Exploring the Usefulness of a Non-Random Holdout Sample for Model Validation: Welfare Effects on Female Behavior

Michael P. Keane

Yale University

and

Kenneth I. Wolpin

University of Pennsylvania

May, 2005

Revised, April 2006

The authors are grateful for support from NICHD under grant HD-34019 and from several grants from the Minnesota Supercomputer Institute. Part of Keane's work on this project was completed while he was visiting Arizona State University as the Goldwater Chair in American Institutions. We thank two anonymous referees and a guest editor for helpful comments.

I. Introduction

An important goal of empirical research in economics is to provide evidence on the validity of decision-theoretic models that describe the behavior of economic agents. There are two approaches to this endeavor that stem from different epistemological perspectives. The first stems from a view that knowledge is absolute, that is, there exists a “true”decision-theoretic model from which observed data are generated. This leads naturally to a model validation strategy based on testing the validity of the behavioral implications of the model and/or testing the fit of the model to the data. A model is not deemed invalid if it is not rejected in such tests, according to some statistical criterion. Rejected models are deemed invalid and discarded.

The second approach stems from a pragmatic epistemological view, in which it is assumed that all models are necessarily simplifications of agents’ actual decision-making behavior. Hypothesis testing as a means of model validation or selection is eschewed because, given enough data, all models would be rejected as true models. In this pragmatic view, there is no true decision-theoretic model, only models that perform well or poorly, or better or worse, in addressing particular questions. Models should be chosen that are “best” for some specific purpose, and alternative models may co-exist or be valid for different purposes.

Decision-theoretic models are typically designed and estimated with the goal of predicting the impact on economic agents of changes in the economic environment. Thus, one criterion for model validation/selection that fits within the “pragmatic” view is to examine a model’s predictive accuracy, namely, how successful the model is at predicting outcomes of interest within the particular context for which the model was designed. In contrast, in the absolutist view, a model would be considered useful for prediction only if it were not rejected, despite the fact that non-rejection does not necessarily imply predicted effects will be close to actual effects. Nor will non-rejected models necessarily outperform rejected models in terms of their (context-specific) predictive accuracy.

A particularly challenging use of decision-theoretic models in economics is to forecast the impact of large changes in the environment. Often, these changes are related to policy interventions that are outside of the scope of current policies, such as the recent 1996 welfare reform or the new expansion of Medicare to cover prescription drugs. A “good” in-sample fit of a model is unlikely, in itself, to give us much confidence in its forecasting ability in such contexts. This problem arises in large part because of the common practice of using the same data both for estimation and for model development.[1] The problem we explore in this paper is how to gain confidence in a model's ability to predict the impact of such large changes.

Below, we show that an approach to validation and model selection that includes the choice of a "non-random holdout sample", a sample that differs significantly from the estimation sample along the policy dimension that the model is meant to forecast, can be fruitful. The idea is that if a model can provide a good forecast for a holdout sample that faces a policy regime well outside the support of the data (and that is not used in model formulation), then we gain confidence that it can provide a good forecast of impacts of other policy changes along that same dimension.[2] We explore the usefulness of this external validation approach, as a component of an overall validation strategy, using the model of female life-cycle behavior in Keane and Wolpin (2006). We choose a holdout sample that faced very different welfare policy rules from the estimation sample.[3]

---

[1] This practice is ubiquitous in both structural and non-structural estimation.

[2] It is useful to contrast the non-random holdout sample with cross-validation (CV), which relies on random holdout samples. A typical cross validation procedure is to split the data into K mutually exclusive, randomly chosen subsets of (approximately) equal size, estimate the model on each possible group of K-1 subsets, and assess the model's predictive accuracy based on each left out set. By design, cross validation methods are based on predicted outcomes that, although out of sample, are always within the support of the current regime. Thus, we would argue they do not, by themselves, create confidence in a model's ability to predict effects of large regime changes.

[3] The deliberate choice of a non-random holdout sample for the purpose of model validation is commonplace in time-series analysis. We are, however, unaware of its use in cross-section or panel data settings in economics. The idea seems to be a rather old one in psychology. Mosier (1951) suggested the procedure, naming it "validity generalization," that is, validation by generalizing beyond the sample. Recently, Busemeyer and Wang (2000) have argued for its more widespread adoption in psychology and provide Monte Carlo evidence on its performance in model selection. The use of models to forecast out-of-sample behavior is also common in the marketing literature, where considerable effort has been devoted to forecasting demand for new products. Few of the papers in that literature, however, compare predictions to subsequent demand after the product is introduced.

The most convincing examples in economics of external validations of decision-theoretic models have been based on randomized social experiments or on large regime shifts (that can be treated as experiments for the purpose of model validation), but such opportunities are rare. Among the earliest examples in which such a large regime shift is exploited is work by McFadden (1977) on forecasting the demand for rail rapid transport in the San Francisco Bay area. McFadden estimated a random utility model (RUM) of travel demand before the introduction of the Bay Area Rapid Transit (BART) system, obtained a forecast of the level of patronage that would ensue, and then compared the forecast to actual usage after BART's introduction.[4] Since that work, there have been, to our knowledge, only a handful of papers in the economics literature that have pursued a similar method of model validation.

McFadden's model validation treats pre-BART observations as the estimation sample and post-BART observations as the validation sample. A similar opportunity was exploited by Lumsdaine, Stock, and Wise (1992). They estimated a model of retirement behavior of workers in a single firm who were observed before and after the introduction of a temporary one-year pension window. They estimated several models on data before the window was introduced and compared the forecast of the impact of the pension window on retirement based on each estimated model to the actual impact as a means of model validation and selection. Keane and Moffitt (1998) estimated a model of labor supply and welfare program participation using data after federal legislation (OBRA 1981) that significantly changed the program rules. They used the model to predict behavior prior to that policy change. Keane (1995) used the same model to predict the impact of planned expansions of the Earned Income Tax Credit in 1994-1996.

Randomized social experiments have also provided opportunities for model validation and selection. Wise (1985) exploited a housing subsidy experiment to evaluate a model of housing demand. In the experiment, families that met an income eligibility criterion were randomly assigned to control and treatment groups. The latter were offered a rent subsidy. The

---

[4] A regime shift, as opposed to a randomized experiment, is generally characterized by a time lapse between observations on the estimation sample (the control group) and those on the validation sample (the treatment group). Over that period, changes may have occurred that would affect behavior in ways not captured in the estimation. In addition, whatever assumption is made about the exogeneity of a regime shift becomes part of the validation exercise.

model was estimated using only control group data and was used to forecast the impact of the program on the treatment group. The forecast was compared to its actual impact. More recently, Todd and Wolpin (2002) used data from a large-scale school subsidy experiment in Mexico, where villages were randomly assigned to control and treatment groups. Using only the control villages, they estimated a behavioral model of parental decisions about child schooling and work, as well as family fertility. The validity of the model was then assessed according to how well it could forecast (predict) the behavior of households in the treatment villages.[5] Similarly, Lise, Seitz and Smith (2003) used data from a Canadian experiment designed to move people off of welfare and into work to validate a calibrated search-matching model of labor market behavior.[6]

All of these papers made use of what was, from the researcher's perspective, a fortuitous event. The common and essential element is the existence of some form of a regime change radical enough to provide a degree of distance between the estimation and validation samples. The more different the regimes, the less likely are forecasted and actual behavior in the validation sample to be close purely by chance. However, waiting for such regime shifts to arise, given their rarity, does not lead to a viable research approach to model validation and selection. Our alternative approach attempts to mimic the essential element of regime change by non-randomly holding out from estimation a portion of the sample that faces a significantly different policy regime. This "non-random holdout sample" is then used for model validation/selection.

We illustrate the non-random holdout sample approach to model validation in the context of a model of welfare program participation. The policy heterogeneity that we exploit to generate the hold-out sample takes advantage of the wide variation across U.S. States that has existed in welfare policy. Specifically, we formulate and estimate a dynamic programming (DP) model of

---

[5] When the model provides sufficient structure, and assuming that the model is deemed "valid", it is possible to simulate the impact of regime shifts other than the one used for validation. For example, Wise (1985) and Todd and Wolpin (2002) contrasted the effect of the policies evaluated in the experiments to several alternative policies.

[6] The use of laboratory experiments to validate economic models has, of course, a long tradition. Bajari and Hortascu (forthcoming) provide a recent example of evaluating a structurally estimated auction model by comparing the estimated valuations to those randomly assigned in an experimental setting.

the joint schooling, welfare take-up, work, fertility and marriage decisions of women using data from one group of U.S. states (the estimation or "control" sample) and forecast these same decisions on another state (the validation or "treatment" sample) that differs dramatically in the generosity of its welfare program. Notably, our model extends the literature on welfare participation in several dimensions.[7] We augment the choice set to include schooling and fertility in addition to work, marriage and welfare participation and allow for a richer modeling framework within which those choices are made.

We implement the model using 15 years of information from the 1979 youth cohort of the National Longitudinal Surveys of Labor Market Experience (NLSY79), supplemented with state-level welfare benefit rules that we have collected for each state over a 24 year period prior to the new welfare reform. The model is estimated on five of the largest states represented in the NLSY79 (California, Michigan, New York, North Carolina and Ohio) and validated on data from Texas. In terms of generosity, California, Michigan and New York are high benefit states, North Carolina and Ohio are medium benefit states and Texas is a low benefit state.

Both to establish a performance metric and to provide potential alternatives to the DP model, we also estimate four multinomial logit (MNL) specifications, differentiated by whether they include state fixed-effects and by the complexity of the specification of welfare rules (i.e., whether the rules are characterized by one parameter or five parameters). These MNL models are consistent with a myopic random utility model or a flexible approximation to a DP model.

