

DISCUSSION PAPER SERIES

No. 5937

OPTIMAL WELFARE-TO-WORK PROGRAMS

Nicola Pavoni and Giovanni L Violante

*INTERNATIONAL MACROECONOMICS
and LABOUR ECONOMICS*



Centre for **E**conomic **P**olicy **R**esearch

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP5937.asp

OPTIMAL WELFARE-TO-WORK PROGRAMS

Nicola Pavoni, University College London and IFS
Giovanni L Violante, New York University and CEPR

Discussion Paper No. 5937
November 2006

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL MACROECONOMICS and LABOUR ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Nicola Pavoni and Giovanni L Violante

ABSTRACT

Optimal Welfare-to-Work Programs*

A Welfare-to-Work (WTW) program is a mix of government expenditures on various labor market policies targeted to the unemployed (e.g., unemployment insurance, job search monitoring, social assistance, wage subsidies). This paper provides a dynamic principal-agent framework suitable for analyzing chief features of an optimal WTW program such as the sequence and duration of the different policies, the dynamic pattern of payments along the unemployment spell, and the emergence of taxes/subsidies upon re-employment. The optimal program endogenously generates an absorbing policy of last resort ('social assistance') characterized by a constant lifetime payment and no active participation by the agent. Human capital depreciation is a necessary condition for policy transitions to be part of an optimal WTW program. The typical sequence of policies is quite simple: the program starts with standard unemployment insurance, then switches into monitored search and, finally, into social assistance. The optimal benefits are decreasing during unemployment insurance and constant during both job search monitoring and social assistance. Whereas taxes (subsidies) can be either increasing or decreasing with duration during unemployment insurance, they must decrease (increase) during a phase of job search monitoring. In a calibration exercise, we use our model to analyze quantitatively the features of the optimal program for the U.S. economy. With respect to the existing U.S. system, the optimal WTW scheme delivers sizeable welfare gains to unskilled workers because the incentives to search for a job can be retained even while delivering more insurance, and using costly monitoring less intensively.

JEL Classification: D82, H21, J24, J64 and J65

Keywords: human capital, job search monitoring, recursive contracts, unemployment insurance and welfare-to-work

Nicola Pavoni
University College London
Gower Street
London
WC1E 6BT
Email: n.pavoni@ucl.ac.uk

Giovanni L Violante
Department of Economics
New York University
269 Mercer Street
New York, NY 10003-6687
USA
Email: gianluca.violante@nyu.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=152195

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=145021

* We thank Daron Acemoglu, Marco Bassetto, Chris Flinn, Narayana Kocherlakota, Per Krusell, Matthias Messner, Torsten Persson, Chris Pissarides, Rob Shimer and various seminar participants and discussants for useful comments. Violante thanks the CV Starr Center for research support, and the Federal Reserve Bank of Minneapolis for its hospitality at various stages of this project.

Submitted 17 October 2006

1 Introduction

Public expenditures on labor market policies targeted to the unemployed average 3% of GDP, across OECD countries (Martin, 2002). These policies face a delicate trade-off between providing unemployed workers with consumption insurance and re-employment incentives. In order to strike the right balance, governments use a wide range of policy instruments such as unemployment insurance, social assistance policies providing income support of last resort once unemployment benefits have expired, job-search monitoring, training, and wage subsidies. The share of expenditures on “active” labor market policies (i.e., job search monitoring, training and subsidized jobs) has risen substantially over the past 10 years, and this type of government intervention is now a pivotal ingredient of social welfare schemes.

It is instructive to consider how these various policy instruments have been assembled together in the United States and in the United Kingdom. In the United States, since 1935 there exists an unemployment insurance system with vast coverage which replaces on average 60% of the pre-displacement wage. Upon expiration of the unemployment compensation (usually after 26 weeks), other forms of financial support become available. The Temporary Assistance for Needy Families (TANF) program is the most notable example. With the Personal Responsibility and Work Opportunities Reconciliation Act (PRWORA) of 1996, the federal U.S. government limited the payment of TANF benefits to a maximum of 5 years—with tighter limits in several states—and imposed strict participation requirements to labor market programs for welfare recipients. In some programs the emphasis is on training and skill formation, whereas in others the focus is on quick reemployment through monitoring during job search and temporary subsidized jobs. The Food Stamps program represents the main form of social assistance, once the TANF benefits have expired. Finally, the Earned Income Tax Credit, introduced by the federal government in 1975, represents today the major wage subsidy program for low-income workers (see Moffitt, 2003, for a survey of the U.S. welfare system).

An example of an even more structured mix of policies is the U.K. “New Deal,” a mandatory program for all the unemployed workers introduced in 1998. Formally, the New Deal is organized in four sequential stages. Stage 1 consists of a standard unemployment insurance policy lasting 6 months for younger workers and 2 years for older workers. In stage 2 (the “gateway”), a personal adviser regularly meets with the worker to assist with and monitor her job search. If after a period of 3-6 months the worker is still unemployed, she moves to stage 3 (the “options”) where she enters a phase of training or subsidized jobs for up to one year. At the end of this phase, the worker enters stage 4 (the “follow-through”), another period of job-search assistance/monitoring. Also the U.K. welfare system contains an assistance policy of last resort, called Income Support. Finally, the U.K. system features a large earnings subsidy program for poor households, the Working Family Tax Credit (see Blundell and Meghir, 2001, for an overview of the U.K. system).

It is useful to note two common features of the U.K. and the U.S. Welfare-to-Work programs, shared by many other countries as well. First, active labor market policies tend to be implemented

after an initial spell of unemployment insurance. Second, policies of pure social assistance are usually ordered last in the sequence of policy interventions.

A Welfare-to-Work (WTW) program is precisely a government expenditure program that combines together several policy instruments, as in the U.S. and the U.K. examples. Clearly, every WTW program implicitly promises a certain level of ex-ante welfare to the unemployed agent. An *optimal* WTW program is a mix of policies that maximizes the expected discounted utility of the unemployed agent, subject to a government budget constraint. The fundamental trade-off to be solved in the design of an optimal WTW program is the one between offering consumption insurance and eliciting the right effort level in the search activity.

The first objective of this paper is to develop a theoretical framework suitable to study the key features of optimal WTW programs. Our point of departure is the classic setup—originated largely from the seminal article of Shavell and Weiss (1979)—where the optimal unemployment insurance contract is studied in the presence of a repeated moral hazard problem: the risk-neutral principal (planner/government) cannot observe the risk-averse unemployed agent’s job search effort (hidden action).¹ Following the most recent contributions in the literature (Wang and Williamson, 1996; Hopenhayn and Nicolini, 1997; Pavoni, 2003a), we exploit the recursive representation of the planner’s problem where the expected discounted utility promised by the contract to the unemployed agent becomes a state variable.

We generalize this standard setup in two directions. First, we enrich the economic environment of the standard setup with the introduction of a costly monitoring technology that can be used to observe the worker’s search effort.² As discussed earlier, job-search monitoring is a key component of actual WTW schemes, but the standard setup (e.g., Hopenhayn and Nicolini, 1997) is unable to address the optimality of such an important policy instrument.

Second, following Pavoni (2003b), we let workers’ wages and their job finding probabilities depend on *human capital* (skills) and allow human capital to depreciate along the unemployment spell. Human capital is our second key state variable in the recursive representation. The introduction of human capital depreciation in the problem permits a better representation of labor market data along two

¹Bailey (1978), Flemming (1978) and more recently Fredricksson and Holmlund (2001) and Acemoglu and Shimer (1999) follow a different approach. They study economies where search effort is observable (i.e., there is no moral hazard) and model unemployment insurance as the optimal policy of a benevolent government that chooses the benefit *level* (not the entire time-path) associated to the welfare-maximizing labor market equilibrium. Recently, Saez (2002) analyzed the optimality of certain income maintenance programs in the U.S. and France within static models where the worker type is private information.

²Extending Atkeson and Lucas (1995), Alvarez and Aiyagari (1995) study the optimal level of random monitoring of agents’ reports on job offers and the long-run stationary distribution of consumption in a pure adverse selection setup with temporary (one-period) job offers, and no human capital dynamics. Recently, Boone et al. (2002) have addressed the issue of the optimality of job-search monitoring within a structural search model of the labor market. They numerically compute the optimal level of monitoring in an economy where every worker is subject to the same monitoring probability, but they do not fully solve for the optimal contract where the use of each instrument is history-dependent, as we do. Their simpler framework has clearly some other advantages, for example they can study general equilibrium effects which we abstract from.

important dimensions. First, since wages depend on human capital, in our economy workers experience wage losses upon separation, consistently with the findings of a vast set of empirical studies (for a survey, see Fallick, 1996). Second, since we let the job-finding probability depend on human capital, search effort becomes less effective as the unemployment spell progresses, inducing negative duration dependence in the unemployment hazard—a common feature of the data, as reported by Machin and Manning (1999) in their survey. In particular, several studies (e.g., Blank, 1989, for welfare recipients; Bover, Arellano and Bentolila, 2002, for UI benefits recipients) continue finding a rapidly declining hazard even after explicitly controlling for unobserved heterogeneity.³

The planner’s recommendations on effort choices and on the use of the costly search monitoring technology map naturally into four distinct policy instruments: 1) unemployment insurance, where the planner elicits positive search effort from the agent; 2) job search monitoring, where the planner can observe the search effort upon payment of a cost; 3) social assistance, defined as an income-assistance program of “last resort” with the planner inducing zero search effort and simply insuring the worker; and 4) wage taxes and subsidies, generated by choosing the agent’s consumption during employment, along the lines of Hopenhayn and Nicolini (1997).

Our qualitative characterization of the optimal WTW program yields three main results. First, in absence of human capital depreciation, the optimal program does not contemplate switching between policy instruments: the program may initially select different policies for workers with different human capital levels, but each policy is absorbing within the unemployment spell. With human capital depreciation, the typical policy sequence in the optimal WTW program starts from unemployment insurance, switches into job-search monitoring, and then turns into social assistance, which remains the only absorbing policy. The faster the human capital depreciation, the more rapidly the optimal WTW program transits between policy phases. Second, we can entirely describe the dynamics of the government transfers and taxes/subsidies associated to each policy phase of the program. Under social assistance and job-search monitoring, benefits are constant because of full insurance, while due to the presence of the incentive constraint they are decreasing during unemployment insurance. Third, the nature of re-employment wage taxes and/or subsidies is largely determined by competing forces which we explain intuitively. One fairly general result is that, as the optimal program approaches the monitoring phase, earnings subsidies become more and more likely, and during job-search monitoring the subsidy tends to increase with duration. This finding is linked to human capital depreciation and contrasts with the well-known result of Hopenhayn and Nicolini (1997) that, in an economy without human capital dynamics, taxes should rise along the unemployment spell.

The second objective of the paper is to study quantitatively the features of the optimal WTW program for the typical welfare recipient in the U.S. economy and contrast them to the actual welfare

³Skill depreciation is also a central ingredient in a popular explanation of the comparative unemployment experience of the U.S. and Europe (e.g., Ljungqvist and Sargent, 1998).

system.⁴ We calibrate the parameters of our model to match some key labor market statistics. Next, we solve numerically for the optimal program and, by simulation, derive the optimal sequence of policies, their duration, the pattern of optimal benefits, taxes and subsidies. We then calculate the welfare gains for the worker (or, equivalently, the budget savings for the government) of shifting from the current scheme to the optimal scheme.

When we use our theoretical framework to study the optimality of the current U.S. welfare system, we find that the current WTW scheme should provide workers with more consumption insurance and should use the costly monitoring technology less intensively. The size of the welfare gains of switching from the current scheme to the optimal program varies across skill type. More skilled workers who rejoin the employment ranks quickly have small gains, around 0.11% of lifetime consumption; however, unskilled workers with low hazard rates can achieve gains beyond 2.9% of lifetime consumption.

The rest of the paper is organized as follows. Section 2 presents the economic environment (available to the agent in autarky). Section 3 describes the contractual relationship between planner and agent, and presents the recursive formulation of the planner’s problem. Section 4 characterizes the key features of the optimal WTW program. Section 5 develops the quantitative analysis applied to the U.S. labor market. Section 6 concludes the paper.

2 The Economy

Preferences: Workers have period utility over consumption c and effort a given by $u(c) - a$. Preferences are time-separable and the future is discounted at rate $\beta \in (0, 1)$. We impose that $c \geq 0$, and that $u(\cdot)$ is strictly increasing, strictly concave and smooth, with $\lim_{c \rightarrow \infty} u'(c) = 0$. A technical assumption that will prove useful in our characterization is that the first derivative of u^{-1} is convex. This condition is satisfied by a wide range of utility functions, including the CRRA class with risk-aversion parameter greater than one half, and the entire CARA class.⁵

Employment status and effort: The agent can be either unemployed ($z = z^u$) or employed ($z = z^e$). During unemployment, she searches for new jobs, and search effort can be either low or high, i.e., $a \in \{0, e\}$, with $e > 0$. Employment is an absorbing state with low effort ($a = 0$).⁶

Human capital: Workers are endowed with a time-varying stock of human capital (skills) $h \geq 0$. During unemployment, if search fails, human capital depreciates geometrically at rate $\delta \in [0, 1]$.

⁴There is an extended econometric literature on the evaluation of active labor market programs (for surveys, see Meyer, 1995; Heckman, LaLonde and Smith, 1999; Moffitt, 2003) which can provide useful guidance on the effectiveness of different types of interventions, but cannot address the complex question of how to design a government program which combines various instruments and follows the worker throughout her non-employment experience.

⁵Newman (1995) highlights the role of the curvature of the first derivative of the cost function for the principal (the inverse of the agent’s utility function) in a static moral hazard model of entrepreneurship.

⁶In the quantitative section we also study the case where during work the agent suffers a positive effort cost $a = e$.

During employment human capital remains constant.⁷ Let y denote the outcome of the worker search activity during unemployment, with $y \in \{s, f\}$, where s denotes “success”, and f “failure”. Then,

$$h^f = (1 - \delta) h. \tag{1}$$

Production technology: An employed worker of type h produces output $\omega(h)$. We let $\omega(\cdot)$ be a continuous and increasing function, with $\omega(h) \in [0, \omega_{\max}]$ and $\omega(0) = 0$.

Search technology: During search, both effort a and human capital h affect the job finding probability of an unemployed worker. Denote the unemployment hazard rate as $\pi(h, a)$. We assume that $\pi(h, 0) \equiv 0$ and that $\pi(\cdot, e) \equiv \pi(\cdot) \in (0, 1)$ is continuous and increasing.

Insurance markets: During unemployment the agent would like to insure against the outcome y of her search activity, which affects both her employment status and her level of human capital, thus her future income. We study the optimal contract when the worker has no access to storage, insurance or credit markets—in particular we abstract from self-insurance. We return on this point in the Conclusions.

It is useful to further discuss some of the assumptions we made in laying out our environment. The discrete effort choice is coherent with a long tradition in labor economics stressing the importance of fixed costs and the extensive margin in participation decisions (e.g., Hausmann, 1980; Cogan, 1981). The assumption that employment is an absorbing state is made, as in Hopenhayn and Nicolini (1997), to focus the analysis of the optimal dynamic contract during unemployment.⁸ The absence of effort cost on the job ensures that the value of employment always dominates the value of remaining unemployed so that no job offer is turned down. These two assumptions have no bearing on the qualitative characterization of the optimal WTW program during unemployment and can be relaxed, as long as employment remains dominant.

The monotonicity of π in h is a natural property, consistent with the overwhelming evidence that unemployment duration is longer for workers with lower pre-displacement wage (e.g., Meyer 1990). Strictly speaking, it is a “reduced form” that can be microfounded, for example, in a model where skilled workers have access to wider labor market opportunities, or in a model where productivity on the job is positively correlated with search skills. This monotonicity, together with skill depreciation, induces duration dependence in the unemployment hazard.⁹

⁷In Pavoni and Violante (2005) we assume that during employment human capital accumulates, for example through learning-by-doing. None of our qualitative results change.

⁸As long as the layoff rate is exogenous, the qualitative predictions of our model, within the same unemployment spell, are unchanged. Zhao (2001) and Hopenhayn and Nicolini (2004) introduce a job-separation probability that is under agents’ control and show that the optimal dynamic contract must also take the employment history into account.

⁹None of our analytical results relies on the *strict* monotonicity of π in h . For example, in Proposition 6, we will provide a full analytical characterization of the optimal sequence of policies for the case where π is constant (and only ω is affected by h).

Our specification of π will also display complementarity between the stock of human capital h and the search effort level a : increasing effort has a larger impact on the hazard rate the larger the level of h . Bover, Arellano and Bentolila (Table 4, 2002) document that the receipt of benefits significantly reduces the hazard rate of leaving unemployment and this effect tends to fade away as the unemployment spell progresses (and h depreciates). It is logical to associate the receipt of benefits with a lower search effort, thus their finding provides a possible foundation for our assumption.

