032)
$\log G$	0.163	0.825***	0.172	0.797***
	(0.108)	(0.154)	(0.106)	(0.157)
$\log \hat{W}(1-t)$	0.460*	0.677*	0.499*	0.713*
	(0.257)	(0.371)	(0.268)	(0.416)
Age	-0.003**	0.009	-0.003**	0.009
	(0.001)	(0.008)	(0.002)	(0.008)
Black	-0.008	0.142	-0.004	0.142
	(0.012)	(0.090)	(0.013)	(0.090)
Family Size	-0.022	-0.039	-0.025	-0.034
	(0.018)	(0.034)	(0.018)	(0.037)
Number of Children < 6	0.004	0.295***	0.005*	0.294***
	(0.003)	(0.041)	(0.003)	(0.041)
Food Stamp Guarantee	-0.308	1.873*	-0.341	1.628
	(0.463)	(1.131)	(0.488)	(1.099)

Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$. This table reports estimates of an OLS regression of the barrier index onto a series of individual and state characteristics and a probit model for AFDC participation using those same variables. "Gov Vote Share" measures the share of the vote the Democratic or Republican candidate for governor received in the last election. "Close Election" is an indicator for whether the winning Democratic or Republican candidate's vote share was under 60% or 55%. "State Legislature" variables are indicators for the partisan control of the state legislature. Observations for states that aggregated by the SIPP or had outlier values for the political variables are omitted (i.e., Washington DC, Colorado, Maine, Vermont, Iowa, North Dakota, South Dakota, Alaska, Idaho, Montana, and Wyoming). Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level. The second panel reports F-statistics from an OLS version of the probit model and within different participation probability ranges based on the probit estimates. To calculate the F-statistic within a specific range of $\hat{F}$, define $RSS(q)$ as the residual sum of squares, equal to the sum of $[P - \hat{F}]^2$ taken over all observations in the range. The F-statistic is calculated as (1) the difference in $RSS(q)$ for the restricted model excluding the instruments and the unrestricted model $RSS(q)$ including the instruments divided by the degrees of freedom, divided by (2) the residual variance computed over all observations in the sample, using $\hat{F}$ from the restricted model. *Table continues onto next page.*

Table A5: First Stage Estimates Using Close Election RD—Detailed Estimates (continued)

	Dem Govs within 60%		Rep Govs within 55%	
	Barrier Index OLS	AFDC Probit	Barrier Index OLS	AFDC Probit
Northeast	-0.473***	0.471*	-0.464***	0.548**
	(0.159)	(0.251)	(0.151)	(0.260)
Midwest	-0.325***	0.275	-0.327***	0.325*
	(0.118)	(0.172)	(0.123)	(0.190)
West	-0.232	0.198	-0.219	0.271
	(0.148)	(0.240)	(0.153)	(0.251)
Elections:				
Gov Vote Share	-0.212	0.517*	0.168	-0.841***
	(0.162)	(0.302)	(0.256)	(0.322)
Close Election	-0.066*	0.035	0.038	0.369***
	(0.035)	(0.079)	(0.073)	(0.131)
State Leg:				
Party Opposite Gov	0.056	0.100	-0.120*	0.368**
	(0.064)	(0.148)	(0.073)	(0.181)
Split	0.236***	-0.136	0.138**	0.140
	(0.044)	(0.094)	(0.063)	(0.157)
Interactions:				
Party Opposite Gov × Close Election	0.180**	-0.814***	0.048	-0.397***
	(0.076)	(0.171)	(0.071)	(0.127)
OLS F-Stat for Instruments		13.40		6.41
Pseudo F-Stat by Part. Prob. Range				
0.00–0.25		4.60		3.10
0.25–0.66		9.10		4.21
0.66–1.00		1.70		-0.39
0.00–0.33		7.12		0.94
0.33–0.66		6.58		6.36
Observations	3,152	3,152	3,152	3,152

Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$. This table reports estimates of an OLS regression of the barrier index onto a series of individual and state characteristics and a probit model for AFDC participation using those same variables. "Gov Vote Share" measures the share of the vote the Democratic or Republican candidate for governor received in the last election. "Close Election" is an indicator for whether the winning Democratic or Republican candidate's vote share was under 60% or 55%. "State Legislature" variables are indicators for the partisan control of the state legislature. Observations for states that aggregated by the SIPP or had outlier values for the political variables are omitted (i.e., Washington DC, Colorado, Maine, Vermont, Iowa, North Dakota, South Dakota, Alaska, Idaho, Montana, and Wyoming). Standard errors are in parenthesis and obtained using a weighted bootstrap procedure with 1,000 *iid* exponential weights drawn at the state-level. The second panel reports F-statistics from an OLS version of the probit model and within different participation probability ranges based on the probit estimates. To calculate the F-statistic within a specific range of $\hat{F}$, define $RSS(q)$ as the residual sum of squares, equal to the sum of $[P - \hat{F}]^2$ taken over all observations in the range. The F-statistic is calculated as (1) the difference in $RSS(q)$ for the restricted model excluding the instruments and the unrestricted model $RSS(q)$ including the instruments divided by the degrees of freedom, divided by (2) the residual variance computed over all observations in the sample, using $\hat{F}$ from the restricted model.

Appendix B Cubic Spline

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

$$g(\hat{F}) = g_1 + g_2\hat{F} + g_3S3 + g_4S4 + g_5S5 \tag{12}$$

where

$$S3 = d_1 - d_4 \tag{13}$$

$$S4 = d_2 - d_4 \tag{14}$$

$$S5 = d_3 - d_4 \tag{15}$$

where

$$d_1 = \frac{\max(0, \hat{F} - F_1) - \max(0, \hat{F} - F_5)}{F_5 - F_1} \tag{16}$$

$$d_2 = \frac{\max(0, \hat{F} - F_2) - \max(0, \hat{F} - F_5)}{F_5 - F_2} \tag{17}$$

$$d_3 = \frac{\max(0, \hat{F} - F_3) - \max(0, \hat{F} - F_5)}{F_5 - F_3} \tag{18}$$

$$d_4 = \frac{\max(0, \hat{F} - F_4) - \max(0, \hat{F} - F_5)}{F_5 - F_4} \tag{19}$$