Welfare Work Requirements

with Paternalistic Government Preferences

Robert Moffitt
Johns Hopkins University and NBER

June, 2005
Revised, March, 2006

This manuscript is a revised version of paper delivered at the Conference on Social Interactions in Profiling, Northwestern University, October 29-30, 2004. The author would like to thank the participants at that conference, particularly John Weymark, for comments. Comments from participants at workshops at Brown University, Johns Hopkins University, William and Mary, and the Institute for Research on Poverty are appreciated. The correction of an error by Michael Keane is also gratefully acknowledged.

cmwq_v9.wpd
3/22/06

Abstract

Work requirements in means-tested transfer programs have grown in importance in the U.S. and in some other countries. The theoretical literature which considers their possible optimality generally operates within a traditional welfarist framework where some function of the utility of the poor is maximized. Here it is argued that society instead has preferences over the actual work allocations of welfare recipients and that the resulting paternalistic social welfare function is more consistent with the historical evidence. With this social welfare function, optimality of work requirements is possible but depends on the accuracy of the screening mechanism which assigns work requirements to some benefit recipients and not others. Numerical simulations show that the accuracy must be high for such optimality to occur. The simulations also show that earnings subsidies can be justified with the type of social welfare function used here.

Work requirements in means-tested (or "welfare") programs have seen a resurgence in the last twenty years in the U.S. and in a few European countries. These requirements generally stipulate that benefit recipients must work or engage in some work-like activity for some minimum number of hours each week. Most economists who have studied the welfare system have opposed work requirements. Milton Friedman, for example, believed that welfare should be administered through a negative income tax with benefits based solely on income. Other economists have criticized work requirements for being administratively unworkable, for introducing discretion into the system on the part of caseworkers, and for generating incentives on the part of individuals to avoid the requirements (Barth and Greenberg, 1971; Browning, 1975; Lurie, 1975). In addition, the classic optimal tax framework of Mirrlees (1971), which generates a negative income tax as an optimal solution in a utilitarian maximization model, allows only the manipulation of guarantees and tax rates and permits no other policy instrument to be used.

However, there is a body of research papers operating within the optimal tax framework which does consider the possibility that work requirements might be part of a solution. Akerlof (1978) noted that if those with low job skills who have extreme difficulty in becoming employed--that is, those who are deemed, by some definition, to be unable to work--can be partly identified by some measurable, observable characteristic (or "tagged"), then it is possible that social welfare could be improved if those who are untagged, and hence able to work, are denied support because the tagged individuals can then be paid higher benefits. Diamond and Sheshinski

(1995), Parsons (1996), and Salanié (2002) have similar models which broaden the set of assumptions and applications of the general idea of using an imperfect observable indicator of need to separate the population. There is also a literature which assumes no tagging (i.e., no observable characteristic to differentiate individuals) but which finds work requirements to be optimal only under restricted conditions in that case (Brett (1998), Beaudry and Blackorby (2004), Cuff (2000), Fortin et al. (1993), Besley and Coate (1992,1995); see Kaplow (2004) for further discussion). Of these, only Besley and Coate are able to provide a strong case for work requirements, and only then by assuming a nonwelfarist social welfare function which an income maintenance criterion which puts little value on leisure.

This study provides a new model of optimal work requirements. The primary departure from the previous literature is to use a nonwelfarist social welfare function which represents society as caring directly about the level of work of the poor, not just their utility. Unlike Besley and Coate, the major previous model with a nonwelfarist social welfare function, it is assumed that work, rather than income, enters the social welfare function directly. Society is assumed to value of the leisure time of some families, but only if their true skill (=ability) is low; if their true skill is sufficiently high, society expects them to work. The assumption that work per se is valued is based upon observation of the history of redistributional policy in the U.S. over the last thirty years, which, it is argued here and has been argued elsewhere (Moffitt, 2002a, 2003a), makes clear that voters do not accept the preferences of the poor but instead wish to force onto them specific allocations toward food, medical expenditure, housing, and other goods. It is commonly observed that this attitude is responsible for the rise in in-kind transfer programs for food, medical expenditure, and housing, for example, and for the decline in pure cash programs

which allow recipients to spend their transfers according to their own preferences. A similar paternalistic attitude can be argued to be responsible for the rise in the desire for society to have recipients work even if though those recipients could achieve an equal, or higher, level of utility with less work and the same governmental expenditure.

With this change of objective function, work requirements follow directly as a possible instrument for improving social welfare. In fact, work requirements may be optimal even if ability is completely observed. But in the more realistic case of partially unobserved ability, work requirements become a policy instrument that is more likely to be used within a standard screening model if the screening mechanism is accurate enough. The paper lays out several models of this type. Numerical simulations are presented which show the utility gains and losses from each type of policy under different assumptions about the accuracy of the screening mechanism. An important feature of the simulations, not always a focus of attention in prior work, is the distribution of utility gains and losses across those of high and low ability resulting from each scenario.

A by-product of the model and the simulations is a new justification for earnings subsidies and for the Earned Income Tax Credit (EITC) in the U.S. While earnings subsidies are never optimal in the standard Mirrlees framework, they can be optimal with the relaxation of some of the assumptions in that framework, either the assumption that labor supply decisions are at the intensive margin or that home productivity is high (Saez, 2002; Beaudry and Blackorby, 2004; Choné and Laroque, 2005; Laroque, 2005). Earnings subsidies are often optimal here, instead, because work is included directly in the social welfare function.

The next section provides institutional background on work requirements in the U.S.

welfare system and more discussion of the evidence in favor of paternalism. The following section presents several models of paternalistic social welfare preferences, and the subsequent section presents the results of the simulations. A summary and conclusion ends the paper.

Institutional Background

Proposals for various types of work requirements in welfare have been growing in frequency over the last thirty years in the U.S. but, until the 1990s, were not enforced (Moffitt, 2003b).[1] However, Congressional legislation in 1996 required work requirements of all states operating cash programs for single mothers, stipulating that at least 50 percent of single-parent recipients work or be in work-related activities. A minimum of 20 hours of work per week were required and exemptions were significantly narrowed. By 2000, about one-third of TANF recipients were in work or in a work-related activity (U.S. DHHS, 2002), an historic high. The impact of work requirements was made possible by the extensive use of sanctions, or benefit reductions for failure to comply with the requirements. Nationwide, about 5 percent of welfare recipients are sanctioned each month (Bloom and Winstead, 2002). Thus, for the first time in the history of the program, the work requirements were fairly vigorously enforced.

