Educational Attainment and the Changing U.S. Wage Structure: Dynamic Implications on Young Individuals' Choices[†]

Moshe Buchinsky[‡]

UCLA, CREST–INSEE and NBER

Phillip Leslie

Stanford University and NBER

July, 2009

Abstract

This study is an analysis of individuals' schooling decisions in an environment that is changing overtime. This is a key issue for a forward looking individual; an issue that has been largely overlooked in the literature. We present a dynamic model of individuals' educational investments that allows us to explore alternative modeling strategies for forecasting future wage distributions. The key innovation we propose is an approach to forecasting that relies only on the information that would be available at the actual time decisions are made, and which incorporates the role of parameter uncertainty into the decision making process. This model is then used to analyze the effects from: (a) actual changes in conditional wage distributions over the years 1980 to 1994; (b) the use of alternative forecasting behaviors; and (c) altering the individual's aversion to risk (or intertemporal substitution). We also examine several other effects affecting the individuals' decision making in that environment. These issues are studied in a variety of simulations using the *March Current Population Surveys*. We find that the forecasting method used by individuals has a significant effect on schooling choices and the predictions of the model, and we provide new insights into the role of risk aversion, rising tuition costs, liquidity constraints, state dependence, and individuals' unobserved heterogeneity.

JEL Classification: J24 and I21

Keywords: returns to education, return to experience, dynamic programming, forecasting wage distributions, Gibbs sampling, human capital.

[†]We thank Dan Ackerberg, Lanier Benkard, Gary Chamberlain, Paul Devereux, Ron Goettler, Joe Hotz, Guido Imbens, Ariel Pakes, John Rust, and Chris Sims for helpful discussions. We also wish to thank an anonymous referee, and the editor, Chris Taber, for extremely helpful comments and suggestions. Financial support to the first author from the NSF under grant SBR-9320386 is acknowledged. The second author is grateful to the Cowles Foundation for financial support.

[‡]Corresponding author, Department of Economics, University of California, Los Angeles, CA 90095-1477. Email: buchinsky@econ.ucla.edu.

1 Introduction

The wage structure of the U.S. economy has changed tremendously since the 1970s.[1] In particular, it is well documented that the returns to both education and experience have dramatically risen. There are several explanations for this phenomena, such as skill-biased technological change and a general shift from manufacturing to service industries. But what have been the consequences of this turbulence for educational investments of individuals during this period? To what degree did these changes in the wage structure drive increases in schooling?

The challenge to analyzing these issues is incorporating a plausible model of how individuals form expectations about the future returns to education and experience—it is unlikely that individuals could foresee these dramatic changes in the 1980s, say. We therefore propose a dynamic programming model of educational choices in which individuals rely only on information that is contemporaneously available in determining their forecasts regarding future wage distributions, and that also accounts for the inherent uncertainty in forecasting due to the reliance on parameter estimates. Expectations of future distributions, and hence forecasting, are key assumptions in dynamic models. It is therefore vital that we examine the sensitivity of predicted behavior to alternative forecasting methods.

Economists have long recognized the importance of expectations to individuals' educational choices.[2] However, it remains a difficult challenge for empiricists because expectations are obviously unobserved, and because we can never be certain how individuals use the available information to assess the future. One approach is to assume individuals know all future conditional wage distributions (but of course do not know the particular wage draws they will obtain from these distributions). This is a convenient assumption that helps impose a degree of internal consistency on the model, and if the wage structure were reasonably stable over time (which it was not) this may actually be a good approximation to reality. Such an approach may be termed rational expectations (although this term can be interpreted more broadly as we discuss below). We are not the first to question the empirical validity of rational expectations—see Manski (2004) for a critique of the empirical relevance of rational expectations (and the references cited therein).[3]

The method we propose incorporates forecasting of future wage distributions based on contemporaneously available information into a dynamic programming framework which naturally emphasizes the importance of forward-looking decision making. A key element of our framework is that we explicitly take account of the inherent uncertainty associated with forecasting. That is to say, the agent in our model is aware that their forecast of future wage distributions

[1] See for example Juhn, Murphy and Pierce (1993), Katz and Murphy (1992), and Buchinsky (1993).

[2] See for example: Belzil and Hansen (1997,1999), Keane and Wolpin (1997), and Taber (2000,2001). Other prominent studies that investigated investment in human capital under uncertainty include Becker (1964), Ben-Porath (1967), Blinder and Weiss (1976), Griliches (1977), Heckman (1993), Heckman, Roselius and Smith (1994), Rosen (1976), and Willis and Rosen (1979). For a variety of interpretations on the theory of human capital see the survey by Willis (1986).

[3] Manski (2004) emphasizes the point: "If experts disagree on the returns to schooling, is it plausible that youth have rational expectations? I think not." (p. 1336).

is an uncertain forecast, which may also impact their optimal choices. This complication is not merely an artifact of our framework, but is conceivably an important issue for individuals' actual decision making. From an agent's point of view, the uncertainty that stems from having to draw wages from a known distribution is augmented by the uncertainty over which distribution they will be drawing from. This framework is potentially useful for other applications in which the analysis of individuals' choices incorporates anticipated future returns, such as: family planning, savings and financial investments, investment in health, investment in children, retirement decisions, immigration and relocation choices, and other specific forms of human capital investment.

We compare the performance of our method with alternative models of forecasting behavior. To implement this we develop a finite horizon dynamic optimization model in which a risk averse agent decides in each period his current consumption level and whether to attend school or to work. As is standard in such models, the valuation of alternative choices is obtained by integrating over uncertain future outcomes, including wages which depend on the individual's level of education and experience. The different forecasting models imply different distributions of expected future outcomes.

Our approach is part estimation and part calibration. The estimated components are as follows. An input to the dynamic optimization problem is the underlying conditional wage distributions. We flexibly estimate these distributions using the semi-parametric technique of quantile regression, in which we account for self-selection using the method of Buchinsky and Hahn (1998). The data come from the *March Current Population Survey* for the years 1964 to 2004. As we explain in Section 2 below, there is no additional estimation if we assume the agents knows all future conditional wage distributions (at any point in time). In the model where agents rely on past information to forecast future wage distributions we estimate a VAR model, as detailed in Section 2. The model also includes a parameter of the utility function (the coefficient of relative risk aversion) and a discount rate. While it is conceptually possible it is highly burdensome to estimate these parameters, and beyond the scope of this paper. Instead, we specify values for these parameters that are within the range of what has been previously obtained in the literature (as described in Section 5).[4]

We also examine three issues of policy-relevance that our preferred model is particularly well-suited to analyzing. First, we consider how individuals with differing degrees of risk aversion make educational choices. We find that increasing the degree of risk aversion leads to lower educational investment. This is because, according to our analysis, education is a more risky investment than experience, at least during the period we study. The second issue we analyze is the importance of financial resources for individuals considering higher education. We assess this by examining the potential impact of changes in initial wealth on educational choices. We find that raising initial wealth from $5,000 to $50,000 for individuals who graduated high school

[4] In our model each period new information arrives, forecasted future wage distributions are updated, and the value function is re-computed by backward recursion. In fact the computational burden is even more severe once we also account for parameter uncertainty in the forecasting rule, as we explain below.

in 1985 leads to an increase in the average accumulated education of about 0.7 years. Higher initial wealth also increases the speed at which education is accumulated. The third issue of policy-relevance we explore is the impact of tuition costs. Incorporating actual tuition costs into the model leads to a sharp drop in the average accumulated education by almost one year.

One concern with the approach we have outlined above is that unobserved heterogeneity or state-dependence in wages may change the qualitative conclusions of our analysis. It is possible that either of these factors may reduce the uncertainty in the returns to education and/or experience, leading us to systematically either over- or under-predict average investment in education. We therefore extend the model to include both, as is explained below. In both of these cases the predictions from the model are qualitatively unchanged, although there are some quantitative differences.

The remainder of the paper is organized as follows. In Section 2 we detail our modeling framework, including the information environment, the forecasting model and the role of parameter uncertainty. The data is described in Section 3, and Section 4 summarizes the results from the estimation of the conditional wage distributions. In Section 5 we examine the results of a large set of simulations designed to contrast the implications of the various forecasting methods. In Section 6 we examine the role of tuition costs, unobserved heterogeneity and state dependence in wages. Concluding remarks are offered in Section 7. An Appendix provides a detailed description of the computation of the posterior distribution of the model's parameters via Gibbs sampling, which is needed for the model in which we incorporate parameter uncertainty.

2 The Framework for Educational Choices under Uncertainty

In this section we present our framework for analyzing educational choices of individuals in a changing economic environment. Our approach incorporates forecasting of future of conditional wage distributions into a dynamic model of individuals' educational choices, in a manner that explicitly accounts for the inherent uncertainty associated with predicting the future. In particular, we incorporate:

1. uncertainty over realizations of future wage offers from *known* future conditional wage distributions; and

2. uncertainty because future conditional wage distributions are *unknown*, stemming from two distinct sources:

   (a) uncertainty in the parameters of the forecasting rule that maps past *known* conditional wage distributions into future *unknown* conditional wage distributions; and

   (b) uncertainty due to the use of *estimated* rather than *known* past conditional wage distributions, which are an input into the forecasting rule for future conditional wage distributions.

Prior research incorporates the first source of uncertainty. One key innovation of this paper is to propose a method that also incorporates the sources of uncertainty in the second category.

Our goal is to incorporate a realistic forecasting model into an analysis of individual's schooling decisions. In doing so we focus on incorporating uncertainty about aggregate parameters (future wage distributions), while presuming certainty over individual parameters (such as parameters of the agent's utility function). Of particular importance in our approach is the information set and its revelation over time, that is relied upon by the economic agent as well as the econometrician. As is explain in detail below, the agent is treated equivalently to the econometrician so that both have the same information at each point in time with more information becoming available to both over time. Another key feature is that our approach is part estimation and part calibration and it is important that we are clear about these components.

2.1 Information Environment and Agents' Decision Problem

The agent in our model makes decisions about consumption, investment in general education and labor supply. The agent's objective is to maximize at each point in time (t) the present discounted value of the expected future utility from a single consumption good: c_t. One approach to this problem is to assume that the underlying distributions are stationary, so that future distributions can be directly estimated from past data. However, there are many cases in which stationarity is unrealistic, especially in periods where the underlying distributions go through dramatic changes, such as the changes we observed in the wage structure in the U.S. over the past few decades.

We do not assume stationarity in our approach. Rather, we assume that the agent has access to estimates of past wage distributions from which they forecast future distributions. These forecasts may be subject to errors of various kinds, including errors in the forecasting model and errors in the estimates of past wage distributions. We make no explicit assumption about the true data generating process (DGP) for the evolution of actual conditional wage distributions, which presumably depend on many factors (including macroeconomic variables). The agent and the econometrician are agnostic about this aspect of the underlying DGP. To be clear, however, our characterization of agents use of past data to estimate past wage distributions, which is used as an input for computing estimates of future conditional wage distributions, should be interpreted as the true DGP governing the agent's decision process.

It is also worth noting that the individual does not know the aggregate wage distributions and thus needs to estimate these, but he/she knows his own wage at any point in time. Also, the individual has no uncertainty about the parameters that correspond to his/her own preferences. The individual knows his/her discount factor, the parameter the correspond to his/utility function, etc.

Assume that individuals have access to data from which they estimate parameters of past wage distributions. Denote the *true* past conditional distribution at year t by $f(w|s;\beta_t)$, where s is a vector of observable attributes (for both the individual agent and the econometrician) at year t, and β_t is the vector of true parameters that correspond to year t. Hence, given β_t the conditional distribution of wage $f(w|s;\beta_t)$ is known. Given this structure, let β_h^t denote the set of past true vectors of parameters that correspond to the wage distributions from year $t-h$ to the current year t:

$$\beta_h^t = (\beta_{t-h}, \beta_{t-h+1}, ..., \beta_t). \tag{1}$$

Importantly, as mentioned above, neither the agent or the econometrician know the values in β_h^t, and thus need to estimate them from data available for past years. Let $\hat{\beta}_t$ denotes the estimate for the year t parameter vector β_t, and let $\hat{\beta}_h^t$ denote the estimate for β_h^t. The distinction between *true* parameters and *estimated* parameters is crucial, because it implies the agent is uncertain about past conditional wage distributions, which are the key ingredient to forecasting future wage distributions. Accounting for this *parameter uncertainty* into the forecasting method of future distributions is a central feature of our framework.

We assume that the agent estimates past wage distributions using quantile regression. Specifically, the agent obtains $\hat{\beta}_{t-h}$ from running a sequence of quantile regressions for each of the years $t-h$ through t. The entire information obtained from this procedure for year t is contained in $\hat{\beta}_h^t$. Below we detail the structure and estimation of $\hat{\beta}_h^t$, and for the remaining explanation in this section we can assume that the individuals are provided with $\hat{\beta}_h^t$.

To forecast future wage distributions we assume the agent employs a vector auto-regressive (VAR) model. Let θ_{VAR} denote the parameter vector containing all true values for the VAR model's parameter, and let $\hat{\theta}_{VAR,t}$ denote the estimate for θ_{VAR}, obtained at time t (using all the available information from $t-h$ to t). The VAR model provides the mapping from $\hat{\beta}_h^t$ to forecasted future values of β_m (for $m = t+1, t+2, ..., T$). Similar to β_h^t, the fact that the agent does not know the true VAR model's parameter but only has estimates of these parameters, implies an additional sources of parameter uncertainty that one also needs to account for when forecasting future distributions.

There is an important distinction between past and future wage distributions at any point in time. Consider an agent at time t. The agents has estimates of past wage distributions, which are based on data, denoted by $\hat{\beta}_h^t$. The agent inputs these estimates into the estimated VAR model to generate forecasts of future wage distributions, that we denote by $\hat{\hat{\beta}}_{t+1}, ..., \hat{\hat{\beta}}_T$. Note, in order to clearly distinguish between past and future, we generically denote the former by $\hat{\beta}$, and the latter by $\hat{\hat{\beta}}$. Formally, predicted future parameter vectors are obtained recursively via:

$$\hat{\hat{\beta}}_m = h\left(\hat{\hat{\beta}}_{m-1}, \hat{\theta}_{VAR,t}, \hat{\beta}_h^t\right),$$

where $h(\cdot)$ denotes the function that is implied by the VAR model (the exact definition is provided below). It is important to note that as the agent (and the econometrician) moves forward in time more information regarding past wage distributions becomes available (i.e., the data for the most recent year). Hence, in order to incorporate all available information at any given time the individual re-estimates the VAR model's parameters every period.

Lastly, a point of clarification about the DGP we have outlined in this section. In principle there can be a distinction between: (i) the true DGP; (ii) what the agent uses to approximate the true DGP; and (iii) what the econometrician uses to approximate the true DGP.[5] It should be clear from the above description that in our approach (ii) and (iii) are the same. Furthermore, we assume the above model is a complete description of the true DGP for how individuals make decisions. But we are agnostic about the underlying DGP of the evolution of conditional wage distributions—the VAR model is not intended as a true description of the DGP for the evolution of conditional wage distributions over time, but rather a flexible-form model that approximate the historical evolution, and which, we assume, individuals utilize for forecasting future distributions. In other words, we assume that the VAR model is the true model used by individuals in making education choices, but this is different to assuming that the VAR model is the true model of how conditional wage distributions are evolving over time.[6]

2.2 The Behavioral Model

As outlined above, the agent is forward looking and relies on contemporaneously available information to forecast the necessary future wage distributions. This forecasting behavior is embedded in a life-cycle model in which the objective of the agent is to maximize their life-time discounted expected utility from a single consumption good, c. Conditional on the information available to the individual at each time t, the agent chooses their optimal level of consumption and whether to attend school or to work. Education choices have an immediate effect on utility via the current period budget constraint, i.e., foregone wage. Education choices also have an inter-temporal effect by altering the agents attributes (level of education and labor market experience) which determine future wage distributions. The immediate and inter-temporal effects imply a trade-off the agent must decide upon.

At time t, given the state vector z_t (defined below), the individual solves the problem given by

$$\begin{aligned}
&\max_{\{c_\tau,e_\tau\}_{\tau=t}^{T}} \quad \sum_{\tau=t}^{T} \delta^{\tau-t} E\left[u(c_\tau) \mid z_t\right] \\
&\text{subject to:} \\
&B_\tau = (1+r_\tau)\left[B_{\tau-1} + (1-e_{\tau-1})w^P_{\tau-1} - c_{\tau-1}\right]; \\
&c_\tau \in \left[0, B_\tau + (1-e_\tau)w^P_\tau\right]; \quad \text{and} \\
&S_\tau \leq \overline{S},
\end{aligned} \tag{2}$$

where $u(\cdot)$ is a per-period concave utility function. The expectation is taken with respect to all sources of uncertainty associated with future wage distributions, including uncertainty because the agent faces a wage distributions (rather than a sequence of known wages); and uncertainty

[5] We are grateful to the editor for proposing this classification.

[6] One interpretation is of the distinction made at the start of this paragraph is that if (i), (ii) and (iii) are all the same, then the model can be said to include rational expectations. Since we do not provide a description of the true DGP for the evolution of future wages, we would argue that (i) is not the same as (ii) and (iii). Therefore, the model does not include rational expectations. But that may be merely a semantic issue.

because future wage distributions are unknown. It is important to note that the agent has no uncertainty regarding their utility function. That is, the agent knows all the parameters associated with his utility function, including his discount factor and taste for risk.

The variables c_τ and e_τ are the agent's choices of consumption and education, respectively, at time τ ($e_\tau = 1$ if the individual chooses to attend school and $e_\tau = 0$ otherwise). The term B_τ denotes the individual's level of wealth at the beginning of period τ, r_τ denotes the interest rate paid on wealth held from time $\tau - 1$, and δ is the individual's discount factor.[7] The quantity w_τ^P is the wage offer (or potential wage) at time τ that will be realized only if the agent chooses to work. Consequently, the individual's observed wage is given by $w_\tau = (1 - e_\tau)w_\tau^P$.

Lastly, S_τ denotes the level of education acquired by the beginning of period τ (i.e., $S_\tau = \sum_{j=0}^{\tau-1} e_j$). We impose an upper limit ($\overline{S}$) on the level of education an individual can acquire. The cost of education in the base model is the opportunity cost of foregone income. In the results section below we also incorporate the direct cost of education (tuition fees).[8] Note also that in this model individual is not allowed to borrow against future earnings. Consequently, the level of initial wealth is likely to play a major role. We investigate below the extent of this borrowing constraint on individuals' educational attainments.

