Marginal Treatment Effects as a Random Coefficients Model, with an Application to the Labor Supply Effects of Welfare Program Participation*

Robert A. Moffitt† Johns Hopkins University

Matthew V. Zahn Johns Hopkins University

August 11, 2024

Abstract

We extend the random coefficients marginal treatment effect (MTE) model introduced by Björklund and Moffitt (1987) to a modern nonparametric framework with clearer identification conditions. The approach captures the MTE with a coefficient on the propensity score in the outcome equation which varies with the score. We provide an illustrative application to the impact of welfare program participation on labor supply, first modifying the standard labor supply model of transfers to incorporate marginal effects. We estimate a U-shaped MTE curve, initially increasing and then decreasing as participation rates rise. Marginal work disincentives are essentially zero in some ranges but sizable in others.

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. Research support from the National Institutes of Health is gratefully acknowledged.

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

Marginal treatment effects are now an established part of the literature on causal models. Given a continuous instrument, the marginal treatment effect (MTE) is the causal response to an offered treatment of individuals, or units in general, who are on the margin of participating in the treatment. Thus the MTE is a continuous version of the standard causal effect identified by a discrete instrument (e.g., a LATE). The origin of the MTE dates back to Björklund and Moffitt (1987), which appeared in this journal, and was framed as a random coefficients model. While the authors noted that, when properly specified, the random coefficient formulation was equivalent to what was then called the generalized Roy model—now called the Rubin Causal Model specified in terms of potential outcomes—they argued that the random coefficient formulation had interpretative advantages because it directly specified the heterogeneous response in terms of a single well-defined function, similar to a LATE with a discrete instrument.

The MTE literature has evolved in many directions since Björklund and Moffitt (1987). The MTE framework was under utilized until Heckman and Vytlacil (1999, 2005) formalized the concept, clarified its identification, and demonstrated that familiar treatment effect concepts could be characterized as weighted averages of underlying MTEs.[1] However, those authors formulated the MTE in a control function framework, with outcomes as a function of treatment conditional on the value of the error term at the indifference margin rather than as a random coefficients model (as noted in a comment by Moffitt, 1999). In the applications that have followed (see e.g., Doyle, 2007; Carneiro et al., 2011; Maestas et al., 2013; Kowalski, 2016; Cornelissen et al., 2018; Bhuller et al., 2020), the MTE is generally expressed either as a reduced form in the propensity score or in a control function approach. The random coefficient formulation has not been generally employed.

[1] Angrist et al. (2000) also showed that a continuous elasticity curve estimated with IV yields a weighted average of underlying continuous effects.

This paper updates the Björklund-Moffitt random coefficients model with a more modern and now well-understood nonparametric formulation. Björklund-Moffitt assumed multivariate normality for all unobservables and applied traditional parametric maximum likelihood rather than the simple instrumental variable (IV) approach in use today. We show that the random coefficient formulation results in a model where the outcome is a function of the propensity score but that the coefficient on that score can vary with the propensity score itself. It is the coefficient on the propensity score that is the object of interest, and nonparametric estimation of the model requires nonparametric estimation of that coefficient. The MTE is then calculable from the manner in which that coefficient varies with the propensity score. This formulation has intuitive appeal and is different than that used in most other MTE applications, although we stress that it has a one-to-one equivalence with most other approaches.

A second contribution of our paper is to provide an illustration of the random coefficient approach with an application to the literature on the effects of welfare programs on labor supply. While there have been a wide variety of applications of the MTE approach—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), and incarceration (Bhuller et al., 2020)—this paper is the first to apply the method to the effect of welfare programs on labor supply. The vast literature on welfare program work disincentives, which represents responses as a function of welfare guarantees and tax rates (see Moffitt, 1992, 2003; Ziliak, 2016, for reviews) assumes constant coefficients on the variables of interest and allows only limited forms of unobserved heterogeneity. In a MTE framework, instead, heterogeneity allows the labor supply disincentives of those on the margin of participating to differ from the responses of those already on the program, and to change as a program expands or contracts, thereby bringing (removing) those on the margin into (from) the program. Our paper develops a new but simple theoretical static labor supply model which shows that the sign of the MTE for labor supply is ambiguous even when utility functions are well-behaved and when positive selection occurs.

We then illustrate how to apply the random coefficients MTE framework empirically with an application to the pre-1996 version of the Aid to Families with Dependent Children (AFDC) program—the last U.S. cash welfare program to take the classic negative-income-tax form. The random coefficient model is specified and estimated using instruments based on three distinct sources of variation in the probability of program participation. The results show a U-shaped MTE curve: as participation in the program expands from low participation rates to modest ones, labor supply supply disincentives of those brought into the program grow (i.e., become more negative) but, as participation rates rise beyond a certain threshold, the marginal work disincentives fall (i.e., become less negative). We provide an economic interpretation of this U-shape by arguing that it is likely due to changing responses of those in full-time and part-time work. The policy implication is that work disincentives of welfare program participation are not large on average but the average response masks some margins where effects are close to zero and some where the effects are sizable. Beyond our methodological contribution, our analysis also contributes to the empirical literature that studies the effects of welfare programs on labor supply.

The paper proceeds as follows. Section 1 sets up the simple random coefficient model and Section 2 has the application, consisting of a short theoretical section adapting the standard labor supply model with welfare programs to have heterogeneous responses, and then an empirical illustration. Section 3 concludes the paper.

1 The Random Coefficient Model

The Björklund and Moffitt (1987) model was the following:

$$Y_i = X_i\beta + \alpha_i T_i + \epsilon_i \quad (1)$$

$$\alpha_i = W_i\gamma + \omega_i \quad (2)$$

$$T_i^* = \alpha_i - \phi_i \quad (3)$$

$$\phi_i = Z_i\psi + \nu_i \quad (4)$$

$$T_i = 1(T_i^* > 0) \quad (5)$$

where Y_i is the outcome for individual i, X_i is a vector of observables, and T_i is a binary treatment indicator. The effect of the treatment on the outcome is the random coefficient α_i, which varies across individuals. Equation (2) shows the effect to be related to observables W_i and an unobservable ω_i. Equation (3) specifies the propensity to be in treatment as a function of the "gains" (α_i) and the "costs" (ϕ_i) of participation, which is a particular Roy model interpretation that is not necessary for the econometrics of the problem. Costs are a function of an observable vector Z_i and an unobservable ν_i, with Z_i presumed to contain at least one element not in X_i or W_i. Equation (5) expresses treatment determination as a threshold function of an index, but the threshold formulation is not necessary for anything here.[2] All unobservables were assumed to be mean zero and distributed independently of X_i, W_i, and Z_i but not necessarily independently of each other.

[2]Vytlacil (2002) shows that the latent index model familiar in selection models is equivalent to the LATE model's assumptions of monotonicity and separability.

