2009 Presidential Address

Optimal Tax Theory: The Journey from the Negative Income Tax to the Earned Income Tax Credit*

Jonathan H. Hamilton

What lessons do we learn from optimal tax theory for the design of income redistribution programs? I modify a standard model of optimal nonlinear income taxation with discrete types to consider differences in both earning ability and the disutility of effort. This gives a role for “workfare” in the optimal tax policy. The existence of screening mechanisms can play a role in explaining nonparticipation in cash and in-kind redistribution programs, including *Progresa-Oportunidades*, Lifeline Telephone subsidies, and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). Stigma can increase efficiency of a redistribution program by discouraging participation by individuals near the eligibility thresholds. The Family Assistance Program proposed in the early 1970s lacked adequate stigma for nonworkers, which contributed to a lack of political support. In contrast, the current Earned Income Tax Credit (EITC) provides greater benefits to workers than to nonworkers. Thus the EITC does not require any stigma to screen out individuals who do not work from obtaining benefits. Reasons for separate income support programs for nonworkers and for workers are discussed.

JEL Classification: H21, I38

1. Introduction

As a Yale undergraduate in the 1970s, it was nearly impossible not to be aware of the contrasting views of Milton Friedman and James Tobin on macroeconomic policy. At the same time, Professors Friedman and Tobin were among the most prominent economists supporting proposals for a “Negative Income Tax” (NIT). Many graduate students I knew placed great importance on the fact that Friedman and Tobin were in agreement on this issue. “If they agree on this, it must be the right policy” is a fair summary of their views. Because the Yale graduate students did not accept Friedman’s views on monetary policy, it was certainly fascinating to see them hold Friedman’s support for any policy in such high esteem. Then again, Milton Friedman was such a passionate advocate for many policies—the students may just have been overcome with joy to have Milton on their side for once. More likely, the graduate students still shared the public’s common misconception that economists agree on very little. As most of us know, we agree on quite a bit—it is with noneconomists that we agree on fewer issues in the realm of economic policy.

* This was the Presidential Address delivered at the Southern Economic Association Meeting, 22 November 2009, in San Antonio.

I thank many people for sharing thoughts in the development of this essay, including Sarah Hamersma and Katherine Cuff. I also thank Celeste Carruthers and Edward See for most helpful research assistance. I thank the Mr. and Mrs. Wendell P. Ayres Business Enrichment Endowment for financial support of this research.
Not only was Milton Friedman a passionate advocate, in many cases, he was a successful one. Rent control policies now meet considerable skepticism. School vouchers find support from many different groups. In contrast, the NIT is not an ideal toward which many continue to strive. I wish to explore what we’ve learned in economic theory that might explain why society has not come to share Friedman and Tobin’s views on this issue.

First, we need to recall the alternatives at the time. The Aid to Families with Dependent Children (AFDC) program in the 1950s made no provisions for two-parent households. When a recipient entered the labor force, she faced a 100% marginal tax rate on earnings because benefits were reduced dollar-for-dollar with wages.\(^1\)

Tobin (1965) argued that welfare gave recipients incentive to withdraw from the labor force and gave fathers incentives to desert their families. Tobin, Pechman, and Mieszkowski (1967, p. 1) also found the “numerous indignities by the procedures employed to enforce the means test…” an objectionable feature of the existing system. They also sought a system of uniform benefits across the nation (states chose benefit levels at the time).

Friedman’s support for the NIT had other motivations. First, he pushed for a program to replace all existing programs (including Social Security and many other forms of income maintenance). Consistent with his passion for liberty, he sought a system that “treats [the] indigent as responsible individuals, not incompetent wards of the state” (Friedman 1968, p. 211). Furthermore, he argued that “if social workers are hired on government funds they should devote their energies to helping the indigent, not spying on them” (p. 205).\(^2\)

During the Nixon Administration, the Family Assistance Plan was actively considered in welfare reform discussions. At the same time, the NIT experiments were conducted to determine the incentive effects and costs of these proposals.

In 1975, the Earned Income Tax Credit was introduced. Initially, the benefit was quite small, and only a small number of individuals received it. By 2000, after several expansions of the program, it supported 55 million people (Moffitt 2003).

In 1996, President Clinton and a Republican Congress agreed on a major welfare reform bill, which included work requirements (although some waivers and exceptions do exist). With lifetime limits on benefits, this really did “end welfare as we knew it.”

What I will discuss here is what we have learned in economic theory about income redistribution since Friedman and Tobin, some policy innovations in the U.S. and elsewhere, and I hope to shed light on how and why the public and policy makers have come to regard the NIT as not ideal.

