INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2007-14

The Structural Change and the Endogeneity Bias of the College Premium in the United States 1968-2001

F. MURTIN[1]

---

[1] Paris-Jourdan Sciences Economiques (joint research unit CNRS-ENS-EHESS-ENPC), CREST-INSEE and Centre for the Economics of Education (London School of Economics). fabrice_murtin@hotmail.com
– I thank Francis Kramarz, Thierry Kamionka, Thomas Piketty, Jean-Marc Robin, Bernard Salanié and seminar participants at CREST (INSEE), Paris-Jourdan Sciences Economiques and University Paris I for helpful discussions.

Abstract

This paper derives and tests a procedure to estimate a general class of panel data models that display a change of regime, as well as potential correlation between specific effects and explanatory variables. This model detects a structural change for all cohorts on the US labor market in the early 80s, and shows that some unobserved skills correlated with schooling have become rewarded in the second regime. As a result, I quantify the bias in the college premium when it is calculated in cross-section: it amounts to circa 30% of the cross-sectional estimates after 1980. Interestingly, this bias is carried almost entirely by the two most educated cohorts of individuals. A simple model where skilled individuals have a comparative advantage to graduate from college can account for this fact, since it implies that the correlation between unobserved ability and college graduation increases with overall college graduation.

Résumé

Cet article introduit et teste une procédure d'estimation de modèles à changements de régime et à effets aléatoires corrélés sur données de panel. Ce modèle détecte un changement structurel pour toutes les cohortes présentes sur le marché du travail américain au début des années 1980, et montre que des aptitudes inobservées corrélées à l'éducation sont devenues rémunérées dans le second régime. Ainsi, l'article quantifie le biais d'endogénéité associé au diplôme universitaire lorsque l'estimation est conduite sur des données en coupe: ce biais est égal à environ +30% du rendement du diplôme après 1980. Il est porté presque entièrement par les deux cohortes les plus éduquées. Ce dernier fait est expliqué par un modèle simple de sélection où les individus talentueux ont un avantage comparatif dans l'obtention d'un diplôme universitaire, et où la corrélation entre aptitudes inobservées et diplôme universitaire augmente avec la proportion de diplômés de l'université au sein de chaque cohorte.

JEL Classification: C11, C15, C23, C24, C34, J24, J31, O15.

Keywords: MCMC, Regime-switching, College premium, Inequality.

1 Introduction

There has been a large body of research focusing on the increase in inequality in the United-States over the last two decades. This trend as well as some correlated events have been recently surveyed in Eckstein and Nagypal (2004), and discussed in several others. Indeed, the role of skill-biased technological change as a unifying explanation for the increase in within and between-groups inequality has been challenged in the recent past. Measurement problems in the CPS have been pointed at by Card and DiNardo (2002), the differential behaviour of the bottom and the top halves of the income distribution has been underlined by Autor-Katz-Kearney (2006), Lemieux (2006a) has shown the crucial role of composition effects in the increase of within-groups inequality, while Piketty-Saez (2002) and Lemieux (2006b) have emphasized that most of overall inequality was concentrated in the top of the income distribution.

Most of the literature mentionned above is based on cross-sectional Mincerian regressions, pointing at the modification of the return to observed and unobserved skills across years. Nevertheless, it is possible that the rise in the return to schooling may be due to modifications of unobserved determinants that are correlated with education: this rise might be governed by endogenous effects. Focusing on inequality between educational groups, this paper shows that the rise in the college premium occuring from the 80s has been smaller than that depicted in the literature, and it proposes new estimates of the college premium.

Some studies have already tried to gauge this endogeneity bias. On one hand, Taber (2001) uses a dynamic selection model and finds a strong role of unobserved ability in the increase in the college premium. On the other hand, Chay and Lee (2000) have shown that the rise in the return to unobserved ability could explain at most 30-40% of the rise in the college premium. They use the time variations of earnings heteroskedasticity across educational groups to derive an *upper bound* for the contribution of unobserved ability. Deschênes (2006) finds compatible results, using the rise in the convexity of the earnings-schooling profile as a source of identification. In order to do so, he assumes that there is no heterogeneity in the convexity of the latter function[1]. Also, these two studies are not informative on the

[1] Also, the reconstruction of the years of schooling variable from the degrees variable

nature of unobserved skills and do not mention how the latter are distributed across cohorts.

In contrast, this paper uses the simple strategy of computing explicitely the unobserved heterogeneity variable using the longitudinal dimension. Identification stems from the changing of educational attainment of workers along their life cycle. Then, I run cross-sectional regressions including the estimated unobserved ability variable, which purges the coefficient of the college premium from any endogeneity bias. The major problem likely to appear in this procedure is that earnings dynamics have been non-stationnary across the past thirty years as demonstrated by Meghir-Pistaferri (2004). As a result, a structural change in the US wages' dynamics is a phenomenom that one would like to account for, within a correlated random effects model.

I present a general empirical framework where all coefficients as well as the variance of the residuals are allowed to vary across two different regimes governed by an unobserved variable, which follows an hidden Markov Chain. Though potentially richer, this regime switching structure is used to capture the structural change. Markov-Chain Monte-Carlo methods have recently been developped for numerous models[2], and are of particular interest when the likelihood of the model is complicated and/or leads to numerical unfeasibility. In the present case, this approach is much easier to implement than maximum likelihood methods, which would rely on unaccurate two-steps methods.

The paper develops an estimation procedure for both uncensored and censored data. It is tested and validated on different simulated datasets, then is applied to wage dynamics in the United-States from 1968 to 2000, using the PSID (SRC) database. The estimation is conducted on 5-years cohorts of age and shows that there has been a change of regime in the early 80s, and none since that date. For two cohorts in particular - the most educated - unobserved ability is significatively and positively correlated with education, leading to a 30% upward bias in the annual increase in the college premium after 1980.

taken from the CPS could be a source of measurement errors. Using IUPMS Census data and a detailed variable for grades achieved, Morrisson-Murtin (2007) find that the earnings-schooling profile is concave over the whole period 1940-1980, while Deschênes (2006) finds a convex earnings function in 1980. The detailed grades variable is not available in subsequent IUPMS surveys.

[2]see Chib (2001) for a review

More precisely, four conclusions can be derived: (i) the major part of the observed rise in the cross-sectional return to schooling reflects an increase in the *causal* effect of schooling on earnings; (ii) the correlation of unobserved ability across regimes is about 0.5 for any cohort having experienced the structural change, implying that unobserved ability is multi-dimensional - otherwise this correlation would be equal to one; (iii) whatever the cohort, sorting of unobserved ability and schooling is null in the first regime, indicating that the unobserved skills correlated with schooling are exclusively rewarded in the second regime; (iv) sorting of unobserved ability and schooling is higher among most educated cohorts, a fact that is explained by a simple model of imperfect sorting where skilled individuals have a comparative advantage to graduate from college.

Although it cannot provide a representative picture of the US labor market, this paper integrates well in the recent literature on inequality. It puts the emphasis on the multi-dimensionality of productive skills, and argues that the organizational changes occurring in the early 80s have increased the value of some unobserved skills refered to as “managerial skills”. The paper finds an endogeneity bias in the college premium in tune with the literature, but also explains why this bias takes place among the most educated cohorts.

Section 1 presents the statisical model and section 2 the Bayesian estimation, which is tested in section 3. Section 4 introduces the treatment of censoring. Section 5 presents the results for the college premium and last section concludes.

2 The model

The most general regime switching model is considered:

$$\begin{aligned} y_{i,t} &= x_{i,t}\beta_{s_t} + z_{i,t}b_{i,s_t} + \sigma_{s_t}u_{i,t}, \qquad i \leq N, t \leq T \qquad (1) \\ b_{i,s_t} &\rightsquigarrow \mathcal{N}_L\left(0, D_{s_t}\right) \\ u_{i,t} &\rightsquigarrow \mathcal{N}\left(0, 1\right) \end{aligned}$$

where $x_{i,t}$ is a vector of observations of dimension 1xK, $z_{i,t}$ a vector of observations of dimension 1xL, $u_{i,t}$ is a strictly exogenous iid white noise, and s_t is a two-states Markov chain starting from its invariant distribution with transition probabilities $\pi = (\pi_{k,l})_{k,l\in\{0,1\}}$

$$\mathrm{P}\ (s_t = l|s_{t-1} = k) = \pi_{k,l} \tag{2}$$

From a bayesian point of view parameters $\beta_{s_t}, \sigma_{s_t}, D_{s_t}, \pi$ are assumed to be random variables with independant prior distributions as it is detailed below. The random variables b_{i,s_t} are assumed to be iid, and given D_{s_t}, conditionally independant from $u_{i,t}, \beta_{s_t}, \sigma_{s_t}, \pi$.

In a basic specification that excludes regime switching, this kind of structure is called "mixed models" in the statistical literature[3]. They differ from the traditional random-coefficient model introduced by Swamy (1970) or Hsiao (1974) because they do not assume orthogonality between explanatory variables and specific effects (here the b_{i,s_t} variables).

All distributions are allowed to vary between the two regimes, including the random-effects b_{i,s_t}. From an economic perspective, it can be interpretated as a constant level of unobserved ability for each individual, but with a time-varying return across regimes; as a consequence the distributions of random-effects, including their variance, might differ across both cohorts and regimes. A larger number of states can be considered[4], but given the small panel length T, this could make estimation harder. Also this model is motivated by its empirical application and it is very likely that a single structural change has occurred over the last thirty years. At this stage the model is not identified because of invariance by permutation of the regimes, so that an additional identification constraint $\sigma_0 < \sigma_1$ is set.

The density of observations is gaussian conditionally on the regime and on the corresponding parameters

$$f(y_{i,t}|s_t = k, \theta_k) \rightsquigarrow \mathcal{N}\left(x_{i,t}\beta_k + z_{i,t}b_{i,k}; \sigma_k^2\right) \tag{3}$$

where $\theta_k = (\beta_k, b_{i,k}, \sigma_k, D_k, \pi_k)$. Bayesian estimation treats the parameters of interest $\Theta = (\theta_0, \theta_1)$ as random variables, and aims at inferring the posterior distribution of parameters conditionally on the data $p(\Theta, s_1...s_T|Y)$.

Modelling the unconditional prior distribution of parameters is an important step since improper priors generally lead to ill-defined posterior distri-

[3] in the econometric literature, "fixed-effects models" are a particular case.