The authors noted that the model implies that:

$$Y_i = \begin{cases} X_i\beta + W_i\gamma + \epsilon_i + \omega_i & \text{if } T_i{=}1 \\ X_i\beta + \epsilon_i & \text{if } T_i{=}0 \end{cases} \tag{6}$$

and hence is equivalent to the conventional potential outcomes framework in the literature, with suitable changes of notation and variable definitions.

Björklund and Moffitt assumed the unobservables had a multivariate normal distribution and estimated the model with maximum likelihood. The goal in this paper is to relax those parametric assumptions, remove unnecessary identifying assumptions, and specify a more transparent model in line with modern approaches. To simplify notation, assume we are implicitly conditioning on X_i and W_i, and assume that Z_i is a scalar. We write a basic model as:

$$Y_i = \beta_i + \alpha_i T_i \tag{7}$$

$$T_i^* = m(Z_i) + \delta_i \tag{8}$$

$$T_i = 1(T_i^* > 0) \tag{9}$$

Each individual i has a value of Y if participating and if not participating. The function $m(\cdot)$ is not allowed to be individual-specific. The unobservable δ_i is assumed to be separable from the $m(\cdot)$ function. These assumptions are needed to satisfy the monotonicity condition of Imbens and Angrist (1994).[3]

The object of interest is the distribution of α_i. Selection in this model can occur either on the intercept (β_i) or the slope coefficient (α_i) or both because both may be related to δ_i. Identification is most easily seen in the reduced form; obtained by writing the

[3]See Vytlacil (2002) and Heckman and Vytlacil (2005) for a discussion.

means of Y and T conditional on Z_i:

$$E(Y_i \mid Z_i = z) = E(\beta_i \mid Z_i = z) + E(\alpha_i \mid T_i = 1, Z_i = z)\,\Pr(T_i = 1 \mid Z_i = z) \tag{10}$$

$$\Pr(T_i = 1 \mid Z_i = z) = \Pr[\delta_i \geq -m(z)] \tag{11}$$

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

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

$$E(\alpha_i \mid T_i = 1, Z_i = z) = g[E(T_i \mid Z_i = z)] \tag{A2}$$

where g is the effect of the treatment on the treated conditional on Z_i. That effect depends on the shape of the distribution of α_i and how different fractions of participants are selected from different portions of that distribution. While equation (A1) is familiar, equation (A2) may be less so. The usual assumption in the literature is that the two potential outcomes, β_i and $\beta_i + \alpha_i$, are fully independent of Z_i, which implies that α_i is as well. Equation (A2) is a weaker condition which states that all that is required is that the mean of α_i conditional on participation is independent of Z_i conditional on the participation probability (i.e., the propensity score). Variation in Z_i generates variation in participation which induces variation in the conditional mean of α_i, but there should be no other channel by which Z_i affects that conditional mean.

Inserting equations (A1) and (A2) into the main model in equations (10)–(11), and denoting the propensity score as $F(Z_i) = E(T_i \mid Z_i)$, where F is the cdf of δ, we obtain two estimating equations:

$$Y_i = \beta + g[F(Z_i)]F(Z_i) + \epsilon_i \tag{12}$$

$$T_i = F(Z_i) + \xi_i \tag{13}$$

where ϵ_i and ξ_i are mean zero and orthogonal to the RHS by construction. No other restriction on these error terms is necessary as this is a reduced form of the model.

Equation (12) is the key to our approach and forms the basis for estimation. It shows that the population mean of Y_i—taken over participants and nonparticipants—equals a constant plus the mean response of those in the program times the fraction that is in the program. The implication of this model specification—that is, as a random coefficient model—is that preference heterogeneity is detectable by a nonlinearity in the response of the population mean of Y_i to changes in the participation probability. If responses are homogeneous and hence the same for all members of the population, the function g reduces to a constant and therefore a shift in the fraction on the program has a linear effect on the population mean of Y_i. However, if the responses of those on the margin vary, the response of the population mean of Y_i to a change in participation will depart from linearity. Estimation can proceed by assuming the function g is a nonparametric function of F, thereby allowing the data to determine the shape of g and hence the MTE curve (defined below).[4]

Nonparametric identification of the parameters of the model—β, the function g at every point of the propensity score F, and the propensity score F itself—has been extensively discussed in the literature, so we only briefly restate those conditions. The propensity score F is identified at every data point Z_i from the second equation from the mean of T_i at each value of Z_i (apart from sampling error). With identification of the propensity score F, the LATE of Imbens and Angrist (1994) is identified by the discrete difference in Y between two points z and z' divided by the difference in the propensity score between those two points. With multiple values of z, multiple LATE values are identified.

[4]The term $g[F(Z_i)]F(Z_i)$ in equation (12) can be collapsed into a single function of $F(Z_i)$ and estimation can be conducted by a direct nonparametric estimation of that equation. Equation (12) factors $F(Z_i)$ out and labels its coefficient as $g[F(Z_i)]$. Testing for homogeneity is slightly easier in this formulation because it only requires testing whether g varies with the score instead of testing, equivalently, whether the outcome is quadratic in the score.

A MTE is a continuous version of this and requires some smoothing method across discrete values of Z. The MTE is computed by $\partial Y/\partial F = g'(F)F + g(F)$. However, while the MTE $\partial Y/\partial F$ is identified, g and g' are not unless there is a value of Z_i in the data for which $F(Z_i) = 0$. In that case, β is identified from the mean of Y_i at that point and hence g is identified pointwise at every other value of z since the propensity score is identified. If no such value is in the data, then g can only be identified subject to a normalization of its value at a particular value of z or if the value of g is known at some value.

2 Illustration Using the Labor Supply Effects of a Transfer Program

We first modify the standard static model of the labor supply response to transfer programs then provide an illustrative estimation of the model.

2.1 Modifying the Static Labor Supply Model for Heterogeneous Response

The canonical static model of the labor supply response to transfers (Moffitt, 1983; Chan and Moffitt, 2018) assumes utility to be:

$$U(H_i, Y_i; \theta_i) - \phi_i P_i \tag{14}$$

where H_i is hours of work for individual i, Y_i is disposable income, P_i is a program participation indicator (switching from T_i to P_i to use the standard notation of the transfers literature), θ_i is a vector of labor supply preference parameters, and ϕ_i is a scalar representing fixed costs of participation in utility units whose distribution is in the positive domain. The presence of P_i allows for the presence of fixed costs of participation—in money, time, or utility (stigma), with the exact type unspecified and scaled in units of utility (Moffitt, 1983; Daponte et al., 1999; Currie, 2006). Some type of cost is required to fit the data on almost all transfer programs because many individuals who are eligible for transfer programs do not participate in them. The presence of participation costs also provides a potential source of exogenous variation in participation, and such variation will be used in the empirical work in the next section.