2. The Optimal Income Tax

The 1996 Nobel Memorial Prize in Economics was unique in one way—the two men who shared the prize were not coauthors or scholars pursuing different lines of inquiry. William

---

1 In 1967, the AFDC program moved in the direction of a negative income tax, with a reduction in the marginal tax rate from 100% to 67%, although categorical eligibility requirements continued.

2 Tobin (1970) also discusses the merits of specific egalitarianism as an approach to redistribution, guaranteeing consumption levels of particular commodities. He discusses how support for such an approach is stronger among noneconomists than economists; although, he argues that there is a role for it. Friedman argues strongly against such approaches (Friedman and Friedman 1980, pp. 115–9).
Vickery in 1945 posed a problem, and James Mirrlees solved it in 1971 (Vickery 1945; Mirrlees 1971). The essential insight is that the optimal income distribution question is one of asymmetric information. Ideally, one’s tax payment depends on one’s earning ability. However, the tax authority only observes earned income, and earnings depend on choices that the tax schedule affects. Mirrlees’s breakthrough is often called the Revelation Principle. The tax authority does not know each person’s ability, but it develops a tax schedule for which each person would truthfully reveal his or her type.

With a continuous ability distribution (as Vickery and Mirrlees considered), the mathematics constrain us to consider individuals who differ only in ability but share a common utility function. Stiglitz (1982) introduced the discrete ability distribution version, and Brito et al. (1990) relaxed many of the restrictive assumptions.

Let us start with a simple model. Two types of individuals differ in earning ability (wages), with \( w_1 < w_2 \). There are \( N_1 \) type 1s and \( N_2 \) type 2s. Initially assume that both types have a common utility function with consumption and labor supply as the arguments, \( U(C, L) \). The tax authority only observes income, so it helps to rewrite the utility functions in terms of consumption and earned income, or

\[
V^i(C, Y) \equiv U\left(C, \frac{Y}{w_i}\right).
\]

Because wages differ, the utility functions differ. Stiglitz (1982) discusses the usefulness of considering the information-constrained Pareto frontier, so let us write our problem as follows:

\[
\begin{align*}
\text{Max} & \quad V^1(C_1, Y_1) \\
\text{s.t.} & \quad V^2(C_2, Y_2) \geq \bar{v}^2 \\
& \quad N_1(Y_1 - C_1) + N_2(Y_2 - C_2) \geq 0 \\
& \quad V^2(C_2, Y_2) \geq V^2(C_1, Y_1) \\
& \quad V^1(C_1, Y_1) \geq V^1(C_2, Y_2).
\end{align*}
\]

The first constraint is the Pareto constraint, and the second is the revenue constraint. These last two constraints are the self-selection (or incentive compatibility) constraints—each individual prefers the bundle intended for her to the bundle intended for the other type.

When the tax authority redistributes toward the low-ability type, the solution has the following properties (see Figure 1):

1. Type 1’s marginal rate of substitution (MRS) between \( C \) and \( Y \) is less than 1.
2. Type 2’s MRS is equal to 1.

Type 2 thus has a marginal tax rate equal to zero, while type 1 has a bundle that is distorted by the tax schedule.\(^3\) All the tax paid by type 2 is effectively a lump-sum tax. In contrast, type 1’s labor supply is less than what he would choose if given the transfer in lump-sum form. The distortion is to prevent type 2 from mimicking type 1. Notice that the role of the tax wedge differs from a linear income tax—the distortion serves to collect money from someone else, not to raise money directly.

\(^3\) The tax schedule to support these choices by both types will have a kink at \((C_1, Y_1)\), so the marginal tax rate is not defined. It is useful to think of MRS\(^1\) as 1’s marginal tax rate.
One problem jumps out—our tax system looks nothing like this. Sadka (1976) shows that the highest ability type faces a zero marginal tax rate in the continuous case. However, Diamond (1998) shows that optimal tax rates may be quite high near the top of the income distribution. There are also results on when the marginal tax rate at the bottom of the earning schedule equals zero, but the conditions are unlikely to hold in practice (Brito and Oakland 1977; Seade 1977).

However, many individual redistributive programs fit the model better. Many programs have income cutoffs for eligibility that may reduce the labor supply of the recipients, while nonrecipients simply support the programs through the normal tax system. Let’s explore further by expanding our model.

With discrete types, we can solve the model in more general settings than with a continuous ability distribution. One result continues to hold—the taxpayer who pays the largest tax faces a zero marginal tax rate because no one will envy his bundle in a constrained efficient allocation. Now relax the assumption that all taxpayers have the same utility function over consumption and labor effort (a two-characteristic model presents many mathematical complexities with continuous distributions, but it is straightforward with discrete distributions).