3 The Contractual Relationship

We now introduce a risk-neutral planner/government (principal) who faces an intertemporal budget constraint with a real interest factor equal to β^{-1} . At time $t = 0$, the planner offers the unemployed worker (agent) an insurance/credit contract that maximizes the expected discounted stream of net revenues (fiscal revenues minus expenditures) and guarantees the agent at least an expected discounted utility level U_0 . The value of U_0 should be thought of as an exogenous parameter measuring the “generosity” of the welfare system (e.g., the outcome of voting).

Information structure: The planner can perfectly observe the employment status z , and the outcome y of the search activity.¹⁰ However, the agent’s effort choice a during search is private information of the agent, so the planner faces a moral hazard problem.

A search-effort monitoring technology is available to the planner when the worker seeks job opportunities: upon payment of a cost $\kappa > 0$, the job-search effort of the agent can be perfectly observed and enforced by the planner. The monitoring technology can be interpreted as the situation where the planner pays the services of a “caseworker” who closely monitors the search activity of the agent.¹¹

Contract: In each period t , the contract specifies transfers of resources to the worker, recommendations on search effort level to exert, and the choice of using the effort-monitoring technology. The period t components of the contract are contingent on all publicly observable histories up to t and, whenever the monitoring technology is not used, search-effort recommendations must be incentive compatible. Moreover, at every t , we allow the planner to specify the contract contingent on the publicly observable realization $x_t \in [0, 1]$ of a uniform random variable X_t . We will explain later how this “randomization” may be used in the optimal contract to convexify the planner’s problem and, thus, enhance welfare (see also Phelan and Townsend, 1991; and Phelan and Stacchetti, 2001).

¹⁰Since the law of motion for human capital dynamics is known and deterministic, it is enough knowing the pre-displacement wage and unemployment duration in order to recover the individual level of human capital h at every point in the worker’s history.

¹¹Our monitoring technology could be interpreted in a more general way, as “random monitoring”, with one additional assumption. Suppose that the government observes the agent’s search effort only with some probability $q < 1$. When $u(0) = -\infty$, then if the government threatens to sanction the shirking worker by not paying any benefit, this random monitoring acts exactly as a perfect monitoring, since the worker will never want to be without consumption for a period. Such monitoring activity will be essentially identical to ours whenever the monitoring cost has a fixed component bounded away from zero.

The components of the contract as policies of the Welfare-to-Work (WTW) program:

The combination of unmonitored search, monitored search, together with the high and low effort recommendations configures four possible options. Notice first that the planner will never choose to pay the monitoring cost and suggest the minimal effort level. The reason is that, since $\pi(h, 0) = 0$, the observable realization of a successful search activity perfectly detects a deviation from the zero-effort recommendation at no additional cost.¹² As a result, the planner is left with three options, which we label “policy instruments” of the WTW program, and we index with i .

We denote as “Unemployment Insurance” ($i = UI$) the joint recommendation of search activity and positive search effort. When positive search effort is suggested together with the use of the monitoring technology, the policy will be labelled “Job-search Monitoring” ($i = JM$). The zero-effort recommendation in the search activity denotes the “Social Assistance” policy ($i = SA$). Finally, during employment, the difference between the wage and the planner’s transfer defines implicitly the employment tax (if positive) or subsidy (if negative).

3.1 Recursive Formulation of the Planner’s Problem

The Technical Appendix describes the sequential formulation of the problem and explains that the problem can be written recursively through a state vector composed by the expected discounted utility U promised to the agent by the continuation of the contract and the level of human capital h of the worker (e.g. Spear and Srivastava, 1987; and Abreu, Pearce and Stacchetti, 1990).¹³

Exploiting this recursive representation, consider an unemployed worker who enters the period with state (U, h) . At the beginning of the period, the planner selects the optimal policy instrument $i(U, h)$ by solving

$$V(U, h) = \max_{i \in \{JM, SA, UI\}} V^i(U, h), \quad (2)$$

where the function V is the upper envelope of the values associated to the different policies. In choosing a particular policy, implicitly, the planner also chooses an effort recommendation $a(U, h)$, the transfer $c(U, h)$, and the continuation utilities $U^y(U, h)$ conditional on the outcome y of the search activity. We describe these additional choices in the next section.

As anticipated, the planner in general may decide to use randomizations through X . In this case,

¹²Put differently, the incentive-compatibility constraint associated to the zero effort recommendation is a trivial one, since the planner can punish without limits the worker upon finding a job, an outcome which is off the equilibrium induced by the optimal contract.

¹³Strictly speaking, the employment status z is also a state variable, but since employment is absorbing and presents no incentive problems, the contract is nontrivial only as long as the worker is unemployed and $z = z^u$. Therefore, in what follows, to lighten the notation, we omit z from the state vector.

the value function for the planner solves

$$\begin{aligned} \mathbf{V}(U, h) &= \int_0^1 \max_{U(x) \in D} V(U(x), h) dx, \\ s.t. &: \\ U &= \int_0^1 U(x) dx, \end{aligned} \tag{3}$$

where the constraint says that the planner is committed to keep his promises: he must deliver to the agent continuation utility U in (ex-ante, with respect to the shock x) expected value terms.

3.2 The Policies

We now describe in detail the planner problem during employment and during unemployment, under the four policies we considered.

Employment state (wage tax/subsidy): Consider an employed worker with state (U, h) . Since employment is an absorbing state without informational asymmetries, the planner simply solves

$$\begin{aligned} W(U, h) &= \max_{c, U^s} \omega(h) - c + \beta W(U^s, h) \\ s.t. &: \\ U &= u(c) + \beta U^s. \end{aligned} \tag{4}$$

The planner will provide full consumption smoothing for the agent, thus promised utility is constant over time. The promise-keeping constraint implies that in every period the optimal transfer c^e is constant and satisfies $c^e(U) = u^{-1}((1 - \beta)U)$. Therefore, the magnitude

$$\tau(U, h) = \omega(h) - c^e(U) \tag{5}$$

is the implicit tax (or subsidy, if negative) the government imposes on employed workers. State-contingent taxes and subsidies are a key component of an optimal WTW plan.

Unemployment Insurance (UI): When the worker is enrolled by the planner in the unemployment insurance scheme, the problem of the planner is

$$\begin{aligned} V^{UI}(U, h) &= \max_{c, U^f, U^s} -c + \beta \left[\pi(h)W(U^s, h^f) + (1 - \pi(h))\mathbf{V}(U^f, h^f) \right] \\ s.t. &: \\ U &= u(c) - e + \beta \left[\pi(h)U^s + (1 - \pi(h))U^f \right], \\ U &\geq u(c) + \beta U^f, \end{aligned} \tag{6}$$

where h^f is generated through the law of motion (1). The pair (U^s, U^f) are the lifetime utilities promised by the planner contingent on the outcomes (s or f) of search. Given the observability of

the employment status, the outcome of search is verifiable. For notational simplicity we have denoted $\pi(h, e)$ as $\pi(h)$. The first constraint above describes the law of motion of the state variable U (promise-keeping constraint), and the second constraint states that payments have to be incentive compatible. The expressions for \mathbf{V} and W are given by equations (3) and (4), respectively.

Job Search Monitoring (JM): The problem of the planner that chooses to monitor the search effort of the agent is

$$\begin{aligned} V^{JM}(U, h) &= \max_{c, U^f, U^s} -c - \kappa + \beta \left[\pi(h)W(U^s, h^f) + (1 - \pi(h))\mathbf{V}(U^f, h^f) \right] \\ \text{s.t.} &: \\ U &= u(c) - e + \beta \left[\pi(h)U^s + (1 - \pi(h))U^f \right]. \end{aligned} \tag{7}$$

Notice the similarity between problem (JM) and problem (UI): the former is identical to (UI) except for the fact that there is no incentive-compatibility constraint in exchange for the additional per period cost κ . This cost can be interpreted as the salary of the government employee (“caseworker”) who monitors and enforces the search activity of the unemployed agent, plus the additional administrative expenditures associated to this task.¹⁴

Social Assistance (SA): In social assistance, the worker is “released” by the planner for the current period, in the sense that the planner does not demand high effort, but simply transfers some income to the worker. The problem of the planner is

$$\begin{aligned} V^{SA}(U, h) &= \max_{c, U^f} -c + \beta \mathbf{V}(U^f, h^f) \\ \text{s.t.} &: \\ U &= u(c) + \beta U^f. \end{aligned}$$

The expression for \mathbf{V} is given by equation (3) and the constraint describes how the promised utility U can be delivered by a combination of current and future payments.

Below we will prove that if at any point during the contract the planner makes the “zero effort” recommendation, it is optimal to do so from that point onward: *SA* is an absorbing policy. In light of this characterization, it is natural to think of *SA* as a pure income-assistance program of last resort.

In what follows, it is convenient to state some basic properties of these value functions. By inspecting problem (4), it is easy to see that the value of employment has the following form:

$$W(U, h) = \frac{\omega(h)}{1 - \beta} - \frac{u^{-1}((1 - \beta)U)}{1 - \beta}. \tag{8}$$

¹⁴The objective of the job-search monitoring program we modelled is exclusively that of enforcing search effort. Often, these programs also aim at improving workers’ search effectiveness—for any given effort level—by providing counselling and explicit job placement services, an aspect we have neglected in our analysis. The challenge, as emphasized by Meyer (1995), is to obtain reliable data to separate these two aspects of job search assistance policies. In the rare instances where the data allow it (e.g., van den Berg and van der Klaauw, 2001, for the Netherlands), the conclusion is that the aspect we have emphasized in this paper, monitoring, is far more effective.

It is easy to see that W is a continuous function, which is increasing in h , and decreasing, concave and continuously differentiable in U .

By applying fairly standard results in dynamic programming, one obtains the same continuity, monotonicity and concavity properties for the convexified upper envelope \mathbf{V} as well, but two caveats are worth mentioning. First, monotonicity in U is guaranteed whenever at (U, h) the consumption level c associated to the optimal program is positive (e.g., whenever $u(0) = -\infty$). Second, the concavity of \mathbf{V} in U is guaranteed because of the randomization in (3). In the Technical Appendix (Proposition 0) we report the proof of all such properties. Finally, the properties of \mathbf{V} are inherited by the value functions of each single policy V^i . In particular, all the problems defining policies $i \in \{JM, SA, UI\}$ are also concave, and each V^i is continuously differentiable in U .

4 The Optimal WTW Program

We are now ready to study the key characteristics of an optimal WTW program. We begin with a discussion of the economics behind the choice among alternative policies, a useful step to understand two important preliminary results. First, without human capital depreciation the optimal program involves only one policy stage, chosen optimally among JM, SA or UI, and no transition across policies (i.e., each policy is absorbing). Second, even in the presence of human capital dynamics, social assistance maintains its absorbing nature.

We then turn to studying the optimal sequence of policies in the model with human capital dynamics. Here, we heavily exploit the recursive formulation of the optimal contracting problem. By projecting the upper envelope (2) on the (U, h) state space, we obtain a graphical representation of which policy is optimally implemented at every (U, h) pair. The state space can then be divided into different connected areas, each corresponding to a specific policy whose value dominates all the others. A key step to this characterization is ranking the slopes of the value functions $V^i(U, h)$ for each policy i with respect to its two arguments. Finally, we describe the dynamics of the government transfers associated to each policy phase of the program.

4.1 Choosing Among Policies: Economic Forces at Work

Within our model, there are several key economic forces that induce the planner to select one particular policy over the others. For ease of exposition, we divide them into costs and returns. We begin by describing the costs and then move to the returns.

First of all, a planner who wants to implement JM will have to incur certain additional expenses (κ) associated to the administration of the monitoring activity.

Second, the planner must compensate the agent for her effort. Since u is concave, the higher the promised utility U , the lower the marginal utility of consumption. Hence, the larger the payments of

the planner must be to compensate workers for their fixed disutility of the search effort cost e . This “wealth effect” makes SA more attractive, compared to UI and JM , for high enough levels of U .

The third cost component is the cost induced by the presence of the incentive compatibility constraint. By using the promise-keeping constraint, the incentive compatibility constraint during UI can be conveniently reformulated (independently of the unemployment benefit c), as

$$U^s - U^f \geq \frac{e}{\beta\pi(h)}. \quad (\text{IC1})$$

The difference between the state-contingent utilities U^s and U^f is increasing as h falls, through the hazard rate $\pi(h)$. Large utility dispersions are associated to large consumption dispersions. Since the agent is risk-averse, in order to compensate her for the wider spread of payments across states, the planner has to deliver a higher average transfer for any given promised utility U . In sum, incentive costs for the planner (i.e., resource costs of satisfying incentive-compatibility during UI) increase as human capital h depreciates.

Moreover, when the third derivative of u^{-1} is positive, the cost of providing the utility lottery associated to (IC1), in terms of consumption payments of the planner, increases with U . Hence, the incentive costs during UI increase with the level of promised utility U .¹⁵

We now turn to the returns associated to each optimal policy. Social assistance has no direct returns for the planner. In UI and JM , the returns to search in terms of job finding rate via $\pi(h)$, and earnings once employed via $\omega(h)$, are increasing in h .

4.2 The “Absorbing” Nature of Social Assistance

Consider a situation where, along some history, social assistance turns out to be optimal for the planner. Intuitively, given the absence of IC constraints, during SA the planner offers full insurance to the agent, and U is constant. Because of depreciation, however, h declines over time. As h gets smaller, the incentive cost in UI rises and the returns to search fall, hence both UI and JM become less attractive compared to SA , which reinforces the optimality of SA . This mechanism delivers the following result.

Proposition 1 (SA absorbing): *SA is an absorbing policy. That is, if it is chosen at any period t , choosing it thereafter is optimal.*

Proof: *See Appendix A.*

Proposition 1 already establishes one key property of the optimal sequence of policies in a WTW program, as it rules out programs where incentive-provision or monitoring is offered after a spell of

¹⁵In the Proof of Lemma A5, Appendix A, we provide an intuitive argument for this property.

social assistance. As discussed in the Introduction, pure income support policies (like Food Stamps in the U.S.) tend to be interventions of last resort, in accordance with our result.

A consequence of Proposition 1 is that, because of its absorbing nature, and since $\pi(h, 0) = 0$, the equilibrium value of SA does not depend on h and can be written as

$$\hat{V}^{SA}(U) = -\frac{c^{SA}(U)}{1 - \beta}, \quad (9)$$

where $c^{SA}(U) = u^{-1}((1 - \beta)U)$ is the constant benefit paid to workers in SA by the planner.

4.3 Human Capital Dynamics Are Necessary for Policy Transitions

Suppose there is no human capital depreciation (i.e., $\delta = 0$), then human capital is not a state variable (h remains always constant). The planner's values of each policy are defined as in Section 3.2, without dependence on h . In the next proposition, we show that in this case the structure of an optimal WTW program in an economy is very sharp: each policy is absorbing.

Proposition 2 (No human capital dynamics): *If $\pi(\cdot)$ and $\omega(\cdot)$ do not depend on h , every policy (JM , SA , UI) is absorbing: if policy i is chosen at the beginning of the program, choosing it thereafter is optimal. If, in addition, \mathbf{V} is strictly concave, any optimal program must possess such absorbing characteristics.*

Proof: *See Appendix A.*

Consider the problem of a planner facing an agent with initial utility entitlement equal to a level U_0 where only one policy between UI , JM and SA is implemented with certainty.¹⁶ For U_0 high enough, the search effort compensation cost is prohibitive and the planner will release the agent immediately into social assistance, which is absorbing.

Suppose now that U_0 is such that the planner decides to require the agent to supply positive search effort: the choice would be between either facing the IC constraint or paying κ to monitor the agent's effort perfectly. As the utility entitlement falls, the IC constraint becomes "cheaper" to satisfy, so for low enough initial levels of U_0 , the planner will begin by enrolling the agent in UI , while for intermediate values of U_0 the planner will choose JM as its initial policy. During UI , because of incentive compatibility, the state variable U is decreasing which reinforces the optimality of UI compared to the other available policies, since a smaller U reduces the incentive costs. During JM the agent is fully insured: since \mathbf{V} is concave, keeping the promised utility constant over time and never switching out of JM is always an optimal policy.

¹⁶In general at U_0 the optimal program might involve random assignment to different policies in period zero. The proof of Proposition 2 establishes that, in a stationary economy, period zero is the only period in which such random assignment might occur.

The same logic used in Proposition 2 can be used to rank the slopes of the value function with respect to U , a result that will be extended in the next section.