To highlight the results, we find that the non-random holdout sample approach provides useful information for model validation and selection. But, we also find that it can be quite misleading when relied on exclusively. Specifically, as expected, it is difficult to distinguish among the DP model and the four logits in terms of within-sample fit, applying a RMSE criterion to the five states used in estimation (the models have a similar number of parameters). On the other hand, using the same RMSE criterion, the holdout sample does differentiate among models. In particular, one of the two logit models without state fixed-effects clearly gives unreasonable forecasts for Texas. The other logit without state fixed-effects performs better for minorities than

[7] See Moffitt (1992) for a review of the early literature based on static models. Previous DP models of welfare participation include Sanders (1993) and Swann (1996).

the DP model but worse for white women. The two logits with state-fixed effects perform better than both the DP model and the logits without state effects.[8] Based on these external validation results alone, one might well favor the logit specifications with state fixed-effects, and conclude they would provide more reliable forecasts of other counterfactual experiments related to welfare generosity than would the DP model.

However, based on our earlier work (Keane and Wolpin (2001a)), it seemed implausible that the state fixed-effects logit specifications would provide the best forecast of Texas. In that paper, we note that, in models with state fixed-effects, estimates are based only on within-state over-time variation in welfare benefits. Thus, such models provide estimates of the impact of transitory changes in welfare rules. Yet, the external validation performed by forecasting behavior under Texas' rules requires prediction of behavior under a permanently different welfare regime, one with both a change in the "long-run" level of benefits and their transitory fluctuations over time. Why were fixed effects models successful in such an exercise? As noted earlier, in carrying out the external validation, we estimated the state fixed-effects for Texas, thus effectively fitting mean behaviors. To see if the estimation of Texas dummies was critical to the superior performance of the state fixed-effect specifications, we performed a further exercise, a counterfactual experiment in which we gave the five estimation states the same welfare rules as Texas, maintaining the previous estimates of the state fixed-effects for the five states.

Of course, our expectation was that welfare participation would decline; no reasonable behavioral model would predict otherwise. But, consistent with our concern, the logit models with state fixed-effects predicted a substantial increase. Moreover, the one logit model without state fixed-effects that performed about equally to the DP model in our previous validation also produced unexpected results, forecasting that the decline in welfare participation would be considerably less than the increase in employment, a result that also seems inconsistent with a reasonable behavioral model.[9] The DP model, which imposes restrictions on behavior that the

[8] The Texas state fixed-effects were estimated while all of the other parameters were fixed at the values obtained for the estimation sample.

[9] Consider a simpler random utility setting in which there are three mutually exclusive discrete alternatives: work, welfare and home (no welfare). Given a reduction in welfare benefits,

more flexible logit does not, produced plausible forecasts. Hence, it seems to be the preferable model for analyzing large policy interventions.

The next section describes the structure of the DP model. Section 3 describes the data, Section 4 the estimation method and section 5 the results. The final section concludes.

II. Model

In this section, we provide an outline of our model.[10] We consider a woman who makes joint decisions at each age "a" of her life about the following discrete alternatives: whether or not to attend school, s_a, work part-time, h_a^p, or full-time, h_a^f, in the labor market (if an offer is received), be married (if an offer is received), m_a, become pregnant (if she is of a fecund age), p_a, and receive government welfare (if she is eligible), g_a. Thus, a woman chooses from as many as 36 mutually exclusive alternatives at each age during her fecund life cycle stage, and 18 during her infecund stage.[11] The fecund stage is assumed to begin at age 14 and to end at age 45; the decision period extends to age 62. Decisions are made at discrete six month intervals, i.e., semi-annually. A woman who becomes pregnant at age a has a birth at age a+1, with n_{a+1} representing the discrete birth outcome.[12] Consumption, C_a, is determined uniquely by the alternative chosen.

The woman receives a utility flow at each age that depends on her current choices and consumption, and on state variables that reflect past choices, for example, the number of children already born, N_a, and their current ages (which affect child-rearing time costs), and the

if the increase in the proportion who work exceeds the decline in the proportion who are on welfare, then the proportion taking leisure must decline. However, it is not possible in this setting for a reduction in welfare benefits, which reduces only the value of the welfare alternative, to reduce the prevalence of either of the other two alternatives.

[10] A complete description of the model with exact functional forms is provided in our companion paper, Keane and Wolpin (2006).

[11] Being married and receiving welfare is not an option. Although the AFDC-Unemployed Parent (AFDC-UP) program provided benefits for a family with an unemployed father, it accounts for only a small proportion of total spending on AFDC.

[12] In keeping with the assumption that pregnancies are perfectly timed, we only consider pregnancies that result in a live birth, i.e., we ignore pregnancies that result in miscarriages or abortions. We assume a woman cannot become pregnant in two consecutive six month periods.

current level of completed schooling, S_a (which affects utility from attendance). We also allow preferences to evolve with age. The five choices (school, work, marriage, pregnancy, welfare) are subject to a vector of five age-varying serially independent preference shocks ϵ_a. And, we allow for permanent heterogeneity in tastes across individuals by birth cohort, race and U.S. State of residence, and in a permanent unobservable characteristic we refer to as a woman's "type."[13] Conditional on current choices, the utility of an individual at age a of type k can be written:

$$(1) \quad U_a^k = U_a(C_a,\ s_a,\ m_a,\ p_a,\ g_a,\ h_a^p,\ h_a^f;\ \epsilon_a,\ I(\text{type}=j),\ \Omega_a^u),$$

where Ω_a^u represents the subset of the observed state variables that affects utility.[14, 15]

The budget constraint, assumed to be satisfied each period, is given by:

$$(2) \quad \begin{aligned} C_a &= y_a^o(1-m_a)(1-z_a) + [y_a^o + y_a^m]m_a\tau_a^m + [y_a^o + y_a^z\tau_a^z]z_a \\ &\quad + \beta_1 g_a b_a - [\beta_3 I(S_a \geq 12) - \beta_4 I(S_a \geq 16)]s_a, \end{aligned}$$

where y_a^o is the woman's own earnings at age a, y_a^m is the spouse's earnings if the woman is married, τ_a^m is the share of household income the woman receives if she is married, y_a^z is her parents' income, a share, τ_a^z, of which she receives if she co-resides with her parents, b_a is the amount of welfare benefits the woman is eligible to receive. β_1 is a fraction that converts welfare dollars into a monetary equivalent consumption value.[16] β_3 is the tuition cost of college and β_4

---

[13] In the model, we assume women do not change their State of residence, and we restrict our estimation to a sample with that characteristic.

[14] $I(\cdot)$ is the indicator function equal to one when the argument is true and zero otherwise.

[15] Monetary costs, when unmeasured, are not generally distinguishable from psychic costs. It is thus somewhat arbitrary what to include in the utility function or the the budget constraint. For example, we include in (1): (i) a fixed cost of working; (ii) a time cost of rearing children that varies by their ages; (iii) a time cost of collecting welfare (waiting at the welfare office); (iv) a school re-entry cost; and (v) costs of switching welfare and employment states.

[16] β_1 reflects the fact that welfare recipients are restricted in what they may purchase with welfare benefits, e.g., food stamps cannot be used to purchase alcohol.

the cost of graduate school. S_a is the completed level of schooling at age a, and $I(\cdot)$ is an indicator function equal to unity when the argument in the parentheses is true. By assumption, income is pooled when married, but not when co-residing with parents.

Living with parents and being married are taken to be mutually exclusive states. In particular, if a woman receives a marriage offer (see below) and chooses to be married she cannot live with parents. A single woman lives with her parents according to a draw from an exogenous probability rule, π_a^z. We assume that the probability of co-residing with parents, given the woman is unmarried, depends on her age and lagged co-residence status. The woman's share of her parents' income, when co-resident, depends on her age, her parents' schooling and whether she is attending post-secondary school (the latter captures parental transfers to help finance college).

In each period a woman receives a part-time job offer with probability π^{wp} and a full-time job offer with probability π^{wf}. Each of these offer rates depends on the woman's previous-period work status. If an offer is received and accepted, the woman's earnings is the product of the offered hourly wage rate and the number of hours she works, $y_a^o = 500 \cdot w_a^p h_a^p + 1000 \cdot w_a^p h_a^f$. The hourly wage rate is the product of the woman's human capital stock, Ψ_a, and its per unit rental price, which is allowed to differ between part- and full-time jobs, r^j for j=p, f. Specifically, her ln hourly wage offer is

$$\text{(3)} \quad \ln w_a^j = r^j + \Psi(\cdot) + \epsilon_a^w, \quad j = p, f.$$

The woman's human capital stock is modeled as a function of completed schooling, accumulated work hours up to age a, H_a, whether or not the woman worked part- or full-time in the previous period, her current age and her skill endowment at age 14. As with permanent preference heterogeneity, the skill endowment differs by race, State of residence and unobserved type. Random shocks to a woman's human capital stock, $\epsilon_a{}^w$, are assumed to be serially independent.

There is stochastic assortative mating. In each period a single woman draws an offer to marry with probability π_a^m, that depends on her age and welfare status. If the woman is currently married, with some probability that depends on her age and duration of marriage, she receives an offer to continue the marriage. If she declines to continue, the woman must be single for one period (six months) before receiving a new marriage offer.

The husband's earnings depends on his human capital stock, Ψ^m_a. If a woman receives a marriage offer, the potential husband's human capital is drawn from a distribution that depends on a subset of the woman's characteristics: her schooling, age, race/ethnicity, State of residence and unobserved (to us) type. In addition, there is an iid random component to the husband human capital draw that reflects a permanent characteristic unknown to the woman prior to meeting, μ^m. Thus, a woman can profitably search for husbands with more human capital, and can also directly affect the quality of her potential husbands by her choice of her schooling. There is a fixed utility cost of getting married, which augments a woman's incentive to wait for a good husband draw before choosing marriage (we allow for a cohort effect in this fixed cost). After marriage, the woman receives a utility flow from marriage, and a share of husband income. His earnings evolve with a fixed trend subject to a serially independent random shock, ϵ^m_a. Specifically,

$$(4) \quad \ln y^m_a = \mu^m + \Psi^m_{0a}(\cdot) + \epsilon^m_a$$

where Ψ^m_{0a} is the deterministic component of the husband's human capital stock.