Assume that there are three consumer types:

Type a has a low wage and a high disutility of labor;
Type b has a high wage and a high disutility of labor; and
Type c has a high wage and a low disutility of labor.

---

See Brito et al. (1990).
To keep things simple, assume that utility is additive in consumption and labor supply:
\[ U^i(C, L) = v(C) - g_i t(L). \]
The function \( t(L) \) is the same for all consumers, but \( g_i \) is a constant that differs across types (so that we have a multiplicative shift on the disutility of hours). We can keep things simple by further assuming that \( w_1 < w_2 \), where \( w_1 \) is a’s wage and \( w_2 \) is b’s and c’s wages, while \( g_1 > g_2 \), where \( g_1 \) multiplies a’s and b’s disutilities of labor and \( g_2 \) multiplies c’s disutility of labor.

Thus, the utility functions of the three types in terms of \( C \) and \( Y \) are as follows:

\[
V^a(C, Y) = v(C) - g_1 t \left( \frac{Y}{W_1} \right),
\]

\[
V^b(C, Y) = v(C) - g_1 t \left( \frac{Y}{W_2} \right),
\]

\[
V^c(C, Y) = v(C) - g_2 t \left( \frac{Y}{W_2} \right).
\]

In some applications, we need to worry whether differences in the disutility of labor arise from disability or taste for leisure. Here, I want to focus on taste for leisure as what certainly distinguishes group b from group c. Figure 2 displays one possibility for the optimal allocation chosen by the tax authority. Note that groups a and b differ only on one dimension, and groups

---

5 See, for example, Cuff (2000) for a discussion of some different implications. Some of the ideas I present here are discussed in Brett (1998), Cuff (2000), and Besley and Coate (1992), although my focus differs.
b and c differ only on one dimension. An allocation where c envies b and b envies a is the normal case—the only way to have c envy a at the optimum is for a and b to get the same bundle. What we can’t tell without more structure is whether type b is taxed or subsidized.

The work disincentive for type a does not discourage type b particularly strongly because of the difference in earning abilities. What would discourage type b individuals from mimicking type a is having to work as much as type a workers, in contrast to having to earn the same amount of money income.

I want to use this model to illustrate a key point. With self-selection constraints, anything that weakens the constraints can lead to a more efficient outcome. The solution in Figure 2 simply uses the conventional approach. What else might be considered? Certainly, it is at least plausible to observe more about benefit recipients than other taxpayers, or to impose additional rules on them. Suppose we consider a different type of “workfare”—anyone who wishes to receive a net benefit must work in a job with a fixed wage, $w_1$. This really has no impact on type a because that is their wage in any job. Now consider the binding self-selection constraints:

$$v(C_b) - g_1t \frac{Y_b}{w_2} \geq v(C_a) - g_1t \frac{Y_a}{w_1} = V^a(C_a, Y_a)$$

$$v(C_c) - g_2t \frac{Y_c}{w_2} \geq v(C_b) - g_2t \frac{Y_b}{w_2}.$$  

The first constraint shows that we can give type b an allocation for which type b is no better off than type a. In the previous solution, type b’s utility is greater than type a’s. As long as $C_b \leq Y_b$ at the optimum, the new self-selection constraint allows the tax authority to transfer more income to type a consumers (and reducing b’s utility makes it possible to collect more tax from type c consumers).

Another way to achieve a similar gain would be to impose an hours requirement on individuals receiving a net benefit. Suppose that the tax authority selects a consumption-``private'' earned income bundle $(C_a, \hat{Y}_a)$ and requires any recipient to work an additional $h_a$ hours. Then, the self-selection constraint for type b consumers with respect to the type a bundle becomes

$$v(C_b) - g_1t \frac{Y_b}{w_2} \geq v(C_a) - g_1t \frac{\hat{Y}_a + h_a}{w_2}.$$  

The work requirement relaxes the self-selection constraint and allows a greater transfer to type a consumers. If type a consumers had a lower disutility of labor than type b consumers ($g_a < g_b$), the work requirement could even be less productive than private sector work by type a.  

The notion that differences in attachment to the labor force among individuals with adequate earnings ability are a major constraint on income support programs is not only a theoretical curiosity. Berman (2000a) discusses the case of ultra-Orthodox Jews in Israel who attend yeshivas (Talmud study schools) until past age 40. The presence of a large group of individuals (approximately 5% of the Israeli population) who choose not to work or

---

6 Assume that the work in these additional hours has the same productivity as type a workers do in the private sector.

Then the budget constraint is as follows: $N_a(Y_a + w_1h_a - C_a) + N_b(Y_b - C_b) + N_c(Y_c - C_c) \geq 0$.