Corollary to Proposition 2: *The (negative) slopes of the value functions with respect to U satisfy*

$$\hat{V}_U^{SA}(U) \geq V_U^{JM}(U) \geq V_U^{UI}(U),$$

where the first inequality holds for any U , whereas the second inequality holds at the crossing point, i.e., at the unique (if any) U where $V^{JM}(U) = V^{UI}(U)$.

Proof: See Appendix A.

The value of unemployment insurance for the planner V^{UI} falls more steeply than JM with respect to U because of the incentive cost, and V^{JM} is steeper than \hat{V}^{SA} because of the effort-compensation cost. These properties imply that the relevant value functions, if they cross, can cross at most once.

4.4 Graphical Representation in the (U, h) Space

In this section we begin the characterization of the optimal WTW scheme in the presence of human capital dynamics. Our ultimate aim is to answer questions such as: What is the optimal sequence of policies? What are the dynamics of unemployment benefits and taxes/subsidies upon re-employment, within each policy?

The recursive formulation naturally suggests the following two-step strategy to approach these questions. We will first derive a simple graphical representation of the regions within the (U, h) space where each policy arises as optimal, a result of its own independent interest. Next, by reading the (U, h) state space as a phase diagram—whose dynamics are driven by the policy functions $U^f(U, h)$ describing the law of motion for U , and by the exogenous law of motion for human capital (1), we will be able to recover the optimal sequence of policies.

Finally, the policy functions $c^i(U, h)$ and $c^e(U)$, together with the laws of motion for the two states, fully describe the optimal sequence of unemployment benefits, as well as wage taxes/subsidies during the optimal WTW program, since $\tau(U, h) = \omega(h) - c^e(U)$.

In the Corollary to Proposition 2, we have established the relative slopes of the value functions with respect to U in the stationary case. The next proposition establishes conditions under which the result remains true for the general case with human capital dynamics.

Proposition 3 (Slopes of the value functions with respect to U): *Let $\eta(U, h)$ be the real number that solves*

$$\mathbf{V}_U(U, h) = -g'((1 - \beta)(U + \eta(U, h))),$$

where $g \equiv u^{-1}$ is the inverse utility function. Assume that $\eta(U, h)$ is non-increasing in U for every h . Then, the (negative) slopes of the value functions V^i with respect to U satisfy

$$V_U^{UI}(U, h) \leq V_U^{JM}(U, h) \leq \hat{V}_U^{SA}(U) = -g'((1 - \beta)U) \quad \text{for all pairs } (U, h).$$

Proof: See Appendix A.

Given the convexity of g' , the required condition essentially establishes an upper bound on the curvature of \mathbf{V} which is, for example, satisfied by log-utility.

In the next proposition we establish an intuitive ranking on the slope of the value functions V^i across the different policies $i = JM, SA, UI$, with respect to human capital h .

Proposition 4 (Slopes of the value functions with respect to h): Assume that the value function \mathbf{V} is submodular and that both \mathbf{V} and V^i are differentiable in h . Then, their slopes with respect to h satisfy

$$V_h^{UI}(U, h) \geq V_h^{JM}(U, h) \geq \hat{V}_h^{SA}(U) = 0 \quad \text{for all } U, h.$$

Proof: See Appendix A.

The logic of this proposition is that V^{UI} is steeper than V^{JM} because incentive costs rise with h and V^{JM} is steeper than \hat{V}^{SA} (invariant to h) because of the returns to search.

To understand the role of submodularity, recall that in the twice-differentiable case submodularity means $\mathbf{V}_{U,h}(U, h) \leq 0$. The shape of \mathbf{V} is generated by two contrasting forces. First, “within-policy” there is a tendency toward *supermodularity*: an increase in h raises $\pi(h)$ and reduces the marginal cost of delivering a given level of utility U since incentives are more easily provided. However, a high h makes it more attractive to implement active policies like *JM* or *UI*, and we saw that search-intensive policies have more negative slopes with respect to U . This “between-policy” force that tends to generate *submodularity* of \mathbf{V} hinges on the fact that the expected flow return $\pi(h)\omega(h)$ is an increasing function of h . The assumption in Proposition 4 holds whenever the second force dominates the first, for example, when $\pi(\cdot)$ is constant and the wage increases with skills, a case that will be analyzed in Proposition 6.¹⁷

Given our characterization of the relative slopes of V^i with respect to U and h , when the upper envelope $V(U, h) = \max_i V^i(U, h)$ is projected onto the (U, h) space, as done in Figure 1, we obtain immediately the regions in the state space where each policy emerges as optimal.¹⁸ Note that in

¹⁷General conditions on the primitives for \mathbf{V} to be submodular are hard to establish. One technical reason is that the nature of the *max* operator is to preserve *supermodularity*, but not necessarily *submodularity* (e.g., see Hopenhayn and Prescott, 1992).

¹⁸Lemma A7 in Appendix A shows that under the conditions of Proposition 4 the functions V^i are totally differentiable. If we let $d = (d_U, d_h)$ be any direction of movement in the (U, h) space, under the assumptions of Propositions 3 and 4, the slopes can also be ranked for all directions d with $d_h \leq 0 \leq d_U$ since $V_d^{UI}(U_0, h_0) \leq V_d^{JM}(U_0, h_0) \leq \hat{V}_d^{SA}(U_0) \leq 0$. In the case $d_h \geq 0 \geq d_U$, all the inequalities are reversed.

the Figure these regions appear to be connected.¹⁹ We start by interpreting Figure 1 as we move “horizontally” in the (U, h) space, i.e., we let U change for a given h . Next, we study the optimal policies as we move “vertically” through the diagram, i.e., we change h for a given level of utility entitlement U .

Moving horizontally (along U): Given any h , start from the highest utility level in the diagram. For high enough U , compensating the agent for the high effort is prohibitively costly, and SA is optimal. Moving leftward, as we decrease U the effort compensation cost falls and it becomes optimal to choose a program with high-effort requirement. For intermediate levels of U , the incentive cost is still high and the value of JM dominates the value of UI . As we keep decreasing U , gradually the planner finds it more profitable to face the incentive cost rather than pay the fixed monitoring cost κ , and UI becomes optimal.

Moving vertically (along h): Given any U , for high levels of h (i.e., high $\omega(h)$ and $\pi(h)$), returns from search are high and incentive costs are low, so UI is optimal. Moving downward, as h falls incentive costs increase and the planner finds it optimal to pay the monitoring cost and implement JM . For very low levels of h , the returns to search are so low that the planner prefers to save the effort-compensation costs as well, and SA is the optimal program.²⁰

4.5 Optimal Sequence of Policies

The optimal sequence of policies is dictated by the evolution of the state vector (U, h) . Conditional on unemployment, h declines monotonically, that is, whenever search is unsuccessful we have $h^f \leq h$. The evolution of U depends on the specific policy. Formally, we have the following proposition:

Proposition 5 (Optimal policy sequence): *Assume that \mathbf{V} satisfies all properties required in Propositions 3 and 4, and recall that \mathbf{V} is a concave function. If \mathbf{V} is either strictly concave or strictly submodular or both, then, if at some period t the optimal program selects JM , then next period it is always optimal to either repeat JM or switch to SA . In particular, an optimal WTW program never switches from JM into UI .*

Proof: *See Appendix A.*

It is straightforward to show two properties of the dynamics of U . First, because of its absorbing nature and full insurance, during SA the continuation utility U is constant. Second, during UI , the

¹⁹We consistently found this feature in our simulations. One can formally show that both SA and UI areas must constitute connected sets. Our formal characterization of the (U, h) space, however, cannot rule out the possibility of a non-connected region for JM .

²⁰Figure 1 in the Technical Appendix describes the optimal use of lotteries. It appears that lotteries are mostly used along the borders of the policy regions, in particular between JM and SA , whereas they are only seldom used between UI and JM .

utility entitlement U^f promised by the planner to the unemployed worker when search fails tends to decline monotonically to satisfy the incentive constraint.

Third, perhaps surprisingly, during JM the utility entitlement of the agent will *increase* (see Lemma A8 in Appendix A). Because of full insurance, during JM the planner equates marginal costs intertemporally. Under submodularity, this implies that lower levels of human capital are optimally associated with higher levels of lifetime utility. Mechanically, as h decreases along the unemployment spell, the optimal program approaches the social assistance option. Full insurance requires the benefits c to be constant between JM and SA , but once in social assistance, the agent will also save the search-effort cost e . Hence, the overall utility is higher in SA , and U^f gradually increases during JM as the implementation of SA becomes more and more likely in the program.

Combining the dynamics of h and U with our graphical representation, we immediately obtain another key qualitative characteristic of an optimal WTW program: as soon as the program enters into the JM region, it can either remain in JM or switch to SA . In particular, a transition from JM to UI is never optimal because human capital depreciation during JM worsens the incentive cost and makes UI less attractive over time.

Conditional on failure of search, the typical sequence of an optimal WTW program begins with UI followed by JM followed, in turn, by SA . Figure 1 uses the state space as a phase diagram and illustrates a simulated individual history of an unemployed worker leading to this optimal sequence of policies.²¹ Interestingly, as discussed in the Introduction, this is also the typical observed sequence of actual policies in many countries.

Of course, with the same parameterization, for higher levels of initial promised utility U_0 (and/or lower levels of h_0) the optimal program will start in JM and skip UI altogether, and for even higher levels of U_0 (and/or even lower levels of h_0) it may start and end with SA . In contrast, for parameterizations where the monitoring cost κ is especially high, JM may not arise as optimal at all and UI will be directly followed by SA .

Part of our characterization of the state space and the optimal sequence of policies relies on restrictions on \mathbf{V} , which are endogenous.²² Given the discussion following Proposition 4 on the submodularity of \mathbf{V} , it should be clear that one important special case of the model where such property is satisfied is an economy where the agent is subject only to wage depreciation during unemployment, while the hazard rate function $\pi(\cdot)$ is fixed at some constant $\pi > 0$. This case is also much easier to handle analytically, and we are able to provide a full characterization of the policy sequence by only imposing conditions on primitives of the model.

Proposition 6 (No duration dependence): *Assume that the hazard rate does not change with*

²¹The parameterization of this example is exactly as in Section 5.1.

²²Our numerical simulations have widely verified that the slopes of the value functions are those guaranteed by our assumptions on \mathbf{V} .

h and that it is fixed at $\pi > 0$. Then, if at some period t the optimal program selects JM , next period it is always optimal to either repeat JM or switch to SA . In particular, an optimal WTW program never switches from JM into UI .

Proof: See Appendix A.

It may appear surprising that even though the hazard rate is constant, JM should follow UI since we argued that the rise in incentive costs as h depreciates is associated to the fall in $\pi(h)$ and the corresponding widening of the utility spread in (IC1). There is, however, an additional reason why incentive costs go up as h depreciates that survives as long as $\omega(\cdot)$ is increasing in h . The optimal incentive provision is shaped by the tension between intra- and inter-period consumption smoothing. The planner can improve intra-period consumption insurance (across employment states) and keep incentive costs low by moving part of the punishment burden forward into the future. The emergence of SA , where U^f cannot decline further, limits the use of the inter-period margin, forcing the UI payments to be biased toward the static component of incentive provision. This effective shortening of the time horizon during UI occurs as h declines, returns to work fall and the entry in SA becomes closer (see also Pavoni, 2003b).

4.6 Optimal Benefits and Wage Taxes/Subsidies

We now turn to the qualitative characterization of the optimal sequence of the unemployment benefits and wage taxes/subsidies. We have the following result:

Proposition 7 (Payments): (i) During unemployment insurance (UI), benefits are decreasing; (ii) during job search monitoring (JM), the benefits are constant and the wage tax (subsidy) is decreasing (increasing); (iii) during social assistance (SA) benefits are constant; (iv) if π and ω do not depend on h (fixed human capital case), during UI the wage tax is increasing over time, and during JM both the benefits and the wage tax/subsidy are constant.

Proof: See Appendix A.

The need for incentive provision under UI implies that payments should decrease with unemployment duration. Benefits are constant in SA and JM because, within these policies, the absence of incentive problems allows the planner to implement full insurance.

The result on the structure of payments and taxes during UI in the model without human capital dynamics is a re-statement of Hopenhayn and Nicolini (1997) specialized to our environment. A direct consequence of (iv) is that wage subsidies are either paid at the beginning of the unemployment spell (for particular combinations of high U_0 and low h_0), or otherwise they will never be used: the government will never switch from a wage tax to a wage subsidy during the program. The presence

of human capital depreciation changes predictions in two dimensions. First, the behavior of the wage tax during *UI* becomes a quantitative issue, which will be discussed below. Second, since the expected gross wage $\omega(h)$ decreases during unemployment and c^e is constant during *JM*, the wage tax $\tau(U, h) = \omega(h) - c^e(U, h)$ must now decrease and could become a subsidy.

In order to graphically illustrate the key features of the benefits paid across the various policies, it is useful to simulate the model. The bottom right panel of Figure 2 shows the path of wage depreciation. In particular, given the complexity of the algorithm, to reduce the grid points for h , we allow depreciation to be stochastic in the numerical simulations: upon realization of the outcome $y = f$, human capital either remains constant or falls by one grid point (the h grid is spaced geometrically). The top left panel shows the behavior of the *UI* benefits as a fraction of the pre-displacement wage, and the net wage (wage minus tax, or plus subsidy) that the unemployed would earn if she found a job in that period. The top right panel depicts the implied tax/subsidy as a fraction of the current wage; the bottom left panel shows the dynamics of U^f , which are those depicted in Figure 1.

As previously discussed, benefits (consumption during unemployment) decrease during *UI* and remain constant throughout *JM* and *SA* because of consumption smoothing. The net wage (consumption during employment) first decreases and then tends to rise as *UI* approaches *JM*. There are two main reasons for these dynamics. First, as human capital depreciates, expected utility dispersion must rise to satisfy the IC constraint (*IC1*) and part of this additional dispersion is generated in terms of a wider gap in current consumption across employment states. Second, as explained above, when h depreciates, the effective time horizon of the *UI* problem shortens and the planner must give incentives creating a wider spread between consumption across states. As a result, the planner starts using heavily wage subsidies in order to reward employment and to widen the difference between *UI* payments and net wage upon job finding.²³

When the worker enters *JM*, there is complete insurance also across employment states, hence the net wage and unemployment benefits coincide and remain constant. The wage subsidy increases if and only if human capital depreciates in order to keep consumption perfectly smooth across states. In the deterministic depreciation case, the subsidy would rise monotonically during *JM*.

5 Quantitative Analysis: Assessing the Optimality of Some U.S. Welfare Programs

In this section we illustrate how our framework can be used for quantitative work. The first step is the calibration of the model to match some salient features of the U.S. labor market and of the current

²³An advantage of allowing for stochastic depreciation in the numerical simulations is that we can choose a history of human capital where for several periods (1 to 11) h is constant, as assumed by Hopenhayn and Nicolini (1997), in order to illustrate that the features of the optimal contract they emphasize arise as a particular case of our setup: benefits, net wage, and U_t^f never stop decreasing (and taxes never stop increasing) over this sub-period.

U.S. welfare system. We then simulate histories of unemployed workers with different initial skill levels h_0 undergoing the existing welfare program and choosing optimally their effort level. This allows us to calculate the initial utility entitlement $\bar{U}_0(h_0)$ promised by the current system. Next, we solve for the optimal WTW program that delivers the same expected utility and contrast its chief features to the current program. Finally, we calculate the budget savings for the government associated to switching to the optimal WTW scheme.²⁴

5.1 Calibration

The unit of time is set to one month. We focus on the period 1991-1999, since the main source of our data on job search monitoring policies refers to that period (see below). Our household of reference in the exercise is composed by two parents and two dependent children.²⁵ We restrict our attention to workers in the age range 18-50, and with at most a high-school degree.

It is useful to divide the parameters of the model into three groups. First, the preference parameters $\{u(\cdot), \beta, e\}$. Second, the labor market parameters $\{\omega(h), \delta, \pi(h)\}$. Third, the set of parameters characterizing the U.S. welfare system, i.e., program costs $\{\kappa\}$, observed policy durations $\{\bar{d}^{UI}, \bar{d}^{JM}\}$, observed payments $\{\bar{c}^{UI}, \bar{c}^{JM}, \bar{c}^{SA}\}$, and the observed tax/subsidies rates $\{\bar{\tau}\}$. Several pieces of legislation over the years have built a network of federal and state government interventions in the U.S. labor market. In Appendix C, we list in detail the major components of the U.S. welfare system, here we keep the exposition to a minimum. Table 1 summarizes parameter values.