[4] see for instance Chopin-Pelgrin (2004).

butions[5]. The following priors are natural because they enable closed-form expressions of the posterior distribution:

$$\begin{aligned} \beta_k &\rightsquigarrow \mathcal{N}_K\left(\beta^0, B_0\right) \\ b_{i,k}|D_k &\rightsquigarrow \mathcal{N}_L\left(0, D_k\right) \\ 1/\sigma_k^2 &\rightsquigarrow \mathcal{G}\left(\frac{\nu_0}{2};\frac{\delta_0}{2}\right) \\ D_k^{-1} &\rightsquigarrow \mathcal{W}_L\left(\rho_0; R_0\right) \end{aligned} \tag{4}$$

where $\mathcal{G}$ stands for a Gamma distribution, $\mathcal{W}$ for a Wishart distribution, and $(\beta^0, B_0, \nu_0, \delta_0, \rho_0, R_0)$ are hyperparameters, which could eventually be treated as random variables just as the parameters of interest above.

Classical priors on the transition matrix specify the ith row of the transition matrix as a Dirichlet distribution

$$\pi_k = (\pi_{k,0}, \pi_{k,1}) \rightsquigarrow \mathcal{D}(\alpha_{k,0}; \alpha_{k,1}) \tag{5}$$

The joint probability density function of $(\Theta, s_1...s_T, |Y)$ is

$$\begin{aligned} p\left(\Theta, s_1...s_T|Y\right) &= p\left(Y|\Theta, s_1...s_T\right)p\left(s_1...s_T|\Theta\right)p\left(\Theta\right) \qquad (6) \\ &\propto \prod_{i=1}^{N} f(Y_i|\Theta, s_1...s_T)\prod_{t=2}^{T}\pi_{s_{t-1},s_t}g(s_1)\prod_{k=0}^{1}p\left(\pi_k\right)p\left(\theta_k\right) \end{aligned}$$

where $\Theta = (\theta_0, \theta_1)$ are the parameters of interest and $g(.)$ is the invariant distribution of the Markov Chain[6].

3 Gibbs sampling estimation

Gibbs sampler techniques have been widely used in Markov Chain Monte Carlo methods. They consist in three steps:

1. Setting initial values for all parameters.

2. Drawing random sequences of the parameters of interest according to the conditional posterior distribution. Parameters are generally drawn sequentially and by group: for instance in the first regime, one draws

[5] see Hobert and Casella (1996) for an exemple relevant to this paper.

[6] the eigenvector associated to the eigenvalue 1

random effects $b_{i,0}$ conditionally on subsequent realizations of other parameters $(\beta_0, \sigma_0, D_0, \pi_0)$ and hidden variables $(s_1...s_T)$.

3. Iterate M times the sampling of Θ. The resulting distribution $(\Theta_1, \Theta_1, ...\Theta_M)$ is a Markov chain that converges towards the target distribution under fairly general conditions (see Roberts and Smith (1994) and Tierney (1994)). Generally a burn-in phase is implemented and the corresponding values of Θ are discarded from the final sample.

This procedure is a natural tool for identification of models that allow for a non-null correlation between unobserved heterogeneity and regressors. Let us consider the simplest case where $L = 1$ and $z_{i,t} = 1$ for all i and t, which corresponds to the case of unobserved heterogeneity in the levels of the dependant variable; in that case, the usual GLS estimator is no more consistent, as well as any estimator using the Between variance of the observations. The Within estimator or First Differencing methods provide consistent estimates because they eliminate the source of bias, namely unobserved heterogeneity. Nevertheless, this comes at the price that all explanatory variables must be time-varying in order to fulfill the standard rank condition, i.e. inversibility of the matrix E $\tilde{x}_{i,t}^{'}\tilde{x}_{i,t}$ where $\tilde{x}_{i,t}$ stands for the transformed explanatory variables. Instrumental variables have been proposed in case some regressors are exogenous (e.g. Amemiya and MaCurdy (1986)), whereas the Chamberlain approach[7] extracts information from the moments of the variables implied by the model to identify the parameters of interest.

In the following empirical application, identification stems from the fact that a significant proportion of workers - between 12% and 40% - increases their level of education during adulthood, providing a relevant source of identification. But it is worth mentioning that identification is rarely an issue in a Bayesian framework thanks to the specification of proper prior distributions on the parameters[8]. Even in the case of time-constant and endogenous regressors, the specification of priors, albeit vague and misspecified, enables to achieve the estimation of a correlated random model as demonstrated by Murtin (2006)[9].

[7] see Chamberlain (1982) and Chamberlain (1984).

[8] see Gelfand-Sahu (1999) for extensive description of identication in mixed models.

[9] though estimation is unaccurate with low levels of correlation between fixed-effects and endogenous variables, and is more sensible to distributional forms.

The following Bayesian method has a crucial advantage over maximum likelihood methods: states and parameters from the equations of interest are determined simultaneously. A maximum likelihood method would consist in first deriving a test of structural change for panel data, which was not available at the time this paper was built[10]; second, in running some tobit panel model with fixed-effects on each of the distinct estimated periods, holding into account that some correction of standard errors might be brought to these second-step estimates. Despite a slight "entry cost", it is clear that MCMC procedures are much easier to implement[11].

The main difficulty arises from missing data, mainly the unknown regime states. As the target distribution is $p(\Theta, s_1...s_T|Y)$ one extracts information on the states $(s_1...s_T)$ by inferring the conditional distribution $p(s_1...s_T|\Theta, Y)$. The procedure begins with a *data augmentation step,* which consists in simulating the unobserved states $\hat{s}_t$ from the former distribution. Following Chib (1996), this is achieved with a forward pass through the data, which stores the probability distributions $p(s_t|Y, \Theta)$ for all $t \leq T$, and with a backward pass where the states $\hat{s}_t$ are simulated from the above distributions. Then parameters are drawn from the conditional distribution $p(\Theta|Y, \hat{s}_1...\hat{s}_T)$ updated via Bayes' rule

$$\begin{aligned} p(\Theta|Y, \hat{s}_1...\hat{s}_T) &= \prod_{k=0,1} p(\pi_k|\hat{s}_1...\hat{s}_T) \prod_{t \in T_k = \{t/s_t=k\}} p(\theta_k|\hat{s}_t = k, Y_t) \qquad (7) \\ &\propto \prod_{k=0,1} p(\pi_k|\hat{s}_1...\hat{s}_T)\, p(\theta_k) \prod_{t \in T_k} p(Y_t|\hat{s}_t = k, \theta_k) \\ &= \prod_{k=0,1} p(\pi_k|\hat{s}_1...\hat{s}_T)\, p(\theta_k) \prod_{t \in T_k} \prod_i f(y_{i,t}|\hat{s}_t = k, \theta_k) \end{aligned}$$

In practice the algorithm is the following:

Algorithm

1. Step 1 (Forward pass): Set $p(s_1|Y_0, \Theta)$ to be the stationary distribution of π, which is drawn from its unconditional distribution. Compute

[10] xxx

[11] Though this study has been conducted independantly, Tsionas-Kumbhakar (2004) look at a roughly similar model designed for a macroeconomic application. However, they do not consider the important specificities of microeconomic data such as endogeneity and censoring, which are the key ingredients in this study.

recursively for $t = \{1, 2...T\}$

$$p(s_t = k|Y_t, \Theta) = \frac{p(s_t = k|Y_{t-1}, \Theta) f(y_t|Y_{t-1}, \theta_k, \pi_k)}{\sum_{l=0,1} p(s_t = l|Y_{t-1}, \Theta) f(y_t|Y_{t-1}, \theta_l, \pi_l)}$$

where

$$p(s_t = k|Y_{t-1}, \Theta) = \sum_{l=0,1} p(s_t = k|s_{t-1} = l, \Theta) p(s_{t-1} = l|Y_{t-1}, \Theta)$$

2. Step 2 (Backward pass): Simulate s_T from $p(s_T|Y, \Theta)$, and compute recursively for $t = \{T-1, T-2...1\}$

$$p(s_t = k|Y_t, \Theta, \hat{s}_{t+1}) = \frac{p(s_t = k|Y_t, \Theta) p(\hat{s}_{t+1}|s_t = k, \pi)}{\sum_{l=0,1} p(s_t = l|Y_t, \Theta) p(\hat{s}_{t+1}|s_t = l, \pi)}$$

where $p(\hat{s}_{t+1}|s_t, \pi)$ is the first column of π when $\hat{s}_{t+1} = 0$, the second otherwise. Then $\hat{s}_t$ can be drawn from the above distribution.

3. Step 3 (Parameters sampling): Given $(\hat{s}_1...\hat{s}_T)$, simulate θ_k from its posterior conditional distribution

$$p(\theta_k) \prod_{t \in T_k = \{t/s_t = k\}} \prod_i f(y_{i,t}|\hat{s}_t = k, \theta_k)$$

With the subsequent priors and independance assumptions, the posterior distributions admit closed-forms given by:

- $\beta_k \rightsquigarrow \mathcal{N}_K \left(B_k(B_0^{-1}\beta^0 + \frac{1}{\sigma_k^2} \sum_{i=1, t\in T_k}^N x_{i,t}^{'} (y_{i,t} - z_{i,t}b_{i,k})), B_k = (B_0^{-1}\beta^0 + \frac{1}{\sigma_k^2} \sum_{i=1}^N x_{i,t}^{'} x_{i,t})^{-1} \right)$
- $b_{i,k} \rightsquigarrow \mathcal{N} \left(D_i \frac{1}{\sigma_k^2} \sum_{t\in T_k}^N z_{i,t}^{'}(y_{i,t} - x_{i,t}\beta_k), D_i = (D_k^{-1} + \frac{1}{\sigma_k^2} \sum_{t\in T_k}^N z_{i,t}^{'} z_{i,t})^{-1} \right)$.
- $D_k^{-1} \rightsquigarrow \mathcal{W}_L \left(\rho_0 + N; (R_0^{-1} + \sum_{i=1}^N b_{i,k} b_{i,k}^{'})^{-1} \right)$
- $\frac{1}{\sigma_1^2} \rightsquigarrow \mathcal{G} \left(\frac{\nu_0 + N \text{ . card } (T_1)}{2}; \frac{\delta_0}{2} + \frac{1}{2} \sum_{i=1, t\in T_1}^N v_{i,t}^2 \right)$
  where $v_{i,t} = y_{i,t} - x_{i,t}\beta_1 - z_{i,t}b_{i,1},\ t \in T_1$
- $\frac{1}{\sigma_0^2} \rightsquigarrow \mathcal{TG}_{[\frac{1}{\sigma_1^2}, +\infty)} \left(\frac{\nu_0 + N \text{ . card } (T_0)}{2}; \frac{\delta_0}{2} + \frac{1}{2} \sum_{i=1, t\in T_0}^N v_{i,t}^2 \right)$
  where $v_{i,t} = y_{i,t} - x_{i,t}\beta_0 - z_{i,t}b_{i,0},\ t \in T_0$ and $\mathcal{TG}_A$ represents a truncated Gamma distribution on the interval A.