7 Besley and Coate (1992) analyze the design of income redistribution from the perspective of minimizing the cost of benefits. In effect, they study the least cost program to guarantee all individuals a minimum level of consumption. Workfare is completely unproductive in their model, but the planner ignores the utility cost of work.
accumulate significant market human capital has become a source of tension in Israeli politics. Berman (2000b) argues persuasively that ultra-Orthodox yeshiva attendance in Israel is a result of the Israeli social welfare system. Ultra-Orthodox communities in other countries do not exhibit such extended yeshiva attendance nor such low rates of labor force participation among prime-age males. The subsidies available in Israel induce prolonged yeshiva attendance as a device to display commitment to the religious community (which supplements government financial support). Berman (2000a) calls this “subsidized sacrifice” and suggests that welfare (and draft deferment) reform in Israel could reduce the deadweight loss of the signaling sacrifice of the ultra-Orthodox.

3. Participation in Redistribution Programs

I now want to use the self-selection model to think about some income redistribution programs. To do so, I will move away from the formal model of the previous section, keeping in mind the idea of relaxing the self-selection constraints of those who would choose to mimic low-ability workers. To quote from last year’s Distinguished Guest Lecture by Peter Diamond on how to use theory:

In particular, I am concerned that too many economists take the findings of individual studies literally as a basis for policy thinking, rather than seeking inferences from an individual study to be combined with inferences from other studies that consider other aspects of a policy question, as well as with intuitions about aspects of policy that are not in the models. To me, taking a model literally is not taking the model seriously. (Diamond 2009, p. 2)

One issue I wish to explore is nonparticipation in redistributive programs. A concern among many policy experts is that many apparently eligible individuals do not enroll in programs that would give them cash or in-kind assistance. Table 1 summarizes some findings in the literature.

An early contribution toward the theory of nonparticipation is Moffitt’s (1983, p. 1023) model of “welfare stigma.” His definition was “disutility arising from participating in a welfare

<table>
<thead>
<tr>
<th>Program</th>
<th>Estimated Take-Up Rate</th>
<th>Reasons</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplemental Security</td>
<td>&lt; 60% of poor elderly</td>
<td>Information (?)</td>
<td>a</td>
</tr>
<tr>
<td>Income (SSI)</td>
<td>Lower for nonelderly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earned Income Tax</td>
<td>&gt; 80% of eligibles</td>
<td>Low cost to apply</td>
<td>b</td>
</tr>
<tr>
<td>Credit (EITC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFDC</td>
<td>65% of eligible single</td>
<td>Long-term need drives take-up</td>
<td>c</td>
</tr>
<tr>
<td>mothers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food stamps</td>
<td>70% of eligibles</td>
<td>Transactions costs, stigma</td>
<td>d</td>
</tr>
<tr>
<td>WIC</td>
<td>75% of eligibles</td>
<td>Transactions costs, value</td>
<td>d</td>
</tr>
<tr>
<td>Lifeline Telephone</td>
<td>3.3%—&gt;100%</td>
<td>Information, quality (?)</td>
<td>e</td>
</tr>
<tr>
<td>Service</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Burkhauser and Daly (2003).
* Scholz (1994).
* Blank and Ruggles (1996).
* Currie (2003).
* Hauge, Jamison, and Jewell (2008).
program per se.” Hoynes (1996) broadens the concept to include transactions cost of participation (to some extent, one cannot disentangle them in studying participation).

Thinking back to my model raises the possibility that a program designer might find stigma desirable. Anything that discourages participation by high-wage individuals has the potential to reduce the budget costs of an income support program. The “best” stigma would be one that affects the potential mimickers but not the program participants—requiring work in a low-wage job to receive the benefit might be gold-medal stigma. Moffitt modeled two components of stigma—a flat component (independent of the size of the benefit) and a variable component (which increases with benefit size). His empirical results support only the flat component. To discourage participation by the marginally eligible, flat stigma is useful, while variable stigma is not. Moffitt did not directly consider stigma as a function of income (eligibility), but stigma that increases with income would relax self-selection constraints while imposing lesser costs on beneficiaries. Can program designers develop efficiency-enhancing stigma?

I now wish to discuss some possible interpretations of nonparticipation in several programs.

The Progresa-Oportunidades Programs

Researchers at the World Bank have taken great interest in a new form of antipoverty programs, which they call “conditional cash transfers” (CCTs). Mexico has the most fully developed (and analyzed) program, Progresa (now renamed Oportunidades). In 1997, the first experiments were launched as Mexico sought a way out of its economic crisis while maintaining support for its poorest citizens. One element of the program is quite close to Milton Friedman’s desired approach—replacing a large number of programs with one that better targets the poor. Previously, Mexico had an extensive program of food subsidies that did a poor job of focusing on the poor, especially the rural poor (Levy 2006).

