Computing the Ability Bias of the College Premium: an Application of Linear Mixed Models with Regime Switching*

Fabrice Murtin†

January 29, 2005

Abstract

This paper derives and tests a procedure to estimate a general class of panel data models that display a regime switching structure, as well as potential correlation between random effects and explanatory variables. An application to wages' dynamics in the United-States from 1968 to 2001 is presented. It is likely that a change of regime has occurred for the earnings process, as documented by a vast body of research on the rise of inequality. This model detects such a change of regime in the early 80s for all cohorts of age on the labor market at that date, and emphasizes that the rise in the return to education along the life cycle has been shifted differently across cohorts. Moreover, as the model accounts for correlation between unobserved heterogeneity and regressors, I quantify the bias in the return to schooling when it is calculated in cross-section: it amounts to circa 30% of the college premium after 1980, and is carried almost entirely by the two or three postwar cohorts of individuals.

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

1 Introduction

There has been a large body of research insisting on the rise of inequality in the United-States over the last two decades. Some correlated events have been exhibited, such as the rise of the skill premium despite the augmented supply of college graduates, the rise of unobserved determinants of wages, the increase of women's participation in the labor force, and the growing inequality across occupational groups. These trends have been recently surveyed in Eckstein and Nagypal (2004), and discussed in several

*I thank Francis Kramarz and seminar participants at CREST (INSEE) for helpful discussions.

†CREST (INSEE) and PSE (joint research unit CNRS-ENS-EHESS-ENPC), Paris. E-mail: fabrice.murtin@ensae.fr.

others (Card-DiNardo (2002), Lemieux (2004), Autor-Katz-Kerney (2004)). Indeed, the classical explanation based on skill-biased technological change turns out to be challenged by some empirical problems (measurement problems in the CPS), puzzles (the slowdown of inequality in the 90s despite the technological acceleration, the closing-gender gap), or challenging explanations (composition effects, minimum wage).

This paper shows that the rise of the college premium since the 80s has been probably smaller than that often depicted in this litterature, and it proposes new estimates of the college premium.

Indeed, most of the papers on this topic estimate a classical Mincerian equation in each year t

$$\ln Y_{i,t} = X_{i,t}\beta_t + u_{i,t}$$

where $Y_{i,t}$ is the real wage of individual i at time t, $X_{i,t}$ a vector of individual characteristics and $u_{i,t}$ a white noise. The modification of some returns β_t is pointed out, especially for education. Nevertheless it is possible that the rise of these coefficients may be associated to modifications of unobserved determinants that are correlated with observed ones: the rise of the return to education might be governed by endogenous effects. Actually Taber (2001) uses a dynamic programming selection model and finds a strong role of unobserved ability in the increase of 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% to the rise of the college premium. They use the variations across groups and time of the within groups income variance to derive this conclusion. This paper argues that a general way to control for this potential problem is to account for correlated random-effects, i.e to estimate a Mincerian equation whose residuals are potentially correlated to the regressors. Therefore the following equation is considered

$$\ln Y_{i,t} = X_{i,t}\beta + b_i + u_{i,t}$$

where b_i is a zero-mean random-variable that is potentially correlated to some regressors such as education or race.

In addition, when the structure of the economy is deeply modified through technological revolutions or institutional evolutions, then the return to human capital, the importance of luck or unobserved heterogeneity might vary consequently. Therefore a change of regime in the US wages' dynamics is a plausible hypothesis that one would like to account for within the framework of a correlated random-coefficient model. All coefficients, as well as the variance of the residuals $u_{i,t}$, are allowed to vary across time through different regimes governed by an unobserved variable which follows an hidden Markov Chain. Markov-Chain Monte-Carlo methods have recently been developed for numerous models[1], and are of particular interest when the likelihood

[1] see Chib (2001) for a review

of the model is complicated and/or leads to numerical unfeasibility. They constitute in this case a natural approach.

The paper develops an estimating 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 estimates conducted on 5-years cohorts of age show that there has been a switch of regime in the early 80s, and none since that date. The profile of the return to education has varied differently accross cohorts, following a downward shift for younger cohorts. I find that for some particular cohorts unobserved ability is significatively correlated with education, leading to a 30% upward bias in the annual rise of the return to education.

2 The model

The following regime switching model is considered:

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

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 $\pi = (\pi_{k,l})_{k,l\,\in\{0,1\}}$

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

From a bayesian standpoint 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 which excludes regime switching, this kind of structure is called "mixed models" in the statistical litterature. 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 random variables (here the b_{i,s_t} variables). These models have been extensively used in many fields, such as animal breeding in biostatistics[2].

Here all parameters are allowed to vary between the two regimes, including the random-effects b_{i,s_t}. From an economic perspective each individual i has two levels of unobserved heterogeneity, and the variances of those heterogeneity distributions differ as well. With $L = 1$ and $z_{i,t} = 1$ for all i and

[2]see Wang (1994) for example.

t, these are reduced to unobserved heterogeneity in the levels of the dependant variable. A larger number of states can be considered, but given the generally small panel length T, this could make estimation harder. At this stage the model is not identifiable because of invariance by permutation of the regimes, so that the additional identification constraint $\sigma_0 < \sigma_1$ is set.

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

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

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 conditional on the data $p\left(\Theta, s_1...s_T \mid Y\right)$.

Modelling the unconditional prior distribution of parameters is an important step since improper priors generally lead to ill-defined posterior distributions[3]. The following priors are natural because they enable closed-form expressions of the posterior distribution:

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

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[4].

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

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

where $\Theta = (\theta_0, \theta_1)$ are the parameters of interest and $g(.)$ is the invariant distribution of the Markov Chain (the eigenvector associated to the eigenvalue 1).

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

[4] they could eventually be treated as random variables just as the parameters of interest above.

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 will draw 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 the sampling of Θ for M times. The resulting distribution $(\Theta_1, \Theta_1, ... \Theta_M)$ is a Markov chain that converges to 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 the correlation between unobserved heterogeneity and regressors to be nonzero. Let us consider the simplest case where $L = 1$ and $z_{i,t} = 1$ for all i and t; the basic econometric results are the following: first, 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. As soon as race or education are included into the set of regressors, this procedure does not guaranty anymore the identification of the model. Instrumental variables have been proposed in case some regressors are exogenous (e.g. Amemiya and MaCurdy (1986)), whereas the Chamberlain approach[5] extracts information from the moments of the variables implied by the model to identify the parameters of interest.

In a Gibbs sampling framework identification of β_{s_t} follows from orthogonality between residuals $u_{i,t}$ and regressors $x_{i,t}$ conditionally on unobserved heterogeneity b_{i,s_t} and variance σ_{s_t}; identification of b_{i,s_t} follows from orthogonality between residuals and regressors $z_{i,t}$ conditionally on β_{s_t}, σ_{s_t}, D_{s_t}, and so on for all other parameters of interest (see the algorithm below). Thus the blocking scheme of the Gibbs sampling, basically an iterative estimation of groups of parameters, exploits the exogeneity conditions without

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

assuming anything on the joint distribution of explanatory variables and unobserved heterogeneity.