The 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 nonlabor income, respectively, and hence the budget constraint is:

$$Y_i = \begin{cases} W_i(1-t)H_i + G + (1-r)N_i & \text{if } P_i = 1 \\ W_iH_i + N_i & \text{if } P_i = 0 \end{cases} \tag{15}$$

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

$$H_i = H[W_i(1-tP_i), N_i + P_i(G - rN_i); \theta_i] \tag{16}$$

$$P_i^* = V[W_i(1-t), G + N_i(1-r); \theta_i] - V[W_i, N_i; \theta_i] - \phi_i \tag{17}$$

$$P_i = 1(P_i^* \geq 0) \tag{18}$$

where $H[\cdot]$ is the labor supply function, $V[\cdot]$ is the indirect utility function and $1(\cdot)$ is the indicator function. The direct utility gain from participation (the change in $V[\cdot]$) must be greater than costs to generate participation.

The labor supply response to the program for individual i conditional on the budget constraint parameters is the change in hours worked from participating:

$$\triangle_i(\theta_i|C_i) = H[W_i(1-t), G + N_i(1-r); \theta_i] - H[W_i, N_i; \theta_i] \tag{19}$$

where $C_i = [W_i, N_i, G, t, r]$ is the set of budget constraint variables. The marginal labor supply response is the value of $\triangle_i$ for individuals whose θ 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:

$$0 = V[W_i(1-t), G + N_i(1-r); \theta_D] - V[W_i, N_i; \theta_D] - \phi_D \tag{20}$$

and it is the set of values (θ_i, ϕ_i) that fall on one side of this (θ_D, ϕ_D) locus that generate participation. If we use this locus to define a function $\theta_D = f(\phi_D|C)$, then the marginal labor supply response conditional on the budget constraint is the integral of this function over the distribution of participation costs ϕ.

One difference with the textbook Roy model arises here. In that model, positive selection is said to occur when individuals who have greater gains in the outcome variable (e.g., a higher earnings payoff from college) are more likely to participate (e.g., attend college). But here there is no necessary relationship between the labor supply response, θ, and the probability of participating; if "positive" selection means that those with greater labor supply reductions are more likely to participate, such selection does not necessarily hold. Positive selection only occurs on indirect utility $V[\cdot]$, with greater gains in V leading to higher participation. However, greater gains in V could come either from greater gains in consumption or leisure. As a consequence, as participation rises, the marginal labor supply response could rise, fall, or stay the same as relative preferences for consumption and leisure change over new parts of the distribution. This is an additional reason that a nonparametric specification of that response is needed for any empirical work.

To estimate the marginal labor supply response, let $S_{\theta\phi}$ denote the set of parameters that generate participation. Then the mean effect of the transfer program over the entire population, participants and non-participants combined, conditional on the budget constraint, is:

$$\begin{aligned}\widetilde{\triangle} &= E(\triangle_i P_i | C_i) \\ &= \int\limits_{S_{\theta\phi}} \int \triangle_i(\theta_i | C_i) dJ(\theta_i, \phi_i)\end{aligned} \tag{21}$$

where $J(\theta_i, \phi_i)$ is the joint distribution function of θ_i and ϕ_i. With an exogenous shift in the distribution of ϕ_i (e.g., a reduction in costs), the mean labor supply response is shifted but so is the participation rate:

$$\begin{aligned}P &= E(P_i | C_i) \\ &= \int\limits_{S_\phi} \int\limits_{S_\theta} 1\{V[W_i(1-t), G + N_i(1-r); \theta_i] - V[W_i, N_i; \theta_i] - \phi_i\} dJ(\theta_i, \phi_i)\end{aligned} \tag{22}$$

where S_θ and S_ϕ represent the unconditional supports of the two parameters. The marginal labor supply at that point is the change in mean labor supply from a small increment in participation, or $\partial\widetilde{\triangle}/\partial P$. Thus estimating mean labor supply responses as a function of the participation rate will permit estimation of the MTE.

2.2 Empirical Illustration

The general form of the random coefficient model is presented in equations (12)–(13). In this section, we convert our theoretical model into an empirical one that we can take to data. We start by defining a vector X_i that consists of the budget constraint variables $(W_i, N_i, G_i, t_i, r_i$, where we now allow the benefit formula variables to vary across individuals) plus exogenous demographic covariates (e.g., age, family size, etc.). We re-specify the model as:

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

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

where, for sample size reasons, we remain parametric on all components except g, which we will estimate nonparametrically using sieve methods (see below).

Equation (24) is the propensity score equation, which is a function of the budget constraint variables, demographics, and a observable scalar Z_i proxying for the costs of participation.[5] Equation (23) explicitly shows the arguments of the propensity score function $F(X_i\eta + \delta Z_i)$ but allows the treatment effect to also interact with observables X_i. Allowing treatment effects to vary with exogenous observables is common in the literature.[6] A minor restriction for theoretical reasons is that the intercept term in the outcome equation is specified as a function of a vector X_i^β which excludes the benefit formula variables G_i, t_i, and r_i because that intercept represents hours of work when off welfare, and those should not be a function of the benefit parameters.

With these two functions specified, we will employ two-step estimation of the model, with a first-stage probit estimation of equation (24) and second-stage nonlinear least squares estimation of equation (23) using fitted values of the propensity score F from the first stage. Consistency and asymptotic normality of two-step estimation of nonlinear conditional mean functions with estimated first-stage parameters is demonstrated in Newey and McFadden (1994). Standard errors are obtained by jointly bootstrapping equations (23) and (24) as well as a wage equation using weights randomly drawn at the state-level (Rubin, 1981; Cheng et al., 2013).

<u>Data.</u> The type of open-ended, cash transfer program illustrated in the theoretical model no longer exists in the U.S. (at least for the non-elderly non-disabled). The last program to take this form was the Aid to Families with Dependent Children (AFDC) program, which was transformed starting in 1993 with the introduction of work requirements, time limits, and other features that made it a different type of program than

[5] We will estimate the equation with a probit model. Estimating with OLS gives almost identical results.

[6] $X_i\lambda$ will be normalized to have mean zero to allow the $\tilde{g}(\cdot)$ function to have an intercept. We estimate some specifications which interact X with g.

that illustrated above. We therefore provide an illustration of the pre-1993 version of the AFDC program.

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 (only Spring to avoid seasonal variation) and pool them into one sample, excluding overlapping observations by including only the first interview when the person appears to avoid dependent observations.

Eligibility for AFDC in this period required sufficiently low assets and income and, for the most part, required that eligible families be single mothers with at least one child under 18. The sample is therefore restricted to such families, similar to the practice in prior AFDC research. To concentrate on the AFDC-eligible population, we restrict the sample to those with completed education of 12 years or less, non-transfer non-labor income less than $1,000 per month, and between the ages of 20 and 55. The resulting data set has 3,381 observations.

The means of the variables used 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), whether the mother was on AFDC at any time in the prior month (P), and covariates for education, age, race, and family structure (several state characteristics are also used as conditioning variables).[7] Thirty-seven percent of the observations were on AFDC. 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.[8] For N, the weekly value of non-transfer non-labor income reported in the survey is used. AFDC guarantees and tax rates by year, state, and family size are taken from estimates by Ziliak (2007), who used administrative caseload data to estimate "effective" guarantees and tax rates. The effective guarantees and tax rates in the AFDC program 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

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

[8]These results are similar to those presented in the text and are available upon request.