- $\pi_i \rightsquigarrow \mathcal{D}\left(\alpha_{k,0}+n_{k,0};\alpha_{k,1}+n_{k,1}\right)$

  where $n_{k,0}$ (resp. $n_{k,1}$) is the number of transitions from state k to state 0 (resp. 1): this updates the transition matrix given $(\hat{s}_1...\hat{s}_T)$.

Treating the sampling of b_i in one block independently from the slopes β can be somewhat tricky because of mixing problem of the Gibbs algorithm. In practice one should use a large number of iterations[12] and choose reasonable hyperparameters.

When independance between unobserved heterogeneity and explanatory variables is imposed, the algorithm can be adapted. Chib and Carlin (1999) propose interesting blocking schemes, a simple one consisting in sampling β marginalized over b_i and then sampling b_i conditionally on β. In practice this scheme is very simple because the density of the observations marginalized over b_i is gaussian as well

$$f(y_{i,t}|s_t=k,\beta_k,\sigma_k,D_k) \rightsquigarrow \mathcal{N}\left(x_{i,t}\beta_k;V_{i,t,k}\right), \quad \text{with } V_{i,t,k}=\sigma_k^2+z_{i,t}D_k z_{i,t}^{'} \tag{8}$$

The sampling of β_k in the former algorithm is modified by taking $b_{i,k}=0$ and adapting the scheme to the new covariance matrix.

[12] in what follows $M = 50000$ for the real data case.

4 Test of the procedure

In order to test the estimator, four datasets are simulated with the following structure

$$
\begin{aligned}
y_{i,t} &= x_{i,t}\beta_{s_t} + z_{i,t}b_{i,s_t} + \sigma_{s_t}u_{i,t} \\
x_{i,t} &= \upsilon_i + \varepsilon_{i,t} \qquad \varepsilon_{i,t}\perp u_{i,t}, \varepsilon_{i,t}\perp b_{i,s_t} \\
(b_{i,0}, b_{i,1}, \upsilon_i) &\rightsquigarrow \mathcal{N}(0;V) \\
V &= \begin{bmatrix} D_0 & \Gamma_{0,1} & \Gamma_{0,\upsilon} \\ & D_1 & \Gamma_{1,\upsilon} \\ & & \sigma^2_\upsilon \end{bmatrix}
\end{aligned}
$$

Random effects and a time-constant component of explanatory variables have a joint normal distribution with non-trivial covariance matrix. It allows for correlation between random effects of the two regimes as well as with regressors. The choice of prior parameters marginally affects final estimates because they are sufficiently vague. I practice I take $\rho_0 = \nu_0 = 12$ while R_0 and δ_0 are calibrated on a coefficient of variation of variances' priors of 0.5[13]. Then $\forall\ i,j \quad \alpha_{i,j} = 1$ so that the prior transition probabilities are uniform on [0,1]. Each estimation consists in 10 000 iterations, and the first five hundred ones are discarded from the final sample. Table I compares the estimated values with the true ones.

The first model considers the case without unobserved heterogeneity. Gibbs sampling perfectly estimates the underlying parameters and detects the hidden states, although one transition probability is imprecisely estimated. The second model introduces unobserved heterogeneity, while the third model examines the multivariate case with random coefficients uncorrelated across regimes. All estimates fit the true values, with few exceptions for the transition probabilities, which anyway are not parameters of interest in the empirical application. Figure 1 depicts the values simulated by the Gibbs sampling algorithm in both regimes: the coefficient of the first regressor, the variance of residuals and of the first random component, the correlation between random effects, the transition probabilities, the expected and real states.

[13]more precisely I specify $\delta_0 = \frac{\nu_0^2}{4+\nu}m_0$ and $R_0 = \frac{\rho+4}{\rho^2}m_1^1$, with m_0 and m_1 being prior values on respectively the average variances σ_k^2 and D_k.

Last model is the same excepted that it introduces some correlation between explanatory variables and the corresponding random coefficient equal to 0.6 (resp. 0.3) in the first (resp. second) regime, as well as some correlation between random components across the two regimes (equal to 0.5). All underlying parameters are consistently estimated, as well as most of parameters of secondary interest such as the correlations. Figure 2 depicts the convergence of estimates in the fourth model, displaying in particular the correlation between random components and regressors in both regime.

5 Accounting for censored data

Due to confidentiality constraints or unemployment many economic individual files are censored, which can lead to serious bias in the estimates if the censoring rate is too high. As a goal is to apply the estimator to wage dynamics over thirty years, this problem is likely to appear because of unemployment or exit from the labor force. A basic view is that individuals do not work if the wage they might earn falls below a certain level, called the reservation wage. As a first approximation, I will assume that the reservation wage does not depend on any observed or unobserved characteristic. It is indeed the case that individuals could have some heterogeneous preferences or that education, familial background, household composition or former participation could affect current participation to the labour market. This is exemplified in a detailed study by Hyslop (2003) for married women in the US. Introducing those refinements would not make the model numerically intractable because the algorithm relies on simulations. But one would have to introduce quite different tools such as the Metropolis-Hastings step in order to deal with non-linearity. For the sake of simplicity, the reservation wage is kept constant across individuals.

Thus, the wage distribution is left-censored, and a simple way to correct the former algorithm is to use a latent variable model following Chib (1992):

$$
\begin{aligned}
y^*_{i,t} &= x_{i,t}\beta_{s_t} + z_{i,t}b_{i,s_t} + \sigma_{s_t}u_{i,t} \qquad i \leq N, t \leq T \\
y_{i,t} &= y^*_{i,t}1_{y^*_{i,t} > \tau_t} \\
b_{i,s_t} &\rightsquigarrow \mathcal{N}(0, D_{s_t}) \quad \text{and} \quad u_{i,t} \rightsquigarrow \mathcal{N}(0,1)
\end{aligned} \tag{9}
$$

Although the level of censoring might differ across time and individuals, it is assumed to be constant in the following simulations. Interestingly, the algorithm is only marginally modified. The posterior distribution becomes

$$\begin{aligned}
p\left(\Theta|Y^{*},\hat{s}_1...\hat{s}_T\right) &\propto \prod_{k=0,1} p\left(\pi_k|\hat{s}_1...\hat{s}_T\right) p\left(\theta_k\right) \prod_{i,t\in T_k=\{t/s_t=k\}} f(y^{*}_{i,t}|\hat{s}_t = k,\theta_k) \\
&= \prod_{k=0,1} p\left(\pi_k|\hat{s}_1...\hat{s}_T\right) p\left(\theta_k\right) \text{ x} \\
\prod_{i,t\in T_k} f(y_{i,t}|\hat{s}_t &= k,\theta_k, y_{i,t} > 0) \prod_{i,t\in T_k} f(y^{*}_{i,t}|\hat{s}_t = k,\theta_k, y_{i,t} = 0) \qquad (10)
\end{aligned}$$

Then Bayes rule provides

$$\begin{aligned}
f(y^{*}_{i,t}|\hat{s}_t = k,\theta_k, y_{i,t} = 0) &\propto f(y^{*}_{i,t}|\hat{s}_t = k,\theta_k) f(y_{i,t} = 0|\hat{s}_t = k,\theta_k, y^{*}_{i,t}) \\
&= f(y^{*}_{i,t}|\hat{s}_t = k,\theta_k) 1_{y^{*}_{i,t}\leq\tau_t}
\end{aligned}$$

Given the gaussian specification of f, the unobserved values of $y^{*}_{i,t}$ are thus drawn from a truncated normal $\mathcal{TN}_{(-\infty,\tau_t]}\left(x_{i,t}\beta_k + z_{i,t}b_{i,k};\sigma^2_k\right)$.

A correction of the above algorithm immediately follows from including the unobserved values of $y^{*}_{i,t}$ into the sampling. The former algorithm remains the same except that it ends with a data augmentation step to simulate censored observations:

Algorithm for censored data

1. Steps 1 to 4 are the same as before provided that $y_{i,t}$ is remplaced by $y^{*}_{i,t}$ in the sampling of parameters.

2. Step 5 (Censoring correction): Sample $y^{*}_{i,t} \rightsquigarrow \mathcal{TN}_{(-\infty,\tau_t]}\left(x_{i,t}\beta_k + z_{i,t}b_{i,k};\sigma^2_k\right)$ for any censored observation.

In practice I test this model with three different datasets while τ_t corresponds to the 10% quantile of the vector Y. As before, the sampling consists in 10000 Gibbs iterations. Table 2 depicts the results for multivariate cases with (model III) and without (models I and II) correlation between explanatory variables and random components. The first model uses the same data as model 3 from Table 1, excepted that some observations are censored. The second model is virtually the same excepted that it introduces a deterministic trend and increases the degree of correlation between random effects in the second regime. The third model uses the same data as model

4 from Table 1, adding only censoring. Estimates converge reasonnably well towards the right values for all the models. A look at Figure 3 shows how the censored data is simulated at final iteration in the third model. The left tail of the distribution might be a little bit too thick, but overall the results are satisfactory.

6 Application to US wages dynamics from 1968 to 2001

6.1 Model and data

I use the PSID (SRC part) from 1968 to 2001 and study seven different cohorts. The range of age spans from 41-45 years old in 1968 to 26-30 years old in 1983. The dynamics of hourly earnings are assessed in a Mincerian equation with regime-varying coefficients. This model also accounts for censored data, and potentially endogenous regressors. Its general form is

$$\begin{aligned}
y^*_{i,t} &= a_{0,s_t} + a_{1,s_t} E_{i,t} + (a_{3,s_t} + a_{4,s_t} E_{i,t}) A_{it} + (a_{5,s_t} + a_{6,s_t} E_{i,t}) A^2_{it} + v_{i,t} \\
v_{i,t} &= b_{i,s_t} + \sigma_{s_t} u_{i,t} \\
y_{i,t} &= y^*_{i,t} 1_{y^*_{i,t} > \tau_t} \\
b_{i,s_t} &\rightsquigarrow \mathcal{N}(0, D_{s_t}), \; u_{i,t} \rightsquigarrow \mathcal{N}(0,1)
\end{aligned} \tag{11}$$

where s_t is the unknown state to infer; covariates are age (A_{it}), squared age (A^2_{it}), the number of years of schooling (E_{it}), as well as the interaction of age and squared age with education. Indeed, interaction effects are both theoretically and empirically motivated as suggested by Heckman-Locher-Todd (2005). They imply that the return to education varies with age. With this specification, one controls for time effects on the return to schooling through the structural change, for age effects through the interaction mentionned above, and for cohorts effects since all estimates are cohort-specific.