All states allow some exemptions from work requirements. Most often exemptions are granted if the welfare recipient has a very young child or is disabled or in poor health, and sometimes if the recipients is caring for an older family member or lacks transportation to get to

[1] Work requirements have a long history in welfare reform. How and why they were used in the 1834 English Poor Law has generated a particularly large literature (e.g., Boyer, 1990).

work. However, while official exemptions exist in the law in each state, local offices and individual caseworkers exercise considerable discretion in appying the exemptions and therefore in deciding who should be required to work (Diller, 2000, Lurie, 2001), as well as discretion in choosing when to apply sanctions (Bloom and Winstead, 2002 and Pavetti and Bloom, 2001). There is very systematic and statistically-based empirical evidence on the extent of this discretion, but there is a large literature documenting its existence with case studies.

The increasing use of work requirements is prima facie evidence that voters care directly about work, for if it were merely a matter of voters becoming more conservative about redistribution, they could have instead simply reduced benefits. The most articulate spokesman for work and other requirements is Mead (1986,1997), who argues that the welfare system should be used to change the behavior of recipients in a way that will improve their well-being even though they do not realize it. Indeed, the debate over the 1990s welfare reform was widely characterized at the time as concerning not the efficiency of the system or the amount of redistribution, but rather about "values," namely, the importance of imposing society-wide values governing regarding work behavior, self-betterment, and marriage and childbearing decisions onto the poor. This paternalistic attitude is consistent with the long-run trend away from cash transfers and toward in-kind transfers such as those for food, housing, and medical care (Moffitt, 2002a, 2003a), reflecting a desire on the part of voters to specify the types of goods consumed by the poor and, implicitly, a desire to override the preferences of the poor themselves.

\ An interesting question, but one beyond the scope of this paper, is why society values work so highly. While it is not difficult to understand voters' desire to see poor families have minimum amounts of food, clothing, and housing, it is not obvious why voters would think that

work per se has intrinsic value to a poor person. One explanation is that the poor are lazy, as in Cuff (2000), and the model below could be interpreted as an alternative formulation of that notion. It has also often been suggested that work may lead to human capital investment and lead to a lower likelihood of being poor and on welfare in the future, although it not clear why subsidies to formal education and training are explicitly prohibited under the 1996 law. It has also been suggested that work will improve the self-esteem of the poor; that their utility should increase if their income comes from earnings rather than welfare benefits per se; and that working and being off welfare (holding income fixed) will have beneficial effects on children. None of these explanations suggest why work requirements have become more popular over time, however, and here the major suggestion that has been made is that the rising rate of female labor force participation in the general society has led to a greater social norm that women with children should work. These issues would be useful for a separate exploration.

Some Paternalistic Models of Welfare Work Requirements

Illustrative Models with Full Welfare Participation. We begin with a simple illustrative model and state a more general model subsequently. Assume we have a poor population with preferences over leisure and income given by U(L,I) where L is a dichotomous indicator for working (L=1) or not working (L=0) and I is income. While preferences are homogeneous, the wage is heterogeneous and is given by W_α, where α has a two-point distribution with α=1 for high ability individuals and α=0 for low ability individuals. In the development below, we will take α to be either completely observed, partially observed, or completely unobserved.

Assuming there is no private income other than earnings, those off welfare have income $I=W_{\alpha}L$. The government pays a welfare benefit G to nonworkers and a benefit B to workers, where B will generally be set according to a negative income tax formula B=G-tW, where t is the tax rate. Income on welfare is $I= (W_{\alpha}+B)L+G(1-L)$. G and B may be set at different levels for those of different ability, if that ability can be fully or partially observed, e.g., G_{α} and B_{α} in the fully observed case. In this initial model, we will assume that $W_1<(G/t)$, which implies that high-ability individuals (and therefore low-ability individuals as well) who work will be eligible for a positive benefit. We will make the model more realistic below by relaxing this assumption, for it has little effect on the main results. We also assume that the high-ability individuals always work in the absence of any transfer program.

The government (or the voters, society, or the welfare agency, all taken as equivalent here) has a preference function over the utilities of the two ability types but also cares whether they work. If p is the fraction of the population of type $\alpha=1$ and if L_{α} is the work level of type α, we assume the government preference function to be

$$V = pL_1U(1,W_1+B) + (1-p)(1-L_0)\,U(0,G) \qquad (1)$$

This function is a polar case where the society desires that high-ability individuals work and that low-ability individuals not work. The government therefore may wish to override individual preferences. The low-ability individuals may consist only of the disabled or those otherwise in

poor health, or may extend up the ability distribution to an arbitrary point.[2] The society obtains no utility from high-ability individuals who do not work or from low-ability individuals who work; this extreme assumption could be relaxed by letting society put lesser weight on the alternative states, without changing the tenor of the results.[3]

In this illustrative model, it will be assumed for simplicity that p=1/2. Hence the government objective function is

$$V = L_1 U(1,W_1+B) + (1-L_0)\, U(0,G) \qquad (2)$$

This object of this paper is to work out the theoretical and numerical implications of assuming that (2) or some variant of it represents societal preferences.

The government is assumed to know the parameters of U as well as the values W_1 and W_0 in the population. Its information beyond this depends on the observability of α: if it is fully observed, the government knows the W_α for each individual; if it is partially observed, the government only knows the value of a characteristic that is correlated with W_α; and if it is

[2] Low-ability single mothers with young children are intended for this group as well, although for them the presumption is that their home productivity (i.e., the value to their staying at home with their children) is high relative to their market wage. We do not introduce home productivity for simplicity.

[3] A more general model would allow $V=V_1[L_1,U(1,W_1+B)] + V_0[L_0,U(0,G)]$ and hence permit substitutability between L and U. But then the numerical simulations in the next section would have to assume values for those substitution parameters. We therefore work only with the extreme case in this paper.

completely unobserved, the government has no additional information.[4] Given its information set, the government then sets a benefit schedule {G,B} and decides whether to impose a work requirement to maximize V.