The agent's wage offer at time t is a draw from the conditional wage distribution, conditional on their vector of characteristics s_{it}:

$$w_{it}^P \sim f\left(\cdot | s_{it}; \beta_t\right).$$

We assume that $s_{it} = (S_{it}, X_{it})$, where X_{it} denotes the individual's labor market experience at time t. In principle, we can also account for an unobservable individual-specific characteristic. In the results section below we provide results from simulations that incorporate unobserved heterogeneity.

The vector of state variables is given by

$$z_{it} = \left(S_{it}, X_{it}, B_{it}, w_{it}^P, \hat{\beta}_h^t\right). \tag{3}$$

Note that z_{it} does not include the quantile regression, or the VAR model, parameter estimates since all these can be derived from $\hat{\beta}_h^t$, as explained below.

The maximization problem in (2) involves optimally choosing the consumption and education paths for all years $\tau = t, \ldots, T$. The maximal expected value of the discounted life-time

[7] Here, we do not estimate δ, but rather calibrate it as is explained below.

[8] In principle one could also include additional pecuniary, as well as non-pecuniary costs, which depend upon observable and unobservable characteristics of the individual. For example, there may be a non-pecuniary cost associated with returning to school, or costs which depend on the unobservable health status of the individual and/or his family.

utility from consumption at time t is given by

$$V(z_t, t) = \max_{\{c_\tau, e_\tau\}_{\tau=t}^{T}} \left\{ \mathrm{E}\left[\sum_{\tau=t}^{T} \delta^{\tau-t} u(c_\tau) \middle| z_t \right] \right\},$$

where, due to the finite horizon nature of the problem, the value function depends not only on z_t, but also on the time period t. Note, the agent in our model decides whether they will work or attend school, but not the number of hours supplied to the labor force. Another implication of the model is that if the agent has achieved $\overline{S}$ years of schooling then they must work, and the only decision is about the level of consumption.

It is convenient to write the associated Bellman's equation in two parts:
Value from attending school

$$\begin{aligned} &V^S(z_t, t) \equiv V(z_t, t \mid e_t = 1) = \\ &\sup_{c \in \Gamma(z_t | e_t = 1)} \left\{ u(c_t) + \delta \mathrm{E}\left[\max\left\{ V^S(z_{t+1}, t+1),\ V^L(z_{t+1}, t+1) \right\} \middle| z_t, e_t = 1 \right] \right\}. \end{aligned} \tag{4}$$

Value from working

$$\begin{aligned} &V^L(z_t, t) \equiv V(z_t, t \mid e_t = 0) = \\ &\sup_{c \in \Gamma(z_t | e_t = 0)} \left\{ u(c_t) + \delta \mathrm{E}\left[\max\left\{ V^S(z_{t+1}, t+1),\ V^L(z_{t+1}, t+1) \right\} \middle| z_t, e_t = 0 \right] \right\}, \end{aligned} \tag{5}$$

where the correspondence $\Gamma(z_t)$ defines the feasible set of choices at time t.

The corresponding value function at time t is simply

$$V(z_t, t) = \max\left\{ V^S(z_t, t),\ V^L(z_t, t) \right\}, \tag{6}$$

for $t = 0, \ldots, T-1$. At $t = T$ (final period of the model) we let

$$V(z_T, T) = u\left(B_T + w_T\right), \tag{7}$$

That is, in the final period the agent consumes all remaining wealth.

2.3 Reservation Wage Property

The agent chooses to attend school only if $V^L(z_t, t) > V^S(z_t, t)$. Under mild assumptions this implies a reservation wage property. That is, there exists a reservation wage $\varphi\left(z_{it}\right)$ such that, for all $w_t^P \geq \varphi\left(z_{it}\right)$, the agent chooses to work.[9] We now briefly discuss the conditions the imply the reservation wage property and their implications. This is provided for illustration purposes assuming that the individual wage follows an $AR\left(1\right)$ process, but it has no implication

[9] In practice the computation of the value functions associated with the choice of education and work do not make use of the reservation wage property.

about the way we estimate the population wage distribution as is explained below.

As stated above, we assume $u(\cdot)$ is a concave utility function, so that $u'(\cdot) > 0$. We also assume that the conditional wage distributions exhibit first order stochastic dominance in education, S, and experience X.[10] Furthermore, we assume that the (log) wage equation takes the canonical form

$$\log w = \gamma_{0t} + \gamma_{1t} S + \gamma_{2t} X + \gamma'_{3tt} M_t + \xi_t, \quad \text{with}$$
$$\xi_t = \begin{cases} \rho \xi_{t-1} + \xi_t & \text{if } e_{t-1} = 0, \\ \xi_t & \text{if } e_{t-1} = 1. \end{cases},$$
$$\rho > 0, \quad \text{and}$$
$$\xi_t \sim i.i.d.\left(0, \sigma_\xi^2\right),$$

where the vector M_t consists of relevant characteristics.[11] Note that the parameters γ_j ($j = 0, 1, 2, 3$) are allowed to change over time, but are common to all individuals at any given time t. There is strong support in the literature for the assumption that $\rho > 0$ (see Abowd and Card, 1989, and Gottchalk and Moffitt, 2002).

Under the assumptions stated above it follows that the value function $V(z_t, t)$ is continuous, bounded, increasing, and strictly concave (e.g., Atakan, 2003, Corollary 1, p. 453). It follows that $V^L(z_t, t)$ is strictly increasing and concave in ξ_{t-1}, while $V^S(z_t, t)$ is non-increasing in ξ_{t-1}, so that

$$\frac{\partial \left[V^L(z_t, t) - V^S(z_t, t)\right]}{\partial \xi_{t-1}} > 0.$$

The implication of this result is that there is a reservation wage $w_0 \equiv \varphi(z_{it})$, such that for any draw $w_1 > w_0$ the agent will choose to work instead of a year of schooling. By construction, this reservation wage is a function of all the state variables. It is not necessary to know the exact form of the function $\varphi(\cdot)$ to use the censored quantile regression estimation method we describe below.

The conditional wage distribution can be identified by first estimating the probability that an individual will participate in the labor market, and then plugging the estimated probability into the objective function of the censored quantile regression. That is, the probability and the regression depend on the same variables. However, for efficiency reasons we also include additional variables, which are not included in z_t (as described below).

[10] This simply means that the distribution of w_t^p conditional on E and X first order stochastically dominates the distribution of w_t^p conditional on E', X' for any $E \geq E'$ and $X \geq X'$, with strict inequality for at least one of the two variables. Also, it is worth noting that, literally speaking, this assumption violates Mincer's specification in which the value of experience falls toward the end of the life-cycle.

[11] This form of the wage equation is consistent with the estimation of the conditional wage distribution described below. Nevertheless, we make no attempt to recover the "structural" parameters, i.e., γ's, ρ, and σ_ξ, but rather to obtain a parsimonious characterization of the conditional wage distribution using quantile regression.

2.4 The Role of Parameter Uncertainty

Parameter uncertainty plays a major role in our analysis and this section we provide a more formal description of how it is incorporated into the analysis. Recall that in order to compute the value functions in (4) and (5) it is necessary to calculate $E\left[V(z_{t+1}, t+1)|\, e_t, z_t\right]$, where $V(z_{t+1}, t+1)$ is defined in (6). This expectation is taken with respect to all random variables: w_{t+1}^P as well as $\hat{\hat{\beta}}_{t+1}$:

$$\begin{aligned} E[V(z_{t+1}, t+1)|e_t, z_t] &= \int\int V(t+1, z_{t+1}) dP\left(w_{t+1}^P, \hat{\hat{\beta}}_{t+1} \mid e_t, z_t\right), \\ &= \int\left[\int V(t+1, z_{t+1}) dP\left(w_{t+1}^P|\hat{\hat{\beta}}_{t+1}, e_t, z_t\right)\right] dP\left(\hat{\hat{\beta}}_{t+1}|e_t, z_t\right), \\ &= \int\left[\int V(t+1, z_{t+1}) dP\left(w_{t+1}^P|\hat{\hat{\beta}}_{t+1}, e_t, z_t\right)\right] dP\left(\hat{\hat{\beta}}_{t+1}|z_t\right), \end{aligned} \tag{8}$$

where the last equality stems from the fact that

$$\Pr\left(\hat{\hat{\beta}}_{t+1}|e_t, z_t\right) = \Pr\left(\hat{\hat{\beta}}_{t+1}|z_t\right),$$

i.e., the distribution of future wage distributions' parameters does not depend on any specific individual's choices.

The most common practice in the literature is to treat $\hat{\hat{\beta}}_{t+1}$ as if it were the true parameter vector β_{t+1}, and instead of computing (8), to compute:

$$E_{\beta_{t+1}}\left[V(z_{t+1}, t+1)|\, e_t, z_t\right] = \int V(z_{t+1}, t+1) dP(w_{t+1}^P|\hat{\hat{\beta}}_{t+1}, e_t, z_t), \tag{9}$$

where $E_{\beta_{t+1}}(\cdot)$ denotes that the expectation is taken holding $\hat{\hat{\beta}}_{t+1}$ constant. How important it is to incorporate parameter uncertainty is an empirical question, and in the analysis below we attempt to shed light on this aspect.

2.5 Forecasting Future Distributions

2.5.1 Estimating β_t from past wage data using quantile regression

There are many possible ways of estimating conditional wage distributions using parametric, non-parametric, and semi-parametric approaches. We prefer to use quantile regression for several reasons. First, quantile regression provides a parsimonious characterization of all conditional wage distributions that is particularly convenient for implementation in the dynamic choice model due to the discretization of wages. Second, it is a semi-parametric method that assumes only certain conditional quantile restrictions.[12] Third, there are censoring problems

[12] One can certainly estimate conditional distributions non-parametrically, but forecasting these distributions into the future becomes intractable given the framework we adopted here.

with the CPS data (described in Section 3) used in this study that the quantile regression method is well-suited to deal with.

We assume that the wage distribution can be approximated arbitrarily closely by a sequence of quantile regressions. In this study we discretize the distributions into five quantiles, assigning each of these quantiles, located at the .1, .3, .5, .7, and .9 quantiles, a probability mass of .2.[13] We assume that wages depend on education and experience, and we allow for a flexible functional form for the conditional quantiles (by including polynomials in education and experience as regressors).

The qth conditional quantile of earnings, conditional on s_{it}, is given by $\text{Quant}_q(w_{it}^P|s_{it}) = s_{it}'\beta_{t,q}$, where $0 < q < 1$, and $\beta_{t,q}$ is a $K \times 1$ parameter vector. The corresponding conditional quantiles are

$$s'\beta_{t,q}, \quad \text{for } q = .1, .3, .5, .7, .9,$$

and for all years for which we have data.

We denote the $5K \times 1$ stacked vector of the population parameters for year t by

$$\beta_t = \left(\beta_{t,.1}', \beta_{t,.3}', \beta_{t,.5}', \beta_{t,.7}', \beta_{t,.9}'\right)', \tag{10}$$

and denote its estimate by $\hat{\beta}_t$. Note that β_t and $\hat{\beta}_t$ correspond to β_t and $\hat{\beta}_t$ defined generically in Section 2.1. For convenience we denote B_{β_t} as the matrix version of β_t, and let $\hat{B}_{\beta_t}$ denote the estimate for B_{β_t}, that is:

$$\hat{B}_{\beta_t} = (\hat{\beta}_{t,.1}, \ldots, \hat{\beta}_{t,.9}),$$

where $\hat{\beta}_{t,q}$ is the quantile regression estimate for $\beta_{t,q}$, for $q = .1, .3, .5, .7, .9$. Then the estimated conditional distribution of earnings, conditional on any vector of characteristics, say s_0, is provided by

$$\left\{s_0'\hat{\beta}_{t,.1}, s_0'\hat{\beta}_{t,.3}, s_0'\hat{\beta}_{t,.5}, s_0'\hat{\beta}_{t,.7}, s_0'\hat{\beta}_{t,.9}\right\} = s_0'\hat{B}_{\beta_t},$$

where each point of support has a probability mass of .2. As we explained above, the individual is presumed to undertake this procedure using data from years up to the current decision year to obtain a sequence of estimates: $\hat{\beta}_{t-h}, ..., \hat{\beta}_t$ for $\beta_{t-h}, ..., \beta_t$. The individual then uses these estimates to forecast future β_t's. These forecasts allow one to characterize all future conditional distributions that are necessary for computing the expectation in (9). However, to be able to compute the expectation in (8) we must also obtain the joint distribution of all parameter estimates. Before explaining how to obtain this joint distribution, we first explain how to use $\hat{\beta}_{t-h}^t$ to predict $\hat{\hat{\beta}}_m$, for $m \geq t$.

[13] Discretizing continuous distribution by a finite points of support is a common practice in the literature, especially in studies that use dynamic programing models. This is essentially what we do here: each conditional distribution is being discretized.

2.5.2 Forecasting future distributions via a VAR model

We assume agents uses the widely accepted vector autoregression (VAR) model to form forecasts about future distributions. Formally:[14]

$$\beta_{\tau+1} = a + A\beta_\tau + \nu_\tau, \quad \text{for } \tau = 0, \ldots, t, \tag{11}$$

where $\nu_\tau \sim \text{i.i.d.}N\left(0, \Sigma_\nu\right)$. Recall that β_τ is a $5K \times 1$ vector, so a is a $5K \times 1$ parameter vector and A is a $5K \times 5K$ matrix of parameters. For convenience we define $C'_\beta = (a, A)$ and the stacked vector of the columns of C_β by $c_\beta = \text{vec}(C_\beta)$. As discussed above, we assume this is the true DGP for how individuals forecast future conditional wage distributions. However, we do not assume this is the true DGP for the actual evolution of conditional wage distributions (rather as a flexible-form model that provides individuals with a good approximation for the evolution of β_t).

If $\beta_{t-h}, \ldots, \beta_t$ were known, it would be straightforward to obtain an estimate for C_β, and consequent forecasts for all the necessary β_t's into the future. The problem the agent (and the econometrician) face is that we only observe $\hat{\beta}_{t-h}, \ldots, \hat{\beta}_t$. Ignoring that these are only estimates of the true parameters would yield inconsistent estimates for C_β. However, note that $\hat{\beta}_t$ may be expressed as

$$\hat{\beta}_\tau = \beta_\tau + \epsilon_\tau \quad \text{for } \tau = t-h, \ldots, t. \tag{12}$$

That is, the estimate $\hat{\beta}_\tau$ is decomposed into the true parameter β_τ and and an estimation error ϵ_τ. We further assume, as is commonly done in the literature, that $\epsilon_\tau \sim \text{i.i.d.}N\left(0, \Sigma_\epsilon\right)$, and that ϵ_τ and ν_τ are uncorrelated with each other in every period. Equations (11) and (12) provide a special case of the Kalman filter framework, whereby (11) is the *state variable* (or transition) equation, while (12) is the *measurement* equation.[15] For convenience let θ_{VAR} denote the vector containing all the VAR model's parameters: a, A, Σ_ν, and Σ_ϵ. Having specified the distributions for both ϵ_τ and ν_τ, we use the well-known Kalman filter framework for estimating C_β.[16]

The agent uses all available data up to the year in which he makes the decisions, say t. In period $t+1$ additional data become available. From that data the individuals can estimate $\hat{\beta}_{t+1}$. The agent then re-estimates the VAR model using the additional estimates. If there are no dramatic changes in $\hat{\beta}_t$, as was the case in the U.S. during most of the 1970s, there may be minor changes in the estimate $\hat{C}_\beta$. However, for more dramatic changes in the observed wage structure, such as those that happened during the 1980s, there may be considerable change in

[14] We also tried an alternative specification in which we considered a VAR model, but in differences. That is, estimating a VAR model for $\Delta\beta_t \equiv \beta_t - \beta_{t-1}$. This model performed badly in terms of within-sample predictions as well as of out-of-sample predictions. Another alternative we considered was a simple AR(1) model fitted to each coefficient separately. However, the problem with this model is that it does not take into account the dependence between the various coefficients over time, and especially the high correlations between the coefficients of the same variable across the various quantiles.

[15] For details see, for example, Harvey (1991, pp. 82–105).

[16] The derivation of the maximum likelihood estimate for C_β has been described and discussed in num,erous text books. For a complete detailed dicussion of the Kalman filter and the estimation see, for example, Chapter 13 of Hamilton (1994).

the estimate $\hat{C}_\beta$ as the agent moves forward one year at a time.

In principle (with enough data) the model outlined above may be fully estimated. Since the CPS is an annual survey which began in 1964, we have a limited amount of data and we therefore impose some further simplifications for the model's parameters. Specifically, we allow each parameter in each of the five quantile parameter vectors to depend only on past values of that particular parameter and the value of that particular parameter for the other four quantile vectors, that is:

$$\beta_{\tau,q,k} = \phi_{qk} + \sum_p \delta_{q,k}\beta_{\tau-1,p,k} + \zeta_{qk},$$

where $k = 1, \ldots, K$; $q = .1, .3, .5, .7, .9$; $p = .1, .3, .5, .7, .9$; and where the ϕ's and δ's are the unrestricted parameters (in C_β) to be estimated. The covariance matrices Σ_ν and Σ_ϵ are restricted to be block diagonal matrices of the form $\Sigma_\nu = I_5 \otimes \Sigma_{\nu 0}$ and $\Sigma_\epsilon = I_5 \otimes \Sigma_{\epsilon 0}$, where $\Sigma_{\epsilon 0}$ and $\Sigma_{\nu 0}$ are $K \times K$ symmetric matrices.

