EIEF WORKING PAPER SERIES

IEF

EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 05/11

March 2011

On the Distribution of College Dropouts: Wealth and Uninsurable Idiosyncratic Risk

by

Ali K. Ozdagli

(Federal Reserve Bank of Boston)

Nicholas Trachter

(EIEF)

On the Distribution of College Dropouts: Wealth and Uninsurable Idiosyncratic Risk*

Ali K. Ozdagli

Federal Reserve Bank-Boston

Ali.Ozdagli@bos.frb.org

Nicholas Trachter

EIEF

nicholas.trachter@eief.it

March 6, 2011

Abstract

We present a dynamic model of college education where the students face uncertainty about their income stream after graduation due to unobserved heterogeneity in their innate scholastic ability. As students write exams, they reevaluate their expectations and may find it optimal to drop out and join the workforce without reaping the whole benefit of college education. The model shows that, in accordance with the data, poorer students are less likely to graduate and are more likely to drop out earlier than wealthier students. Our model generates these results without introducing credit constraints. Conditioning on measures of innate ability, we find in the data that poor students are at least 31% more likely to drop and they do so around a year before rich students.

*We thank Andrea Pozzi. We also benefited from comments by seminar participants at the Federal Reserve Bank of Boston. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Boston or the Federal Reserve System. Sarojini Rao provided excellent research assistance. All mistakes are our own.

1 Introduction

Around 40% of every cohort that enrolls in 4-year U.S. colleges drops out and there is a higher concentration of dropouts among the students from lower income families. We also observe that students from low income families tend to drop out earlier than students from high income families. Given the high return to graduation, the skewed distribution of dropouts generates a channel that perpetuates and exacerbates income inequality.

A widespread explanation of low college enrollment rate among high school graduates from poor families is credit constraints. However, credit constraints of poor college students are not likely to explain their relatively high dropout rates in college: The selection in initial enrollment due to credit constraints would imply that poor high school graduates, that choose to enroll in college, should have higher ability than rich students. Therefore, poor college students should be more likely to end up in high paying jobs which should give them more incentive to graduate. Nevertheless, borrowing constraints may explain the skewed distribution of dropouts if poor students are more likely to receive negative wealth shocks. In this paper, we propose a story complementary to borrowing constraint.

We present a dynamic model of educational choice to explore the relationship between household wealth and dropout behavior. In the model, the differences in the unobserved innate scholastic ability and initial wealth are the driving force behind the high and skewed dropout rate. At every period, risk-averse students write an exam of which outcome provides both human capital and information that can be used to update beliefs about students' ability level. Given the outcome of the exams and their income level, the students decide optimally if and when to drop out. Therefore, the optimal dropout behavior of a student is characterized by the distance between her belief about her abilities and a belief threshold at which the student drops out. We show that this threshold depends endogenously on the wealth level of the student, therefore providing the link between wealth and dropout behavior.

In order to model the evolution of beliefs and the decision to drop out, we take Miao and Wang (2007) framework of entrepreneurial learning and survival and extend it with realistic features that are important for the dropout decision faced by college students. In particular, we allow the workers' lifetime wage profile to depend on their experience, measured by the time spent on the job, and on their tenure in college as in Mincer (1974). Unlike Mincer, however, we also let lifetime wage profile depend on whether the worker has graduated from college and allow the experience premium to interact with the graduation status.

Using data from the *National Longitudinal Survey of Youth 1979* or NLSY, and the *National Longitudinal Study of the High School Class of 1972* or NLS-72, we test the model's predictions. We find that poor students are at least 31% more likely to drop and they do so around one year before rich students, conditional on the measures of unobserved ability.

Our model implies that college education is a risky investment because the outcome is uncertain and that richer students behave as if they are less risk averse due to constant relative risk aversion preferences.[1] This framework combined with the learning mechanism generates the result that poor students are less willing to take the risk associated with

[1] Chen (2001) finds that college investment is indeed risky after correcting for selection bias and accounting for permanent and transitory earnings risks.

college education and that they do not want to continue their education for as long as the rich students in order to learn their ability. Our result is robust to different specifications of the lifetime wage profile for students with different tenure at college.