We will also assume that the government has a fixed amount of expenditure per poor person (the size of the poor population is fixed), Y, to spend on transfer benefits, and hence we will sidestep the full optimal tax problem. This is a major simplication which allows us to focus on the low income population. One could vary this expenditure amount and calculate the utility loss from taxpayers arising from of each its levels, and thereby conduct a fuller analysis than is conducted in this paper.

Fully Observed α. In this case the government can make lump sum transfers to each individual. The question is whether adding a work requirement can improve V. Assuming that the optimum occurs at $L_1=1$, $L_0=0$, which we shall show to always occur, the government budget constraint is $Y=B_1+G_0$ where B_1 is the lump sum transfer to high-ability individuals and G_0 is the lump sum transfer to low-ability individuals (since the transfers are lump sum, the welfare tax rate for both groups is zero and hence $B_1=G_1$ and $B_0=G_0$). Without a work requirement, the government maximizes (2) by choosing lump sum transfers B_1 and G_0 subject to this budget constraint and subject to constraints to insure that $L_1=1$ and $L_0=0$ are chosen:

[4] We ignore the fact that, even in the completely unobserved case, the government can deduce the wage of those individuals who choose to work (as in Laroque(2005)). We assume instead that the government must set benefits and work requirements in this case prior to observing the L and WL that will determine benefits. We also assume that individuals cannot hide, or affect the reporting of, their true W or L. See Maderner and Rochet (1995) for a discussion of the consequences of allowing such behavior.

$$U(1,W_1+B_1) \geq U(0,B_1) \quad (3)$$

$$U(1,W_0+G_0) \leq U(0,G_0) \quad (4)$$

The potential problem with the solution is that setting $L_1=1$ and $L_0=0$ in (2) and maximizing it w.r.t. the budget constraint may lead to a solution which violates (3) or (4). The marginal condition for the maximization is

$$U_2(0,G_0)=U_2(1,W_1+B_1) \quad (5)$$

which can be solved for B_1 and G_0 by combining it with the budget constraint.[5] This condition equalizes marginal utilities of income across groups; if L and I are separable in the utility function, incomes are also equalized (see the simulations below). If this condition holds at a point where $L_1=1$ and $L_0=0$ voluntarily, the optimum can be achieved without a work requirement. However, there are cases where this will not hold and where the provision of the lump sum benefits dictated by (5) will induce the high-ability group to not work and hence will violate (3). If a lump sum program is to prevent this, B_1 must be lowered (and hence G_0 must rise to satisfy the budget constraint) until (3) is satisfied in equality, but this will lead to a violation of (5). A work requirement, on the other hand, will be superior in this case because it will impose $L_1=1$ for the high-ability group and benefits can then be set according to (5), thereby

[5] With the assumption that leisure is a normal good, (5) will exhibit the single crossing-point property and a unique solution will exist. We will assume that it is an internal solution and that so that (5) holds with equality, even though it is possible that one or the other benefit might be driven to zero without equating marginal utilities.

achieving the first-best solution. High-ability individuals will not reject this benefit because it is better than nothing. This work requirement solution will dominate the lump sum solution because a lower B_1 and a higher G_0 than that dictated by (5) will lead to a lower V. The distributional implications of the work-requirement solution are unfavorable because it allows higher benefits to be paid to the better-off, high-ability individuals and hence lower benefits to the worse-off, low-ability individuals.[6] One implication of these results is that the optimality of work requirements does not depend upon the existence of uncertainty about α, given the government preference function we have assumed.

Completely Unobserved α. With α completely unobserved, the solution without a work requirement is a negative income tax (NIT) in which only B and G (or, equivalently, G and t) are available as policy instruments. B and G are not indexed by α because true wages and ability are not observed. Not surprisingly, an income-equalizing choice of B and G (i.e., t=1) will result in an inferior solution because all high ability individuals will not work, and the lower B that may be required to induce work in that population in the known α case will only make matters worse. Here, just the opposite will happen, for B must be raised and G lowered, relative to the first-best values, in order to induce work among the more able. Thus the tax rate will fall from t=1 to some lesser value. The exact value to which it will fall is determined by the incentive

[6] The societal preference functions in (1) and (2) imply distributional weights because marginal utility declines with income. However, there may be distributional considerations above and beyond what is implied by these functions. Note that the result that work requirements result in more distribution toward the better-off is seemingly inconsistent with the general result in the literature that tagging allows more expenditure to be concentrated on the worse-off. But the literature result is comparing an NIT program under completely unknown α to a tagging program under a completely or partially known α, and in that comparison the work requirement program does result in more distribution toward the worse off (see below).

compatibility constraints, which are now:

$$U(1,W_1+B_1) \geq U(0,G) \quad (6)$$

$$U(1,W_0+B_0) \leq U(0,G) \quad (7)$$

where $B_\alpha = G-tW_\alpha$. At t=1, (7) is satisfied but (6) is not. The government will lower t and G until (6) is satisfied in equality (this will always occur before (7) is violated) and hence this will determine optimal benefits. Optimal V will be lower than what it was in the case of known α. Also compared to the known α case, the NIT has unfavorable distributional implications because it shifts funds to the better off; gaining more information about α therefore could be seen as an investment in assisting the worse-off individuals in the population. Note as well that it is possible that the optimum will occur at $t<0$, which is a type of earnings subsidy solution.

Not knowing α means that work requirements, if imposed at all, must be imposed on everyone. The optimum in this case is for the government to pay all benefits to the working, high-ability individuals; under the assumption that no social utility is obtained from work by low-ability individuals, the government should prefer them to not work but then not receive benefits. High-ability workers receive a benefit of 2Y and low-ability nonworkers receive nothing. This does not seem to be a very interesting solution from a practical viewpoint, but it logically follows from the assumptions.[7] This optimum is even farther from the first-best than

[7] It is possible that the benefit of 2Y paid to workers is sufficiently high to induce low-ability workers to prefer to work. In that case, the government could either discard some of its funds and pay workers a sufficiently lower benefit to induce the low-ability individuals to return to nonwork, or it could adopt a nonlinear benefit formula which simply denied benefits to any worker with earnings less than W_1.

the NIT solution is, and hence will always be inferior, and the distributional consequences are also clearly unfavorable.

