INTERNATIONAL ECONOMIC REVIEW
Vol. 50, No. 4, November 2009

TIME-INCONSISTENCY AND WELFARE PROGRAM PARTICIPATION: EVIDENCE FROM THE NLSY*

BY HANMING FANG[1] AND DAN SILVERMAN

University of Pennsylvania, Duke University, U.S.A.; University of Michigan, U.S.A.

We empirically implement a dynamic structural model of labor supply and welfare program participation for agents with potentially time-inconsistent preferences. Using panel data on the choices of single women with children from the National Longitudinal Surveys (NLSY) 1979, we provide estimates of the degree of time-inconsistency, and of its influence on the welfare take-up decision. With these estimates, we conduct counterfactual experiments to quantify a measure of the utility loss stemming from the inability to commit to future decisions, and the potential gains from commitment mechanisms such as welfare time limits and work requirements.

1. INTRODUCTION

Economists studying choice over time typically assume that decision makers are impatient and, traditionally, this impatience is modeled in a very particular way: Agents discount future streams of utility or profits *exponentially* over time. Strotz (1956) showed that exponential discounting is not just an analytically convenient assumption; without it, intertemporal marginal rates of substitution will change as time passes, and preferences will be time-inconsistent.

A literature has built on the work of Strotz and others to explore the consequences of relaxing the standard assumption of time-consistent discounting. Drawing both on experimental research and on common intuition, economists have built models of quasi-hyperbolic discounting to capture the tendency of decision makers to seize short-term rewards at the expense of long-term preferences.[2] This literature studies the implications of time-inconsistent preferences, and their associated problems of self-control, for a variety of economic choices and environments.[3]

This article is an empirical investigation of the relationship between time discounting and work and welfare program participation decisions. Using panel data on the choices of never-married women with dependent children, we estimate a dynamic structural model of labor supply

* Manuscript received March 2006; revised March and August 2007.

[1] We are deeply indebted to Ken Wolpin for his advice and encouragement on this project. We also thank Steve Berry, John Bound, Raj Chetty, Stefano Della Vigna, Zvi Eckstein, Michael Keane, David Laibson, Donghoon Lee, Ulrike Malmandier, Robert Miller, Daniele Paserman, Andrew Postlewaite, Matthew Rabin, Mark Rosenzweig, John Rust, Kent Smetters, and participants at many seminars and conferences for helpful suggestions and discussions. Finally, we thank three anonymous referees for careful comments that much improved the article. We are responsible for the remaining errors and shortcomings. Please address correspondence to: Hamming Fang, Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297. E-mail: *hanming.fang@econ.upenn.edu*.

[2] A body of experiments, reviewed in Ainslie (1992) and in several papers in Loewenstein and Elster (1992), indicate that hyperbolic time discounting may parsimoniously explain some basic features of the intertemporal decision making that are not consistent with simple models with exponential discounting. Specifically, standard decision models with exponential discounting are not easily reconciled with commonly observed preference reversals: Subjects choose the larger and later of two prizes when both are distant in time, but prefer the smaller but earlier one as both prizes draw nearer to the present (see Rubinstein, 2003, and Halevy, 2008 for alternative explanations of preference reversals).

[3] For examples, models of time-inconsistent preferences have been applied by Laibson (1997) and O'Donoghue and Rabin (1999a,b) to consumption and savings; by Barro (1999) to growth; by Gruber and Koszegi (2001) to smoking decisions; by Krusell, Kuruşçu, and Smith (2002) to optimal tax policy; by Carrillo and Mariotti (2000) to belief formation; and by Della Vigna and Paserman (2005) to job search.

and welfare program participation that allows present-biased time preferences. Our estimates, which also allow for unobserved heterogeneity in skills and tastes, indicate a time-inconsistent discount function. Implementing the quasi-hyperbolic form, we estimate a present-bias factor considerably less than one, and reject a standard exponential discounting model.

The article makes two contributions to the literature on present-biased preferences. First, by applying a model that allows quasi-hyperbolic preferences to the problem of labor supply and welfare program participation, we provide an economically significant setting for an evaluation of the importance of time-inconsistency. As a source of information about time-preferences, this context has the advantage that labor supply decisions are among the most consequential economic choices that individuals make: They drive time use for working-age adults. A disadvantage of focusing on welfare decisions is that it leads us to examine a special segment of the population (never-married women with children): Our results may not extend beyond that group.[4] It may be, however, that time-inconsistency is particularly consequential for this population. Recent welfare reforms and anecdotal evidence indicate a commonly held view: The trade-off between the short-term costs of entering the labor force at a low wage relative to the welfare benefit, and the long-term reward of higher wages from the accumulation of work experience, may generate problems of self-control. These self-control problems may, in turn, provide a rationale for the common belief that the decision to rely on welfare for many years is, somehow, suboptimal. Our previous research makes clear how and when this common belief might be justified if preferences are time-inconsistent, and shows how self-control problems may produce important observable differences in the behavior of time-consistent and time-inconsistent agents (Fang and Silverman, 2004).[5]

The article's second contribution is methodological. Economists have so far largely calibrated models of time-inconsistent preferences to match important moments of aggregate data sets (for example, Laibson et al., 1998). In this article, we estimate the structural parameters of the model, including the present-bias parameter, from a single panel data set.[6] Two recent papers also use field data to structurally estimate discount factors. Paserman (2008) estimates a structural job search model using data on unemployment spells and accepted wages from the National Longitudinal Surveys (NLSY). Laibson et al. (2007) estimate a structural model of consumption and saving. They calibrate some of their model parameters and estimate the time discount factors using the method of simulated moments.

More generally, our attempt to quantify consequences of time-inconsistency distinguishes this article from many in the literature on quasi-hyperbolic discounting. That literature often uses stylized models to demonstrate the potentially large behavioral effects of time-inconsistency in preferences. Quantitative assessments are relatively few.[7] Simulating our estimated model allows us to quantify the effects, in terms of behavior and utility, of obtaining perfect commitment ability or imperfect commitment via a welfare reform. This analysis is important because whether even profound present-bias in preferences implies economically substantial behavioral consequences is an empirical question. In order to illustrate this point, consider the extreme case in which agents' choices among discrete options are made very far from the margin;[8] then, even if they were highly present-biased, their choices would be unlikely to change even when they are able

[4] There is reason to think that lower income groups will reveal higher rates of time discount. See, e.g., Hausman (1979), Lawrance (1991), and Paserman (2008).

[5] We investigate just one mechanism (time-inconsistent preferences) that may lead to suboptimal welfare dependence. Externalities from welfare receipt could, for example, also make long-term dependence socially, if not individually, suboptimal. Moreover, there are other cognitive biases, such as optimism or misprediction of future preferences or returns from work, that could also lead recipients to receive welfare "too long." See Fang and Silverman (2006) for a discussion.

[6] Prior research has tested the reduced-form implications of hyperbolic discounting. For example, Della Vigna and Paserman (2005) consider the influence of self-control problems on job search; and Della Vigna and Malmandier (2006) find evidence of time-inconsistent preferences in data on health-club contracts and usage.

[7] Laibson et al. (1998), Angeletos et al. (2001), Gruber and Koszegi (2001), Della Vigna and Paserman (2005), and Paserman (2008) are notable exceptions.

[8] That is, if an agent chooses alternative A over B, her utility from A is much larger than that from B; and vice versa.

to commit to their future choices. Similarly, even when the ability to commit would dramatically alter agents' *behavior,* it need not imply large *utility gains*. In order to illustrate this point, consider another extreme case in which agents' choices are made very close to the margin; then their choices are likely to change when they can commit their future selves' behavior. But such changes in choices will have little utility consequence because these consumers were initially close to the margin. Our simulations of perfect commitment ability suggest that, for many women in our data, this latter case applies. For many, the ability to commit leads to substantial changes in behavior but relatively small changes in discounted lifetime utility.

Finally, this article is related to a literature on labor supply and welfare participation that structurally implements models of dynamic decision making. Miller and Sanders (1997) estimates a dynamic discrete choice model in which women decide monthly whether to work or receive welfare. In an effort to explain both the low welfare take-up rate among eligible families and the persistence of welfare choices among the families who do enroll, Miller and Sanders incorporate wage growth through work experience and preferences that adapt both to labor supply and to welfare experience. As in our article, fertility and marriage are exogenous. Swann (2005) adds marriage to the choice set, and looks at women's decisions annually. Keane and Wolpin (2005) endogenize education, employment, fertility, and marriage decisions. These prior papers all assume exponential time discounting. Our article contributes to this literature and to the welfare reform debate with, to our knowledge, the first empirical examination of the relationship between time-inconsistency and the welfare take-up decision.

The remainder of the article is structured as follows. Section 2 presents our model and describes both the intrapersonal game played by the decision maker and the numerical method for obtaining the game's solution. Section 3 presents the estimation strategy and discusses identification. Section 4 describes the data and variable definitions. Section 5 presents the estimation results and associated simulations. Section 6 provides estimates of both the behavioral consequences and the utility effects of commitment and of various policy changes such as time limits and workfare. Section 7 offers conclusions.

2. THE MODEL

We consider a discrete time model of work-welfare decisions by a single parent (agent). Each agent has a finite decision horizon starting from her age at the birth of her first child, a_0, and ending at age A.[9] At each age $a \in \{a_0, \dots, A\}$, the agent must choose from a set D that includes three mutually exclusive and exhaustive alternatives: receive welfare, work in the labor market, or stay at home without work or welfare. The alternatives of welfare, work, and home are, respectively, referred to as choices 0, 1, and 2, thus $D = \{0, 1, 2\}$.[10] The agent's decision at age a is denoted by $d_a \in D$.

The return from choosing alternative d for an agent of age a represents all of the current-period benefits and costs associated with the choice, and it is denoted by $R_a(d; \mathbf{s}_a, \epsilon_{da})$, where s_a is a vector of state variables at age a, detailed below, and ϵ_{da} is a random shock to the value of alternative d at age a. We parameterize $R_a(d; \mathbf{s}_a, \epsilon_{da})$ as follows.

Welfare. At age a, an agent's payoff relevant state $\mathbf{s}_a$ includes, among other things, her state of residence j, the number of her children in period a, denoted by n_a, and her age-$(a-1)$ choice

[9] The agent will obtain a continuation value at age A as a function of her endogenous state variables. The empirical implementation sets $A = 34$, which guarantees that, up to age A each woman in the data sample continues to have children younger than 18 and thus meets the minimum requirement for receiving Aid to Families with Dependent Children (AFDC).

[10] In reality, an agent may choose more than one action in any period, and there are distinctions between part- and full-time work. For example, Edin and Lein (1997) report that, in their study of 379 low-income single mothers, many welfare recipients both work in the (unofficial) labor market and rely on family and neighborhood resources.

d_{a-1}. In the absence of a time limit, the age-a return to welfare, $R_a(0; \mathbf{s}_a, \epsilon_{0a})$, is given by[11]

$$R_a(0; \mathbf{s}_a, \epsilon_{0a}) = e(n_a) + G_j(n_a) - \phi(d_{a-1}) + \epsilon_{0a}, \tag{1}$$

where $e(n_a)$ is the monetary value of her home production skills or leisure as a function of the number of her children; $G_j(n_a)$ is the monetary value of the cash and food welfare benefits in state j as a function of the number of her children; $\phi(d_{a-1})$ is the net stigma associated with welfare participation denominated in dollars; and ϵ_{0a} is an idiosyncratic, choice-specific shock.

The value of home production skills (leisure) is allowed to depend on the number of children to capture the additional demands or rewards of having more children. We assume a quadratic function for $e(n_a)$:

$$e(n_a) = e_0 + e_1 n_a + e_2 n_a^2, \tag{2}$$

where e_0 may be heterogeneous in the population. The welfare benefits schedule $G_j(n_a)$ is assumed to be an affine function of the number of children[12]

$$G_j(n_a) = \theta_{j0} + \theta_{j1} n_a. \tag{3}$$

The welfare benefit schedule $G_j(n_a)$ is estimated separately for each state of residence j. Finally, the net welfare stigma $\phi(d_{a-1})$ is specified as

$$\phi(d_{a-1}) = \begin{cases} 0, & \text{if } d_{a-1} = 0 \\ \phi, & \text{otherwise.} \end{cases} \tag{4}$$

Thus, we assume stigma lasts for just one period after switching into welfare from some other choice.[13] In our empirical estimation, ϕ is also an element of the unobserved heterogeneity we allow. The specification of welfare stigma in (4) is natural if we interpret the stigma as the psychic and administrative costs associated with welfare take-up. If we take a more general interpretation of stigma, then (4) imposes a particular form of stigma decay with continued participation.

If there are welfare time limits, the cumulative number of periods an agent has received welfare prior to age a is also payoff relevant. This variable is denoted by κ_a. Given a lifetime limit of L periods, the return to welfare is then[14]

$$R_a(0; \mathbf{s}_a, \epsilon_{0a}) = \begin{cases} e(n_a) + G_j(n_a) - \phi(d_{a-1}) + \epsilon_{0a}, & \text{if } \kappa_a < L \\ -\infty, & \text{otherwise.} \end{cases}$$

Work. An agent's age-a return from work $R_a(1; \mathbf{s}_a, \epsilon_{1a})$ is her wage. Following a standard theory of human capital, we model this wage as the product of a (constant) rental price of human capital,

[11] In order to decrease the dimension of the state in our empirical estimation, we select a sample of women who have children younger than age 18 throughout the period they are observed and are therefore eligible for welfare, during all periods that we analyze.

[12] The actual welfare benefits schedule deviates from a linear function approximation by a few dollars at most. We abstract from asset and income restrictions on welfare eligibility.

[13] The net stigma parameter $\phi(d_{a-1})$ has, since Moffitt (1983), become standard in empirical studies of welfare participation. Its primary function is to help explain the fraction of welfare-eligible adults who remain at home without work or welfare.

[14] We assume that time limits are perfectly and uniformly enforced. In reality, the implementation of time limits has been complex, with many states providing exemptions to large fractions of recipients who reach the limits (Bloom et al., 2002). Moreover, there is no national database for preventing recipients who migrate across states from receiving more than 5 years of benefits. Such interstate migrants will, however, often face other restrictions on their eligibility.

r, and the quantity of skill units held by the individual $h_a(\mathbf{s}_a, \epsilon_{1a})$

$$R_a(1; \mathbf{s}_a, \epsilon_{1a}) = r h_a(\mathbf{s}_a, \epsilon_{1a}).$$

When the state is $\mathbf{s}_a$, an agent's age-a skills are given by

$$h_a(\mathbf{s}_a, \epsilon_{1a}) = \exp\left[h_0 + \alpha_1 g_0 + \alpha_2 x_a + \alpha_3 x_a^2 + \alpha_4 \mathtt{I}(x_a > 0) + \alpha_5 \mathtt{I}(d_{a-1} \neq 1) + \epsilon_{1a}\right], \tag{5}$$

where h_0 is the agent's (unobserved) skill endowment at the birth of her first child; g_0 is her completed years of schooling at the birth of her first child; x_a is her total work experience prior to age-a; d_{a-1} is her choice in the previous period; and ϵ_{1a} is the age-a skill shock. In specification (5), $\mathtt{I}(\cdot)$ is an indicator function equal to one if the expression in parentheses is true. Thus $\alpha_4 \mathtt{I}(x_a > 0)$ takes value α_4 if the agent acquired any work experience before age a and captures a persistent first-year experience effect. The term $\alpha_5 \mathtt{I}(d_{a-1} \neq 1)$ takes value α_5 (which is presumably negative) whenever the agent did not work in the previous period. Thus the parameter α_5 represents the one-time depreciation of human capital that occurs whenever the agent leaves work to choose welfare or home. Note that the functional form (5) implies that the sum $(\ln r + h_0)$, but not $\ln r$ and h_0 separately, can be identified.

Home. An agent's current-period return from staying home without work or welfare $R_a(2; \mathbf{s}_a, \epsilon_{2a})$ is specified as follows:

$$R_a(2; \mathbf{s}_a, \epsilon_{2a}) = e(n_a) + \eta \mathtt{I}(d_{a-1} = 2) + \epsilon_{2a},$$

where $e(n_a)$ is the same monetary value of home production as in (2); η captures the possible decay or appreciation of the value of home production when a woman stays home without welfare (it, too, will be an element of the unobserved heterogeneity we allow); and ϵ_{2a} is a choice-specific shock.

We assume that the choice-specific shocks $\boldsymbol{\epsilon}_a = (\epsilon_{0a}, \epsilon_{1a}, \epsilon_{2a})$ are distributed according to a joint normal distribution $\mathrm{N}(0, \boldsymbol{\Omega})$, and they are serially uncorrelated.

Observed and Unobserved Heterogeneity and the State. So far we have described the choices for a typical never-married mother. Now, we describe the state variable $\mathbf{s}_a$ and its transition in greater detail, as well as the observed and unobserved heterogeneity among single mothers that we allow in the estimation.