Welfare eligibility and the benefit amount for a woman residing in State s at calendar time t depends on her number of minor children and her household income. For any given number of children under the age of 18 residing in the household, N^{18}_a, the schedule of benefits can be accurately approximated by two line segments. The first corresponds to the guarantee level; it is assumed to be (approximately) linearly increasing in the number of minor children and, in the case of a woman co-residing with her parents, linearly declining in parents' income, y^z_a. The second line segment is negatively sloped as a function of the woman's own earnings, y^o_a, plus parents' income if she is co-resident, and also linearly increasing in the number of minor children. The negative slopes reflect the benefit reduction (or tax) applied to income.

In general, benefits are equal to the guarantee level until the woman's earnings reach some positive amount, as the rules provide a child care/work expense allowance for working mothers. Denoting this (State-specific) "disregard" level of earnings as $y^{s1}_{at}(N^{18}_a)$, and the level of earnings at which benefits become zero (where the second line segment intersects the x-axis) as $y^{s2}_{at}(N^{18}_a)$, the benefit schedule for a woman with $N^{18}_a > 0$ children is given by:

$$
\begin{aligned}
(5)\qquad b_t^s(N_{at}^{18}, y_{at}^o, y_{at}^z) &= b_{0t}^s + b_{1t}^s N_{at}^{18} - b_{3t}^s \beta_2 y_{at}^z z_{at} && \text{for } y_{at}^o < y_{at}^{s1}(N_a^{18}), \\
&= b_{2t}^s + b_{4t}^s N_{at}^{18} - b_{3t}^s[(y_{at}^o - y_{at}^{s1}) + \beta_2 y_{at}^z z_{at}] && \text{for } y_{at}^{s1}(N_a^{18}) < y_{at}^o < y_{at}^{s2}(N_a^{18}) \\
&= 0 && \text{otherwise.}
\end{aligned}
$$

We refer to $b_t^s(N_{at}^{18}, y_{at}^o, y_{at}^z)$ as the benefit rule, and the b_{kt}^s's as the State/time specific benefit rule parameters. We exclude β_2 from this set because it is treated as a fixed parameter in our model.[17]

As the benefit rule parameters, and thus benefits themselves, change over time, women who are at all forward-looking should incorporate forecasts of future values of the rule parameters into their decision rules. We assume that benefit rule parameters evolve according to the following general vector autoregression (VAR), and that women use the VAR to form their forecasts of future benefit rules:

$$(6)\quad \mathbf{b}_t^s = \boldsymbol{\lambda}^s + \boldsymbol{\Lambda}^s \mathbf{b}_{t-1}^s + \mathbf{u}_t^s$$

where $\mathbf{b}_t^s$ and $\mathbf{b}_{t-1}^s$ are 5×1 column vectors of the benefit rule parameters, $\boldsymbol{\lambda}^s$ is a 5×1 column vector of regression constants, $\boldsymbol{\Lambda}^s$ is a 5×5 matrix of autoregressive parameters and $\mathbf{u}_t^s$ is a 5×1 column vector of iid innovations drawn from a stationary distribution with variance-covariance matrix $\boldsymbol{\Xi}^s$. We call (6) the evolutionary rule (ER) and $\boldsymbol{\lambda}^s$, $\boldsymbol{\Lambda}^s, \boldsymbol{\Xi}^s$ the parameters of the ER. Evolutionary rules are specific to the woman's state of residence.[18]

The woman is assumed to maximize her expected present discounted value of remaining lifetime utility at each age. The maximized value (the value function) is given by

---

[17] β_2 represents the fraction of parent's income that is included when calculating a woman's benefits if she lives with her parents. The exact treatment of parents' income is quite complicated, and varies even within States. Thus, rather than attempting to model this explicitly, as an approximation we chose to estimate β_2 as a fixed parameter.

[18] As noted earlier, it is assumed that a woman remains in the same location from age 14 on. Clearly, introducing the possibility of moving among states in a forward-looking model such as this would greatly complicate the decision problem.

$$(7)\quad V_a(\Omega_a) = \max\ E\,[\sum_{\tau=a}^{62} \delta^{\tau-a}\, U_\tau(\cdot)\,|\,\Omega_a]\,,$$

where the expectation is taken over the distribution of future preference shocks, labor market, marriage and parental co-residence opportunities, and the distribution of the future innovations of the benefit ER. The decision period is six months until age 45, the assumed age at which the women becomes infecund, but one year thereafter.[19] In (7), the state space Ω_a denotes the relevant factors known at age a that affect current or future utility or that affect the distributions of the future shocks and opportunities.

The solution to the optimization problem is a set of age-specific decision rules that give the optimal choice at any age, conditional on the elements of the state space at that age. Recasting the problem in a dynamic programming framework, the value function, $V_a(\Omega_a)$, can be written as the maximum over alternative-specific value functions, denoted $V_a^j(\Omega_a)$, i.e., the expected discounted value of choice $j \in J$, that satisfy the Bellman equation, namely

$$\begin{aligned} & V_a(\Omega_a) = \max_{j\in J}[V_a^J(\Omega_a)] \\ (8)\quad & V_a^j(\Omega_a) = U_a^j + \delta E(V_{a+1}(\Omega_{a+1})|j\in J,\Omega_a) \quad \text{for } a < A, \\ & \qquad\quad\ = U_A^j \qquad\qquad\qquad\qquad\qquad \text{for } a = A. \end{aligned}$$

A woman at each age a chooses the option j that gives the greatest expected present discounted value of lifetime utility. The value of each option depends on the current state Ω_a, which depends on the State of residence and the ER for that State (6), as well as the current realizations of the benefit rule parameters, preference shocks, own and husband's earnings shocks, parental income shocks, and labor market, marriage and parental co-residence opportunities.

The solution of the optimization problem is in general not analytic. In solving the model numerically, one can regard its solution as consisting of the values of $EV_{a+1}(\Omega_{a+1}\,|\,j\in J,\Omega_a)$ for

[19] Allowing for a longer decision period after age 45 reduces the computational burden of the model (see Wolpin (1992)).

all j and elements of Ω_a. We refer to this function as Emax for convenience. As seen in (8), treating these functions as known scalars for each value of the state space transforms the dynamic optimization problem into the more familiar static multinomial choice structure. The solution method proceeds by backwards recursion beginning with the last decision period.[20]

III. Data

The 1979 youth cohort of the National Longitudinal Surveys of Labor Market Experience (NLSY79) contains extensive information about schooling, employment, fertility, marriage, household composition, geographic location and welfare participation for a sample of over 6,000 women who were age 14-21 as of January 1, 1979. In addition to a nationally representative core sample, the NLSY contains oversamples of blacks and Hispanics. We use the annual interviews from 1979 to 1991 for women from the core sample and from the black and Hispanic oversamples.

The NLSY79 collects much of the relevant information on births, marriages and divorces, periods of school attendance, job spells, and welfare receipt, as dated events. We adopt a decision period of six months. Periods are defined on a calendar year basis, beginning either on January 1 or on July 1 of any given year. The first decision period is the first six month calender period that the woman turns age 14. The last period we observe is the second six month calendar period in 1990 (or, if the woman attrited before then, the last six-month period in which the data are available). The first calendar period observation, corresponding to that of the oldest NLSY79 sample members, occurs in the second half of 1971. There are fifteen other birth cohorts who turned age 14 in each six month period through January, 1979.

We restrict the sample to the six States in the U.S. that have the largest representations of NLSY79 respondents: California, Michigan, New York, North Carolina, Ohio and Texas. The estimation is performed using only the first five States. Texas is used as a holdout or validation

[20] Because the size of the state space is large, we adopt an approximation method to solve for the Emax functions. The Emax functions are calculated at a limited set of state points and their values are used to fit a polynomial approximation in the state variables consisting of linear, quadratic and interaction terms. See Keane and Wolpin (1994, 1997) for details. As a further approximation, we let the Emax functions depend on the expected values of the next period benefit parameters, rather than integrating over the benefit rule shocks.

sample on which to perform out-of-sample validation tests of the model. The reason for this choice is that Texas is by far the least generous State in terms of welfare benefits and thus requires an fairly extreme out-of-sample extrapolation.

As noted, we consider the following choices: whether or not to (i) attend school (ii) work (part- or full-time), (iii) be married, (iv) become pregnant and (v) receive welfare (AFDC). The variables are defined as follows:

School Attendance: The NLSY79 collects a monthly attendance record for each women beginning as of January, 1979. A woman was defined to be attending school if she reported being in school each month between January and April in the first six-month calendar period and each month between October and December in the second calendar period. Given the sample design of the NLSY79, school attendance records that begin at age 14 exist only for the cohort that turned 14 in January, 1979.

Employment Status: Using employment event history data, we calculated the number of hours worked in each six month period. A woman was considered working part-time in the period (500 hours) if she worked between 260 and 779 hours and full-time (1000 hours) if she worked at least 780 hours. As with school attendance, employment data does not extend back to age 14 for many of the cohorts. We assume that initial work experience, that is, at age 14, is zero.

Marital Status: The NLSY79 provides a complete event-dated marital history that is updated each interview. However, dates of separation are not reported. Therefore, for the years between 1979 and 1990, data on household composition was used to determine whether the woman was living with a spouse. But, because these data are collected only at the time of the interview, marital status is treated as missing during the non-interview periods, in most cases for one six-month period per year. Marital event histories were used for the periods prior to 1979 even though it is uncertain from that data whether the spouse was present in the household.

Pregnancy Status: Although pregnancy rosters are collected at each interview, conception dates are noisy and miscarriages and abortions are under-reported. We ignore pregnancies that do not lead to a live birth, dating the month of the conception as occurring nine months prior to the month of birth. Except for misreporting of births, there is no missing information on pregnancies back to age 14 for any of the cohort.