Preferences: We pick a value for the monthly discount factor β of 0.9957 in order to match an interest rate of 5% on a yearly basis, and use a logarithmic specification for the period-utility over consumption. We assume that the disutility of work effort equals the disutility of search effort. To set a value for e , we follow a common practice in calibrated macroeconomic models. Suppose that the disutility of the fraction of time n spent working/searching is also logarithmic, and let ϕ be the preference parameter measuring the relative weight on leisure versus consumption. From the static optimality condition of the agent and the observation that the labor share is 0.60, the consumption-income ratio is 0.75 and the fraction of time worked is $n = 0.3$, one obtains $\phi = 1.87$ (Cooley, 1995). Then, the disutility of effort is $e = \phi [\ln(1) - \ln(1 - n)] = 0.67$ which represents a cost of 49% in consumption equivalent terms. Our value is within the range of the existing estimates. Hausman (1980), Cogan (1981) and Eckstein and Wolpin (1993) compute participation costs of, respectively, 27%, 41% and 62% in consumption equivalent terms ($e = 0.31, e = 0.51$ and $e = 0.97$), within female labor force participation models. Being based on static models, the Hausman and Cogan's values

²⁴In the numerical solution, we discretize h while we treat U as a continuous variable. We assume that h can take only values over a discrete set and choose a grid with 30 points geometrically spaced over the range $[0, 20]$. The grid for U has 500 points in the interval $[90, 700]$. For given h , the value functions with respect to U are computed using Chebyshev polynomials up to the 20th order. Details of the computation procedure are available upon request.

²⁵Differently from the AFDC which was targeted to single parents, no restriction on who is to be included the assistance unit is imposed by the TANF legislation (Moffitt, 2003).

are lower bounds as they neglect the cost from non-participation due to the foregone accumulation of experience. Because they are based on models with risk-neutrality, all these estimates represent a lower bound. Since they refer to female workers, they may represent an upper bound for men. However, Keane and Wolpin (1997) estimate similar participation costs for young men, around 50% of consumption ($e = 0.69$).²⁶ At the end of section 5.2, we present a sensitivity analysis on e .

Wage function: We assume a linear (monthly) wage function $\omega(h) = h$, where human capital h is interpreted as efficiency units of labor in a competitive labor market. Thus, changes in human capital map directly into observable wage changes. We normalize h so that one unit corresponds to \$100, e.g., $h = 15$ means a monthly wage of \$1,500.

Wage depreciation: The existing microeconomic estimates of wage depreciation span a wide range. At the high end, Keane and Wolpin (1997) within a structural model estimate an average annual skill depreciation rate of 23% (36% for white collars and 9.6% for blue collars). At the other end, the average annual earnings loss upon displacement computed by Jacobson, LaLonde and Sullivan (1993, Table 3) is roughly 10%. In the middle of the spectrum, Neal (1995, Table 3) reports that an additional week of unemployment reduces the re-employment wage by 0.37%, implying an annual decay rate of 17.5%. Addison and Portugal (1989) find that a rise in duration by 10% reduces post-displacement wages by 1%. Since the average duration of unemployment corresponds to 8.5 months in their sample, they implicitly estimate a yearly skill depreciation rate of 16%. Overall, choosing a value of 15% per year seems appropriate.

The numerical load required to solve for the optimal dynamic contract in our setup is quite heavy. Thus, to reduce the number of grid points for human capital, it is useful to allow depreciation to be stochastic: workers either keep their human capital level with probability q^f , or move down one step on the human capital grid, with probability $1 - q^f$. In order to have a constant expected depreciation rate for all levels of human capital, we set a geometrically spaced grid.²⁷

Hazard function: To estimate the hazard function, we exploit information on the length of the current unemployment spell for workers who report to be unemployed in the Basic Monthly Current Population Survey (CPS). Within the window 1991-1999, the year from May 1995 to April 1996 witnessed a very stable unemployment rate, always between 5.5% and 5.7%. We choose these 12 months for our estimation in order to avoid issues of non-stationarity in the parameters. We restrict our sample to married workers with at most a high-school degree, between 18 and 50 years old.

²⁶See Hausmann (1980), page 189. See Cogan (1981), Table III and page 957. See Eckstein and Wolpin (1993), Table II for the sample means, Table IV for the estimated coefficients and equation (6) for the specification of the utility function. See Keane and Wolpin (1997), Tables 8 and 9 for estimated participation costs, and Table 4 for sample means. Further details are available upon request.

²⁷If the grid is geometrically spaced at rate Δ , and x^f is the estimated monthly depreciation rate of human capital, q^f solves $q^f = 1 - x^f/\Delta$.

Our estimation strategy follows closely the method outlined by Flinn (1986) and assumes a Weibull distribution with hazard function $\mathbf{p}(t) = \alpha\lambda t^{\lambda-1}$, where t is the length of the unemployment spell.

The estimated hazard displays negative duration dependence ($\hat{\lambda} = 0.66$). The top panel of Figure 3 shows that the bulk of the drop in the job finding probability occurs in the first 12 months, after which the hazard keeps decaying more slowly. This qualitative feature is in line with the existing empirical evidence.²⁸ The bottom panel maps the hazard $\mathbf{p}(t)$, as a function of duration, into a function of human capital $\pi(h)$, based on a depreciation rate of 15% per year and an initial earnings level equal to the median of our CPS sample ($h_0 = 15$). See Appendix B for more details on the estimation procedure.

An important issue in the estimation of the hazard function is that the identification of the structural parameters hinges on the assumption that workers observed in the data exercise search effort. There are two reasons why, in practice, this concern is not a major one. First, non-employed workers who declare no search effort in the CPS survey are considered to be out of the labor force and, as such, they were excluded from our estimation. Second, when we simulate workers' histories under the current system to compute \bar{U}_0 , we find that all workers search with high effort, except for those with $h < 4.1$ who, at times, select zero effort. In the CPS data, only 3% of our sample has monthly earnings below \$411, thus this problem seems quantitatively small in the estimation. As a further precaution, we will compare actual and optimal programs for workers initially endowed with skill levels high enough that they would only be expected to reach the region of human capital for which the estimate of the hazard function may be downward biased at a point where the optimal program recommends no search effort (i.e., social assistance), thus the hazard becomes irrelevant.

Monitoring cost: The U.S. Department of Education and the U.S. Department of Health and Human Services jointly sponsored a large-scale evaluation of welfare-to-work policies, the *National Evaluation of Welfare-to-Work Strategies* (NEWWS), based on data pertaining to over 40,000 individuals followed for 5 years, (falling between 1991 and 1999, depending on the location), in 11 separate locations. In particular, in three locations, Atlanta (GA), Grand Rapids (MI), and Riverside (CA), monitored job search was the first mandatory activity for all clients. An average across these three locations reveals that the job search monitoring cost is roughly \$478 per month/per worker (NEWWS, 2001, Tables 13.2 and 13.3). This is the value we adopt for κ .

Policy durations: As they become unemployed, for the first 6 months workers are entitled to unemployment insurance. At the expiration of the UI phase, workers enter the TANF regime and are subject to mandatory active labor market programs—monitored job search in our case. Based on the strictest time limits rules for TANF discussed in Appendix C, we assume that this regime lasts for 24 months. We use these observed durations to calibrate $\{\bar{d}^{UI}, \bar{d}^{JM}\}$.

²⁸For the U.S., van den Berg and van Ours (1996) report that the exit probability from unemployment falls by 30% after 3 months. Blank (1989) finds that the bulk of the decline in the welfare hazard occurs before two years. Our estimated hazard is consistent with these findings.

Table 1: Summary of the Parameterization

Parameter	Value	Moment to Match
β	0.9957	Interest rate (Cooley, 1995)
e	0.670	Fraction of time spent working (various sources)
δ	0.0135	Monthly human capital depreciation (various sources)
$\pi(h)$	see text	Unemployment hazard function (Basic CPS 1995-1996)
κ	\$478	Monthly cost of JSM (NEWWS, 2001)
\bar{d}^{UI}	6	Duration of UI in months (U.S. Department of Labor)
\bar{d}^{JM}	24	Duration of JSM in months (NEWWS, 2001)
\bar{c}^{UI}	$0.60 \cdot \omega$	UI benefits (U.S. Department of Labor)
\bar{c}^{JM}	\$867	Median welfare benefits (Moffitt, 2002)
\bar{c}^{SA}	\$397	Median allotment of food stamps (Moffitt, 2002)
$\bar{\tau}(h)$	see text	Taxes and EITC (Hotz and Scholtz, 2001)

Benefits: During UI, workers are entitled to benefits comprising a replacement ratio of 60% on their past earnings, plus food stamps. To calibrate the benefits received on the welfare rolls, we use the detailed state-level data collected by Moffitt (2002). From this data set, we compute that in 1996, on a monthly basis, the median welfare payment (TANF benefits plus food stamps) was \$867 for a family of four. When the TANF time limits are reached, we assume that households receive only food stamps as social assistance. The median monthly allotment of food stamps for a family of four in 1996 was \$397. These values allow to calibrate $\{\bar{c}^{UI}, \bar{c}^{JM}, \bar{c}^{SA}\}$.

Taxes/Subsidies: In the event individuals find employment, they are subject to the federal and state unemployment tax, and to the Medicare tax. Workers' earnings are subsidized exactly as indicated by the Earned Income Tax Credit (EITC) legislation for a family with two dependent children in the period under consideration—see Appendix C for details. These institutions determine jointly the function $\bar{\tau}(h)$.

Two key inputs of the normative analysis are the initial utility entitlement promised implicitly by the actual U.S. welfare program $\bar{U}_0(h_0)$, and the associated stream of public expenditures $\bar{V}(h_0)$. In computing these values, we follow closely the strategy suggested by Hopenhayn and Nicolini (1997). Since both employment and social assistance are absorbing states, by backward induction it is easy to reconstruct the initial expected utility entitlement a worker with initial human capital h_0 would receive under the current U.S. welfare system and the net government expenditures associated to this level of utility.²⁹ Expenditures include benefits and wage subsidies paid to the worker (during unemployment and employment, respectively), and the costs of operating job search monitoring programs for the durations specified above, net of the tax levied on earnings.

²⁹In computing these values, we always let the worker optimally choose between high and low effort, except when she is in the job search monitoring phase of the program, where high search effort is perfectly enforced.

5.2 Results

We begin by studying the optimal sequence of policies and payments for two types of workers, a skilled worker with $h_0 = 15$ and an unskilled worker with $h_0 = 10$, corresponding roughly to the median and the 25th percentile of the monthly wage distribution in our sample of workers with at most a high-school diploma—respectively \$1,540 and \$1,020. The average public expenditure per month (during the entire lifetime of the worker) implicit in the current system amounts to \$142 for the skilled type and \$283 for the unskilled type, suggesting that unskilled workers are more expensive, mainly due to their lower job-finding rates.

Figure 4 contrasts some key features of the optimal WTW program to the features of the actual welfare system for these two types of workers. Since the evolution of h is stochastic, in order to illustrate the main quantitative features of the optimal WTW program, for each type of worker we generate 25,000 histories of human capital shocks, and shocks to search outcomes (success/failure). We then calculate sample averages of the optimal time-path of payments upon unemployment, taxes/subsidies upon employment, and compute the fraction of workers assigned to each policy since displacement.

Policy sequence: The two left hand side panels display the fractions of skilled and unskilled workers assigned to the various policies at each unemployment duration. The optimal program starts off workers in *UI*, exactly like the actual scheme. As human capital depreciates, workers begin to be gradually moved to *JM*, and the fraction of workers in *UI* decreases while that in *JM* steadily increases. After roughly 1.5 years of unemployment, the fraction of workers in *JM* also begins to fall as the flow from *JM* into *SA* more than counterbalances that from *UI* into *JM*. For sufficiently long durations, all unemployed workers end up in *SA*. The typical policy sequence of the optimal WTW program is therefore UI-JM-SA for both types of workers.

For skilled workers, the median duration of each stage is 25 months for *UI* and 8 months for *JM*. For unskilled ones, the median duration of each policy phase is 25 months for *UI* and 6 months for *JM*. The reason for this lower transition rate across policies lies in the shape of the hazard function. As is clear from Figure 3, since for high levels of human capital the hazard is much steeper, skill depreciation induces a bigger rise in the incentive cost for skilled types, making *JM* more attractive.

Recall that in the actual program, UI lasts for 6 months and JM for 24 months. Therefore, the median duration of the pre-SA phase of the optimal program is remarkably close to the actual one, estimated at 30 months. However, the optimal WTW program features longer unemployment insurance and shorter job search monitoring compared to the actual scheme. So the actual scheme overspends in monitoring, whereas a well designed UI system that uses contingent payments achieves the same objective (i.e., to induce high search effort) more cheaply.

Unemployment benefits and wage subsidies: The upper-middle panel shows that, for skilled workers, the average optimal replacement ratio for welfare benefits (the solid line) is more generous

than the actual scheme (the dotted line). The optimal payments decrease smoothly from 96% to 75% of the pre-displacement wage, while the payments in the existing U.S. program never exceed 86% of the pre-displacement wage for our benchmark worker, and drop to 26% after 30 months. This higher benefit level of the optimal WTW program is entirely linked to the expected initial utility entitlement offered by the current system, but the smoother decline is a feature of optimality.³⁰

In the upper-right panel, we compare the optimal and actual structure of earnings subsidies as a fraction of the re-employment wage. Note that the dependence of the actual subsidy on duration is not directly imposed by the legislation, but it occurs anyway through the depreciation of earnings power during the unemployment spell.³¹ The optimal scheme traces quite closely the actual subsidy (between 7% and 15%) for the first 1.5 years, but subsequently it increases much less with duration, so for long durations the optimal subsidy ends up being around half of the actual one. The presence of a subsidy in the optimal program is in sharp contrast with what Hopenhayn and Nicolini (1997) found for the stationary model without human capital depreciation. In Section 4.4 we discussed in detail the mechanism generating such result.

In the bottom-middle and bottom-right panels of Figure 4 we report benefits and subsidies for the unskilled workers with pre-displacement skill level $h_0 = 10$. Note that optimal and actual benefits are very close for the first 30 months or so, but afterwards the decline of actual benefits is much more pronounced. The actual subsidy upon re-employment is more generous than the optimal one, roughly by a factor of 2 at every duration, and is increasing, while the optimal one is roughly flat.

To conclude, for both types of workers in our group, the current system exceeds in providing incentives. This occurs through the static margin (across states) by over-rewarding employment for the unskilled and over-punishing unemployment for the skilled worker, as well as through the dynamic margin (over time) by letting benefits decline too rapidly along the unemployment spell. In doing so, the actual program does not provide enough insurance. Put differently, high search effort can be still induced by offering more insurance.

Budget savings and welfare gains: The government budget savings are calculated by comparing the actual expenditures $\bar{V}_0(h_0)$ to the expenditures $\mathbf{V}(\bar{U}_0(h_0), h_0)$ that the planner would incur by delivering utility $\bar{U}_0(h_0)$ under the optimal program. For our skilled workers, cost savings are only of the order of 1.2%, whereas for unskilled workers they are 10.7% and they can be largely achieved by less generous subsidies. The small budget saving for skilled workers is due to the fact that they quickly rejoin the employment ranks (over 80% find a job within 6 months) and, upon employment, the actual wage subsidy is quite close to the optimal subsidy at short durations. The larger gain for

³⁰Also Hopenhayn and Nicolini (1997) uncovered very high replacement rates, close to 100%, in the optimal UI scheme (with employment taxes) that offers the same discounted utility as the actual program.

³¹The tax/subsidy paid by the current U.S. scheme is computed by applying the rules for taxes and EITC to each simulated history, and averaging out across histories. The calculation of the optimal tax/subsidy is similar, except that we excluded observations where the worker is in *SA* since it is an absorbing state with no re-employment possibility.

Table 2: Sensitivity Analysis ($h_0 = 15$)

Effort Level e	Median Duration		Payments			Subsidy			Budget Savings
	(months)		(fraction of initial wage)			(fraction of wage)			
	UI	JM	$t = 1$	$t = 15$	$t = 30$	$t = 1$	$t = 15$	$t = 30$	
0.84	4	8	0.92	0.82	0.80	0.02	-0.05	-0.08	8.1%
0.67	24	9	0.98	0.87	0.81	0.07	0.13	0.15	1.2%
0.50	51	4	0.99	0.94	0.87	0.05	0.24	0.39	0.9%

low-skill individuals suggests instead that, for this group, the existing scheme could be substantially improved through a less intense use of monitoring, and a transfer of resources from earnings subsidies (EITC) to unemployment insurance (UI system). More precisely, for the unskilled individuals the lower monitoring costs account for 23.6% of budget savings, whereas the rest is obtained by delivering the same expected utility with less consumption dispersion, hence lower average payments.³²

This thought experiment keeps worker’s welfare unchanged between current and optimal program and measures the efficiency gains from the point of view of the government. An alternative way to assess the distance between actual and optimal scheme is through the eyes of the worker. The welfare gains for workers are computed by comparing the actual utility entitlement of the current system $\bar{U}_0(h_0)$ with the level $U_0(h_0)$ that the planner can deliver by spending exactly as much as the actual program on the optimal scheme, i.e., $U_0(h_0)$ solves the equation $\mathbf{V}(U_0(h_0), h_0) = \bar{V}(h_0)$. The welfare gain is then expressed in terms of fraction of lifetime consumption. For high-skill workers, the welfare gain is 0.11% of lifetime consumption. For low-skill workers, the welfare gain of switching to the optimal program amounts to 2.93% of lifetime consumption.