These two polar cases bracket the case where α is partially observed. They imply that work requirements are more likely to be optimal, and more likely to have favorable distributional implications, the more observed is the ability parameter α. The accuracy of the observability mechanism is therefore the critical empirical question.

Partially Observed α. We assume that the government observes a continuous signal s for each individual which is positively correlated with α but nevertheless contains noise. The government is assumed to know the distribution function of the errors. We focus on how the availability of this signal can improve targeting of work requirements. A work requirement can be assigned if $s>s^*$ and not otherwise, where s^* is a government-chosen cutoff for the signal. Letting $F(s^*)$ be the probability that an $\alpha=1$ individual has $s>s^*$ and $H(s^*)$ be the probability that an $\alpha=0$ individual has $s<s^*$ ($F'<0$, $H'>0$), and assuming the government to be an expected utility maximizer, its objective function is now

$$E(V) = F(s^*)\, U(1,W_1+B_1) + [1-F(s^*)]\, 1[U(1,W_1+B_1)>U(0, G)]\, U(1,W_1+B_1)$$
$$+ H(s^*)1[U(0,G)>U(1,W_0+B_0)]\; U(0,G) \qquad (8)$$

where 1 is the indicator function and where $B_\alpha=G-tW_\alpha$. Here we assume that individuals who are told they must work always comply (this assumption is relaxed when we allow welfare nonparticipation below). The government chooses G, t, and s^* subject to the constraint

$$
\begin{aligned}
Y \;=\; & F(s^*)B_1 \\
& + \; [1-F(s^*)] \; \{ \; 1[U(1,W_1+B_1)>U(0,G)] \; B_1 \; + \; 1[U(1,W_1+B_1)\leq U(0,G)] \; G \; \} \\
& + \; H(s^*) \; \{ \; 1[U(0,G)>U(1,W_0+B_0)]G \; + \; 1[U(0,G)\leq U(1,W_0+B_0)]B_0 \; \} \\
& + \; [1-H(s^*)]B_0 \qquad (9)
\end{aligned}
$$

The first order conditions are cumbersome to specify because of the nonlinearity of the choice functions but the influences on the choices are clear enough. Increases in G raise benefits for nonworkers and hence improve the utility of low-ability workers not subject to a work requirement but also lead to utility losses from those high-ability workers not subject to work requirements who quit work. Increases in t also drive more high-ability workers not subject to a work requirement to nonwork. Loosening the stringency of the requirement by raising s* lets more high-ability individuals off the hook but puts more low-ability individuals into the correct, no-requirement category. The government utility from this program will be in between those in the two polar cases.

A More General Model. The next section of the paper will report numerical calculations from a calibrated model which is generalized from these simple models. The main generalization is to allow a continuum of types rather than just two. Let $F(\alpha)$ be the c.d.f. of α. Assume that society desires that those with $\alpha>\alpha^*$ work and that those with $\alpha<\alpha^*$ not work. As discussed in the early sections of this paper, how voters arrive at a particular α^* is not considered here but is instead taken as given. In the simulations in the next section, α^* will be taken to be a particular percentile point in the distribution, and that point will be varied.

The government objective is the following:

$$\underset{S}{\text{Max}} \quad V = \int_{\alpha^*}^{\infty} L_\alpha U(1, W_\alpha + P_\alpha B_\alpha)\, dF(\alpha) + \int_{-\infty}^{\alpha^*} (1 - L_\alpha) U(0, G)\, dF(\alpha) \qquad (10)$$

where $B_\alpha = G - tW_\alpha$, P_α is a dummy variable indicating participation in the welfare system, and S is the government instrument set, either {G,t} alone or this set plus work requirements. Here we now allow individuals to choose whether to participate in welfare, and thereby allow low-ability individuals who do not accept the work requirement to opt out of the system as well as allow the government to make some high-ability individuals ineligible. This objective is maximized subject to the individual choice functions for L and P, which are a function of S, and the government budget constraint.

No attempt is made here to characterize the possible solutions to this generalized model for either arbitrary or specific parametric forms of the ability distribution and signal error distribution. However, the calibrations in the next section, which will assume normality for the distributions of both, will result in substantive modification of the solutions outlined in the two-point ability distribution assumed earlier in this section. While there will always be a low-ability group which does not work and receives benefits, and a high-ability that works and receives benefits, two other groups will often appear in the solution. One is a group of high-ability individuals who work and receive no benefits and hence are off welfare, as described previously; but another is a group of low-ability individuals who neither work nor receive benefits but who are not the lowest-ability individuals. This case will be discussed in more detail in the next section.

Simulations

The model described in (10) is calibrated to a labor supply model estimated by Keane and Moffitt (1998). Keane and Moffitt (KM) used a 1984 cross-sectional data set containing information on weekly hours of work, hourly wages, and welfare program participation on a sample of 968 single mothers with low levels of assets and who are therefore likely to be eligible for welfare if not working. The model assumed a quadratic utility function $U(H,I)=\beta H-\gamma H^2+I-\delta I^2+c$, where c is a constant. Interactions between H and I were tested and rejected. The parameter β was allowed to be stochastic by KM but here it is taken as fixed for all individuals; we calibrate its conditional mean to fit the observed fraction of women in the KM data who were working (.50). We define L=1 to equal 20 hours per week. With this modification, and with the necessary translations of parameters to a dichotomous outcome, the utility function becomes

$$U(L,I) = \beta L + I - \delta I^2 + c \qquad (11)$$

where the redefined β=-80 and δ=.000315. With the normalization of the coefficient on I to 1, β=-80 implies that 20 hours of work reduces utility by about $80. The constant term in (11) is set to zero.

KM jointly estimated an hourly wage equation of the form $\log W=X\lambda+\nu$, where X is a set of regressors (age, education, etc.) and ν is a normally distributed error. Adjustments for selection bias from the absence of observed wages for nonworkers were made, so that the estimates are to be interpreted as offer wages for all individuals whether working or not. The predicted wage at the KM mean of the X vector was $4.00 per hour and the estimated standard

deviation of ν was estimated to be .51.[8] We take ν to be the equivalent of α in the theoretical model. We take 1,000 random draws from a $N(0,.51^2)$ distribution and use them to create a wage distribution. With the assumed utility function parameters we can predict a value of utility-maximizing choice of L for each individual for each budget constraint.