Card (2001), Lemieux (2006b) and Deschênes (2006) introduce two sources of heterogeneity with a classical unobserved ability component but also heterogeneous returns to schooling. This is particularly attractive in cross-sectional regressions from the 80s because this specification naturally links the rises in inequality between and within-groups. Indeed, a growth in the return to schooling magnifies the heterogeneity in this return and the residual inequality associated to it. But this study starts in the 70s during which between and within-groups inequalities have varied in a opposite way, so that the relevance of heterogeneous returns is not obvious in this framework.

Similarly, a potential model for earnings dynamics allows for unobserved heterogeneity interacted with a time trend[14]. As discussed by Meghir and Pistaferri (2004), who use the PSID as well, this would imply long-term autocorrelations of the first-differenced residuals, which cannot be empirically detected. One step further, the authors decompose the residuals into a transitory MA noise and a random walk that represents the dynamics of permanent income - and indeed they show that allowing for shocks on permanent income is empirically motivated. Interestingly, they graph the variance of the permanent shocks, which follow an inverted U-curve centered on the beginning of the 80s. The model considered here is compatible with those empirical evidence since it accounts for permanent shocks through the structural change: it simply substitutes a Dirac distribution to the variance's inverted U-curve. For simplicity, transitory components are ignored, which might only affect efficiency of estimates.

Last, it is worth underlining that the change of regime is specific to each cohort. As Card-Lemieux (2001) have underlined the imperfect substituability of workers of different age, it is relevant to investigate differences in the timing of the structural change across cohorts. Even if it is maybe unlikely, the model leaves opened the possibility that a cohort experiences a *transitory* increase in the variance of residuals, hence a regime switching rather than a structural change. This possibility could be easily removed by constraining the coefficient $\pi_{1,1}$ in the transition matrix to 1, thus creating an absorbing state. The estimation of other parameters would be exactely the same. Also, it is worth underlining that a single transition from state 0 to state 1 is sufficient to identify the whole transition matrix because there are only two states and coefficients of the transition matrix are linearly linked.

In the data, some observations are missing and others are outliers with high probability, typically when income falls at very low levels. I censore all observations below a threshold equal to a zero log-hourly income in real terms (1968 prices), which means 4.3$ per hour in 2001 prices. With this assumption, the percentage of censored data is typically around 10%.

[14]for instance, because of heterogenous returns to experience.

Table 3 provides some elementary descriptive statistics[15]. The increase in the mean educational level is a well-known tendency. The slowdown of higher education in the beginning of the 80s is also much debatted as a potential explanation for the increase in the college premium at that period. The percentage of people acquiring some education along their lifetime - subsequently called the "Movers" - varies from 12.4% to 38.2%. Cohorts of workers older than 45 years old in 1968 were not included because the percentage of "Movers" was too low, as it was also the case for younger cohorts of the late 80s. I also reported the percentages of college graduates and post-graduates at the end of working life, as well as the percentages of college or post-graduate degrees among acquired years of education[16]. It will be useful in the discussion of the results.

6.2 Results

This model enables to calculate a return to education at a given age for each cohort, as well as an unobserved ability variable for each individual in each regime. It shows whether a structural break has affected each cohort, and how the age-profile of the return to education has been modified across regimes. In a second step, I can aggregate all individuals aged between 26 and 64 and compute the college premium year by year using Mincerian regressions. Whether I introduce or not the individual unobserved ability variable into those cross-sectional equations, the estimated college premium will be purged or not from the endogeneity bias. The difference between both estimates will indicate how much of the college premium is explained by unobserved heterogeneity. The procedure would remain the same whatever the number of heterogenity sources included in the model.

As a result, the global impact of the structural change on the return to education is depicted in Figures 4 to 13: for the oldest cohort in 1968 and the youngest cohort after 1980, the model detects only one regime. For other cohorts, the model detects a change in the regime occuring in the beginning of the 80s. The second and current regime is characterized by higher variances of both residuals and unobserved heterogeneity, and

[15]Statistics for education refer to percentages of total observations along the working life.

[16]hence one "Mover" can acquire several post-graduate degrees and counts for several moves.

different profiles of the return to education across age: the shift of the return to education profile goes upward for the workers aged between 31 and 40 years old in 1968, but goes downward for the younger ones. This suggests that older cohorts have benefited from the organizational change of the early 80s relatively more than younger cohorts. If there has been a re-organization process in which skilled individuals were concentrated within some firms as described by Kremer-Maskin (1999) among others, then it is likely that the oldest skilled workers were able to obtain the best positions and even capture some rent from this re-organization.

The main result is given in Table 4, which displays the level of correlation between education and unobserved ability, the correlation of unobserved heterogeneity across both regimes , as well as the size of the ability premium - i.e. the average unobserved ability of college graduates minus the average unobserved ability of other workers.

First, the correlation of unobserved heterogeneity across the two regimes is about 0.5 for any cohort experiencing the structural change. It pleads for a multi-dimensional vision of ability: some skills have been rewarded differently in the second regime compared to the first one, but this change in the return has not been uniform across skills. Otherwise, the correlation of unobserved ability across regimes would be equal to one. Bowles-Gintis-Osborne (2000) review many factors with a significant impact on earnings such as IQ, psychological or physical traits, parental background and adaptability to a new economic environment in general. This suggests that some of them, which remain to be isolated, have been rewarded differently in the new managerial era of the 80s.

Second, the two cohorts of workers aged between 26 and 30 years in 1973 and 1978 display significant and positive levels of correlation between education and unobserved heterogeneity in the second regime. The ability premium amounts to respectively 0.14 and 0.17, namely a rough 30% of the college premium. Why is this significant correlation existing only in the second regime? The answer could follow from what precedes: some skills associated to schooling have become rewarded after the structural change, while their return was null in the first regime. If viewed as innate, those skills were likely existing in the first regime, so that only an increase in their return can explain both the null and positive correlations in respectively regimes 1 and 2. Admittedly, it is hard to put forward a particular kind of skill given the lack of knowledge on this point in the data. But it is true that "managerial skills" are a good candidate, given that they are both associated with education and the changing forms of production in the early 80s.

An examination of the composition of education and educational mobility in Table 3 goes into this direction and provides a piece of answer to the following question: why is the positive sorting of unobserved ability and college graduation taking place in only two cohorts? It is striking that the two relevant cohorts display the largest amount of higher education. Consider in particular the percentage of post-graduates at the end of working life, denoted as "PG end" in Table 3. The two highest values are 19.4% and 25.2% and correspond to the cohorts of workers respectively born in 1943-1947 and 1948-1952, the two particular cohorts we are interested in. These higher levels are partly explained by adult educational mobility since respectively 49% and 66% of "Moves" concern a post-graduate degree. Thus, it is plausible that in the re-organization process of the early 80s, a large proportion of college graduates from these two cohorts found profitable to acquire a Master degree or a PhD. Those workers were likely to be the most endowed with "managerial skills", which return had increased due to new forms of production.

Importantly, this taste for higher education was much lower among former cohorts and has decreased in subsequent ones. Therefore, the most simple hypothesis is that the sorting of ability and college graduation depends positively on the percentage of college graduates within each cohort. The intuition is the following: if individuals with "managerial skills" have a comparative advantage to graduate from college relatively to other individuals, then the increase in overall graduation rates should benefit relatively more to those skilled individuals, and the sorting of ability should increase. This would explain why the ability premium is higher among the cohorts with the largest percentage of college graduates.

In his essay, Acemoglu (2002) suggests that unless particular distributional assumptions on unobserved ability, this intuition is not true under perfect sorting, namely when the fact of obtaining a college degree is purely an increasing function of ability[17]. I show in the following that a simple

[17] Acemoglu (2002) shows that under perfect sorting the ability premium is constant if the distribution of unobserved ability is uniform. In fact one can show that under perfect

model of imperfect sorting can support this intuition.

Consider two types of workers $1, 0$ endowed respectively with ability $a = 1$ and $a = 0$. The "skilled type" 1 is in a fixed proportion p. Notice p_1^S (resp. p_0^S) the probability that type 1 (resp. 0) has at least a college degree. Then the proportions of college graduates and the average levels of ability among college graduates (group S) and non college graduates (group U) are respectively

$$\begin{aligned} p^S &= pp_1^S + (1-p)p_0^S \\ a^S &= \frac{pp_1^S}{p^S} \\ a^U &= \frac{p(1-p_1^S)}{1-p^S} \end{aligned} \tag{12}$$

The comparative advantage of the skilled type is that $p_0^S = \lambda p_1^S$ with $\lambda < 1$. Hence skilled individuals have a higher probability of graduation relatively to type 0. This leads to $p^S = p_1^S(p + (1-p)\lambda)$ and $a^S = \frac{p}{(p+(1-p)\lambda)}$ which is a constant. In contrast, basic algebric manipulations lead to

$$\frac{\partial a^U}{\partial p^S} = \frac{p(1-p)}{(p+(1-p)\lambda)}(\lambda - 1) < 0$$

So it is clear that the ability premium $a^S - a^U$ augments when overall college graduation p^S augments. This potentially explains why the ability premium is higher in the two cohorts with the highest percentages of college graduates and post-graduates.

The last question is about the empirical bias in the college premium when it is computed in cross-sections. Once each worker is attributed her level of unobserved heterogeneity in each year, it is possible to compare the coefficient of a College dummy in a Mincerian cross-sectional regression, including or excluding unobserved heterogeneity. For that purpose a Tobit estimator is used in order to take into account censoring. Last figure shows the result: the causal return to college is approximatively 30% lower than that usually calculated. This conclusion is compatible with Chay and Lee (2000) and similar to Deschênes (2002).

---

sorting the ability premium is a decreasing (resp. increasing) function of the overall college graduation rate if the distribution of ability is skewed to the right (resp. to the left).

7 Conclusion

This paper presents a general class of panel models that encompasses correlated random coefficients models and regime switching models and accounts for censored data. This is the minimal statistical framework in order to model earnings trajectories of workers on the US labour market since 1968. It proposes an original Gibbs sampling procedure to estimate the model. I apply the estimator to seven US 5-years cohorts of workers, and then compute the cross-sectional return to schooling with and without accounting for the correlation between a college dummy and estimated unobserved skills. The endogenous part of the college premium, called the ability premium, represents circa 30% of its value and is carried by the two most educated cohorts. A simple model where skilled workers have a comparative advantage to graduate from college can account for this fact.

References

[1] Acemoglu, D. (2002). Technical Change, Inequality and the Labor Market. Journal of Economic Literature, 40, pp..7-72

[2] Amemiya, T. and T.E. MaCurdy (1986), Instrumental Variable Estimation of Dynamic Models Using Panel Data, *Econometrica*, 54, 869-880.

[3] Angrist, J. and W. Newey (1991). Over-Identification Tests in Earnings Functions with Fixed-Effects. *Journal of Business and Economic Statistics*, Vol 9, pp. 317-323.