[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.

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. We require instruments Z_i that proxy fixed costs of participation that affect participation but not labor supply directly and satisfy the mean independence conditions in equations (A1) and (A2). 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 were politically driven at the gubernatorial and state legislature level and were aimed at keeping caseloads in the program down. 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 the states. We will use them as instruments for AFDC participation in our SIPP data in three alternative ways. First, we will simply use cross-state variation in the error rates and will 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. There are obvious 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 but 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. Consequently, we use construct two other instruments as well. The second is based on a differences-in-differences (DD) approach. Although there was mostly no change in government policy in the states over the period, so a general DD approach is not possible, there was one piece of federal legislation in 1989 that altered the federal monitoring process. We find this law had differential effects on welfare participation across states. 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 approach. We draw on the literature arguing that political differences across the states were responsible for the differences in error rates, and use a traditional close election regression discontinuity (RD) as an instrument, using narrowly elected Democratic governors combined with a Republican legislature, 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 different sources of variation, nevertheless yield a MTE curve with a similar shape. Taken together, this consistency across identification strategies increases confidence in the results of our empirical illustration of the random coefficient MTE framework.

2.3 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]

[11] The administrative variables bounce around 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

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 second specification includes an interaction with non-labor income, which we found to be highly predictive of AFDC participation. 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.[14] But these statistics are not relevant to the MTE model which estimates a continuous treatment effect. Such a model requires a measure of the instrument's strength at each point along the continuous MTE curve. Such statistics have 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). To illustrate the relevance of this issue, we estimate what we term "pseudo-F statistics" for different discrete ranges of the propensity score distribution, first for quartiles and then for terciles. 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 tails. This pattern is

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.

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

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 simply restrict our estimates of the MTE 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.[15]

Results. Estimating equation (23) using the fitted values of the participation probabilities for F yields estimates of β, λ, and the parameters of the g function.[16] The g function is estimated with conventional cubic splines, hence $g(F) = g_0 + \sum_{j=1}^{J} g_j \max(0, F - \pi_j)^3$, where the π_j are J preset spline knots. For a given J, the knots will are chosen to be regularly spaced within the (0.25, 0.66) range. The estimation starts 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

[15] We note that Angrist and Kolesar (2024) have recently noted that the Stock-Yogo analysis is motivated by the case of many instruments, and that specifications with a low number of instruments do not suffer from the same problem, as well as noting that Stock and Yogo are concerned with maximum possible bias, which occurs when the correlation coefficient measuring the degree of endogeneity is ± 1. Angrist and Kolesar show that that coefficient can be estimated by comparing OLS and IV estimates, but we cannot estimate that coefficient in our application because our "OLS" equation is equation (1) and would have to be estimated allowing a distribution of α_i. We have generated some approximate estimates of an average coefficient which suggest that it may be quite low. These results are available upon request.

[16] We use the second specification in Table 2 but the results are very similar for the first specification.

Appendix B.[17]

To illustrate our choice of the number of knots, Figure 1 shows the estimated MTE curves in the (0.25, 0.66) range with 90 percent confidence intervals for three-to-six knots.[18] The 3-knot and 4-knot specifications show monotonic MTE 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 Figure 1. Column (1) has only the budget constraint variables in the λ vector, which are not very strong predictors of hours, 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 no effects are found there. Column (3) adds Age and Black to the 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 1. The estimates in columns (1) and (2) yield similar MTE curves.

We return to Figure 1 for substantive interpretation. The marginal labor supply responses are nonmonotonic and U-shaped, starting off at $F = 0.25$ slightly including zero in the confidence interval but then growing in (negative) size as participation increases

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

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

with confidence intervals excluding zero. The marginal response peaks at a participation probability about 0.36, when it reaches a labor supply disincentive of those on the margin of approximately -21 hours per week. It then declines, becoming insignificantly different from 0 at approximately $F = 0.42$. 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 zero at many margins but non-trivially negative at other margins, depending on the initial participation rate at the time any expansion would have begun.

Some insight into the mechanics behind the U-shaped pattern of responses we have found can be gained by examining marginal responses between full-time work, part-time work, and non-work. Figure 2 shows the results of estimating the hours worked equation by successively replacing the dependent variable for H with binary variables for not working, working part-time, and working full-time. The leftmost panel shows that the probability of non-work rises sharply as participation goes from 0.25 to 0.35, the same range where the MTE for average hours falls the most. The middle panel shows that the MTE for part-time work actually starts off at a positive level (albeit small), implying an increase in part-time work that can only come from full-time workers reducing labor supply to the part-time level. The part-time MTE becomes less positive as participation increases and eventually becomes zero or negative, implying that some part-timers move at that point to non-work. But the right panel shows that the MTE for full-time work is large and negative in the 0.25 to 0.35 participation rate range implying, when combined with the other panels, that a large part of the reduction in labor supply over that range is from full-time workers moving to part-time but mostly to non-work upon participation. Eventually, however, after participation rises high enough, movements out of full-time work fall to zero. Thus the decline in the labor supply reductions in average hours when participation rates rise sufficiently high reflects a decline in movements out of full-time work.[19]

2.4 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 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 MTE 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 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

[19]A separate analysis shows that those on the margin of participation at low participation rates have higher than average wage rates, are less likely to be black, are older, and have fewer young children, all of which are correlated with higher hours of work (and hence are likely to be starting off at full time work). Those on the margin at higher participation rates have the opposite characteristics, and are more likely to not work, at which point further reductions are not possible.

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 4 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.[20] 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. While below 10, we note again that Angrist and Kolesar (2024) found that just identified models need only have correlation coefficients for endogeneity only below about 0.50 for robust inference, and our explorations of the possible magnitude of that coefficient indicates its value to be possibly quite low.

The estimated MTE curve from the hours worked equation using this instrument in the hours equation is shown in Figure 3, 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 0.50 range, and with a peak (negative) work disincentive of -28 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 different source of variation used with this instrument, the substantive result is the same as for the first instrument, that marginal work disincentives are small or insignificantly different from zero at many margins of participation but substantial at other margins.

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

2.5 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 the 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 upon the literature on regression discontinuity designs in political economy research which use 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 is 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 elected in a close election, which we define as having been elected with at most 60% of the vote. We control for the Democratic share of the vote as the running variable. Using the Republican party instead of the Democratic party yields similar results. We also gather information on the political makeup of the state legislature, which we hypothesize could affect the ability of a 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 legislature is entirely Republican (both chambers) or whether it is split, with one chamber controlled by Democrats and one controlled by Republicans (a "split" legislature).[21] We test whether the impact of a Democrat governor who has been elected in

[21] Nebraska has a unicameral state legislature, so splits are only possible if that chamber is equally divided.

a close election varies with these legislative party control variables.