We analyze a never-married woman's work/welfare/home decisions from the age when she was first surveyed in the NLSY or when she gave birth to her first child, whichever occurred later.[15] When an agent first enters our analysis at age a_0, we observe a set of initial conditions, including her state of residence j, her years of completed schooling g_0, her prior work experience x_0, and her decision in the period prior to the birth of her first child d_{a_0-1}.[16] We assume that an agent's state of residence remains unchanged during the course of the data, and she does not complete further schooling; thus (j, g_0) are constant over time.[17] We do not model the process that generated the differences in initial conditions among agents; instead we assume that the differences are captured by persistent, unobserved heterogeneity. Specifically, and as noted

[15] Among all women in the NLSY, 21% had already given birth to a child before 1979, the first year of the survey. In order to insure that these women maintain minimum eligibility for AFDC until 1991, we exclude all who had children older than age 3 when first interviewed in 1979. As a result, among those women with children when first interviewed, the oldest who remained in our sample was 20 years old in 1979.

[16] We rely on the NLSY's retrospective questions (going back as far as 1975) regarding work/welfare experience to provide information about the initial conditions of those women in the sample who already had children when first interviewed in 1979.

[17] In fact, 85% of the sample described below continued, throughout the period observed, to reside in their state of residence at age a_0. In the same sample 34% went on to acquire additional schooling after the birth of their first child. Of this fraction, approximately half acquired less than one additional year of schooling.

above, we allow for heterogeneity in the labor market and home skill endowments (h_0, e_0), the nonwelfare home production decay η, and welfare stigma ϕ. Section 3 describes how our method allows for a correlation between unobserved heterogeneity and observable initial conditions. Because we treat state of residence, schooling, and work/welfare decisions prior to age a_0 as predetermined, the intertemporal trade-offs that dictated these decisions do not inform our estimates of the (homogeneous) discount function.

An agent's period-a state variables include her prior work experience x_a, the number of her children n_a, the number of prior periods she had participated in welfare κ_a, and her last period decision d_{a-1}. Thus, $(x_a, n_a, \kappa_a, d_{a-1})$ represents the potentially time-varying elements of the period-a state. In order to summarize, an agent's period-a state variable is denoted as $\mathbf{s}_a = (j, g_0, x_a, n_a, \kappa_a, d_{a-1})$, and we write the space for the state at age a as $\mathcal{S}_a$. The evolution of the elements of the state is straightforward except for n_a, the number of children. We treat the arrival of additional children as exogenously determined and model births as a process that satisfies

$$n_{a+1} = \begin{cases} n_a + 1, & \text{with probability } \rho(a, n_a, d_a) \\ n_a, & \text{with probability } 1 - \rho(a, n_a, d_a), \end{cases}$$

where $\rho(a, n_a, d_a)$ is a logistic function

$$\rho(a, n_a, d_a) = \frac{\exp\left[\gamma_0 + \gamma_1 a + \gamma_2 n_a + \gamma_3 \mathrm{I}(d_a = 0) + \gamma_4 \mathrm{I}(d_a = 1)\right]}{1 + \exp\left[\gamma_0 + \gamma_1 a + \gamma_2 n_a + \gamma_3 \mathrm{I}(d_a = 0) + \gamma_4 \mathrm{I}(d_a = 1)\right]}. \tag{6}$$

Preferences. We now move on to describe an agent's intertemporal preferences. We assume that an agent consumes all of the returns from her choice d in each period and obtains an instantaneous utility $u_a = R_a(d; \mathbf{s}_a, \epsilon_{da})$. An agent in period a is concerned about both her present and future instantaneous utilities. Let $U^a(u_a, u_{a+1}, \dots, u_A)$ represent an agent's intertemporal preferences from the perspective of period a. We adopt a simple and now commonly used formulation of agents' potentially time-inconsistent preferences: (β, δ)-preferences (Phelps and Pollak, 1968; Laibson, 1997, and O'Donoghue and Rabin, 1999a):

DEFINITION 1. (β, δ)-preferences are intertemporal preferences represented by

$$U^a(u_a, \dots, u_A) \equiv \delta^a u_a + \beta \sum_{t=a+1}^{A} \delta^t u_t,$$

where $\beta \in (0, 1]$, $\delta \in (0, 1]$, and $a \in \{a_0, a_0 + 1, \dots, A\}$.

Following the terminology of O'Donoghue and Rabin (1999a), the parameter δ is called the *standard discount factor* and captures long-run, time-consistent discounting; the parameter β is called the *present-bias factor* and captures short-term impatience. The standard model is nested as a special case of (β, δ)-preferences when $\beta = 1$. When $\beta \in (0, 1)$, (β, δ)-preferences capture "quasi-hyperbolic" time discounting (Laibson, 1997). We say that an agent's preferences are time-consistent if $\beta = 1$ and are present-biased if $\beta \in (0, 1)$.

Following previous studies of time-inconsistent preferences, we will analyze the behavior of an agent by thinking of the single individual as consisting of many autonomous *selves,* one for each period. Each period-a self chooses her current behavior to maximize her current utility $U^a(u_a, \dots, u_A)$, whereas her future selves control her subsequent decisions. The literature on time-inconsistent preferences distinguishes between *naive* and *sophisticated* agents (Strotz, 1956; Pollak, 1968; O'Donoghue and Rabin, 1999a,b). An agent is *partially naive* if the self in every period a underestimates the present-bias of her future selves, believing that her future selves' present-bias is $\tilde{\beta} \in (\beta, 1)$; in the extreme, if the present self believes that her future selves are time-consistent, i.e., $\tilde{\beta} = 1$, she is said to be *completely naive.* On the other hand, an agent is *sophisticated* if the self in every period a correctly knows her future selves' present-bias β and anticipates their behavior when making her period-a decision.

2.1. *Strategies, Payoffs, and Equilibrium.* We restrict our attention to Markov strategies and define a *feasible strategy* for a period-a self as a mapping $\sigma_a : \mathcal{S}_a \times \mathbb{R}^3 \to D$, where $\sigma_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a) \in \{0, 1, 2\}$ is simply the choice of the agent's period-a self over welfare, work, or home when her state is $\mathbf{s}_a$ and the period-a shock vector is $\boldsymbol{\epsilon}_a = (\epsilon_{0a}, \epsilon_{1a}, \epsilon_{2a})$. With slight abuse of notation, we write $R_a(\sigma_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a); \mathbf{s}_a, \boldsymbol{\epsilon}_a)$ as the instantaneous period-a utility the agent obtains from strategy σ_a when the state is $\mathbf{s}_a$ and shocks are $\boldsymbol{\epsilon}_a$.

A *strategy profile* for all selves is $\boldsymbol{\sigma} \equiv \{\sigma_t\}_{t=a_0}^{A}$. It specifies for each self her action in all possible states and under all possible realizations of shock vectors. For any strategy profile $\boldsymbol{\sigma}$, write $\boldsymbol{\sigma}_a^+ \equiv \{\sigma_t\}_{t=a}^{A}$ as the *continuation strategy profile* from period a to A. In order to define and characterize the equilibrium of the intrapersonal game of an agent with potentially time-inconsistent preferences, we first introduce a useful concept. Write $V_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a; \boldsymbol{\sigma}_a^+)$ as the agent's period-a expected continuation utility when the state is $\mathbf{s}_a$ and the shock vector is $\boldsymbol{\epsilon}_a$ under her *long-run* time preference for a given a continuation strategy profile $\boldsymbol{\sigma}_a^+$. We can think of $V_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a; \boldsymbol{\sigma}_a^+)$ as representing her intertemporal preferences from some prior perspective when her own present-bias is irrelevant. Specifically, $V_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a; \boldsymbol{\sigma}_a^+)$ can be calculated recursively as follows. First, let

$$V_A(\mathbf{s}_A, \boldsymbol{\epsilon}_A; \boldsymbol{\sigma}_A^+) = R_A(\sigma_A(\mathbf{s}_A, \boldsymbol{\epsilon}_A); \mathbf{s}_A, \boldsymbol{\epsilon}_A) + \delta \mathrm{E}[W(\mathbf{s}_{A+1}) \mid \mathbf{s}_A, \sigma_A(\mathbf{s}_A, \boldsymbol{\epsilon}_A)], \tag{7}$$

where $W(\mathbf{s}_{A+1})$ is the continuation value at the terminal age A as a function of the period-$(A+1)$ state; and the expectation is taken over the fertility shock conditional on $\mathbf{s}_A$, as modeled by (6), and decision $\sigma_A(\mathbf{s}_A, \boldsymbol{\epsilon}_A)$.[18] Recursively, for $a = A-1, \ldots, a_0$,

$$V_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a; \boldsymbol{\sigma}_a^+) = R_a(\sigma_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a); \mathbf{s}_a, \boldsymbol{\epsilon}_a) + \delta \mathrm{E}\big[V_{a+1}(\mathbf{s}_{a+1}, \boldsymbol{\epsilon}_{a+1}; \boldsymbol{\sigma}_{a+1}^+) \,\big|\, \mathbf{s}_a, \sigma_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a)\big], \tag{8}$$

where the expectation is taken over both the conditional fertility shock and $\boldsymbol{\epsilon}_{a+1}$.

We will define the equilibrium for a partially naive agent whose period-a self believes that, beginning next period, her future selves will behave optimally with a present-bias factor of $\tilde{\beta} \in [\beta, 1]$.[19] Following O'Donoghue and Rabin (1999b, 2001), we first define the concept of an agent's *perceived continuation strategy profile* by her future selves.

DEFINITION 2. The *perceived continuation strategy profile* for a partially naive agent is a strategy profile $\tilde{\boldsymbol{\sigma}} \equiv \{\tilde{\sigma}_a\}_{a=a_0}^{A}$ such that for all $a \in \{a_0, \ldots, A\}$, all $\mathbf{s}_a \in \mathcal{S}_a$, and all $\boldsymbol{\epsilon}_a \in \mathbb{R}^3$,

$$\tilde{\sigma}_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a) = \arg\max_{d \in D} \big\{ R_a(d; \mathbf{s}_a, \epsilon_{da}) + \tilde{\beta}\delta \mathrm{E}\big[V_{a+1}(\mathbf{s}_{a+1}, \boldsymbol{\epsilon}_{a+1}; \tilde{\boldsymbol{\sigma}}_{a+1}^+) \,\big|\, s_a, d\big]\big\}.$$

[18] In the empirical implementation, we approximate the continuation value by the following function of state variables:

$$W(s_{A+1}) = \omega_1 n_{A+1} + \omega_2 n_{A+1}^2 + \omega_3 x_{A+1} + \omega_4 x_{A+1}^2 + \omega_5 \mathrm{I}(d_A = 1) + \omega_6 \mathrm{I}(d_A = 2).$$

Our approach follows, for example, Keane and Wolpin (2001) by approximating terminal valuations with a parsimonious polynomial. The Monte Carlo evidence in Keane and Wolpin (1994) indicates that such polynomials approximate the value function quite well.

[19] Note, we define equilibrium for partially naive agents to ease exposition; it has the virtue of incorporating the naive and sophisticated agents as special cases. We are *not,* however, estimating the naivety parameter $\tilde{\beta}$ in our empirical analysis.

That is, if an agent is partially naive with perceived present-bias by future selves of $\tilde{\beta}$, then her period-a self will anticipate that her future selves will follow strategies $\tilde{\boldsymbol{\sigma}}_{a+1}^{+} \equiv \{\tilde{\sigma}_t\}_{t=a+1}^{A}$. Given this perception, the period-a self's best response is called *perception-perfect strategy profile*.

DEFINITION 3. A *perception-perfect strategy profile* for a partially naive agent is a strategy profile $\boldsymbol{\sigma}^* \equiv \{\sigma_a^*\}_{a=a_0}^{A}$ such that, for all $a \in \{a_0, \ldots, A\}$, all $\mathbf{s}_a \in \mathcal{S}_a$, and all $\boldsymbol{\epsilon}_a \in \mathbb{R}^3$,

$$\sigma_a^*(\mathbf{s}_a, \boldsymbol{\epsilon}_a) = \arg\max_{d \in D} \left\{ R_a(d; \mathbf{s}_a, \epsilon_{da}) + \beta\delta \mathrm{E}\left[V_{a+1}\left(\mathbf{s}_{a+1}, \boldsymbol{\epsilon}_{a+1}; \tilde{\boldsymbol{\sigma}}_{a+1}^{+}\right) \,\middle|\, \mathbf{s}_a, d \right] \right\}.$$

When the agent is sophisticated, the perceived continuation strategy profile is correct. That is, for a sophisticated agent

$$\tilde{\sigma}_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a) = \arg\max_{d \in D} \left\{ R_a(d; \mathbf{s}_a, \epsilon_{da}) + \beta * \delta \mathrm{E}\left[V_{a+1}\left(\mathbf{s}_{a+1}, \boldsymbol{\epsilon}_{a+1}; \tilde{\boldsymbol{\sigma}}_{a+1}^{+}\right) \,\middle|\, \mathbf{s}_a, d \right] \right\},$$

and thus $\tilde{\boldsymbol{\sigma}} = \sigma^*$. For sophisticates, then, the perception-perfect strategy profile is the familiar subgame perfect equilibrium of the intrapersonal conflict game. In our empirical implementation, we will report results for both completely naive ($\tilde{\beta} = 1$) and sophisticated agents ($\tilde{\beta} = \beta$).

2.2. *Numerical Solution of* $\boldsymbol{\sigma}^*$. In our empirical implementation, the terminal age A is finite. This allows us to solve numerically the perception-perfect strategy profile $\boldsymbol{\sigma}^*$ recursively. The solutions for sophisticated and completely naive agents are merely special cases of the partially naive solution, so we describe how $\boldsymbol{\sigma}^*$ can be numerically solved for a partially naive agent.

First, consider the terminal period A. For any $\mathbf{s}_A \in \mathcal{S}_A$ and $\boldsymbol{\epsilon}_A \in \mathbb{R}^3$, the period-$A$ self's optimal strategy is simple:

$$\sigma_A^*(\mathbf{s}_A, \boldsymbol{\epsilon}_a) = \arg\max_{d \in D} \{ R_A(d; \mathbf{s}_A, \epsilon_{da}) + \beta\delta \mathrm{E}[W(\mathbf{s}_{A+1}) \mid \mathbf{s}_A, d] \}.$$

A partially naive agent at period-$(A-1)$, however, would perceive that her period-A self would follow

$$\tilde{\sigma}_A(\mathbf{s}_A, \boldsymbol{\epsilon}_A) = \arg\max_{d \in D} \{ R_A(d; \mathbf{s}_A, \epsilon_{da}) + \tilde{\beta}\delta \mathrm{E}[W(\mathbf{s}_{A+1}) \mid \mathbf{s}_A, d] \}.$$

Now for every $a = A-1, \ldots, a_0$, every $\mathbf{s}_a \in \mathcal{S}_a$, and every $\boldsymbol{\epsilon}_a \in \mathbb{R}^3$, we will have, recursively,

$$\tilde{\sigma}_a(\mathbf{s}_a, \boldsymbol{\epsilon}_a) = \arg\max_{d \in D} \left\{ R_a(d; \mathbf{s}_a, \epsilon_{da}) + \tilde{\beta}\delta \mathrm{E}\left[V_{a+1}\left(\mathbf{s}_{a+1}, \boldsymbol{\epsilon}_{a+1}; \tilde{\boldsymbol{\sigma}}_{a+1}^{+}\right) \,\middle|\, \mathbf{s}_a, d \right] \right\}$$
$$\sigma_a^*(\mathbf{s}_a, \boldsymbol{\epsilon}_a) = \arg\max_{d \in D} \left\{ R_a(d; \mathbf{s}_a, \epsilon_{da}) + \beta\delta \mathrm{E}\left[V_{a+1}\left(\mathbf{s}_{a+1}, \boldsymbol{\epsilon}_{a+1}; \tilde{\boldsymbol{\sigma}}_{a+1}^{+}\right) \,\middle|\, \mathbf{s}_a, d \right] \right\},$$

where $V_{a+1}(\cdot, \cdot; \cdot)$ is recursively defined by (7) and (8). This completes the recursion.