Note that the model in (11) and (12) implies that a forecast for future β_t's is provided by

$$\hat{\hat{\beta}}_m = \hat{a} + \hat{A}\hat{\hat{\beta}}_{m-1} + \left(\epsilon_m - \hat{A}\epsilon_{m-1} + \nu_{m-1}\right), \tag{13}$$

for $m > t$, where $\hat{a}$ and $\hat{A}$ are the estimates for a and A, respectively. Since the best predictions for ϵ_m and ν_m are zeros, it follows that best prediction for β_m is simply

$$\hat{\hat{\beta}}_m = \hat{a} + \hat{A}\hat{\hat{\beta}}_{m-1}.$$

In order to account for the prediction error in (13) one needs to integrate over the distribution of $\epsilon_m - \hat{A}\epsilon_{m-1} + \nu_{m-1}$. Since each of these random vectors is $5K \times 1$ the computation of the inner integral in (8) is infeasible. Hence, in the computation of the value function we do not integrate over this joint distribution. Nevertheless, what we do integrate over provides a significant advancement over what has been previously done in the literature, because we do treat (in the VAR-Gibbs forecasting method) $\hat{a}$ and $\hat{A}$ as random variables that contribute to the parameter uncertainty the individuals are faced with.[17]

The process of re-estimating the VAR model each model each period given the new information can be characterized as a learning process. That is, in each period t the agent has a prior belief over the joint distribution of the models parameters, then examines the information that becomes available at this time ($\hat{\beta}_h^t$) and modifies this belief to yield a posterior distribution of those same parameters. In the appendix we specify the prior beliefs and explain how to obtain the posterior joint distribution for all of the model's parameter, i.e., $\left(\hat{\hat{\beta}}_{t+1}, ..., \hat{\hat{\beta}}_T, C_\beta, \Sigma_\nu, \Sigma_\beta | \hat{\beta}_h^t\right)$, using a Gibbs sampler. This posterior distribution is then used in (8) for computing the current period value function.

[17] It is worthwhile noting that the estimated variation of bot $\epsilon_t$and ν_t were found to be relatively small in comparison with the other random elements in the model. This strongly suggests that the problem we do solve closely approximates the one in which one integrates over the distribution of $\epsilon_t + \widehat{A}\epsilon_{t-1} + \nu_{t-1}$.

2.6 The Sequence of Events

In the above sections we have described a framework that includes several components: a model of behavior, estimation of past distributions' parameters, and a forecasting procedure. We now clarify the exact sequence of events within the framework:

1. The agent arrives at the beginning of period t with the following components of the state vector: S_t, X_t, and B_t.

2. The agent is presented with a wage offer w_{it}^P, and the additional wage data for year t.

3. The agent estimates the vector of parameters pertaining to the current year (i.e., $\hat{\beta}_t$). The relevant information for all past distribution is now given by $\hat{\beta}_h^t = \left(\hat{\beta}_{t-h}, \hat{\beta}_{t-h+1}, ..., \hat{\beta}_t\right)$.

4. The agent estimates the VAR model parameters, using the "data" provided in $\hat{\beta}_h^t$, by applying the Kalman filter framework.

5. The agent computes forecasts for all parameter vectors of future distributions (i.e., $\hat{\hat{\beta}}_{t+1}, ..., \hat{\hat{\beta}}_T$).

6. Given $\hat{\hat{\beta}}_{t+1}, ..., \hat{\hat{\beta}}_T$, the agent computes their value function and maximize it with respect to consumption (c_t) and education (e_t).

The individual agent and the econometrician face identical information and sequence of decisions, with two crucial differences. First, while the econometrician observes the agent's choice of consumption and education, the econometrician observes the wage offer only if it is accepted. Second, the econometrician has no direct observation about the preference parameters. In principle, these parameters may be estimated. The current study is partially estimation partially calibration, so that we assume particular values for the individual preference parameters and estimate all the parameters associated with the underlying wage distributions.

3 The Data

The empirical analysis in this study is based on extracts from the March-CPS for the years 1964 through 2004. We use the data for the years 1964 through 1994 for estimation, and the data from 1995 through 2004 for out-of-sample validation. To be consistent with the model specification, we use data only on males between the ages of 14 and 65 who were either working or in school. Since most of the relevant variables in the March-CPS refer to the year preceding the sample year, the dates in this study are the actual years to which the questions referred. Thus, for example, CPS year 1964 is referred to as 1963. We regard an individual to have been in school in a particular year if at least one of the following conditions were true: (a) the reason for not working was being at school; (b) the major activity last year was going to school; and/or (c) the individual worked less than 13 weeks and in the remainder of the year was at school.

For individuals' earnings we use the total earnings from wages and salaries. This variable, as with all nominal variables, was deflated by the implicit price deflator for personal consumption expenditure from gross domestic income.[18]

The basic statistics of the data we used in the estimation (i.e., 1964 trough 1994) are reported in Table 1. This table suggests that, for at least the younger age groups, there exists cyclical behavior in school enrollments. For the 19–24 years old, the group we are most interested in, the percentage of respondents attending school exhibits distinct peaks in the years 1976, 1984 and 1992; a pattern which is similarly shared by the 14–18 age group. Also in the case of the 19–24 years old group, the minimum percentage of those attending school for the entire sample occurs in 1980, 15.2%, while the maximum occurs in 1984, 20.7%, highlighting the fact that the first half of the 1980s was a period of significant change in individuals' behavior. Lastly, notice the percentage of those attending school for all age groups never rises above 1% before 1981, and then after 1981 the percentage only drops below 1% in two of the 14 sample years. Overall, this percentage has more than tripled from 1964 to 1994.

Two major problems hamper use of the pre-1976 March-CPS surveys. First, massive changes in sampling frame, coverage and imputation methods were introduced in 1976. Second, some of the most important variables for our analysis were only bracketed. To deal with the latter problem we adopt the imputation method suggested in Buchinsky (1994). In this procedure the bracketed variables are imputed using the frequencies in the CPS years 1976, 1977 and 1978, conditional on race, age and education.[19]

The problems which stem from the changes in the sampling frame and coverage are harder to deal with. Nevertheless, as shown in Juhn, Murphy and Pierce (1993), these changes affect mainly the top and bottom ten percent of the wage distribution; which remains true after controlling for education and experience. Since the method employed here is quantile regression, the results from Juhn, Murphy and Pierce (1993) suggest the above changes are not likely to introduce any systematic bias to our analysis. Indeed, as is shown below we do not observe any discrete jump in the extreme quantiles (.10 and .90) relative to the return at the median.

4 The Conditional Wage Distribution

In this section we elaborate on the estimation of the quantile estimates, $\hat{\beta}_1, \cdots, \hat{\beta}_t$, for the years 1964 through 1994 and discuss the main features of the results. Using the March-CPS data our estimates are based only on the sub-population of people who choose to work in any given year. To overcome this sample selection problem we resort to the method proposed by Buchinsky and Hahn (1998), in which each individual has a threshold wage offer below which the individual will prefer to acquire an additional year of education, and above which the individual will choose to

[18] This deflator is taken from *The Economic Report of the President*, 2005.

[19] For example, the number of weeks worked by an individual is imputed, conditional on their sex-race-age-education composition, so that the distribution of weeks worked will match the average frequency for 1976 through 1978. A detailed description of the imputation method is provided in an appendix in Buchinsky (1994).

work. Furthermore, as is the case here, the threshold wage is an unknown function of a known set of regressors given by $\varphi(z_t^*)$. The vector z_t^* includes all the variable included in the wage equation, as well as several variables that are excluded from the wage equation: individual and family non-earned income, marital status dummy variable, and a dummy variable for children.[20]

The probability of being censored, $\pi(z_t^*)$ in Buchinsky and Hahn's notation, is estimated by the Nadaraya-Watson kernel with a normal kernel function and a least-squares cross-validated bandwidth. For the quantile regressions themselves only a subset of those variables are included: education, education squared, experience, experience squared, an interaction term between experience and education, race dummy variable and race dummy variable interacted with both education and experience.

Table 2 presents some representative results for the returns to education for mid-career (15 years of experience) white males. The returns to education are simply the derivatives of the conditional quantiles with respect to education when evaluated at the level of education denoted in the table. Figure 1 presents the returns to education for two education groups of white males at the five quantiles. Table 2 and Figure 1 present only a small fraction of the results. Nevertheless, they reveal the most common features. They clearly demonstrate the enormous changes in the returns to education over the years at all five quantiles as well as some noticeable differences across the various quantiles.

Several results stand out. After general decreases in the returns to education through most of the 1970s, there was a sharp and unprecedented increase during most of the 1980s, at all quantiles. The increase, however, was not uniform, in absolute terms, for all quantiles. Greater increases at lower quantiles led to the returns at the various quantiles to converge for most groups. Furthermore, in the early 1970s the returns for the less educated groups tend to be higher at the lower quantiles, while the order reverses itself for the more educated. Similarly dramatic changes are also apparent for the returns to experience.

In the simulations below, we pay attention to the model's predictions, under various forecasting rules and parameter values, for accumulated education of particular cohorts, and to the fraction of individuals in these cohorts in school at any particular time, among other things.[21] Based on the CPS, we show the actual time path of average accumulated education in Figure 2a, and the fraction in school over time in Figure 2b. These are intended only to be rough guides to help in evaluating the simulations. The main features of these figures will be discussed in

[20] Non-earned income includes all sources of income other than work related transfers or direct income from wages and salaries. The dummy variable for race takes the value 1 if the individual is non-white, and 0 otherwise. The dummy variable for metropolitan area takes the value 1 if an individual lives in the CPS standard statistical metropolitan area, and 0 otherwise. Education is defined as the number of completed years of schooling. Potential labor experience is defined as: $\min(\text{age} - \text{education} - 6, \text{age} - 18)$. The division dummy variables are for the CPS divisions: New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific. Each of these variables takes the value 1 if an individual lives in that division, and 0 otherwise. The marital status dummy variable takes the value 1 if an individual is married, and 0 otherwise. The children dummy variable take the value 1 if an individual has children, and 0 otherwise.

[21] The cohorts are defined as individuals who graduated from high school in 1979, 1984 and 1989.

relation to the simulations below.

5 Model Computation and Policy Simulations

The numerical computation of the value function in any period t is via backward recursion based upon the observed information up to period t. From the Gibbs sampling algorithm the agent obtains multiple forecasts, each of which is applied in every backward recursion. The eventual value function for period t is the average over the sample of all possible forecasts. More formally, the computational version of the Bellman equation is

$$V(z_t, t) = \max_{c_t, e_t} \left\{ u(c_t) + \frac{\delta}{5R} \sum_{i=1}^{R} \sum_{j=1}^{5} V\left(\cdot, w_j^p(\hat{\hat{\beta}}_{i,t+1}), t+1\right) \right\},$$

with $\hat{\hat{\beta}}_{i,t+1} = a_i + A_i(a_i + A_i\hat{\hat{\beta}}_{i,t-1})$, where $(\hat{\hat{\beta}}_{i,t-1}, a_i, A_i)$ is the ith draw from the Gibbs sampler, and R is the number of draws from the Gibbs sampler.

The time horizon for the agent is always 40 years (i.e., $T = 40$). However, we vary the starting date ($t = 1$) to correspond with the calendar years 1980, 1985, and 1990. For all cohorts, we utilize historical data from 1964, so, for example, when the model is started in 1980 the individual has 16 years of previous observations upon which to compute his forecasts in the very first year. Similarly, for the 1985 and 1990 cohorts there are 16 and 21 years of data, respectively. For validation purposes we use data only for the years 1964 through 1994, and compare the predictions of the model with those of the real data, namely the data from the CPS of 1995 through 2004. For computational purposes we allow individuals to obtain education only in the first 20 years after completion of their high school education. In reality, very few, if at all, obtain education after the age of 38, which is our implicit limit. For the last 20 years of the individual's horizon consumption is the sole choice variable. We verified and found, by employing an unconstrained model, that this constraint bares virtually no consequence on the results obtained.

We already discussed the assumption about the evolution of $\hat{\beta}^t$from the point of view of the agent. Of the remaining state variables, education and experience are discrete, while wealth and wages are continuous. The latter two are discretized with linear interpolations used to compute the value function at any point in the state space. We allow for nine possible levels of education, 12 through 20 years (i.e., $\overline{S} = 20$), 40 levels of experience, one for each year of the time horizon. Wealth and wages are discretized into 30 and 45 points, respectively. This gives rise to 12,150 elements in the state space. For all these points we compute the exact value function. For all points that do not fall exactly on the grid described above we use linear interpolation to compute the value functions at these points.

The individual' s utility function is assumed to be of the constant relative risk aversion

(CRRA) form:

$$u(c) = \frac{c^{1-\alpha}}{1-\alpha},$$

where $\alpha > 0$. There are two ways to interpret the parameter αin this specification. First, $\alpha$is the coefficient of relative risk aversion (i.e., $\alpha = -(u''(c) \cdot c)/u'(c))$, where the greater is $\alpha$the more risk averse the individual. Second, it is straightforward to show that the intertemporal elasticity of substitution of consumption is $\eta = 1/\alpha$. In this context, more curvature in the utility function (i.e., larger α) implies that the individual would like to smooth consumption more over time relative to a person with smaller α.

The parameters of the current computational model are the rate of interest, r, the individual's discount factor, δ, the CRRA parameter, α, and the calendar year in which the model begins. For given values of these parameters we compute the sequence of policy functions which are then used for simulating the behavior of an individual over their life-cycle from some specified initial state. In performing the simulations the individual receives wage draws based upon the estimated and forecasted β's described above. For α we use values that lie between 0.2 and 2.0. These values are consistent with the prior literature (see the survey of prior research by Bliss and Panigirtzoglou, 2003).

For all of the results shown in the current version of this study it is assumed $r = 0.07$ and $\delta = 0.95$, while the remaining parameters are varied, and stated in each figure and table below. The reason we chose the discount rate to be $\delta = .95$ is because this is the parameter that is typically obtained when estimation takes place. Furthermore, this is the value that has been used by virtually all studies in the literature in which $\delta$is not estimated. Of course, the results might be sensitive to the choice of δ, but it is also well-known that this parameter is very hard to pin down.

The interest rate was chosen to be $r = 0.07$ based on the average secondary market interest rate for one-year Treasury Bill. Over the period from 1964 to 1994 the interest fluctuated enormously from a low rate of about 2% to over 14%. The average for the whole period was about 7%. The large variation in interest rate may suggest that one should include the interest rate as a stochastic term in the state vector. However, it is not feasible to incorporate because it would complicate the computation significantly. Also, the examination of the effect of interest rate on educational choice is certainly beyond the scope of this paper.

In addition to using the VAR model with Gibbs sampling (VAR-Gibbs) to integrate over the distribution of future wage forecasts, we consider three other, simpler, forecasting behaviors: No foresight (NF), perfect foresight (PF) and the VAR forecast without Gibbs sampling (VAR). In the NF case, the individual uses the most recent wage distribution as his forecast for all future wage distributions (as if wages are a random walk). In the PF case the agents uses *actual* future wage distributions for his forecasts.

A limitation in both cases arises from the fact that the planning horizon, and also the periods in which the individual is choosing education and consumption, extends beyond our

March-CPS sample. Consequently, whenever the time period in the model corresponds to a year after 1994, we assume the March-CPS data equals the prediction from the VAR-Gibbs methods. That is, we take the mean of the posterior distribution for the parameter vector and forecast future conditional distributions of wage based on that set of values. Two important facts are worth noting in this context. First, the PF forecast is not the same as that of any of the other forecasts, and specifically not as the VAR forecast. This is obviously true for years up to 1994, but it is also true for years after 1994. We conducted some sensitivity analyses in which we used the available data from the March-CPS for more recent years, i.e., 1995 through 2004. We find that the changes in the results from doing so are trivial and they do not alter our general conclusions. The most important reason for this fact is that most of the educational decisions are made early in life and with a discount rate of $\delta = .95$the future is discounted heavily. In any event, the alternative forecasting methods lead to important comparisons in our results.

5.1 Optimal Policies—Choice of Education

In this subsection we examine the effect of the various factors on schooling choices in the first year after high school graduation. Figure 3 depicts the optimal choice of education, under various forecasting models, in the first year after high school, with the CRRA parameter of $\alpha = 1.2$. The figure depicts the optimal education choices—enroll in the first year of college or not—for every wage-wealth combination, given the information available at the time. All the wage-wealth combinations below the graphs represent the states in which an individual chooses to attend college, while all wage-wealth combinations above the graphs correspond to states where the individual would choose to work.

In Figure 3a we show the results for those who graduated from high school in 1980 under the four alternative forecasting rules outlined above. Figure 3b does the same for those who graduated from high school in 1985. A few important results are apparent from these graphs. First, if high school graduates in 1980 had perfect foresight of the sharp increase in the return to education, then, relative to the NF case, approximately a 10% higher wage offer would have been necessary to attract them into the workforce at that time. Moreover, a larger fraction of the new high school graduates would have decided to attend college in 1985 than in 1980, because they are to enjoy longer period with higher wages over their life-cycle.

Under the VAR forecasting method, there are fewer states of nature for which attending college is an optimal decision than under the NF case. Moreover, when parameter uncertainty is taken into account, namely under the VAR-Gibbs forecasting, then educational choice is optimal in even fewer states than under the VAR model. This suggests that parameter uncertainty may introduce relatively more uncertainty with respect to education, than it does with respect to labor market experience. We discuss this issue in more detail below.

The prediction of the naive NF model is neither consistently below or above the predictions obtained for the VAR or VAR-Gibbs models. Comparing Figures 3a and 3b reveals that the

NF model generally over predicts the fraction of people attending college in good times when the implied return to education seems high (1985), while under predicting college attendance in relatively bad times (1980).

Figure 3c presents the results for the VAR-Gibbs model for the three starting years 1980, 1985 and 1990 with the CRRA parameter of $\alpha = 1.2$. This is our leading example throughout the remainder of this section. Clearly, there are many more wage-wealth combinations in which college choice is optimal in 1985 than in 1980, and even more so in 1990. This represents the general right shift in the conditional wage distributions, throughout the 1980's, particularly the vast increases in the return to education and in the return to experience for the more highly educated individuals.

Note that in all the graphs depicted in Figure 3 it is optimal for a larger fraction of the wealthier individuals to attend school. In the model considered here this is merely due to the fact that individuals are faced with liquidity constraints, since borrowing is not allowed. In reality there are other factors that are associated with wealthier individuals getting more education, but we abstract from these aspects in this study.

5.2 Simulation Results

The above discussion regarding optimal decisions was limited to behavior in the first period for an individual with 12 years of schooling and no work experience. However, there are other interesting issues concerning behavior in later periods, especially because one might change his decision in light of new information in subsequent years.

Using the sequence of policy functions computed for each possible state of nature, we simulate 10,000 individuals through the first 20 years of their post high school careers. The results reported below are for various combinations of forecasting methods, initial endowments, level of risk aversion (or alternatively the intertemporal substitution), and year of high school graduation. However, most of the attention is given to the first two aspects, as they have attracted very little attention in the literature. The wage offer the individual receives in each period is a random draw from the estimated distribution for that period, conditional on their attained levels of education and experience.[22]

Since all individuals are high school graduates in the first period, in all simulations reported here the initial state vector includes 12 years of education and zero experience. In the first set of simulations we let each individual have initial wealth of \$10,000.[23] We consider three key starting years, namely 1980, 1985 and 1990—1980 is the year right before the sharp increase in the returns to education, 1985 is the year by which the sharpest increase was achieved, and

[22] While incorporating the fact that wages are correlated over time is of great importance, the issues of wage mobility and its affect on educational choices are beyond the scope of this study. Nevertheless, we also analyze below some aspects of state dependence in wages.