A difficulty arises from the missing data, namely the unknown regime states. Different approaches and refinements have been proposed , see for instance Billio, Monfort and Robert (1999) or Chopin (2002). As the target distribution is $p(\Theta, s_1...s_T \mid Y)$ one extracts information on the states $(s_1...s_T)$ by inferring the conditional distribution $p(s_1...s_T \mid \Theta, Y)$. 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 during which one stores the probability distributions $p(s_t \mid Y, \Theta)$ 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 $p(\Theta \mid Y, \hat{s}_1...\hat{s}_T)$ updated via Bayes' rule

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

In practice the algorithm is the following:

Algorithm

1. Step 1 (Forward pass): Set $p(s_1 \mid Y_0, \Theta)$ to be the stationary distribution of π, which is drawn from its unconditional distribution. Compute recursively for $t = \{1, 2...T\}$

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

   where

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

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

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

   where $p(\hat{s}_{t+1} \mid s_t, \pi)$ 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[6] 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

[6] in what follows $M = 100000$ for the real data case.

scheme is very simple because the density of the observations marginalized over b_i is gaussian as well

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

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.

4 Test of the procedure

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

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

Random effects and a time-constant component of explanatory variables have a joint normal distribution with non-trivial covariance matrix. It allows for correlation of random effects between regimes as well as with regressors. For a large number of iterations, the choice of prior parameters does not alter convergence. In practice I take $\rho_0 = \nu_0 = 12$ while R_0 and d_0 vary with priors on the mean variances σ_k^2 and D_k. Then $\forall\ i,j \quad \alpha_{i,j} = 1$ so that the prior transition probabilities are uniform on [0,1]. Each estimation consists in 10000 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 univariate case without unobserved heterogeneity. Gibbs sampling perfectly estimates the underlying parameters, as well as it detects the hidden states though one transition probability is imprecisely estimated (T=30 is somewhat small). Other priors on transition probabilities could maybe improve the latter estimate. The second model introduces unobserved heterogeneity, while the third model examines the multivariate case. All estimates fit the true values, with few exceptions for the transition probabilities. Last model is a matter of concern since it introduces some correlation between explanatory variables and unobserved heterogeneity. With traditional methods, this correlation would have rendered the estimates largely biased, especially in the first regime since the correlation amounts 0.6. In the current framework, all underlying parameters are consistently estimated, as well as most of parameters of secondary interest such as the correlations. Figures 1 and 2 depict the convergence of estimates in the third and fourth models during the first thousand iterates.

5 Accounting for censored data

Due to confidentiality constraints or economic realities many economic individual files are censored, which can lead to serious bias in the estimates if censoring is too important. 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. Thus the wage distribution is left-censored, and the knowledge or the estimate of reservation wages across years can potentially account for this problem. A simple way to correct the former model is use a latent variable model, following Chib (1992). Equations 1 are modified the following way

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

Although the level of censoring might differ across individuals, it is only assumed to vary across time. Interestingly, the algorithm is only marginally modified. The posterior distribution becomes

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

Then Bayes rule provides

$$
\begin{aligned}
f(y_{i,t}^{*} | \hat{s}_t &= k, \theta_k, y_{i,t} = 0) \propto f(y_{i,t}^{*} | \hat{s}_t = k, \theta_k) f(y_{i,t} = 0 | \hat{s}_t = k, \theta_k, y_{i,t}^{*}) \\
&= f(y_{i,t}^{*} | \hat{s}_t = k, \theta_k)\ 1_{y_{i,t}^{*} \le \tau_t}
\end{aligned} \tag{11}
$$

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

A correction of the above algorithm immediately follows from including the unobserved values of $y_{i,t}^{*}$ into the sampling. The former algorithm is unchanged except that it finishes by a data augmentation step to correct 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 and construct $y^*_{i,t}$.

In practice I test this model with three different datasets while τ_t corresponds to the 10% quantile of the Y distribution - thus is constant over time. 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 unobserved heterogeneity. Estimates converge fairly well towards the right values for all the models. A look at Figure 3 shows how the censored data is simulated at final iteration.

6 Application to US wages dynamics from 1968 to 2001

I use the PSID (SRC part) from 1968 to 2001 and study ten different cohorts. The range of age spans from 51-55 years old in 1968 to 26-30 years old in 1988. The dynamics of hourly earnings are modeled for cohort through a Mincerian equation with regime-varying coefficients. This model also accounts for censored data, and potentially endogenous regressors. Its general form is

$$\begin{aligned}
y^*_{i,t} &= a_{0,s_t} + a_{1,s_t}E_i + \left(a_{3,s_t} + a_{4,s_t}E_i\right) A_{it} + \left(a_{5,s_t} + a_{6,s_t}E_i\right) A^2_{it} + a_{7,s_t}D_i + v_{i,t}\\
v_{i,t} &= b_{i,s_t} + \sigma_{s_t}u_{i,t}\\
y_{i,t} &= y^*_{i,t}\ 1_{y^*_{i,t}>\tau_t}\\
b^j_{i,s_t} &\rightsquigarrow \mathcal{N}\left(0, D_{s_t}\right), \quad j = 0,1,2\\
u_{i,t} &\rightsquigarrow \mathcal{N}\left(0,1\right) \qquad (12)
\end{aligned}$$

where covariates are age (A_{it}), squared age (A^2_{it}), education as the number of years of schooling (E_{it}), as well as a Black dummy ($D_{i,.}$), and interaction of age and squared age with education. A potential model for earnings dynamics allows for unobserved heterogeneity interacted with a time trend. As discussed in Meghir and Pistaferri (2004) who use the PSID as well, this would imply long-term autocorrelations of the first-differenced residuals that cannot be empirically detected. This motivates the choice of a sole dimension for unobserved heterogeneity in the model above. On the other hand the latter authors decompose those residuals into a transitory MA noise and a random walk that represents the dynamics of the permanent income - indeed they show that allowing for shocks on permanent income is empirically motivated. Interestingly, they graph the variance of the permanent shocks that bursts in the beginning of the 80s and is diminished afterwards. The model considered here is somewhat compatible with those empirical evidence since it accounts for permanent shocks through the switching structure[7].

In what extent can this framework measure the importance of unobserved skills in determining the college premium? This model enables us to calculate a return to education at a given age for each cohort, as well as an unobserved ability for each individual of that cohort. It shows whether a structural break has affected each cohort, and how the return to education across time has been modified. In a second step, I can aggregate all individuals aged between 26 and 64 and compute the college premium year by year. Whether I introduce or not the individual unobserved ability variables in those cross-sections equations, the estimated return to education will account or not for unobserved ability. The difference between both estimates will indicate how much of the college premium is explained by unobserved ability.

In the data, some observations are missing and others are outliers with high probability, typically when income falls at very low standards. I censore all observations below a certain threshold, namely a log-hourly income of 0 in real terms (1968 prices), which is equal to 4.3$ per hour in 2001 prices. With this assumption, the percentage of censored data is typically around 10%, and never exceeds 15%.