Oportunidades is in many ways a remarkable program. First, there was extensive randomization in the geographic rollout of the program to make evaluation easier. Household eligibility for program benefits did not depend directly on income, and eligibles are guaranteed benefits for three years (Levy 2006, p. 22). Thus, work disincentives of the program are minimized. Households attending a health clinic regularly receive cash and in-kind benefits, including nutritional supplements. Clinic visits include preventive health care, treatment, and health education. Households receive additional cash transfers for school attendance (which increase with each school year, and are larger for girls than for boys) and for completion of high school. A household can receive between $15 (US) and $153 per month through the program, and the mean is about $45 (2005 data from Levy 2006, pp. 24 and 32).

Many economists have studied Oportunidades, with a primary focus on how the conditional cash transfers influence school enrollment, nutrition, and health care and status. Overall, the program appears to be a success in improving key indicators of well-being for poor households.

Oportunidades was introduced in urban areas after its spread across rural areas. The take-up rate fell from 97% in rural areas to about 50% in urban areas (Angelucci and Attanasio 2009). Some of the difference arises from how the programs were introduced in urban areas—potential beneficiaries had to apply to learn their eligibility. Among eligibles, who participated
in urban areas? Households with children in school before the program started were more likely to participate; households with teenage children in the work force were less likely to participate. The authors state “results indicate that program participation is much higher among families whose kids would have gone to school irrespective of the program” (p. 497). They conjecture that the payments for school attendance were too small relative to opportunity cost in urban areas. Another interpretation is that Oportunidades incorporates an effective screening device and that this is a sign of success. If some eligible choose not to participate, program savings result. Because children are affected, we might worry. However, presumably parents who don’t send kids to school with a small subsidy wouldn’t send kids to school with an equivalent (or a smaller) lump sum grant! Hence, the money spent in the program flows to the parents as a reward for good behavior.

Morley (2009, in a review of Levy’s book) expresses skepticism on the value of CCTs as a screening mechanism. He describes the urban Progresa program as “more like a straight transfer program for a subset of the poverty population, rather than a dual-purpose program” (to increase human capital and to reduce poverty). From a screening perspective, reducing poverty among families who are investing in human capital (the intersection, not the union) could be an appropriate goal.

Some researchers on CCTs view their success depending on whether the payments induce behavior changes by a large fraction of households.8 Households who enroll children in school only with large CCTs may not be households from whom the return to schooling will be particularly high—their behavior may arise from a lack of interest in pursuing and supporting their children’s education. Smaller payments that only some families qualify for might contribute more to human capital investment—not directly from school enrollment, but from relieving financial pressures for families with school children.

New York City Mayor Michael Bloomberg has introduced a pilot program of conditional cash transfers in New York City. Opportunity NYC includes education rewards, health rewards, and work and training rewards that can amount to $4000 to $6000 in a year. Unlike Mexico’s program where the central goal is to move children from work to school, this program offers incentives for school attendance, passing scores on tests, and parent participation in the student’s education. High school students share rewards with parents. The program randomizes participation, but unfortunately has only one set of rewards. Researchers will not learn how large the rewards need to be to induce improved behavior.

**Lifeline Telephone Subsidies**

In comparison to the relatively large payments that households can receive in the Oportunidades program, the lifeline telephone subsidy is quite small. A household can receive a subsidy of up to $13.50 per month (in 2005) for local telephone service (Hauge, Jamison, and Jewell 2008). Holt and Jamison (2007) report that 88% of low-income households subscribe to local telephone service, but that only one-third receive the Lifeline subsidy. This low take-up rate has led to considerable research and policy discussions about modifying the plan.9

It seems clear that something is awry in this program—it has a low take-up rate and does not seem to have much effect on increasing the number of subscribers. Keep in mind—the low

---

8 See, for example, Das, Do, and Özler (2004).
9 See papers cited in Holt and Jamison (2007) and Hauge, Jamison, and Jewell (2008).
participation is for the subsidy, not the phone service that is being subsidized. Several authors raise the possibility of “stigma,” but Hauge, Jamison, and Jewell (2008, p. 138) point out that one can claim the benefit easily and without disclosing it to others (Lifeline telephone numbers are not identifiable, for example). A lack of information about the benefit seems to predominate as the best explanation for low take-up.