[23] There is huge variation of estimates of initial endowment for young adults in the U.S. Most estimates for mean initial endowment are around \$6,000 in 1980 dollars. We use data from the National Longitudinal Survey of Youth (NLSY) to govern our calibration following Zagorsky (1997, 1999).

1990 is the year by which the increase in the returns to education leveled off.

The results from the primary simulations are presented in Figures 4 to 9, and in each case there are 20 periods marked along the horizontal axis. These 20 periods correspond to different calendar years depending on the starting date of the model.[24]

Figure 4 depicts the fraction of people attending school for the three cohort groups and for each of the four forecasting methods discussed above. The results clearly indicate that for the first few years, the fraction of people from the 1980 cohort who attend college is much lower, as one might expect, than for the 1985 and 1990 cohorts. As expected, smaller differences are observed for the individuals with perfect foresight. Restricting attention to Figures 4c and 4d, we see that this trend changes when the 1980 cohort is 10 years out, as some individuals return to school in response to the sharp increase in the return to education. It is further apparent that when parameter uncertainty is taken into account (Figure 4d), a lower fraction attends school at any given period, due to the apparently greater uncertainty (i.e., parameter uncertainty) associated with education relative to that associated with experience. Note that both the VAR and VAR-Gibbs versions of the model do capture the fact that the 1980 cohort begin with a noticeable lower fraction attending school.

The cumulative effect of school attendance is shown in Figure 5, which depicts the average level of accumulated education (over the first 20 years of the individual horizon after graduating from high school). Clearly, the alternative forecasting methods lead to vastly different predicted behavior. For example, the perfect foresight method (Figure 5b), yields hardly any difference among the three cohorts. With perfect foresight and no cost to education, beyond foregone earnings, all cohorts accumulate virtually the same amount of education. In contrast, under the other three alternatives there are much larger differences in the average level of accumulated human capital among the three cohorts. Note that even though some of the individuals from the 1980 cohort return back to school, on average, they never "catch up" with the individuals from the 1985 and 1990 cohorts who made the college enrollment decision at periods when education seemed to be a lot more attractive. This is true, even though the only cost of education introduced so far is the foregone wage, because the later cohorts get to benefit from the rising returns to education over a longer period of time.

The differences in predicted behavior depending on the forecasting method is also intuitive. Under the NF method (Figure 5a) individuals obtain, on average, the highest level of education relative to all other methods. This is because under the NF method (given the data over the sample period) the individual systematically over estimates the returns to education in future years. The average level of education decreases significantly under the VAR method (Figure 5c) by almost one full year, by the 20th period. And, when parameter uncertainty is accounted for (Figure 5d), the average level of education declines by an additional half a year.

The main goal of the paper is not estimation, but is rather focused on examining the

[24] Specifically, the year marked as 2 for example is 1981 for the cohort whose starting year is 1980, while for the cohort starting in 1985 the year marked as 2 is 1986, and so forth.

implications of alternative forecasting scenarios. Thus, issues regarding the in-sample and out-of-sample fitness of the various models are not key here. Nevertheless, given our focus on differential forecasting models it is of interest to assess which of the competing models best explains changes across the cohorts over the sample period. One way of evaluating this is by comparison of the Pearson χ^2statistics for the models against the data. All models' predictions are quite accurate, which suggest that the key issue in modeling educational choices is the forward looking aspect that exist in all models. Nevertheless, overall, the VAR-Gibbs model, the most general model, which account for all sources of uncertainty, seems to perform the best, both in terms of the within-sample as well as the out-of-sample predictions. For this reason we use it as our leading model in presenting the remainder of the results in this paper.[25]

In Figure 6 we show the model's prediction of the average annual wages for the three cohorts over the first 20 years of their planning horizon. During the first few years after graduating from high school, all cohorts earned similar wages under each of the various forecasting scenarios. However, as time progressed the cohorts who had greater incentives to attend school started realizing the returns on their investment. The patterns of changes in the wage profile reflect the pattern of changes in college attendance depicted in Figure 4. Consequently, the average wage in period 20 (for those who work) is significantly higher for the 1985 and 1990 cohorts than for the 1980 cohort. While, as expected, there is very little difference in the pattern of changes between the three cohorts for the perfect foresight case, there are significant differences for all other cases. For example, in the VAR-Gibbs case average wages in the 20th period are lower by approximately $2,000 (or over 7%) for the 1980 cohort relative to the 1985 cohort, and even more so relative to the 1990 cohort.

As noted above, there is evidence in our simulations that the returns to education may be more risky than the returns to experience, driven, at least partially, by relatively higher parameter uncertainty. We explore this issue more carefully in Figure 7. Here we depict the results when the parameter $\alpha$is increased to $\alpha = 2.0$ (from $\alpha = 1.2$in Figures 3 to 6). The changes in the results induced by an increase in the CRRA coefficient are qualitatively and quantitatively striking. More risk averse individuals choose less education than the relatively less risk averse people. For example, under the VAR-Gibbs model, less risk averse people ($\alpha = 1.2$) from the 1980 cohort accumulated, on average, close to 17 years of education by the end of the 20-year horizon, compared with only 14.5 for the more risk averse ($\alpha = 2.0$) individuals from this cohort. Consequently, the more risk averse agents also obtain lower average wages (figure is omitted). It is worth noting that some of the affect of the parameter α can be interpreted as the effect of intertemporal substitution. That is, individuals with larger αwould like to smooth consumption more than individuals with smaller α. Since we do not allow for borrowing in our model, individuals with larger αprefer to join the labor market earlier, in order to obtain wages, which, in turn, will allow them to smooth consumption more relatively to those with smaller α.

[25] We omit the results for brevity. But, all tables with the results are available from the authors upon request.

The inclusion of parameter uncertainty exacerbates the effect of α on educational investments, as shown in Figure 7b. With $\alpha = 1.2$, the difference in accumulated education after 20 years under perfect foresight versus the VAR-Gibbs model is about one year. Meanwhile, if $\alpha = 2.0$ this difference is about 1.5 years. This implies that not only are the returns to education relatively more uncertain given known distributions of future wages, but the uncertainty inherent in forecasting future conditional wage distributions is greater for higher levels of education.

It is important to note some caveats to these findings about the role of α. Our analysis does not take into account involuntary unemployment. If the probability of involuntary unemployment is higher among low education individuals (e.g. high school graduates) then it is conceivable that educational investments may be associated with relatively less uncertainty. Perhaps even more importantly, we have not incorporated unobserved individual heterogeneity into this analysis (although we do explore this issue, below). For example, high ability individuals may face less uncertainty in the expected returns to education than low ability individuals. We have to be cautious, though, about the exact implications of our results. We have assumed CRRA, and it is conceivable that alternative utility functions, or alternative approaches to incorporating risk, may yield different conclusions.

Tables 3 and 4 provide a more complete summary of the full set of simulations. Table 3 reports the distribution of the accumulated education levels at the end of period 20 for the 10,000 simulated individuals, for a variety of scenarios with alternative forecasting methods, starting years, and attitudes toward risk. Table 4 reports the distribution of the number of years it took to complete the first four years of college, for the same alternative scenarios as in Table 3.[26]

Consistent with the discussion above regarding risk aversion, Table 3 shows that the higher the degree of CRRA the more skewed to the right is the distribution of completed years of education. A significant number of individuals with relatively high α(i.e., $\alpha = 2.0$) never finish college, especially for the high school graduate cohort of 1980. Comparing the results for cohorts with the same attitude toward risk across the various forecasting methods indicates the importance of taking account of uncertainty, in general. For example, the VAR-Gibbs model with $\alpha = 2.0$ in 1985 (line 34) yield a distribution of completed years of education far more skewed to the right than that implied by the VAR model (line 25). In fact, the VAR-Gibbs model indicates that over 80% of the individuals of the 1980 cohort never completed four years of college, while the VAR model predicts that less than 70% will never complete college. In contrast, the PF model predicts that over 60% will complete four years of college. Again, these results, when compared with known observed trends of college enrollment documented in the literature, seem to indicate that individuals are highly risk averse, and are more likely to behave in a fashion consistent with the VAR-Gibbs model, that is, take into account all sources

[26] The numbers in each row of Table 6 need not add up to 10,000, as some of the simulated individuals never completed four years of college education.

of uncertainty, including *parameter uncertainty.*

From a policy standpoint one would like to analyze not only the eventual educational attainments, but also the amount of time needed to reach these attainments. Table 4 reports the speed by which four years of college education were completed, for those who finished a four-year program. The actual speed observed in the population, for any of the groups analyzed, is undoubtedly shorter, but with less people actually completing college. This is mainly due to the fact that tuition, and other observed and unobserved costs, are not integrated into the results reported in Table 4.

Evidently, most of the individuals from the 1980 cohort completed their college education in more than four years. For the other two cohorts, the speed is faster, yet there is a significant number of individuals who required more than 7 years to complete college. While the pattern is similar across the four alternative forecasting models, the magnitudes are somewhat different. Under the VAR model more years are required to complete the first four years of education. This is in sharp contrast to the PF case where individuals complete their education relatively faster. Under the VAR-Gibbs model completion of the first four years takes substantially longer than under any of the other scenarios. There are several reasons for the relatively long spell until education is completed. First, there is no loss incurred by skipping school for a number of years. Second, there is no borrowing in the current model, so that an individual may have to leave school in order to accumulate enough wealth in order to survive.

6 Implications of Other Factors

The framework proposed in this paper is well suited for examining other important aspects affecting educational decisions. In this section we briefly examine several of these issues.

Tuition Costs:

We can use our model to analyze the impact of liquidity constraints (or the availability of funding from government or family) and tuition costs on educational choices, both of which are important policy issues. To do so, we utilize data on resident undergraduate tuition and/or required fees from a nationwide sample of colleges and state universities for the years 1972 through 1994, collected by the State of Washington Higher Education Coordinating Board.[27]Figure 8 provides the historical data from 1972 to 1992. We see that the pattern of changes in the resident tuition is very similar to that of non-resident tuition, but the latter is much higher at all years. For the analysis here we use the resident tuition. To examine the effect of liquidity constraints on the accumulation of human capital we solve the model varying the levels of initial wealth at the time the individuals finish high school, namely \$5,000, \$20,000, and \$50,000.[28]

[27]We thank Tom Kane for making this data available to us. Kane analyzes these data extensively in Kane (1995).

[28]These values correspond to estimates provided by the National Center for Education Statistics for the low, medium, and high level of funding for freshmen in four-year public universities.

Figure 9 depicts the results for accumulated education when tuition costs are added to the analysis. Comparing Figure 9 with Figure 5d clearly illustrates the impact of tuition costs over the 20-year horizon. With no tuition costs an individual from the 1980 cohort accumulates almost a full year of education more than an individual from that cohort who is faced with tuition costs. Similar, but not as strong affects are also observed for the other two cohorts. Consequently, tuition costs cause average annual wages to drop for all cohorts (figure is omitted). This also suggests that the rising tuition cost along with the higher uncertainty in the reward of education relative to that of accumulated experience is a major reason for individuals not to attend colleges even in face of the rising return to education.

Table 5 describes the distribution of education levels when tuition cost is incorporated, under the VAR-Gibbs forecasting model. The table indicates that there is a significant left shift of the education distribution. For example, without tuition costs approximately 90% of the 1985 cohort completed four years of college; but with tuition this fraction falls below 75%. The results also indicate that tuition costs lengthened the time until completion of a four-year college program. Although not shown in the table, these results apply for all forecasting methods, although it takes more time to complete four years of college education under the method that incorporates all sources of uncertainty—the VAR-Gibbs method.

Next we examine the effect of liquidity constraints on the accumulation of human capital. The key results for the VAR-Gibbs model, with $\alpha = 1.2$, are depicted in Figure 10. As expected, individuals with larger initial wealth accumulate dramatically more education. The effect on wages is equally dramatic (not reported). Furthermore, while average wages for the richer individuals are similar to the wages of poor individuals in the early periods, the wage dispersion between rich and poor individuals gets larger as time progresses. This is a direct consequence of the fact that richer individuals benefit from their earlier higher investment in human capital. This seems to suggest that differences in educational achievements stem to large extent because of liquidity constraints. Poorer individuals would benefit from accumulating more human capital, but they simply cannot afford to have the same level of investment as the reacher individuals.

This is clearly demonstrated in Table 6, which provides the distribution of education levels for a larger set of simulations based on alternative initial wealth levels. The table illustrates that the groups with higher levels of initial wealth achieve higher levels of education. We also find (table is omitted for brevity) that these levels of education are achieved in a shorter time. For example, most of the individuals from the 1980 cohort who started with initial wealth of $50,000 complete more than four years of college education. But more importantly, a relatively large fraction did so in less than five years. In contrast, a large fraction of the poor individuals from the 1980 cohort never completed four years of education, and for those who did it takes a significantly longer time.

Unobserved Heterogeneity:

We now examine the role of unobserved heterogeneity. To do that we employ in the simulations three alternative types of individuals that differ from each other by the distribution from which they draw their wages. That is, the support of the distribution is the same as before, only that we assign the three groups different probability distributions. As noted above we use the VAR-Gibbs method and employ the CRRA parameter $\alpha = 1.2$.

As in all previous simulations the wage offers were randomly drawn from the estimated conditional distribution for that period, conditional on education and experience. That is, the probability of each conditional quantile was .2. Here we change these probabilities, assigning the following vectors of probabilities, at the .10, .30, .50, .70, and .90 quantiles for the three types of individuals considered:

$$\begin{aligned} &\text{Type I:} \quad (.2, .4, .2, .2, 0)\,; \\ &\text{Type II:} \quad (.1, .2, .4, .2, .1)\,; \quad \text{and} \\ &\text{Type III:} \quad (0, .2, .2, .4, .2)\,. \end{aligned}$$

In this setup Type I will tend to systematically get *unfavorable* wage draws relative to all other types, regardless of whether or not they completed college education. Type II has symmetric distribution, but relative to our base case will tend to get wage draws closer to the center of the conditional distributions. Finally Type III tend to systematically get *favorable* wage draws relative to all other types.

Figure 11a provides several important results of these set of simulations. First, we observe that the least favorable individuals (Type I) obtain on average more education than the other two groups, even though they also get less favorable wage draws after completing higher education. This is mainly because the only costs of education in these sets of simulations is forgone wage and it is relatively low for this group of individuals. In contrast, Type III individuals get the lowest education; by over one year relative to Type I individuals. Also, for Type II individuals even though they have a symmetric wage distribution, just like in the base case, they tend to get on average lower education. This is because individuals of that type have higher probability of drawing high enough wage offers that draw them to work rather than obtain additional education. Also, the individuals are risk averse, so distribution with smaller variance are preferred. Consequently, they obtain lower education relative to Type I individuals, but much more than Type III. In any case, the introduction of relatively simple unobserved heterogeneity can also play a major role in one's educational choices, and consequently in wage outcomes (figure is omitted).

While we do not explain here why it is that different individuals have different probability distributions, the literature of the U.S. wage structure has given ample examples that will lead to such a situation, the most important one being differences in inherent ability. Nevertheless, the results clearly show that while unobserved heterogeneity can play a major role, the overall pattern of changes observed in our simulations is very similar to that of the base case. In other words, unobserved heterogeneity can certainly explain variation in outcomes, but the randomness in the model produces significant amount of ex-post observed heterogeneity, even though all individuals start with the very same endowment under the base case.

State Dependence:

Finally, we consider the possible effects of having state dependence in wages, that is, having correlated wages over time. To do so we use two different transition matrices for high school graduates and college graduates. These are given, respectively, by[29]

$$TR_{HS} = \begin{pmatrix} .2 & .4 & .2 & .1 & .1 \\ .15 & .2 & .4 & .15 & .1 \\ .45 & .25 & .15 & .1 & .05 \\ .15 & .1 & .5 & .15 & .1 \\ .05 & .05 & .1 & .15 & .65 \end{pmatrix} \quad \text{and } TR_{COL} = \begin{pmatrix} .65 & .15 & .1 & .05 & .05 \\ .15 & .55 & .15 & .1 & .05 \\ .1 & .15 & .55 & .1 & .1 \\ .05 & .1 & .1 & .65 & .1 \\ .05 & .05 & .1 & .10 & .7 \end{pmatrix}, \quad (14)$$

where the rows are the quantiles at time t, while the columns are the quantiles at time $t+1$. For all the other parameters we use those in the base case, namely the VAR-Gibbs method. Note, there is a high degree of stability in wages for college graduates (i.e., the transition probabilities on the diagonal of TR_{HS} are quite large). In contrast, for high school graduates the transition probabilities in TR_{HS} indicate that there is a lot more mobility across the distribution than for college graduates.

The results for these set of simulations are given in Figure 11b and in Table 7. Interestingly, serial dependence in wages need not lead to higher education, even when the serial correlation in wage for graduate students is stronger than that for high school graduates. In fact, serial dependence in wages leads to average accumulated education that is almost two years lower than that under the base case. Consequently, average wages are lower (figure is omitted), as are the consumption and wealth levels (not shown).

From the transition matrices given in (14), note that the probabilities for obtaining good wage offers when individuals graduate from high school are quite high, especially relative to the probabilities of getting low offers. Moreover, mobility of individuals through the wage distribution is relatively low at that stage of their career (i.e., there is significant serial correlation in wages). In turn, individuals tend to get large enough wage offers that induce them to work, and when they do so their wages tend to be positively serially correlated. This leads them to stay in the labor market rather than attend school and acquire higher education. Needless to say, incorporating the direct cost of education will lead to an even larger discrepancy between the base case and the case considered here.

[29] These matrices are based on estimates obtained in Buchinsky and Hunt (1999).

7 Conclusion

We have presented a dynamic model of individuals' educational investments that allows us to explore alternative modeling strategies for forecasting future wage distributions, which are essential in any dynamic programing model. The forecasting procedures we explore are based only on the information that would be available at the actual time these decisions were made. Typically, studies the employ dynamic programing models assume some stationary distribution for wages. Since the returns to education were dramatically rising during the period of our analysis, such an approach has the feature that individuals systematically under-predict the (future) returns to schooling for at least some periods when educational choices were being made. In contrast, our approach provides a more plausible framework for studying educational investments for which the returns will be realized for a very long time after the actual investment is made.