Table 3 provides some elementary descriptive statistics. The increase of 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 of the college premium at that period.

The global impact of the regime transition on the return to education is adressed by Figures 4 to 13: for old cohorts in 1968 and young cohorts 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 variance of residuals and unobserved heterogeneity, and different profiles of the return to education across age: for the cohorts experiencing a switch of regime, the shift of the return to education's profile is directed towards depreciation for most of them. Two cohorts, those aged 26-30 in 1973 and 1978, display significant levels of correlation between education and unobserved heterogeneity in the second regime. Table 4 indicates the level of correlation and the size of the ability premium, i.e. the mean of unobserved ability conditional on holding a College degree.

Once each worker is attributed her level of unobserved heterogeneity in each year, it is possible to compare the coefficient of a College dummy

[7] the variance of the permanent shock will be distributed as a Dirac centered on the date of regimes change.

in a Mincerian regression run in cross-section, including or excluding the variable unobserved heterogeneity. Last picture shows the result: the true return to college is approximatively 30% lower than that usually calculated. This conclusion is quite similar to that of Chay and Lee (2000) who used inter-groups variations of the within-group earnings variances to identify the ability premium.

7 Conclusion

This paper presents a general class of panel models that encompasses correlated random coefficients models, regime switching models, as well as individual-specific slopes models ("mixed models"). It proposes a Gibbs sampling procedure to estimate this kind of models, tests and validates it for both censored and uncensored data. Then I run an application to the US wages' dynamics, which consists in computing the return to schooling with and without accounting for the correlation between schooling and unobserved skills. It is shown that the ability premium represents circa 30% of the college premium and is carried by two particular cohorts. An more careful analysis of the determinants of unobserved ability for these two cohorts, including parental and spatial determinants, is currently performed.

References

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

[2] Autor, D., L. Katz and M. Kearney (2004), Trends in U.S. Wage Inequality: Re-Assessing the Revisionists, mimeo.

[3] 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.

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

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

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

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

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

[9] 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.

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

[11] 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.

[12] 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.

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

[14] Lemieux, T. (2004), Increasing residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? CLE Working Paper No.69.

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

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

[17] 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.

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

[19] Wang, C.S. (1994), Bayesian Analysis of Mixed Linear Models via Gibbs Sampling With an Application to Litter Size of Iberian Pigs, *Genetique, Selection, Evolution*, 26, 1-25.

Table 1 - Test on 4 datasets

		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	$\binom{200}{100}$
	$\hat{b}$	$[\underset{(0.02)}{0.78}\ \underset{(0.04)}{1.01}\ \underset{(0.05)}{1.19}]$	$[\underset{(0.04)}{0.53}\ \underset{(0.06)}{1.03}\ \underset{(0.06)}{1.39}]$	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	$\binom{200}{30}$
	$\hat{b}$	$[\underset{(0.01)}{0.79}\ \underset{(0.03)}{0.98}\ \underset{(0.05)}{1.18}]$	$[\underset{(0.03)}{0.62}\ \underset{(0.05)}{0.96}\ \underset{(0.04)}{1.37}]$	0.22 (0.01)	0.40 (0.02)	$\begin{bmatrix} \underset{(0.02)}{0.20} & \underset{(0.03)}{0.15} \\ & \underset{(0.05)}{0.48} \end{bmatrix}$	$\begin{bmatrix} \underset{(0.07)}{0.46} & \underset{(0.05)}{0.09} \\ & \underset{(0.07)}{0.43} \end{bmatrix}$	0.39 (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	$\binom{200}{30}$
	$\hat{b}$	$[10.06\ \ 0.023\ \ 0.28\ \ 0.08]$ (0.02) (0.001) (0.04) (0.03)	$[10.13\ \ 0.048\ \ 0.42\ \ 0.18]$ (0.03) (0.001) (0.05) (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)	0.42 (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 third and fourth regressors. The correlation between random-effects is equal to 0.5.

[3]random coefficients correspond to the third and fourth regressors. The correlation between random-effects is equal to 0.2.

[4]random coefficients correspond to the third and fourth regressors. The correlation between random-effects is equal to 0.8.

Table 3- Descriptive statistics

Born in $\longrightarrow$	1913-1917	1918-1922	1923-1927	1928-1932	1933-1937	1938-1942	1943-1947	1948-1952	1953-1957	1958-1962
$\%\ of\ Censoring$	11.4	9.4	7.1	6.8	14.4	13.2	9.8	8.0	9.9	8.3
% of HSD	48.1	38.3	32.5	25.2	23.7	13.4	12.8	4.8	6.3	5.4
% of HSG	29.9	28.2	29.9	36.8	34.5	35.7	28.6	28.7	38.8	39.0
% of SC	10.7	15.2	12.4	14.3	16.1	23.0	21.3	23.7	24.0	21.9
% of CG	4.9	12.1	15.6	10.6	15.3	17.5	25.3	24.0	18.1	23.1
% of PG	6.3	6.2	9.6	13.1	10.5	10.3	12.0	18.7	12.8	10.5
N	125	145	143	154	128	115	206	301	339	333

HSD=high-school dropouts, HSG=high-school graduates, SC=some college, CG=college graduates, PG=post-graduates

Table 4 - The Bias of the College Premium

	Regime 1			Regime 2			Correlation of unobserved heterogeneity between regimes
	Years	$\rho\,(E_i, b_{i,s_0})$	E $(b_{i,s_0}\| E_i \geq 16)$ $-$E $(b_{i,s_0}\| E_i < 16)$	Years	$\rho\,(E_i, b_{i,s_1})$	E $(b_{i,s_1}\| E_i \geq 16)$ $-$E $(b_{i,s_1}\| E_i < 16)$	
51-55 in 1968	-	-	-	1968-1976	0.01 (0.10)	0.19 (0.11)	-
46-50 in 1968	-	-	-	1968-1981	0.01 (0.08)	0.06 (0.06)	-
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)	-
26-30 in 1988	-	-	-	1988-2001	-0.02 (0.05)	0.06 (0.04)	-