In a typical Mincerian framework, a student would choose the number of years of education by comparing the marginal gain of an extra year in school with the marginal gain of joining the workforce immediately. Within this framework, a non-linear relationship between schooling and returns to education can potentially explain why many students decide to drop out even though returns to graduation are high. Still, this framework does not fully explain the data. First, the Mincerian model is silent about the relationship between wealth and educational profiles. Second, when college students are confronted with questions regarding their expectations about postsecondary educational outcomes, almost all of them respond that they intend to obtain four years of college education.[2] The failure of this basic Mincerian model suggests that it is necessary to have a story where information unfolds as time passes by, in order to explain the dropout behavior.

The literature provides three different stories where information unravels over time. The first one is binding credit constraints, which is theoretically not very compelling due to the selection in initial enrollment as explained above. The second one is learning about scholastic taste, as in Stange (2007) or Heckman and Urzua (2008). However, there is no clear channel that relates scholastic taste with the wealth level of the family.

The third one, used in this paper, is learning about unobserved ability through experimentation. Our modeling choice follows from the empirical work of Stinebrickner and Stine-

[2]See the data manual of *National Longitudinal Survey of Youth 1979*.

brickner (2008) and Stinebrickner and Stinebrickner (2009), who construct a panel study in order to understand the dropout decisions of students in a particular 4-year college, Berea College. Stinebrickner and Stinebrickner (2008) find that binding credit constraints can not explain dropouts. Stinebrickner and Stinebrickner (2009) find that academic performance, a proxy for learning about scholastic ability, is a good predictor of dropout behavior to the detriment of learning about scholastic taste.[3]

2 Evidence

To motivate our model, this section presents some statistics regarding dropout behavior based on the *National Longitudinal Survey of Youth 1979* or NLSY, and the *National Longitudinal Study of the High School Class of 1972* or NLS-72. For the NLSY we focus on individuals who joined a four-year college during or after 1979 with no discontinuities in their education spells upon starting college, and those who joined the workforce upon high school graduation and never attended a four-year college. For the NLS-72 we focus on individuals who joined a four-year college during 1972 with no discontinuities in their education spells upon starting college, and those who joined the workforce upon high school graduation and never attended a four-year college.[4]

[3]Another mechanism that can explain the skewed distribution of dropouts is that richer students choose longer duration of education when education is a normal consumption good. However, this cannot explain why academic performance matters so much for dropout decisions, as shown in Stinebrickner and Stinebrickner (2009), unless we believe that academic performance affects marginal utility of education and that its effect on rich and poor students is different.

[4]Note that we also discarded students that attended community colleges. The separation of community colleges from four-year colleges is important because the salary profile of graduates from both types of institutions is quite different. Within the context of our model, community colleges may serve as a stepping stone to four-year colleges by giving students more information about their skills so that high school graduates

Our analysis follows from comparing students' dropout profiles from rich and poor families. For NLS-72 we use as a measure of wealth the socioeconomics status of the respondent's family at the moment of high-school graduation. For the NLSY such a variable is not available so we construct it ourselves. We look at the respondent's family income at high-school graduation and we classify families by it. We classified the lowest 33% of families in terms of income as of low socioeconomic status and the top 33% as of high socioeconomic status.[5]

Table 1 presents some aggregate statistics regarding dropout behavior for both rich and poor students. In both data sets, students from poor families have higher attrition rates. Furthermore, students from poor households tend to drop out before than students with rich families. As shown in the table, they tend to drop around an year before rich students.

Table 1

Dropout rates and mean time before dropping out by socioeconomic status of family

	Socio. status[b]	% that drop	Mean tenure in college	st. dev. of tenure
NLSY	Low	62.03	2.58	1.34
	High	34.64	3.39	1.45
NLS-72	Low	47.26	2.33	1.46
	High	38.58	3.21	1.66

[a] We include individuals that dropped out at most after 6 years since initial enrollment.
[b] For the NLSY we constructed the measure of socioeconomic status thorugh the income level of the family prior to the respondent's enrollment in college. We choose the top 33% to be of High socio-economic status while the bottom 33% to be of low socioeconomic status.