Informally, in equilibrium the agent's decision making proceeds as follows. Beginning at age a_0, the period-a_0 self observes her state $\mathbf{s}_{a_0}$ and then draws three choice-specific shocks $\boldsymbol{\epsilon}_{a_0} \in \mathrm{N}(0, \Omega)$. Given the anticipated behavior of her future selves, represented by $\tilde{\boldsymbol{\sigma}}_{a_{0+1}}^{+}$, she calculates the realized current rewards and the expected future rewards from each of her three alternatives, using her own discount factors (β, δ). This calculation yields $\sigma_{a_0}^*(\mathbf{s}_{a_0}, \boldsymbol{\epsilon}_{a_0})$, representing the alternative that offers the highest discounted present value. Then, the state variable is updated for period-(a_0+1) according to the alternative chosen and the process is repeated. The perception-perfect strategy at each age a, for each $\mathbf{s}_a \in \mathcal{S}_a$, is identified by the region in the three-dimensional space of $\boldsymbol{\epsilon}_a$ over which each of the alternatives is optimal, for the given state $\mathbf{s}_a$. Because there is no closed-form representation of this solution, we will, in the estimation and simulations below, solve the game numerically by backward recursion using crude Monte Carlo integration to approximate the expected continuation values $\mathrm{E}[V_{a+1}(\mathbf{s}_{a+1}, \epsilon_{a+1}; \tilde{\boldsymbol{\sigma}}^{+}_{a+1}) \mid \mathbf{s}_a, d]$.[20]

3. ESTIMATION STRATEGY

The solution to the intrapersonal game described above provides the inputs for estimating the parameters of the model by the following method. We first describe the structure of our data (see also Section 4). We have data on choices, state variables, and related outcomes (such as welfare benefit levels and accepted wages) from a sample of agents, each of whom solves the intrapersonal conflict game. In what follows, we use superscript $i \in \{1, \ldots, N\}$ to index the agents. Our data set consists of three sets of information: (1) agent i's sequence of states represented by $\mathbf{s}^i \equiv \{\mathbf{s}^i_a\}^{A^i}_{a=a^i_0}$, where a^i_0 denotes the time individual i becomes part of our analysis, which is the latter of the age at which she gave birth to her first child and the date of the first interview; and A^i is the age at which we last observe the agent;[21] (2) agent i's sequences of choices $\mathbf{d}^i \equiv \{d^i_a\}^{A^i}_{a=a^i_0}$; (3) if agent i chooses to work, we observe her accepted wages, which we write as $\mathbf{w}^i \equiv \{w^i_a\}^{A^i}_{a=a^i_0}$ with the understanding that $w^i_a = \emptyset$ if $d^i_a \neq 1$. We also have a separate data set that provides the welfare benefit levels for families of different sizes for all the states of residence, denoted by G_j where j indexes the state of residence. We denote our data set by $\mathcal{D}$.

The decision at any age a is deterministic for the agent for a given vector $(\mathbf{s}_a, \epsilon_a) \in \mathcal{S}_a \times \mathbb{R}^3$, but it is probabilistic from our perspective because we do not observe the shock vector $\boldsymbol{\epsilon}_a$. As we described in the last paragraph of Subsection 2.2, for given parameters of the model $\boldsymbol{\Theta}$, we can numerically solve for the perception-perfect strategy profile as the solution to the game, and it then provides the probability of choosing alterative d^i_a at state $\mathbf{s}^i_a$ and, if $d^i_a = 1$, receiving wage w^i_a, denoted by

$$\Pr\left[d^i_a, w^i_a \,\middle|\, \mathbf{s}^i_a; \boldsymbol{\Theta}\right].$$

We can therefore consistently estimate $\boldsymbol{\Theta}$ by maximizing with respect to $\boldsymbol{\Theta}$ the sample likelihood

$$\prod_{i=1}^{N} \prod_{a=a^i_0}^{A^i} \Pr\left[d^i_a, w^i_a \,\middle|\, \mathbf{s}^i_a; \boldsymbol{\Theta}\right].$$

We ease the implementation by estimating two parts of the model outside of our basic choice framework. First, the parameters $\boldsymbol{\theta}_j \equiv (\theta_{0j}, \theta_{1j})$ in the welfare benefits function $G_j(\cdot)$ (see Equation (3)) are taken as the mean of estimated real benefit function in the agent's state of residence j over the period observed.[22] Table A2 of the Appendix presents these estimated parameters and summary statistics for the 20 U.S. states represented in the sample. Second, we estimate the parameters $\gamma \equiv (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$ of the fertility function $\rho(a, n_a, d_a)$ (see Equation (6)) separately by estimating a logit.[23]

[20] The numerical solution method we employ follows closely Keane and Wolpin (1994). However, because the state space of our model is, conditional on type, relatively small (roughly 150,000 elements at age $A = 34$), we do not use Keane and Wolpin's method for approximating the expected continuation values using only a subset of the state space. Instead we approximate the expected continuation value for *every* element of the state space by Monte Carlo integration. Based on sensitivity analysis, we chose to rely on 150 draws from the ϵ distribution to perform this integration.

[21] The age at which we last observe the agent in the data A^i is not, in most cases, the terminal age in the model A. We solve for all decisions up to A, but for each woman with children, only her decisions up to age A^i inform the likelihood.

[22] We thank Ken Wolpin for providing us with these estimates. The state of residence is defined as the state in which the respondent resided at the birth of her first child.

[23] We treat the estimates of these functions as determined. In fact, ignoring the sampling errors associated with these estimates will tend to make the calculated standard errors of our structural estimates too small. We assume that this effect is modest and does not affect our qualitative conclusions.

Given this set of estimated parameters, the remaining parameters of the model, including those in the utility function, the returns functions, and the variance–covariance matrix of the shocks Ω, denoted by $\tilde{\Theta}$, are estimated by maximizing $\tilde{\Theta}$ over a restricted likelihood function:[24]

$$\mathcal{L}\left(\tilde{\Theta};\mathcal{D}\right)=\prod_{i=1}^{N}\prod_{a=a_0^i}^{A^i}\Pr\left[d_a^i,w_a^i\,|\,\mathbf{s}_a^i;\left(\tilde{\Theta},\hat{\theta}_j,\hat{\gamma}\right)\right]. \tag{9}$$

For each observation i, $\Pr[d_a^i, w_a^i \mid \mathbf{s}_a^i; (\tilde{\Theta}, \hat{\theta}_j, \hat{\gamma})]$ is a three-dimensional integral that we approximate using 300 Monte Carlo draws to form kernel-smoothed simulators of the probabilities.[25]

3.1. *Unobserved Heterogeneity.* The likelihood function in (9) applies to a sample that is homogenous except for the following observable initial conditions at the latter of the birth of the first child and the date of first interview: age a_0^i, education g_0^i, work experience $x_{a_0^i}^i$, previous period choice $d_{a_0^i-1}^i$, and the state of residence j^i.[26] The skills and preferences of individuals are likely to vary, however, in unobserved ways that are both persistent and correlated with observed initial conditions. For example, those with greater endowments of unobserved human capital may be more likely to prolong schooling and postpone both childbirth and entry into the workforce.

In order to allow for the possibility of persistent heterogeneity correlated with initial conditions, we posit that agents can be of K possible types, indexed by $k \in \{1, \ldots, K\}$, and allow different types of agents to differ, as we briefly mentioned in Section 2, in their home production skill endowment e_0, unobservable labor market skill endowment h_0, welfare stigma ϕ, and nonwelfare home production decay parameter η. In our estimation, these parameters will be type specific and denoted by $e_0^{(k)}$, $\phi^{(k)}$, $h_0^{(k)}$ and $\eta^{(k)}$ for each $k \in \{1, \ldots, K\}$, respectively.[27]

The ex ante probability that an agent i is of type k is denoted by P_k^i. In order to capture correlation between an agent's unobservable type and her initial conditions, we allow P_k^i to depend on all of her observable initial conditions except state of residence in the form of a multinomial logit.[28] That is, for $k = 2, \ldots, K$,

[24] In order to ease identification and the computational burden, we make the relatively standard assumption that $\mathrm{Cov}(\epsilon_{0a}, \epsilon_{1a}) = \mathrm{Cov}(\epsilon_{1a}, \epsilon_2 a) = 0$. The remaining elements of the variance–covariance matrix ($\mathrm{var}(\epsilon_{0a})$, $\mathrm{var}(\epsilon_{1a})$, $\mathrm{var}(\epsilon_{2a}, \mathrm{Cov}(\epsilon_{0a}, \epsilon_{2a}))$) are estimated.

[25] We chose 300 draws after tests for sensitivity of the simulated probabilities and data fit to changes in the number of repetitions. The kernel of the simulated integral is given by

$$\exp\left[\frac{Q_d^a-\max\limits_{d\in D}(Q_d^a)}{\tau}\right]\Big/\sum_{d=0}^{2}\exp\left[\frac{Q_d^a-\max\limits_{d\in D}(Q_d^a)}{\tau}\right],$$

where $Q_d^a = R_a(d; \mathbf{s}_a, \epsilon_a) + \beta\delta \mathrm{E}[V_{a+1}(\mathbf{s}_{a+1}, \epsilon_{a+1}; \tilde{\boldsymbol{\sigma}}_{a+1}^{+}) \mid \mathbf{s}_a, d]$ is the present value of choosing alternative d at period a and τ is the smoothing parameter. In the estimation results that follow, τ is set to 150, again based on sensitivity analysis. For a related application of this kernel smooother, see Eckstein and Wolpin (1999).

[26] In fact the initial levels of work and welfare experience do not vary much in our NLSY79 subsample. Just 14% had more than a year of work experience before entering our subsample, and just 8% had received welfare before first being surveyed in 1979.

[27] Given that estimation of discount factors has proven problematic in some settings roughly similar to ours (e.g., van der Klaauw, 1996; Rust and Phelan, 1997; Eckstein and Wolpin, 1999) we did not allow heterogeneity in discount factors. In our estimation, we choose $K = 3$ after sensitivity analysis; experimenting with a model with fourth types substantially increased computation time but did not show promise of significantly improving within-sample fit, or the likelihood. Computation costs dissuaded us, however, from pursuing the model with four types to the point where likelihood maximization routine converged.

[28] We omit state of residence because the variation in welfare benefits in the data provides an important source of identification for the model's parameters. Allowing type to depend on state of residence would weaken our ability to identify, in particular, unobserved home production and welfare stigma parameters from variation in decisions correlated with variation in initial conditions, welfare benefits, and wages.

$$P_k^i = P_k\left(\mathbf{s}_{a_0^i}; \boldsymbol{\pi}\right)$$
$$= \frac{\exp\left[\pi_0^{(k)} + \pi_1^{(k)} a_0^i + \pi_2^{(k)} g_0^i + \pi_3^{(k)} x_{a_0^i}^i + \pi_4^{(k)} \mathrm{I}\left(d_{a_0^i-1}^i = 0\right) + \pi_5^{(k)} \mathrm{I}\left(d_{a_0^i-1}^i = 1\right)\right]}{1 + \sum_{l=2}^{K} \exp\left[\pi_0^{(l)} + \pi_1^{(l)} a_0^i + \pi_2^{(l)} g_0^i + \pi_3^{(l)} x_{a_0^i}^i + \pi_4^{(l)} \mathrm{I}\left(d_{a_0^i-1}^i = 0\right) + \pi_5^{(l)} \mathrm{I}\left(d_{a_0^i-1}^i = 1\right)\right]}$$

and normalize $P_1^i(\mathbf{s}_{a_0^i})$ as

$$P_1^i = P_1\left(\mathbf{s}_{a_0^i}; \boldsymbol{\pi}\right)$$
$$= \frac{1}{1 + \sum_{l=2}^{K} \exp\left[\pi_0^{(l)} + \pi_1^{(l)} a_0^i + \pi_2^{(l)} g_0^i + \pi_3^{(l)} x_{a_0^i}^i + \pi_4^{(l)} \mathrm{I}\left(d_{a_0^i-1}^i = 0\right) + \pi_5^{(l)} \mathrm{I}\left(d_{a_0^i-1}^i = 1\right)\right]},$$

where $\boldsymbol{\pi} \equiv \{\pi_0^{(l)}, \ldots, \pi_5^{(l)}\}_{l=2}^{K}$. Now write $\tilde{\boldsymbol{\Theta}}^k$ as the set of model parameters for type-k agent to be estimated by simulated maximum likelihood; the sample likelihood, integrating over all types, can be written as

$$\tilde{\mathcal{L}}(\tilde{\boldsymbol{\Theta}}^1, \ldots, \tilde{\boldsymbol{\Theta}}^K, \boldsymbol{\pi}; \mathcal{D}) = \prod_{i=1}^{N} \sum_{k=1}^{K} P_k\left(\mathbf{s}_{a_0^i}; \boldsymbol{\pi}\right) \prod_{a=a_0^i}^{A^i} \Pr\left[d_a^i, w_a^i \mid \mathbf{s}_a^i; (\tilde{\boldsymbol{\Theta}}^k, \hat{\theta}_j, \hat{\gamma})\right]. \tag{10}$$

3.2. *Identification of β and δ.* We now consider the issue of identification of the discount parameters β and δ. In some models, the decisions of sophisticated present-biased agents are observationally equivalent to those of time-consistent exponential discounters, and identification of these two parameters is thus precluded. For example, Barro (1999) demonstrates the observational equivalence in a growth model with sophisticated agents, perfect credit markets, and log utility. This equivalence does not hold more generally. Harris and Laibson (2001) show, for example, that observational equivalence is not obtained when the assumption of perfect credit markets is relaxed. This illustrates that, as is true in any structural empirical paper, the ability to separately identify β and δ results from both the structure imposed by the model and the variation in the data. In what follows, we approach the identification questions from three different angles.

Formal Arguments for Distinguishing Exponential and Hyperbolic Discounting in a Simpler Model. In a related paper (Fang and Silverman, 2006), we studied the identification of exponential and hyperbolic discounting in a somewhat simpler model of welfare program participation for single mothers that contains most of the central elements of the one we estimate here. In that paper we show that, if there are three or more periods of observations, then a present-bias model with $\beta \in (0, 1)$, $\delta \in (0, 1)$ can be distinguished from exponential discounting model with $\beta = 1$ and $\delta \in (0, 1)$ using standard data and without making parametric assumptions on the distribution of the stochastic shocks to payoffs (see Proposition 2 of Fang and Silverman, 2006).[29] In other words, if standard data were generated by a model of (β, δ) discounting, there exist no parameters of the model with time-consistent discounting ($\beta = 1$) that could rationalize those data. The identification argument we presented there uses ideas first presented in Hotz and Miller (1993): The standard data contain information about individuals' choices in each period and thus provide information about conditional choice probabilities.

[29] Standard data sets are formally defined there as data sets that contain information about individuals' choices each period, all relevant state variables, welfare benefit levels, and accepted wages.

The conditional choice probabilities, together with the accepted wage and welfare benefit information in the data, allow us to calculate the continuation values for each choice à la Hotz and Miller. These continuation values then put restrictions on the choice probabilities in the previous period. With three or more periods of data, an exponential discounting model could not rationalize the choice probabilities if the data were generated by a model of (β, δ) discounting.

The argument for identification presented in Fang and Silverman (2006) would, with suitable adaptation, apply to our current setting, *if we did not introduce unobserved heterogeneity*.[30] As is well known, unobserved heterogeneity of agents would prevent "direct observation" of conditional choice probabilities, a key step of Hotz and Miller (1993)-style identification argument. What is clear from Fang and Silverman (2006), however, is that (β, δ)-discounting does not, per se, create problems for a formal identification proof in this context.

Formal identification in a model with unobserved heterogeneity is hard to establish; but we note two factors that aid us in the current context. First, we have assumed a parametric functional form (normality) on the joint distribution of the shocks to payoffs; second, we have, for the typical member of our sample, many more than three periods of data.

Intuitive Arguments for Identification. Less formally, there are three important patterns in the data that together reflect, in the context of our model, time-inconsistency in behavior. The first pattern is the very low levels of work when young: By as late as age 20, just 11% of the sample is working. The second important pattern is the relatively high levels of work when older: Among 32-year-olds in our data, 42% are working.[31] Finally, the data reveal substantial returns to experience in the labor market.[32]

Given the returns to experience and the modest AFDC benefits in most states, the low levels of work when young imply considerable impatience, though not necessarily the present-bias that would generate time-inconsistency. The eventual transition by many women into work, however, demands substantial future orientation—workers have to anticipate the relatively steep *growth* in wages as they accumulate experience. We assume that tastes for leisure do not depend on age over the relevant age range. Thus, given that real wages do not fall with experience,[33] in a deterministic setting this combination of behaviors would clearly be time-inconsistent: If an agent had planned to work eventually, she should have started working immediately. Such a delay is a hallmark of naive, present-biased agents whose false belief that they will embark on a career next year leads them to postpone entry until a time when the immediate costs are low enough. As Fang and Silverman (2004) show, however, even sophisticated present-biased agents may delay entry into work. This can happen when, for example, the relative return to work is increasing with work experience, but at a decreasing rate.[34] In this circumstance, the sophisticated agent may optimally choose to delay entry into work until a time when the steepest part of the experience-return-to-work profile looms large.

[30] Fang and Silverman (2006) also abstracted from *observed* heterogeneity for expositional simplicity, but it is clear that the identification proof goes through unchanged as long as we condition on the observed heterogeneity.

[31] These differences in work levels by age are not merely due to the different patterns of work and welfare among those women who first gave birth, and thus entered the sample, at older ages. For example, among the women in our sample who had children and were not working at age 20, 36% were working at age 32.