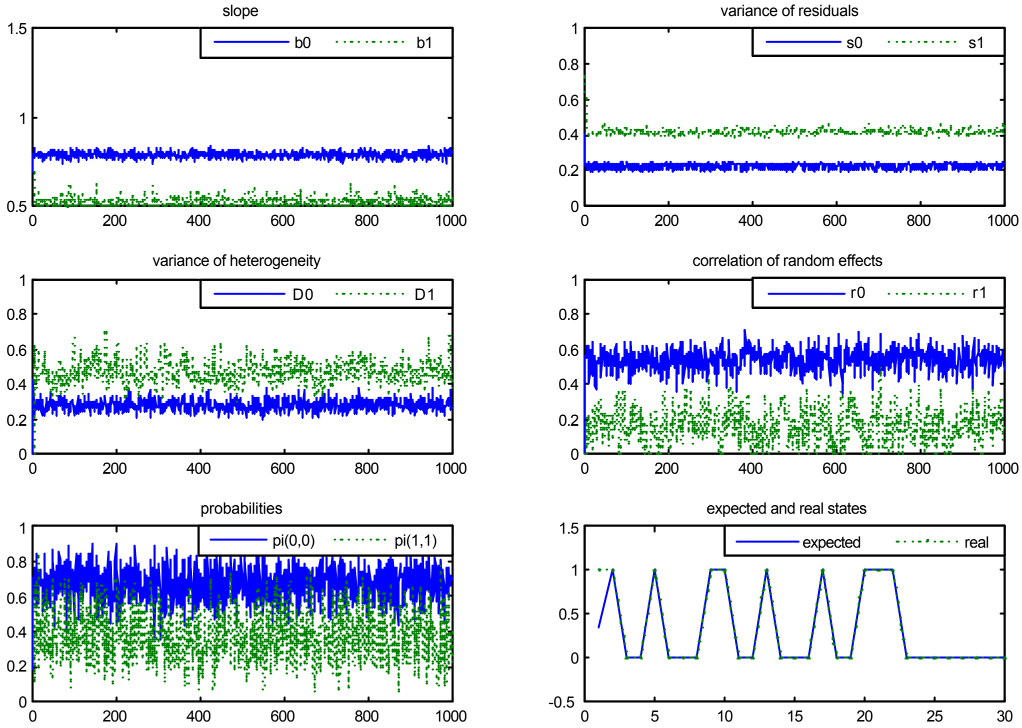

Figure 1: Model III (non-censored)

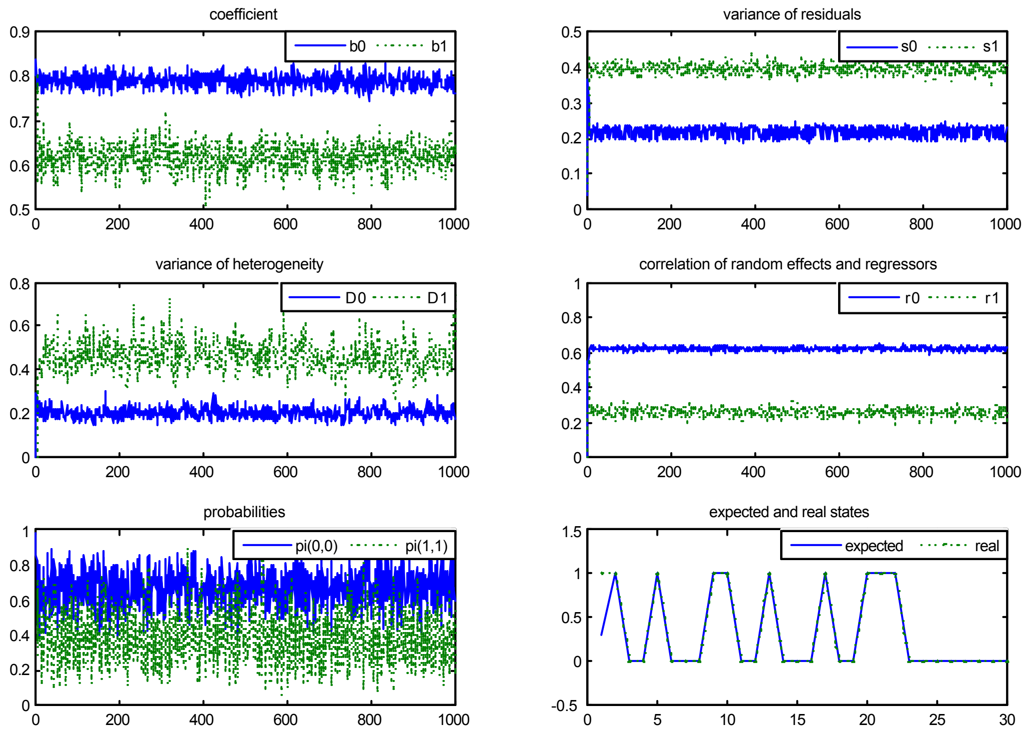

Figure 2: Model IV (non-censored)

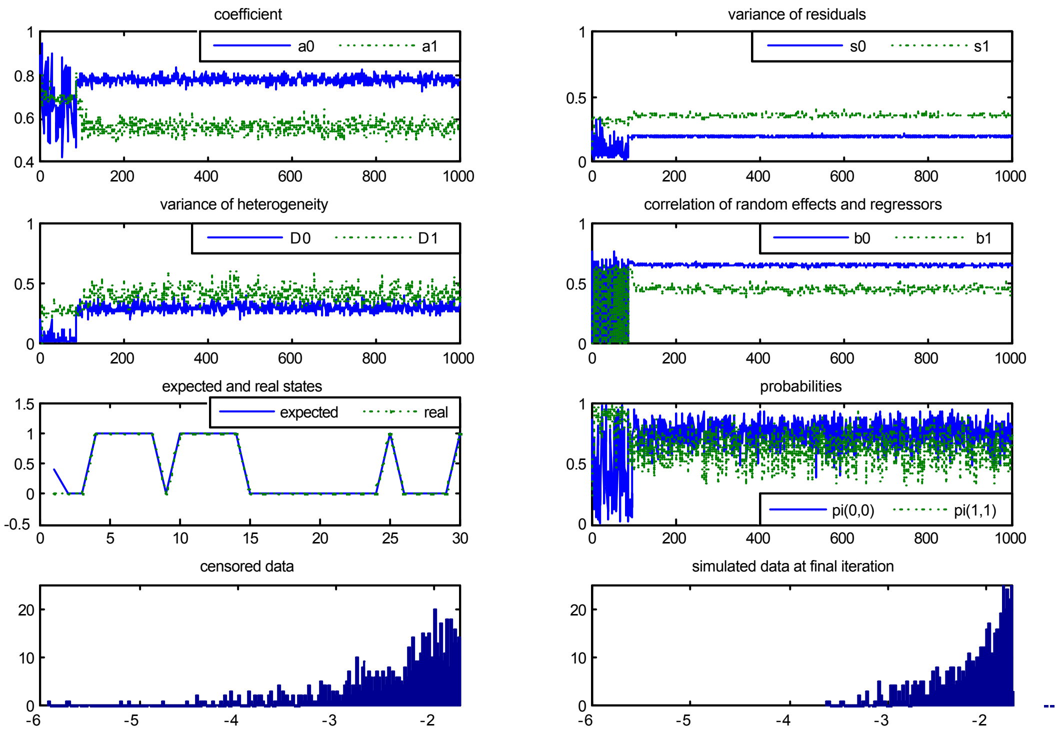

Figure 3: Model III (censored)