The two columns of Table 5 show the results of the relevant first stage estimation.[22] The first column tests whether a narrowly elected Democratic governor affects the level of the barrier index in the state. Preliminary testing revealed it to have little impact on its own but to have a significant impact depending on whether the legislature was fully controlled by Republicans. 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 second column shows the first-stage AFDC participation probit, showing that the same interacted variable has a negative effect on that participation, consistent with the higher barrier index effect in the first column. 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.

Figure 4 shows the estimated MTE curves using this close election variable as an instrument using 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 MTE 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 -33 hours per week, and then falling in absolute value as the caseload expands. The confidence interval includes zero at a participation rate of 0.53 or above.

[22] Appendix Table A5 contains estimates for all variables.

3 Summary

This paper has provided an illustration of how marginal treatment effect models can be set up in an intuitive and appealing way as a random coefficients model. Nonparametric estimation of the coefficient on the propensity score can be conducted using a variety of methods from in the literature; our empirical illustration uses sieve methods with natural cubic splines. We also provide an application to the marginal labor supply effects resulting from expansions and contractions of a welfare program, moving beyond the constant treatment effects in the bulk of the literature. We provide an empirical illustration to the historic AFDC program, the last program in the U.S. to take the classic, open-ended cash negative-income-tax form. Our analysis shows U-shaped marginal work disincentive curves, with the disincentive becoming more negative as the program expands from low participation rates to modest participation rates, then becoming less negative as the program expands further.

We suggest that future research on marginal treatment effects consider using the random coefficient model for the interpretative advantages we have outlined and for the clarity of identification conditions it offers. For research on welfare programs in particular, we suggest that the model be considered for current and more recent programs. Our finding for the pre-1993 AFDC program that marginal responses greatly vary depending on the scale of the program and who is and who is not on the margin may apply to welfare programs in general as well as to outcomes other than labor supply. Our analysis also reveals a gap in the weak IV literature, which has only considered constant treatment effect models. For continuous treatment effect models, particularly those with propensity score curves which necessarily flatten out at high and low values of the score, the strength of the instrument is likely to vary over the range of the score. New research on this issue is needed to provide confidence in the results of MTE models.

References

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

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

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

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

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

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

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

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

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

Cheng, G., Z. Yu, and J. Z. Huang (2013). The custer bootstrap consistency in generalized estimating equations. *Journal of Multivariate Analysis 115*, 33–47.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schoenberg (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 A. J. Auerbach, D. Card, and J. M. Quigley (Eds.), *Public Policy and the Income Distribution*, pp. 80–148. New York: Russell Sage Foundation.

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

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

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

Heckman, J. J. and E. Vytlacil (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences 96*(8), 4730–4734.

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

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

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

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

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

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

Maestas, N., K. J. Mullin, and A. Strand (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. (1983, December). An Economic Model of Welfare Stigma. *American Economic Review 73*(5), 1023–35.

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

Moffitt, R. (1999). Models of treatment effects when responses are heterogeneous. *Proceedings of the National Academy of Sciences 96*(8), 6575–6576.

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

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

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

Rubin, D. B. (1981). The bayesian bootstrap. *The Annals of Statistics 9*(1), 130–134.

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

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

Vytlacil, E. (2002). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica 70*(1), 331–41.

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

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

Tables and Figures

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$\times N$		-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	182.172***	177.437***	157.417**
	(67.429)	(67.264)	(68.634)
$\hat{F}$	-1,634.866***	-1,574.373***	-1,495.355***
	(508.930)	(516.896)	(509.494)
S3	41,338.290***	39,272.001***	38,339.154***
	(13,775.924)	(13,898.910)	(13,808.760)
S4	-57,133.232***	-53,988.278***	-52,966.936***
	(19,334.775)	(19,463.784)	(19,400.364)
S5	16,122.761***	14,906.941***	14,906.921***
	(5,765.008)	(5,760.739)	(5,801.959)
$\boldsymbol{\lambda}$			
$\log \hat{W}$	-8.056	41.269	-16.921
	(20.039)	(48.684)	(20.383)
N	0.304**	0.303	0.340**
	(0.120)	(0.249)	(0.133)
$\log G$	-4.511	-18.052	-5.550
	(7.660)	(16.089)	(8.425)
$\log \hat{W}(1-t)$	-10.391	-7.961	-15.558
	(9.084)	(29.306)	(10.663)
Age			0.841***
			(0.302)
Black			-0.241
			(3.801)
Interactions			
$\log \hat{W} \times \hat{F}$		-54.193	
		(53.784)	
$N \times \hat{F}$		0.043	
		(0.348)	
$\log G \times \hat{F}$		15.473	
		(16.386)	
$\log \hat{W}(1-t) \times \hat{F}$		-11.983	
		(40.297)	

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)
$\boldsymbol{\beta}$			
$\log \hat{W}$	40.047***	37.680***	49.904***
	(7.987)	(9.182)	(11.383)
$\log N + 10$	-2.720***	-2.191**	-2.835***
	(0.892)	(0.924)	(0.936)
Age	-0.267**	-0.298**	-0.595***
	(0.115)	(0.127)	(0.208)
Black	0.143	0.010	0.359
	(0.839)	(1.026)	(1.956)
Family Size	-0.626	-0.413	-0.845
	(0.583)	(0.639)	(0.571)
Number of Children < 6	-2.379***	-2.797***	-2.581***
	(0.811)	(0.893)	(0.902)
Food Stamp Guarantee	-25.889	-32.118*	-27.344
	(16.266)	(18.390)	(17.537)
State Unemployment Rate	-0.359	-0.386	-0.384
	(0.227)	(0.248)	(0.239)
State Pct. Urban	-0.295	-0.329***	-0.319
	(0.083)	(0.091)	(0.093)
State Pct. Black	-4.299***	-5.156	-4.195
	(4.766)	(4.872)	(5.126)
State Per-Capita Income	0.431	0.584*	0.420***
	(0.320)	(0.355)	(0.331)
Northeast	-14.465***	-15.465***	-16.398***
	(2.838)	(3.061)	(3.461)
Midwest	-4.860*	-5.140**	-5.940**
	(2.279)	(2.503)	(2.616)
West	-6.021*	-5.896*	-7.286**
	(3.131)	(3.324)	(3.629)
Constant	18.490	25.911	18.212***
	(20.086)	(21.233)	(21.560)
GCV	318.08	318.47	317.25
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: 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	✓
State 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