The government budget constraint is taken from KM as well, for their data had information on average welfare benefits and the percent of the population receiving benefits. The data showed that the mean weekly benefit among recipients was $64 which, combined with the welfare participation rate of .30 in their sample, implies a mean benefit over entire population of $19.2 per week per capita. With a simulation sample of 1,000 individuals, we therefore set the weekly budget constraint for the government to $19,200. With all parameters specified and with a government budget constraint, we can find the values of G and t that produce the maximal value of V, and whether work requirements improve or lower that value. We can also produce the distribution of utilities among gainers and losers from each government scheme for the 1,000 simulated individuals.

The first column of Table 1 shows the values of utility at decile points of the wage distribution if there is no transfer program present, in which case each individual utility is the maximum of zero (the value of (11) with L=I=0; this is a normalized value) and $\beta+W-\delta W^2$, where W is equal to 20 times the individual's hourly wage. About half of the population does not work at all, a typical figure for low-wage single mother samples in the U.S. and equal to that in KM by construction.

Known α. The two programs to be considered are, as described previously, a program of

[8] All dollar amounts in this paper are taken from KM and hence are 1984 dollars.

lump sum transfers and one with work requirements (benefits tied to work or nonwork. We will consider cases, in turn, where the government wishes that the highest 90 percent, 75 percent, and 50 percent of the ability distribution in the poor population work; or, equivalently, where it exempts the lowest 10 percent, 25 percent, and 50 percent, respectively. The last of these will be the easiest to achieve because it is closest to the utilitarian solution; the others will be more difficult. No percents lower than 50 will be considered because about 50 percent of the poor population is working in the absence of any program.

As described in the theoretical discussion, the initial inclination of the government, at least if it desires more work (90 or 75 percent) than results under the utilitarian solution, will be to lower benefits for high-ability individuals and raise them for low-ability individuals to induce the former to work; or, in the work requirement case, to simply require that the former work to receive benefits. But, unlike the case of a two-point distribution of ability, here there may be a portion of the population which does not work and is not given any benefit because the government wishes them to work but has insufficient funds to induce them to do so. In the case in which 90 percent of the poor population is desired to work, for example, offering financial support to those in the 11^{th} percentile of the ability distribution is a waste of money if they are not induced to work because the government receives no social utility from those individuals if they do not work.

There are, in total, four possible categories of how individuals are treated. The solution for V will be of the general form:

$$V = \sum_{i=1}^{n_1} U(0,G) + \sum_{i=n_2+1}^{n_3} U(1,W_i+B_i) + \sum_{i=n_3+1}^{1000} U(1,W_i) \qquad (12)$$

where n_1 is the number of nonworkers receiving benefits (n_1 will generally constitute the lower 10 percent, 25 percent, or 50 percent of the population whom the government wishes not to work), n_2-n_1 is the number of nonworking individuals in the next highest portion of the ability distribution who receive nothing and hence do not contribute to V, n_3-n_2 is the number of recipient workers, and 1000-n_3 is the number of high-ability working nonrecipients. The benefit is B_i=G-W_i because incomes are equalized among those who receive benefits.

Table 1 shows the simulated optima under the lump sum and work requirement programs. If the government desires 90 percent of the poor population to work and tries to do so with a lump sum program, it will concentrate all of its benefits on the bottom 10 percent of the ability distribution and will give nothing to nonworkers or workers above that point. Benefits are not given to workers because the benefits that they would receive would induce some of them to quit work, if given as a lump sum. In the work requirement program, however, workers can be offered benefits only if they work; in the resulting optimum, the incomes of individuals in the (50.7,81.1) portion of the ability distribution are given benefits. This results in a lowering of benefits for the bottom 10 percent of the ability distribution. However, government utility (V) is necessarily higher under the work requirement program because it has an additional policy instrument which loosens the incentive compability constraint faced under the lump sum program.

While this result may seem unrealistic, the U.S. experience can be interpreted as exactly reflecting this solution. Welfare reform in the 1990s in the U.S. concentrated benefits on a more disadvantaged population, and it has been documented repeatedly that many of those former recipients who have left welfare are not working (Fang and Keane, 2004; Moffitt, 2002b). At

the same time, the rise of the EITC can be interpreted as supplementing the benefits of those who are of sufficiently high ability to work but only up to a maximum level, beyond which no additional subsidy is provided. Thus the four categories of individuals described in this solution appear to exist in the current U.S. situation.

For the cases where 75 percent or 50 percent of the population is desired to work, again benefits are paid to the lowest 25 and 50 percent of the ability distribution, respectively; no benefits are paid to nonworkers in a portion of the ability distribution above this; and no benefits are paid to workers. No benefits are paid to workers because the government expenditure is entirely absorbed by paying benefits to those whom it desires not to work; the resulting average benefit is sufficiently low that the wages of workers exceed this benefit (recall that taxation is not allowed, and hence incomes of higher-wage workers cannot be equalized with those below if their wages exceed the benefit level).[9] The solution is the same under the lump sum and work requirement programs because the only difference in those programs is whether benefits are paid to workers, and in neither of these cases is that done. Thus we find that work requirements are optimal only if the percent of the population desired to work is high.