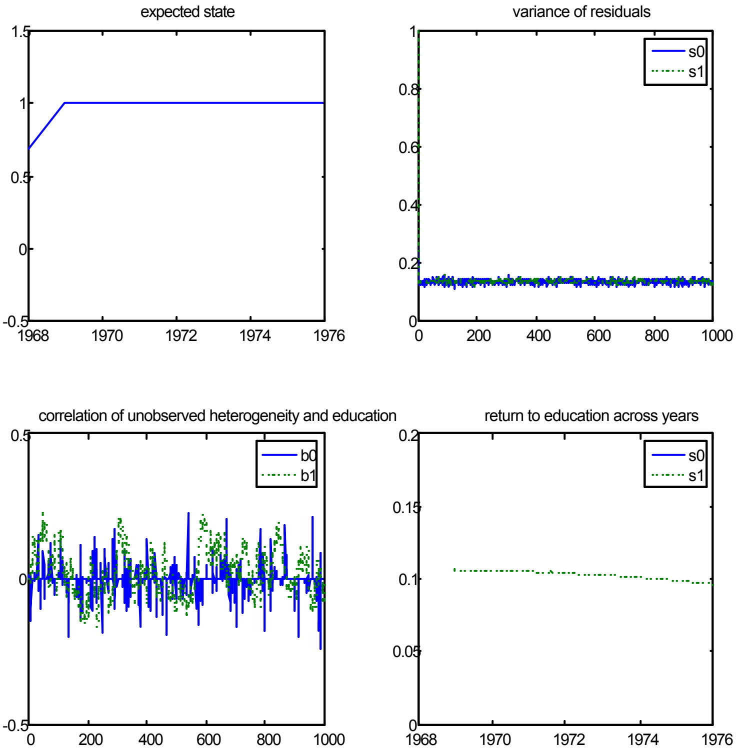

Figure 4: 51-55 in 1968

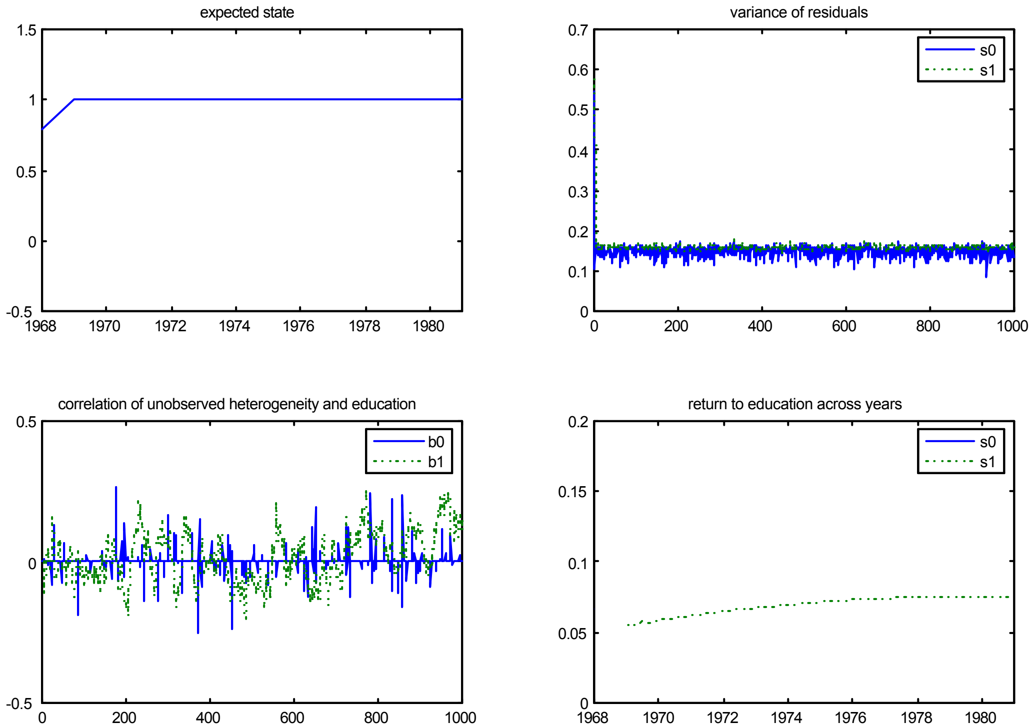

Figure 5: 46-50 in 1968

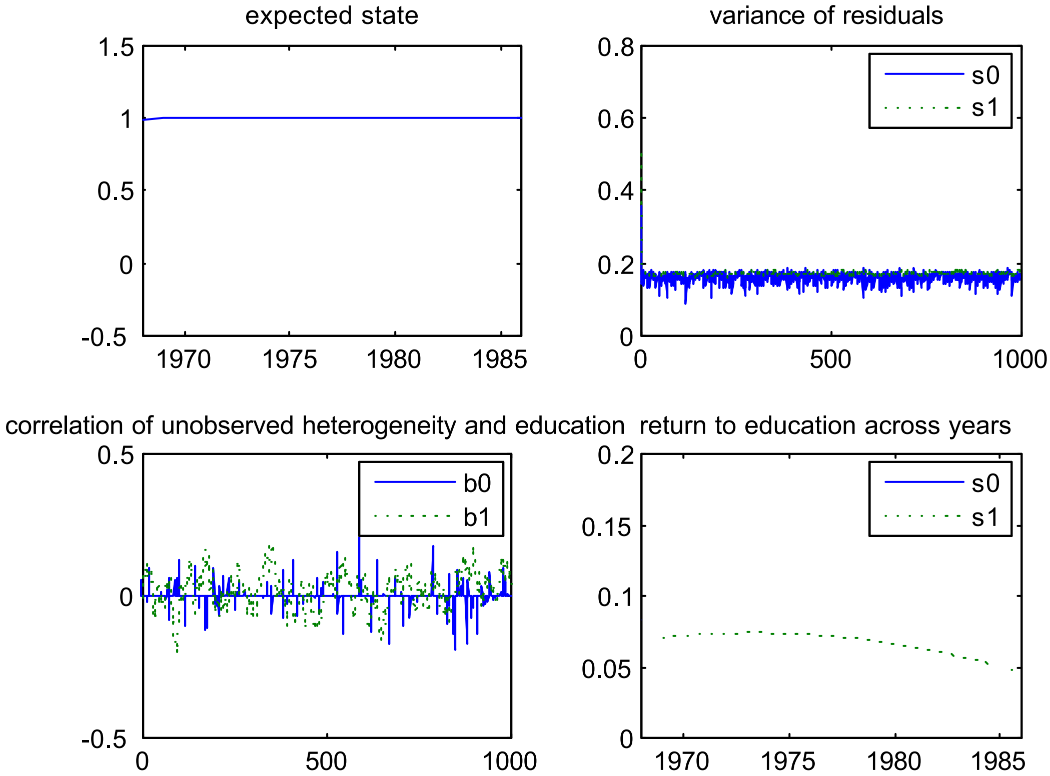

Figure 6: 41-45 in 1968