An important innovation in this study is that we also incorporate the role of parameter uncertainty into the decision making process. In other words, the risk-averse individual in our framework not only takes into the account the uncertainty of future wage draws (conditional on education and experience), but also the uncertainty in the distributions themselves, i.e., the uncertainty in the parameters of the (estimated) conditional wage distributions. This adds to the plausibility of our forecasting model—any prediction of future unknown distributions inherently involves uncertainty. We are unaware of prior research in the dynamic programming literature that has included this feature.

Our analysis yields a number of findings. Firstly, the particular forecasting method of future wage distributions has a major impact on predicted individual behavior. We find distinctive differences across forecasting methods in the fraction of individuals that attend school in any given year, the average level of accumulated education, and the time elapsed until a college degree is completed. The most general model of forecasting we propose—the VAR-Gibbs model—incorporates parameter uncertainty into the decision process. This model reveals interesting time paths of educational choices. High school graduates in 1980 are less likely to attend college than their 1985 counterparts. Moreover, the 1980 cohort never catch-up by returning to school later in their careers in order to benefit from the rising returns to education. Similarly, the 1990 high school graduates attain higher levels of education than the 1985 cohort. These predictions are also evident in the data, providing basic verification of the limited ability of individuals in the early 1980s to foresee the dramatic increases in the returns to education.

Secondly, the impact of the observed (and previously documented) increase in returns to education after 1980 on educational investments is mitigated by two factors: (i) individuals' inability to foresee these changes based on the information available at the time; and (ii) along with the rising returns to education there has been an increase in the uncertainty associated with education. That is, there is less variation in the forecast of wage distributions that is due to experience than due to education. In our model an individual has to be engaged in either schooling or in the labor market. Consequently, more risk averse individuals tend to invest less in education and more in on-the-job acquisition of experience. While this aspect is generally ignored in the literature, we find it to be of first order importance. When we also include data on tuition costs into the model, we find that the observed rising tuition costs are a significant deterrent to educational investments during the period of our analysis.

Thirdly, our results suggest that appropriate subsidies can be instrumental in encouraging individuals to accumulate more education. While this has been discussed previously in the literature, our study highlights the potential impact of uncertainty on educational choices. Our results indicate that there can be potential policies that would lead to increase in educational attainment. One example of such a policy is for college graduates to repay tuition fees only if their income exceeds some minimum pre-specified level.[30] This would reduce the uncertainty that is associated with obtaining higher education relative to that associated with obtaining labor market experience.

Structural dynamic models of schooling, such as the one developed here, incorporate a host of assumptions. Foresight of future wage distributions and unobserved individual heterogeneity aside, there are other questionable assumptions that are rarely discussed, such as treating education as homogeneous, and assuming that individuals' unobservable ability is constant over time. Several extensions to our current approach are in order. For example, one may allow for borrowing at any point in time, in addition to analyzing the implications of changing the individual's level of initial wealth.

Overall, this study provides further verification of the usefulness of dynamic programming methods for analyzing individuals' educational choices. An important aspect of these kinds of models is their ability to capture forward-looking behavior, which require some assumption regarding the evolution of future wage distribution. The main point of this paper is that the choice of the forecasting method to achieve this goal is not an innocuous assumption. We therefore propose a general method of forecasting that is embedded in a dynamic optimization framework, which relies upon plausible assumptions about what information is available at the time of forecasting, and which accounts for sources of uncertainty faced by individuals that have not been previously studied.

[30] This is, for example, the current policy in Australia.

Appendix A: Evaluating the Posterior Distribution of $(\beta_h^t, C_\beta, \Sigma_\nu, \Sigma_\beta)$

In this appendix we first specify the prior beliefs and then explain how to obtain the posterior joint distribution for $(\beta_h^t, C_\beta, \Sigma_\nu, \Sigma_\beta \mid \hat{\beta}_h^t)$ using a Gibbs sampler. The posterior distribution is then used for integrating over the true β_t's.

Prior Distribution for the Model's Parameters:

Following the definition of β_τ in Section 3, one can estimate the components of $\beta_\tau = (\beta_{.1,\tau}, \beta_{.3,\tau}, \beta_{.5,\tau}, \beta_{.7,\tau}, \beta_{.9,\tau})$ from five separate quantile regressions of earnings on the individual's characteristics for each of the sample years available at the time the individual makes his decision in period t. We adopt a diffuse prior for β_τ, for all $\tau = 0, \ldots, t$.

Conditional on the true values $\beta_0, \ldots, \beta_t$, C_β is simply given by

$$C_\beta = (X_\beta' X_\beta)^{-1} X_\beta' Y_\beta,$$

where

$$Y_\beta' = \begin{pmatrix} \beta_1 & \cdots & \beta_t \end{pmatrix}, \quad \text{and } X_\beta' = \begin{pmatrix} 1 & \cdots & 1 \\ \beta_0 & \cdots & \beta_{t-1} \end{pmatrix}.$$

We let $c_\beta = \text{vec}(C_\beta)$ have a diffuse prior.

The prior distribution employed for Σ_ϵ^{-1}, associated with ϵ_t in (12), is a suitably diffuse prior given by $p\left(\Sigma_\epsilon^{-1}\right) \propto |\Sigma_\epsilon|^{L/2}$, where L is the dimension of Σ_ϵ.

Unlike the case for Σ_ϵ^{-1}, since we do not have a direct observation on β_t, we cannot employ an improper prior for Σ_ν^{-1}, because such a prior would result in an improper posterior distribution for Σ_ν. We therefore adopt a Wishart distribution as the prior distribution for Σ_ν^{-1}, that is, $\Sigma_\nu^{-1} \sim \text{Wishart}\,(1, I_L)$, where I_L is an identity matrix of dimension L, the dimension of Σ_ν.

Given the above specification our goal is to obtain the posterior distribution of $(\beta_h^t, C_\beta, \Sigma_\nu, \Sigma_\beta)$, given the "data" $\hat{\beta}_h^t$; that is $p(\beta_h^t, C_\beta, \Sigma_\nu, \Sigma_\beta \mid \hat{\beta}_h^t)$. We do so using a Gibbs algorithm, drawing from the conditional distribution of only a subset of the parameters at a time, conditioning on all other parameters, as explained below.

The Gibbs Sampler:

In order to cycle through the various steps of the Gibbs sampler, one needs to choose starting points for the model's parameters. The natural starting values for Σ_ϵ, Σ_ν, and C_β are their maximum likelihood estimates. For β_h^t we simply take $\hat{\beta}_h^t$ as the initial value. Given these initial starting values we cycle through the following four steps of the Gibbs sampler:

Step 1:

Sample sequentially from the conditional distribution of β_h^t given $(\hat{\beta}_h^t, C_\beta, \Sigma_\nu, \Sigma_\epsilon)$. If we define $\beta_{h,-\tau}^t$ as β_h^t, excluding β_τ, and similarly for the other variables; then we first sample from the conditional distribution of β_1 given $(\hat{\beta}_h^t, \beta_{h,-1}^t, C_\beta, \Sigma_\nu, \Sigma_\epsilon)$. For the VAR model outlined above and the assumed prior distributions, this distribution is normal with mean and variance given, respectively, by

$$\text{E}\left[\beta_1 \mid \hat{\beta}_h^t, \beta_{h,-1}^t, C_\beta, \Sigma_\nu, \Sigma_\epsilon\right] = W_{11}\hat{\beta}_1 + W_{21}\beta_1^f; \quad \text{and}$$
$$\text{Var}\left(\beta_1 \mid \hat{\beta}_h^t, \beta_{h,-1}^t, C_\beta, \Sigma_\nu, \Sigma_\epsilon\right) = \left(\Sigma_\epsilon^{-1} + \left(A'\Sigma_\nu^{-1}A\right)\right)^{-1},$$

where

$$
\begin{aligned}
W_{11} &= \left(\Sigma_\epsilon^{-1} + \left(A'\Sigma_\nu^{-1}A\right)\right)^{-1}\Sigma_\epsilon^{-1};\\
W_{21} &= \left(\Sigma_\epsilon^{-1} + \left(A'\Sigma_\nu^{-1}A\right)\right)^{-1}\left(A'\Sigma_\nu^{-1}A\right); \quad \text{and}\\
\beta_1^f &= A^{-1}\left(\beta_2 - a\right).
\end{aligned}
$$

Now sequentially sample all other β_τ for $\tau = 2, \ldots, t-1$. The conditional distribution of β_τ, given $(\hat{\beta}_\tau, \hat{\beta}^t_{h,-\tau}, C_\beta, \Sigma_\nu, \Sigma_\epsilon)$, is normal with mean and variance given, respectively, by

$$
\begin{aligned}
\mathrm{E}\left[\beta_\tau \mid \hat{\beta}^t_h, \hat{\beta}^t_{h,-\tau}, C_\beta, \Sigma_\nu, \Sigma_\epsilon\right] &= W_{1\tau}\hat{\beta}_\tau + W_{2\tau}\beta^f_\tau + W_{3\tau}\beta^b_\tau; \quad \text{and}\\
\mathrm{Var}\left(\beta_\tau \mid \hat{\beta}^t_h, \hat{\beta}^t_{h,-\tau}, C_\beta, \Sigma_\nu, \Sigma_\epsilon\right) &= \left(\Sigma_\epsilon^{-1} + \Sigma_\nu^{-1} + \left(A'\Sigma_\nu^{-1}A\right)\right)^{-1},
\end{aligned}
$$

where

$$
\begin{aligned}
W_{1\tau} &= \left(\Sigma_\epsilon^{-1} + \Sigma_\nu^{-1} + \left(A'\Sigma_\nu^{-1}A\right)\right)^{-1}\Sigma_\epsilon^{-1};\\
W_{2\tau} &= \left(\Sigma_\epsilon^{-1} + \Sigma_\nu^{-1} + \left(A'\Sigma_\nu^{-1}A\right)\right)^{-1}\left(A'\Sigma_\nu^{-1}A\right);\\
W_{3\tau} &= \left(\Sigma_\epsilon^{-1} + \Sigma_\nu^{-1} + \left(A'\Sigma_\nu^{-1}A\right)\right)^{-1}\Sigma_\nu^{-1};\\
\beta^b_\tau &= a + A\beta_{\tau-1}, \qquad \text{and}\\
\beta^f_\tau &= A^{-1}\left(\beta_{\tau+1} - a\right).
\end{aligned}
$$

Finally, sample from the conditional distribution of β_t, given $(\hat{\beta}^t_h, \beta^t_{-t}, C_\beta, \Sigma_\nu, \Sigma_\epsilon)$. This distribution is normal with mean and variance given, respectively by

$$
\begin{aligned}
\mathrm{E}\left[\beta_t \mid \hat{\beta}^t_h, \beta^t_{-t}, C_\beta, \Sigma_\nu, \Sigma_\epsilon\right] &= W_{1t}\hat{\beta}_t + W_{3t}\beta^b_t; \quad \text{and}\\
\mathrm{Var}\left(\beta_t \mid \hat{\beta}^t_h, \beta^t_{-t}, C_\beta, \Sigma_\nu, \Sigma_\epsilon\right) &= \left(\Sigma_\epsilon^{-1} + \Sigma_\nu^{-1}\right)^{-1},
\end{aligned}
$$

where

$$
\begin{aligned}
W_{1t} &= \left(\Sigma_\epsilon^{-1} + \Sigma_\nu^{-1}\right)^{-1}\Sigma_\epsilon^{-1},\\
W_{3t} &= \left(\Sigma_\epsilon^{-1} + \Sigma_\nu^{-1}\right)^{-1}\Sigma_\nu^{-1}, \quad \text{and}\\
\beta^b_t &= a + A\beta_{t-1}.
\end{aligned}
$$

Step 2:

Sample from the conditional distribution of c_β, given $(\hat{\beta}^t_h, \beta^t_h, \Sigma_\epsilon, \Sigma_\nu)$. This conditional distribution is normal with mean and variance given, respectively, by

$$
\begin{aligned}
\mathrm{E}\left[c_\beta \mid \hat{\beta}^t_h, \beta^t_h, \Sigma_\nu, \Sigma_\epsilon\right] &= \mathrm{vec}\left(\left(X'_\beta X_\beta\right)^{-1} X'_\beta Y_\beta\right); \qquad \text{and}\\
\mathrm{Var}\left(c_\beta \mid \hat{\beta}^t_h, \beta^t_h, \Sigma_\nu, \Sigma_\epsilon\right) &= \Sigma_\nu \otimes (X'_\beta X_\beta)^{-1}.
\end{aligned}
$$

Step 3:

Sample from the conditional distribution of Σ_ϵ^{-1}, given $(\hat{\beta}^t_h, \beta^t_h, C_\beta, \Sigma_\nu)$. This conditional distribution is Wishart:

$$
\Sigma_\epsilon^{-1} \mid \hat{\beta}^t_h, \beta^t_h, C_\beta, \Sigma_\nu \sim \mathrm{Wishart}\left(K_\beta, V_\beta\right),
$$

where $K_\beta = t$, and

$$V_\beta^{-1} = \sum_{t=1}^{t} \left(\beta_t - \hat{\beta}_t\right) \left(\beta_t - \hat{\beta}_t\right)'.$$

Step 4: Sample from the conditional distribution of Σ_ν^{-1}, given $(\hat{\beta}_h^t, \beta_h^t, C_\beta, \Sigma_\epsilon)$. This conditional distribution is Wishart:

$$\Sigma_\nu^{-1} \mid \hat{\beta}_h^t, \beta_h^t, C_\beta, \Sigma_\epsilon \sim \text{Wishart}\left(K_\nu, V_\nu\right),$$

where $K_\nu = T + 1$, and

$$V_\nu^{-1} = I_L + (E'E)^{-1}, \quad \text{and } E = Y_\beta - X_\beta C_\beta.$$

Cycling through these four steps of the Gibbs sampler will eventually lead to draws that can be considered as coming from the posterior distribution of $(\beta^t, C_\beta, \Sigma_\nu^{-1}, \Sigma_\epsilon^{-1})$, given $\hat{\beta}_h^t$.

References

Abowd, J.M. and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, **57**(2), 411–45.

Atakan, A.E. (2003): "Stochastic convexity in Dynamic Programming," *Economic Theory*, **22**, 447–455.

Becker, G. (1964): *Human Capital*, New York: National Bureau of Economic Research.

Belzil, C. and J. Hansen (1997): "Estimating the Returns to Education from a Non- Stationary Dynamic Programming Model," Working paper, Centre for Labour Market and Social Research, University of Aarhus, 97-06.

Belzil, C. and J. Hansen (1999): "Subjective Discount Rates, Intergenerational Transfers and the Return to Schooling," IZA Discussion Paper No. 60.

Belzil, C. and J. Hansen (2004): "Earnings Dispersion, Risk Aversion and Education," in Polachek, S.W. (ed.) *Accounting for worker well-being*, Research in Labor Economics, Amsterdam; San Diego and Oxford: Elsevier, JAI, 23, 335–358.

Ben-Porath, Y. (1967): "The Production of Human Capital and Life Cycle Earnings," *Journal of Political Economy*, 75, 352–365.

Berkovec, J. and S. Stern (1991): "Job Exit Behavior of Older Men," *Econometrica*, 59(1), 189–210.

Blinder, A.S. and Y. Weiss (1976): "Human Capital and Labor Supply: A Synthesis," *Journal of Political Economy*, 84(3), 449–472.

Bliss, R.R and N. Panigirtzoglou (2003): "Option-Implied Risk Aversion Estimates," forthcoming in *Journal of Finance*.

Bray, M. (1982): "Learning Estimation and the Stability of Rational Expectations Equilibria," *Journal of Economic Theory*, 26, 318-339.

Bray, M. and N. Savin (1986):"Rational Expectations Equilibria, Learning, and Model Specification," *Econometrica*, 54, 1129–1160.

Buchinsky, M. (1994): "Changes in the U.S. Wage Structure 1963–1987: Application of Quantile Regression," *Econometrica*, 62, 405–458.

Buchinsky, M. and J. Hahn (1998): "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66(3), 653–71.

Buchinsky, M. and J. Hunt (1999): "Wage Mobility in the United States," *Review of Economics and Statistics*, 81, 351–368.

Chow, G.C. and Y.K. Kwan (1997): "Rational Expectations Is Not Generally Valid for Econometric Models: Evidence from Stock Market Data," *Pacific Economic Review*, 2(3), 149–63.

DeCanio, S.J. (1979): "Rational Expectations and Learning from Experience," *Quarterly Journal of Economics*, 94, 47–57.

Dominitz, J. and C.F. Manski (1997): "Using Expectations Data to Study Subjective Income Expectations," *Journal of the American Statistical Association*, 92(439), 855–67.

Dwyer, G.P., Williams, A.W., Battalio, R.C. and T.I. Mason (1993): "Tests of Rational Expectations in Stark Setting," *The Economic Journal*, 103, 586–601.

Farmer, R.E.A. (1991): "The Lucas Critique, Policy Invariance and Multiple Equilibria," *Review of Economic Studies*, 58, 321–332.

Fourgeaud, C., Gouriéroux, C. and J. Pradel (1986): "Learning Procedures and Convergence to Rationality," *Econometrica*, 54, 845–868.

Fuller, D., Manski, C.F. and D. Wise (1982): "New Evidence on the Economic Determinants of Postsecondary Schooling Choices," *Journal of Human Resources*, 17, 477–495.

Gouriéroux, C., Lafont, J.J. and A. Monfort (1982): "Rational Expectations in Dynamic Linear Models: Analysis of the Solutions," *Econometrica*, 50, 409–426.

Griliches, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45, 1–21.

Harvey A.C. (1993): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Heckman, J.J. (1993): "Assessing Clinton's Program on Job Training, Workfare, and Education in the Workplace," National Bureau of Economic Research Working Paper: 4428.

Heckman, J.J., Roselius, R.L. and J.A. Smith (1994): "U.S. Education and Training Policy: A Re-evaluation of the Underlying Assumptions behind the 'New Consensus'," in Solomon, L.C. and A.R. Levenson (eds.) *Labor Markets, Employment Policy, and Job Creation*, Milkin Institute Series in Economics and Education, Westview Press, 83–121.

Hotz, V.J. and R.A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

Juhn, C., Murphy, K. and B. Pierce (1993): "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, 101(3), 410–42.