Some screening also occurs. Low-income households are more likely to have only a landline or only cell phone service than higher-income individuals. Only some eligible households can use the Lifeline subsidy for cell phone service. In some states, only single-line residential service qualifies for the subsidy, thus excluding households purchasing multiple lines or premium services. Burton, Macher, and Mayo (2007) find that such restrictions lower participation rates. Limiting the subsidy in this way avoids subsidizing those most likely to buy without the subsidy.

**WIC**

The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) provides vouchers to buy particular categories of food (milk, juice, etc.) as a nutrition assistance program. Households with incomes below 185% of the poverty line are eligible when the mother is pregnant or nursing or when a child under age 5 is present (Currie 2003). For a married couple with two children, the cutoff is $38,203 in 2008 (with four children, the cutoff is $51,079). The U.S. Department of Agriculture estimates that 75% of eligible individuals participate in the program. The participation rate appears to be higher among the more disadvantaged. At first glance, this seems to be a desirable outcome. Some households near the maximum income thresholds may benefit greatly from the program; for others, the program benefit may be considerably less than the cash value of the food products. WIC imposes several conditions on the purchase of the groceries—buying the least expensive brand of milk and eggs, buying from a list of 33 ready-to-eat breakfast cereals (in Florida in 2009), buying canned light tuna in place of white or albacore. A mother cannot replace low-quality items with higher-quality ones, so some parts of the voucher may go unused. The transactions costs and the low value of the benefit may be screening moderate-income eligibles from participating.

It is useful to keep in mind that income as measured for WIC eligibility (or any other program) will only be a proxy for “permanent income” (or another measure of ability to pay). First, there is the issue of earnings distortion in order to qualify for WIC. Second, a household’s earnings may be temporarily low, and thus a household with high lifetime resources might be eligible, even if not particularly deserving. If we observe a drop in take-up for households closer to the eligibility threshold, that may not be a flaw in program management. Unfortunately, we may have a difficult time determining whether the

---

10 The National School Lunch Program has the same income cutoffs.
12 While I am unaware of papers searching for such effects in the context of WIC eligibility, Ham and Shore-Sheppard (2005) analyze the question of whether households limit hours worked in order to qualify for Medicaid. They do not find substantial evidence of hours reductions. The WIC limits coincide with the school lunch program income limits, so a mother using infant formula who also has a school-age child could qualify for a large subsidy by staying under the earnings limit.
13 I am thinking of a graduate student household, for example.
nonparticipants are households for whom the social value of their participation is lower than for participants with similar measured incomes.

For an in-kind program such as WIC, with benefits having a cash value of $35 per month (in 1998, excluding infant formula; Currie 2003), it seems likely that the coupons influence consumption more for low-income than for moderate-income households. Moderate-income households who do not anticipate any benefit from the nutrition education component of WIC would seem to be the more likely nonparticipants.

One way to increase the take-up rate for WIC would be to lower the income threshold for eligibility. Such a reform would screen out the nonparticipants and some moderate-income households that the program would hope to reach. The higher threshold allows the latter group to participate, but then screening out some households becomes desirable. Note that there are two gains from discouraging participation in WIC—the program saves resources and fewer households have an incentive to reduce earnings to qualify for the benefit.

The open question is whether the nonparticipants are the households for whom the social value of participation is the lowest. I will return to this in the conclusion.

4. The Negative Income Tax and the Earned Income Tax Credit

Milton Friedman’s pure negative income tax (replacing all income redistribution programs) never became a legislative proposal. In 1969, President Richard Nixon proposed the Family Assistance Plan (FAP), which had some properties of a guaranteed income plan. The guaranteed income is essentially the base amount in a negative income tax—the benefit for someone who earns zero. The House of Representatives passed Nixon’s FAP, but it died in the Senate.14 Along the way, the proposal suffered cruel treatment from both those who favored more income redistribution and those opposed. Different sides saw different things in the FAP—and most saw things they didn’t like. To some opponents, a guaranteed income stood for “the proposition that people ought not to have to work for a living” (Moynihan 1973, p. 10). On the other hand, a Yugoslav Marxist said “it would finally free the individual and his family from the myriad and inescapable forms of coercion which society exerts through the employment nexus” (Moynihan 1973, p. 4). I would not relish serving on a committee with representatives from these two groups!

The “guaranteed income” notion raised many hackles. One senses that the problem was too little stigma. In this period, Hawaii established a residency requirement for its general assistance program to discourage “welfare hippies” (to use the lieutenant governor’s words) or “parasites in paradise” (from a state senator) (Moynihan 1973, p. 34).