[32] Our estimates of the returns to experience are determined, in part, by the selection mechanism implied by our model. However, Loeb and Corcoran (2001), using very different methods, estimate remarkably similar returns to experience for a similar sample of women.

[33] Likewise, the continuation payoff from additional experience should be nonnegative. Our point estimates of the continuation value of experience are, indeed, positive in the unrestricted models.

[34] This circumstance is captured by "free ride" outcome in Fang and Silveman (2004). The diminishing marginal return is obtained in the model estimated in the present article because of diminishing marginal wage returns to experience, increasing benefits of welfare and home production as more children arrive, and the decay of welfare stigma.

In a setting with uncertainty, these transitions might also reflect random shocks to the returns to various choices; but the age pattern is clear—the transitions tend to go from home and welfare into work.[35] Thus, simple uncertainty that would generate random transitions would not generate this particular age trend toward work.[36] A similar logic applies to transitions from home into welfare. If a "career" in welfare will be superior to staying home, that career should begin immediately. To the extent that women postpone entry into long-term welfare spells as a way of avoiding stigma, this too reflects, in the context of our model, time-inconsistency. The size of the delays and the eventual long-run rates of work and welfare program participation jointly function to identify short-term and long-term impatience in the model.

An important element of this intuitive argument is that, in this setting, neither work nor welfare is an unavoidable task. Time-consistent agents would postpone unpleasant tasks, if those tasks were unavoidable. They would not postpone tasks that are, on net, valuable to them. Time-consistent agents would therefore either wait forever to work or take welfare or do it immediately. They would not postpone these decisions.

The above simple argument that identification comes from delayed entry into work is complicated in the plausible case that the relative value of work increases with the number of children in a household or with the age of those children. We have accommodated the first possibility by allowing the value of home production to be a (quadratic) function of the number of children at home. In order to keep the state space a manageable size, however, we did not let the age of the children directly affect payoffs. However, even among the oldest women in our sample most continued to have quite young children at home; at age 32, half of the sample had a child younger than 6, and the average age of the youngest child was 6.33 years. More important, the women with younger children were only slightly less likely to work than those with older children; at age 32, 40% of those with a child younger than 6 were working, this number is 43% among those whose children are all older than 6. It thus seems unlikely that the positive relationship between age and work that we observe is driven largely by the increasing ability of women to leave their children unsupervised or in the care of others.

To the extent that identification of the discount parameters depends on the relative payoff to welfare, it depends on how welfare stigma and its decay are modeled. Stigma is an important component of our model in that it helps explain the large fractions who remain at home without work or welfare. We have assumed that stigma lasts for only one period after switching into welfare from some other choice. Estimates of a more flexible form of stigma decay would be of considerable intrinsic interest, but are beyond the scope of this article. Here our goal is simply to approximate stigma and its decay in a reasonable way, and estimate the parameters of the model under the assumption that stigma takes this form.

Practical Identification. Finally, in practice, whether the two discount parameters are separately identified depends on the curvature of the likelihood surface as we vary β and δ. In Figure 1, we present two slices of the log-likelihood surface as a function of δ only, for $\beta = 1$ and $\beta = \hat{\beta} = 0.338$ when other parameters are set at their respective estimates. In order to clearly show the curvature, we use two different scales for the curves. This figure shows that, along these dimensions, the likelihood exhibits considerable curvature and that when β is set to 1 the maximum log likelihood is substantially lower than its maximum when $\beta = \hat{\beta}$.

[35] A model with habituation or, more generally, state-dependent preferences, might also explain this pattern. We view present-biased time discounting as one source of what would appear as state-dependent utility. Also note that our model accommodates several other forms of "structural" state dependence, for example, wages and welfare stigma are allowed to depend on the recent behavior of the agent.

[36] Note, however, that the increasing wage returns from experience imply that, once the agent has been working for a substantial period, the combination of shocks that would induce a switch to welfare or home are rarer. But this does not fully explain the age trend toward work because this same mechanism creates "structural" state dependence in welfare and home choices as well; the value of switching into work from one of these choices declines the longer it has been postponed.

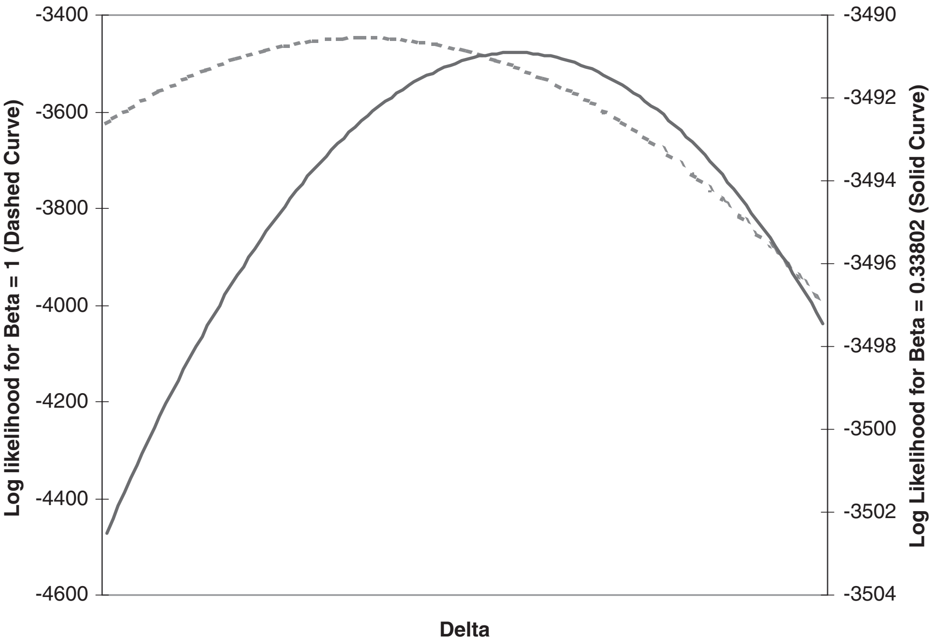


FIGURE 1

THE LOG-LIKELIHOOD SLICES AS A FUNCTION OF δ, FOR $\beta = 1$ (DASHED CURVE, LEFT SCALE) AND $\beta = \hat{\beta} = 0.33802$ (SOLID CURVE, RIGHT SCALE)

4. DATA

4.1. *Sample Definition.* The data are taken from the 1979 youth cohort of the NLSY. The NLSY began in 1979 with 6,283 women (age: 14–22 years), and has interviewed this cohort annually up to 1994 and biannually since 1994. We restrict attention to the 675 women who, as of their interview in 1992, had both remained unmarried and had at least one child during the years they were surveyed. We then consider only the decisions each individual made after the birth of her first child and during the calendar years 1978–1991.

Our purpose in selecting this subsample of individuals and years is threefold. First, to be consistent with our model, we want to restrict attention to those who, if they do not work, are almost certainly eligible for welfare by virtue of having a child and being unmarried.[37] Second, to justify better our assumption that anticipated changes in marital status are not driving work and welfare decisions, we restrict attention to women who never marry during the period observed.[38] Third, we want to limit our analysis to decisions made before the changes in welfare eligibility rules beginning in 1993 and perhaps anticipated by 1992. Finally, again to ease the computational burden, we further limit our sample to residents of the 20 U.S. states best represented in the data. This final restriction leaves us with 483 individuals taken from the NLSY's core random sample and its oversamples of blacks and Hispanics.[39] These sample selection criteria naturally suggest

[37] During the sample period, a parent's AFDC eligibility was determined largely by family structure (the number of dependent children living at home) and income and asset levels. Although in many states married couples were technically eligible for benefits, in practice income and asset restrictions made it very unlikely that married couples would receive benefits.

[38] In fact, by 1993, 2.9% of the sample is observed to be married. This number rises to 10.1% by 1996 and 16.4% by 2002. The potentially interesting effects of anticipated changes in marital status are beyond the scope of this article.

[39] The restriction to never-married women with children is especially important and leaves us with a subsample that is disproportionately drawn from the survey's oversamples of blacks and Hispanics (just 32% of our sample is drawn from the core of the NLSY). Our sample is therefore disproportionately non-white (80%). For purposes of comparison, during the period we study, approximately 50% of the parents in AFDC families were never married and 62% were non-white (Department of Health and Human Services, 1996).

caution in generalizing the estimates in this article to the overall population.[40] The women in our subsample were observed with at least one child for an average of 9.3 of the 14 years from 1978 to 1991, providing us with 4,487 state-choice observations for the estimation.

4.2. *Period and Variable Definition.* At each interview, the NLSY collects welfare participation data as a monthly event history recorded back to the preceding interview. The survey's employment data are collected as a weekly event history. We assume that the decision period of the model corresponds to a calendar year, and identify an agent as age a in a year if she was a years old for at least half that year. The decisions at each age a are defined as follows: An individual chose welfare at age a if she received AFDC for at least six months of the year during which she was a years old. An individual chose work at age a if she was employed for at least 1,500 hours of the year during which she was a years old. An agent chose to stay home if she chose neither of the above.[41,42]

4.3. *Descriptive Statistics.* Descriptive statistics of the subsample are presented, by age, in Table A.1 of the Appendix. Because none of the women in the subsample marries during the period she is observed, the group we study is not typical of the general U.S. population. In order to better understand the ways in which members of the subsample differ from the average population, Table A.1 also compares their statistics with those of the entire sample of women in the NLSY from 1978 to 1991. Broadly, this comparison suggests that although the subsample represents the targets of the U.S. welfare policy, it is atypical of the population as a whole.

The distribution of choices among welfare, work, and home is presented by age in Table 1. We concentrate on the decisions made at ages 16–32, which represents 98% of the data. The fraction of the subsample choosing welfare increases considerably between ages 16 and 22. Of the 16-year-olds with at least one child, 32% chose welfare whereas 54% of 22-year-olds with children chose welfare. The proportion choosing work exhibits a comparable increase over the same period, rising from 0% of 16-year-olds with children to 17% of 22-year-olds. Given these changes in welfare and work participation we, by definition, observe a more dramatic decline in the fraction of women with children choosing to remain at home, with 68% choosing to stay home at age 16 and just 29% choosing to stay at home at age 22.

Although these basic trends continue for the fractions choosing work and home beyond age 22, the fraction choosing welfare stops increasing and instead exhibits a slow decline after age 22. By age 25, 47% of the sample is now choosing welfare, despite having on average more children. By age 29, the fraction is 43%.

Not all of the movements in these age-decision profiles reflect the changing choices of the same individuals. The observed transitions are partly due to the fact that the composition of the sample is changing as the women of the NLSY age and, by virtue of having a child, join the subsample. In order to investigate the degree to which the choices of same individuals change over time, Table 2 presents the one-period transition rates between decisions by the same agent. Here, we see evidence of considerable persistence in individuals' choices. The rows of Table 2

[40] Selection on time preferences may be a particular concern. Women enter the sample only if they have children. Thus our data consist disproportionately of those who had their children at earlier ages. To the extent that this group is more present-biased than average, our estimates will be less applicable to other groups. We thank an anonymous referee for pointing this out.

[41] Although some who are coded as on AFDC or as staying home also report working for pay, their work hours are low. The average annual hours worked among those classified as choosing AFDC or home were 179 and 167, respectively. (The low work levels among those on welfare may be due to a high effective tax on earnings. During the sample period welfare recipients could keep the first $30 per month in earnings. Beyond $30, earnings were taxed at a minimum rate of 67%.) On 19 occasions a respondent reported that she both received AFDC for at least 6 months of the previous calendar year *and* worked more than 1,500 hours that year. In these cases the agent was defined as having chosen welfare.

[42] In about 9% of our observations, the respondent was attending school. This part of the sample is concentrated among agents younger than 18. These observations are also concentrated in the sample we classify as "choosing home," among whom 15.7% were attending school.

TABLE 1

CHOICE DISTRIBUTION, AGES 16–32, NLSY SAMPLE OF NEVER-MARRIED WOMEN WITH AT LEAST ONE CHILD, 1979–1991

	Welfare		Work		Home		Total	
Age	Percent	Number	Percent	Number	Percent	Number	Percent	Number
16	31.9	15	0.0	0	68.1	32	100.0	47
17	38.2	34	0.0	0	61.8	55	100.0	89
18	38.5	60	1.9	3	59.6	93	100.0	156
19	46.6	109	8.6	20	44.9	105	100.0	234
20	50.2	143	11.9	34	37.9	108	100.0	285
21	50.5	165	14.1	46	35.5	116	100.0	327
22	53.7	188	16.9	59	29.4	103	100.0	350
23	51.2	191	20.6	77	28.2	105	100.0	373
24	48.5	182	25.6	96	25.9	97	100.0	375
25	47.3	187	27.1	107	25.6	101	100.0	395
26	48.6	196	30.5	123	20.8	84	100.0	403
27	44.3	167	32.1	121	23.6	89	100.0	377
28	45.1	142	33.0	104	21.9	69	100.0	315
29	42.8	109	37.7	96	19.6	50	100.0	255
30	47.9	91	35.3	67	16.8	32	100.0	190
31	43.1	62	39.6	57	17.4	25	100.0	144
32	35.6	32	42.2	38	22.2	20	100.0	90
Total	47.1	2073	23.8	1048	29.1	1284	100.0	4405

TABLE 2

YEAR-TO-YEAR CHOICE TRANSITION MATRIX, NLSY SAMPLE OF NEVER-MARRIED WOMEN WITH AT LEAST ONE CHILD, 1979–1991

	Choice at t		
Choice at $t-1$	Welfare	Work	Home
Welfare			
Row %	84.3	3.5	12.3
Column %	76.7	6.3	17.9
Work			
Row %	5.3	79.3	15.3
Column %	2.6	76.4	12.1
Home			
Row %	28.3	12.0	59.7
Column %	20.7	17.3	70.0

represent the choices made in period $t-1$; the columns describe the choices made in period t. The top figure (Row %) in each cell represents the fraction of the subsample that made the row choice in period $t-1$ who went on to make the column choice in period t. The bottom figure (Column %) in each cell shows the fraction of the subsample that made the column choice in period t who made the row choice in the previous period. We find that 84.3% of those who chose welfare in period $t-1$ went on to choose it again in period t. Conversely, of those who chose welfare in period t, 76.7% had chosen welfare in the previous period. Of those who chose work in period $t-1$, 79.3% went on to choose it again in period t. Decisions to remain at home are considerably less persistent. Of those who chose to stay home in period $t-1$, 59.7% chose it again in period t.

5. RESULTS

5.1. *Estimates of Welfare Benefit Function G_j and Fertility Function ρ.* Table A.2 in the Appendix presents the parameters of the benefit rule for the 20 selected U.S. states used in our estimation. The benefits include the cash value of AFDC plus food stamps. As has been often noted, there is considerable variation in benefits levels across the U.S. states. In our sample, the estimated average annual benefit for a mother with two children ranges from \$4,856 (1987 dollars) to \$9,490. Patterns of welfare participation vary with the level of benefits in ways consistent with optimizing behavior. In our sample, 56% of the residents in the five states with the highest benefits received welfare, whereas 37% of these in the five states with the lowest benefits were on welfare.[43]

Table A.3 in the Appendix presents the parameter estimates of the fertility function (6). These estimates suggest that the probability of an additional birth is decreasing with age and with the number of children. The estimates also indicate that, relative to those who stay home, the probability of an additional birth is lower for workers and higher for those on welfare. We note, however, that our simple exogenous model of subsequent fertility beyond the first child explains very little of the variation in the timing of births in this subsample. The pseudo-R^2 is less than 0.02.

5.2. *Parameter Estimates.* In our estimation, we assume that agents are of three possible types, i.e., $K = 3$. Tables 3 and 4 present the parameter estimates under three different restrictions of the model. Column (1) presents estimates when we restrict agents to be time-consistent, that is, restricting $\beta = 1$; Column (2) presents estimates when agents are assumed to be sophisticated (i.e., $\tilde{\beta} = \beta$) and allowed to be present-biased; and Column (3) presents the estimates when agents are assumed to be completely naive ($\tilde{\beta} = 1$) and allowed to be present-biased.[44] We present both the point estimates and their asymptotic standard errors.[45]

In the sophisticated present-bias model, the estimated present-bias factor β equals 0.338 with a reasonably small standard error of 0.069. A Wald test rejects the hypotheses of time-consistency (t-statistic 9.53 against the null of $\beta = 1$). Our estimate of the standard discount factor δ equals 0.88 with a standard error of 0.016. Allowing for present-bias improves the data fit in a statistically significant way—a likelihood-ratio test easily rejects the time-consistent model (the χ^2 statistic for the likelihood-ratio test is more than 32). However, the likelihood-ratio test does not yield overwhelming evidence in favor of the completely naive or sophisticated model.[46] In what follows, we will focus on the results from the sophisticated present-biased agent model.[47]

Besides the discount factors, Tables 3 and 4 also present estimates, by (unobservable) type, of the net welfare stigma, home production functions, wage functions, continuation value functions, and variance–covariance matrix of the shocks, etc. Of particular interest are the substantial estimated return to experience in the wage offer function and the considerable variation in the estimated skills and tastes across types. There is an important average gain in wages for every year of additional work experience. The unobservable skill levels that determine those wages vary importantly, however, by type.