[4] Autor, D., L. Katz and M. Kearney (2006). The Polarization of the US Labor Market. *American Economic Review, Papers and Proceedings*, Vol 96, pp. 189-194.

[5] Bowles, S., Gintis, H. and M. Osborne (2000). The Determnants of Earnings: Skills, Preferences, and Schooling.

[6] Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistant Econometric Problems. Econometrica, 69(5), pp. 1127-1160.

[7] Card, D. and J. DiNardo (2002), Skill Biased technological Change and Rising Wage Inequality: Some Problems and Puzzles, *Journal of Labor Economics*, 20, 733-783.

[8] Card, D. and T. Lemieux (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? *Quaterly Journal of Economics*, 116, pp. 705-46

[9] Chamberlain, G. (1982), Multivariate Regression Models for Panel Data, *Journal of Econometrics*, 18, 5-46.

[10] Chamberlain, G. (1984), Panel Data, in Handbook of Econometrics, Z. Griliches and M.D. Intriligator eds., North-Holland.

[11] Chay, K. and D. Lee (2000), *Journal of Econometrics*, 99, 1-38.

[12] Chib, S. (1992), Bayes Inference in the Tobit Censored Regression Model, *Journal of Econometrics*, 51, 79-99.

[13] Chib, S. (1996), Calculating posterior distributions and model estimates in Markov mixture models, *Journal of Econometrics*, vol. 75, pp. 79–97.

[14] Chib, S. (2001), Markov Chain Monte Carlo Methods: Computation and Inference. In Heckman, J.J. and Leamer, E. (eds), Handbook of Econometrics, Volume 5, pp. 3569-3649, North Holland, Amsterdam.

[15] Chib, S. and BP Carlin. (1999), On MCMC Sampling in Hierarchical Longitudinal Models, *Statistics and Computing*, vol. 9, pp. 17-26.

[16] Chopin, N. & Pelgrin, F. (2004). Bayesian inference and state number determination for hidden Markov models: an application to the information content of the yield curve about inflation, Journal of Econometrics 123:2, 327-344.

[17] Deschênes, O. (2001). Unobserved Abiity, Comparative Advantage, and the Rising Return to Education in the United States: a Cohort Based Approach. Princeton University Industrial Relations Section, wp 465.

[18] Eckstein, Z. and E. Nagypal (2004), The Evolution of U.S. Earnings Inequality: 1961-2002, *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 28, No. 2, December 2004, pp. 10-29.

[19] Gelfand, A.E. and S. K. Sahu (1999). Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. *Journal of the American Statistical Association*, Vol. 94, No. 445.

[20] Heckman, J., Lochner L. and P. E. Todd (2005). Earnings Functions, Rates of Returns and Treatment Effects: the Mincer Equation and Beyond. IZA DP 1700.

[21] Hobert, J.P. and G. Casella (1996), The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models, *Journal of The American Statistical Association*, Vol. 91, No. 436.

[22] Hsiao, C. (1975), Some Estimation Methods for a Random Coefficients Model, *Econometrica*, 43, 305-325.

[23] Hyslop, D. (2003). State Dependance, Serial Correlation and Heterogeneity in Intertempora Participation Behavior: Monte Carlo Evidence and Empirical Results for Married Women. *Econometrica*, 67(6), p. 1255-1294.

[24] Kremer, and E. Maskin (1999). Segregation by Skill ad the Rise in Inequality. *Quaterly Journal of Economics*, Vol

[25] Lemieux, T. (2006 a). Composition Effects, Wage Measurement, and the Growth in Within-Group Wage Inequality. *American Economic Review*, Vol .

[26] Lemieux, T. (2006 b). Postsecondary Eucation and Increasing Wage Inequality. *American Economic Review, Papers and Proceedings*, Vol. 96 pp.195-199. .

[27] Morrisson, C. and F. Murtin (2007). Education Inequalities and the Kuznets Curves: 1870-2000. Mimeo.

[28] Murtin, F. (2006). Bayesian Estimation of Panel Data Models with Time-Constant Endogenous Regressors. Mimeo.

[29] Meghir, C. and L. Pistaferri (2004), Income Variance Dynamics and Heterogeneity, *Econometrica*, Vol. 72, No. 1, pp. 1-32.

[30] Swamy, P.A.V.B. (1970), Efficient Inference in a Random Coefficient Regression Model, *Econometrica*, 38, 311-323.

[31] Taber, C. (2001), The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability ?, *Review of Economic Studies,* 68, 665-691.

[32] Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions, *The Annals of Statistics*, 22, 1701-1762.

[33] Tsionas, E.G. and S.C. Kumbhakar (2004). Markov Switching Stochastic Frontier Model. *Econometrics Journal*, vol. 7, pp398-425

Table 1 - Test on 4 datasets (time-varying regressors)

		β_0	β_1	σ_0^2	σ_1^2	D_0	D_1	ρ_{b_0,b_1}	ρ_{X,b_0}	ρ_{X,b_1}	π_{00}	π_{11}	$\binom{N}{T}$
I	b_0	1	1.2	0.3	0.5	0	0	-	-	-	0.7	0.5	$\binom{500}{30}$
	$\hat{b}$	1.01 (0.01)	1.17 (0.01)	0.30 (0.01)	0.49 (0.02)	0.01 (0.00)	0.01 (0.00)	-	-	-	0.70 (0.09)	0.36 (0.14)	
II	b_0	1	1.2	0.3	0.5	0.3	0.5	-	-	-	0.8	0.5	$\binom{200}{30}$
	$\hat{b}$	0.98 (0.04)	1.23 (0.05)	0.30 (0.01)	0.50 (0.02)	0.35 (0.04)	0.54 (0.07)	-	-	-	0.73 (0.10)	0.58 (0.13)	
III	b_0	[0.8 1.0 1.2]	[0.6 1.0 1.4]	0.2	0.4	$\begin{bmatrix} 0.25 & 0.18 \\ & 0.5 \end{bmatrix}^1$	$\begin{bmatrix} 0.5 & 0.10 \\ & 0.5 \end{bmatrix}^2$	-	-	-	0.8	0.5	$\binom{200}{100}$
	$\hat{b}$	[0.78 1.01 1.19] (0.02)(0.04)(0.05)	[0.53 1.03 1.39] (0.04)(0.06)(0.06)	0.22 (0.02)	0.42 (0.02)	$\begin{bmatrix} \underset{(0.03)}{0.27} & \underset{(0.03)}{0.20} \\ & \underset{(0.05)}{0.48} \end{bmatrix}$	$\begin{bmatrix} \underset{(0.06)}{0.47} & \underset{(0.05)}{0.07} \\ & \underset{(0.06)}{0.42} \end{bmatrix}$	-	-	-	0.67 (0.09)	0.37 (0.15)	
IV	b_0	[0.8 1.0 1.2]	[0.6 1.0 1.4]	0.2	0.4	$\begin{bmatrix} 0.25 & 0.18 \\ & 0.5 \end{bmatrix}^3$	$\begin{bmatrix} 0.5 & 0.10 \\ & 0.5 \end{bmatrix}^3$	0.5	0.6	0.3	0.8	0.5	$\binom{200}{30}$
	$\hat{b}$	[0.79 0.98 1.18] (0.01)(0.03)(0.05)	[0.62 0.96 1.37] (0.03)(0.05)(0.04)	0.22 (0.01)	0.40 (0.02)	$\begin{bmatrix} \underset{(0.02)}{0.20} & \underset{(0.03)}{0.15} \\ & \underset{(0.05)}{0.48} \end{bmatrix}$	$\begin{bmatrix} \underset{(0.07)}{0.46} & \underset{(0.05)}{0.09} \\ & \underset{(0.07)}{0.43} \end{bmatrix}$	0.49 (0.03)	0.62 (0.01)	0.26 (0.03)	0.66 (0.10)	0.38 (0.14)	

[1]random coefficients correspond to the second and third regressor.
The correlation between random-effects is equal to 0.5.
[2]random coefficients correspond to the second and third regressor.
The correlation between random-effects is equal to 0.2.
[3]random coefficients correspond to the second and third regressor; they display some correlation across regimes as well as with regressors.

Table 2 - Test on 3 censored datasets

		β_0	β_1	σ_0^2	σ_1^2	D_1	D_2	ρ_{b_0,b_1}	ρ_{X,b_0}	ρ_{X,b_1}	π_{00}	π_{11}	$\binom{N}{T}$
I	b_0	[0.8 1.0 1.2]	[0.6 1.0 1.4]	0.2	0.4	$\begin{bmatrix} 0.25 & 0.18 \\ & 0.5 \end{bmatrix}^2$	$\begin{bmatrix} 0.5 & 0.10 \\ & 0.5 \end{bmatrix}^3$	-	-	-	0.8	0.5	$\binom{200}{30}$
	$\hat{b}$	[0.78 1.00 1.19] (0.01)(0.04)(0.05)	[0.62 0.98 1.29] (0.02)(0.06)(0.06)	0.19 (0.01)	0.39 (0.06)	$\begin{bmatrix} \underset{(0.03)}{0.28} & \underset{(0.03)}{0.19} \\ & \underset{(0.05)}{0.48} \end{bmatrix}$	$\begin{bmatrix} \underset{(0.07)}{0.46} & \underset{(0.05)}{0.13} \\ & \underset{(0.07)}{0.51} \end{bmatrix}$	-	-	-	0.69 (0.10)	0.52 (0.14)	
II[1]	b_0	[10.0 0.025 0.3 0.1]	[10.1 0.05 0.5 0.2]	0.2	0.4	$\begin{bmatrix} 0.25 & 0.11 \\ & 0.20 \end{bmatrix}^2$	$\begin{bmatrix} 0.50 & 0.36 \\ & 0.40 \end{bmatrix}^4$	-	-	-	0.8	0.6	$\binom{200}{30}$
	$\hat{b}$	[10.06 (0.02) 0.023 (0.001) 0.28 (0.04) 0.08] (0.03)	[10.13 (0.03) 0.048 (0.001) 0.42 (0.05) 0.18] (0.05)	0.20 (0.02)	0.38 (0.01)	$\begin{bmatrix} \underset{(0.02)}{0.21} & \underset{(0.02)}{0.12} \\ & \underset{(0.02)}{0.18} \end{bmatrix}$	$\begin{bmatrix} \underset{(0.06)}{0.53} & \underset{(0.05)}{0.41} \\ & \underset{(0.06)}{0.50} \end{bmatrix}$	-	-	-	0.62 (0.12)	0.66 (0.11)	
III	b_0	[0.8 1.0 1.2]	[0.6 1.0 1.4]	0.2	0.4	$\begin{bmatrix} 0.25 & 0.18 \\ & 0.5 \end{bmatrix}^2$	$\begin{bmatrix} 0.5 & 0.10 \\ & 0.5 \end{bmatrix}^3$	0.5	0.6	0.3	0.8	0.6	$\binom{200}{30}$
	$\hat{b}$	[0.78 0.97 1.17] (0.01)(0.04)(0.06)	[0.56 0.93 1.26] (0.03)(0.05)(0.05)	0.19 (0.01)	0.36 (0.01)	$\begin{bmatrix} \underset{(0.03)}{0.28} & \underset{(0.03)}{0.22} \\ & \underset{(0.06)}{0.56} \end{bmatrix}$	$\begin{bmatrix} \underset{(0.03)}{0.41} & \underset{(0.03)}{0.17} \\ & \underset{(0.06)}{0.51} \end{bmatrix}$	0.58 (0.03)	0.60 (0.01)	0.47 (0.02)	0.75 (0.10)	0.63 (0.13)	

[1]the first regressor is a constant and the second a linear trend.
[2]random coefficients correspond to the second and third regressors. The correlation between random-effects is equal to 0.5.
[3]random coefficients correspond to the second and third regressors. The correlation between random-effects is equal to 0.2.
[4]random coefficients correspond to the second and third regressors. The correlation between random-effects is equal to 0.8.