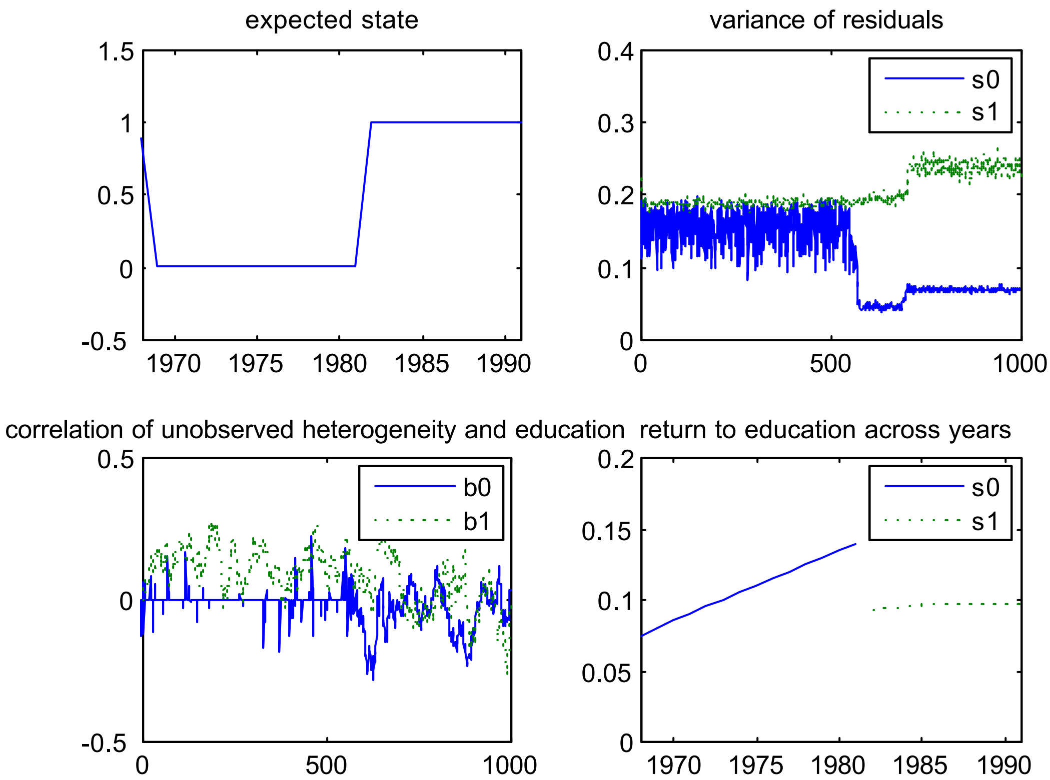

Figure 7: 36-40 in 1968

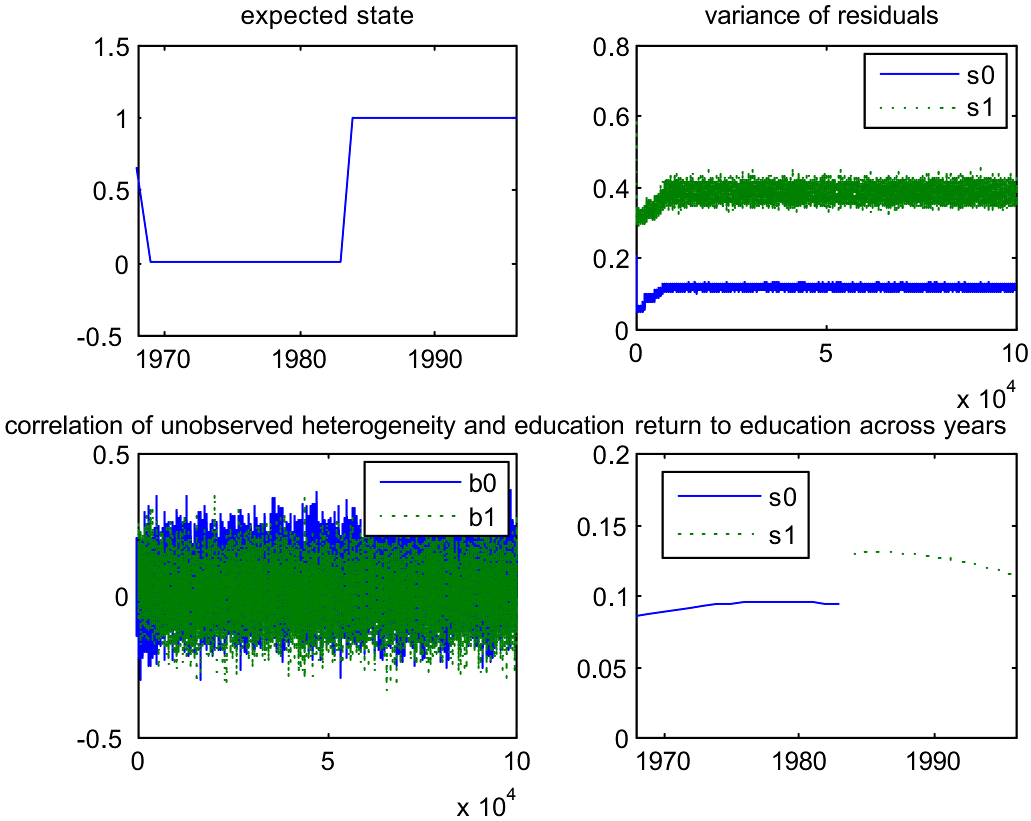

Figure 8: 31-35 in 1968

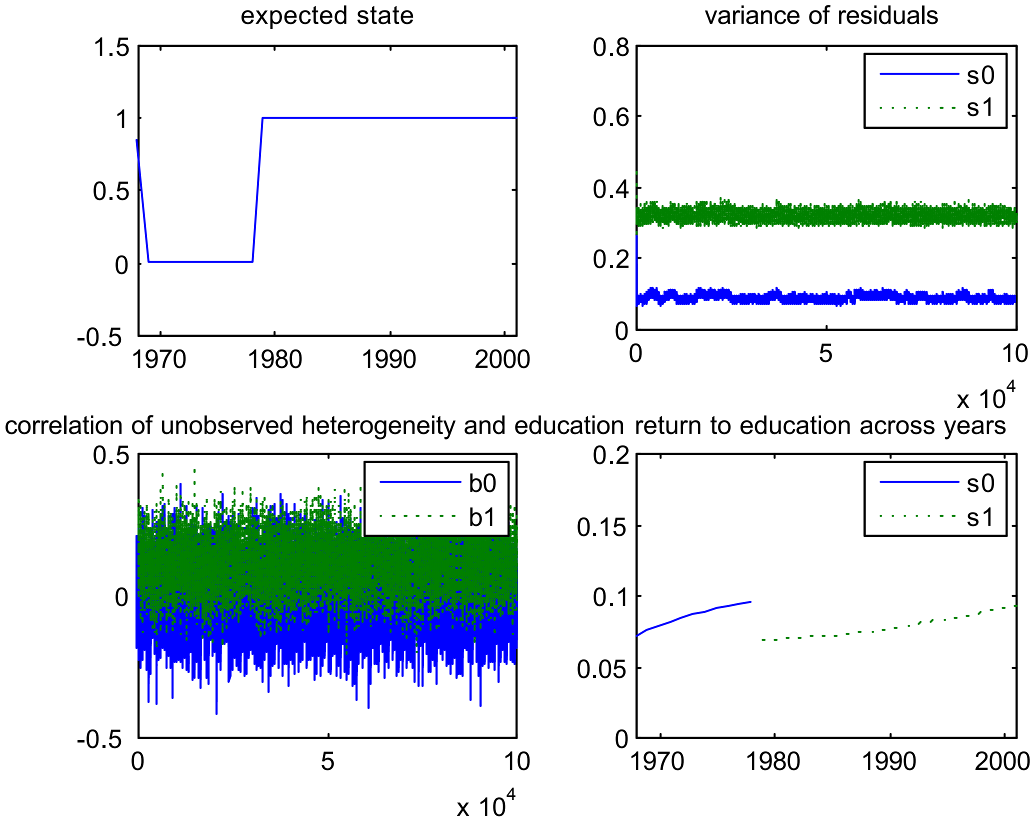