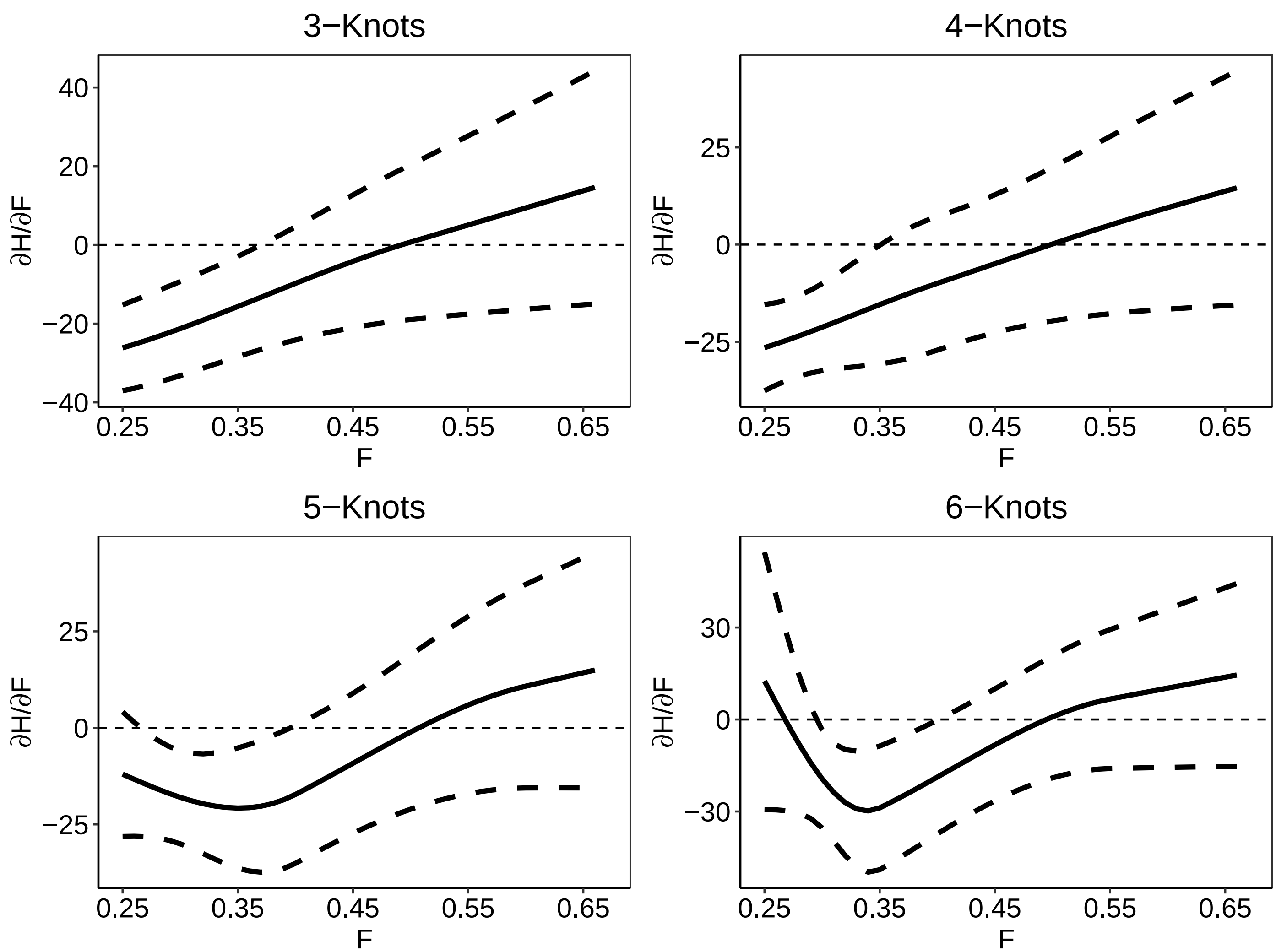
Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$. This table reports estimates of a probit model for AFDC participation onto a series of individual and state characteristics. The "Barrier Index" is the inverse variance weighted average of the log of the individual administrative barrier variables in Table 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 5: First Stage Estimates Using Close Election RD

	Barrier Index OLS	AFDC Probit
Elections:		
Dem Gov Vote Share	-0.212	0.517*
	(0.162)	(0.302)
Dem Share Under 60%	-0.066*	0.035
	(0.035)	(0.079)
State Legislature:		
Republican	0.056	0.100
	(0.064)	(0.148)
Split	0.236***	-0.136
	(0.044)	(0.094)
Interactions:		
Republican × Dem Share Under 60%	0.180**	-0.814***
	(0.076)	(0.171)
Budget Constraint	✓	✓
Demographic Controls	✓	✓
State Controls		
Region FEs	✓	✓
OLS F-Stat for Instruments		13.40
Pseudo F-Stat by Part. Prob. Range		
0.00–0.25		4.60
0.25–0.66		9.10
0.66–1.00		1.70
Observations	3,152	3,152

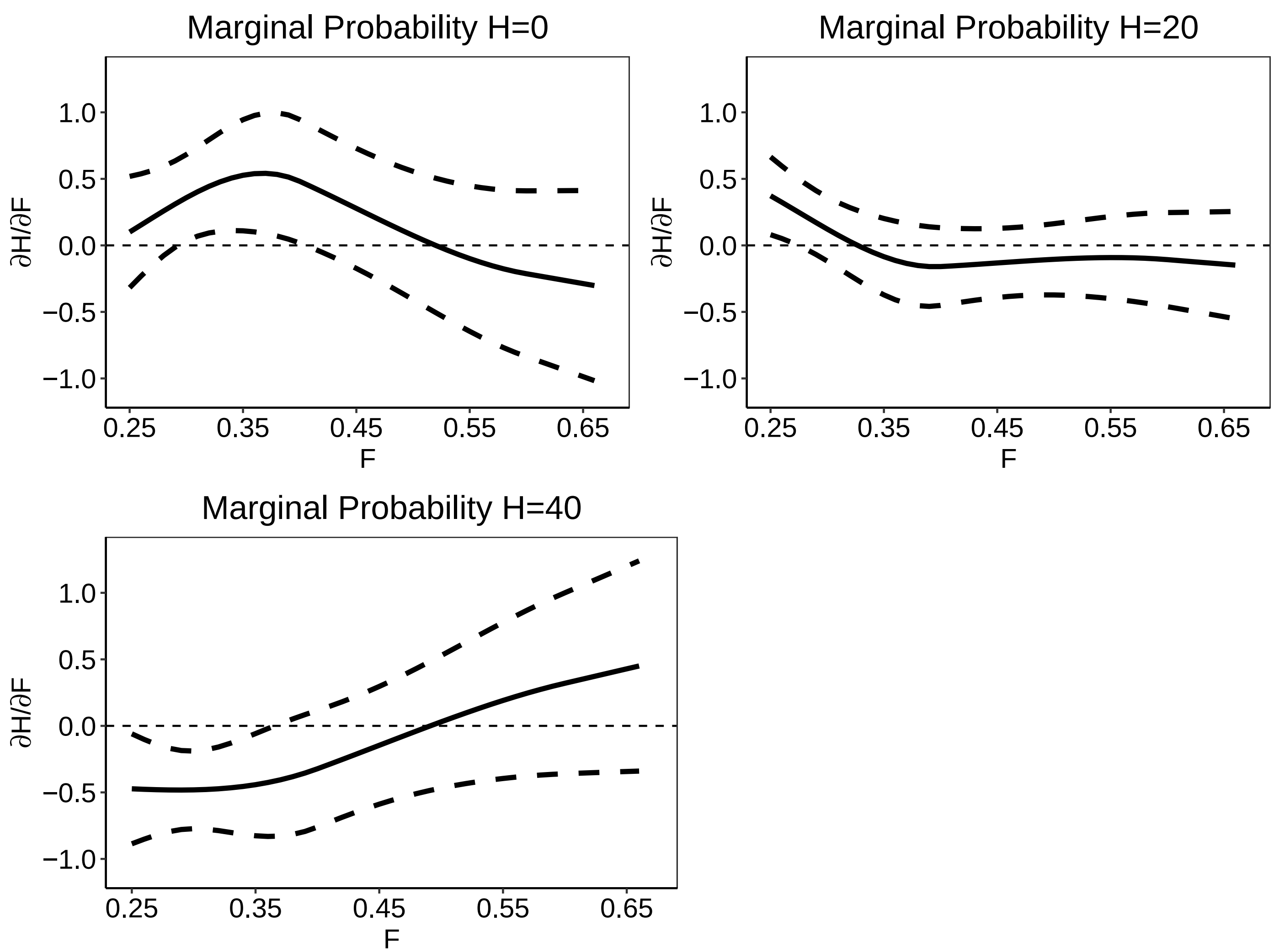
Notes: $***p < 0.01$; $**p < 0.05$; $*p < 0.1$.This table reports estimates of an OLS regression of the barrier index onto a series of individual and state characteristics and a probit model for AFDC participation using those same variables. "Dem Gov Vote Share" measures the share of the vote the Democratic candidate for governor received in the last election. "Dem Share Under 60%" is an indicator for whether the winning democratic candidate's vote share was under 60%. "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.