Table 3- Descriptive statistics - Males

Born in $\longrightarrow$	1923-1927	1928-1932	1933-1937	1938-1942	1943-1947	1948-1952	1953-1957
% of Censoring	6.8	8.2	13.4	12.0	9.0	8.0	10.0
% of Movers	12.4	14.7	20.8	24.1	36.9	38.2	25.9
% of HSD	30.1	24.6	23.5	12.4	12.9	4.5	6.3
% of HSG	1.1	35.8	33.7	35.7	27.9	31.0	38.8
% of SC	12.7	15.8	15.4	23.5	21.6	25.2	24.0
% of CG	16.4	10.6	16.6	17.6	25.2	23.1	18.1
% of PG	9.7	13.1	10.8	10.7	12.3	16.1	12.8
% of CG end	13.4	7.8	13.6	13.5	19.9	19.3	18.2
% of PG end	13.4	16.9	15.2	15.3	19.4	25.2	13.6
% of CG Moves	5	11	7	4	8	8	18
% of PG Moves	33	39	39	37	49	66	24
N	145	156	125	112	206	301	339

HSD=high-school dropouts, HSG=high-school graduates, SC=some college, CG=college graduates, PG=post-graduates. CG/PG end: degree at the end of working life. CG/PG Moves: percentage of college/post-college degrees among acquired years of education.

Table 4- The bias of the College Premium

	Regime 1			Regime 2			Correlation of unobserved ability between regimes
	Years	$\rho\left(E_i, b_{i,s_0}\right)$	E $(b_{i,s_0}\|E_i\geq 16)$ −E $(b_{i,s_0}\|E_i< 16)$	Years	$\rho\left(E_i, b_{i,s_1}\right)$	E $(b_{i,s_1}\|E_i\geq 16)$ −E $(b_{i,s_1}\|E_i< 16)$	
41-45 in 1968	-	-	-	1968-1986	−0.01 (0.09)	0.02 (0.06)	-
36–40 in 1968	1968-1981	−0.01 (0.09)	0.02 (0.07)	1982-1991	0.01 (0.08)	0.06 (0.09)	0.55 (0.03)
31-35 in 1968	1968-1981	0.03 (0.09)	0.02 (0.06)	1982-1996	0.05 (0.09)	0.09 (0.09)	0.55 (0.03)
26-30 in 1968	1968-1979	0.00 (0.10)	0.01 (0.05)	1980-2001	0.11 (0.08)	0.08 (0.08)	0.54 (0.03)
26-30 in 1973	1968-1980	0.02 (0.07)	−0.03 (0.04)	1981-2001	0.17 (0.05)	0.14 (0.04)	0.52 (0.03)
26-30 in 1978	1968-1982	−0.01 (0.06)	−0.04 (0.04)	1983-2001	0.23 (0.04)	0.17 (0.04)	0.57 (0.02)
26-30 in 1983	-	-	-	1983-2001	0.09 (0.04)	0.08 (0.04)	-

