Simulating a progressive loan system for Spain with real labor market data

Antonio Cabrales∗ Maia Guell † Rocio Madera ‡ Analia Viola §

June 15, 2017

Preliminary and Incomplete. Do not distribute. Comments Welcome.

Abstract

This paper analyzes the distributional implications of introducing income-contingent student loans subsidized by the government to finance higher education in Spain. For that purpose, we simulate the lifetime earnings of graduates using a model of earnings dynamics and employment transitions. We estimate the model to match different dynamic relations observed in administrative tax and social security data. One challenge of this analysis for the case of Spain is the disproportionate share of graduates that start their working histories in temporary forms of employment, which translate into lower-paying and more volatile jobs. In contrast with previous literature, we thus model permanent and temporary contracts separately. Our model can replicate the dynamics of employment and earnings remarkably well. We use the simulated profiles to calculate the burden of introducing public loans for individuals at different points of the earnings distribution and for the government. We find that (1) our proposed structure is highly progressive under all specifications, with the top quarter of the distribution paying close to the full amount of the tuition and the bottom 10% paying almost no tuition; and (2) the cost for the government is between 90% and 60% lower than under the current system.

JEL Codes: I22, I23, I24

∗UCL, FEDEA; acabrales@ucl.ac.uk
†University of Edinburgh, FEDEA; maia.guell@gmail.com
‡University of Minnesota; made0088@umn.edu
§FEDEA; aviola@fedea.es
1 Introduction

The university system in Spain has suffered continuous budget cuts and, given the current fiscal situation, it is not likely that will reverse soon. It is likely that the Spanish university system will run out of public resources in the near future. According to Education at a Glance (OECD, 2016), Spain is at the tail in education spending compared to other OECD countries. In particular, education accounts for 11% of total public spending across OECD countries. In Spain, the budget assigned to education is 8%. Moreover, the percentage allocated to scholarships and student aid is quite small, below 5% overall and 2% for tertiary education.

Several studies have focused on aspects of university financing in Spain (Beneito, Boscá & Ferri 2016, De la Fuente & Jimeno 2009, Del Rey & Racionero 2012, Escardíbul Ferrá & Pérez Esparrells 2013, Mora, García Montalvo & García-Aracil 2000, Nuñez 2006, Vázquez 2013). However, the analysis of the impact of student loans in Spain has been limited to one paper, which focuses on the specific case of loans-to-masters that was implemented in 2007 and lasted only until 2011 (Callado Muñoz, Del Rey & Utrero González 2015). The loans-to-masters program did not prove to be very successful, partly due to the lack of consistency of the conditions (interest rate, repayment horizon, and the like) across years. There was also a grace period stipulated independently of the income level and a monthly fixed repayment, which imposed a heavy burden to graduates at the lower end of the income distribution.

Under this context, it is crucial to think about alternatives sources of resources and ways of financing higher education in Spain. The United Kingdom (UK hereafter) has been one of the precursors, among other countries in Europe, in designing a progressive loan system subsidized by the government to finance higher education. The UK has undergone 3 main reforms during the last 20 years that have been related to increasing fees and implementing a loan system (1998, 2007 and 2012 reforms). The evidence so far reveals that the system has been working reasonable well in the UK, especially in its progressive nature (Dearden et al. 2008, Azmat & Simion 2017).

Based on this experience, the main motivation of this paper is to evaluate how a loan system similar to that in the UK starting in 2007 would work for Spain to finance higher education and study the implications for lifetime income, the burden of repayments on workers, and the cost to the government. Indeed, the main objective of this study is to
explore different loan policies and the effects along the income distribution.

The experiment consists of increasing fees and at the same time applying a progressive loan system subsidized by the government with these particular characteristics: fees can be deferred until starting to work, repayments will depend on ex-post labor income and minimum exemption, and there will be a debt write-off and low interest rates. In sum, the different options will result from the interaction of the key loan parameters of the system: (1) Amount of debt at the end of college (total fees); (2) Loan interest rate; (3) Repayment rate i.e. percent of earned labor income to repay debt; (4) Exemption income level, understood as the threshold at which people have to pay the repayment in a given year, that is if income is below this threshold there is no repayment that year; (5) Debt write-off, or number of years after which the debt is forgiven.