[43] In order to keep computation costs in check, we do not accommodate the changes in welfare benefits with calendar year. The time pattern of AFDC benefits differed from that for food stamps. Inflation outstripped substantial increases in nominal AFDC benefits and led to an 11% decline in real average benefits during the period 1980 to 1984. Average real benefits then increased by a total of 3.6% over the next three years before declining again at an average annual rate of about 1.5% for the next four years (Crouse, 1995). The declines in AFDC benefits were somewhat offset, especially at the end of the sample period, by increases in food stamp benefits. Our method approximates these nonmonotonic real benefits profiles with a state-specific flat line.

[44] Except when β is restricted to equal 1, we make no restrictions on the values the discount parameters may take. In particular, β and δ are each allowed to be greater than one or negative.

[45] Asymptotic standard errors are estimated using the BHHH, or outer product of gradients, method. See Berndt et al. (1974).

[46] Technically we cannot use a likelihood-ratio test to distinguish completely naive and sophisticated present-biased models because they are not nested. Note, however, that both the Akaike information criterion and the Bayesian (Schwartz) Information Criterion, which account for differences in the number of parameters in nonnested models, reduce to selecting the model with the highest likelihood in cases like this where each model has the same number of estimated parameters. We are grateful to two referees for pointing this out.

[47] The key simulation results for naive present-biased are, for completeness, included in the Appendix.

TABLE 3

PARAMETER ESTIMATES FOR TIME CONSISTENT, SOPHISTICATED, AND NAIVE PRESENT-BIASED AGENTS

Parameters		(1) Time Consistent		(2) Present-Biased (Sophisticated)		(3) Present-Biased (Naive)	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
		Preference Parameters					
Discount factors	β	1	n.a.	0.33802	0.06943	0.355	0.0983
	δ	0.41488	0.07693	0.87507	0.01603	0.868	0.02471
Net stigma	$\phi^{(1)}$	7537.04	774.81	8126.19	834.011	8277.46	950.77
(by type)	$\phi^{(2)}$	10100.9	1064.83	10242.01	955.878	10350.20	1185.27
	$\phi^{(3)}$	13333.2	1640.18	12697.25	1426.40	12533.69	1685.92
Home production	$e_0^{(1)}$	2684.97	427.85	2209.48	405.26	2224.98	456.85
(by type)	$e_0^{(2)}$	3324.79	516.96	3502.66	509.07	3492.15	617.64
	$e_0^{(3)}$	1729.53	1418.21	2126.86	879.54	2182.17	1227.66
	e_1	84.83	441.45	124.92	48.95	121.58	130.57
	e_2	−36.21	105.61	−603.29	215.67	−608.39	560.31
	$\eta^{(1)}$	2484.69	494.09	4565.06	399.07	4588.88	756.19
	$\eta^{(2)}$	4432.11	573.40	6547.94	503.62	6557.07	933.40
	$\eta^{(3)}$	9858.23	1290.18	12149.5	869.089	12054.63	1670.74
		Wage and Skill Parameters					
Constant	$h_0^{(1)}$	0.12881	0.09963	0.16329	0.0676	0.1672	0.1362
(by type)	$h_0^{(2)}$	0.59176	0.10073	0.6121	0.06828	0.61628	0.13625
	$h_0^{(3)}$	1.11547	0.12045	1.10907	0.08089	1.12299	0.14646
Years of schooling	α_1	0.01995	0.0082	0.02153	0.00501	0.02166	0.00976
Experience	α_2	0.13513	0.01056	0.12252	0.00853	0.12142	0.01203
Experience^2	α_3	−0.00736	0.0009	−0.00623	0.00068	−0.00605	0.00099
1st year experience	α_4	0.09352	0.04291	0.06681	0.02949	0.06742	0.04535
Experience decay	α_5	−0.22702	0.03601	−0.23105	0.03096	−0.23694	0.03731

Discussion of the Discount Factor Estimates. Our estimates of present-bias factor β at 0.338, combined with the estimated standard discount factor $\delta = 0.88$, implies a one-year ahead discount rate of 238%. Our estimate of the present-bias factor is low relative to most of those estimated in experimental studies, though more similar to Paserman's (2008) structural estimate for low-wage workers. Inferential studies such as Hausman (1979) and Warner and Pleeter (2001) estimate discount rates ranging from 0% to 89% depending on the characteristics of the individual and intertemporal trade-offs at stake. Paserman finds, for low-wage workers a discount rate of about 149%. For their benchmark model and calibration, Laibson et al.'s (2007) point estimates of β and δ are, respectively, 0.7031 (with standard error of 0.1093) and 0.9580 (with standard error of 0.0068), which imply a one-year ahead discount rate of 48.5%.

There are two possible explanations for the difference between our finding and others. First, the samples cover different subpopulations. Our sample selection criteria, which restrict our analysis to mostly poor, never-married women who had children at relatively young age, had relatively low schooling, and did not move across states of residence as much as the population (see Table A.1), may have led to a subpopulation who is most susceptible to present-biases.[48] For the purpose of welfare policymaking, however, our subsample may be the relevant subpopulation to study. The second potential explanation for the difference between our findings and others is simply that different papers focus on different spheres of decision making, and it is possible that the magnitudes of present bias differ by a realm of decision.

[48] Hausman (1979), Lawrance (1991), and Paserman (2008) all find that discount rates are higher for low-income groups.

TABLE 4

PARAMETER ESTIMATES FOR TIME CONSISTENT, SOPHISTICATED, AND NAIVE PRESENT AGENTS (CONTINUED FROM TABLE 3)

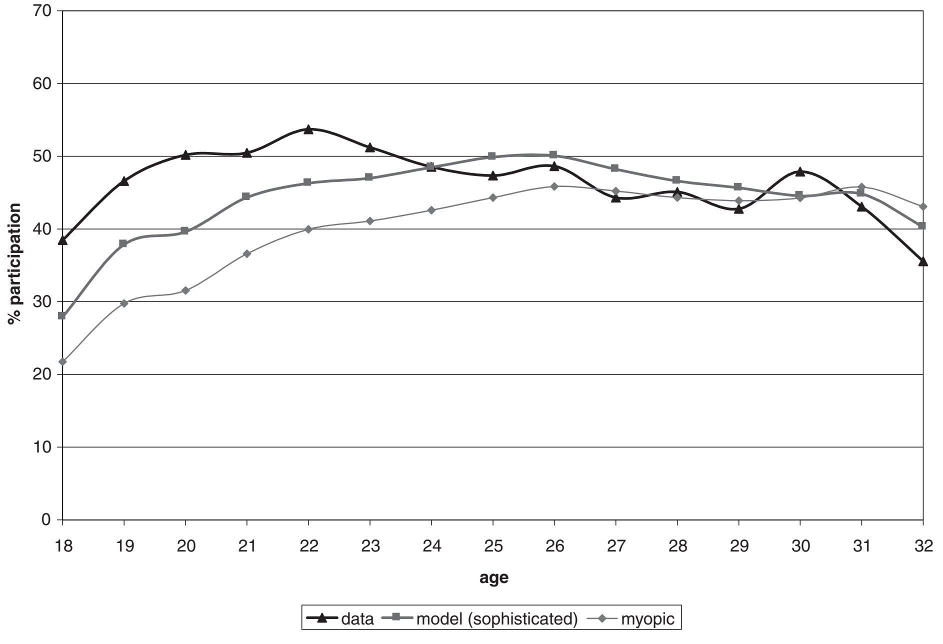
Parameters		(1) Time Consistent Estimate	(1) S.E.	(2) Present-Biased (Sophisticated) Estimate	(2) S.E.	(3) Present-Biased (Naive) Estimate	(3) S.E.
Continuation Value Function at Age 35							
Number of children	ω_1	794.52	743350.2	2618.55	3511.39	2496.75	197163.19
Number of children^2	ω_2	−8938.74	82101.40	−8918.7	5258.05	−8638.95	27929.24
Experience	ω_3	62.74	20429.20	235.24	268.94	231.11	4500.37
Experience^2	ω_4	−54.47	516.04	378.36	115.00	374.21	185.64
Welfare lag	ω_5	2617.59	7515.73	8707.61	6322.23	8725.00	10638.20
Work lag	ω_6	1544.06	13820.09	6151.05	4142.20	6260.41	14140.67
Log Odds as Function of Initial Conditions for Types 2 and 3							
Type 2: Constant	$\pi_0^{(2)}$	−1.842	1.544	−1.070	1.550	−1.179	1.593
Age	$\pi_1^{(2)}$	0.0067	0.087	−0.0406	0.086	−0.0385	0.0867
Years of schooling	$\pi_2^{(2)}$	0.129	0.124	0.133	0.122	0.139	0.127
Experience	$\pi_3^{(2)}$	0.217	0.194	0.227	0.190	0.221	0.187
Welfare lag	$\pi_4^{(2)}$	0.0865	0.662	0.398	0.618	0.406	0.633
Work lag	$\pi_5^{(2)}$	0.0131	0.578	0.062	0.587	0.0576	0.578
Type 3: Constant	$\pi_0^{(3)}$	−3.948	2.423	−5.627	2.328	−5.562	2.273
Age	$\pi_1^{(3)}$	−0.687	0.126	−0.360	0.168	−0.356	0.167
Years of schooling	$\pi_2^{(3)}$	1.303	0.156	0.9322	0.268	0.918	0.263
Experience	$\pi_3^{(3)}$	0.1055	0.2811	0.314	0.278	0.318	0.277
Welfare lag	$\pi_4^{(3)}$	−0.526	1.252	−0.640	1.508	−0.463	1.305
Work lag	$\pi_5^{(3)}$	0.575	0.881	−0.13387	0.874	−0.144	0.846
Variance and Covariance of Shocks							
Std. dev. of ϵ_0	σ_{ϵ_0}	5262.40	548.55	5656.61	446.56	5708.50	579.19
Std. dev. of ϵ_1	σ_{ϵ_1}	0.3751	0.0122	0.3726	0.0071	0.3707	0.0076
Std. dev. of ϵ_2	σ_{ϵ_2}	4168.06	334.76	4116.96	331.49	4074.99	459.82
$\text{cov}(\epsilon_0, \epsilon_2)$	$\sigma^2_{\epsilon_0 \epsilon_2}$	−3046.77	168.32	−2849.19	202.06	−2861.02	247.60
Log-likelihood		−3505.96		−3489.80		−3486.44	
χ^2-Statistics		32.32		n.a.		6.72	

NOTE: χ^2 statistics are calculated under the null hypothesis of the present-biased sophisticated model.

5.3. *Within-Sample Fit*

Age-Choice Profiles. Summarizing the interaction of the potentially complex and countervailing effects of time preferences and basic incentives, Figures 2–4 compare the estimated model's predicted distributions over the three alternatives (welfare, work, and home) to the actual distributions in the data, by age. The model's predictions represent the simulated decisions of 1,000 agents in each of 16 cells defined to reflect the sample variation in initial conditions j, a_0, g_0, x_{a_0}, and d_{a_0-1}. There are four different j categories defined as high, medium-high, medium-low, and low benefits municipality. Similarly there are four g_0 categories defined as 10 years of schooling or less, 11 years of schooling, 12 years of schooling, and some college at the birth of the first child. Within each of these 16 cells, the initial conditions are given by the sample average (benefits, age, schooling, experience) level in the cell. These sample averages imply probabilities of the agent being of the three different unobservable types. The distribution of the 1,000 simulated decisions in each of these cells is then weighted by the probability of each type and the proportion of the data falling into that initial condition cell to generate the predicted distributions appearing in Figures 2–4.

The simulated age profiles match the data reasonably well. Each of the profiles implied by the estimated model assumes approximately the correct shape and often matches the levels of



NOTES: The simulation of myopic agents simply sets β and δ to zero. For that simulation, all other parameters are set to the values estimated under the assumption of sophisticated agents.

FIGURE 2

AGE-WELFARE PARTICIPATION PROFILES: DATA VERSUS MODEL SIMULATION FOR SOPHISTICATED AND MYOPIC AGENTS

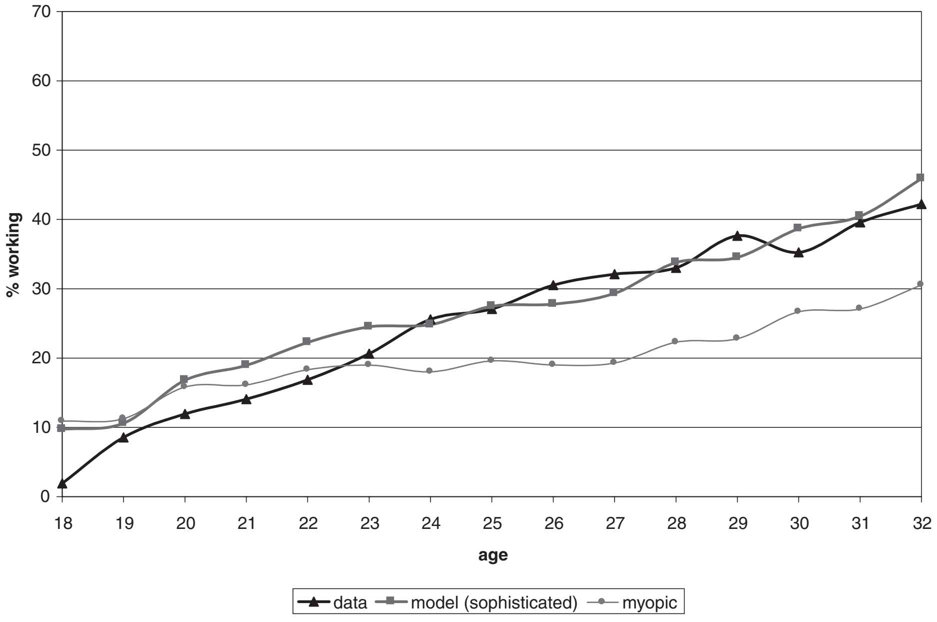
the data quite closely. More formally, Table A.4 of the Appendix presents the within-sample χ^2 goodness-of-fit statistics for the model with respect to the choice distribution, by age.[49] Note that among the 15 age groups we examine, the within-sample χ^2 goodness of fit rejects the null hypothesis of no difference between actual and predicted probabilities in six instances (among a total of 30 statistics, two for each age group). Although the fit is not perfect, we would like to emphasize that our estimation did not directly use these moment restrictions. These statistics confirm the impression given by Figures 2–4.

Our parameter estimates indicate a high degree of short-term impatience; the one-year-ahead discount factor is just 0.29. A natural question is whether, practically speaking, the behavior of agents with such limited patience is meaningfully different from that of agents with no concern for the future. In order to shed some light on this issue, we simulate behavior for the case where, holding all other parameters constant at their estimated values, agents are assumed to be completely myopic ($\beta = \delta = 0$). These are *not* simulations from the estimates of a completely myopic model (with β and δ restricted to zeros); rather, they merely reflect how behavior would look in the previously estimated environment if agents were completely myopic.[50] The age-choice profiles for these simulations are also displayed in Figures 2–4. The behavior of myopic agents is qualitatively different. Most important, myopic agents eventually enter the labor force at a rate (31% by age 32) substantially lower than is predicted for even modestly forward-looking agents (46% by age 32).

Transition Probabilities. Table 5 presents the simulated one-period transition probabilities for the sophisticated present-biased agent model. This table is to be compared with the transition probability matrix in the data (see Table 2). The model matches the persistence and relative

[49] This goodness-of-fit test does not correct for a sampling error.

[50] Thus, we do not attempt to answer the question of whether a model with forward-looking agents fits the data significantly better (in terms of the likelihood function) than one in which agents are assumed to be myopic.



NOTES: The simulation of myopic agents simply sets β and δ to zero. For that simulation, all other parameters are set to the values estimated under the assumption of sophisticated agents.