Welfare Receipt: AFDC receipt is reported for each month within the calendar year preceding the interview year, i.e., from January 1978. We define a woman as receiving welfare in a period if she reported receiving an AFDC payment in at least three of the six months of the period.[21] As with school attendance and employment, data are missing back to age 14 for most of the cohorts. It is assumed that none of the women received welfare prior to age 14, as is consistent with the fact that none had borne a child by that time.

Descriptive Statistics:

Table 1 provides (marginals of) the sample choice distribution by full-year ages and by race aggregated over the five states used in the estimation. Notice that school attendance is essentially universal until age 16, drops about in half at age 18, the normal high school graduation age, and falls to around 10 percent at age 22. About 3 percent of the sample attends school at ages after 25. The implied school completion levels that result from these attendance patterns are, at age 24, 12.9 for whites, 12.7 for blacks and 12.2 for Hispanics. Employment rates for white and Hispanic women (working either part- or full-time) increase rapidly through age 18 but only slowly thereafter. They are higher for whites throughout by about 10-20 percentage points. In contrast, employment rates for black females rise more continuously, roughly doubling between age 18 and 25, and are comparable to that of Hispanics at ages after 25.

Marriage rates rise continuously for whites and Hispanics, reaching about 60 percent by age 25 for whites and 50 percent for Hispanics. However, for blacks, marriage rates more or less reach a plateau at about age 22, at between 20 and 25 percent. With respect to fertility, by age 20, white, black and Hispanic females had, on average, .28, .47 and .40 live births, respectively. By age 27, these figures are 1.06, 1.36 and 1.39, respectively, and by age 30, they are 1.54, 1.61 and 1.76, respectively. Welfare participation increases with age, at least through age 24. Race differences are large; participation peaks at 7 percent for whites, 28 percent for blacks and 17 percent for Hispanics. Teenage pregnancies that lead to a live birth are higher by 68 percent for blacks than for whites and by 43 percent for Hispanics than for whites.

---

[21] The use of almost any cutoff in establishing welfare participation would have only a small effect on the classification; most women who report receiving welfare in any one month during a six month period report receiving it in all six months.

There are large differences between women in the five States used in estimation (the estimation sample) and Texas (the validation sample). The largest differences are for AFDC take-up and for full-time employment. For example, among black women, welfare receipt peaked at about 30 percent in the estimation sample, while it peaked at only about 10 percent in the validation sample. Full-time employment is much higher in Texas. For example, at age 25, the difference in the proportion engaged in full-time work is 14.3 percentage points for whites, 18.9 percentage points for blacks and 19.6 percentage points for Hispanics.[22]

Benefit Rules:

In order to estimate the benefit schedules (5), and the evolutionary rules governing changes in benefit parameters (6), we collected information on the rules governing AFDC and Food Stamp eligibility and benefits in each of the 50 states for the period 1967-1990. We then simulated a large sample of hypothetical women, with different numbers of children and different labor and non-labor income, and calculated their benefits according to the exact rules in each State and each year. We took the sum of monthly AFDC and Food Stamp benefits, and expressed them in 1987 New York equivalent dollars. This data set was used to estimate the (approximate) benefit rule parameters, $b_{t0}^s, b_{t1}^s, b_{t2}^s, b_{t3}^s, b_{t4}^s$, for each State s and year t.[23] Given the estimates of the benefit rule parameters, we then estimated (6), the evolutionary rule, for each State s.

Table 2 transforms the benefit parameters obtained from the estimates of (5) into a more convenient set of benefit measures, namely the total monthly income of non-working women (with zero non-earned income) who have either one or two children, and the total monthly income of women with one or two children who have part-time monthly earnings of 500 dollars or full-time earnings of 1000 dollars.[24] Referring to table 2, among the six states, NY, CA and MI are considerably more generous than NC, OH and TX. Among the first group Michigan is the most generous, with average benefits over the 24 years for a woman with one child being 654

[22] See the working paper version, Keane and Wolpin (2005), for further details.

[23] The approximation given by (5) fits the monthly benefit data quite well, with R-squared statistics for the first line segment mostly above .99 and for the second, mostly about .95. See Keane and Wolpin (2005) for the regression estimates.

[24] See Keane and Wolpin (2005) for summary statistics on the benefit parameters.

(1987 NY) dollars per month. Among the second group Texas is the least generous, with the same average benefits figure only 377 dollars. CA and NY were about equally generous on average (589 and 574 dollars) over the period as were NC and OH (480 and 489 dollars).[25]

As table 2 reveals, there was a steep decline in benefit amounts between the early 1970's and the mid 1980's, and relative constancy thereafter. For example, in Michigan monthly benefits fell from 735 dollars for a woman with no earnings and two children in 1975 to 561 dollars in 1985. For the same woman with 500 dollars in monthly earnings, benefits fell from 762 dollars in 1975 to 405 dollars in 1985, and then rose slightly to 484 dollars in 1990.

IV. Estimation Method:

Numerical solution to the agents' maximization problem provides (approximations to) the Emax functions appearing on the right hand side of (8). The alternative-specific value functions, V_t^j for j=1,..,J, are known to the agents. But, in general, the econometrician does not observe the random preference shocks, full- and part-time wage offer shocks, the earnings shock of the husband, and the income of parents. Also, the choice set depends on shocks governing whether a marriage offers, job offers and parental co-residence offers are received.

Thus, conditional on the deterministic part of the state space, the probability an agent is observed to choose option j is an integral over the region of the several-dimensional error space such that k is the preferred option. The order of integration needed to form the likelihood contribution depends on the observed choice. For example, if the agent chooses a work option, then the wage offer is observed, and we need not integrate over the wage shock. In that case, the likelihood of the observation includes the density of the wage error.

As the choice set contains up to 36 elements, evaluation of the choice probabilities is computationally burdensome. Efficient smooth unbiased probability simulators, such as the GHK method (see, e.g., Keane (1993, 1994)), are often useful in such situations. Unfortunately, smooth unbiased simulators like GHK generally rely on a structure in which the value of each alternative

---

[25] Benefit reduction rates for AFDC and for Food Stamps are federally set. They differ across states in our approximation due to the fact that AFDC payments terminate at different income levels among the states while food stamp payments are still non-zero and the two programs have different benefit reduction rates. There is thus a kink in the schedule of total welfare payments with income that our approximation smooths over.

is a strictly monotonic function of a single stochastic term, and where (J-1)•(J-1) variance matrix of the error terms has full rank. Also, as discussed in Keane and Moffitt (1998), when the number of choices exceeds the number of errors, boundaries of the regions of integration for the choice probabilities are often intractably complex. Thus, given our model, the most practical method to simulate the choice probabilities would be to use a kernel smoothed frequency simulator. These were proposed in McFadden (1989), and have been successfully applied to models with large choice sets in Keane and Moffitt (1998) and Keane and Wolpin (1997).[26]

In the present context, however, standard simulated maximum likelihood methods are not feasible because of severe problems created by unobserved state variables. As noted, we do not generally have complete histories of employment, schooling or welfare take-up back to age 14, so the state variables for work experience, schooling and welfare dependence cannot be constructed. Parental co-residence and marital status are also only observed once a year.

A further complication is that a youth's initial schooling level at age 14 is observed only for one of the 16 cohorts. Unobserved initial conditions (see Heckman (1981)), and unobserved state variables in general, pose formidable computational problems for estimation of dynamic discrete choice models. If some elements of the state space are unobserved, one must integrate over their distributions to construct the conditional choice probabilities. Even in much simpler dynamic models than ours, such distributions are typically computationally intractabe.

Keane and Wolpin (2001b) develop a simulation algorithm that deals in a practical way with unobserved state variables. The idea is based on simulating complete outcome histories (age 14 to the terminal age) for a set of artificial agents. A history consists of initial schooling, S_0, parental schooling, S^z, and simulated values in all subsequent periods for all of the outcome variables in the model (i.e., school attendance, part- or full-time work, etc.). Denote by $\tilde{O}^n$ the simulated outcome history for the nth such person, $\tilde{O}^n = (S_{14}^n, S^z, \tilde{O}_{a=1}^n, \ldots, \tilde{O}_{a=A}^n)$, for n = 1,..., N.

To motivate the algorithm, it is useful to ignore for now the complication that some outcomes are continuous. Let O^i denote the observed outcome history for person i, which may include missing elements. An unbiased frequency simulator of the probability of the observed

[26] Kernel smoothed frequency simulators are biased for positive values of the smoothing parameter, so consistency requires it to approach zero as sample size increases.

outcome history for person i, $P(O^i)$, is just the fraction of the N simulated histories that are consistent with O^i. In this construction, missing elements of O^i are counted as consistent with any entry in the corresponding element of $\tilde{O}^n$. The construction of this simulator relies only on unconditional simulations: it does not require evaluation of choice probabilities conditional on state variables. Thus, unobserved state variables create no problem for this procedure.

Unfortunately, because the number of possible outcome histories is huge, consistency of a simulated history with an actual history is an extremely low probability event. Hence, simulated probabilities will typically be 0, as will be the simulated likelihood, unless an impractically large simulation size is used (see Lerman and Manski 1981). In addition, the method breaks down if any outcome is continuous, e.g., the woman's wage offer, regardless of simulation size, because agreement of observed with simulated wages is a measure zero event.

We solve this problem by assuming, as is apt, that all observed quantities are measured with error. With measurement error there is a nonzero probability that any observed outcome history might be generated by any simulated outcome history. Denote by $P(O^i|\tilde{O}^n)$ the probability that observed outcome history O^i is generated by simulated outcome history $\tilde{O}^n$. Then $P(O^i|\tilde{O}^n)$ is the product of classification error rates on discrete outcomes and measurement error densities for wages that are needed to make O^i and $\tilde{O}^n$ consistent. Observe that $P(O^i|\tilde{O}^n)>0$ for any $\tilde{O}^n$, given suitable choice of error processes. The specific measurement error processes that we assume are described below. The key point here is that $P(O^i|\tilde{O}^n)$ does not depend on the state variables at any age a, but only depends on the outcomes.