Our benchmark economy is one where agents do not save. The relevance of our welfare calculations depends on how large, empirically, wealth holdings are relative to the expected income loss from unemployment. Gruber (2001) finds that for over 1/3 of the unemployed, savings cannot even replace 10% of their income loss and, in line with this result, Browning and Crossley (2001) find that for 10% of the unemployed the elasticity of consumption expenditures to UI benefits is 0.8 (it is one in our model). Given that our calculations refer to a group of particularly disadvantaged workers (i.e., the bottom half of the earnings distribution of those with at most a high-school degree) we expect our welfare estimates to be quite accurate, even though they represent an upper bound.

Sensitivity analysis on effort: Based on the above discussion, in Table 2 we report the key features of the optimal WTW program for our skilled worker under two alternative effort costs representing perturbations of 25% on the baseline value ($e = 0.67$).

³²As explained when we discussed the hazard function estimation, whenever the optimal program recommends high effort, the worker under the current program would also actively search. There are states, though, where the optimal scheme chooses SA for the worker (and no effort), while under the current program the unemployed would keep searching with high effort because of the larger consumption dispersion imposed by the current program across employment and unemployment states. In our calculations, the budget saving associated to this reduction in effort compensation cost for the planner is negligible, even after accounting for the foregone returns from search effort.

Reducing the effort cost decreases the incentive cost and makes JM less attractive compared to UI. At the same time, it lowers the effort compensation cost, making UI and JM more attractive relative to SA. As a result, the median duration of JM is shorter compared to UI, but the median duration of pre-SA policies increases from 33 months in the baseline to 55 months. Increasing the effort disutility has the opposite effect. The dynamics of the payments are quite stable across experiments. At short durations, payments decline more steeply for the high effort cost as required by the incentive compatibility constraint (*IC1*), whereas at long durations payments fall more steeply for the low effort case due to the less intense use of JM.

In the high effort cost case, the optimal subsidy falls along unemployment duration and becomes a tax, in contrast to the other two parameterizations. The reason is that when the disutility of effort is high, the optimal program transits fast across policies. First, in the transition between UI and JM, the subsidy tends to shrink because there is no need to provide incentives. Second, human capital depreciation induces a rapid transition into SA, thus workers who are still actively searching (i.e., not in SA) at long durations are those whose human capital has not depreciated. Recall from Figure 2 that when h remains constant, due to dynamic incentives, the tax should increase over time.

In comparison to the benchmark, budget savings are quite similar in the low effort case while they are larger almost by a factor of six in the highest effort disutility example: when the effort is so costly to compensate, the optimal program makes heavier use of SA where high utility can be delivered cheaply by inducing the worker to forego her search effort.

6 Concluding Remarks

Welfare-to-Work (WTW) programs combine various labor market policies in an attempt to solve a complex trade-off between providing insurance to jobless workers and offering an incentive structure that will move them quickly among employment ranks.

In this paper we have provided a theoretical framework to study WTW programs from a pure normative standpoint. We have tackled a large set of important questions, such as: What is the optimal sequence of policies in an optimal WTW program? And, how long should each policy stage be? What is the optimal level and dynamics of payments in each phase of the program? Should wages upon re-employment be taxed or subsidized? Our theoretical characterization offers sharp answers to some of these questions, but only general guidelines to other questions. In this latter case, we showed how a numerical analysis does an exhaustive job and can be used for quantitative policy analysis.

Overall, our work represents a first step toward a better understanding of the optimal design of WTW programs. As such, it has a number of limitations that future research should address.

By introducing a separation probability during employment, one can make both labor market states transient and characterize the stationary distribution of workers over the state vector (U, h, z) . This

more general problem would also allow one to characterize the optimal degree of cross-subsidization between workers with different skill levels. A richer model would also incorporate incentive problems during employment and assign a more prominent role to the firm in the analysis, as in Zhao (2001) and, more recently, Blanchard and Tirole (2004). Looking further ahead, equilibrium effects of large-scale programs should be considered in the analysis.

An important set of policies which are missing from our analysis are formal training and on the job training (see Heckman, LaLonde and Smith, 1999, for a comprehensive evaluation of training programs), as well as publicly funded transitional jobs (see Kirby et al., 2002). In Pavoni and Violante (2006a) we introduce these additional instruments in our framework.

Even though we compare quantitatively the optimal WTW program to the current U.S. welfare scheme, we have not questioned in depth the role of the government in providing insurance to the unemployed when the latter can access credit markets. In Pavoni and Violante (2006b), we show that when agents have *anonymous access* to credit markets, but face a no-borrowing constraint, the optimal contract outlined here can be implemented with a simple additional instrument: an interest tax on savings. When a tax on savings is ineffective (e.g., in the presence of hidden storage), Ábrahám and Pavoni (2004) prove that, when the source of moral hazard is hidden action (hence, different from the hidden information case of Cole and Kocherlakota, 2001), the constrained-efficient allocations improve upon self-insurance so, at least qualitatively, in our economy there is scope for government-sponsored programs. Quantitatively, the size of these welfare gains remains to be established.

Finally, one could choose to answer the set of questions we laid out from a different angle, i.e., adverse selection instead of moral hazard. Unobserved heterogeneity in “types” of workers (e.g., with high/low productivity, or high/low job finding rate) will induce the planner to optimally screen workers through a menu of alternative programs and payments. We defer all these considerations to ongoing and future research.

References

- [1] Ábrahám, Á. and N. Pavoni (2004a), “Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending,” *Duke Economics Working Paper* 2004-05.
- [2] Abreu D., D. Pearce and E. Stacchetti (1990), “Towards a Theory of Discounted Repeated Games with Imperfect Monitoring,” *Econometrica*, 58(5), 1041-1064.
- [3] Acemoglu D. and R. Shimer (1999), “Efficient Unemployment Insurance,” *Journal of Political Economy*, 107(5), 893-928.
- [4] Addison J., and P. Portugal (1989), “Job Displacement, Relative Wage Changes and Duration of Unemployment,” *Journal of Labor Economics*, 7(3), 281-302.
- [5] Alvarez F., and R. Aiyagari (1995), “Efficient Dynamic Monitoring of Unemployment Insurance Claims,” *mimeo* University of Chicago.
- [6] Apostol, T. M. (1974), *Mathematical Analysis*, Addison-Wesley, Second Edition.
- [7] Atkeson A., and R. E. Lucas (1995), “Efficiency and Equity in a Simple Model of Efficient Unemployment Insurance,” *Journal of Economic Theory*, 66, 64-88.
- [8] Bailey, M. N. (1978), “Some Aspects of Optimal Unemployment Insurance,” *Journal of Public Economics*, 10, 379-402.
- [9] Blanchard O., and J. Tirole (2004), “The Optimal Design of Unemployment Insurance and Employment Protection: A First Pass,” NBER Working Paper 10443.
- [10] Blank, R. (1989), “Analyzing the Duration of Welfare Spells,” *Journal of Public Economics*, 39(3), 245-73.
- [11] Blundell R., and C. Meghir (2001), “Active Labour Market Policy vs Employment Tax Credits: Lessons from Recent UK Reforms,” *Swedish Economic Policy Review*, 8, 239-266.
- [12] Boone J., P. Fredriksson, B. Holmlund, and J. van Ours (2002), “Optimal Unemployment Insurance with Monitoring and Sanctions,” CEPR Discussion Papers 3082.
- [13] Bover O., M. Arellano, and S. Bentolila (2002), “Unemployment Duration, Benefits Duration and the Business Cycle,” *Economic Journal*, 112, 223-265.
- [14] Browning M., and T.F. Crossley (2001), “Unemployment Insurance Benefit Levels and Consumption Changes,” *Journal of Public Economics*, 80, 1-23.
- [15] Cogan, J. F. (1981), “Fixed Costs and Labor Supply,” *Econometrica*, 49(4), 945-63.

- [16] Cole H., and N. Kocherlakota (2001), "Efficient Allocations with Hidden Income and Hidden Storage," *Review of Economic Studies*, 68, 523-542.
- [17] Cooley, T. F. (1995), "Frontiers of Business Cycle Research," Princeton, NJ: Princeton University Press.
- [18] Department of Health and Human Services (2004), *TANF: Sixth Report to Congress*, Washington D.C.: U.S. Government Printing Office.
- [19] Eckstein and Wolpin (1989), "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience," *Review of Economic Studies*, 56(3), 375-390.
- [20] Fallick, B. C. (1996), "A Review of Recent Empirical Literature on Displaced Workers," *Industrial and Labor Relations Review*, 50(1), 5-16.
- [21] Flemming, J. S. (1978), "Aspects of Optimal Unemployment Insurance," *Journal of Public Economics* 10, 403-425.
- [22] Flinn, C. (1986), "Econometric Analysis of CPS-Type Unemployment Data," *Journal of Human Resources*, 21(4), 457-484.
- [23] Fredriksson P., and B. Holmlund (2001), "Optimal Unemployment Insurance in Search Equilibrium," *Journal of Labor Economics*, 19, 370-399.
- [24] Gruber J. (2001), "The Wealth of the Unemployed," *Industrial and Labor Relations Review*, 55(1), 79-94.
- [25] Hausman, J. (1980) "The Effect of Wages, Taxes, and Fixed Costs On Womens' Labor Force Participation," *Journal of Public Economics*, 14, 161-194.
- [26] Heckman J., R. LaLonde, and J. Smith (1999), "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, Vol. 3, Ashenfelter, A. and D. Card, eds., Amsterdam: Elsevier Science.
- [27] Hopenhayn H., and J. P. Nicolini (1997), "Optimal Unemployment Insurance," *Journal of Political Economy*, 105(2), 412-438.
- [28] Hopenhayn H., and J. P. Nicolini (2004), "Optimal Unemployment Insurance and Employment History," *mimeo*, Universidad Torquato Di Tella.
- [29] Hopenhayn H., and E. Prescott (1992), "Stochastic Monotonicity and Stationary Distributions for Dynamic Economies," *Econometrica*, 60(2), 1387-1406.
- [30] Hotz V. J, and J. K. Scholtz (2001), "The Earned Income Tax Credit," NBER Working Paper 8078.

- [31] House Ways and Means Committee (1996), *1996 Green Book*, Washington, D.C.: U.S. Government Printing Office.
- [32] Jacobson L. S., R. J. LaLonde, and D. G. Sullivan (1993), "Earning Losses of Displaced Workers," *American Economic Review*, September, 83(4), 695-709.
- [33] Keane M. P., and K. I. Wolpin, (1997), "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473-522.
- [34] Kirby G., H. Hill, L. Pavetti, J. Jacobson, M. Derr and P. Winston (2002), *Transitional Jobs: Stepping Stone to Unsubsidized Employment*, Mathematica Policy Research Inc.
- [35] Ljungqvist L., and T.J. Sargent (1998), "The European Unemployment Dilemma," *Journal of Political Economy*, 106(3), 514-550.
- [36] Machin S., and A. Manning (1999), "The Causes and Consequences of Long-Term Unemployment in Europe," in O. Ashenfelter and D. Card (Eds) *Handbook of Labor Economics*, Vol. 3C, 3085-3139.
- [37] Martin, J. P. (2002), "What Works Among Active Labour Market Policies: Evidence from OECD Countries' Experiences," *OECD Economic Studies*.
- [38] Meyer, B. D. (1990), "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58(4), 757-782.
- [39] Meyer, B. D. (1995), "Lessons from the U.S. Unemployment Insurance Experiments," *Journal of Economic Literature*, 33(1), 91-131.
- [40] Moffitt, R. A. (2002), "Welfare Benefits File", <http://www.econ.jhu.edu/People/Moffitt/>
- [41] Moffitt, R. A. (2003), "The Temporary Assistance to Needy Families Program," in *Means tested Transfers programs in the United States*, R. A. Moffitt (Ed.), Chicago: Chicago University Press.
- [42] National Evaluation of Welfare-to-Work Strategies (2001), "How Effective Are Different Welfare-to-Work Approaches? Five-Year Adult and Child Impacts for Eleven Programs," U.S. Department of Health and Human Services, and U.S. Department of Education.
- [43] Neal D. (1995), "Industry-Specific Human Capital: Evidence from Displaced Workers," *Journal of Labor Economics*, 13(4), 653-677.
- [44] Newman A. (1995), "Risk-Bearing, Entrepreneurship, and the Theory of Moral Hazard," *mimeo*, Boston University.
- [45] Pavoni N. (2003a), "On Optimal Unemployment Compensation," *Journal of Monetary Economics*, forthcoming.

- [46] Pavoni N. (2003b), “Optimal Unemployment Insurance, with Human Capital Depreciation and Duration Dependence,” *University College London Discussion Papers no. 03/08*.
- [47] Pavoni N., and G.L. Violante (2005), “Optimal Welfare-to-Work Programs,” *Federal Reserve Bank of Minneapolis Discussion Paper 143*.
- [48] Pavoni N., and G. L. Violante (2006a), “Optimal Use of Training Within Welfare-to-Work Programs,” *mimeo*, New York University.
- [49] Pavoni N., and G. L. Violante (2006b), “Optimal Insurance Contracts with Moral Hazard: A Simple Implementation with Anonymous Access to the Credit Market,” *mimeo*, New York University.
- [50] Pavoni N., and G. L. Violante (2006c), “Optimal-Welfare-to-Work Programs: Technical Appendix,” available on <http://www.restud.com/supplements.htm>
- [51] Phelan C., and E. Stacchetti (2001); “Sequential Equilibria in a Ramsey Tax Model,” *Econometrica*, 69, 1491-1518.
- [52] Phelan C., and R. M. Townsend (1991), “Computing Multi-Period, Information-Constrained Optima,” *Review of Economic Studies*, 58, 853-881.
- [53] Saez E. (2002); “Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses,” *Quarterly Journal of Economics*, 117: 1039-1073.
- [54] Shavell S., and L. Weiss (1979), “The Optimal Payment of Unemployment Insurance Benefits over Time,” *Journal of Political Economy*, 87, 1347-1362.
- [55] Spear S. E., and S. Srivastava (1987), “On Repeated Moral Hazard with Discounting,” *Review of Economic Studies*, 53, 599-617.
- [56] van den Berg G.J., and B. van der Klaauw (2001), “Counseling and Monitoring of Unemployed Workers: Theory and Evidence from a Social Experiment,” CEPR Working Paper 2986.
- [57] van den Berg G. J., and J.C. van Ours (1996) “Unemployment Dynamics and Duration Dependence,” *Journal of Labour Economics*, 14, 100-125.
- [58] Wang C., and S. Williamson (1996), “Unemployment Insurance with Moral Hazard in a Dynamic Economy,” *Carnegie-Rochester Conference Series on Public Policy*, 44, 1-41.
- [59] Zhao, R. (2001), “The Optimal Unemployment Insurance Contract: Why a Replacement Ratio?,” *mimeo*, University of Illinois.

7 Appendix A: Proofs

Some steps in the proofs of this Appendix are only sketched. The full-length proofs are contained in the companion Technical Appendix (Pavoni and Violante, 2006c).

PROPOSITION 0: PROPERTIES OF THE VALUE FUNCTIONS

Proposition 0 below states formally several technical properties of the value functions which have been briefly discussed in the main text.

Proposition 0: *(i) The functions W, \mathbf{V} and V^i are concave in U (for all h). (ii) They are jointly continuous in (U, h) and monotonically increasing in h . (iii) Let $i^*(U, h)$ be the implemented policy and $c^*(U, h) > 0$ be the optimal payment at (U, h) . Then both \mathbf{V} and V^i are decreasing and continuously differentiable with respect to the first argument at (U, h) with*

$$\mathbf{V}_U(U, h) = V_U^{i^*(U, h)}(U, h) = -\frac{1}{u'(c^*(U, h))}.$$

Proof: See Technical Appendix.

PROOF OF PROPOSITION 1: SA ABSORBING

Although the result is very intuitive, the proof is non-trivial. The reason is that we must rule out situations where SA is used at an early stage of the program for the sole purpose of allowing randomizations over policies with different implicit effort levels. The statement of Proposition 1 must therefore be re-formulated in probabilistic terms. Formally, we will show the following (equivalent, in probabilistic terms to the statement of Proposition 1).