FIGURE 3

AGE-WORK PROFILES: DATA VERSUS MODEL SIMULATION FOR SOPHISTICATED AND MYOPIC AGENTS

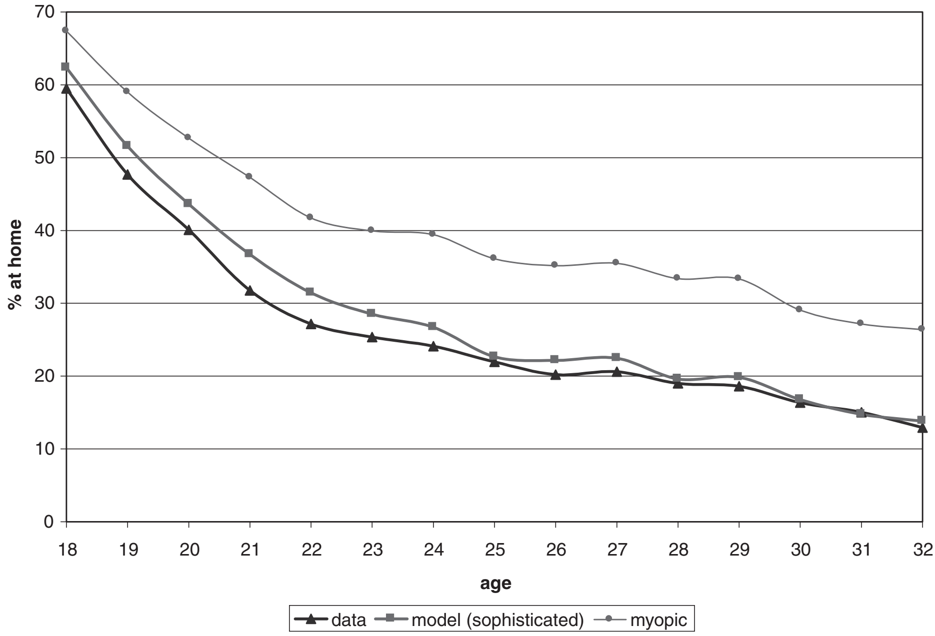
rates of transition quite well. In order to illustrate, the estimated model predicts that 84.4% of those who chose welfare in period $t-1$ will go on to choose it again the following period whereas 11.4% will choose to stay at home. These figures should be compared with 84.3% and 12.3% observed in the data. Similarly the model predicts that 57.0% of those choosing home in period $t-1$ will remain at home next period, whereas 25.9% will switch to welfare, comparable to 59.7% and 28.3%, respectively, in the data.

Wage Profiles. Figures 5 and 6 compare, respectively, the model's mean wage-age and wage-experience profiles, with the parallel moments in the data. Save the outlying wages of age-18 workers, the model somewhat underestimates of average wages for those who choose to work (see Figure 5). Overall, however, the average accepted wages, by age, of the model and data are quite similar. Save the accepted wages of those with no experience, the model slightly underestimates wage levels while replicating the observed shape of the wage-experience profile (see Figure 6).

5.4. *Out-of-Sample Fit.* As we mentioned in Subsection 4.1, we have used only residents of the 20 U.S. states best represented in the NLSY in our empirical estimation. The sample of never-married women with children from the remaining states allows us to examine the estimated model's out-of-sample fit. The "hold-out" sample is much smaller than the estimation sample; it includes 101 individuals and provides just 583 decisions over the relevant years.[51] Although sampling variation will make close quantitative fit unlikely, we view the comparison as informative. Figure 7 compares the proportions of this choosing welfare, work, and home by age predicted by our present-biased sophisticated model using the parameter estimates in Section 5.2 with their empirical counterparts from the hold-out sample.[52] The model captures

[51] The corresponding numbers in the estimation sample are 483 and 4,487.

[52] We do not compare the model's predictions with data from women younger than 20 or older than 29 because, in the left-out sample, the data are especially thin, fewer than 20 observations per age, in these ranges.



NOTES: The simulation of myopic agents simply sets β and δ to zero. For that simulation, all other parameters are set to the values estimated under the assumption of sophisticated agents.

FIGURE 4

AGE-HOME PROFILES: DATA VERSUS MODEL SIMULATION FOR SOPHISTICATED AND MYOPIC AGENTS

TABLE 5

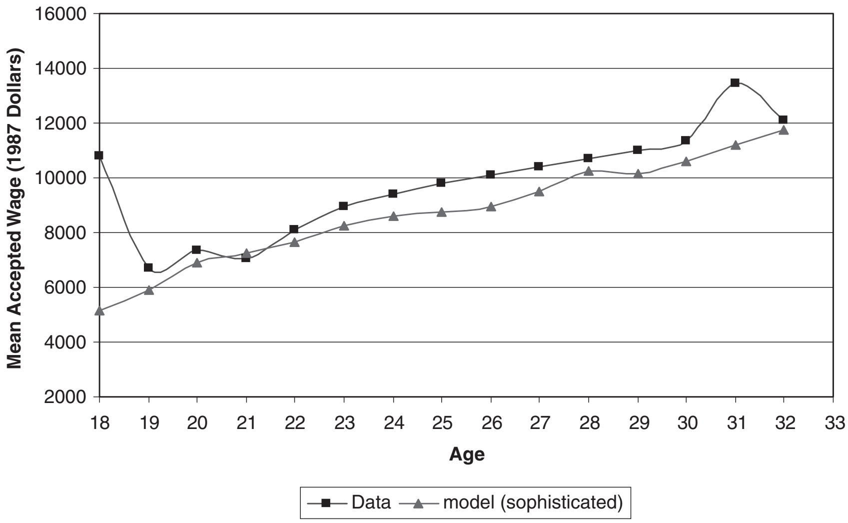
SIMULATED YEAR-TO-YEAR TRANSITION PROBABILITY MATRIX FOR SOHPISTICATED PRESENT-BIASED AGENTS

	Choice at t		
Choice at $t-1$	Welfare	Work	Home
Welfare			
Row %	84.4	4.2	11.4
Column %	78.4	7.4	19.8
Work			
Row %	10.9	74.2	14.9
Column %	5.6	72.7	14.5
Home			
Row %	25.9	17.1	57.0
Column %	15.9	19.9	65.7

the relative shape of changes in the participation rates as the women get older; for example, the model's prediction of the increase in the proportion of working single mothers mirrors that in the data, but it consistently overestimates the proportion of single mothers on welfare and underestimates the proportion at home.[53]

To the extent that the model mispredicts behavior out of sample, it suggests caution in interpreting the counterfactual experiments simulated below. For this reason, and others, we are reluctant to put a great deal of stock in the precise *quantitative* predictions of the policy responses. However, we view the *qualitative* predictions of the estimated model as informative

[53] We choose not to present the formal out-of-sample goodness-of-fit test because it will not be very informative due to the small sample size in the hold-out sample.



NOTES: Accepted wages for 18-year-olds are available for just three observations.

FIGURE 5

MEAN ACCEPTED WAGES FOR WORKERS BY AGE: DATA VERSUS MODEL SIMULATION FOR SOPHISTICATED AGENTS

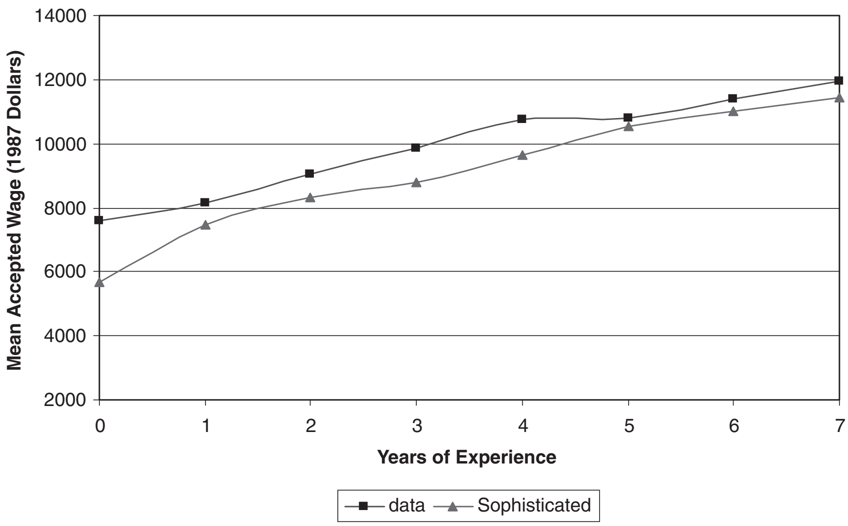


FIGURE 6

MEAN WAGE-EXPERIENCE PROFILES: DATA VERSUS MODEL SIMULATION FOR SOPHISTICATED AGENTS

and useful. They give a qualitative sense of the importance of time-inconsistency and of the effects of imperfect commitment devices on behavior and utility in this setting and in the context of a model whose parameters have been importantly disciplined by data.

6. NUMERICAL SIMULATIONS

The estimates and simulations presented in Subsection 5.2 indicate that the work–welfare–home decisions of never-married women with children reveal time-inconsistency. With a reasonable degree of precision, the estimated model indicates a present-bias factor (β) substantially less than unity. In this section, we present simulation results for sophisticated

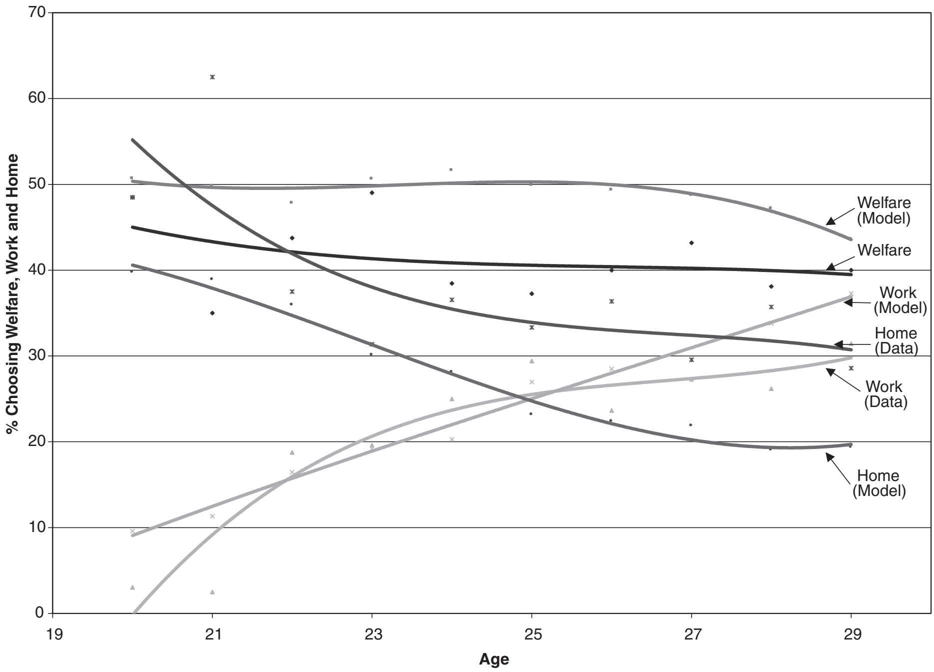


FIGURE 7

AGE-DECISION PROFILE: COMPARISON OF OUT-OF-SAMPLE DATA AND SIMULATION WITH ESTIMATED PARAMETERS (SOPHISTICATED AGENTS)

present-biased agents. Analogous results for completely naive present-biased agents are included in the Appendix.

On its own, an estimated β less than one does not imply that time-inconsistency importantly influences the work-welfare decisions of never-married women with children. It may be that the ability to commit to future decisions influences behavior in statistically identifiable, but economically insubstantial, ways. This possibility is particularly relevant for the model estimated here. In our model, initial conditions such as welfare benefits in state of residence and years of schooling and unobservable skills differ across individuals, and these differences would be expected to importantly influence decision making. Although time-inconsistency in preferences may affect marginal decisions, it may be that the influence of initial conditions typically places individuals far from these margins, and the ability to commit would have little effect on decisions. By extension, if most individuals are little influenced by their inability to commit to future decisions, the behavioral and utility consequences of policies such as time limits or workfare that may serve as commitment devices will not much depend on the time-inconsistency of their targets.[54] In the next sections, we use the estimated parameters of the sophisticated present-biased model to quantify both the behavioral and utility consequences of the ability to commit and consider how different policy reforms affect both behavior and utility in the presence of time-inconsistency.

6.1. *Consequences of an Ability to Commit.* In order to evaluate the consequences of an ability to commit to future decisions, we use the estimated parameters of the sophisticated present-biased model to simulate the decisions of agents with various initial conditions both with and without commitment ability. In this experiment, an individual has commitment ability if, starting from the period in which her first child is born, her future selves behave as though

[54] See Fang and Silverman (2004) for a discussion of how time limits could serve as commitment mechanisms.

they were time-consistent (i.e., $\beta = 1$), and also believed all of their future selves to be time-consistent. Equilibrium behavior represents the optimal plan of an individual considering the sequence of decisions to begin at the birth of her first child.

Evaluating utility effects in a setting with time-inconsistency is often thought to be especially problematic because sequences of utility flows may be valued differently by the different selves of the same individual. In the literature on time-inconsistency, two criteria have been proposed to serve as a basis for comparing an agent's well being: the *Pareto* criterion (Laibson, 1997) and the *long-run utility* criterion (O'Donoghue and Rabin, 1999a). The Pareto criterion asks if all the selves are made better off whereas the long-run utility criterion takes the perspective of an effectively time-consistent agent just prior to the decision-making sequence and asks if she is made better off. From the perspective of policy evaluation, it is not obvious which of these criteria is the more appropriate. Based on its similarity to prior utility evaluations in structural estimation (see, e.g., Keane and Wolpin, 1997), we adopt the latter criterion for estimating changes in well-being. Specifically, we calculate the discounted stream of expected lifetime utility for period-a_0 self, i.e., the self when her first child was born if, counterfactually, $\beta = 1$. We first numerically solve for the perception-perfect strategy profile, denoted by $\boldsymbol{\sigma}^{c*} \equiv \{\sigma_a^{c*}\}_{a_0}^{A}$ for an agent with $\beta = 1$ and $\delta = \hat{\delta} = 0.875$, the point estimate presented in Table 3. Of course, $\boldsymbol{\sigma}^{c*}$ depends on agents' initial conditions at period a_0. Conditional on an agent's initial conditions, her utility with commitment ability is given by

$$U^c = \mathrm{E} \sum_{a=a_0}^{A} \delta^t R_a\big(\boldsymbol{\sigma}^{c*}(s_a, \epsilon_a); s_a, \epsilon_a\big).$$

As a benchmark for comparison, when agents do not have ability to commit, we also numerically solve for the perception-perfect strategy profile, denoted by $\boldsymbol{\sigma}^{n*} \equiv \{\sigma_a^{n*}\}_{a_0}^{A}$ for an agent with $\beta = \hat{\beta} = 0.338$ and $\delta = \hat{\delta} = 0.875$. Conditional on an agent's initial conditions, the utility without commitment ability that we use as benchmark comparison is given by

$$U^n = \mathrm{E} \sum_{a=a_0}^{A} \delta^t R_a\big(\boldsymbol{\sigma}_a^{n*}(s_a, \epsilon_a); s_a, \epsilon_a\big).$$

Note that U^n is not how period-a_0 self would have evaluated the lifetime utility with her (β, δ) preference, but rather how a prior period self would have evaluated the sequence. We report below $(U^c - U^n)/U^n$ as the percentage change in lifetime utility as a result of the ability to commit. Representative results of the simulations are presented, by initial conditions cell, in Table 6. The cells (1–8) vary according to the level of the benefits in the state of residence, age, and years of schooling at first birth, and thus probability of being types 1 and 2. The levels of these initial conditions are presented in second panel in Table 6. The same initial conditions (by cell) are used in subsequent tables.

Panel 1 of Table 6 indicates that, although the behavioral effects of an inability to commit may be large, they differ both in size and sign depending on initial conditions. For example, among individuals in cell 2, who were relatively young and little educated at the birth of their first child and who live in a high benefits state, work is relatively unattractive and commitment ability leads them to work somewhat *less* (2.7%) of the time between ages 18 and 34. For this group, the inability to commit generated costly delay, not in work, but in the takeup of welfare. With commitment ability, they are quicker to endure welfare stigma in exchange for the future benefit of welfare receipt. Compare this effect of commitment to that of similarly young and but better educated individuals in a low benefits state (cell 3). In this second group, for whom working is relatively attractive, the ability to commit leads them to work an additional 23.3% of the time, representing a 66% increase in their probability of working. Comparing across other cells, we observe similar disparities in the behavioral reaction depending on the relative attractiveness of

TABLE 6

SIMULATED EFFECTS OF THE ABILITY TO COMMIT FOR SOPHISTICATED PRESENT-BIASED AGENTS, BY INITIAL CONDITIONS

	Initial Conditions Cell							
	1	2	3	4	5	6	7	8
Panel 1: Simulated Effects								
Change in % working	14.07	−2.74	23.32	−3.38	24.44	−1.71	24.69	13.07
Change in lifetime utility	\$1,908	\$1,632	\$2,756	\$1,660	\$3,318	\$1,713	\$3,702	\$2,508
% Change in lifetime utility	3.41	2.42	4.78	2.37	5.33	2.47	5.03	3.26
Panel 2: Initial Conditions for Different Cells								
Wel. benefits (1 child)	4126.53	7103.51	4103.53	7278.74	4073.25	7116.39	4278.39	7023.98
Wel. benefits (2 children)	5383.13	8781.58	5340.66	8969.49	5315.70	8809.24	5529.23	8746.56
Age at first birth	17	18	19	19	21	20	22	22
Years of schooling	9	9	11	11	12	12	14	14
Work year before first birth	No	No	No	No	Yes	No	No	Yes
Years of work experience at first birth	0	0	0	0	1	0	1	1
Prob (type = 1)	0.622	0.636	0.556	0.556	0.469	0.513	0.335	0.339
Prob (type = 2)	0.356	0.349	0.383	0.383	0.454	0.387	0.383	0.412

NOTE: Panel 1 presents the simulated effects, in terms of behavior (% of time working) and discounted utility (in 1984 dollars), of providing sophisticated agents with perfect commitment ability. These simulations are provided for eight representative cells of initial conditions. The relevant characteristics of those cells are provided in Panel 2, along with estimates of the probabilities that members of that cell are of each of the three possible unobserved types.

welfare and work. Among the more educated and those living in lower welfare benefits states, the ability to commit leads to significantly more work; among those with less education and living in high benefits states commitment generates either little or negative changes in work behavior.