Using N simulated outcome histories we obtain the unbiased simulator

$$(9)\quad \hat{P}_N(O^i) = \frac{1}{N}\sum_{n=1}^{N} P(O^i|\tilde{O}^n).$$

This simulator is analogous to a kernel-smoothed frequency simulator, in that $I(O^i=\tilde{O}^n)$ is replaced with an object that is strictly positive, but that is greater if $\tilde{O}^n$ is "closer" to O^i. However, (9) is unbiased as measurement error is assumed to be present in the true model.

It is straightforward to extend the estimation method to allow for unobserved heterogeneity. Assume that there are K types of women who differ in their permanent preference parameters and skill endowments.[27] Let $\pi_{k,0}$ and π_k denote the frequency of type k according to the simulation and the model, respectively, and let $\tilde{O}_k^n$ indicate that artificial agent n is of type k. Then, we obtain:

$$(10)\quad \hat{P}_N(O^i) = \frac{1}{N}\sum_{n=1}^{N} P(O^i|\tilde{O}_k^n)\frac{\pi_k}{\pi_{k0}}.$$

Observe that in (10), the conditional probabilities $P(O^i|\tilde{O}^n)$ are weighted by the ratio of the proportion of type k according to the model to the proportion of type k in the simulation.

This simulator is not smooth because $P(O^i|\tilde{O}^n)$ will "jump" at points where changes in model parameters cause a simulated outcome history $\tilde{O}^n$ to change discretely. But (10) can be smoothed by applying an importance sampling procedure: hold simulated outcome histories fixed and re-weight them as parameters are varied. Given an initial parameter vector θ_0 and updated vector θ', the appropriate weight for sequence $\tilde{O}^n$ is the ratio of the likelihood of simulated history n under θ' to that under θ_0. This importance sampling weight (i.e., ratio of densities under the target and source distributions) is a smooth function of model parameters. Note it is straightforward to simulate the likelihood of an artificial history $\tilde{O}^n$ using conventional methods because the state vector is fully observed at all points along the history. Choice probabilities along a path $\tilde{O}^n$ are simulated using a kernel smoothed frequence simulator. As $P(O^i|\tilde{O}^n)$ is now smooth in the model parameters, standard errors can be obtained by the BHHH algorithm.

For the measurement error processes, we assume discrete outcomes are subject to a simple form of classification error: there is some probability the reported response category is the truth and some probability that it is not.[28] For continuous variables, we assume that the woman's wage error and the husband's income error are multiplicative and the parents' income error is additive. Measurement errors are assumed to be serially and mutually independent.

---

[27] Initial schooling, parents schooling and the latent type are drawn from a joint distribution that we estimate. In (10), type refers to this vector of three initial conditions.

[28] To ensure that the measurement error is unbiased, the probability that the reported value is the true value must be a linear function of the predicted sample proportion (see Keane and Wolpin (2005) for details). Keane and Sauer (2005) have applied this algorithm successfully with more general classification error processes

V. Results

Both to provide a metric for assessing the fit of the dynamic programming (DP) model, and as possible alternatives to it, we also estimated a multinomial logit (MNL) that relates four of the five choice variables, welfare take-up, school attendance, work and pregnancy, to the state variables of the model at each age. We estimated four different specifications of the MNL, but present the results for now of only the one that best fit the estimation and validation samples.[29]

The variables included in the MNL are the benefit amount for a woman with one child and no earnings, State dummies, age, age squared, parents schooling, whether the woman was on welfare, worked or was pregnant in the previous period, whether she was pregnant two periods before, the number of children already born, the woman's years of schooling and its square, whether the woman was living in a nuclear family at age 14, and race dummies. There are 13 mutually exclusive choices (3 were combined because of small cell size) and 240 parameters.

In comparison, the DP model is more comprehensive, including also a marriage decision and a distinction between full or part-time work. It also embeds additional structural relationships (functions describing the probability of living with a parent, husband's income if married and parent's income if co-resident, and full and part-time wage offers). Nevertheless, that DP model has a similar number of parameters (202).

Table 3 shows the fit to the estimation sample for the MNL and the DP models by four age groups (15-17.5, 18-21.5, 22-25.5, 26-29.5) for each race separately. Although there are clear differences in the fit of the two models, neither seems to be uniformly better. For example, the MNL fits welfare take-up better for blacks than does the DP model, but fits Hispanics worse and whites about the same. Similarly, the MNL model seems to fit the work alternative better for Hispanics at earlier ages, but the DP model fits better at later ages. Both models capture well age trends and quantitative differences by race. The table also compares the fit to two of the state variables, the mean number of children ever born before ages 20, 24 and 28, and the mean highest grade completed by age 24. The performance is similar with respect to these measures, except for the severe overstatement of schooling for Hispanics by the MNL model.

Table 4 presents the same comparison for the validation sample. The MNL clearly does

---

[29] These regressions are available on request.

better than the DP model in terms of welfare participation, especially for blacks in the last two age groups. Other comparisons also seem to favor the MNL, although differences seem to be small. As with the estimation sample, age trends and racial differences are captured well. Neither model is very far off in forecasting children ever born or schooling.[30]

To summarize the overall fit to the estimation and validation samples, table 5 reports the root mean squared error (RMSE), calculated from deviations between actual and forecasted age-specific means, for the DP model and the four alternative MNL models. The MNL model that we discussed above is labeled MNL1 - FE, where FE indicates it includes State fixed-effects, and "1" indicates welfare rules are summarized by a single variable (the benefit level for a woman with one child). MNL1 - No FE drops the fixed effects, but instead includes the mean one-child benefit for the State over the period 1967-1990. In model MNL2 - FE, the "2" indicates that the five state-specific benefit parameters from (5) are included in the specification separately, instead of just of the one-child benefit. MNL2 - No FE drops the State fixed-effects, and instead includes the means of those five benefit parameters over the 1967-1990 period.

With respect to the estimation sample, all of the MNL models and the DP model appear about equally good. Notable exceptions are the better fit of the DP model to school attendance among whites (.028 vs. .044 for MNL1- No FE), the worse fit of the DP model to work (.066 vs. .030 for MNL1- No FE) for blacks, and the better fit of the DP model to welfare (.024 vs. .044 for MNL1 - FE), to work (.048 vs. .059 for MNL2 - FE) and to school attendance (.033 vs. .048 for MNL1 - No FE) for Hispanics.

On the other hand, large differences in fit emerge for the validation sample.[31] Among the

---

[30] We also considered the fit of the DP model to all of the other variables for both the estimation sample and the validation sample (see Keane and Wolpin (2005)). The fit with respect to the estimation sample is uniformly good, capturing well age trends and racial differences. In some cases, the fit is remarkably close. For example, because of selection, fitting accepted wages when working percentages are low is challenging. Nevertheless, the DP model predictions are quite close to the actual data. For example, predicted mean accepted wage rates are often within 5 percent of the actual wage rates.

[31] To forecast Texas for the MNL models with state dummies, we re-estimated the model on Texas data with Texas state dummies, constraining all other parameters to be the same as in the estimation sample.

MNL models, the two that include state dummies (MNL1 - FE and MNL2 - FE) have the lowest root mean squared errors. Although using the five benefit parameters to capture welfare rules (MNL2 - FE) provides a statistically significant improvement in the estimation-sample fit, there is no discernible impact on the root mean squared error for the validation sample.[32]

Replacing the state dummies with the mean one-child benefit does adversely affect the RMSE. For example, comparing the MNL1 - FE vs. MNL1- No FE models, the largest increases in RMSE are from .068 to .093 for work and from .046 to .086 for school attendance for whites, from .021 to .030 for welfare for blacks, and from .050 to .062 for work and from .034 to .059 for school attendance for Hispanics.

But, the deterioration in fit is much greater for the MNL2 models. Dropping the state dummies, and instead including the five state-specific means of the five benefit parameters in (5), increased the RMSE enormously. The fit to welfare was particularly adversely affected, rising from .010 (MNL2 - FE) to .815 (MNL2 - No FE) for whites, from .021 to .844 for blacks and from .014 to .842 for Hispanics. Thus, in specifications that include only the one-child benefit (MNL1) instead of the five benefit rule parameters (MNL2), dropping the state fixed-effects does not lead to such a serious deterioration of the fit to Texas. We take this result as evidence that the validation sample is capable of identifying over-fitting in a way that the estimation sample was not.

The DP model uniformly does not fit as well as MNL1 - FE and overall fits about the same as MNL1 - No FE; fitting better for whites, but worse for blacks and Hispanics. Based on the evidence from this validation exercise, it would therefore appear that MNL1 - FE would be the best model to use for counterfactual experiments.

Given that expectation, table 6 reports on the results from a counterfactual experiment where the estimation sample states are given Texas' welfare benefits. We report on the effects for both MNL1 specifications and for the DP model. The predicted effects from the MNL1 - FE specification are seemingly perverse. Welfare receipt and fertility are predicted to increase substantially, while there is a similarly large decline in work. The predictions from the MNL1-

---

[32] The chi-square statistic for the joint test that all of the additional benefit parameters are zero has a p-value of .000.

No FE specification are exactly the opposite, a large reduction in welfare receipt, a large increase in work and a relatively small reduction in fertility.

Keane and Wolpin (2001a) noted an important distinction between specifications with and without state-specific effects. If women are forward looking, the effect of a change in welfare benefit rules on behavior depends critically on how that change affects expectations about future benefit rules. Changes in welfare benefits can have very different effects depending on whether they are perceived as being permanent or transitory. Estimates that use different sources of variation in benefits, variation across states versus variation within states over time, may result in different estimates simply because they identify responses to benefit changes that may be perceived as having different degrees of permanence.