Completely Unobserved α. Table 2 shows the solutions for the case of completely unobserved α for NIT and work requirement programs. The NIT figures results represent solutions to linear NIT programs which maximize V by manipulating G and t only, while the work requirement add the requirement that benefits can be received only if the recipient works. When 90 percent of the population is desired to work, the NIT solution is to eliminate support for

[9] In many of the cases in the table, utility is lower for many of those off welfare than for those on welfare. This is because those off welfare work and lose the utility value of leisure, and because a marginal redistribution of income toward those workers lowers V.

nonworkers entirely (G=0) and to offer an earnings subsidy with t=-.26. The gain from inducing as many individuals to work as possible leads the guarantee to be reduced to zero. The losses in utility from not paying benefits to the lowest 10 percent of the distribution are outweighed by the gains from inducing all individuals in the percentile range (33.4, 50.6) to work whereas they would not have worked if no program had been offered, and by the gains from paying benefits to workers higher up the distribution. The work requirement solution in this case is identical; G=0 by definition, and therefore the solution must be the same as the NIT.[10]

The programs diverge for the 75 and 50 percent cases. When only 75 percent of the population is desired to work, the NIT solution results in a small but positive G but still an earnings subsidy, t=-.22. Nonworkers in the bottom 25 percent receive some support and less is given to those higher up the distribution as a consequence. But the work requirement program, which cannot pay benefits to nonworkers, now suffers because a larger fraction of the population whom the government does not wish to work (25 percent) receives no support. The work requirement solution is the same as before, but with lower V. When 50 percent of the population is desired to work, the NIT solution offers a higher G and a tax rate of approximately zero, and V reaches the maximum of any program in the table. The value of V for the work requirement program falls further because even more individuals whom the government desires not to work receive no support. Thus, it is clear that when α is completely unknown that work requirement programs are inferior to linear NIT schemes.

[10] In the theoretical section above, it was never necessary to reduce G to zero to induce high-ability individuals to work, but that followed from the assumption that high-ability individuals would always work in the absence of any program. Here, with 90 percent of the population desired to work but only 50 percent working in the absence of the program, about 40 percent of the "high ability" distribution is not working in the absence of the program.

As expected, the values of V are lower for all these cases than obtain under known α. Also as expected, the distributional implications of these cases are unfavorable relative to those under known α.

Partially Observed α. Partial observability is simulated by adding noise to the wage equation and by assuming that the resulting wage is what is observed by the government. This "signal" wage is then used to categorize the population into those who are required to work or not required. The signal error is constructed as

$$s = \upsilon + [(1-k)/k]^{1/2} \epsilon \qquad (13)$$

where υ is the true wage error and ϵ is the noise. The parameter k controls the amount of noise, with k=1 implying no noise and k=0 implying infinite noise. The noise error ϵ is drawn from a normal distribution with the same mean (0) and variance as υ; with this simulation construction, it can be shown by manipulation of (13) that $k=Var(\upsilon)/Var(s)$, or what is usually called the reliability ratio of the signal. The k=1 case is equivalent to the α known case; simulations will be conducted for k=.99, .90, .75, and .50 (reliability less than .50 will not be necessary, as the results below will demonstrate).

Tables 3 shows the solution for the case where 90 percent of the poor population is desired to work. The percentile point utilities are those for the true wage, not the signal wage. When k=.99, the bottom 7.7 percent of the signal wage distribution is exempted from the work requirement and allowed to receive a benefit of $133, while workers in the (54, 89) percentile range receive benefit supplements to equalize their incomes with those of nonworkers (t=1 is

imposed for this group). This solution is not very different from the α known case, with the only difference that in that case exactly 10 percent of the low-ability distribution was given support because there was no inaccuracy in the assignment. Here, the government reduces that percent in order to avoid wasting money on high-wage individuals who are incorrectly assessed as being low-wage and receive benefits, for their utilities do not contribute to societal welfare. As shown by the 'coverage' and 'targeting' rates in the table, the solution takes place at a point where all those who are offered benefits for not working are low-wage, even though this results in only about three-fourths of low-wage individuals to be covered by the program.[11]

The decile points shown in the Table reveal the noise in the signal, for there are many individuals with true wages in the lower 50 percent of the distribution who are working because they had a positive wage error and were "mistakenly" offered a supplemental benefit, and they choose to accept it and to work. The individual at the 50th percentile is of this type, for example; the goverment does not start paying supplemental benefits until the 54th percentile of the signal wage distribution.

As k falls and the noise increases, the government grows more cautious and reduces the fraction of the low-s distribution to whom it offers a nonworking benefit. For the k=.90, .75, and .50 levels of screening accuracy, the percent offered a benefit falls to the bottom 3.9 percent, 1.3 percent, and .5 percent of the signal wage distribution, respectively. Benefits correspondingly increase both for the working groups and for the nonworkers who received G. The reduction in coverage is made to the point where there is no error in targeting. Further reductions in k are not

[11] The coverage and targeting rates in the table are functionally related, for c=t(s/p), where c=coverage rate, t=targeting rate, s=percent of signal wage distribution offered G for not working, and p=percent society desires to not work (=10 percent in this case).

simulated because, by k=.50, the government is on the verge of moving entirely to the program it chose in Table 2 when α was completely unknown and benefits were denied to nonworkers entirely.

The amount of error and noise in the wage distribution increases as k falls, and is demonstrated in the table by the individual at the 60th true wage percentile point, who has a negative error and, for k=.90 and lower, is deemed by the government not to be eligible for supplemental benefits and hence works without benefits (9.2 is the utility level for this individual off welfare). Likewise, the individuals at the 70th and 80th true wage percentile points have negative wage errors and are denied benefits if k falls sufficiently, again reducing their utility to the nonwelfare level.

Table 4 shows the results for 75 and 50 percent work cases, and reveal the same pattern but with more generous benefits paid to nonworkers because a greater fraction of the ability distribution is exempted from the work requirement. For k=.99, benefits are paid up to the 22nd percentile of the signal wage distribution if 25 percent are desired not to work and up to the 42nd percentile of the distribution if 50 percent are desired not to work. However, as k falls, the government rapidly retreats on this generosity and moves quickly toward the near denial of benefits to nonworkers. In the 50-percent-work case where 50 percent of the population is allowed not to work, only the bottom 9.5 percent of the signal wage distribution is offered a benefit if k=.50.[12] The benefit is very large for that group, and no benefits are paid to workers, but the 40 percent of individuals in the (10,50) percentile range whom society would prefer not to

[12] The individual at the 10th percentile of the true wage distribution has a negative wage error and 'mistakenly' receives benefits.

work, do not do so but also do not receive benefits (again, there are individuals with true wages in this range whose signal wages put them in the lower 9.5 percent; they receive benefits; but these individuals are not at the percentile points shown in the table). Again the group who is offered benefits for not working is highly targeted on low-wage individuals as k falls, again with consequent reductions in coverage.