Importantly, the results of Table 6 also indicate that although the behavioral changes produced by an ability to commit may be large, the utility effects are invariably modest. The change in lifetime utility as a result of commitment ranges from \$1,737 (a 5.03% increase) for those in cell 7 with the highest levels of education and medium welfare benefits to \$1,525 (a 5.33% increase) for those in cell 5 with medium levels of education and low welfare benefits and to \$1,092 (a 2.37% increase) for among those in cell 1 with very low levels of education and welfare benefits.

It may seem puzzling that the behavioral effects of commitment could be large whereas the utility gains among the same group are relatively small. This result derives from two mechanisms. First, for those delaying welfare takeup in favor of home, the delay in the absence of commitment is fairly short—typically less than two years. Thus, the cumulative gains are relatively modest. Second, for those delaying entry into the labor force, the delay is typically longer, but the gains are realized only in the relatively distant future. So although it may be optimal from the perspective of the period a_0 self to commit herself to a career of work, the gains from that decision (relative to the decisions made in the absence of commitment) will be realized only after substantial work experience has accumulated and will thus be discounted by time. The costs required in order to acquire that work experience are, on the other hand, realized in the relatively near term and thus are discounted less by time. As a result, from the perspective of the period a_0 self, the net gains from commitment may be relatively small even when the behavioral consequences are substantial. If, however, we evaluate the change in utility from the perspective of the agent in her late 20s, the utility gains from commitment can be as high as 11% of continuation utility.

6.2. *Consequences of Time Limits.* The experiment of the previous section sets an upper bound on the utility gains to potential welfare recipients from commitment. We know that imperfect commitment devices such as time limits and workfare can at best deliver some fraction of these gains.[55] Table 7 presents the results of simulation exercises when we impose welfare

[55] We consider only the utility gains to potential welfare recipients and not the gains to taxpayers from reform.

TABLE 7

SIMULATED EFFECTS OF TIME LIMITS OF VARYING LENGTHS FOR SOPHISTICATED PRESENT-BIASED AGENTS, BY INITIAL CONDITIONS

		Initial Conditions Cell							
Time Limits		1	2	3	4	5	6	7	8
7 years	% change in lifetime util.	−1.86	−9.70	−1.10	−8.10	−0.15	−6.63	−0.14	−1.70
	changes in % working	11.23	22.28	7.44	20.22	2.41	17.80	1.77	9.69
5 years	% change in lifetime util.	−2.52	−13.05	−1.36	−10.83	−0.10	−9.07	−0.17	−2.36
	changes in % working	17.90	33.29	13.01	31.54	4.93	28.48	3.66	16.72
3 years	% change in lifetime util.	−3.53	−17.29	−1.99	−14.23	−0.20	−11.97	−0.23	−3.14
	changes in % working	26.02	43.66	19.74	42.38	8.71	40.73	6.63	24.84
1 year	% change in lifetime util.	−5.54	−22.93	−3.11	−18.96	−0.43	−15.76	−0.32	−4.20
	changes in % working	34.97	55.06	29.18	55.51	14.91	53.74	12.03	35.55
0 year	% change in lifetime util.	−5.55	−23.92	−2.92	−19.66	0.21	−16.12	0.16	−3.51
	changes in % working	40.22	61.03	34.50	61.42	19.40	60.33	15.50	41.90

NOTE: Each row presents the simulated effects, in terms of behavior (change in % of time working) and discounted utility (in 1984 dollars), of introducing a time limit of the length given in the first cell of the row. These simulations are provided for eight representative cells of initial conditions (the columns). See Panel 2 of Table 6 for initial conditions for the different cells.

eligibility time limits of varying lengths. Again we consider the behavioral and utility consequences for individuals with different initial conditions.

Although each of the time limits increases the frequency of work, in doing so they almost always reduce the lifetime utility of individuals in the model. Regardless of the limit's length, the predicted increases in work and decreases in utility are most dramatic for those with little education living in high benefits states (see, e.g., cells 2 and 4). The model implies that time limits are too crude for a commitment device. As they induce more work, time limits fail to increase expected lifetime utility. Though, for those living in low benefits states, and for those with higher levels of skills and education, the utility losses from the actual five-year time limit are quite modest (see cells 1, 3, 5, and 7). Thus, in these lower benefits states, the model suggests that if the policy goal is to promote work while limiting the utility consequences to the welfare eligible, a five-year time limit is a reasonable tool. In higher benefits states and among those with low human capital, however, the estimated utility consequences are relatively severe.

From the perspective of the period a_0 self, the preferred length of the time limit depends somewhat on education level and type. Among those with less education (cell 3), longer limits induce less work but generate more utility. Among those with more education (cells 5, 7, and 8) the longest limit is most preferred, but among the shorter limits, the shorter the better. Indeed, among these groups, eliminating welfare is preferred to a year-long time limit, and in particular, the cells 5 and 7 group may strictly prefer the elimination of the welfare system.

6.3. *Consequences of Workfare.* Table 8 presents the results of a parallel analysis with workfare policies. In these experiments, two dimensions of the policy are varied: (1) the degree to which workfare contributes to human capital and (2) the extent to which workfare compensates for lost home production.

Policy version 1 assumes that workfare is merely "make-work"—participation in the program adds nothing to human capital. In this version of the policy, home production is compensated by 50% while on workfare through, for example, a child care subsidy. (With each policy the stigma of welfare participation is assumed to apply.) **Policy version 2** assumes workfare approximates market work—participation in workfare contributes to work experience just as labor market work would.[56] Again home production is compensated by 50%. Finally, **policy version 3** replicates the human capital structure of policy version 2, but increases home production compensation to 75%.

[56] The decay of human capital still occurs when an individual leaves market work for workfare.

TABLE 8

SIMULATED EFFECTS OF WORKFARE POLICIES FOR SOPHISTICATED PRESENT-BIASED AGENTS, BY INITIAL CONDITIONS

Policy		Initial Conditions Cell 1	2	3	4	5	6	7	8
Workfare Version 1	% change in lifetime util.	−4.88	−14.74	−2.89	−12.90	−0.38	−11.26	−0.16	−3.15
	changes in % working	22.51	13.55	20.49	19.33	12.04	22.22	10.58	23.58
Workfare Version 2	% change in lifetime util.	1.78	−2.33	2.68	−1.75	2.80	−0.77	1.96	2.58
	changes in % working	21.61	20.32	18.93	23.85	8.15	26.35	7.29	18.93
Workfare Version 3	% change in lifetime util.	6.58	6.15	6.36	6.09	5.00	6.36	3.47	6.33
	changes in % working	13.85	13.00	12.88	16.11	2.39	18.84	2.62	9.46

NOTE: Each row presents the simulated effects, in terms of behavior (change in % of time working) and discounted utility (in 1984 dollars) of introducing the workfare policy given in the first cell of the row. In version 1, the policy is make-work: Welfare eligibility requires full-time employment and this work does not contribute to human capital. However, 50% of the value of home production lost from workfare is paid for by, for example, a child-care subsidy. In version 2, the policy is the same except that required employment contributes to human capital just like standard market employment. In version 3, the policy is the same as that in version 2 except that 75% of the value of home production lost from workfare is paid for by child-care subsidies. These simulations are provided for eight representative cells of initial conditions (the columns). See Panel 2 of Table 6 for initial conditions for the different cells.

For the first version of the workfare policy, in which the work requirement adds nothing to human capital while reducing home production by half, the model predicts substantial increases in market work. Among those with less schooling, the increases in time spent in the labor market are somewhat larger for those in low benefit states (see cells 1 and 3). Among those with more schooling, the opposite holds: Make-work policies lead to the largest increases in market work for those living in higher benefit states (see cells 6 and 8). Regardless of education or welfare benefits level, however, this first workfare policy reduces expected lifetime utility, though for those with greater human capital living in low benefits states, the declines are quite modest.

The predicted effects of workfare can be qualitatively different, however, when the work required adds to human capital (policy versions 2 and 3). When workfare provides the opportunity to accumulate human capital, there are two countervailing effects on decision making. On one hand, the access to human capital at a guaranteed "wage" makes welfare a relatively attractive choice. On the other hand, the accumulation of human capital while receiving welfare will make a transition into market work more appealing. The simulations indicate that the dominating effects vary with the initial conditions of the agents. Relative to the "make-work" policy, the second version of the policy leads to greater increases in market work among welfare-eligibles with relatively low human capital in high benefit states (cells 2, 4, and 6). For those with higher human capital, and/or living in low benefits states (cells 1, 3, 5, 7, and 8) the employment gains are smaller with this second policy.

When home production is compensated by half (policy version 2), the utility effects of the policy experiment are somewhat mixed. Among those with more human capital living in low benefits states, the commitment effect of the policy combined with the ability to accumulate human capital while on welfare leads to modest increases in expected lifetime utility. But those with relatively low human capital living in high welfare benefit states (cells 2, 4, and 6), lifetime expected utility declines, though the declines are quite modest. When home production is compensated by 75% (policy version 3), the model predicts, more uniformly, lifetime utility gains. These gains are arguably modest, but mostly derive from increases in employment of a size comparable to those derived from make-work workfare. Thus, these simulations indicate that sizable increases in employment among the welfare eligible can be achieved at relatively low utility cost (or indeed with utility gains) from workfare that both generates marketable human capital and substantially compensates for lost home production.[57]

[57] The gains would be more substantial if, as is plausible, the policy also reduced the stigma of welfare participation.

7. CONCLUSION

Estimates of the structural parameters of a dynamic model of labor supply indicate that the work–welfare–home decisions of never-married women with children reveal time-inconsistent preferences. For this group, we estimate a present-bias factor (β) less than unity and we reject a model of standard discounting at standard levels of confidence.

Simulations of the estimated model indicate that, for this group of largely low-income, single women with children, the behavioral consequences of an inability to commit to future decisions may be substantial, but, by one measure, the utility consequences of the self-control problem are modest. The model suggests that the ability to commit to future decisions would often lead to considerably more work and less welfare participation. However, for those with low levels of human capital and living in high welfare benefits states, procrastination leads to costly delays in welfare takeup. For this group, commitment ability leads to slightly more, not less, welfare participation. Moreover, among those entering the labor force earlier, this entry involves costs in terms of welfare benefits and home production forgone, and the benefits in terms of higher wages are accrued only in the relatively distant future. As a result, the discounted lifetime utility gains from commitment may be small even when the behavioral consequences are large.

Further simulations of the model indicate that behavioral and utility consequences of welfare reform policies that serve as imperfect commitment devices vary according to both the characteristics of the intended targets and the design of the policy. We find that time limits are too crude to enhance expected utility. Although limits serve to substantially increase employment, they do so at a sometimes substantial utility cost for the welfare-eligible. For those living in low benefits states and for those with higher levels of skills and education, however, the utility losses from a five-year time limit are quite modest. The estimated model indicates that workfare policies also better serve those with more education living in states with lower welfare benefits. However, when workfare leads to the accumulation of valuable human capital and includes compensation for lost home production through, for example, child-care subsidies, the estimated model suggests that most potential recipients will increase both their employment and their lifetime utility.

We interpret these results as qualified support for the extension of standard models of dynamic labor supply to allow for time-inconsistency. Our analysis focuses on a special group (never-married women with children) whose preferences may not be typical of the general population, though may be quite representative for the potential welfare population. With respect to this group, however, our analysis indicates that allowing for time-inconsistency may be both feasible and fruitful, adding to our understanding of the potential consequences of policy. We also view our findings as a caution against simple arguments for accounting for the role of psychological biases in public policy. As our simulations indicate, even when individuals display substantial present-bias in preferences, simple policies that resemble commitment devices may not function effectively as such.

APPENDIX

A. *Additional Tables and Estimates*

A.1. *Descriptive statistics of the selected women and all women.* Table A.1 compares the statistics of our selected subsample (never-married women with at least one child) with those of the entire sample of women in the NLSY. It shows that the subsample has on average more children at every age. By age 32, the gap is relatively small with the subsample having on average 2.1 children and the entire sample 1.6. At every age the subsample has an average of 1.25 fewer years of work experience and 2.01 more years of AFDC receipt, and at every age older than 19, full-time workers in the subsample earn on average $1,456 less than their counterparts in the entire sample. On average, the subsample has also completed fewer years of schooling (10.9) than the entire sample (12.6).

TABLE A.1
DESCRIPTIVE STATISTICS FOR ALL WOMEN AND SELECTED SAMPLE (NEVER-MARRIED WOMEN WITH AT LEAST ONE CHILD): AGES 16–32

	Number of Children		Yrs. of Work Experience		Yrs. of Schooling*		Earnings for Workers**		Yrs. Received AFDC	
Age	All Women	Our Sample	All Women	Our Sample	All Women	Our Sample	All Women	Our Sample	All Women	Our Sample
16	0.06 (0.01)	1.23 (0.08)	0.00 (0.00)	0.00 (0.00)	9.38 (0.02)	8.81 (0.17)	5410.34 (617.97)	n.a.	0.00 (0.00)	0.13 (0.06)
17	0.10 (0.01)	1.24 (0.06)	0.03 (0.00)	0.00 (0.00)	10.28 (0.02)	9.37 (0.14)	5808.70 (166.49)	n.a.	0.01 (0.00)	0.24 (0.06)
18	0.17 (0.01)	1.28 (0.04)	0.10 (0.01)	0.00 (0.00)	11.14 (0.02)	10.04 (0.12)	7001.72 (167.58)	10822.56 (2254.07)	0.02 (0.00)	0.34 (0.06)
19	0.25 (0.01)	1.36 (0.04)	0.26 (0.01)	0.05 (0.01)	11.75 (0.02)	10.41 (0.10)	7723.65 (122.65)	6715.04 (766.20)	0.04 (0.00)	0.49 (0.06)
20	0.33 (0.01)	1.43 (0.04)	0.54 (0.01)	0.14 (0.03)	12.19 (0.02)	10.63 (0.09)	8301.49 (102.97)	7361.80 (623.75)	0.07 (0.01)	0.78 (0.07)
21	0.44 (0.01)	1.49 (0.04)	0.87 (0.02)	0.25 (0.03)	12.48 (0.02)	10.77 (0.09)	8819.52 (111.21)	7040.99 (594.02)	0.10 (0.01)	1.09 (0.08)
22	0.54 (0.01)	1.59 (0.05)	1.26 (0.02)	0.43 (0.05)	12.71 (0.03)	10.86 (0.08)	9676.16 (110.01)	8097.39 (505.02)	0.15 (0.01)	1.45 (0.09)
23	0.66 (0.01)	1.70 (0.05)	1.72 (0.03)	0.69 (0.07)	12.87 (0.03)	10.92 (0.08)	10405.64 (107.29)	8929.34 (374.76)	0.21 (0.01)	1.86 (0.10)
24	0.77 (0.01)	1.79 (0.05)	2.24 (0.03)	0.95 (0.08)	12.95 (0.03)	10.96 (0.09)	11086.97 (113.59)	9376.78 (394.42)	0.27 (0.01)	2.30 (0.12)
25	0.89 (0.02)	1.84 (0.05)	2.79 (0.03)	1.34 (0.10)	13.01 (0.03)	11.02 (0.09)	11719.11 (129.55)	9787.95 (411.93)	0.32 (0.02)	2.67 (0.13)
26	1.01 (0.02)	1.90 (0.05)	3.34 (0.04)	1.68 (0.12)	13.07 (0.03)	11.06 (0.09)	12280.13 (136.35)	10099.91 (381.09)	0.38 (0.02)	3.02 (0.14)
27	1.12 (0.02)	1.94 (0.05)	3.85 (0.04)	2.05 (0.14)	13.14 (0.03)	11.10 (0.09)	12674.95 (165.59)	10392.56 (393.77)	0.43 (0.02)	3.38 (0.16)
28	1.23 (0.02)	1.95 (0.06)	4.39 (0.05)	2.44 (0.17)	13.18 (0.04)	11.22 (0.10)	13379.60 (220.25)	10692.78 (448.43)	0.47 (0.02)	3.61 (0.19)
29	1.33 (0.02)	1.99 (0.07)	4.91 (0.06)	2.66 (0.21)	13.20 (0.04)	11.37 (0.11)	13651.76 (278.03)	11004.52 (497.69)	0.50 (0.03)	3.89 (0.23)
30	1.41 (0.02)	2.08 (0.08)	5.48 (0.08)	2.89 (0.26)	13.24 (0.04)	11.44 (0.13)	13531.17 (293.62)	11360.18 (632.93)	0.52 (0.03)	4.48 (0.29)
31	1.54 (0.03)	2.11 (0.10)	5.90 (0.10)	3.22 (0.32)	13.22 (0.05)	11.56 (0.15)	13614.89 (364.01)	13455.48 (1091.30)	0.55 (0.04)	4.67 (0.33)
32	1.63 (0.03)	2.08 (0.13)	6.40 (0.12)	3.99 (0.45)	13.24 (0.05)	11.70 (0.21)	14301.09 (827.82)	12091.61 (871.85)	0.58 (0.04)	4.49 (0.43)

NOTES: Standard errors in parentheses. Means are calculated using the NLSY's 1979 sample weights. Members of the poor white and military oversamples are excluded.
*Years of schooling at the birth of the first child.
**Earnings are full-time equivalent in 1987 dollars.