Figure 9: 26-30 in 1968

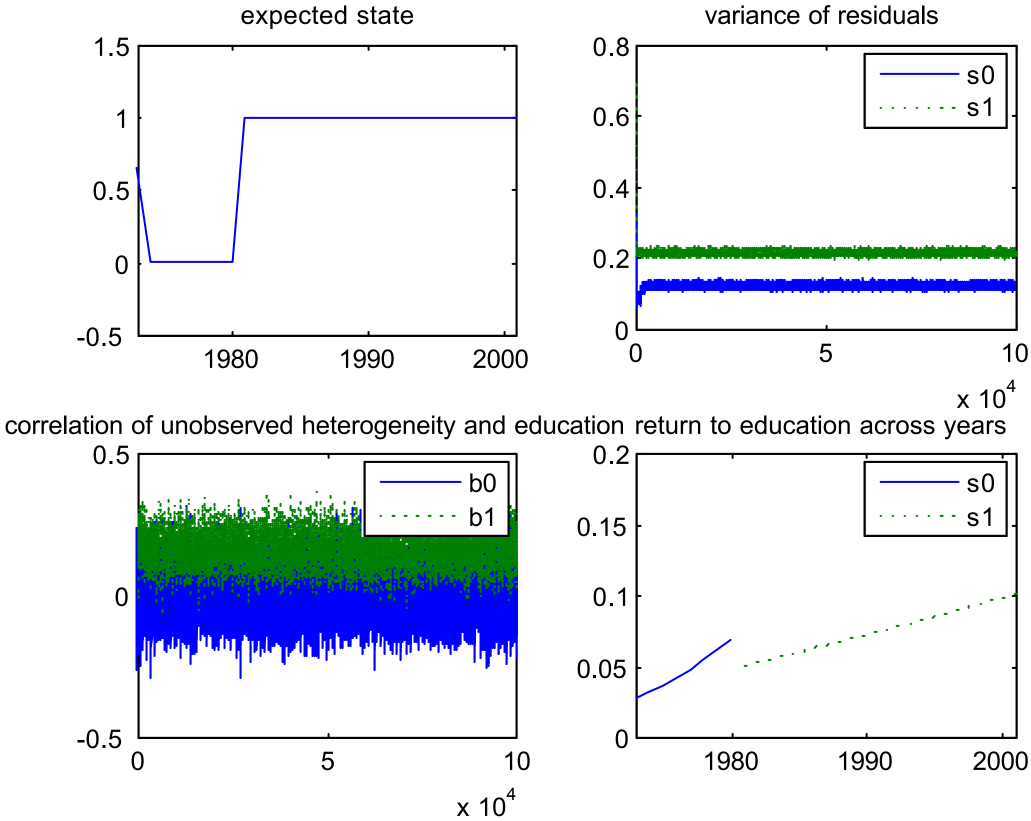

Figure 1: 26-30 in 1973

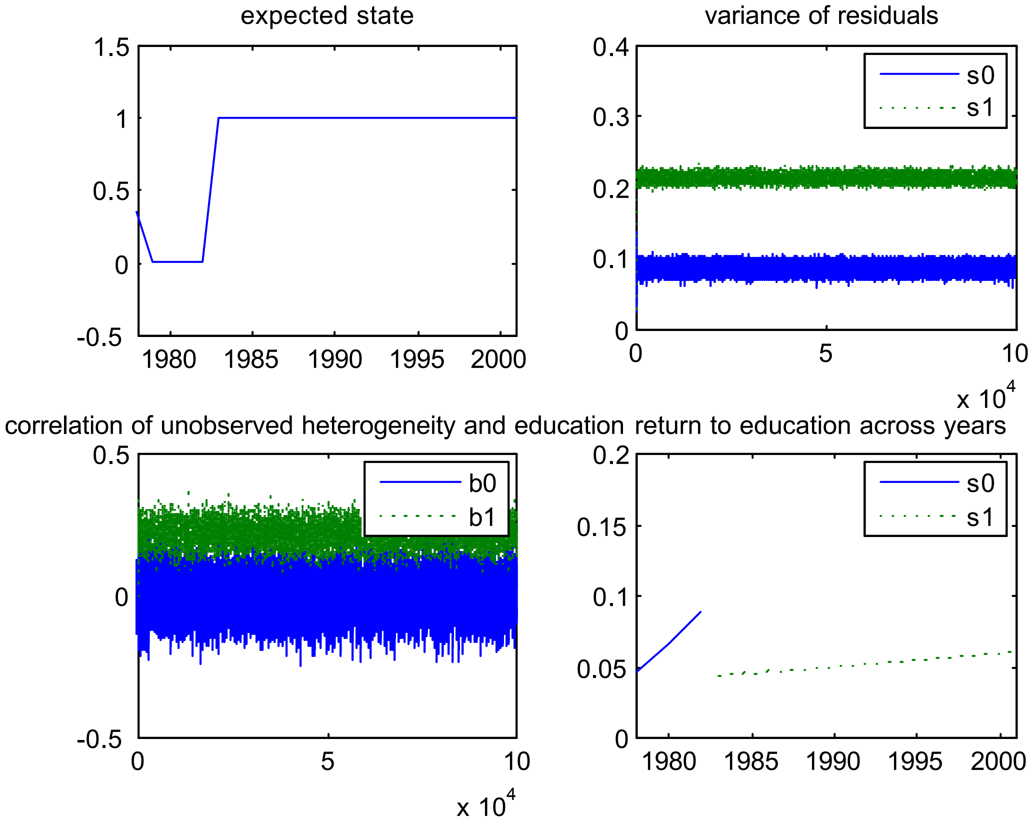

Figure 2: 26-30 in 1978

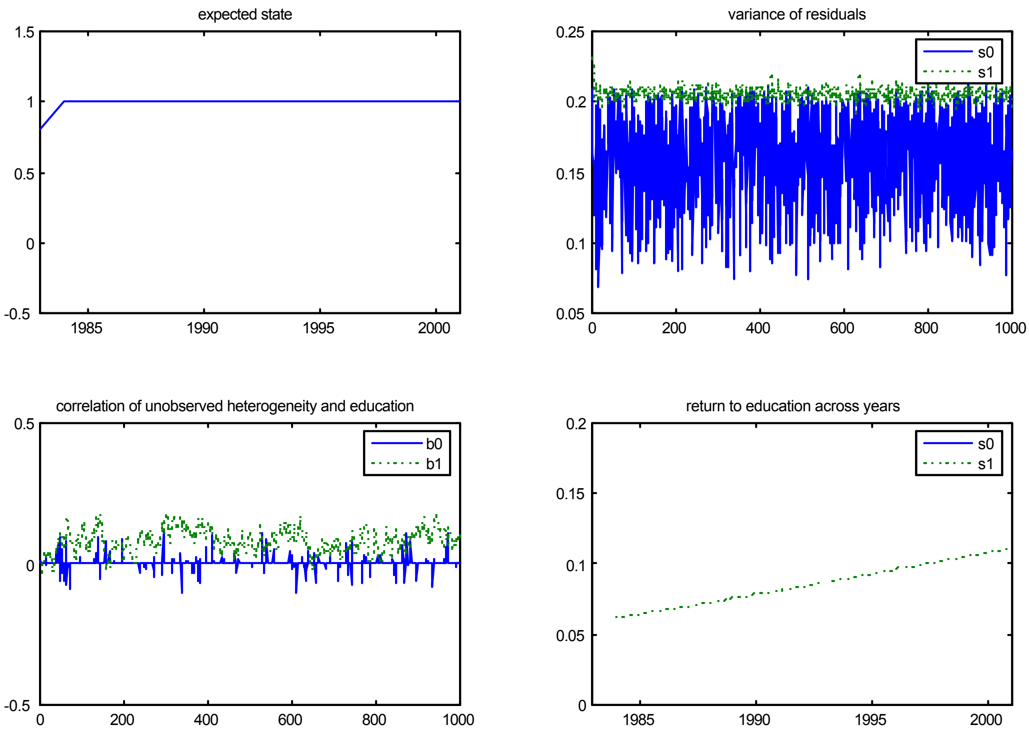