Proposition 1A: *In order for a program to be strictly optimal at node (U, h_t) , it cannot be that at such node SA is implemented with positive probability and, in period $t + 1$, a different policy is implemented after SA , with positive probability.*

Proof: For the sake of contradiction, suppose (without loss of generality) that there is an optimal plan \mathcal{W} implementing SA in period t almost surely for all x_t , and policy $i \neq SA$ in period $t + 1$ for a positive measure of shocks x_{t+1} . The stated sequence (SA followed by policy $i \neq SA$) cannot be part of an optimal program. We will show that the planner can provide the agent with the same expected utility as under \mathcal{W} at a lower cost, by designing an alternative plan \mathcal{W}' which uses a randomization between two branches. *Branch 1:* This branch is implemented with probability β (i.e. equal to the discount factor). In this branch, the new program \mathcal{W}' implements exactly the same (randomization of) policies following SA in the original program \mathcal{W} , and it delivers the same lifetime utility. *Branch 2:* With probability $(1 - \beta)$, the planner implements SA forever and transfers to the agent the same

consumption level as that paid in the first period of the original plan \mathcal{W} (recall that the \mathcal{W} plan always implements SA at time t).

By construction, the new plan \mathcal{W}' provides the agent with the same ex-ante utility as under the original program \mathcal{W} . It turns out, though, that \mathcal{W}' is also cost-reducing compared to \mathcal{W} , essentially because it uses policies $i \neq SA$ (e.g., UI or JM) for higher levels of human capital ($h_t \geq h_{t+1}$), which relaxes the incentive compatibility (IC) constraint and increases expected returns to search. See the Technical Appendix for the detailed proof. **Q.E.D.**

PROOF OF PROPOSITION 2: ECONOMY WITHOUT HUMAN CAPITAL DYNAMICS

SA: Proposition 1 shows that SA is an absorbing policy for the general case.

JM: In order to show the absorbing property of JM , note that the first order conditions are $\mathbf{V}_U(U) = V_U^{JM}(U) = \mathbf{V}_U(U^f)$. This implies that setting $U^f = U$ is optimal. As a consequence, implementing the same policy, i.e. JM , every period is part of an optimal program. Clearly, whenever \mathbf{V} is strictly concave this absorbing policy is the unique optimal one. Given the absence of the IC constraint, the absorbing nature of JM implies that consumption remains constant once the worker enters JM .

UI: The proof that UI is absorbing is quite involved. Let us start by stating the first-order and envelope conditions under UI :

$$\begin{aligned} -\mathbf{V}_U(U) &= -V_U^{UI}(U) = \frac{1}{u'(c)}, \\ -\mathbf{V}_U(U^f) &= \frac{1}{u'(c)} - \mu \frac{\pi}{1-\pi}, \\ -W_U(U^s) &= \frac{1}{u'(c)} + \mu, \end{aligned} \tag{10}$$

where $\mu \geq 0$ is the multiplier on the incentive compatibility constraint. It is useful to begin by stating the conditions under which the incentive constraint is binding and promised utility declines.

Lemma A3: *At any U_0 where UI is optimal, we must have $U^f < U_0$. Moreover, if V is strictly concave to the left of U_0 , then it must be that $\mu > 0$.*

Proof. See Technical Appendix.

It is immediate from (10) that if the incentive compatibility constraint binds, then $U^f < U_0$. When it does not bind, then the planner can fully insure the worker upon success of search. The same result can then be obtained using the constraint (IC1) with strict inequality and the full-insurance version of the continuation utility $U^s = (1 - \beta)u(c)$ into the promise-keeping constraint. Since each function

V^i is continuous by Proposition 0, if for different levels of utility different policies are preferred, then the value functions must cross each other.

Lemma A4: *For every U we have that $V_U^{SA}(U) \geq V_U^{UI}(U)$.*

Proof. See Technical Appendix.

The line of proof of this lemma follows Pavoni (2003b): in UI , a rise in U increases the incentive cost and the effort compensation cost, reducing V^{UI} , whereas SA is a policy without IC problems and where effort is not required, so V^{SA} is flatter with respect to U . Note that Lemma A3 and Lemma A4 imply that SA will not follow UI : if UI is optimal at U_0 , then next period $U^f < U_0$ (by the first part of Lemma A3) which reinforces the optimality of UI relative to SA (by Lemma A4).

Lemma A5: *Let U_0 be such that $V^{JM}(U_0) = V^{UI}(U_0)$. Then, we have $V_U^{JM}(U_0) \geq V_U^{UI}(U_0)$ so V^{UI} crosses V^{JM} from above at most once.*

Proof. See Technical Appendix.

The intuition of the proof is the following. We know that JM is an absorbing policy, so when the worker enters into JM , even though he is asked to always supply positive effort, he will never be subject to random consumption sequences. We also know that after implementing UI , an optimal program never uses policy SA . This implies that once the worker is assigned to policy UI , she will always be required to supply positive effort thereafter, as in JM . In addition to that, she will possibly face random consumption due to the IC constraint. Given our assumption that $1/u'$ is a convex function, this extra randomization on consumption will induce extra costs for the planner which will make V^{UI} more negatively sloped than V^{JM} . This statement is formally shown in Lemma A6 of the Technical Appendix. The proof of Lemma A5 heavily builds on that, and takes care of only ‘off-the-optimum’ behavior.

To understand why the convexity of $1/u'$ leads to incentive costs for the planner that increase in U , consider the two-period version of our model (hence, we abstract from h , and assume that in the last period $a = 0$). Let $g \equiv u^{-1}$ and let $C_i(U)$ be the cost for the planner of delivering utility U under policy i . Then, it is easy to see that $C_{JM}(U) = \kappa + g\left(\frac{U+e}{1+\beta}\right)$. In UI , from the (binding) IC and PK constraints,

$$U = u(c) - e + \beta \left[\pi u(c^s) + (1 - \pi) u(c^f) \right] = z + \beta u(c^f),$$

where $z = u(c)$ and, clearly, $\beta u(c^f) = U - z$, and $u(c^s) = u(c^f) + \frac{e}{\beta\pi}$. This implies that the cost is

$$C_{UI}(U) = c + \beta \left[\pi c^s + (1 - \pi) c^f \right] = g(z) + \beta \left[\pi g\left(\frac{U - z + \frac{e}{\pi}}{\beta}\right) + (1 - \pi) g\left(\frac{U - z}{\beta}\right) \right].$$

Differentiating with respect to U , and using the optimality condition $g'(z) = \pi g' \left(\frac{U-z+\frac{e}{\pi}}{\beta} \right) + (1-\pi)g' \left(\frac{U-z}{\beta} \right)$ we obtain that

$$C'_{UI}(U) - C'_{JM}(U) = \frac{1}{1+\beta}g'(z) + \frac{\beta}{1+\beta} \left[\pi g' \left(\frac{U-z+\frac{e}{\pi}}{\beta} \right) + (1-\pi)g' \left(\frac{U-z}{\beta} \right) \right] - \frac{1}{1+\beta}g' \left(\frac{U+e}{1+\beta} \right)$$

which is positive for any U if and only if g' is convex because

$$\frac{1}{1+\beta}z + \frac{\beta\pi}{1+\beta} \left(\frac{U-z+\frac{e}{\pi}}{\beta} \right) + \frac{\beta(1-\pi)}{1+\beta} \left(\frac{U-z}{\beta} \right) = \frac{U+e}{1+\beta}.$$

Thus, under this condition, the cost of satisfying the IC constraint $[C_{UI}(U) - C_{JM}(U)]$ is increasing in U .

It remains to be shown that UI cannot be followed by JM . By Lemma A3, during UI we have that $U^f < U_0$ and by Lemma A5 if UI dominates JM at U_0 , it still dominates JM at $U^f < U_0$ which concludes the proof: the only policy that can optimally follow UI is UI itself. **Q.E.D.**

PROOF OF COROLLARY TO PROPOSITION 2: SLOPES

The ranking between V^{UI} and V^{JM} at the crossing point has been shown in Lemma A5. The proof that $\hat{V}_U^{SA}(U) \geq V_U^{JM}(U)$ is based on the idea that under JM the agent is required to supply positive effort at least in the first period of JM , therefore the marginal cost for the planner of providing a given level of utility must be larger than that under SA , a policy with full insurance and $a = 0$ forever. See the Technical Appendix for details. **Q.E.D.**

PROOF OF PROPOSITION 3: SLOPES WITH RESPECT TO U IN THE GENERAL CASE

That $\hat{V}_U^{SA}(U) \geq V_U^{JM}(U, h)$ can be easily derived in the general case with human capital dynamics by following exactly the same lines of proof proposed for the Corollary to Proposition 2, without relying on any additional assumption on the curvature of \mathbf{V} .

We now show the first inequality. Consider first the envelope condition in UI at (U, h) , and let $\eta = \eta(U_{UI}^f, h^f)$ the quantity in the proposition for $h = h^f$ and $U = U_{UI}^f$. We have

$$\begin{aligned} -V_U^{UI}(U, h) &= g' \left(U - \beta U_{UI}^f \right) \\ &= \pi(h) g' \left((1-\beta) \left[U_{UI}^f + \frac{e}{\beta\pi(h)} \right] \right) + (1-\pi(h)) \left[-\mathbf{V}_U \left(U_{UI}^f, h^f \right) \right] \\ &= \pi(h) g' \left((1-\beta) \left[U_{UI}^f + \frac{e}{\beta\pi(h)} \right] \right) + (1-\pi(h)) g' \left((1-\beta) \left(U_{UI}^f + \eta \right) \right). \end{aligned} \quad (11)$$

The first row is derived using the envelope condition and the promise-keeping constraint. The second row is obtained directly from the first order conditions under UI . The equality in the last line is due

to the definition of η . Now, consider JM . We have

$$\begin{aligned}
-V_U^{JM}(U, h) &= g' \left(U + e - \beta \left[\pi(h) U_{JM}^s + (1 - \pi(h)) U_{JM}^f \right] \right) \\
&= -W_U \left(U_{JM}^s, h^f \right) = g' \left((1 - \beta) U_{JM}^s \right) \\
&= -\mathbf{V}_U \left(U_{JM}^f, h^f \right).
\end{aligned} \tag{12}$$

The equality in the first row uses the envelope condition and the promise-keeping constraint. The first and second equalities in the second row use the first order conditions under JM and the shape of W , respectively. The equality in the last row uses again the first order conditions under JM . Now, if $-V_U^{UI}(U, h) \geq -V_U^{JM}(U, h)$ we are done. We must therefore rule out the possibility that $-V_U^{UI}(U, h) < -V_U^{JM}(U, h)$. For this purpose, assume that the last inequality is in fact true. We are looking for a contradiction. First of all, recall that because of incentive constraints we have $-V_U^{UI}(U, h) \geq -\mathbf{V}_U \left(U_{UI}^f, h^f \right)$. Hence, from (12) we get $-\mathbf{V}_U \left(U_{UI}^f, h^f \right) < -\mathbf{V}_U \left(U_{JM}^f, h^f \right)$ which, from the concavity of \mathbf{V} , implies that $U_{JM}^f \geq U_{UI}^f$. From our assumption on \mathbf{V} , and the convexity of g' we get

$$-\mathbf{V}_U \left(U_{JM}^f, h^f \right) \leq g' \left((1 - \beta) \left(U_{JM}^f + \eta \right) \right). \tag{13}$$

Now, the convexity of g' also implies that condition (11), together with the Jensen's inequality, yields

$$-V_U^{UI}(U, h) \geq g' \left((1 - \beta) \left[U_{UI}^f + \frac{e}{\beta} + (1 - \pi(h)) \eta \right] \right).$$

Comparing this condition with the first line of (12), it is easy to see that whenever $U_{UI}^f + \frac{e}{\beta} + (1 - \pi(h)) \eta \geq U_{JM}^f + \eta$, we directly get a contradiction to the statement $-V_U^{UI}(U, h) < -V_U^{JM}(U, h)$, and we are done. Finally, assume instead the complementary inequality. Multiplying all by β and rearranging, it becomes

$$\beta U_{UI}^f + e < \beta U_{JM}^f + \pi(h) \eta. \tag{14}$$

Now, since from (13) and the second line of (12) we have $U_{JM}^s \leq U_{JM}^f + \eta$, it must be true that

$$e - \beta \left[\pi(h) U_{JM}^s + (1 - \pi(h)) U_{JM}^f \right] \geq e - \beta U_{JM}^f + \beta \pi(h) \eta.$$

In other words, if condition (14) is true, we get $e - \beta U_{JM}^f + \beta \pi(h) \eta > -\beta U_{UI}^f$. But comparing the first lines of (11) and (12), we get $-V_U^{UI}(U, h) \geq -V_U^{JM}(U, h)$, again a contradiction. **Q.E.D.**

PROOF OF PROPOSITION 4: SLOPES WITH RESPECT TO h IN THE GENERAL CASE

That $\hat{V}_h^{SA} = 0$ is obvious since $\hat{V}^{SA}(U, h) = -\frac{u^{-1}((1-\beta)U)}{1-\beta}$. Since for $i = JM, UI$ the value functions V^i are monotonically increasing in h , we just have to show the first inequality in the proposition. Fix U_0 and recall the planner's problem under JM . By the envelope theorem we have

$$\begin{aligned}
V_h^{JM}(U_0, h) &= \pi'(h) \beta \left[W \left(U_{JM}^s, h^f \right) - \mathbf{V} \left(U_{JM}^f, h^f \right) \right] + \\
&\quad + \beta \left[\pi(h) W_h \left(U_{JM}^s, h^f \right) + (1 - \pi(h)) \mathbf{V}_h \left(U_{JM}^f, h^f \right) \right],
\end{aligned} \tag{15}$$

where the subscript JM indicates that U^s and U^f are the optimal choices under policy JM . Consider now the UI policy. Clearly, if at (U_0, h) the incentive compatibility under UI is not binding, then $V_h^{UI}(U_0, h) = V_h^{JM}(U_0, h)$ as UI and JM solve essentially the same problem and κ^{JM} does not depend on h . In what follows, we assume the incentive compatibility is binding. Substituting the incentive compatibility into the promise keeping constraint, differentiating the value function with respect to h , and using the envelope theorem, after rearranging we obtain

$$\begin{aligned} V_h^{UI}(U_0, h) &= \beta\pi'(h) \left[W \left(U_{UI}^s, h^f \right) - \mathbf{V} \left(U_{UI}^f, h^f \right) \right] - \beta\pi'(h) W_U \left(U_{UI}^s, h^f \right) \left(U_{UI}^s - U_{UI}^f \right) \\ &\quad + \beta \left[\pi(h) W_h \left(U_{UI}^s, h^f \right) + (1 - \pi(h)) \mathbf{V}_h \left(U_{UI}^f, h^f \right) \right], \end{aligned}$$

where we used the subscript UI notation for the optimal choices under policy UI . Recall that from the IC constraint, we have $U_{UI}^s - U_{UI}^f = \frac{e}{\pi(h)}$. We now show that the above expression is lower than the expression in (15). In light of (15), since $\beta\pi'(h) \geq 0$, the desired inequality $V_h^{UI}(U_0, h) \geq V_h^{JM}(U_0, h)$ will be shown for all h if the following two conditions hold:

$$\pi(h) W_h \left(U_{JM}^s, h^f \right) + (1 - \pi(h)) \mathbf{V}_h \left(U_{JM}^f, h^f \right) \geq \pi(h) W_h \left(U_{UI}^s, h^f \right) + (1 - \pi(h)) \mathbf{V}_h \left(U_{UI}^f, h^f \right), \quad (16)$$

and

$$W \left(U_{UI}^s, h^f \right) - \mathbf{V} \left(U_{UI}^f, h^f \right) - W_U \left(U_{UI}^s, h^f \right) \left(U_{UI}^s - U_{UI}^f \right) \geq W \left(U_{JM}^s, h^f \right) - \mathbf{V} \left(U_{JM}^f, h^f \right). \quad (17)$$

We now show (16). First, the separable form of W (displayed by condition (8) in the main text) implies that the derivative W_h does not depend on U^s . Since h^f is the same, W_h is the same across the two policies, and can be omitted. Thus, we just have to show

$$\mathbf{V}_h \left(U_{UI}^f, h^f \right) \geq \mathbf{V}_h \left(U_{JM}^f, h^f \right).$$

Moreover, since \mathbf{V} is submodular it suffices to show that $U_{JM}^f \geq U_{UI}^f$. From the first-order and envelope conditions, we get

$$\begin{aligned} W_U \left(U_{JM}^s, h^f \right) &= -\frac{1}{u'(c^{JM})} = \mathbf{V}_U \left(U_{JM}^f, h^f \right), \\ W_U \left(U_{UI}^s, h^f \right) &< -\frac{1}{u'(c^{UI})} < \mathbf{V}_U \left(U_{UI}^f, h^f \right), \end{aligned} \quad (18)$$

where c^i is the optimal consumption under policy $i = UI, JM$. Now assume, for the sake of contradiction, that $U_{UI}^f > U_{JM}^f$. Since \mathbf{V} is concave, it must be that

$$-\frac{1}{u'(c^{JM})} = \mathbf{V}_U \left(U_{JM}^f, h^f \right) \geq \mathbf{V}_U \left(U_{UI}^f, h^f \right) > -\frac{1}{u'(c^{UI})},$$

which implies $c^{JM} < c^{UI}$. Then, the first-order conditions (18) and the concavity of W imply that $U_{UI}^s \geq U_{JM}^s$ which leads to a contradiction: payments and continuation utilities in JM are lower than

those under UI , and effort is required under both policies, therefore UI and JM cannot deliver to the agent the same promised utility U_0 . As a consequence, the statement $U_{JM}^f \geq U_{UI}^f$ must be true and so does inequality (16).