For example, if benefits are change from year-to-year, the effect of a change in the current year's benefits on fertility will depend on the degree to which the change is viewed as permanent. This, in turn, depends on the process by which benefits evolve and how potential welfare recipients form expectations. If the perceived benefit process is such that an increase in benefits in one year is anticipated to be followed by declines in subsequent years, then it is possible that fertility may actually respond negatively to the transitory increase.

Thus, the counterfactual using MNL1 - FE is not, under this interpretation, identifying the effect of replacing the estimation sample states' welfare systems with Texas' system. Recall that in performing the external validation exercise, we estimated Texas state dummies, which has the effect of rescaling the means. The counterfactual, however, uses the state dummies obtained from the estimation sample. Given these results, the superior performance of the MNL1-FE in forecasting Texas is an artifact of the treatment of the state dummies.

Unlike MNL1 - FE, MNL1 - No FE replaces not only benefit realizations but also the mean, or permanent benefit level as well. However, the effects predicted by MNL1 - No FE appear to be implausible as well. For example, while welfare participation among whites falls by 3.8 percentage points (from 4.5 to 0.7 percent) at ages 26-29.5, employment increases by 12.2 percentage points. Indeed, for all three race groups, the reduction in welfare participation is usually considerably less than the increase in employment at all ages. The prediction that employment rates would reach close to 90 percent (for whites) with the adoption of Texas' welfare benefits is also not credible.

The counterfactual based on the DP model, which accounts for the entire set of welfare parameters, replaces each of the estimation sample state's benefit realizations as well as its evolutionary rule (6) with that of Texas' realizations and rule. The resulting effects are more modest than in the MNL1 - No FE specification. The largest effects are for Hispanics, where welfare participation falls by as much as 5 percentage points (from 15.3 to 10.2 percent) at ages 22-22.5 and employment increases by 3 percentage points at those ages. For all races, within each age group, the increase in employment is no larger than the fall in welfare participation. In addition, for each race, mean schooling by age 25 increases, though very slightly. Overall, the results from the DP model appear more reasonable than the MNL - No FE specification.[33]

VI. Conclusions:

In this paper, we presented and structurally estimated a dynamic programming (DP) model of life-cycle decisions of young women. The model significantly extends earlier work on female labor supply, fertility, marriage, education and welfare participation by treating all five of these important decisions as being made jointly and sequentially within a life-cycle framework. Needless to say, the resulting model is quite complex, and many behavioral and statistical assumptions were needed to make its solution and estimation feasible. Of course, the model is literally false, as our assumptions are designed to abstract from and simplify the full complexity of how people really make life-cycle decisions. Thus, the model is simultaneously both mathematically complex, yet highly stylized as a depiction of actual behavior. Nevertheless, we believe that such models, tightly specified on the basis of very specific theoretical and statistical assumptions, are potentially quite useful for policy analysis. The issue addressed in this paper is how to develop faith, or validate, that such a model is indeed useful.

To that end, we pursued a range of approaches to model validation. First, we examined the within-sample fit of the DP model across a number of dimensions of interest and compared its fit to a group of "flexible" models, specified as four alternative multinomial logits estimated

[33] Effects of the counterfactual experiment for the DP model on additional variables, considered in Keane and Wolpin (2005), are predicted to be quite small. For example, by ages 26-29.5, the marriage rate is predicted to increase from 65.6 to 65.9 percent for whites, from 28.2 to 28.8 percent for blacks and from 55.7 to 56.9 percent for Hispanics.

on a subset of the choice data. Using a RMSE criterion (the number of parameters are similar), there seemed to be no clear winner in this cross-model competition. Based on these results, our view was that the DP model fit the data well enough to continue to consider it as potentially useful for prediction.

Second, we performed an external validation using data from a non-random holdout sample, specifically data from the state of Texas, which had a very different welfare policy regime from the five states that were used in estimation. In terms of the same RMSE criterion, one of the models (MNL2-No FE) produced predictions for Texas that were terribly inaccurate by any standard, leaving us with no faith in its usefulness. Further, the models with state fixed-effects (MNL1-FE, MNL2-FE) fit the data from Texas better than both the DP model and the remaining logit (MNL1 - No FE), although there was no clear ranking of these latter two models.

However, we were surprised that the state fixed-effect specification, which identifies the behavioral effects of only transitory changes in welfare benefits, could best predict the effect of Texas' permanently different welfare rules. This result led us to a third method of validation. We used the models to predict the effect of a policy intervention that has no analogue in the historical data, but where basic economic theory provides tight priors, in a qualitative sense, on certain aspects of what might possibly happen.

The counterfactual experiment was to give the five estimation states the same welfare rules as Texas. Our strong priors were: (i) that welfare participation should drop, since the Texas benefits are less generous and (ii) that work should increase, but that the decline in welfare places an upper bound on the increase in work. The specifications with state fixed effects violated the first, most basic, theoretical prediction that welfare participation be increasing in welfare generosity. The only reason that the state fixed-effects specification had performed better in the external validation was because, in that exercise, we had estimated Texas state dummies. Moreover, the specification without state fixed -effects, whose performance in terms of the fit to the estimation sample and to the Texas hold-out sample was satisfactory, violated the second prior. Thus, we came to view all four MNL models as unreliable for policy prediction. In contrast, the predictions of the DP model were consistent with both priors.

In summary, the DP model, in our judgement, performed well on three different tests of validity. In light of this evidence, we updated our priors about the potential usefulness of the model (for policy prediction) in a favorable direction. Our research strategy is to continue to look for opportunities to further validate the model, and as these opportunities arise they will either increase or reduce our confidence in the model's usefulness.

One opportunity is presented by the important changes in welfare rules that occurred beginning in the mid-1990s, after our sample period ended. This included EITC expansion, imposition of work requirements for receipt of benefits, and benefit receipt time limits. As discussed in Fang and Keane (2004), there was substantially heterogeneity across states in terms of how exactly these policy changes were structured, and we can use our model to simulate the impact of these changes on a state-by-state basis.[34]

As a final observation, we conjecture that most economists would have professed a greater *a priori* faith in the ability of the MNL models to forecast behavior than in the DP model. That is, they would be concerned that, because the many assumptions invoked in setting up the DP model could all be questioned, it is unlikely such a model could forecast accurately. In contrast, they would view the MNL models, which simply model the value of each alternative as a flexible function of the state variables, as being much less "restrictive." Thus, the poor predictions that the MNL models produced should serve as a cautionary tale, from which we draw two morals.

First, economists should be concerned with model validation regardless of the estimation approach; one needs to hold all models to the same standard. Second, our experience illustrates well the potential strengths of DP models for making policy predictions. It is precisely the economic structure of the model that constrains it to make predictions that are reasonable in a qualitative sense. The MNL models' failure is, at least in part, attributable to the fact that they lack sufficient economic structure to impose such reasonable constraints on their predictions. Economics is indeed valuable in econometrics.

[34] Of course, this experiment provides only an imperfect validation tool because other aspects of the economic and social environment may have changed.

References

Bajari, Patrick and Ali Hortacsu. “Are Structural Estimates from Auction Models Reasonable? Evidence from Experimental Data.” Journal of Political Economy, 113, August 2005, 703-741

Bontemps, Christian, Jean-Marc Robin and Gerard J. van den Berg. “Equilibrium Search with Continuous Productivity Dispersion: Theory and Nonparametric Estimation.” International Economic Review, 41, May, 2000, 305-358.

Busemeyer, Jerome R. And Yi-Min Wang. “Model Comparisons and Model Selections Based on Generalization Criterion Methodology.” Journal of Mathematical Psychology, 44, 2000, 171-189.

Eckstein, Zvi and Kenneth I. Wolpin. “Estimating a Market Equilibrium Search Model from Panel Data on Individuals.” Econometrica, 58, July, 1990, 783-808.

Heckman, James J. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence." In Structural Analysis of Discrete Data with Econometric Applications, Charles F. Manski and Daniel McFadden (eds.), Cambridge, MA: MIT Press, 1981.

Keane, Michael P. “Simulation Estimation for Panel Data Models with Limited Dependent Variables.” in G.S. Maddala, C.R. Rao and H.D. Vinod eds., Handbook of Statistics 11, Amsterdam: Elsevier Science Publishers, 1993, 545-572.

___________. “A Computationally Practical Simulation Estimator for Panel Data.” Econometrica, 62, January, 1994, 95-116.

___________. "A New Idea for Welfare Reform," Federal Reserve Bank of Minneapolis Quarterly Review, 1995, 19, 2-28.

Keane, Michael P. and Robert Moffitt. “A Structural Model of Multiple Welfare Program Participation and Labor Supply.” International Economic Review, 39, August 1998, 553-590.

Keane, Michael and Rob Sauer. “A Computationally Pratical Simulation Estimati0on Algorithm for Dynamic Panel Data Models with Unobserved Endogenous State Variables.” mimeo, Yale University, 2005.

Keane, Michael P. and Kenneth I. Wolpin. “The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence.” Review of Economics and Statistics, 76, November 1994, 684-672.

___________. “The Career Decisions of Young Men.” Journal of Political Economy, 105, June 1997, 473-522.

___________. “Estimating Welfare Effects Consistent with Forward-Looking Behavior, Part I: Lessons From a Simulation Exercise.” Journal of Human Resources, 37, Summer 2001a, 600-622.

___________. “The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment.” International Economic Review, 42, November 2001b, 1051-1103.

___________. “Exploring the Usefulness of a Non-Random Holdout Sample for Model Validation: Welfare Effects on Female Behavior.” mimeo, May, 2005.

___________. “The Role of Labor and Marriage Markets, Preference Heterogeneity and the Welfare System on the Life Cycle Decisions of Black, Hispanic and White Women.” mimeo, January, 2006.