Kane, T.J. (1995): "Rising Public College Tuition and College Entry: How Well do Public Subsidies Promote Access to College?," NBER Working Paper No. 5164.

Keane, M. and K. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.

Manski, C.F. (1989): "Schooling as Experimentation: A Reappraisal of the Postsecondary Dropout Phenomenon," *Economics of Education Review*, 8, 305–312.

Manski, C.F. (1993a): "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?," in Clotfelter, C.T. and M. Rothschild (eds) *Studies of Supply and Demand in Higher Education*, NBER Research Report, Chicago: University of Chicago Press, 43–60.

Manski, C.F. (1993b): "Dynamic Choice in Social Settings: Learning from the Experience of Others," *Journal of Econometrics*, 58, 121–136.

Moffitt R.A. and P. Gottschalk (2002): "Trends in the Transitory Variance of Earnings in the United States," *Economic Journal*, **112**, C68–C73.

Rosen, S. (1976): "A Theory of Life Earnings," *Journal of Political Economy*, 84, S45–S68.

Rothschild, M. and J. Stiglitz (1970): "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 3, 66–84.

Rothschild, M. and J. Stiglitz (1972): "Addendum to 'Increasing Risk: I. A Definition'," *Journal of Economic Theory*, 5, 306.

Swenson, C.W. (1997): "Rational Expectations and Tax Policy: Experimental Market Evidence," *Journal of Economic Behavior and Organization*, 32(3), 433–55.

Taber, C. (2000): "Semiparametric Identification and Heterogeneity in Discrete Choice Dynamic Programming Models," *Journal of Econometrics*, 96, 201–29.

Taber, C. (2001): "The Rising College Premium in the Eighties: Return to College or Return to Ability?" *Review of Economic Studies*, 68, 665–91.

Timmermann, A. (1994): "Can Agents Learn to Form Rational Expectations? Some Results on Convergence and Stability of Learning in the UK Stock Market," *The Economic Journal*, 104, 777–797.

Willis, R.J. (1986): "Wage Determinants: A Survey and Reinterpretation of Human Capital Functions," *Handbook of Labor Economics*, Vol. I, 525–602.

Willis, R.J. and S. Rosen (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, S7–S36.

Wolpin, K. (1987): "Estimating a Structural Search Model: The Transition from School to Work," *Econometrica*, 55(4), 801–17.

Zagorsky, J.L. (1997): "The NLSY79 Wealth Data Evaluation." Columbus, OH: CHRR, The Ohio State University.

Zagorsky, J. L. (1999) "Young Baby Boomers' Wealth," *Review of Income and Wealth*, 45, 135–56.

Table 1: Composition of the March-CPS Extract

Year	All		14-18		19-24		25+	
	Total	% in School	Total	% in School	Total	% in School	Total	% in School
1964	17,599	15.6	2,646	88.1	1,898	19.9	13,055	0.3
1968	39,901	14.3	5,907	79.7	4,624	19.5	29,370	0.3
1975	35,759	14.1	5,350	75.8	5,336	15.4	25,073	0.6
1979	44,095	14.4	6,534	76.6	7,038	15.5	30,523	0.9
1980	52,285	14.2	7,674	75.8	8,421	15.2	36,190	0.9
1981	51,621	14.3	7,264	78.9	8,125	15.9	36,232	1.0
1984	44,536	15.3	5,840	85.6	6,659	20.7	32,037	1.4
1987	43,663	13.5	5,437	83.2	5,933	17.6	32,293	1.1
1990	44,029	9.7	3,925	76.0	5,883	16.6	34,221	0.9
1992	42,848	10.8	3,880	81.8	5,542	19.7	33,426	1.1
1994	39,698	10.1	3,386	82.6	4,868	18.9	31,444	1.0

(Age Group)

Note: For definition of "in school" individuals see the text.

Table 2: Return to Eductaion for Mid-Career White Males

Year	Whites .10	.25	.50	.75	.90	Non-Whites .10	.25	.50	.75	.90
	High School Graduates					College Graduates				
1964	9.42	9.32	8.50	8.04	7.91	2.38	5.44	6.74	8.07	9.76
	(2.73)	(0.79)	(0.33)	(0.12)	(0.18)	(3.30)	(0.93)	(0.39)	(0.12)	(0.26)
1968	10.67	10.76	9.53	8.87	9.02	7.71	8.88	9.31	10.16	12.45
	(0.71)	(0.37)	(0.12)	(0.04)	(0.31)	(0.83)	(0.42)	(0.13)	(0.06)	(0.42)
1973	10.48	10.43	9.70	9.06	9.00	8.31	9.14	9.58	10.54	12.78
	(0.73)	(0.32)	(0.15)	(0.06)	(0.40)	(0.84)	(0.36)	(0.15)	(0.09)	(0.53)
1976	9.98	10.88	10.02	9.37	9.29	8.50	9.39	9.22	9.99	11.77
	(0.57)	(0.34)	(0.19)	(0.04)	(0.23)	(0.64)	(0.39)	(0.21)	(0.03)	(0.33)
1979	9.28	10.27	9.55	9.01	8.67	7.64	8.75	8.28	8.80	10.40
	(0.60)	(0.29)	(0.25)	(0.12)	(0.11)	(0.69)	(0.33)	(0.29)	(0.13)	(0.16)
1981	8.71	9.75	9.33	8.99	8.50	7.62	8.35	8.16	9.18	8.87
	(0.41)	(0.29)	(0.19)	(0.07)	(0.06)	(0.45)	(0.32)	(0.22)	(0.07)	(0.05)
1985	9.63	11.03	10.80	10.45	9.66	11.18	10.88	10.77	11.14	12.68
	(0.12)	(0.22)	(0.14)	(0.06)	(0.26)	(0.18)	(0.23)	(0.14)	(0.05)	(0.36)
1988	9.95	11.70	11.54	11.34	10.77	10.97	11.25	11.22	11.70	14.10
	(0.13)	(0.25)	(0.18)	(0.11)	(0.27)	(0.12)	(0.26)	(0.19)	(0.10)	(0.37)
1992	10.36	11.96	12.08	11.79	11.17	11.61	12.19	12.54	12.64	12.09
	(0.13)	(0.15)	(0.10)	(0.05)	(0.07)	(0.14)	(0.15)	(0.09)	(0.05)	(0.06)
1994	11.55	13.25	13.33	13.47	13.16	13.22	15.40	15.49	16.22	15.85
	(0.15)	(0.13)	(0.10)	(0.14)	(0.16)	(0.21)	(0.20)	(0.15)	(0.21)	(0.24)

Note: All returns to education are expressed in percentage.

Table 3: Simulated Distributions of Education Accumulation (after 20 years) **with $10,000 Initial Wealth**

	Cohort	α	Years of Education 12	13	14	15	16	17	18	19	20
			No Foresight								
1.	1980	0.8	0	0	0	0	13	378	2,885	5,324	1,400
2.	1985	0.8	0	0	0	0	0	24	925	5,275	3,776
3.	1990	0.8	0	0	0	0	0	3	421	4,013	5,563
4.	1980	1.2	0	1	7	165	1,173	4,111	3,583	921	39
5.	1985	1.2	0	0	0	20	252	1,831	4,540	3,041	316
6.	1990	1.2	0	0	0	6	73	805	3,426	4,488	1,202
7.	1980	2.0	109	633	2,066	3,913	2,431	779	66	3	0
8.	1985	2.0	6	151	770	2,473	3,281	2,587	671	61	0
9.	1990	2.0	0	41	284	1,381	2,837	3,421	1,686	338	12
			Perfect Foresight								
10.	1980	0.8	0	0	0	0	1	204	2,285	4,064	3,446
11.	1985	0.8	0	0	0	0	8	241	2,306	4,081	3,364
12.	1990	0.8	0	0	0	0	21	348	2,655	3,863	3,113
13.	1980	1.2	0	0	0	14	296	2,706	4,726	1,912	346
14.	1985	1.2	0	0	0	13	305	2,794	4,509	1,939	440
15.	1990	1.2	0	0	0	22	452	2,991	4,222	1,835	478
16.	1980	2.0	0	63	579	2,042	3,577	2,966	704	68	1
17.	1985	2.0	0	50	555	1,973	3,370	3,131	830	87	4
18.	1990	2.0	0	56	637	2,086	3,281	2,958	878	98	6
			VAR Model								
19.	1980	0.8	0	0	0	1	73	1,314	4,557	3,133	922
20.	1985	0.8	0	0	0	0	20	500	3,179	3,973	2,328
21.	1990	0.8	0	0	0	0	28	420	2,967	3,792	2,793
22.	1980	1.2	0	1	37	438	2,170	4,686	2,314	337	17
23.	1985	1.2	0	0	2	92	895	4,078	3,777	1,028	128
24.	1990	1.2	0	0	3	50	652	3,376	4,073	1,499	347
25.	1980	2.0	54	652	2,395	3,771	2,303	747	77	1	0
26.	1985	2.0	9	299	1,451	3,139	3,037	1,774	281	10	0
27.	1990	2.0	1	100	816	2,393	3,307	2,631	687	61	4
			VAR-Gibbs Model								
28.	1980	0.8	0	0	1	168	1,561	4,351	3,356	563	0
29.	1985	0.8	0	0	0	1	45	779	3,548	4,315	1,312
30.	1990	0.8	0	0	0	0	58	828	3,669	4,256	1,189
31.	1980	1.2	0	1	66	752	3,001	4,026	1,844	304	6
32.	1985	1.2	0	0	15	276	1,772	4,009	3,101	784	43
33.	1990	1.2	0	0	7	208	1,466	3,874	3,266	1,091	88
34.	1980	2.0	181	1169	3,117	3,538	1,641	321	33	0	0
35.	1985	2.0	49	542	2,061	3,770	2,581	870	119	8	0
36.	1990	2.0	14	374	1,856	3,699	2,826	1,022	198	11	0

Table 4: Speed of Education Accumulation for High School Graduates, with $10,000 Initial Wealth

			Years to Completion of College											
	Cohort	α	≤ 5	6	7	8	9	10	11	12	13	14	15	16+
			No Foresight											
1.	1980	0.8	18	176	795	1,542	1,716	1,823	1,553	1,179	621	367	132	78
2.	1985	0.8	88	1,077	2,375	2,232	2,255	1,185	453	226	80	22	6	1
3.	1990	0.8	367	2,465	3,202	2,270	1,128	395	126	31	11	5	0	0
4.	1980	1.2	0	0	13	106	265	580	818	1,335	1,162	1,444	1,315	2,790
5.	1985	1.2	0	7	120	419	1,207	1,650	1,781	1,666	1,221	813	495	601
6.	1990	1.2	0	82	605	1,401	1,981	1,961	1,577	1,054	618	343	189	183
7.	1980	2.0	0	0	0	0	0	4	13	31	43	111	207	2,870
8.	1985	2.0	0	0	0	2	21	71	153	273	442	610	779	5,249
9.	1990	2.0	0	0	2	31	94	260	420	663	785	898	1,091	3,050
			Perfect Foresight											
10.	1980	0.8	280	1,667	2,650	2,388	1,501	800	372	225	68	28	14	7
11.	1985	0.8	208	1,728	3,084	2,374	1,404	683	302	119	59	25	7	7
12.	1990	0.8	465	1,916	2,756	2,279	1,362	683	302	134	58	28	12	5
13.	1980	1.2	0	69	445	1,144	1,522	1,666	1,509	1,598	896	514	320	303
14.	1985	1.2	0	130	801	1,352	1,849	1,863	1,324	1,009	742	411	241	265
15.	1990	1.2	18	112	685	1,335	1,667	1,720	1,425	1,095	750	443	297	431
16.	1980	2.0	0	0	2	22	50	176	271	677	754	814	948	3,602
17.	1985	2.0	0	0	10	54	159	317	360	575	688	792	896	3,571
18.	1990	2.0	0	0	3	36	91	218	321	471	588	690	744	4,059
			VAR Model											
19.	1980	0.8	18	175	741	1,491	1,640	1,884	1,726	1,157	587	311	156	113
20.	1985	0.8	88	1,028	2,379	2,196	2,037	1,199	576	280	127	47	29	14
21.	1990	0.8	120	1,574	2,619	2,471	1,661	842	389	185	84	35	15	5
22.	1980	1.2	0	0	7	75	153	520	794	1,307	1,081	1,267	1,262	3,058
23.	1985	1.2	0	7	161	468	951	1,421	1,402	1,465	1,302	936	743	1,050
24.	1990	1.2	0	43	399	943	1,406	1,636	1,522	1,287	937	606	449	719
25.	1980	2.0	0	0	0	0	4	11	28	72	74	162	236	2,541
26.	1985	2.0	0	0	0	3	21	60	88	205	305	435	548	3,437
27.	1990	2.0	0	0	1	19	53	136	219	373	484	614	673	4,118
			VAR-Gibbs Model											
28.	1980	0.8	2	118	576	1,342	1,591	1,756	1,852	1,376	705	368	189	124
29.	1985	0.8	88	888	2,425	2,251	2,085	1,211	568	267	125	48	31	12
30.	1990	0.8	120	1,611	2,569	2,476	1,621	854	404	196	82	39	18	10
31.	1980	1.2	0	0	6	60	142	415	660	1,218	1,039	1,345	1,238	3,058
32.	1985	1.2	0	7	136	436	972	1,345	1,293	1,340	1,179	897	738	1,366
33.	1990	1.2	0	52	357	829	1,239	1,527	1,383	1,194	956	686	570	992
34.	1980	2.0	0	0	0	0	0	6	12	41	44	106	137	1,649
35.	1985	2.0	0	0	0	1	22	40	74	153	216	277	347	2,448
36.	1990	2.0	0	0	1	11	30	55	115	188	239	367	387	2,664

Table 5: Average Education Accumulation for High School Graduates with Tuition Costs (VAR-Gibbs Mehtod with $\alpha = 1.2$)

Initial Wealth	Cohort	Years of Education 12	13	14	15	16	17	18	19	20
$10,000	1980	0	57	509	2,186	4,037	2,460	684	67	0
$10,000	1985	0	16	366	1,752	3,640	3,019	1,137	114	0
$10,000	1990	0	73	622	2,643	3,654	2,239	689	36	0

Table 6: Average Education Accumulation for High School Graduates with Various Initial Wealth Values (VAR-Gibbs Mehtod with $\alpha = 1.2$)

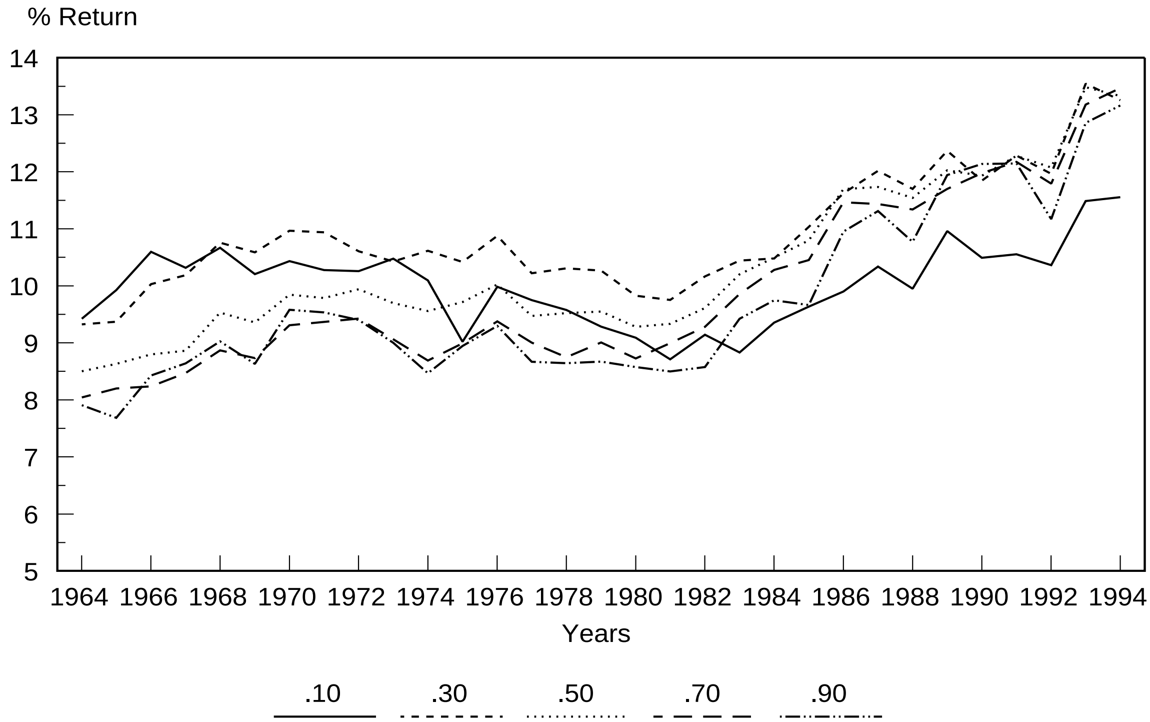
Initial Wealth	Cohort	Years of Education 12	13	14	15	16	17	18	19	20
$5,000	1980	0	11	203	1,386	3,715	3,451	1,098	135	1
$5,000	1985	0	1	30	485	2,411	4,047	2,486	522	18
$5,000	1990	0	0	35	511	2,371	4,003	2,434	615	31
$10,000	1980	0	57	509	2,186	4,037	2,460	684	67	0
$10,000	1985	0	16	366	1,752	3,640	3,019	1,137	114	0
$10,000	1990	0	73	622	2,643	3,654	2,239	689	36	0
$20,000	1980	0	0	7	240	1,764	4,183	2,982	786	38
$20,000	1985	0	0	0	66	772	3,329	4,076	1,591	166
$20,000	1990	0	0	0	52	727	3,130	4,002	1,875	214
$50,000	1980	0	0	0	5	280	2,138	4,505	2,688	384
$50,000	1985	0	0	0	0	58	1,013	3,846	4,025	1,058
$50,000	1990	0	0	0	0	36	844	3,519	4,265	1,336

Table 7: Pearson Within-Sample Statistic for Altenative Specifications For the 1980 Cohort, with VAR-Gibbs Method, $\alpha = 1.2$

Model	Education Level	Fraction in School
Base	0.089	2.320
Hetrogeneity I	0.077	2.448
Hetrogeneity II	0.106	3.245
Hetrogeneity III	0.338	1.478
State Dependence	0.118	2.301