The FAP did not eliminate existing programs and sometimes tried to grandfather provisions to keep people from losing benefits in the reform. As a consequence, some income ranges faced very high marginal tax rates (sometimes over 100%). That these problems were pervasive in the existing system was lost in the debate. It also established a work requirement but weakened the effect by only cutting an individual’s benefit (and not the benefit for dependents).15 Senator Russell Long described the bill as “paying people not to work.”

14 Moynihan (1973) relates the details of what happened to the proposal.
15 The benefit would fall from $1600 to $1300 for a family of four.
(Moynihan 1973, p. 523). Support eroded further as large disincentives to work over some income ranges were discovered.16

As debate on the FAP continued, an important tension became clear: Either the benefit to a family with zero income must be small, the marginal tax rate on work must be high, or the benefit must extend far up the income distribution. Clearly this last alternative makes the program expensive; although many households will only receive a small benefit (and one that is smaller than their regular income tax liability). Some proposed to make the base benefit as high as $5500 in 1970 dollars (or approximately $29,900 in 2008 dollars17). Raising the marginal tax rate from 50% to 67% came under consideration. No amendments would please enough of the Senate.

In the 1972 Presidential campaign, George McGovern proposed a demogrant of $1000 for every American. The details were unclear on what increases in income tax rates would pay for it.18 Opponents focused on the benefit one could obtain with zero earnings, so it became another failure for a guaranteed income program.

What emerged from the FAP debate was far from the original proposal. Russell Long proposed a “work bonus,” essentially an EITC. The original goal was to reduce the impact of the payroll tax on low-income workers with children. In 1978, the EITC was made permanent with a 10% credit up to $5000 of earnings (a maximum benefit of $500). The credit was flat between $5000 and $6000 of earnings, then phased out at a 12.5% rate (back to zero at $10,000). In 1986, the real value of the credit was returned to its 1978 level. In 1993, a smaller credit (only 7.65%, the FICA payroll tax rate) became available to childless households. Since 1996, the tax credit has remained constant in real terms with a 40% credit rate and 21.06% phaseout rate for households with two or more children (30% and 15.98%, respectively, for households with one child) (Ventry 2001). For 2008, the maximum benefit for a couple with two children is $4824; they can earn over $29,000 before income and payroll taxes exceed the EITC payment.19 Note that this is close to what two full-time minimum wage workers would earn in a year.

The legislation expanding the EITC preceded the final debates on AFDC reform (the Personal Responsibility and Work Opportunity Act passed in 1996), but it seems to have followed from the movement toward “making work pay.” The EITC is also separate legislation from any benefits for nonworkers. It thus avoids mixing income support for working poor with “welfare.”20 Steensland (2006, pp. 1286–87) observes:

As a comparison between views of Nixon’s plan and the Earned Income Tax Credit reveals, business leaders and their conservative political allies feared stigmatizing the working poor with “welfare” benefits and the associated repercussions this stigma might have on labor market processes and the national economy. Payments understood as “income supplements” in a separate program did not connote this threat. Just as significantly, the working poor themselves did not lobby on behalf of GAI [Guaranteed Annual Income] proposals, despite the fact that the plans were in their material interest. These members of the working class did not consider themselves to be “welfare” recipients.21

\[^{16}\text{The FAP would interact with existing programs differently across states with different levels of existing benefits.}\]
\[^{17}\text{(211/38.8) \times 5500 = 29,900, using the January 2008 price level.}\]
\[^{18}\text{Leontief (1972) suggests a four-person family’s benefit would phase out at$12,000, which implies a tax rate of 33\% (or higher if the tax rate were zero for low levels of earnings, as he suggests).}\]
\[^{19}\text{We solve for the income level } Y \text{ at which tax liability equals EITC payment or } 0.1(\hat{Y} - 24,900) + 0.0765Y = 4824 - 0.2106(\hat{Y} - 18,740).\]
\[^{20}\text{TANF recipients (Temporary Assistance to Needy Families) are eligible for the EITC, but EITC benefits depend only on earnings and not other benefits one might obtain.}\]
\[^{21}\text{Steensland (2006, p. 1308) also discusses how Supplemental Security Income (created soon after the failure of the FAP) was intended to be devoid of stigma by its designation for certain beneficiary categories.}\]
The EITC has become the income support program that everyone likes. Why? By rewarding work, it avoids transfers to the “parasites.” It operates through the income tax filing procedure, so that it has low administrative costs. Thus, the EITC has little or no stigma.\textsuperscript{22} Nonparticipation is then a puzzle; although, some of that can be attributed to a lack of information by potential beneficiaries.\textsuperscript{23}

However, let us step back for a minute. For a two-parent two-child family, for incomes in the phase-out range, the marginal tax rate is $21.06 + 10 + 7.65 = 37.71\%$ (and then 42.71\%). If one parent works full time at only a bit more than the minimum wage, all of the second parent’s earnings are taxed at these relatively high rates (and for some purposes, we would add the employer’s 7.65\% as well). For these households, the relevant part of the budget set is hard to distinguish from an NIT budget set. All of the difference comes at lower income levels where the credit (a negative marginal rate) and then a zero marginal rate apply. It is not hard to write down an EITC that is identical to an NIT in the EITC phase-out range (compare Figures 3 and 4).