We are now ready to show inequality (17). The first order conditions during JM imply that $W_U(U_{JM}^s, h^f) = \mathbf{V}_U(U_{JM}^f, h^f)$. The concavity of both functions and the fact that W is flatter than \mathbf{V} for all h implies that $U_{JM}^s \geq U_{JM}^f$. Since W is decreasing in U we have

$$W(U_{JM}^f, h^f) - \mathbf{V}(U_{JM}^f, h^f) \geq W(U_{JM}^s, h^f) - \mathbf{V}(U_{JM}^f, h^f).$$

From the above inequality, in order to show (17) it suffices to demonstrate that

$$W(U_{UI}^s, h^f) - \mathbf{V}(U_{UI}^f, h^f) - W_U(U_{UI}^s, h^f) (U_{UI}^s - U_{UI}^f) \geq W(U_{JM}^f, h^f) - \mathbf{V}(U_{JM}^f, h^f). \quad (19)$$

We claim that (19) must be true since $U_{JM}^f \geq U_{UI}^f$. The proof goes as follows. Rewrite first the above inequality as

$$W(U_{UI}^s, h^f) - W(U_{UI}^f, h^f) - W_U(U_{UI}^s, h^f) (U_{UI}^s - U_{UI}^f) + W(U_{UI}^f, h^f) - \mathbf{V}(U_{UI}^f, h^f) \geq W(U_{JM}^f, h^f) - \mathbf{V}(U_{JM}^f, h^f).$$

Now, the concavity of W and the fact that $U_{UI}^s > U_{UI}^f$ imply that

$$W(U_{UI}^s, h^f) - W(U_{UI}^f, h^f) - W_U(U_{UI}^s, h^f) (U_{UI}^s - U_{UI}^f) \geq 0.$$

We are hence left to show that $W(U_{UI}^f, h^f) - \mathbf{V}(U_{UI}^f, h^f) \geq W(U_{JM}^f, h^f) - \mathbf{V}(U_{JM}^f, h^f)$, or—rearranging terms—that

$$\mathbf{V}(U_{JM}^f, h^f) - \mathbf{V}(U_{UI}^f, h^f) \geq W(U_{JM}^f, h^f) - W(U_{UI}^f, h^f).$$

But this inequality must be true since $U_{JM}^f \geq U_{UI}^f$ and \mathbf{V} is steeper than W for all h . **Q.E.D.**

Lemma A7: *If V^i admits partial derivative with respect to h at (U_0, h_0) , then it is differentiable at this point.*

Proof. Since, from Proposition 0, V^i has continuous partial derivative with respect to U , the result follows from Theorem 12.11, page 357, in Apostol (1974). **Q.E.D.**

PROOF OF PROPOSITION 5: OPTIMAL POLICY SEQUENCE

We first describe the dynamics of U during JM :

Lemma A8: *Assume that \mathbf{V} is submodular and recall that it is a concave function. If at (U, h) either one of these properties (submodularity or concavity) is strict, and the implemented policy is JM , then $U^f(x) \geq U$ almost surely for all $x \in [0, 1]$.*

Proof. The first order conditions under JM imply

$$\mathbf{V}_U(U, h) = V_U^{JM}(U, h) = \mathbf{V}_U(U^f, h^f) = \mathbf{V}_U(U^f(x), h^f)$$

for (almost) all $x \in [0, 1]$. Since $h^f \leq h$, sub-modularity implies that $\mathbf{V}_U(U, h^f) \geq \mathbf{V}_U(U, h)$. Since \mathbf{V} is concave, if either one of the properties of \mathbf{V} is strict at (U, h) , we have the desired result. **Q.E.D.**

Since $h^f \leq h$, the proof of the proposition is now trivial from the graphical representation and the utility dynamics of Lemma A8. The only important remark to make is that in this lemma not only have we established that $U^f \geq U$ in expected terms, but we have also shown that the continuation utility increases almost surely for all realizations of X . **Q.E.D.**

PROOF OF PROPOSITION 6: NO DURATION DEPENDENCE

The line of proof can be seen as an extension to that adopted for Proposition 1 (SA absorbing). For the sake of contradiction, assume without loss of generality that at node (U_t, h_t) there is an optimal plan \mathcal{W} implementing JM in period t almost surely for all x_t , and UI in period $t+1$ for a full measure of shocks. We allow for any random plan from period $t+1$ onward.

The idea of the proof is again that we can construct an alternative plan \mathcal{W}' delivering the same ex-ante expected utility to the agent, and same net returns for the planner. Next, we show that in the new plan the IC constraints are relaxed with respect to \mathcal{W} in some states, thus costs can be further reduced and \mathcal{W} cannot be optimal.

The program \mathcal{W}' consists of two branches. *Branch 1* is implemented with probability $\beta(1-\pi)$. To construct this branch, we start from period $t+1$ and look at what policy is implemented in period $t+1$ in the original program \mathcal{W} . Each time in the original plan we see UI or JM we retain UI or JM , but whenever in the old plan we see the sequence $SA \rightarrow SA \rightarrow SA\dots$ (notice that from Proposition 1 this is the only possibility) then we substitute this sequence with $JM \rightarrow SA \rightarrow SA\dots$. Next, we shift the whole plan backward by one period, so this branch will start with UI at time t . *Branch 2* has probability $1 - \beta(1 - \pi)$. To construct this branch, we replicate branch 1 with the following amendment: each time in branch 1 we see UI , we replace it with JM . In particular, since branch 1 starts with UI , in branch 2 the new contract \mathcal{W}' starts with JM at time t .

Consider now the payments in plan \mathcal{W}' . Let c_t be the payment made under \mathcal{W} in period t (notice that because of full insurance during JM , this must be the same for all x_t and across states). In branch 1 we have the following. First, every time the new plan implements UI the planner pays the agent exactly what was paid to her in the old plan during this policy at this node, both in case of success and of failure of search. Whenever \mathcal{W}' implements SA the agent gets exactly the same consumption as in the old plan at that node. The payments under JM are as follows: whenever JM is implemented in order to replace JM in the old plan, the payments are again exactly the same as those in the old

plan, in all states. When JM is implemented (for one period) in order to replace the first period of SA instead, then the transfers are those made in SA in the old plan at that node. In branch 2, the new plan transfers c_t to the agent in any period, history of shocks, and states. Her expected utility is $u(c_t) / (1 - \beta(1 - \pi))$.

It is now straightforward to check that the stated plan \mathcal{W}' delivers the same ex-ante utility to the agent, and the same net return to the planner. We begin by showing the equivalence of the two programs with respect to the monitoring costs for the planner, in two steps.

Step 1: In branch 1, with weight $\beta(1 - \pi)$, the only additional monitoring costs with respect to the continuation of the old plan \mathcal{W} is borne whenever the latter contemplated SA for the first time. Let q be the probability of such event. Note now that this same additional cost occurs also in the second branch, which has weight $1 - \beta(1 - \pi)$. So, up to now, we conclude that the new plan yields additional costs with respect to the $t + 1$ -continuation of the old plan equal to $q\kappa$.

Step 2: Recall that in the first period of program \mathcal{W} the planner pays an initial monitoring cost κ with certainty. Consider now branch 2. If the plan always implemented JM , we would have a cost κ with probability $(1 - \pi)$ every period, so we would get precisely a discounted present value of the monitoring costs equal to κ , which is the initial cost of the old plan. But the cost in branch 2 is smaller since there is a tail, occurring with probability q , where the cost is not paid because the new plan picks SA from then onward. The reduction in cost associated to this tail, compared to the unlimited implementation of JM , is $-q \frac{\kappa}{1 - \beta(1 - \pi)}$ which, weighted with the probability of branch 2 occurring, yields exactly $-q\kappa$, the excess cost of the new plan computed in *Step 1*, which proves that the expected discounted monitoring cost in the two programs must be the same.

The wage returns for the planner must also be the same. The new program implements positive search effort if and only if a positive effort was implemented in the original program for the same level of h , with exactly the same probabilities as in the old plan. In particular, recall that the new plan starts in period t but in both branches it is constructed based on what happens in the original plan at date $t + 1$. Implementing JM for one more period whenever the old plan implemented SA guarantees precisely that $a = 0$ is chosen in the new plan only for values of human capital for which the old plan recommended $a = 0$. The logic used above for the monitoring cost can then be easily extended to also show that - for each h_t - the 'discounted' probability for which either $a = 0$ or $a = e$ is implemented is the same between the two plans. This implies that the agent's effort cost is the same in the two plans as well. See the Technical Appendix for details.

Moreover, the payments are the same in the two programs. Branch 1 of the new plan can be thought of as the continuation of the old plan from time $t + 1$ onward, since it has weight $\beta(1 - \pi)$. It is easy to see that the total value of payments is exactly the same as that paid in the old plan from unemployment duration $t + 1$ onward. In branch 2 payments are like in the old plan, so the discounted present value of payments in branch 2 equals $c_t / (1 - \beta(1 - \pi)) = c_t + \beta\pi \frac{c_t}{1 - \beta}$, the consumption paid

in the old plan during the first period of unemployment and upon employment. Precisely the missing part in order to get the total payment bill.

Finally, notice that under \mathcal{W}' the incentive constraint in UI is relaxed with respect to \mathcal{W} , since \mathcal{W}' reduces utility U^f upon failure of search. This is so since \mathcal{W}' implements JM instead of SA for one more period at later dates, and JM involves a positive effort cost. This allows the planner to further improve upon \mathcal{W}' by reducing consumption dispersion (hence, average transfers). The Technical Appendix contains a fully detailed proof of this Proposition. **Q.E.D.**

PROOF OF PROPOSITION 7: PAYMENTS

(i) The behavior of the payments during UI is easily derived from a set of first-order conditions analogous to (10), i.e.,

$$\begin{aligned} -V_U^{UI}(U) &= \frac{1}{u'(c)}, \\ -\mathbf{V}_U(U^f, h^f) &= \frac{1}{u'(c)} - \mu \frac{\pi(h)}{1 - \pi(h)}, \\ -W_U(U^s, h^f) &= \frac{1}{u'(c)} + \mu, \end{aligned}$$

with $\mu \geq 0$. Since next period the envelope condition is

$$-\mathbf{V}_U(U^f, h^f) = \frac{1}{u'(c^f)}$$

where c^f is the consumption payment in case of unsuccessful search, the result follows immediately from the strict concavity of u .

(ii) The first-order conditions and the envelope condition during JM are

$$\begin{aligned} -V_U^{JM}(U, h) &= \frac{1}{u'(c)}, \\ -\mathbf{V}_U(U^f, h^f) &= \frac{1}{u'(c)} = \frac{1}{u'(c^f)} \\ -W_U(U^s, h^f) &= \frac{1}{u'(c)} = \frac{1}{u'(c^s)}, \end{aligned}$$

hence unemployment payments and net wage c^s are constant during this policy. The wage tax (subsidy) is defined as $\omega(h^f) - c^s$. Since $\omega(h^f)$ increases with h^f , and h^f increases with h , the result is immediate.

(iii) It is obvious and the proof is omitted.

(iv) The shape of W in (8) implies that the net wage c_t^s satisfies $c_t^s = (1 - \beta)U_t^s$, thus the tax in this case is simply $\omega - (1 - \beta)U_t^s$. In JM promised utility is constant, so the result follows easily.

Consider UI . Recall that in the proof of Proposition 2 we have shown that in an economy without human capital dynamics UI is absorbing and the incentive compatibility constraint is binding ($\mu > 0$). From Lemma A3, $U^f < U$. Rearranging the promise-keeping constraint and the incentive constraint over two consecutive periods, we get $U_{t+1}^s = U_{t+1}^f + \frac{e}{\beta\pi} < U_t^f + \frac{e}{\beta\pi} = U_t^s$. The desired result is now immediate since $U_{t+1}^s < U_t^s$. **Q.E.D.**

8 Appendix B: Estimation of the Hazard Function

Data Description: From the Basic Monthly Current Population Survey (CPS), we selected every married worker between 18-50-years-old with at most a high-school degree who reports to be unemployed during the months May 1995-April 1996. This initial sample includes 6,920 observations. However, part of these observations refer to the same individual and the same unemployment spell, since some of the households are interviewed for several consecutive months. In these cases, we selected only the most recent available record on the length of each unemployment spell. The final sample comprises $N = 5,553$ ongoing spells of unemployment. The average duration is 3.82 months.

Estimation: The statistical problems with this type of data are two: right censoring and length bias. We follow Flinn (1986) in dealing with these two issues. We assume that the true distribution of unemployment spells is a Weibull with scaling parameter α and shape parameter λ . Let F be the Weibull distribution function, i.e. $F(t) = 1 - \exp(-\lambda t^\alpha)$, and Γ the Gamma function. The log-likelihood function \mathcal{L}_N for a size- N sample of right-censored, length-biased spells is then

$$\mathcal{L}_N(\alpha, \lambda) = \sum_{i=1}^N \ln[1 - F(t_i|\alpha, \lambda)] + N \ln \alpha + N \ln \lambda - N \ln \Gamma(\lambda^{-1}).$$

The estimation yields $\hat{\alpha} = 0.481$ (0.028) and $\hat{\lambda} = 0.664$ (0.045). The implied hazard function displays negative duration dependence, especially in the first 12 months of unemployment where the exit rate falls by more than half (top panel of Figure 3).

We divided the sample into subgroups based on gender and education, but we did not find any statistically significant evidence of heterogeneity across subgroups. This finding is not too surprising, given the relative homogeneity of our sample.

From durations to human capital: The estimated hazard is a function of unemployment duration, but in the model it is a function of human capital. In order to map duration into human capital all we need is the initial level of human capital (pre-displacement wage) and the monthly rate of human capital depreciation. Median monthly earnings of the workers in the CPS sample are \$1,540, thus we used an initial value of $h = 15$ and reconstructed human capital as a function of duration through the assumed depreciation rate. Figure 3 (lower panel) plots the hazard as a function of human capital.

9 Appendix C: The U.S. Welfare System

In what follows, we describe the pivotal ingredients of the U.S. welfare system.

Unemployment Insurance: The unemployment insurance replacement ratio in the U.S. varies across states. The state-determined weekly benefits generally replace between 50% and 70% of the individual's last weekly pre-tax earnings. The regular state programs usually provide benefits up to 26 weeks. The permanent Federal-State Extended Benefits program, present in every state, extends coverage up to 13 additional weeks, for a combined maximum of 39 weeks. Weekly benefits under the extended program are identical to those of the regular program.³³

TANF: The Temporary Assistance for Needy Families (TANF) program is the main cash assistance program for poor families with children under age 18 and at least one unemployed parent. It was implemented in 1996 as part of *The Personal Responsibility and Work Opportunity Reconciliation Act* (PRWORA) which, at the same time, eliminated all existing Federal assistance programs (the AFDC, in particular). The main innovations of the TANF program were three. First, the emphasis on encouraging self-sufficiency through work. TANF legislation specifies that, with few exceptions, recipients must participate to "work activities", such as un-subsidized or subsidized employment, on-the-job training, community service, job search, vocational training, or education directly related to work. The focus of our analysis is on job search programs. Second, the time limit to benefits: families with an adult who has received TANF assistance for a total of five years are not eligible for further cash aid over their lifetime. A number of states, however, have also imposed a shorter limit over fixed calendar intervals (e.g., 24 months over any given 5-year period). Third, financial incentives were created for states to run mandatory active labor market programs for workers on the TANF rolls. See Moffitt (2003) for a detailed description of the TANF program and Moffitt (2002) for a data set on average TANF payments by state.