Figure 1: Return to Education for Mid Career White Males

a. High School Graduates



b. College Graduates

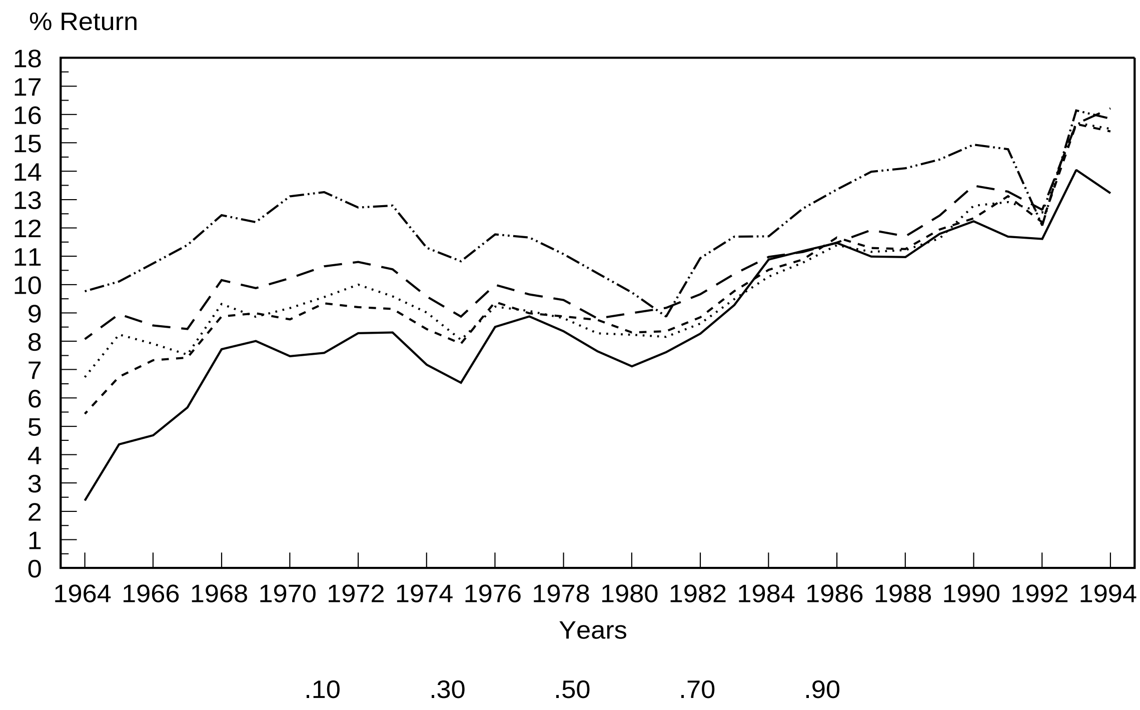
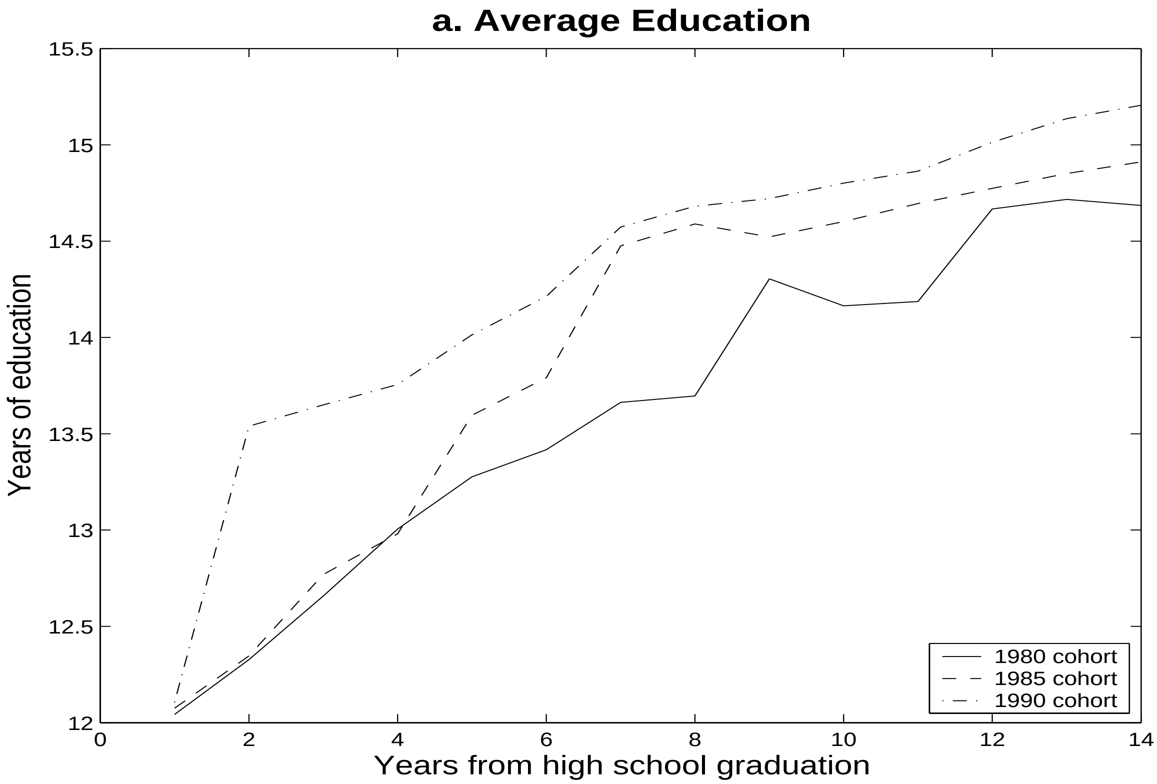


Figure 2: Actual Education Attainments by Cohort
(Source: March CPS)



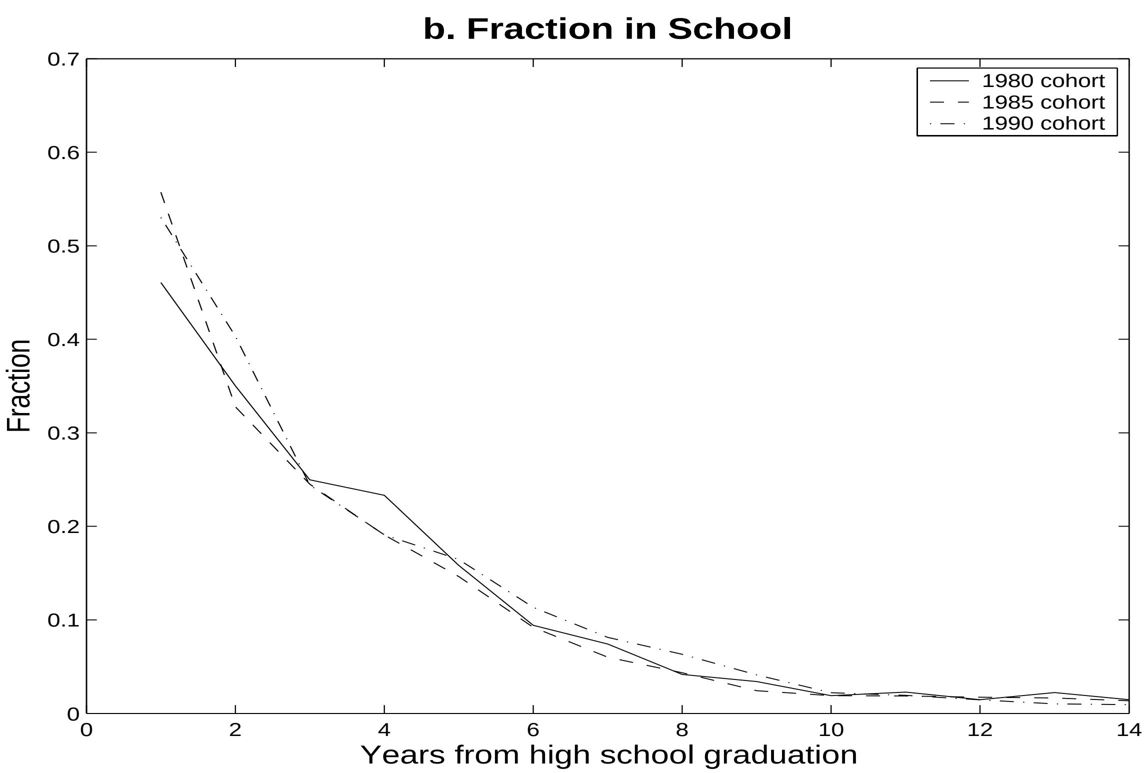


Figure 3: Optimal Choices of Education by Forecasting Method and Starting Year

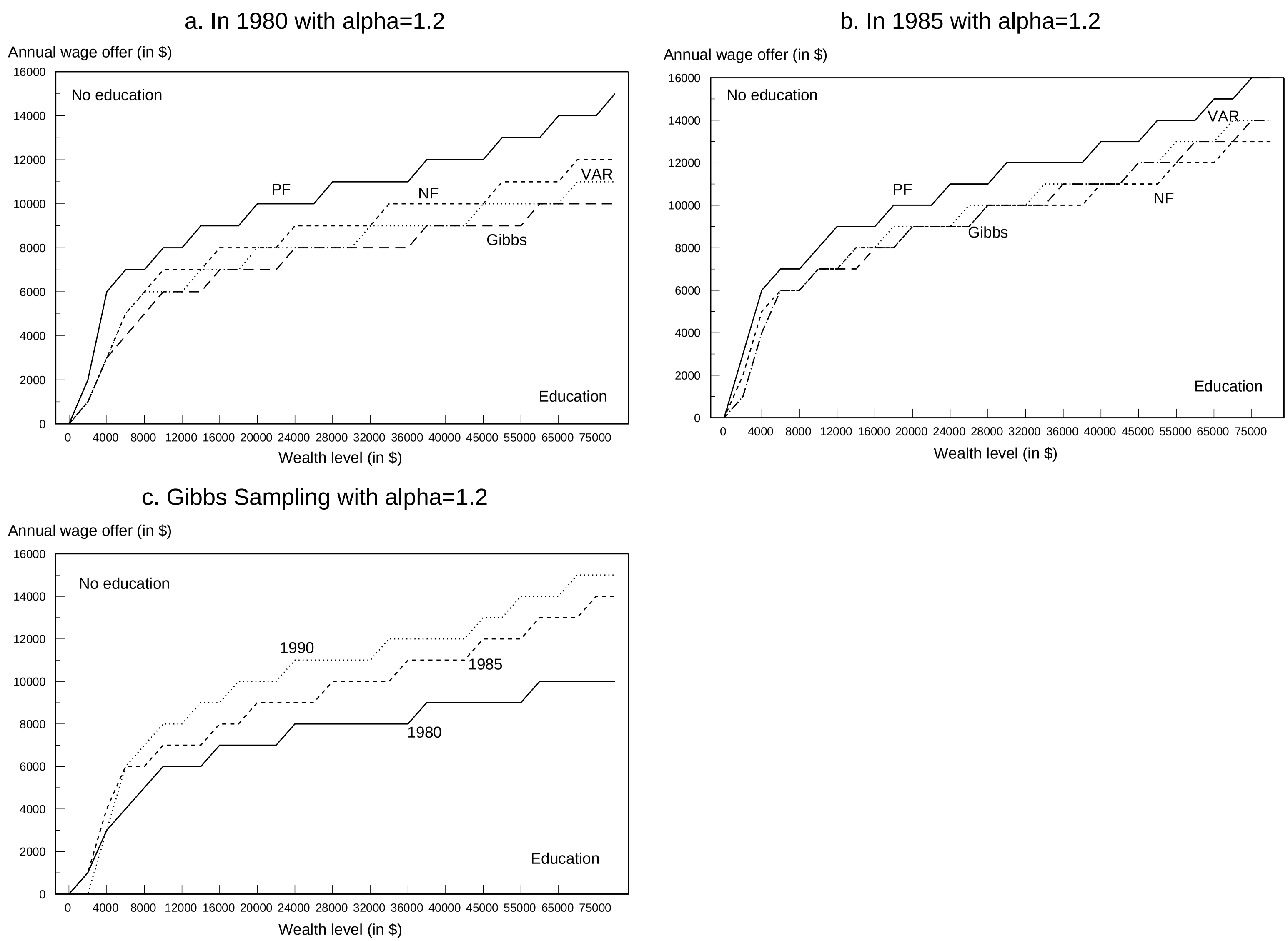


Figure 4: Fraction of High School Graduates Going to School by Forecasting Method and Starting Years

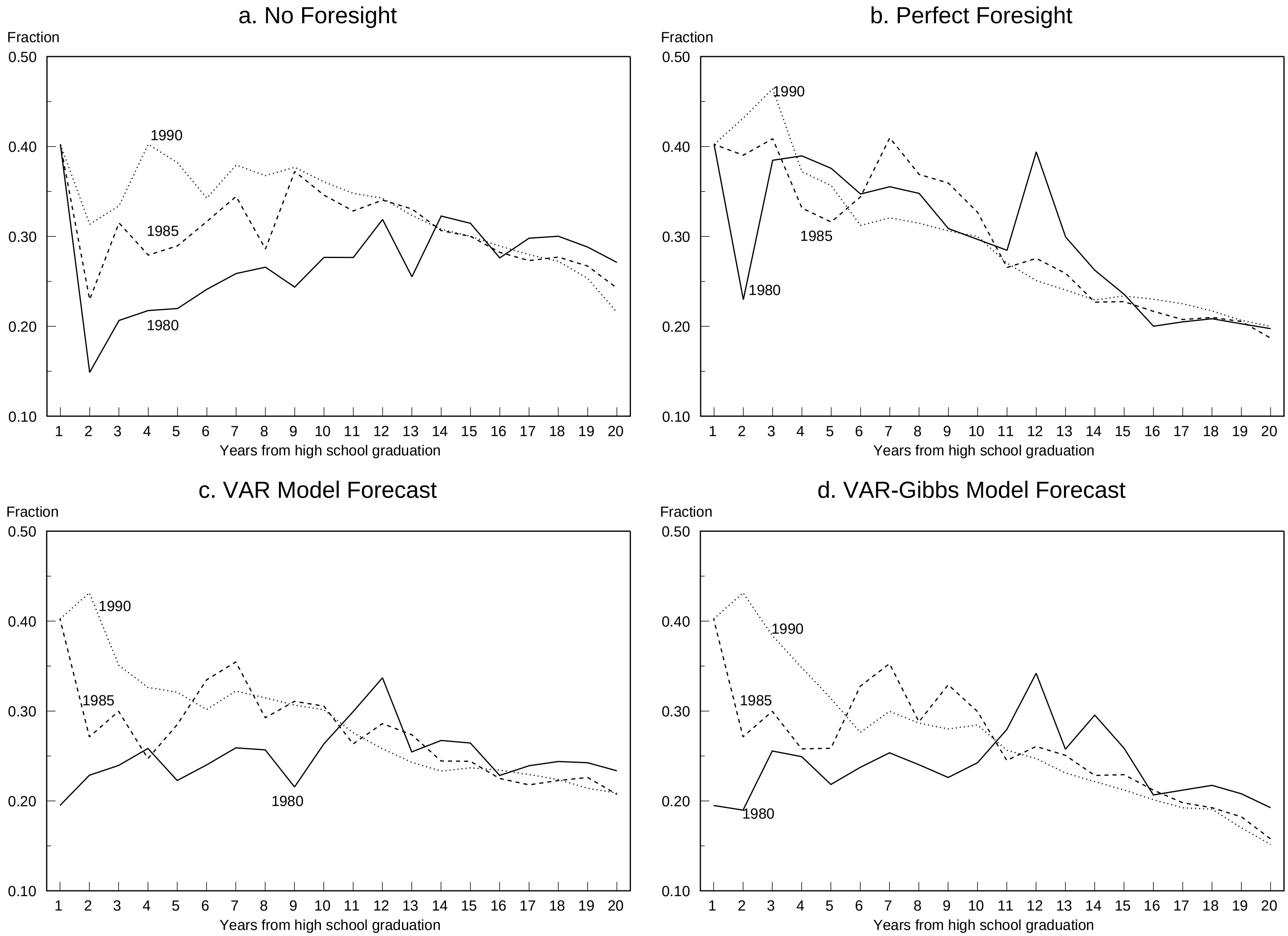


Figure 5: Accumulated Education of High School Graduates by Forecasting Method and Starting Years

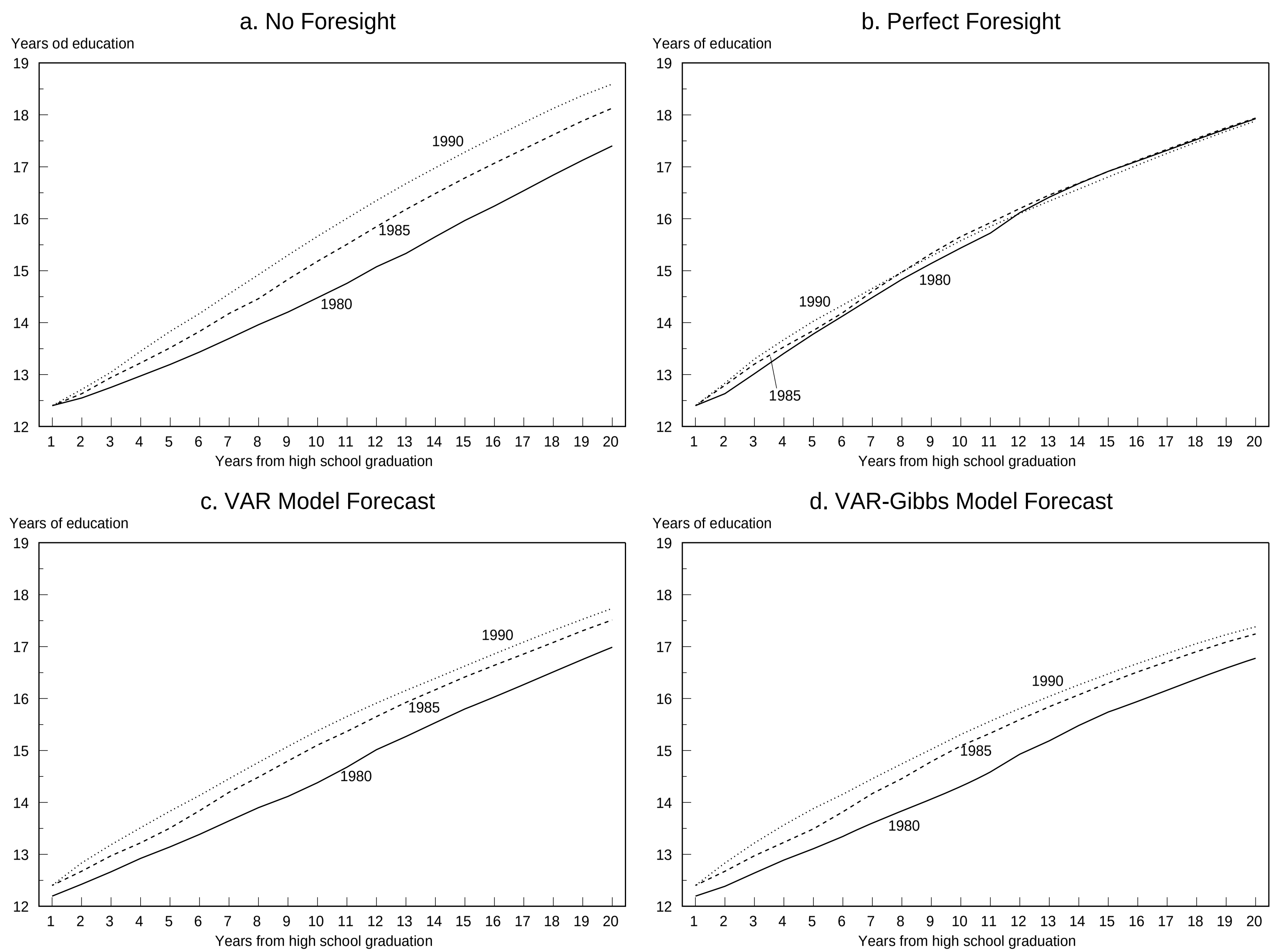


Figure 6: Average Annual Wage of High School Graduates by Forecasting Method and Starting Years