The major conclusion to be drawn from these results is that the desirability of work requirements requires a reasonably accurate screening mechanism. While partial observability of ability always allows a work requirement program to improve social welfare over what it would be if there was no observability, the gains are small if the screening mechanism is not very accurate. In some cases, using an inaccurate screening mechanism may be worse than not imposing work requirements at all but instead simply applying an NIT program which manipulates the guarantee and tax rate.

Summary

This paper has laid out a simple model of optimal work requirements when those requirements arise from a direct preference by society for work among more able individuals. Work requirements are not always optimal, and their distributional effects depend heavily on the degree to which ability is observed. With fully observed ability, work requirements allow benefits to be paid to workers in situations where a lump sum transfer would induce those same individuals to quit work. With fully unobserved ability, work requirements tend to penalize those at the bottom of the ability distribution whom the government cannot induce to work

because of insufficient funds. When ability is partially observed, a reasonably accurate signal allows the government to pay benefits to some of those at the bottom of the wage distribution while simultaneously supplementing the wages of workers, moving some of them over the margin into work. But the capability of the government to do this falls rapidly with declines in the accuracy of the signal. Moreover, the inaccuracy of the signal leads to a considerable amount of noise, with many high-ability individuals inadvertently not required to work and many low-ability individuals required to work. At low levels of signal accuracy, it is possible that a simple negative income tax with no screening on ability would be socially preferable.

Several of the optimal programs reveal a four-part segmentation of the population that has some resemblance to current US distributional policy toward single mothers. A segment at the bottom of the ability distribution is offered a guaranteed payment for nonwork; a segment just above that in the ability distribution is offered nothing, even though they are low-ability, because of insufficient funds; a segment higher up the distribution is offered an earnings supplement; and the highest segment is offered nothing. The US system with a cash transfer program for the very poor with stringent eligibility and work requirements, coupled with a slightly better-off portion of the population which is off welfare altogether, combined with a higher-ability portion which receives significant earnings subsidies from the EITC, resembles this type of system.

These results in the paper would be altered if some of the assumptions of the model were to change. If the government had redistributional preferences (i.e., distributional weights) above those implied by declining marginal utility of income, more support would necessarily flow to low-ability workers, even in cases where ability is only partially observed. If more government funds were devoted to the transfer program, many of the constraints which result in denial of

benefits to families whom the government wishes to support would be relaxed.

There are also many research extensions of the model used in the paper. The strict assumption that no social utility is obtained from high ability individuals who do not work could be relaxed, for example, as could the assumption that no social utility is obtained from low ability individuals who work. Also, the game that is assumed to be played by the government assumes that no learning takes place after individuals demonstrate their earning ability or work preferences. The model in the paper does not represent the fairly complex assessment of sanctions and compliance with sanctions that takes place when work requirements are not met and, relatedly, the motivation of caseworkers is ignored here; caseworkers, welfare agency heads, and society are all assumed to have the same objective function and information. This may not be the case and there may be agency issues among these actors. These are all fruitful areas for more work.

References

Akerlof, G. 1978. "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning." American Economic Review 68 (March): 8-19.

Barth, M. and D. Greenberg. 1971. "Incentive Effects of Some Pure and Mixed Transfer Systems." Journal of Human Resources 6 (Spring): 149-170.

Beaudry, P. and C. Blackorby. 2004. "Taxes and Employment Subsidies in Optimal Redistribution Programs." Mimeo, University of British Columbia and University of Warwick.

Besley, T. and S. Coate. 1992. "Workfare Versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs." American Economic Review 82 (March): 249-26.

Besley, T. and S. Coate. 1995. "The Design of Income Maintenance Programmes." Review of Economic Studies 62: 187-221.

Bloom, Dan and Don Winstead. 2002. "Sanctions and Welfare Reform." In Welfare Reform and Beyond: The Future of the Safety Net, eds. I. Sawhill, R. Weaver, R. Haskins, and A. Kane. Washington: Brookings Institution.

Boyer, G. 1990. An Economic History of the English Poor Law, 1750-1850. Cambridge: Cambridge University Press.

Brett, C. 1998. "Who Should Be On Workfare? The Use of Work Requirements as Part of an Optimal Mix." Oxford Economic Papers 50: 607-622.

Browning, E. 1975. Redistribution and the Welfare System. Washington: American Enterprise Institute.

Choné, P. and G. Laroque. 2005. "Optimal Incentives for Labour Force Participation." Journal of Public Economics 89 (February): 395-425.

Cuff, K. 2000. "Optimality of Workfare with Heterogeneous Preferences." Canadian Journal of Economics 33 (February): 149-174.

Diamond, P. and E. Sheshinski. 1995 "Economic Aspects of Optimal Disability Benefits." Journal of Public Economics 57: 1-23.

Diller, M. 2000. "The Revolution in Welfare Administration: Rules, Discretion, and Entrepreneurial Government." NYU Law Review 75 (November): 1121-1220.

Fang, H. and M. Keane. 2004. "Assessing the Impact of Welfare Reform on Single Mothers." Brookings Papers on Economic Activity 1: 1-95.

Fortin, B.; M. Truchon; and L. Beauséjour. 1993. "On Reforming the Welfare System: Workfare Meets the Negative Income Tax." Journal of Public Economics 51: 119-151.

Kaplow, L. 2004. "Optimal Income Transfers." Mimeo. Cambridge: Harvard University.

Keane, M. and R. Moffitt. 1998. "A Structural Model of Multiple Welfare Program Participation and Labor Supply." International Economic Review 39 (August): 553-589.

Laroque, G. 2005. "Income Maintenance and Labor Force Participation." Econometrica 73 (March): 341-376.

Lurie, I., ed. 1975. Integrating Income Maintenance Programs. NY: Academic Press.

_____. 2001. "Changing Welfare Offices." Policy Brief No. 9. Washington: Brookings Institution.

Maderner, N. and J. Rochet. 1995. “Is It Legitimate to Encourage Work-Sharing?” Scandinavian J of Economics 97:621-633.

Mead, L. 1986. Beyond Entitlement: The Social Obligations of Citizenship. New York: Free Press.

_____, ed. 1997. The New Paternalism: Supervisory Approaches to Poverty. Washington: Brookings.

Moffitt, R. 2002a. “Economic Effects of Means-Tested Transfer Programs in the U.S..” Tax Policy and the Economy 16: 1-35.