One challenge of our analysis that is not as pervasive in the case of the UK is the disproportionate share of graduates that start their working histories in temporary forms of employment (nearly 80%), which tend to involve a lower salary (and, therefore, a higher likelihood of being above the repayment threshold) and higher lifetime uncertainty. In contrast with previous literature and that focused on the UK case, we model permanent and temporary contracts separately.

In our analysis, we simulate the lifetime earnings of graduates using a model of earnings dynamics and employment transitions. We estimate the model to match different dynamic relations observed in Social Security and Tax Records (Muestra Continua de vidas Laborales y el Módulo Fiscal). In particular, we estimate the age profile of the persistence and variance of labor income shocks, in a flexible autoregressive process of earnings dynamics that allows for individual fixed effects and profile heterogeneity, given the panel dimension of the data and the large number of observations. Employment transition probabilities are estimated using probit regressions on past income and contract duration.

Our model can replicate the dynamics of employment and earnings remarkably well. We use the simulated profiles to calculate the burden of introducing public loans for individuals at different points of the earnings distribution and for the government under different combinations of the aforementioned parameters. We find that (1) our proposed structure is highly progressive under all specifications, with the top quarter of the distribution paying close to the full amount of the tuition and the bottom 10% paying almost no tuition; and (2) the cost for the government is between 90% and 60% lower than
under the current system.

2 Data and Sample

2.1 The Data: Social Security and Tax Records
We use administrative data from the Continuous Sample of Working Histories (MCVL hereafter, for its acronym in Spanish) on earnings and working histories of Spanish workers. The data is provided by the Spanish Social Security Administration in cooperation with the IRS counterpart in Spain. In this section we give an overview of the data source and a description of our sample. For the database specifics and more details we refer to Section 2 in ?.

The MCVL consists of a 4% representative random sample of all workers affiliated with the social security administration within a given year between 2004 and 2015. Besides, starting in 2005, the sample has a panel design: all individuals present in each wave subsequently remain in the sample. Retroactive information on the whole working history is provided as early as 1962 for work variables and 1980 for earnings. ? show that the sample is representative at least since the late 1980s. The main drawback is that earnings data is bottom- and top-coded. We, therefore, complement the earnings data with the tax supplement provided by the IRS and matched to the Social Security records. The tax supplement contains non-top-coded information on annual earnings but does not include such detailed job characteristics or earnings a higher-frequency levels.

2.2 Sample Selection
We select college graduates that are at least 22 and at most 60 years old.

2.3 Main Variables

Earnings. The earnings data are extracted from the "Annual summary of retentions and payments for the personal income tax on earnings, economic activities, awards, and income imputations" (known as Modelo 190). All employers are required to fill out Modelo 190 with the total compensation paid to each of their employees during the year, independently of whether or not they pay labor income taxes. To obtain a measure
of total annual labor earnings, we add all the incomes that correspond to each worker during the reference year.

**Annual Employment Status.** Given that our period of observation is one year, it is not uncommon to find workers that hold different simultaneous jobs or that change jobs within the same year. In some cases, some of those contracts are temporary and some permanent. This poses a challenge when defining employment spells at the annual level. To be able to exploit the daily employment information back to the 1980s, we define employment status in terms of share of annual time spent in each kind of job: permanent, temporary, or none. Workers who have zero annual earnings or earn less that the corresponding amount to a month minimum-wage salary are considered unemployed.

3 Age-Specific Earning Dynamics

3.1 Measuring Life-Cycle Earnings Risk

In this section, we pose and estimate a stochastic income process to measure the uncertainty faced by different workers at each age. We group workers based of their exposure to temporary forms of employment during the first 10 years of their working lives. The statistical framework follows ?.