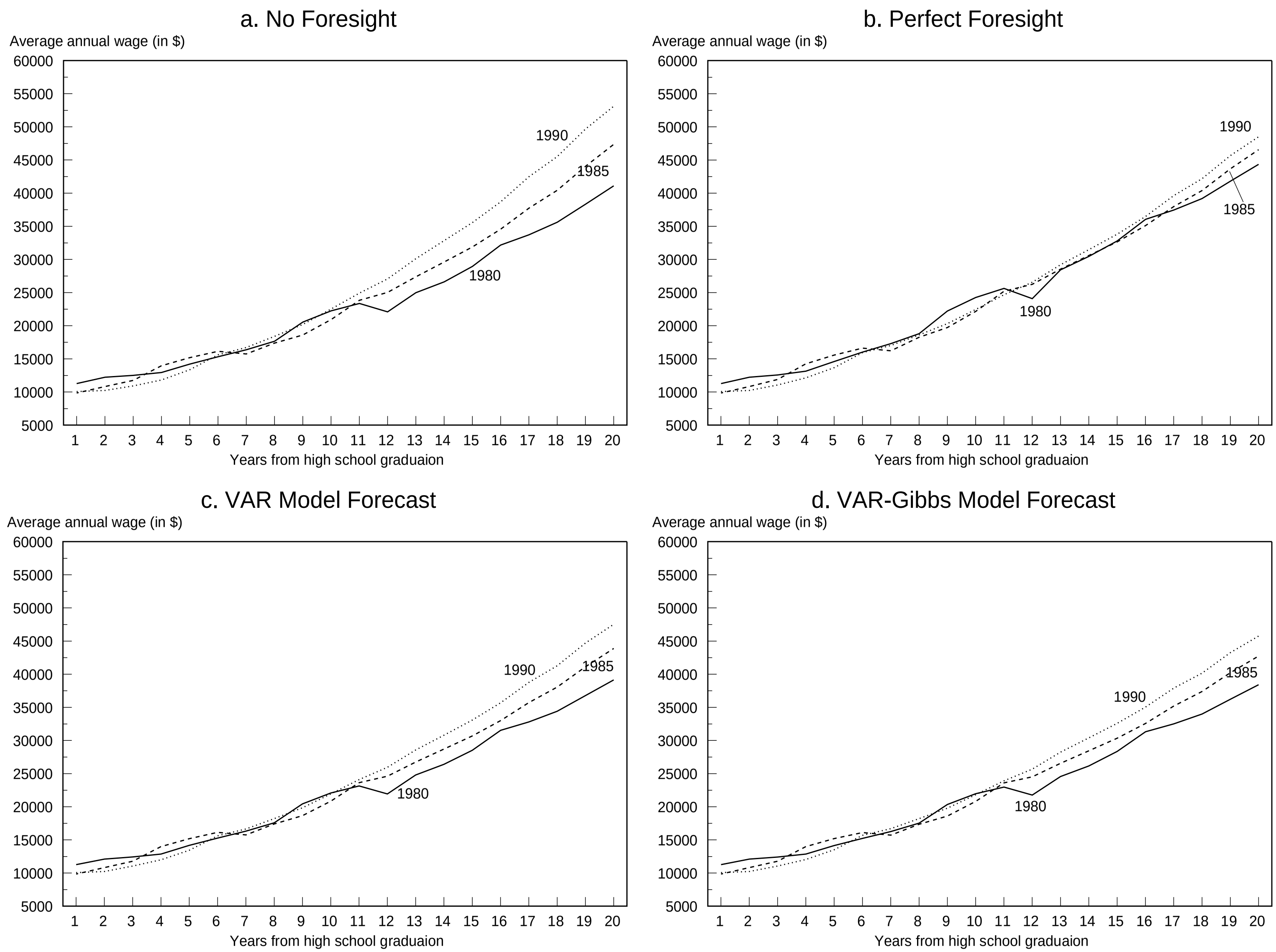
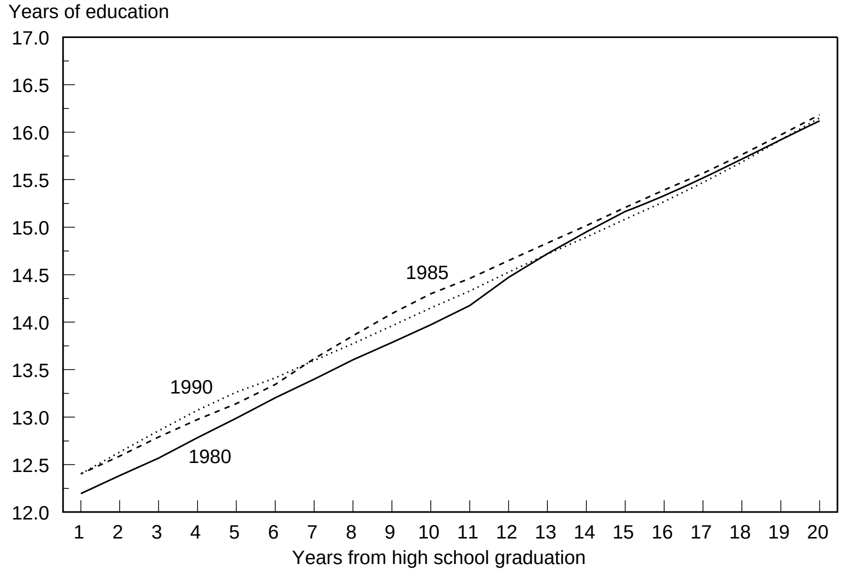


Figure 7: Average Accumulated Education for High School Graduate, by Starting Years ($\alpha = 2.0$)

a. Perfect Foresight



b. Var-Gibbs Forecast

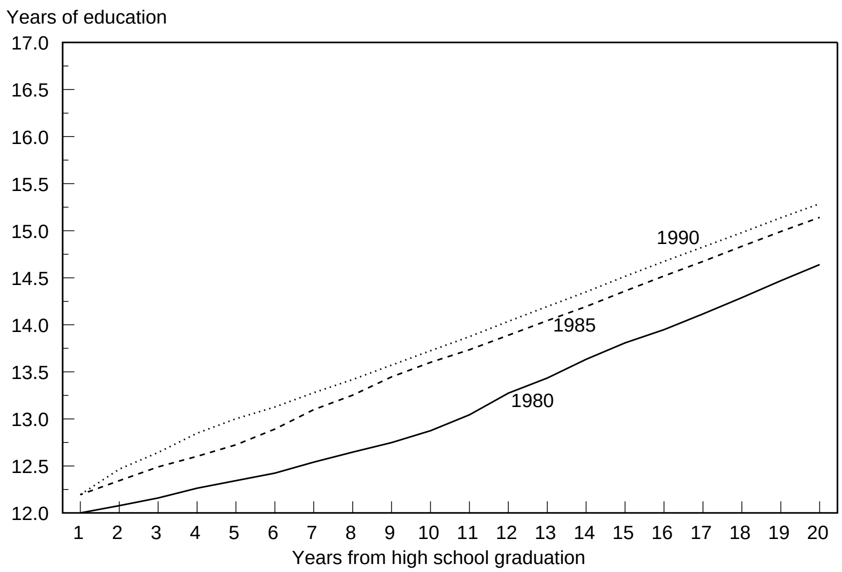


Figure 8: Tuition Costs for National Average for Colleges and State Universities

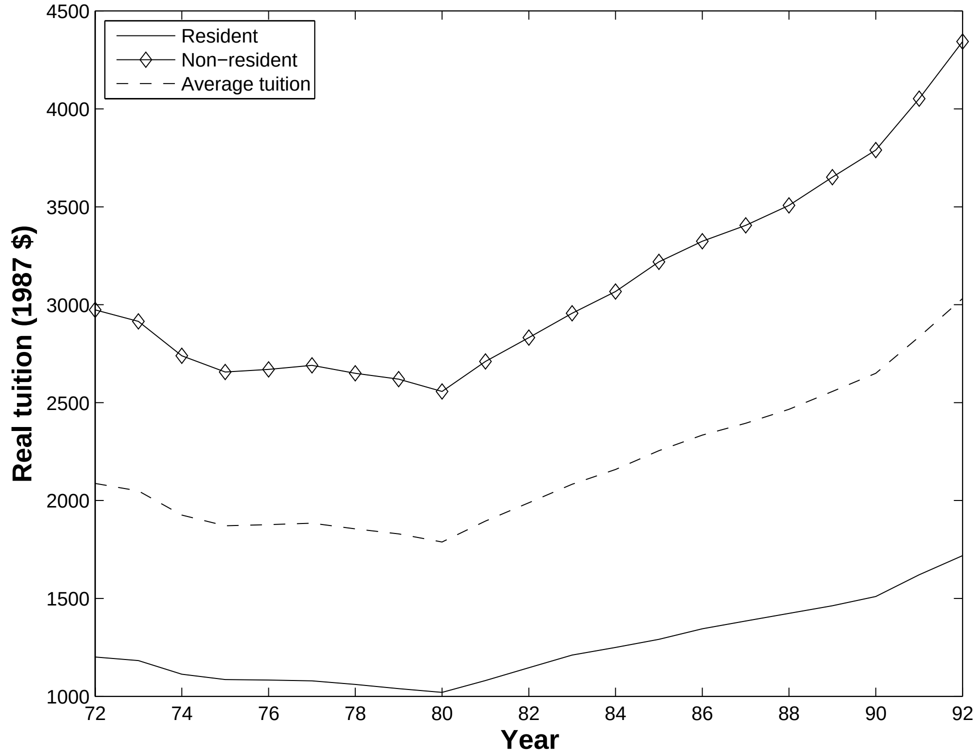


Figure 9: Simulation Results with Tuition Costs for VAR-Gibbs Model by Starting Years ($\alpha = 1.2$)

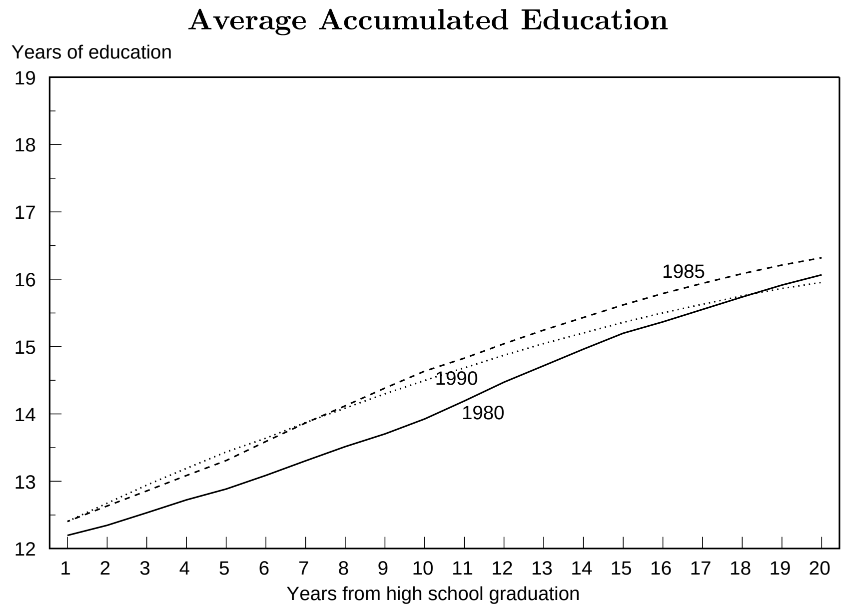
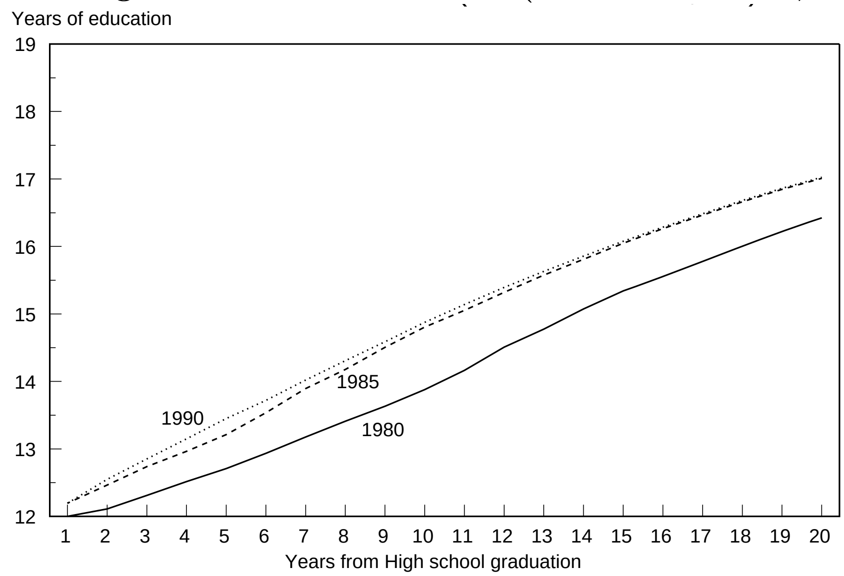


Figure 10: Simulations Results with Varying Initial Wealth for VAR-Gibbs Model by Starting Years ($\alpha = 1.2$)

a. Average Accumulated Education (initial wealth =$5,000)



b. Average Accumulated Education (initial wealth =$50,000)

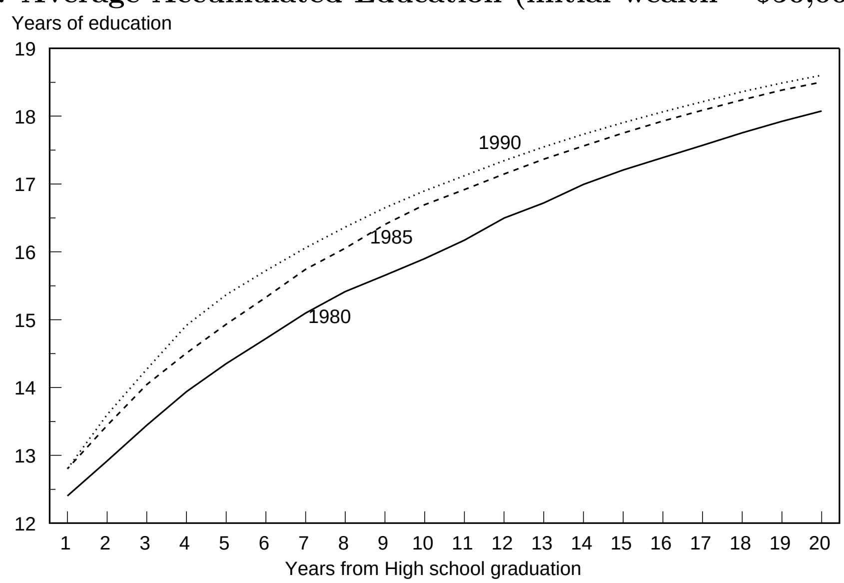
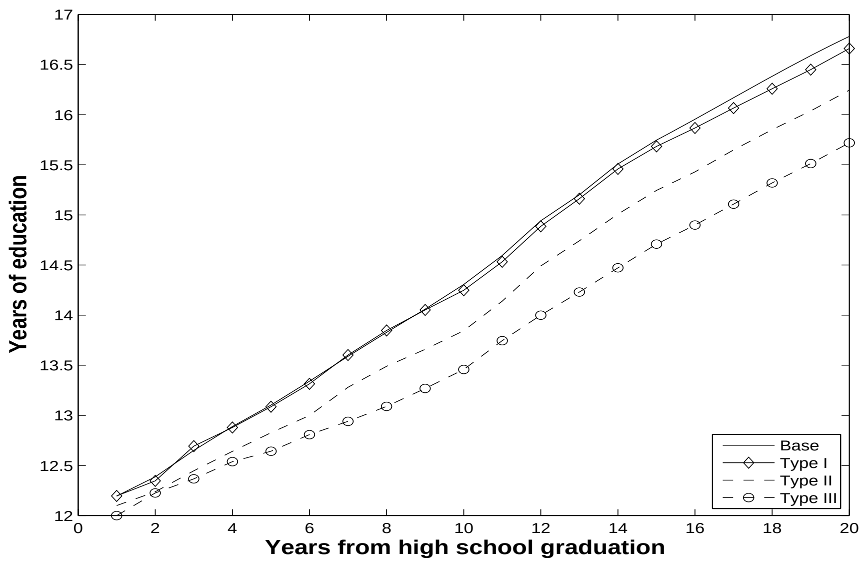


Figure 11: Simulation Results for Average Education Level for 1980 Cohort VAR-Gibbs Method with $\alpha = 1.2$

a. Unobserved Heterogeneity



b. State Dependence in Wages