Figure 1: Marginal Labor Supply Curves for Different Natural Cubic Splines



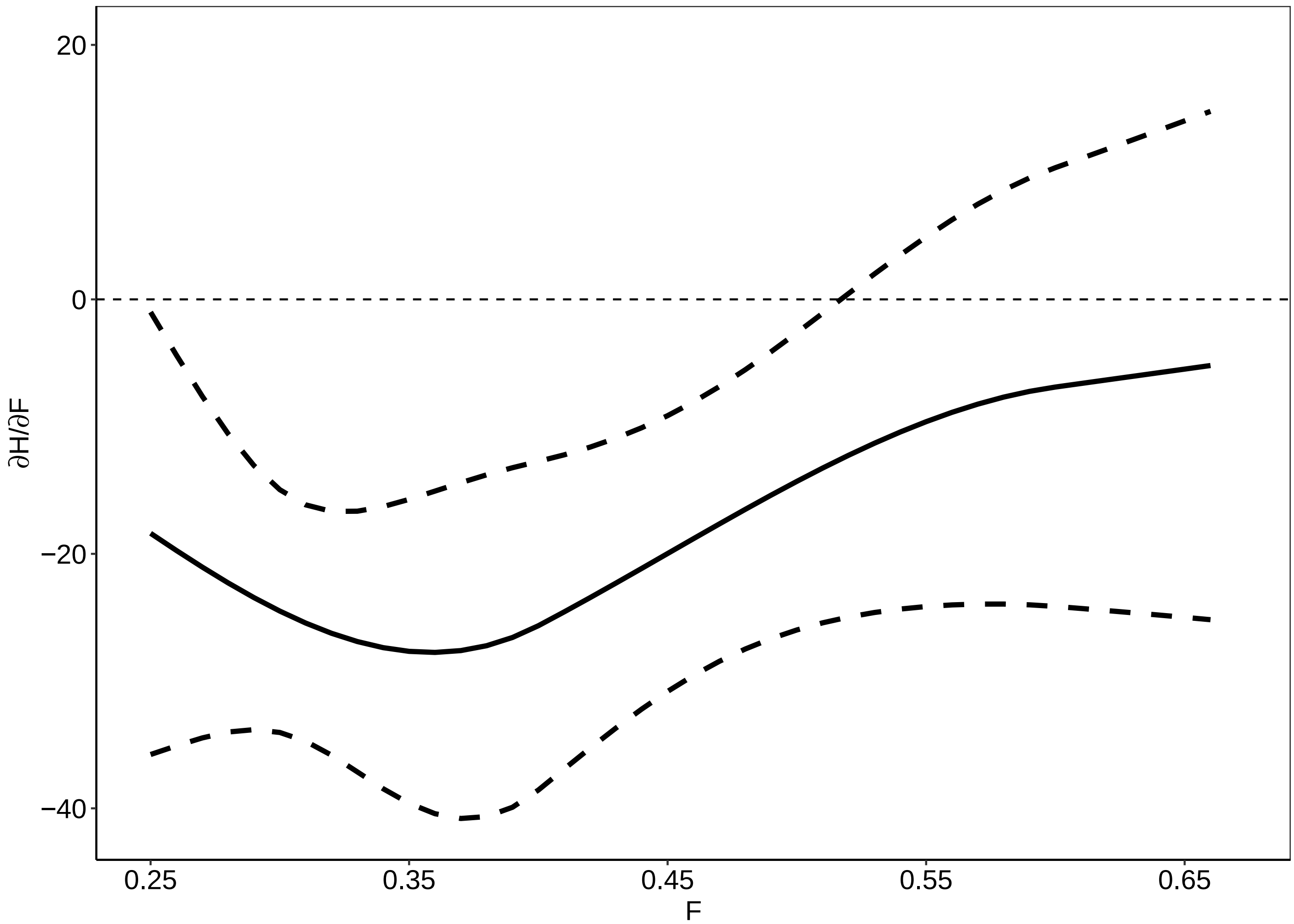
Notes: This figure plots the marginal treatment effect curves using different cubic spline specifications. All specifications use a first stage probit model with the inverse variance weighted log of the AFDC administrative barriers and interactions with N. The dashed lines denote 90 percent confidence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