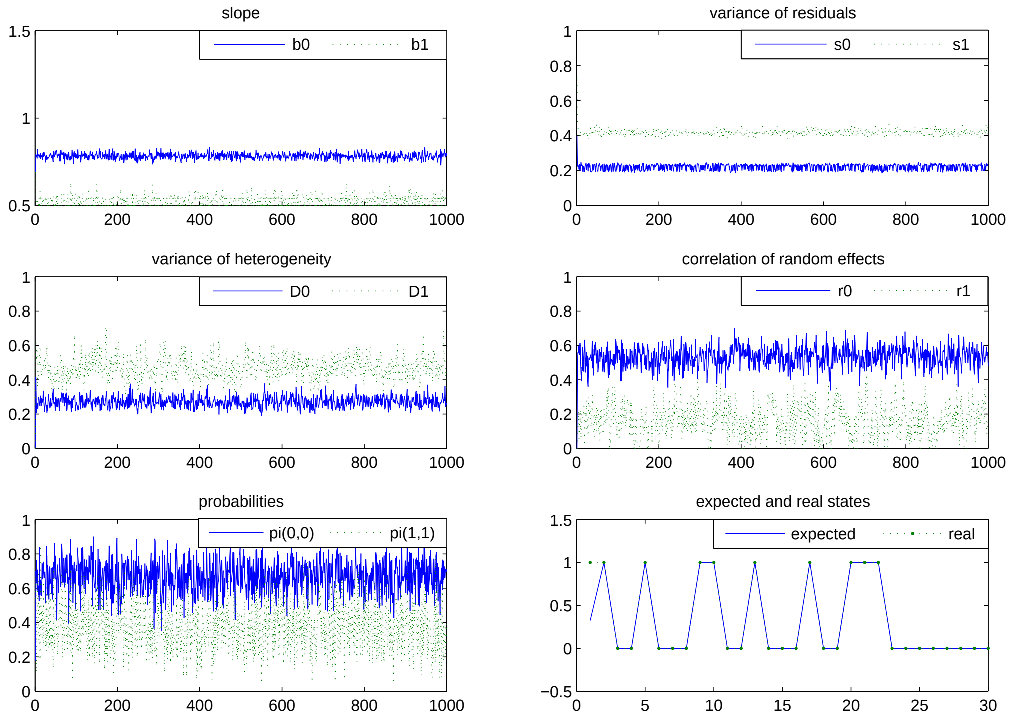

Figure 1: Model III Table 1

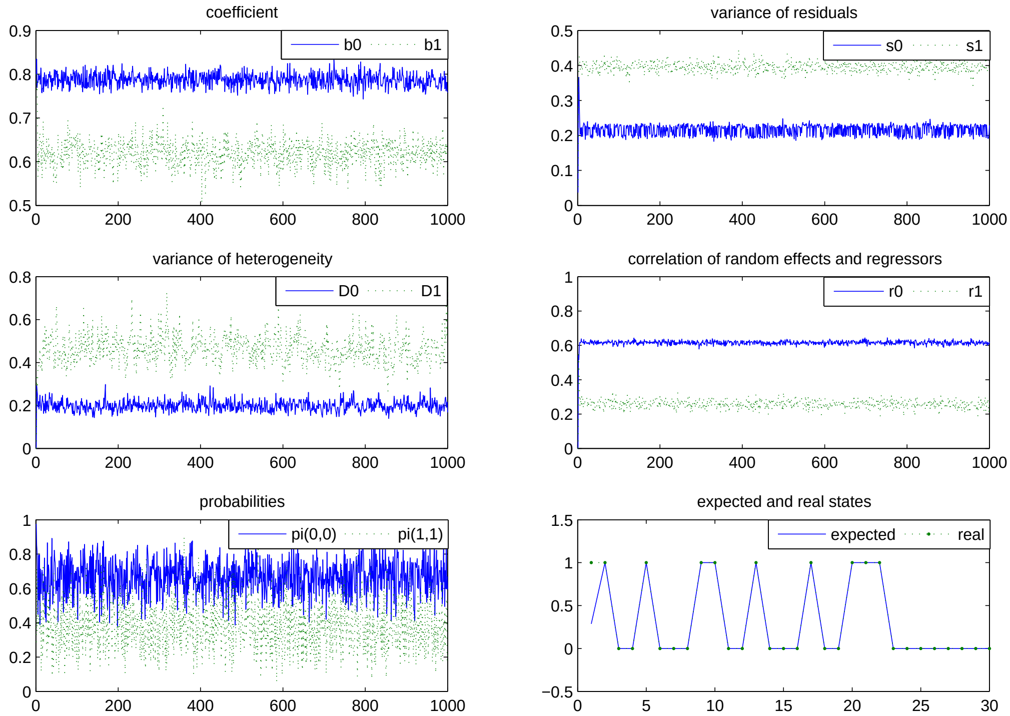

Figure 2: Model IV Table 1

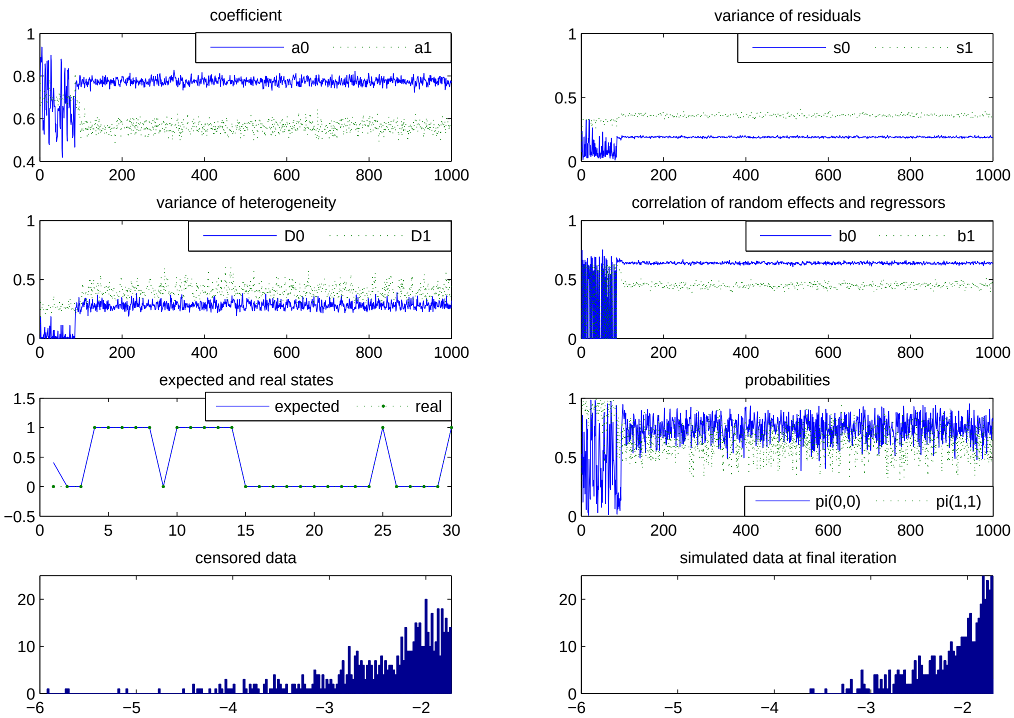

Figure 3: Model III Table 2

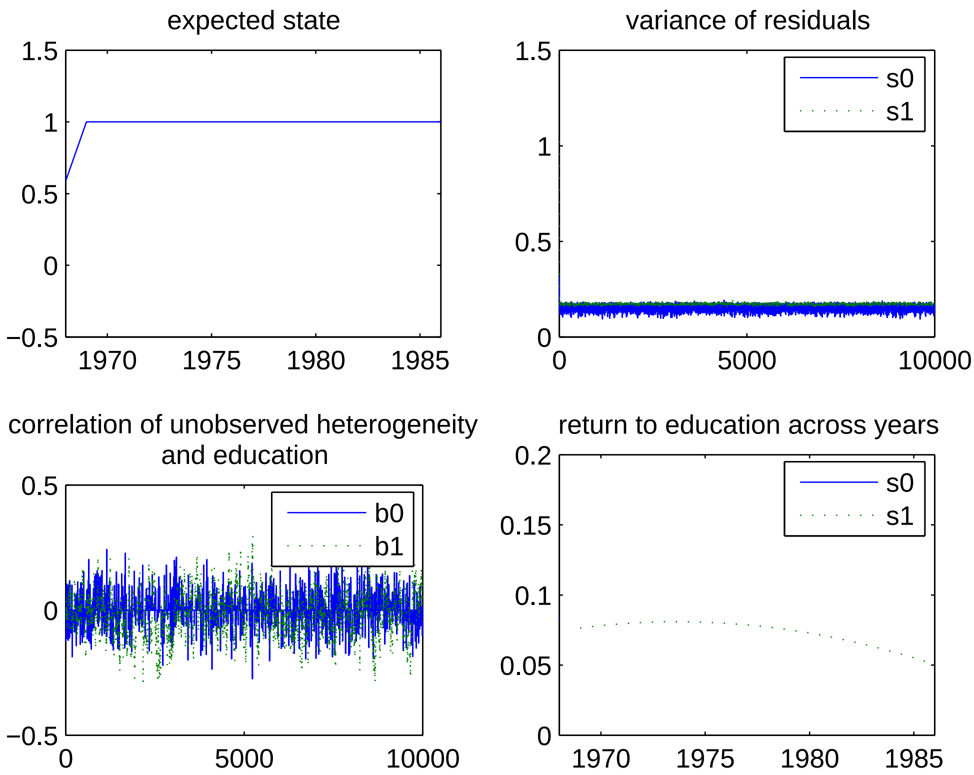

Figure 4: 41-45 years old in 1968

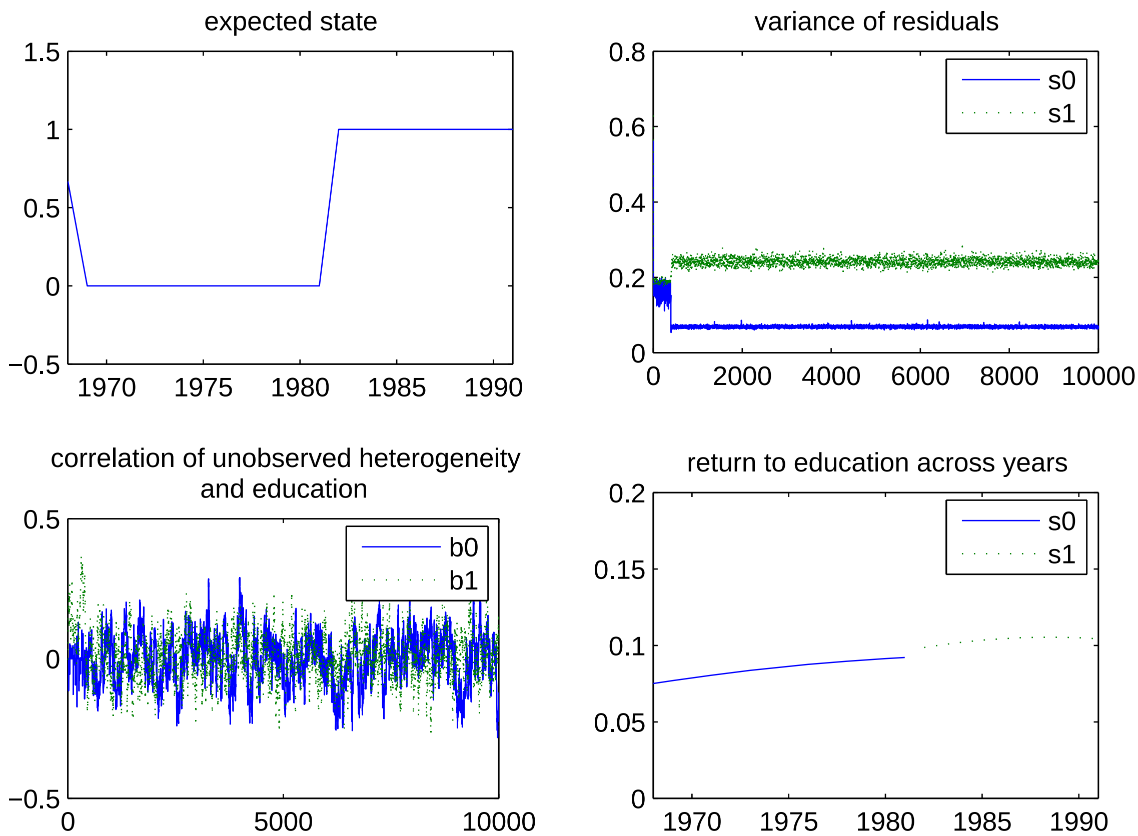

Figure 5: 36-40 years old in 1968

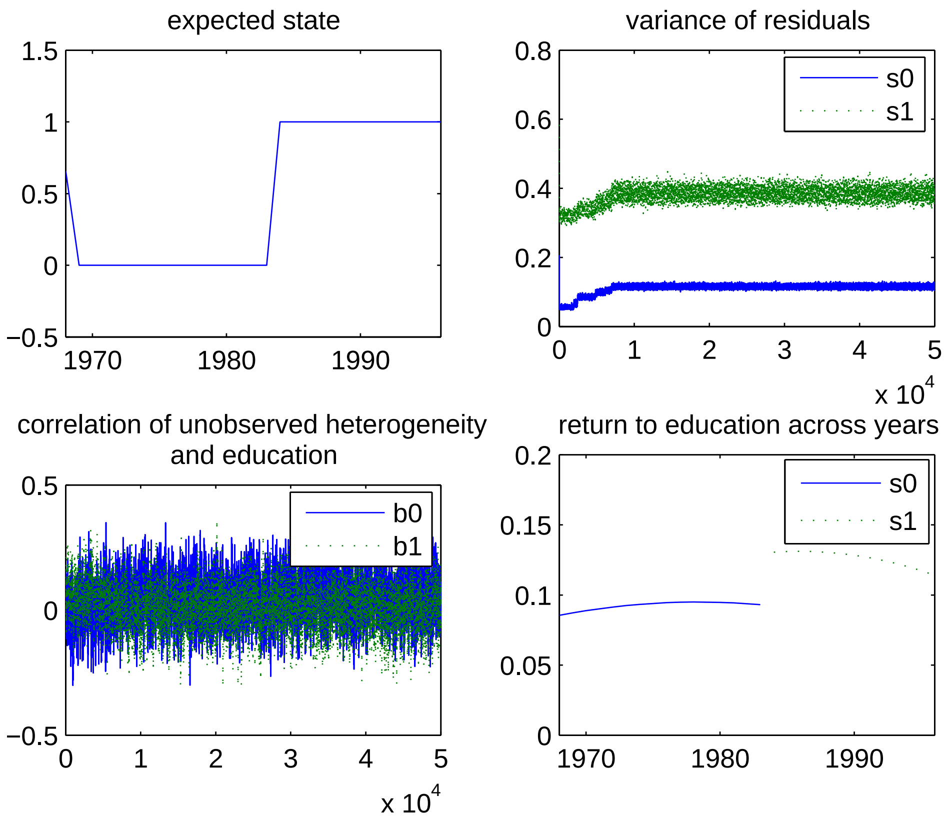