Figure 3: 26-30 in 1983

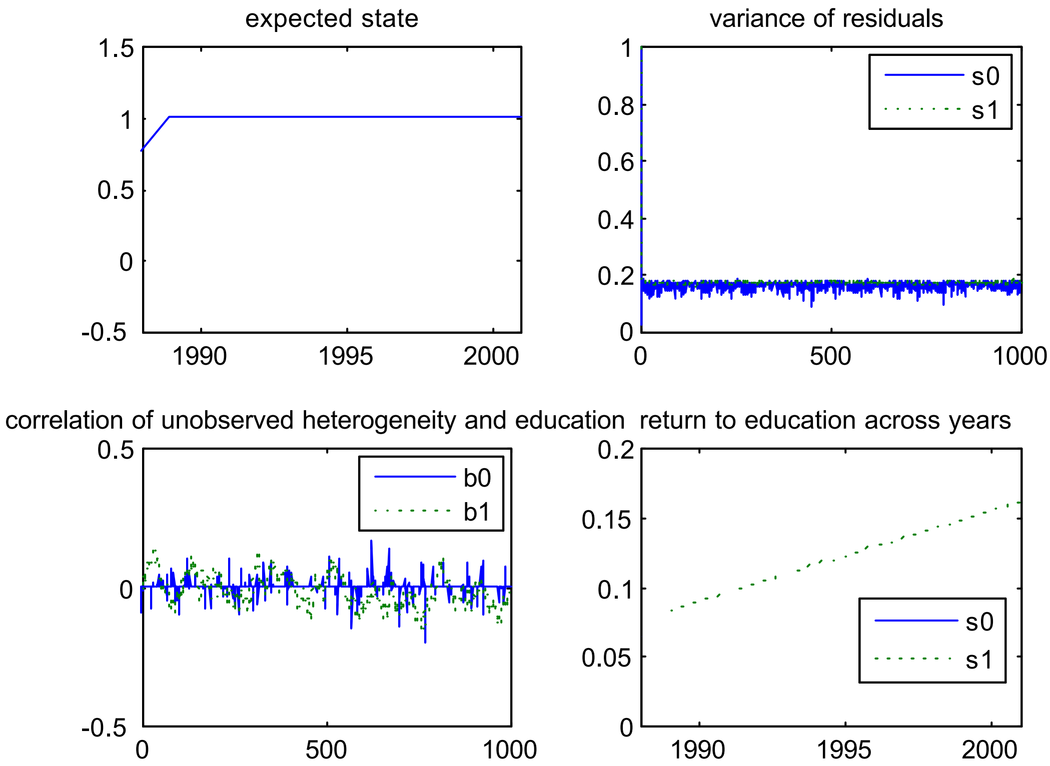

Figure 4: 26-30 in 1988

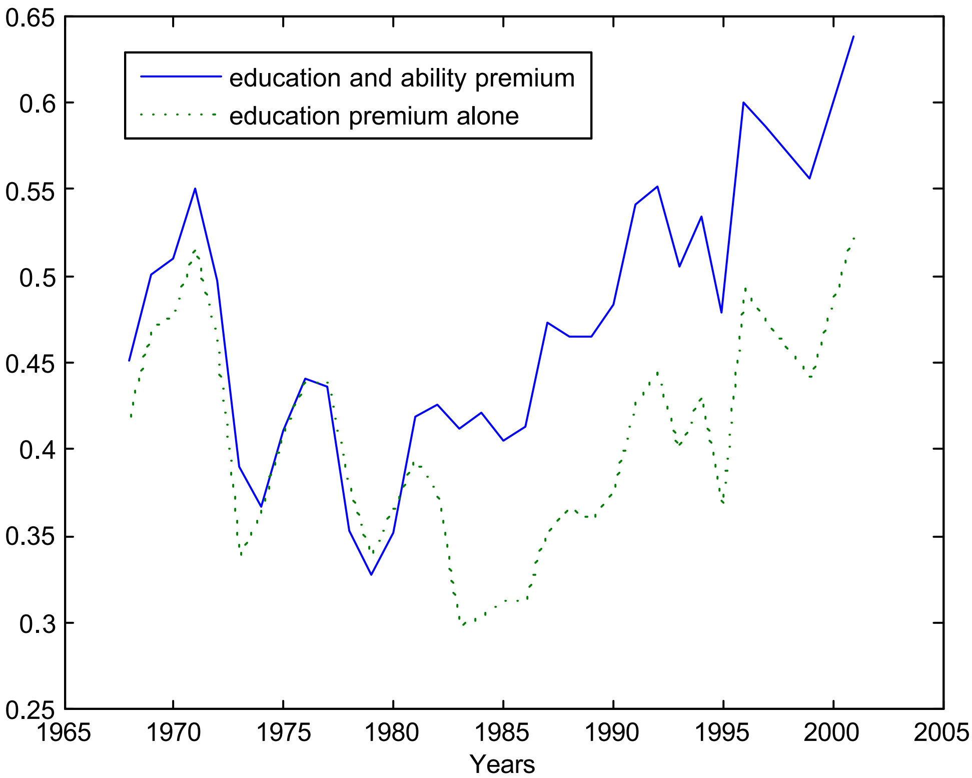

Figure 5: The Ability Bias of the College Premium