The EITC avoids the NIT problem of paying individuals who choose to earn nothing. The maximum dollar credit hits before full-time work for an individual. Perhaps we should not be surprised that the evidence is that the EITC has only a small disincentive effect on the intensive margin (hours) but a strong and positive effect on the extensive margin (working vs. not working).\textsuperscript{24}

Those who are unable to work rely on other programs, removing stigma from the EITC and adding to its political support. These other programs are a mix of high-stigma (think of TANF) and low-stigma programs (think of Supplemental Security Income). It is noteworthy that TANF benefits phase out before a household reaches the maximum benefit under the EITC.\textsuperscript{25} The work incentives of the EITC reach households who earn more than TANF beneficiaries. Steensland’s (2006) analysis suggests that separating the programs serves to increase political support (even though some households obtain both benefits).

Another perspective is that the EITC effectively screens out those who would not work under an NIT benefit program. Think back to the ultra-Orthodox in Israel—spending fulltime in religious studies would let them receive a full NIT benefit, while they would get nothing from an EITC. The negative marginal tax rate portion of the EITC budget line is not where most of the beneficiaries fall.\textsuperscript{26} The credit effectively targets benefits to workers and not to shirkers.

\textsuperscript{22} Individuals can apply for an advance EITC payment via negative withholding, which requires filing a W-4 with an employer. Perhaps the lack of reliance on this is due to avoidance of stigma. Romich and Weisner (2001) discuss how the low rates of advance payment raise the possibility of employer resistance.

\textsuperscript{23} Because the program does not have its own administrative bureaucracy, there are no individuals working with potential participants to explain the program. There were also problems with ineligibles claiming the benefit (often due to misunderstandings about which parent could claim the benefit). There are many reasons to worry whether individuals fully understand the budget set.

\textsuperscript{24} See Hotz, Mullin, and Scholz (2006).

\textsuperscript{25} In Florida, the $303 monthly payment is reduced by 50\% of earnings in excess of $200 per month, so that the benefit falls to zero at $805 of monthly earnings (see Rowe, Murphy, and Kaminski 2008). The EITC provides additional cash to workers receiving TANF (and reduces the implicit tax rate).

\textsuperscript{26} Liebman (2001) reports that 75\% of benefits in 1995 went to workers who earned more than the maximum credit income. Meyer and Sullivan (2009) find that poverty rates among households with children fell from 1990 to 2005, while almost no change occurred for single individuals or married childless couples. They give credit to the EITC expansion.
5. Conclusions and Further Thoughts

Have we found a way to separate the “deserving poor” from the “undeserving poor”?\textsuperscript{27} Such a division would seem an anathema to Friedman’s libertarian beliefs. I do wonder what he thought of Oportunidades with its incentives. Elimination of all the categorical income support programs seems unlikely to happen anytime soon. The EITC also seems to have staying power—supporting working families with a stable tax credit garners support from a broad ideological spectrum. It is an alternative to a higher minimum wage. It weakens the pull of the welfare trap and provides support to those who “play by the rules.”

A pure NIT is vulnerable if too many beneficiaries choose not to work. Such a tax structure does not screen out those who could support themselves but choose not to work. Members of Congress who objected to the FAP may have grasped the optimal tax problem better than many economists at the time.

Let me close by noting one unanswered question—it may be a hard one to resolve. After observing nonparticipation in a program, the question remains whether the screening is excluding the right group (those with the lower benefit–cost ratios, to measure it in one way). In the WIC example, individuals with a small private benefit of participation were less likely to participate. Are those individuals with a low social value of participation? When we consider

\textsuperscript{27} In George Bernard Shaw’s *Pygmalion* (1983), the “undeserving poor” do not “eat less hearty” and “drink a lot more” than the deserving.
additional ways to screen, this question becomes harder to answer. Because unobservable characteristics often drive screening, identifying the differences between participants and nonparticipants may require learning more about eligibles than do the program administrators.

References


Blank, Rebecca, and Patricia Ruggles. When do women use AFDC and food stamps? The dynamics of eligibility vs. participation. *Journal of Human Resources*, 31:57–89.