Lerman, Steven R. and Charles F. Manski. “On the Use of Simulated Frequencies to Approximate Choice Probabilities. ” In Structural Analysis of Discrete Data with Econometric Applications, Charles F. Manski and Daniel McFadden (eds.), Cambridge, MA: MIT Press, 1981.

Lise, Jeremy, Jeffrey Smith and Shannon Seitz. “Equilibrium Policy Experiments and the Evaluation of Social Programs,” mimeo, Queens University, 2005.

Lumsdaine, Robin, James Stock and David Wise. “Three Models of Retirement: Computational Complexity vs. Predictive Validity.” in D. Wise, ed. Topics in the Economics of Aging. Chicago: University of Chicago Press, 1992.

McFadden, Daniel. “A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration.” Econometrica, 57, September, 1989, 995-1026.

McFadden, Daniel. Urban Travel Demand Forecasting Project Final Report, Volume 5, Institute of Transportation Studies, University of California, Berkeley, June 1977.

Moffitt, Robert. “Incentive Effects of the U.S. Welfare System: A Review.” Journal of Economic

Literature, 30, March 1992, 1-61.

Mosier, C.I.. "Problems and Designs of Cross-Validation." Educational and Psychological Measurement. 11, 1951, 5-11.

Rosenzweig, Mark R. and Kenneth I. Wolpin. "Natural 'Natural Experiments' in Economics." Journal of Economic Literature, 38, December, 2000, 827-874.

Swann, Christopher A. "A Dynamic Analysis of Marriage, Labor Force Participation, and Participation in the AFDC Program." Mimeo, University of Virginia, 1996.

Todd, Petra and Kenneth I. Wolpin. "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility." mimeo, University of Pennsylvania, 2002.

Wise, David. "A Behavioral Model Versus Experimentation: The Effects of Housing Subsidies on Rent." in P. Brucker and R. Pauly, (eds.), Methods of Operations Research *50*, Verlag Anton Hain, 1985, 441-489.

Wolpin, Kenneth I. "The Determinants of Black-White Differences in Early Employment Careers: Search, Layoffs, Quits and Endogenous Wage Growth." Journal of Political Economy, 100, June 1992, 535-560.

Table 1
Choice Distributions by Age: Estimation Sample of the Combined Five States

	Attending School			Working (PT or FT)			Married			Becomes Pregnant			Receives AFDC		
Age	W	B	H	W	B	H	W	B	H	W	B	H	W	B	H
14	100	93.3	100	14.3	10.5	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	97.7	100	100	11.4	9.9	5.2	0.0	0.0	0.0	1.0	3.4	1.0	1.0	1.3	0.0
16	88.3	87.5	90.3	30.0	14.5	19.3	3.0	1.0	2.9	3.1	3.8	2.1	1.0	1.0	1.0
17	84.6	80.7	79.2	50.0	26.9	32.4	8.7	1.4	6.4	5.6	5.3	2.5	1.3	2.5	2.3
18	42.8	50.9	41.5	63.0	32.6	50.7	16.4	3.7	11.9	3.7	4.5	6.7	2.6	9.0	3.3
19	32.5	32.1	27.1	65.6	43.4	51.2	24.9	7.1	19.9	4.5	8.6	5.6	3.6	15.6	6.8
20	23.8	22.2	18.8	67.5	46.4	52.2	31.5	11.7	27.1	4.3	6.0	4.9	5.4	17.3	10.3
21	19.4	12.3	12.2	69.6	49.2	58.3	37.1	14.4	34.2	6.0	7.9	6.3	5.1	21.2	13.7
22	10.8	8.3	7.7	70.0	52.5	60.6	37.5	20.3	35.9	4.5	5.3	5.7	6.1	25.6	15.1
23	4.2	6.2	3.9	72.0	54.2	58.5	49.1	22.3	39.7	5.9	6.1	5.3	6.2	27.2	15.3
24	3.8	5.4	4.6	72.7	55.4	57.7	54.1	22.8	45.7	6.6	6.9	7.9	7.0	27.8	17.2
25	4.0	5.9	2.9	73.8	62.8	55.6	58.5	20.9	47.2	7.6	7.0	7.2	6.4	26.8	16.0
26-29	3.2	3.6	2.2	71.5	61.1	56.7	63.6	25.6	52.1	5.8	4.4	5.8	5.0	25.7	15.4
30-33	4.5	2.3	2.6	72.6	63.3	64.9	72.8	32.0	56.7	4.3	2.3	5.3	2.6	22.3	14.5

Table 2
Summary Statistics of Total Monthly Benefits By Numbers of Children and Earnings by State: 1967-1990

	Monthly Earnings					
	Zero		$500		$1000	
	One child	Two children	One child	Two children	One child	Two children
CA						
μ	589	724	351	517	87	196
σ	60	67	85	91	89	151
1970	459	568	416	560	297	440
1975	652	794	441	620	132	311
1980	617	757	405	560	156	311
1985	596	730	260	414	0	46
1990	594	728	303	476	0	110
MI						
μ	654	809	429	621	150	304
σ	92	106	161	179	158	215
1970	671	830	585	799	302	516
1975	735	912	551	762	273	483
1980	660	808	424	602	152	330
1985	561	705	235	405	0	58
1990	551	694	293	484	0	156
NY						
μ	574	718	334	514	92	204
σ	52	71	126	152	98	189
1970	562	726	469	685	189	406
1975	635	798	443	643	172	372
1980	552	679	322	473	61	211
1985	524	644	189	334	0	0
1990	528	649	230	393	0	31

Table 2, continued

NC						
μ	480	566	274	384	35	132
σ	48	58	68	82	40	66
1970	455	513	348	432	143	227
1975	570	679	356	502	50	197
1980	462	553	260	364	31	134
1985	454	543	199	295	0	69
1990	438	530	249	367	13	131
OH						
μ	489	607	270	414	87	128
σ	34	43	69	88	36	87
1970	460	565	361	511	106	256
1975	552	688	339	514	27	202
1980	499	619	284	423	11	151
1985	459	570	185	305	0	0
1990	455	566	218	346	0	0
TX						
μ	377	476	217	329	69	106
σ	50	60	51	73	21	43
1970	417	514	297	429	169	201
1975	445	561	253	398	0	117
1980	334	436	198	295	0	96
1985	375	474	170	264	0	52
1990	343	442	181	287	0	101

Table 3 - Actual and Predicted Choice Probabilities by Age for the Estimation Sample: MNL and DP Models

	White			Black			Hispanic		
	Actual	MNL	DP	Actual	MNL	DP	Actual	MNL	DP
Percent Receiving Welfare									
Age 15-17.5	0.9	0.5	1.3	1.9	2.3	4.8	1.3	0.6	4.4
Age 18-21.5	4.3	3.4	4.7	16.9	16.6	15.5	9.2	5.4	11.2
Age 22-25.5	6.4	5.0	7.1	26.9	23.9	24.6	15.0	10.3	15.1
Age 26-29.5	4.7	4.5	7.1	21.6	21.6	28.0	15.2	10.2	15.8
Percent in School									
Age 15-17.5	86.4	81.4	85.3	86.3	82.0	84.2	84.6	84.2	79.2
Age 18-21.5	27.3	28.9	29.8	26.1	25.2	29.5	22.0	29.2	21.4
Age 22-25.5	5.2	5.4	8.3	6.3	6.3	8.0	5.0	5.2	6.0
Age 26-29.5	3.1	2.2	3.5	3.5	2.5	3.5	2.0	2.1	2.8
Percent Working									
Age 15-17.5	35.2	29.7	28.4	19.2	17.6	18.3	22.2	20.1	26.6
Age 18-21.5	66.7	66.3	63.8	44.1	47.9	53.8	52.8	53.0	58.5
Age 22-25.5	72.4	74.9	70.4	56.8	56.0	59.4	58.7	62.2	57.8
Age 26-29.5	71.1	78.7	69.8	61.1	62.1	57.7	56.1	66.8	55.4
Percent Pregnant									
Age 15-17.5	2.5	2.1	1.9	4.6	2.9	3.0	3.2	3.8	3.2
Age 18-21.5	4.4	5.3	4.7	6.7	5.9	6.5	6.9	7.0	6.5
Age 22-25.5	5.5	6.0	5.1	5.8	6.2	7.3	6.7	7.1	7.7
Age 26-29.5	5.5	5.1	4.8	4.2	5.0	6.6	5.9	5.9	6.6
Children Born Before									
Age 20	0.32	0.32	0.31	0.53	0.39	0.47	0.40	0.43	0.48
Age 24	0.72	0.81	0.72	1.05	0.90	1.02	1.00	1.00	1.03
Age 28	1.26	1.24	1.13	1.41	1.20	1.62	1.60	1.49	1.61
Highest Grade Completed									
By Age 24	12.87	13.03	13.08	12.68	12.90	12.97	12.20	12.83	12.38

Table 4 - Actual and Predicted Choice Probabilities for Validation Sample by Age: MNL and DP Models