More specifically, let $Y_{iat}$ be Annual earnings for individual $i$ at age $a$ in year $t$. We assume $\log Y_{iat}$ is given by

$$\log Y_{iat} = \beta X_{iat} + y_{iat}, \quad (1)$$

where $X_{iat}$ is a vector of observable characteristics that includes a quartic polynomial in age, year dummies, and region dummies. $\beta$ is assumed constant in time.

The residual income form the first stage $y_{iat}$ is decomposed into three components: (1) an individual deterministic component $\alpha_i + \gamma_i a$, formed by a fixed effect and a linear trend; (2) a stochastic persistent component $z_{iat}$, modeled as an AR(1) with persistence $\rho < 1$; and (3) a stochastic transitory component $u_{iat}$, represented by a MA(1). The specific structure is given below by equations (2)-(5):
$$y_{iat} = \alpha_i + \gamma_i a + u_{iat} + z_{iat} \quad \alpha_i \sim N(0, \sigma_{\alpha,a}^2), \gamma_i \sim N(0, \sigma_{\gamma,a}^2) \quad (2)$$

$$u_{iat} = \varepsilon_{iat} + \theta \varepsilon_{i,a-1,t-1} \quad \varepsilon_{iat} \sim N(0, \sigma_{\varepsilon,a}^2) \quad (3)$$

$$z_{iat} = \rho z_{i,a-1,t-1} + \eta_{iat} \quad \eta_{iat} \sim N(0, \sigma_{\eta,a}^2) \quad (4)$$

$$z_{i0t} = 0, \quad \varepsilon_{i0t} = 0. \quad (5)$$

To capture the evolution of uncertainty over life, parameters $\rho$, $\sigma_\varepsilon$, and $\sigma_\eta$ are functions of age. The exact form of age dependence will be discussed next in Section 4.

### 4 Estimation

Our baseline model assumes that $\sigma_{\varepsilon,a}^2$, $\sigma_{\eta,a}^2$, and $\rho$ are all cubic functions of age, and $\theta$, $\sigma_\alpha$, and $\sigma_\gamma$ are fixed across ages.

We minimize the distance between the empirical and the model-implied closed-form covariance matrix to estimated the parameters, using Generalized Method of Moments with efficient weighting matrix, to estimate the parameters.

$$(\theta, \sigma_\alpha, \sigma_\gamma, \rho(\cdot), \sigma_\varepsilon(\cdot), \sigma_\eta(\cdot)) \quad (6)$$

Table I summarizes the average values for the parameters in (6). Before analyzing the dynamics or earnings, it is worth highlighting two points from the time-invariant and average effects.

<table>
<thead>
<tr>
<th></th>
<th>All Workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{\sigma}_\eta$</td>
<td>Var(Persistent Shocks)</td>
</tr>
<tr>
<td>$\bar{\sigma}_\varepsilon$</td>
<td>Var(Transitory Shocks)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Persistence</td>
</tr>
<tr>
<td>$\theta$</td>
<td>MA Coefficient</td>
</tr>
<tr>
<td>$\sigma_\alpha$</td>
<td>Var(Fixed Effects)</td>
</tr>
<tr>
<td>$\sigma_\gamma$</td>
<td>Var(Linear Trend)</td>
</tr>
</tbody>
</table>

Table I: Average Parameters, Males, College Graduates
4.1 Life-Cycle Earnings Dynamics

To ease the interpretation, in this section we present the resulting life-cycle profiles implied by the estimates. We include the average of the estimates in table XX. Our reference group is college-educated males.

![Figure 1: Persistence](image1)

![Figure 2: Variance of Persistent Shocks](image2)

![Figure 3: Variance of Transitory Shocks](image3)
5 Simulating Life-Cycle Earnings Profiles

In this section, we pose a model of employment transitions and earnings dynamics.

The exercise we perform is as follows: We ask the question of what would be the impact of implementing an educational reform of higher education similar to that of the United Kingdom in the early 2000s. That entails a substantial increase in tuition and fees (from roughly $1000 per year to roughly $5000 per year), accompanied by a menu of publicly subsidized income-contingent loans which repayment is linked to the future income of the worker.