_____. 2002b. “From Welfare to Work: What the Evidence Shows.” In Welfare Reform and Beyond, eds. I. Sawhill, R. Weaver, R. Haskins, and A. Kane. Washington: Brookings.

_____. 2003a. “The Negative Income Tax and the Evolution of U.S. Welfare Policy.” Journal of Economic Perspectives 17 (Summer): 119-140.

_____. 2003b. “The Temporary Assistance for Needy Families Program.” In Means-Tested Transfer Programs in the United States, ed. R. Moffitt. Chicago: University of Chicago Press.

Mirrlees, J. 1971. “An Exploration in the Theory of Optimum Income Taxation.” Review of Economic Studies 38 (April): 175-208.

Parsons, D. 1996. "Imperfect 'Tagging' in Social Insurance Programs." Journal of Public Economics 62 (October): 183-208.

Pavetti, L. and D. Bloom. 2001. “Sanctions and Time Limits.” In The New World of Welfare, eds. R. Blank and R. Haskins. Washington: Brookings Institution.

Saez, E. 2002. “Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses.” Quarterly Journal of Economics 117: 1039-1073.

Salanié, Bernard. 2002. “Optimal Demogrants with Imperfect Tagging.” Economics Letters 75 (2002): 319-324.

U.S. Department of Health and Human Services. 2002. TANF: Fourth Annual Report to Congress. Washington: Administration for Children and Families.

Table 1

Maximal V and Distributions of Utility at Percentile Points of the Wage Distribution
Under Optimal Schemes with Known α: Lump Sum and Work Requirement Programs

	No Program	90 Percent Desired to Work		75 Percent Desired to Work		50 Percent Desired to Work	
		Lump Sum Program	Work Req. Program	Lump Sum Program	Work Req. Program	Lump Sum Program	Work Req. Program
V	--	36,847	37,062	37,543	37,543	37,776	37,776
Wage Percentile							
10	0	180.4	117.5	75.0	75.0	38.0	38.0
20	0	0	0	75.0	75.0	38.0	38.0
30	0	0	0	0	0	38.0	38.0
40	0	0	0	0	0	38.0	38.0
50	0	0	0	0	0	38.0	38.0
60	9.2	9.2	37.5	9.2	9.2	9.2	9.2
70	20.5	20.5	37.5	20.5	20.5	20.5	20.5
80	35.6	35.6	37.5	35.6	35.6	35.6	35.6
90	63.0	63.0	63.0	63.0	63.0	63.0	63.0
G	--	192.0	122.0	76.8	76.8	38.4	38.4

Table 2

Maximal V and Distributions of Utility at Percentile Points of the Wage Distribution
Under Optimal Schemes with Unknown α: NIT and Work Requirement Programs

	90 Percent Desired to Work		75 Percent Desired to Work		50 Percent Desired to Work	
	NIT Program	Work Req. Program	NIT Program	Work Req. Program	NIT Program	Work Req. Program
V	34,899	34,899	34,996	34,899	37,057	33,273
Wage Percentile						
10	0	0	3.5	0	17.7	0
20	0	0	3.5	0	17.7	0
30	0	0	3.5	0	17.7	0
40	7.7	7.7	8.1	7.7	17.7	7.7
50	19.2	19.2	19.2	19.2	17.7	19.2
60	31.6	31.6	31.2	31.6	27.8	31.6
70	45.6	45.6	44.7	45.6	39.3	45.6
80	64.2	64.2	62.7	64.2	54.4	64.2
90	97.9	97.9	95.4	97.9	82.1	97.9
G	0	0	4	0	18	0
t	-.26	-.26	-.22	-.26	-.02	-.26

Notes: NIT = negative income tax

Table 3

Maximal V and Distributions of Utility Under Optimal Schemes with Partially Observed α: 90 Percent Desired to Work

	Screening Accuracy			
	k=.99	k=.90	k=.75	k=.50
V	36,919	36,430	36,117	35,678
True Wage Percentile				
10	0	0	0	0
20	0	0	0	0
30	0	0	0	0
40	0	0	0	0
50	47.3	74.3	89.5	105.6
60	47.3	9.2	9.2	9.2
70	47.3	74.3	89.5	20.5
80	47.3	74.3	35.6	35.6
90	63.0	74.3	89.5	105.6
G	133	163	180	198
Percent of signal wage distribution offered G	7.7	3.9	1.3	0.5
Coverage rate[1]	77	39	13	5
Targeting rate[1]	100	100	100	100
Prob(L=1 | $\alpha < \alpha_{10}$)	0	0	0	0
Prob(L=0 | $\alpha > \alpha_{10}$)	.449	.441	.426	.417

Notes:

[1] Coverage rate = percent of those with $\alpha < \alpha_{10}$ who are offered a benefit for not working. Targeting rate = percent of those who are offered a benefit for not working who have $\alpha < \alpha_{10}$.

Table 4

Maximal V and Distributions of Utility Under Optimal Schemes with Partially Observed α: 75 and 50 Percent Desired to Work

	75 Percent Desired to Work		50 Percent Desired to Work	
	k=.99	k=.50	k=.99	k=.50
V	37,387	36,491	37,730	36,067
True Wage Percentile				
10	85.9	0	45.3	172.9
20	85.9	0	45.3	0
30	0	0	45.3	0
40	0	0	45.3	0
50	0	181.0	0	0
60	9.2	9.2	9.2	9.2
70	20.5	20.5	20.5	20.5
80	35.6	35.6	35.6	35.6
90	63.0	181.0	63.0	63.0
G	88	287	46	184
Percent of signal wage distribution offered G	21.7	.8	41.8	9.5
Coverage rate[1]	86	3	84	18
Targeting rate[1]	100	100	100	96
Prob(L=1 | $\alpha < \alpha_p$)	0	0	0	0
Prob(L=0 | $\alpha > \alpha_p$)	.348	.335	.022	.030

Notes:
p=25 or 50

[1] Coverage rate = percent of those with $\alpha < \alpha_p$ who are offered a benefit for not working.
Targeting rate = percent of those who are offered a benefit for not working who have $\alpha < \alpha_p$.