	White			Black			Hispanic		
	Actual	MNL	DP	Actual	MNL	DP	Actual	MNL	DP
Percent Receiving Welfare									
Age 15-17.5	0.0	0.1	0.1	0.6	0.8	1.3	1.3	0.4	0.5
Age 18-21.5	0.0	0.3	0.7	7.3	7.3	6.4	4.2	3.8	2.3
Age 22-25.5	0.8	0.5	1.6	7.8	9.1	13.0	5.0	4.8	4.9
Age 26-29.5	0.7	0.3	1.9	7.3	8.5	17.7	4.7	4.6	5.9
Percent in School									
Age 15-17.5	93.6	88.5	87.0	87.8	82.0	85.4	80.3	81.0	82.0
Age 18-21.5	36.5	38.4	31.1	27.9	25.2	29.1	29.8	31.4	22.5
Age 22-25.5	6.9	7.7	9.4	3.5	6.3	8.5	4.4	5.7	6.5
Age 26-29.5	4.4	3.7	4.0	1.9	2.5	3.8	4.5	3.4	3.0
Percent Working									
Age 15-17.5	39.3	37.3	38.2	24.7	18.6	24.2	24.1	21.6	33.3
Age 18-21.5	68.9	72.8	75.8	60.5	57.4	64.9	55.0	54.4	64.1
Age 22-25.5	80.0	84.2	82.0	73.1	71.5	70.7	68.1	68.5	64.5
Age 26-29.5	79.6	83.5	82.5	72.8	72.3	69.1	64.9	69.5	63.9
Percent Pregnant									
Age 15-17.5	1.3	2.1	1.7	4.5	2.1	2.9	3.8	4.2	3.3
Age 18-21.5	3.7	5.3	4.8	6.9	4.9	6.7	6.7	6.6	7.1
Age 22-25.5	4.5	6.0	4.9	5.8	5.0	7.4	6.4	6.2	7.5
Age 26-29.5	4.2	5.1	4.8	3.5	3.9	6.6	4.9	5.2	7.0
Children Born Before									
Age 20	0.22	0.18	0.29	0.65	0.58	0.46	0.50	0.50	0.52
Age 24	0.49	0.56	0.68	1.12	0.99	1.03	1.06	1.06	1.11
Age 28	0.86	0.92	1.09	1.71	1.45	1.63	1.54	1.54	1.72
Highest Grade Completed By Age 24	13.27	13.47	13.24	12.81	12.71	13.02	12.21	12.41	12.49

Table 5 - Root Mean Squared Error for Alternative MNL Specifications and for DP Model : Selected Choice Variables

	Estimation Sample					Validation Sample				
	MNL1 FE	MNL1 No FE	MNL2 FE	MNL2 No FE	DP	MNL1 FE	MNL1 No FE	MNL2 FE	MNL2 No FE	DP
Whites										
Welfare	.011	.012	.012	.011	.015	.010	.010	.010	.815	.012
(Mean)			(.043)					(.004)		
Work	.054	.051	.049	.048	.046	.068	.093	.068	.255	.077
(Mean)			(.631)					(.688)		
Pregnancy	.012	.012	.013	.012	.012	.019	.022	.019	.442	.021
(Mean)			(.046)					(.036)		
In School	.045	.044	.045	.047	.027	.046	.086	.045	.138	.054
(Mean)			(.268)					(.315)		
Blacks										
Welfare	.030	.028	.027	.026	.026	.021	.030	.021	.844	.063
(Mean)			(.189)					(.061)		
Work	.035	.030	.034	.032	.064	.059	.054	.058	.215	.065
(Mean)			(.470)					(.600)		
Pregnancy	.015	.015	.016	.016	.021	.034	.037	.033	.490	.036
(Mean)			(.054)					(.052)		
In School	.031	.031	.028	.032	.034	.044	.047	.046	.224	.048
(Mean)			(.269)					(.264)		
Hispanics										
Welfare	.044	.052	.049	.050	.024	.014	.018	.014	.842	.019
(Mean)			(.108)					(.040)		
Work	.067	.071	.059	.064	.048	.050	.062	.048	.169	.092
(Mean)			(.491)					(.550)		
Pregnancy	.015	.015	.015	.015	.019	.022	.025	.022	.487	.030
(Mean)			(.059)					(.056)		
In School	.050	.048	.049	.050	.047	.034	.059	.034	.177	.058
(Mean)			(.246)					(.264)		

Table 6- Counterfactual of Other States with Texas Welfare Benefits: MNL and DP Comparison

	Actual	MNL1 FE Baseline	MNL1 FE With Texas	MNL1 No FE Baseline	MNL1 No FE With Texas	DP Baseline	DP With Texas
Whites							
Percent Receiving Welfare							
Age 15-17.5	0.9	0.5	3.0	0.6	0.2	1.3	0.4
Age 18-21.5	4.3	3.4	19.4	3.6	1.1	4.7	3.0
Age 22-25.5	6.4	5.0	25.9	4.8	1.1	7.1	5.5
Age 26-29.5	4.7	4.5	17.1	4.5	0.7	7.1	5.8
Percent In School							
Age 15-17.5	86.4	81.4	82.6	80.4	78.2	85.3	85.4
Age 18-21.5	27.3	28.9	26.5	27.7	21.1	29.8	29.9
Age 22-25.5	5.2	5.4	4.6	5.3	2.8	8.3	8.3
Age 26-29.5	3.1	2.2	2.0	2.4	1.3	3.5	3.5
Percent Working							
Age 15-17.5	35.2	29.7	15.6	29.8	32.9	28.4	27.8
Age 18-21.5	66.7	66.3	37.0	66.5	77.6	63.8	64.1
Age 22-25.5	72.4	74.9	40.4	74.5	87.1	70.4	71.8
Age 26-29.5	71.1	78.7	48.7	77.9	90.1	69.8	71.1
Percent Pregnant							
Age 15-17.5	2.5	2.1	3.6	2.2	1.5	1.9	1.9
Age 18-21.5	4.4	5.3	15.4	5.4	4.7	4.7	4.8
Age 22-25.5	5.5	6.0	16.9	6.0	5.2	5.1	5.1
Age 26-29.5	5.5	5.1	10.8	5.1	4.6	4.8	4.8
Children Ever Born Before							
Age 20	0.32	0.32	0.67	0.34	0.30	0.31	0.31
Age 24	0.72	0.81	1.85	0.82	0.74	0.72	0.71
Age 28	1.26	1.25	2.76	1.27	1.14	1.13	1.13
Highest Grade Completed							
By Age 25	12.87	13.03	12.93	12.97	12.68	13.08	13.09

Table 6, continued

	Actual	MNL1 FE Baseline	MNL1 FE With Texas	MNL1 No FE Baseline	MNL1 No FE With Texas	DP Baseline	DP With Texas
Blacks							
Percent Receiving Welfare							
Age 15-17.5	1.9	2.3	7.3	2.5	1.1	4.8	3.1
Age 18-21.5	16.9	16.6	42.3	17.5	8.2	15.5	12.2
Age 22-25.5	26.9	23.9	57.9	24.9	9.6	24.6	20.4
Age 26-29.5	21.6	21.6	53.0	22.1	7.1	28.0	24.3
Percent In School							
Age 15-17.5	86.3	82.0	78.8	81.6	80.6	84.2	84.6
Age 18-21.5	26.1	25.2	18.0	25.7	19.5	29.5	29.9
Age 22-25.5	6.3	6.3	3.0	6.6	3.0	8.0	8.2
Age 26-29.5	3.5	2.5	1.0	2.7	1.3	3.5	3.6
Percent Working							
Age 15-17.5	19.2	17.6	9.6	17.6	20.1	18.3	18.1
Age 18-21.5	44.1	47.9	22.5	46.4	62.3	53.8	54.9
Age 22-25.5	56.8	56.0	23.3	55.3	75.1	59.4	62.9
Age 26-29.5	61.1	62.1	27.2	61.6	80.7	57.7	61.6
Percent Pregnant							
Age 15-17.5	4.6	2.9	5.4	2.9	1.5	3.0	3.0
Age 18-21.5	6.7	5.9	21.4	5.9	5.1	6.5	6.5
Age 22-25.5	5.8	6.2	22.5	6.1	5.7	7.3	7.3
Age 26-29.5	4.2	5.0	14.4	5.0	4.9	6.6	6.6
Children Ever Born Before							
Age 20	0.53	0.39	0.96	0.40	0.34	0.47	0.47
Age 24	1.05	0.90	2.52	0.91	0.82	1.02	1.02
Age 28	1.41	1.30	3.90	1.33	1.21	1.62	1.62
Highest Grade Completed							
By Age 25	12.68	12.90	12.56	12.92	12.62	12.97	13.00

Table 6, continued

	Actual	MNL1 FE Baseline	MNL1 FE With Texas	MNL1 No FE Baseline	MNL1 No FE With Texas	DP Baseline	DP With Texas
Hispanics							
Percent Receiving Welfare							
Age 15-17.5	1.3	0.6	8.7	0.6	0.1	4.4	1.7
Age 18-21.5	9.2	5.4	49.0	5.1	0.8	11.2	7.0
Age 22-25.5	15.0	10.3	57.5	8.9	1.2	15.1	10.2
Age 26-29.5	15.2	10.2	34.5	9.1	0.9	15.8	11.6
Percent In School							
Age 15-17.5	84.6	84.2	80.9	84.4	82.6	79.2	79.4
Age 18-21.5	22.0	29.2	20.5	28.8	23.3	21.4	21.6
Age 22-25.5	5.0	5.2	4.2	4.9	2.7	6.0	6.1
Age 26-29.5	2.0	2.1	1.3	2.0	1.2	2.8	2.9
Percent Working							
Age 15-17.5	22.2	20.1	8.7	20.2	24.0	26.6	26.4
Age 18-21.5	52.8	53.0	14.6	54.4	70.9	58.5	59.7
Age 22-25.5	58.7	62.2	15.6	63.8	83.6	57.8	61.2
Age 26-29.5	56.1	66.8	31.9	67.2	86.7	55.4	58.9
Percent Pregnant							
Age 15-17.5	3.2	3.8	7.6	3.8	1.5	3.2	3.1
Age 18-21.5	6.9	7.0	30.0	6.9	5.2	6.5	6.6
Age 22-25.5	6.7	7.1	30.2	7.6	6.1	7.1	7.1
Age 26-29.5	5.9	5.9	17.5	5.9	5.4	6.6	6.6
Children Ever Born Before							
Age 20	0.40	0.43	1.29	0.43	0.30	0.48	0.48
Age 24	1.00	1.00	3.35	0.96	0.78	1.03	1.02
Age 28	1.60	1.49	5.06	1.44	1.23	1.61	1.61
Highest Grade Completed							
By Age 25	12.20	12.80	12.50	12.94	12.53	12.38	12.40