As an illustration, we start with a simple statistical model of employment transitions a leave a more structural analysis for future work.

5.1 Model

Employment Transitions

At each point in time, a worker can be in one of three employment statuses: unemployed (U), permanent (P), and temporary (T). Π is the transition matrix that determines changes between statuses.

\[
\begin{bmatrix}
  \pi' \\
  T' \\
  U'
\end{bmatrix} =
\begin{pmatrix}
  \pi^{PP} & \pi^{PT} & \pi^{PU} \\
  \pi^{TP} & \pi^{TT} & \pi^{TU} \\
  \pi^{UP} & \pi^{UT} & \pi^{UU}
\end{pmatrix}
\begin{bmatrix}
  P \\
  T \\
  U
\end{bmatrix}
\]

We estimate these transitions with probit regressions by regressing a dummy variable that takes 1 in the case of a transition on a constant, a quartic in age, and additional covariates depending on the type of the transition. In particular,

\[
\pi^{s_{t-1} \to s_t} = \begin{cases} 
\Phi (\beta_1 y_{t-1} + \beta_2 y_{t-1}^2) 
& \text{if } (s_{t-1}, s_t) \in \{(P, T), (P, U), (T, U)\} \\
\Phi (\beta_1 dur_{t-1} + \beta_2 dur_{2t-1}) 
& \text{if } (s_{t-1}, s_t) \in \{(U, P), (U, T)\} \\
\Phi (\beta_1 y_{t-1} + \beta_2 y_{t-1}^2 + \beta_1 dur_{t-1} + \beta_2 dur_{2t-1}) 
& \text{if } (s_{t-1}, s_t) \in \{(T, P)\}
\end{cases}
\]
where
\[
dur_{1t} \equiv I \{ s_t = s \mid s_{t-1} = r \} \text{ and } s \neq r \\
dur_{2t} \equiv I \{ s_t = s \mid s_{t-1} = s \}
\]
denotes whether the worker had spent one of more years in the initial state. For example, in the case of a transition from unemployed to permanent \((s_{t-1}, s_t) = (U, P)\), \(dur_{1t-1} = 1\) if the worker was unemployed only for one period last year, and \(dur_2 = 1\) if the worker had been unemployed for two or more periods last year.

All specifications include a constant and a quartic in age in the set of regressors, that we have omitted for the sake of exposition.

At the beginning of an employment spell within a contract, each worker draws a level of earnings determined by its previous status and age. This type of transition earnings will be explained in detailed below.

**Transition Earnings Dynamics**

The process defined in ?? characterizes the evolution of earnings if the worker does not change status. We use the parameters corresponding to the *young temporary* when the last year’s status is temporary and to *young permanent* when the last year’s status is permanent.

Whenever there is a change of status of the type \(PT, TP, UT\), or \(UP\), we estimate the new initial earnings as a function of age, duration of previous spell, and past earnings.

More specifically, let
\[
y_{ss'}^t \equiv y(s_t = s' | s_{t-1} = s)
\]
denote log earnings at \(t\) of a worker who just moved from status \(s\) to \(s'\).

We pose the following specification for the log of earnings in the first year at the new status:
\[
\log Y_{ss'}^t = \beta_1 dur_{1t-1}^s + \beta_2 dur_{2t}^r + \beta_2 y_{t-1}^L + \xi_t,
\]
where $y_{t-1}^L$ denotes the level of earnings in the previous status $s$ if $s \in \{P, T\}$, the last earnings observed if $s = U$ and the worker has been unemployed for only 1 year, or a dummy that equals 1 indicating that the last level of earnings is missing in the case of $s = U$ and the worker has spent 2 or more years unemployed. $dur1$ and $dur2$ are as above.

Again, we are omitting a constant and a quartic in age in the set of regressors that is included in this regression.