Figure 2: Marginal Labor Supply Curves for Different Types of Workers



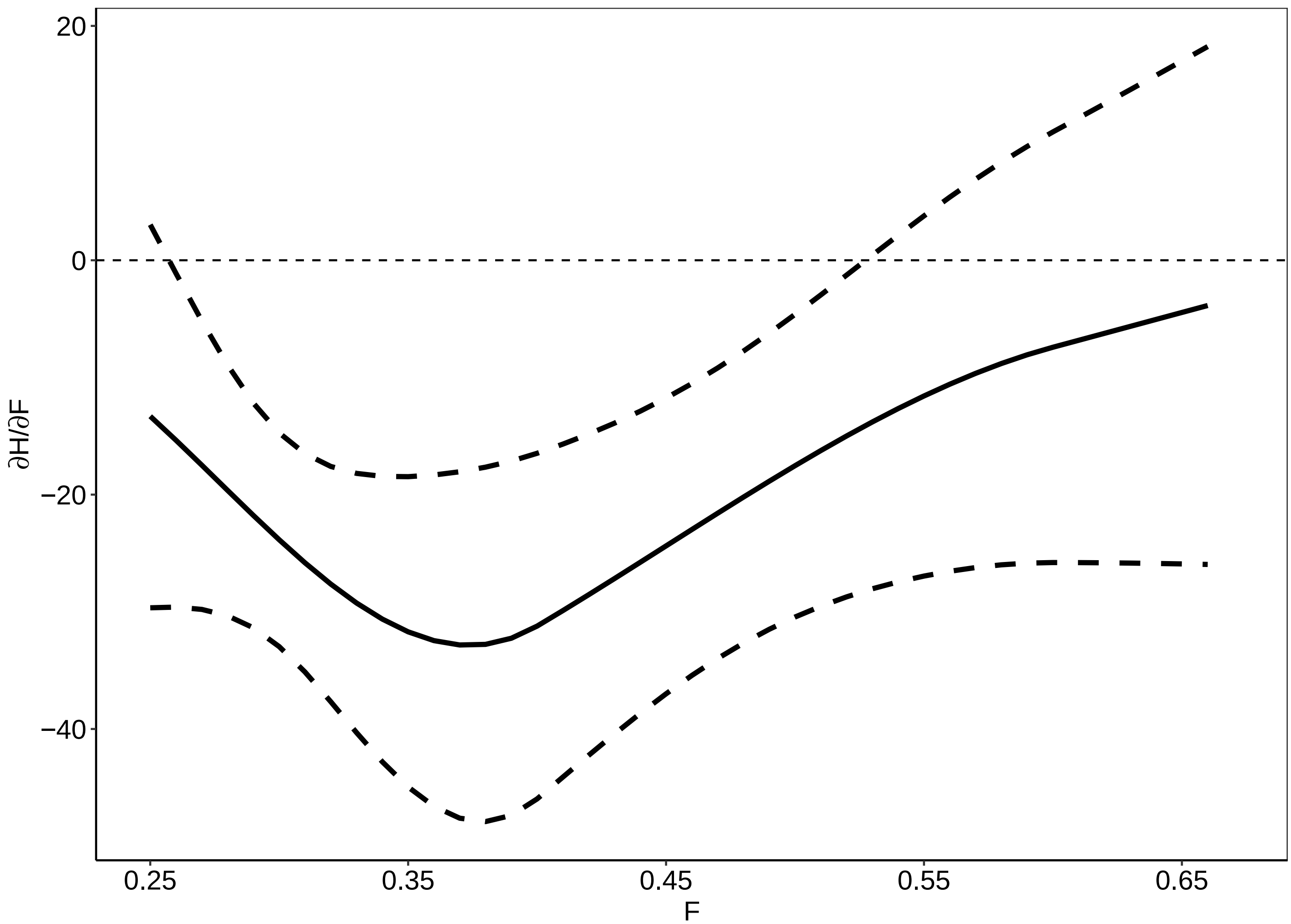
Notes: This figure plots the marginal treatment effect curves for non-workers, part-time, and full-time workers using a 5-knot cubic spline specification. The first stage probit model uses the inverse variance weighted log of the AFDC administrative barriers index and interactions with N. The dashed 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 3: 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 4: 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.

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
P	0.37	1.00	0.00
$\log \hat{W}$	1.78	1.74	1.81
$\log N + 10$	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
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
Northeast	0.28	0.28	0.28
Midwest	0.27	0.27	0.26
West	0.22	0.25	0.20
Observations	3,381	1,238	2,143

Notes: This table reports the means of variables used in our analysis. The sample is composed of single mothers aged 25–55 with a high school education or less with total assets less than $1,500 a week and non-transferable non-labor income less than $1,000 a week drawn from 1988–1992 SIPP interviews. 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

	Barrier Index OLS	AFDC Probit
$\log \hat{W}$	-0.228	-2.564***
	(0.296)	(0.646)
$\log N + 10$	-0.006**	-0.445***
	(0.003)	(0.031)
$\log G$	0.163	0.825***
	(0.108)	(0.154)
$\log \hat{W}(1-t)$	0.460*	0.677*
	(0.257)	(0.371)
Age	-0.003**	0.009
	(0.001)	(0.008)
Black	-0.008	0.142
	(0.012)	(0.090)
Family Size	-0.022	-0.039
	(0.018)	(0.034)
Number of Children < 6	0.004	0.295***
	(0.003)	(0.041)
Food Stamp Guarantee	-0.308	1.873*
	(0.463)	(1.131)

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. "Dem Gov Vote Share" measures the share of the vote the Democratic candidate for governor received in the last election. "Dem Share Under 60%" is an indicator for whether the winning democratic candidate's vote share was under 60%. "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 $\log \hat{W}(1-t)$. 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)

	Barrier Index OLS	AFDC Probit
Northeast	-0.473***	0.471*
	(0.159)	(0.251)
Midwest	-0.325***	0.275
	(0.118)	(0.172)
West	-0.232	0.198
	(0.148)	(0.240)
Elections:		
Dem Gov Vote Share	-0.212	0.517*
	(0.162)	(0.302)
Dem Share Under 60%	-0.066*	0.035
	(0.035)	(0.079)
State Legislature:		
Republican	0.056	0.100
	(0.064)	(0.148)
Split	0.236***	-0.136
	(0.044)	(0.094)
Interactions:		
Republican × Dem Share Under 60%	0.180**	-0.814***
	(0.076)	(0.171)
OLS F-Stat for Instruments		13.40
Pseudo F-Stat by Part. Prob. Range		
0.00–0.25		4.60
0.25–0.66		9.10
0.66–1.00		1.70
Observations	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. “Dem Gov Vote Share” measures the share of the vote the Democratic candidate for governor received in the last election. “Dem Share Under 60%” is an indicator for whether the winning democratic candidate’s vote share was under 60%. “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 $\log \hat{W}(1-t)$. 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

$$g(\hat{F}) = g_1 + g_2\hat{F} + g_3 S3 + g_4 S4 + g_5 S5 \tag{25}$$

where

$$S3 = d_1 - d_4 \tag{26}$$

$$S4 = d_2 - d_4 \tag{27}$$

$$S5 = d_3 - d_4 \tag{28}$$

where

$$d_1 = \frac{\max(0, \hat{F} - F_1) - \max(0, \hat{F} - F_5)}{F_5 - F_1} \tag{29}$$

$$d_2 = \frac{\max(0, \hat{F} - F_2) - \max(0, \hat{F} - F_5)}{F_5 - F_2} \tag{30}$$

$$d_3 = \frac{\max(0, \hat{F} - F_3) - \max(0, \hat{F} - F_5)}{F_5 - F_3} \tag{31}$$

$$d_4 = \frac{\max(0, \hat{F} - F_4) - \max(0, \hat{F} - F_5)}{F_5 - F_4} \tag{32}$$