Figure 6: 31-35 years old in 1968

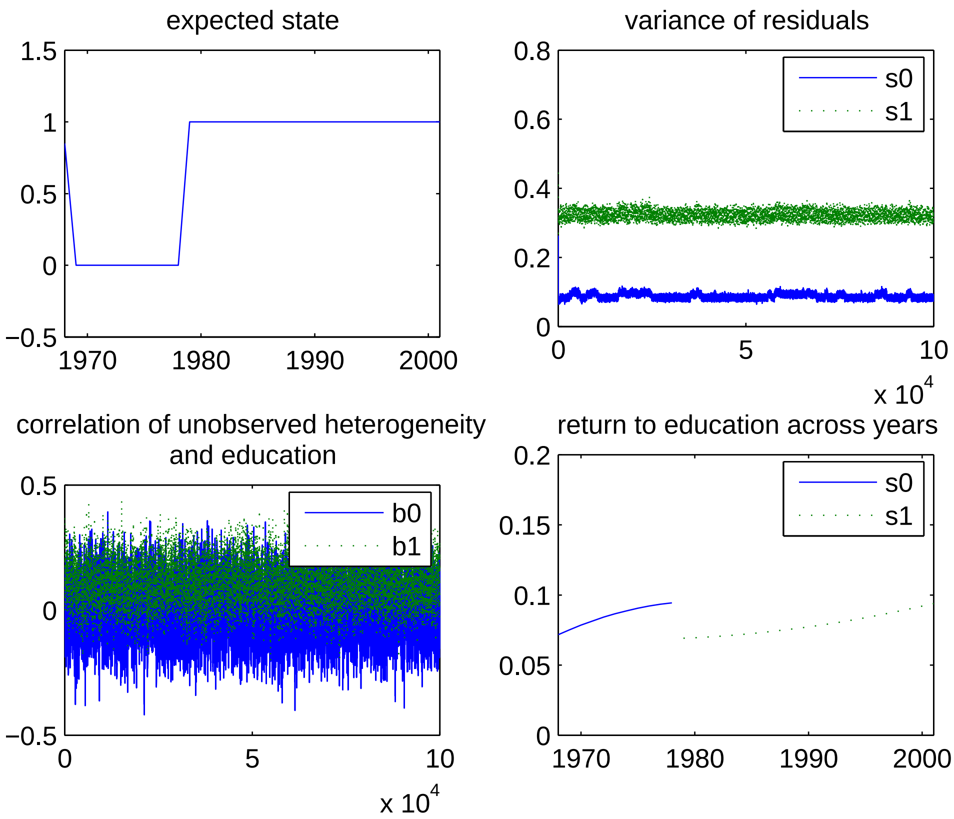

Figure 7: 26-30 years old in 1968

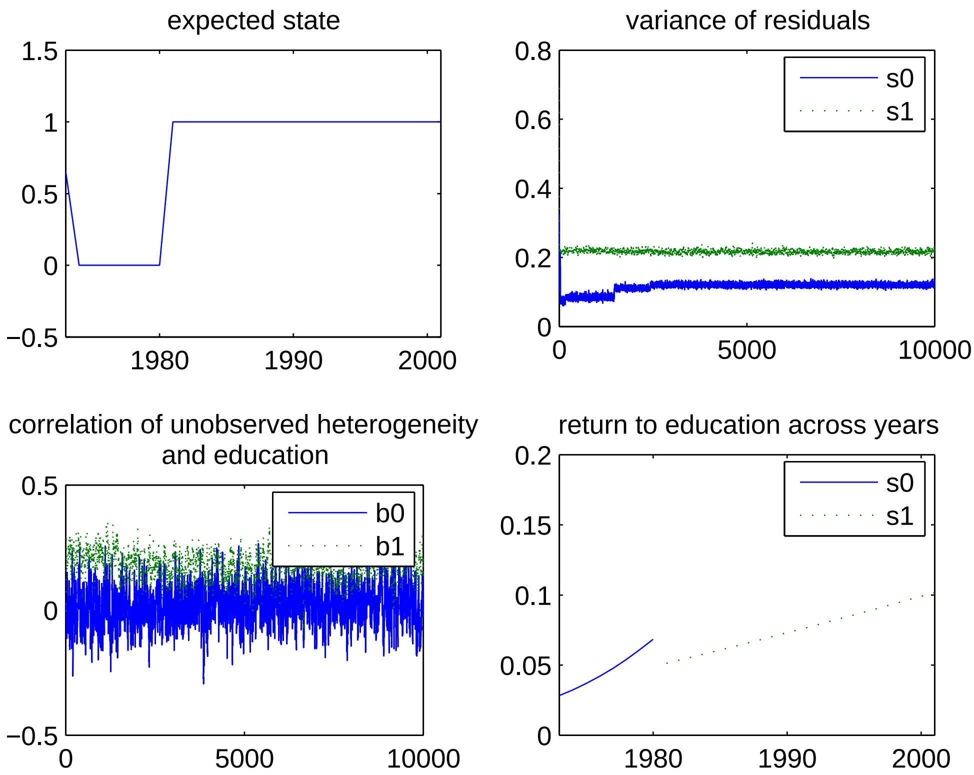

Figure 8: 26-30 years old in 1973

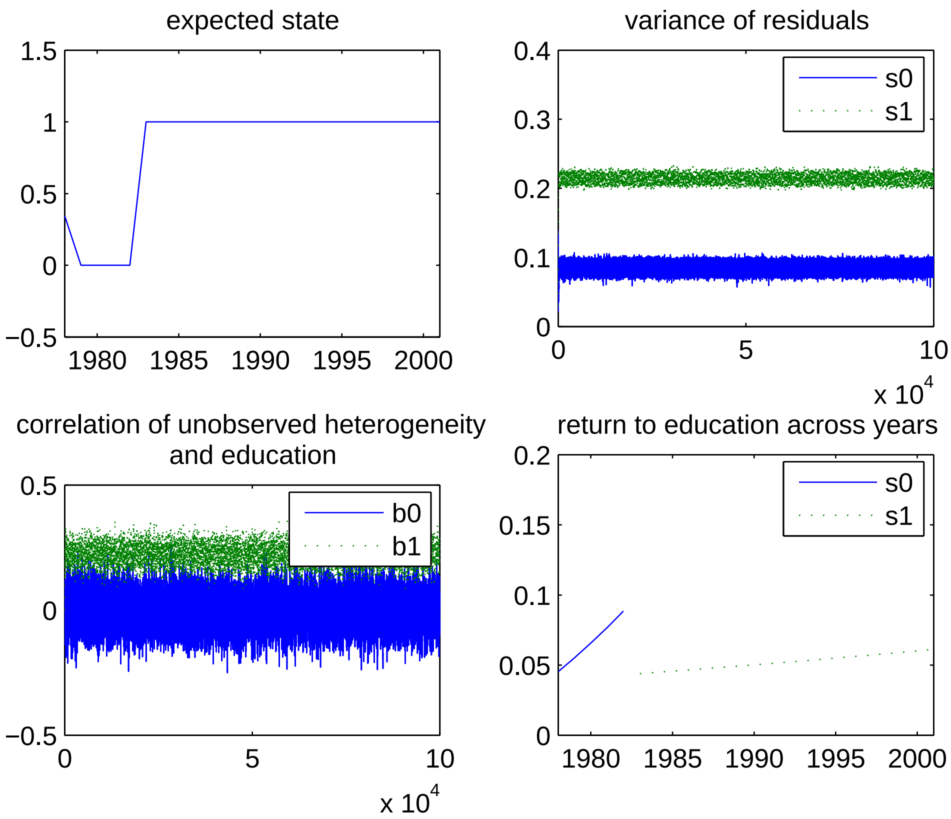

Figure 9: 26-30 years old in 1978

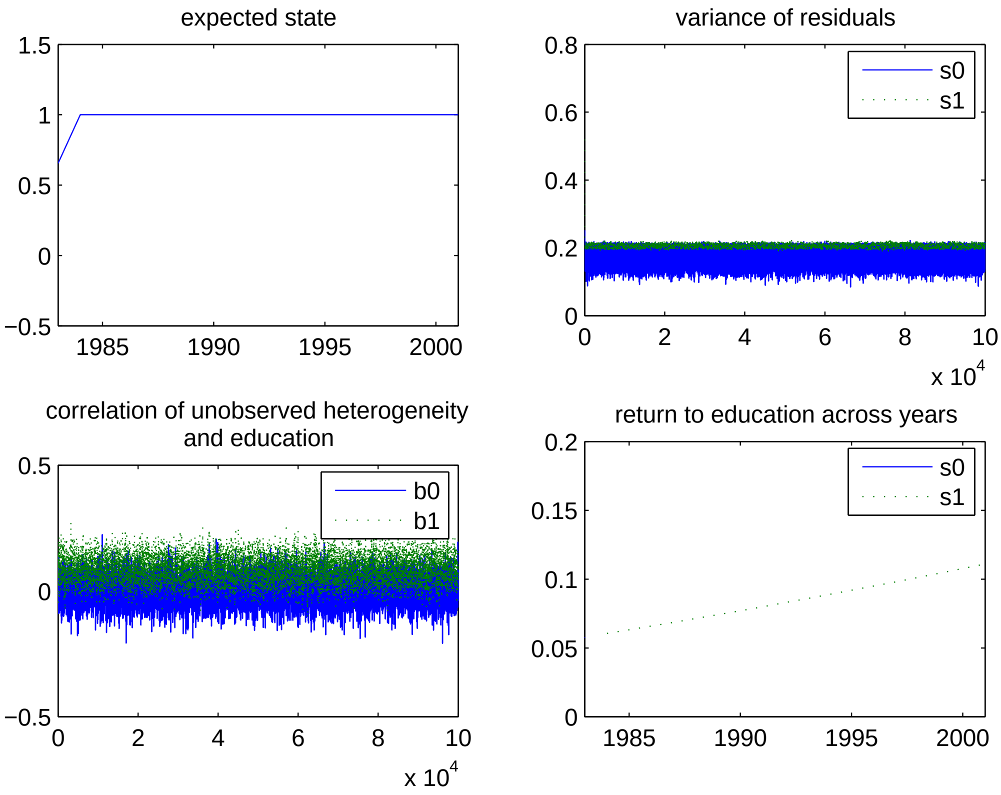

Figure 10: 26-30 years old in 1983

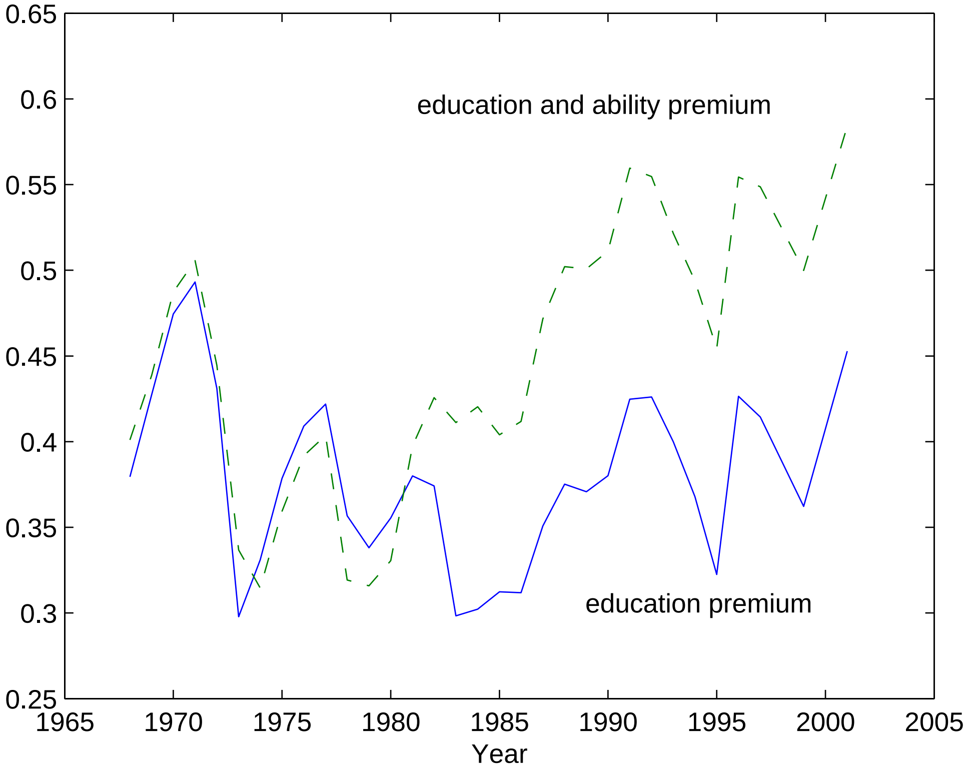

Figure 11: The Ability Bias of the College Premium