5.2 Model Fit

5.2.1 Employment Shares

![Figure 4: Unemployment Share](image)

Figure 4: Unemployment Share

![Figure 5: Permanent Share](image)

Figure 5: Permanent Share
5.2.2 Employment Transitions

Figure 6: Temporary Share

Figure 7: P→T, P→U

Figure 8: T→P, T→U
5.2.3 Earnings Distribution

Figure 9: U->P, U->T

Figure 10: Quantiles of Log earnings
6 A Progressive System of Student Loans

We next use the panel of simulated lifetime income profiles $\{Y_{i,a}\}_{a=22}^{60} \sim \mathcal{N}$ using the estimated parameters and the model to study the implications of introducing a menu of public income-contingent loans. The experiment consists of increasing fees and at the same time applying a progressive loan system subsidized by the government with these particular characteristics: fees can be deferred until starting to work, repayments will depend on ex-post labor income and minimum exemption, and there will be a debt write-off and low interest rates.

The parameters of the loan system are therefore:

- $d$: Principal: total tuition fees over all years + maintenance (maybe)
- $p$: Repayment rate: fraction of gross earnings that is used for repayment
- $x$: Exemption level above which workers start repaying debt
- $m$: Write-off year after which the debt is cancelled
- $r$: Interest rate of debt
- $r^\beta$: Discounting interest rate
Repayment is income contingent and the first $x$ euros are exempt for everybody. That means that those who earn less than $x$ do not repay that year, and the rest pay a fraction of their income above $x$. We define non-exempt earnings as:

$$Y_{i,a}^{NE} = \max \{ Y_{i,a} - x, 0 \}$$

Annual payments at each age before repayment $\bar{a}$ as given by

$$P_{i,a} = pY_{i,a}^{NE} \text{ if } a < \bar{a}_i$$
$$P_{i,\bar{a}} = \min \{ D_{i,\bar{a}}, pY_{i,\bar{a}}^{NE} \}$$

Outstanding debt (plus interests) to be paid at the beginning of each period is calculated as follows:

<table>
<thead>
<tr>
<th>Initial debt (Beginning of period)</th>
<th>Outstanding debt (End of period)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>$D_{i,22} = d - P_{i,22}$</td>
</tr>
<tr>
<td>$(1 + r)D_{i,22}$</td>
<td>$D_{i,23} = (1 + r)D_{i,22} - P_{i,23}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td></td>
</tr>
<tr>
<td>$(1 + r)D_{i,\tau-1}$</td>
<td>$D_{i,\tau} = (1 + r)D_{i,\tau-1} - P_{i,\tau}$</td>
</tr>
</tbody>
</table>

And, given debt and accumulated payments, we calculate three amounts that will be informative of the burden of the loan system under different specifications. We bin individuals based on their position in the total lifetime income distribution. We calculate 100 percentiles.

1. Repayment year

$$\bar{a}_i = \min \left\{ m, \bar{a} \text{ s.t. } \sum_{a=1}^{\bar{a}} P_{i,a} \geq D_{i,\bar{a}} \right\}$$

2. Net present value of all repayments
\[ NPV_i = \sum_{a=1}^{n} \left( \frac{1}{1 + r^a} \right) P_{i,a} \]

3. Subsidy derived from debt write off

\[ S_i = 1 - \frac{NPV_i}{d} \]

6.1 Experiments

In this section, we use our laboratory to understand the impact of changing the parameters. In particular, we are interested in the degree of progressivity and in the average subsidy, as the latter will give us a measure of the burden on the government as compared to the current system, in which the government pays around 80% of the total cost of tertiary education and the student pays the remaining 20%. This will also give us a measure of how feasible our proposed financing structure is as compared to the status quo.

Our baseline scenario follows the 2007 UK reform. In particular, we set

\[ d = 21000 \text{ euros} \]
\[ r = 0\% \]
\[ p = 10\% \text{ annual earnings} \]
\[ x = 15000 \text{ euros} \]
\[ m = 25 \text{ years} \]
6.1.1 Case 1: Changing the level of debt
6.1.2 Case 2: Changing the exemption level
6.1.3 Case 3: Different levels of write-off year
6.1.4 Case 4: Changing the repayment rate
6.1.5 Case 5: Changing the interest rate
7 Conclusion

TBA