TABLE A.2
ESTIMATED ANNUAL WELFARE BENEFITS FUNCTION, SUMMARY STATISTICS (1987 DOLLARS)

States*	$\hat{\theta}_{j0}$	$\hat{\theta}_{j1}$	Annual Benefit for 1 Child	Annual Benefit for 2 Children	Percent on Welfare**
1	2380.45	1238.01	3618.46	4856.48	39.6
2	2467.68	1301.31	3768.99	5070.30	50.0
3	2962.66	1203.84	4166.50	5370.34	32.9
4	2979.62	1280.44	4260.06	5540.50	22.5
5	3128.33	1340.02	4468.35	5808.38	39.2
6	3493.63	1186.81	4680.45	5867.26	29.6
7	3541.08	1251.03	4792.11	6043.13	50.3
8	3985.20	1212.98	5198.18	6411.15	46.6
9	4348.62	1098.98	5447.60	6546.58	28.2
10	4358.47	1318.76	5677.23	6995.99	71.0
11	4279.58	1419.96	5699.54	7119.50	51.2
12	4509.59	1368.62	5878.21	7246.83	29.4
13	4183.05	1539.27	5722.32	7261.59	13.6
14	4592.94	1343.95	5936.89	7280.83	20.2
15	4511.30	1411.63	5922.93	7334.57	66.8
16	5005.98	1480.68	6486.65	7967.33	52.5
17	4988.00	1577.07	6565.07	8142.15	27.0
18	5634.63	1661.86	7296.49	8958.35	61.7
19	5317.42	1851.81	7169.23	9021.04	69.7
20	6264.03	1613.01	7877.04	9490.05	68.5
Mean	4146.61	1385.00	5531.62	6916.62	43.5
Std. Dev.	1042.30	187.46	1182.51	1334.38	17.9

*To preserve the anonyminity of respondents, we were not provided with state names.
**Percent of sample living in the corresponding state that choose welfare.

TABLE A.3
LOGIT ESTIMATES OF THE FERTILITY FUNCTION

Parameter		Estimate	Std. Error
Constant	γ_0	−0.811	0.323
Age	γ_1	−0.044	0.015
Number of existing children	γ_2	−0.077	0.059
Is she on welfare?	γ_3	0.094	0.115
Is she working?	γ_4	−0.494	0.162
Observations:		3911	
Likelihood ratio:		38.20	
Log likelihood:		−1287.16	
Pseudo R^2		0.014	

A.2. *Estimates of welfare benefit function G_j.* Table A.2 presents the estimates of welfare benefit functions for the 20 U.S. states used in our estimation.

A.3. *Fertility function ρ.* Table A.3 presents the estimates of the fertility function ρ.

A.4. *Within-sample goodness-of-fit test.* Table A.4 presents the χ^2 goodness-of-fit test of the within-sample choice distributions by age. The column labeled by "Row" is the χ^2 statistic for the overall choice distribution for the particular age in that row.

B. *Simulation Results for Completely Naive Present-Biased Agents.* We present the simulation results for naive present-biased agents in this section (Tables B.1–B.4).

TABLE A.4

χ^2 GOODNESS-OF-FIT TESTS OF THE WITHIN-SAMPLE CHOICE DISTRIBUTION BY AGE, MODEL WITH SOPHISTICATED AGENTS

	Choice			
Age	Welfare	Work	Home	Row
18	6.09*	†	0.19	6.28*
19	4.56*	0.79	2.05	7.40*
20	7.82*	3.76	2.18	13.76*
21	2.63	3.78	0.16	6.56*
22	3.98*	4.19*	0.50	8.68*
23	1.28	1.94	0.03	3.25
24	0.00	0.19	0.14	0.32
25	0.61	0.00	1.32	1.93
26	0.23	1.46	0.38	2.07
27	1.39	1.38	0.17	2.93
28	0.25	0.00	0.72	0.97
29	0.64	1.06	0.02	1.72
30	0.32	0.36	0.00	0.68
31	0.19	0.00	0.66	0.85
32	0.77	0.12	4.63*	5.52*

*Significant at the 5% level.
†Fewer than five observations.

TABLE B.1

SIMULATED YEAR-TO-YEAR TRANSITION PROBABILITY MATRIX FOR NAIVE PRESENT-BIASED AGENTS

	Choice at t		
Choice at $t-1$	Welfare	Work	Home
Welfare			
Row %	84.4	4.2	11.4
Column %	78.5	7.4	19.7
Work			
Row %	10.7	74.5	14.8
Column %	5.6	72.8	14.3
Home			
Row %	25.7	17.0	57.2
Column %	16.0	19.8	66.0

TABLE B.2

SIMULATED EFFECTS OF THE ABILITY TO COMMIT FOR NAIVE PRESENT-BIASED AGENTS, BY INITIAL CONDITIONS

	Initial Conditions Cell							
	1	2	3	4	5	6	7	8
Changes in % working	9.41	−2.79	17.95	−3.07	22.58	−2.16	23.80	11.30
% Change in lifetime utility	2.82	2.39	4.23	2.43	4.95	2.47	5.00	3.17

NOTE: This table presents the simulated effects, in terms of behavior (% of time working) and discounted utility (in 1984 dollars), of providing naive agents with perfect commitment ability. These simulations are provided for eight representative cells of initial conditions. See Panel 2 of Table 6 for initial conditions for the different cells.

TABLE B.3

SIMULATED EFFECTS OF TIME LIMITS OF VARYING LENGTHS FOR NAIVE PRESENT-BIASED AGENTS, BY INITIAL CONDITIONS

		Initial Conditions Cell							
Time Limits		1	2	3	4	5	6	7	8
7 years	% change in lifetime util.	−1.82	−9.25	−0.99	−7.62	−0.20	−6.24	−0.16	−1.69
	changes in % working	11.31	22.76	7.39	20.77	2.13	17.77	1.55	9.27
5 years	% change in lifetime util.	−2.63	−12.70	−1.37	−10.48	−0.21	−8.78	−0.21	−2.43
	changes in % working	17.60	33.45	12.51	31.82	4.60	28.69	3.25	16.29
3 years	% change in lifetime util.	−3.87	−17.17	−2.15	−14.10	−0.38	−11.89	−0.32	−3.36
	changes in % working	25.48	43.49	19.18	42.47	8.22	40.24	6.38	24.42
1 year	% change in lifetime util.	−6.25	−23.23	−3.62	−19.32	−0.77	−16.08	−0.47	−4.56
	changes in % working	34.03	54.34	28.20	54.78	14.13	52.74	11.69	35.06
0 year	% change in lifetime util.	−6.51	−24.60	−3.65	−20.30	−0.24	−16.80	0.00	−4.08
	changes in % working	39.27	60.10	33.14	60.67	18.83	59.13	15.43	40.89

NOTE: Each row presents the simulated effects, in terms of behavior (change in % of time working) and discounted utility (in 1984 dollars), of introducing a time limit of the length given in the first cell of the row. These simulations are provided for eight representative cells of initial conditions (the columns). See Panel 2 of Table 6 for initial conditions for the different cells.

TABLE B.4

SIMULATED EFFECTS OF WORKFARE POLICIES FOR NAIVE PRESENT-BIASED AGENTS, BY INITIAL CONDITIONS

		Initial Conditions Cell							
Policy		1	2	3	4	5	6	7	8
Workfare	% change in lifetime util.	−5.36	−14.64	−3.25	−12.87	−0.70	−11.39	−0.27	−3.49
Version 1	changes in % working	22.24	13.39	19.78	19.54	11.51	21.79	10.35	23.19
Workfare	% change in lifetime util.	0.76	−3.50	1.96	−2.69	2.40	−1.73	1.85	2.02
Version 2	changes in % working	20.61	19.78	17.78	23.93	7.79	25.63	7.29	18.27
Workfare	% change in lifetime util.	5.49	4.77	5.57	4.91	4.59	5.29	3.32	5.74
Version 3	changes in % working	13.02	12.98	12.06	16.30	2.21	18.52	2.56	9.21

NOTE: Each row presents the simulated effects, in terms of behavior (change in % of time working) and discounted utility (in 1984 dollars), of introducing the workfare policy given in the first cell of the row. In version 1, the policy is make-work: Welfare eligibility requires full-time employment and this work does not contribute to human capital. However, 50% of the value of home production lost from workfare is paid for by, for example, a child-care subsidy. In version 2, the policy is the same except that required employment contributes to human capital just like standard market employment. In version 3, the policy is the same as that in version 2 except that 75% of the value of home production lost from workfare is paid for by child-care subsidies. These simulations are provided for eight representative cells of initial conditions (the columns). See Panel 2 of Table 6 for initial conditions for the different cells.

REFERENCES

AINSLIE, G., *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person* (Cambridge, UK: Cambridge University Press, 1992).

ANGELETOS, G. -M., D. LAIBSON, J. TOBACMAN, A. REPETTO, AND S. WEINBERG, "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation," *Journal of Economic Perspectives* 15 (August 2001), 47–68.

BARRO, R. J., "Ramsey Meets Laibson in the Neoclassical Growth Model," *Quarterly Journal of Economics* 114 (November 1999), 1125–52.

BERNDT, E. R., B. H. HALL, R. E. HALL, AND J. A. HAUSMAN, "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement* 3 (October 1974), 653–65.

BLOOM, D., M. FARRELL, AND B. FINK, WITH D. ADAMS-CIARDULLO, "Welfare Time Limits: State Policies, Implementation, and Effects on Families," Manpower Demonstration Research Corporation report submitted to the U.S. Department of Health and Human Services, 2002.

CARRILLO, J. D., AND T. MARIOTTI, "Strategic Ignorance as a Self-Disciplining Device," *Review of Economic Studies* 67 (July 2000), 529–44.

CROUSE, G. L., "Trends in AFDC and Food Stamp Benefits, 1972–1994," ASPE Research Notes, Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, 1995.

DELLA VIGNA, S., AND U. MALMENDIER, "Paying Not to go to the Gym," *American Economic Review* 96 (June 2006), 694–719.
——, AND M. D. PASERMAN, "Job Search and Impatience," *Journal of Labor Economics* 23 (July 2005), 527–88.
DEPARTMENT OF HEALTH AND HUMAN SERVICES, ACF OFFICE OF FAMILY ASSISTANCE, "Characteristics and Financial Circumstances of AFDC Recipients," Washington, D.C., 1996.
ECKSTEIN, Z., AND K. I. WOLPIN, "Why Youth Drop out of High School: The Impact of Preferences, Opportunities and Abilities," *Econometrica* 67 (November 1999), 1295–339.
EDIN, K., AND L. LEIN, *Making Ends Meet: How Single Mothers Survive Welfare and Low-wage Work* (New York: Russell Sage, 1997).
FANG, H., AND D. SILVERMAN, "On the Compassion of Time-limited Welfare Programs," *Journal of Public Economics* 88 (July 2004), 1445–70.
——, AND ——, "Distinguishing between Cognitive Biases: Belief vs. Time Discounting in Welfare Program Participation," in Joel Slemrod and Edward McCaffery eds., *Behavioral Public Finance: An Agenda* (New York: Russell Sage Foundation, 2006), 47–81.
GRUBER, J., AND B. KOSZEGI, "Is Addiction 'Rational'? Theory and Evidence," *Quarterly Journal of Economics* 116 (November 2001), 1261–303.
HALEVY, Y., "Strotz Meets Allais: Diminishing Impatience and the Certainty Effect," *American Economic Review* 98 (June 2008), 1145–620.
HARRIS, C., AND D. LAIBSON, "Dynamic Choices of Hyperbolic Consumers," *Econometrica* 69 (July 2001), 935–57.
HAUSMAN, J. A., "Individual Discount Rates and the Purchase and Utilization of Energy-using Durables," *Bell Journal of Economics* 10 (Spring 1979), 33–54.
HOTZ, J., AND R. A. MILLER, "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies* 60 (July 1993), 397–421.
KEANE, M. P., AND K. I. WOLPIN, "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *The Review of Economics and Statistics* 76 (November 1994), 648–72.
——, AND ——, "The Career Decisions of Young Men," *Journal of Political Economy* 105 (June 1997), 473–522.
——, AND ——, "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment," *International Economic Review* 42 (November 2001), 1051–103.
——, AND ——, "The Role of Labor and Marriage Markets, Preference Heterogeneity and the Welfare System in the Life Cycle Decisions of Black, Hispanic and White Women," manuscript, University of Pennsylvania, 2005.
KRUSELL, P., B. KURUŞÇU, AND A. SMITH, JR., "Equilibrium Welfare and Government Policy with Quasi-Geometric Discounting," *Journal of Economic Theory* 105 (July 2002), 42–72.
LAIBSON, D., "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics* 112 (May 1997), 443–77.
——, A. REPETTO, AND J. TOBACMAN, "Self-Control and Saving for Retirement," *Brookings Papers on Economic Activity* 1 (1998), 91–196.
——, ——, AND ——, "Estimating Discount Functions from Lifecycle Consumption Choices," manuscript, Department of Economics, Harvard University, 2007.
LAWRANCE, E. C., "Poverty and the Rate of Time Preference: Evidence from Panel Data," *Journal of Political Economy* 99 (February 1991), 54–77.
LOEB, S., AND M. CORCORAN, "Welfare, Work Experience, and Economic Self-Sufficiency," *Journal of Policy Analysis and Management* 20 (February 2001), 1–20.
LOEWENSTEIN, G., AND J. ELSTER, *Choice over Time* (New York: Russell Sage, 1992).
MILLER, R. A., AND S. G. SANDERS, "Human Capital Development and Welfare Participation," *Carnegie-Rochester Conference Series on Public Policy* 46 (June 1997), 1–43.
MOFFITT, R., "An Economic Model of Welfare Stigma," *American Economic Review* 73 (December 1983), 1023–35.
O'DONOGHUE, T., AND M. RABIN, "Doing It Now or Later," *American Economic Review* 89 (March 1999a), 103–24.
——, AND ——, "Addiction and Self-Control," in J. Elster ed., *Addiction: Entries and Exits* (New York: Russell Sage, 1999b).
——, AND ——, "Choice and Procrastination," *Quarterly Journal of Economics* 116 (February 2001), 121–60.
PASERMAN, M. D., "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation," *Economic Journal* 118 (August 2008), 1418–52.
PHELPS, E. S., AND R. A. POLLAK, "On Second-best National Saving and Game-equilibrium Growth," *Review of Economic Studies* 35 (April 1968), 185–99.
POLLAK, R. A., "Consistent Planning," *Review of Economic Studies* 35 (April 1968), 201–08.

RUBINSTEIN, A., "'Economics and Psychology'? The Case of Hyperbolic Discounting," *International Economic Review* 44 (November 2003), 1207–16.

RUST, J., AND C. PHELAN, "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica* 65 (July 1997), 781–831.

STROTZ, R. H., "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies* 23 (1956), 165–80.

SWANN, C., "Welfare Reforms When Agents Are Forward-looking," *Journal of Human Resources* 40 (Winter 2005), 31–56.

VAN DER KLAAUW, W., "Female Labour Supply and Marital Status Decisions: A Life-Cycle Model," *Review of Economic Studies* 63 (April 1996), 199–235.

WARNER, J. T., AND S. PLEETER, "The Personal Discount Rate: Evidence from Military Downsizing Programs," *American Economic Review* 91 (March 2001), 33–53.