Food Stamps: The Food Stamp program provides monthly coupons to eligible low-income families, which can be used to purchase food. Over 80% of TANF recipients also receive Food Stamps (DHHS, 2004). Once TANF benefits expire, households remain virtually without any other form of benefits and have the right to the maximum allotment of food stamps.

Unemployment Tax and Medicare Tax: The Federal Government imposes a net payroll tax on employers (FUTA) of 0.8% on the first \$7,000 of earnings paid annually to each employee. States finance their welfare programs with an additional State Unemployment Tax of 2.2%. The tax base varies by state: we use the median tax base in 1996 which was \$9,000. (House Ways and Means Committee, 1996, Table 5-10). In 1996, the total Medicare Tax rate was 2.9%.

³³Extended programs can be activated when unemployment is "relatively high".

EITC: The Federal Earned Income Tax Credit (EITC) is the major wage subsidy program in the United States. It is a refundable tax credit that supplements the earnings of low-income workers. In 1996, for a household with two dependent children, the subsidy rate was 40% up to \$741 per month. In the range \$741 – \$967, the subsidy is fixed at \$296. For monthly earnings over \$967, workers start paying a tax rate of 21% over and above the \$296 subsidy, until the break-even income such that the net subsidy is exactly zero, i.e., \$2,377. See Hotz and Scholtz (2001, Table 1) for details.

Figure 1

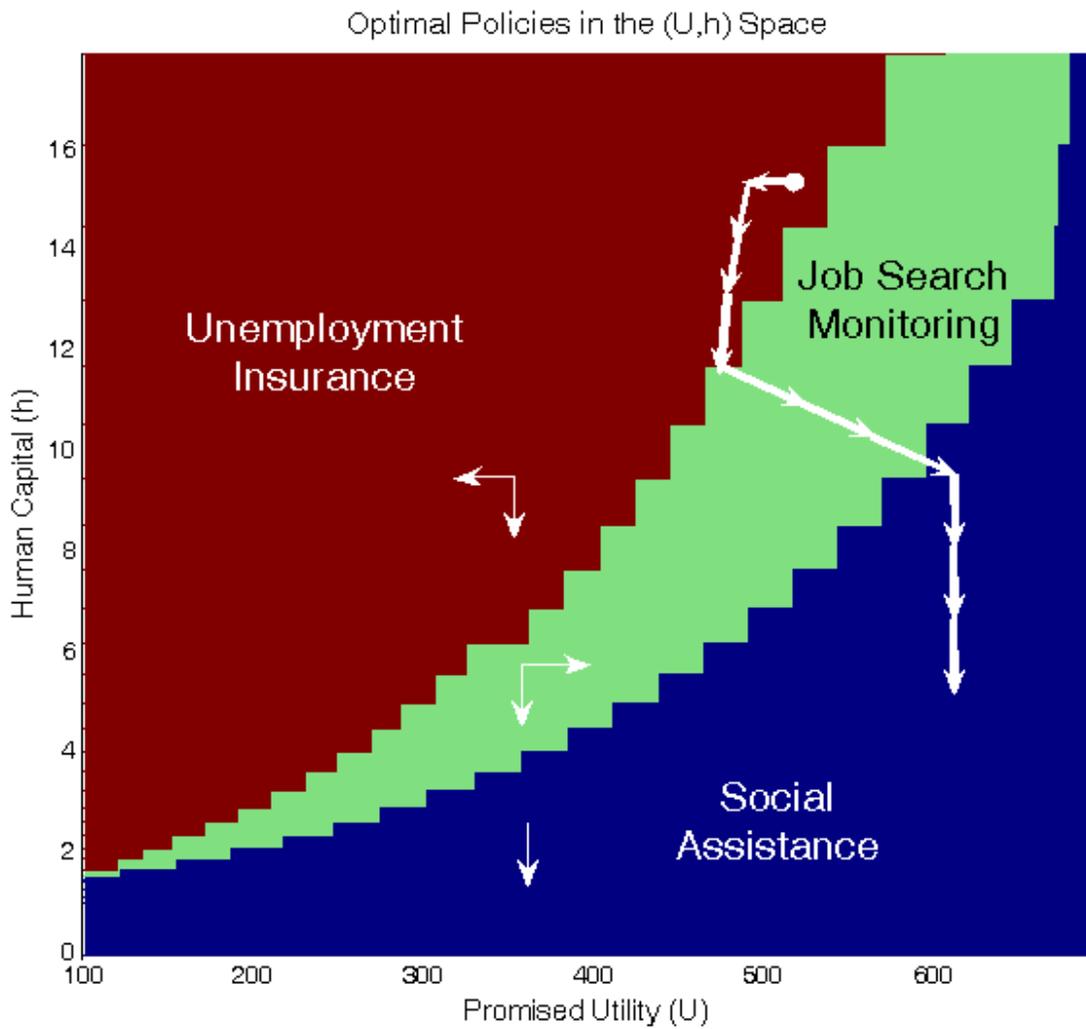


Figure 1: The policies of the optimal WTW program in the state space of human capital h and promised utility U .

Figure 2

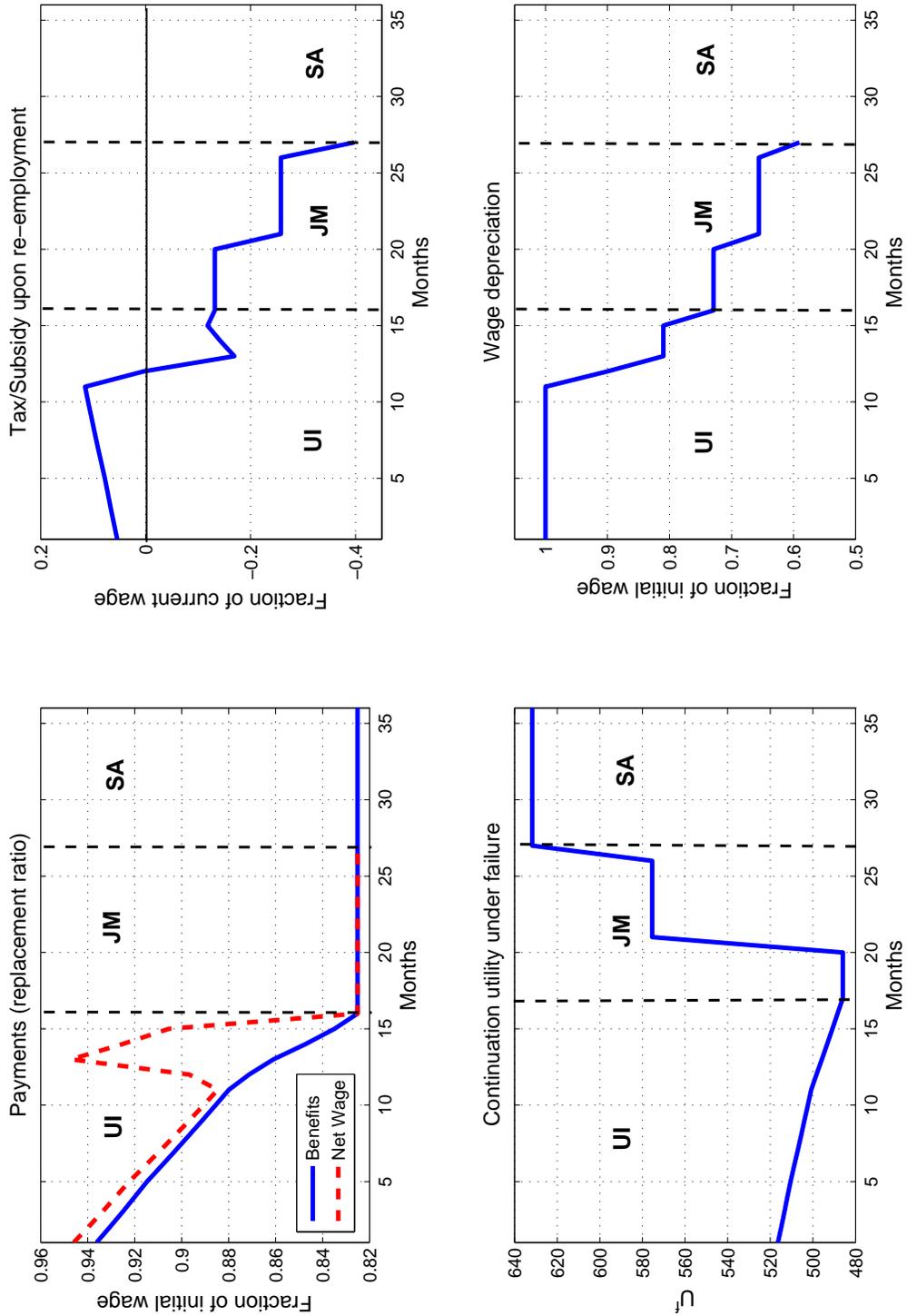


Figure 2: A representative history of the optimal WTW program.

Figure 3

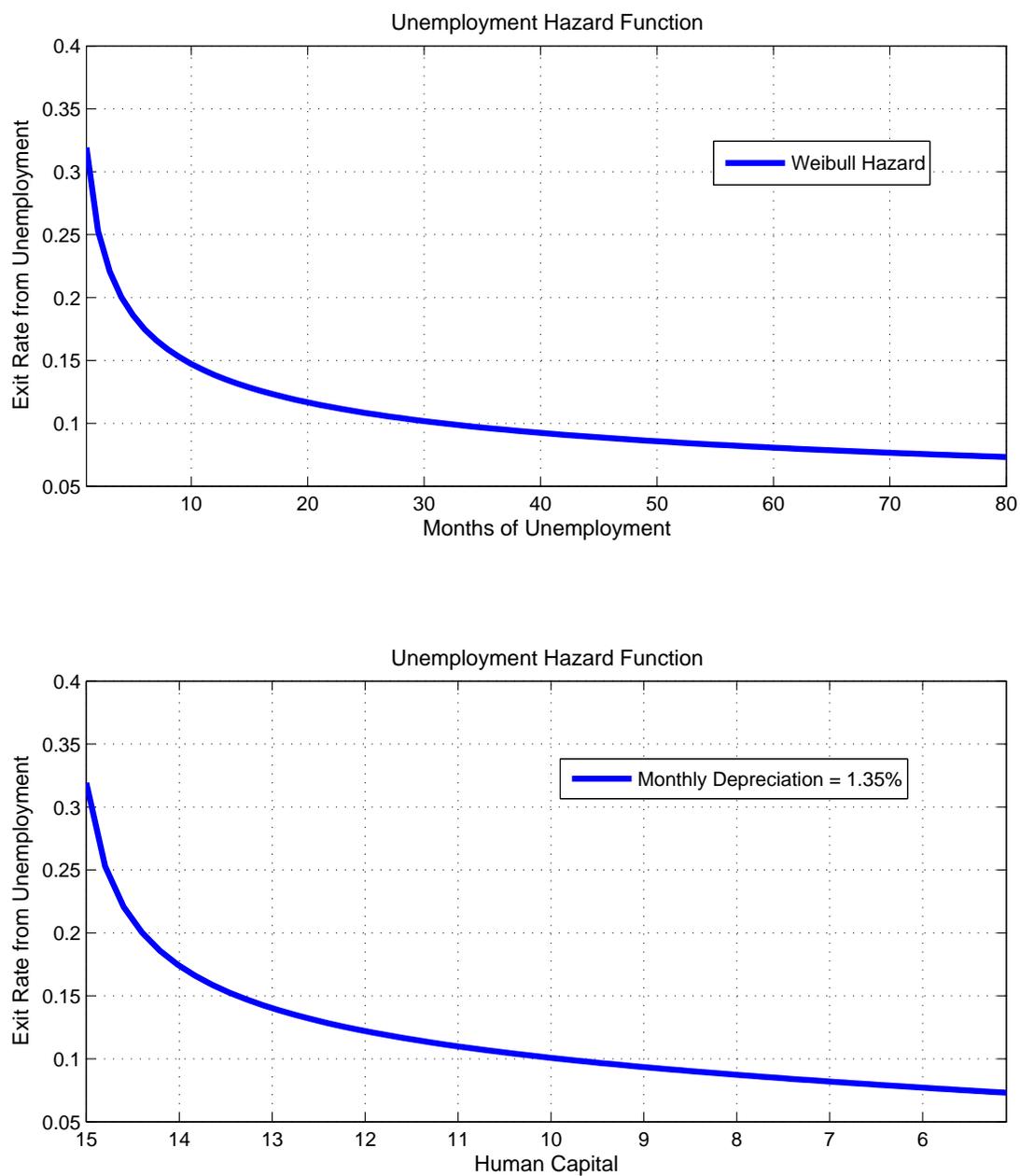


Figure 3: Exit rate from unemployment (Weibull hazard estimated on monthly CPS data May 1995-April 1996).

Figure 4

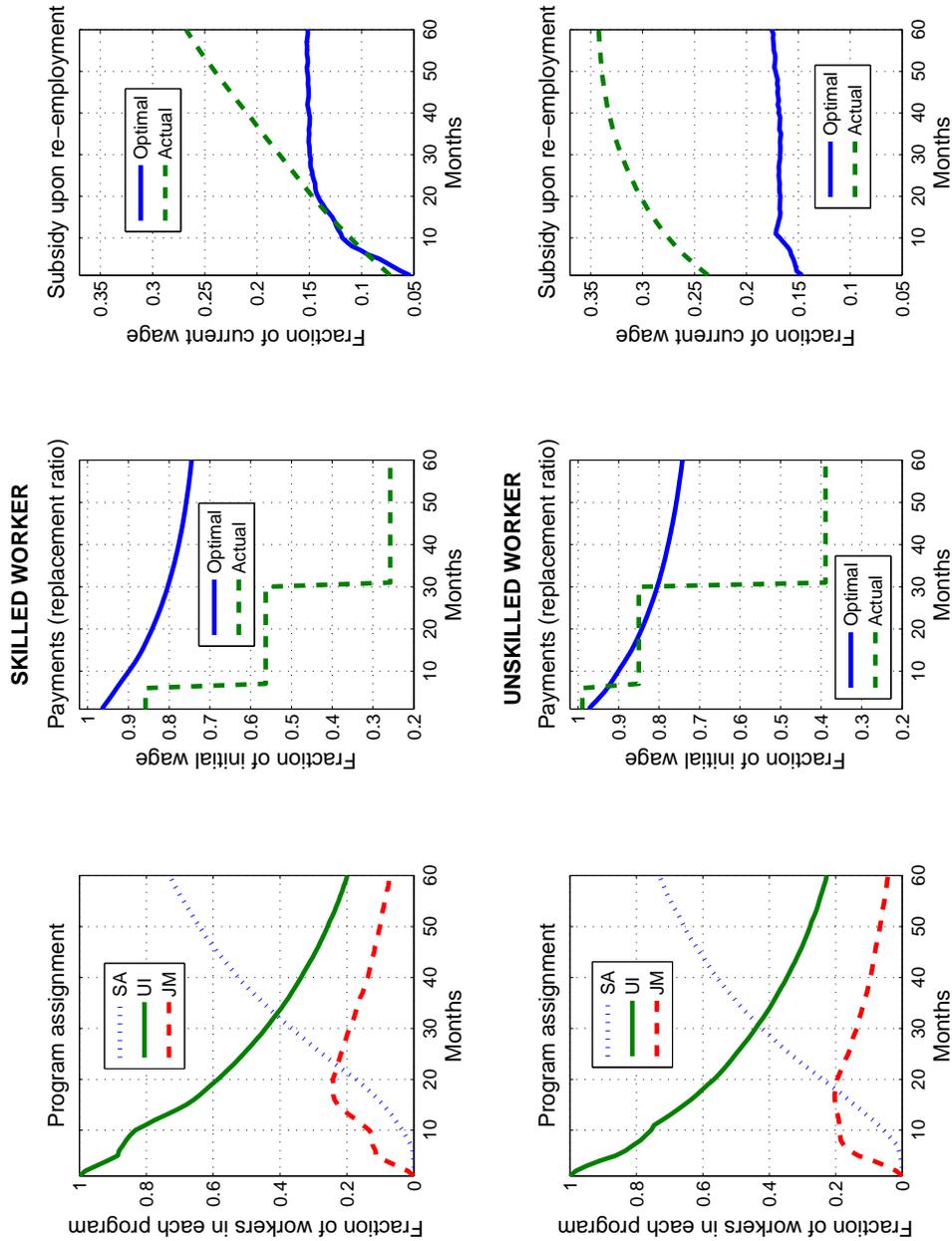


Figure 4: Features of the optimal WTW program compared to the actual U.S. welfare system for a high-skilled worker (monthly earnings of \$1,500) and an low-skilled worker (monthly earnings of \$1,000).