To explore the skewed distribution of dropouts with respect to wealth a little bit further,

that are not optimistic enough to go to college may still enroll community colleges. Trachter (2010) formalizes this idea to study the transition between community and four-year colleges.

[5]We dropped the students between the 33th and 66th percentiles as the variable was very noisy because family income is not perfectly correlated with family's wealth, decreasing the confidence about the results. For the NLS-72, where students imputed their family's socioeconomic status, all of our calculations and results still hold when people with medium socioeconomic status are included. We excluded them to keep the two data sets comparable.

we compare the dropout rates of rich and poor students in different years of college. If poor students tend to drop out earlier than rich students, as we suggest, then a larger proportion of dropouts among poor students should occur in earlier years of college, whereas a larger proportion of dropouts among rich students should occur in later years of college. Table 2 provides a parsimonious way of checking this argument in the data by reporting the ratio of proportion of dropouts among poor students to the proportion of dropouts among rich students in different years of college. This statistic decreases from a number greater than one to a number less than one as we go from earlier to later years of college, providing support to our claim.

Table 2

Dropout rates of low vs. high income students

	tenure between					
	0 and 1 years	1 and 2 years	2 and 3 years	3 and 4 years	4 and 5 years	5 and 6 years
NLSY	3.86	0.83	0.83	0.64	0.33	0.24
NLS-72	1.78	1.46	0.69	0.54	0.63	0.17

Each number in the table represents the dropout rate of low income students as a share of the total dropout rate of low income students divided by the yearly dropout rate of high income students as a share of the total dropout rate of high income students.

We also extend out analysis by controlling for available proxies of ability that do not seem to be strongly co-linear with the socioeconomic status of the household.[6] If students form their beliefs rationally, these proxies are also positively correlated with students' initial belief about their ability because the belief distribution of higher ability students should first-order stochastically dominate the belief distribution of lower ability students.

[6]Education of the student's father and mother are strongly co-linear with the wealth of the household and therefore are not used.

Table 3 presents the marginal effect and percentage effect of having high socioeconomic status relative to low socioeconomic status for the NLSY and NLS-72. For both data sets, the marginal effect is significant and states that, conditioning on observable measures of ability, a student is between 13% and 16% more likely to drop if she comes from a poor household relative to a rich household, confirming the prediction of previous tables. In other words, given the fraction of students that drop out, the probability of dropping out increases by at least 31% for students from poor households relative to those from rich households.

Table 3

Marginal and % effect of high socioeconomic status on probability of dropout

	dF/dx	std. error	fraction of dropouts	% effect	N
NLSY	-0.163	0.064	0.4248	38.3	383
NLS-72	-0.132	0.056	0.4145	31.8	509

Controls: for the NLSY we used sex, race, AFQT scores, while for the NLS-72 we used sex, race, and rank in high-school senior class. N: number of observations.

In Table 4 we present an OLS regression of the time to drop (measured in years) on the socioeconomic status of the student's household and the same set of controls for those students that eventually dropped out.[7] We present the results for both the NLSY and NLS-72 data sets. Dropouts from rich households tend to drop between 0.84 and 1.11 years (depending on the data set) after dropouts from poor households, consistent with the results in Table 1.

[7] The most natural way of capturing the effect would be to run a probit regression on the dropout probability as in Table 3 by adding a variable that accounts for the years spent in college and an interaction variable between the years spent in college and the socioeconomic status of the student's household. The problem is that both dropouts and graduates are clustered. Those that drop do so often early and those that graduate do so later. Therefore, it is not possible to identify the effect of the interaction on the dropout probability in this way.

Table 4
Effect of high socioeconomic status on time to drop

	coefficient	std. error	N
NLSY	0.844	0.328	102
NLS-72	1.11	0.271	211

Controls: for the NLSY we used sex, race, AFQT scores, while for the NLS-72 we used sex, race, and rank in high-school senior class. N: number of observations

3 Model

At $t = 0$ students are enrolled in college, endowed with wealth level x_0. Students differ in their ability to acquire human capital at college. Their ability can be low or high. Let $\mu \in \{0, 1\}$ denote the ability level, where $\mu = 0$ denotes low ability. The ability level is not observable at $t = 0$. Instead, individuals inherit a signal about their true type $p(0) = \Pr(\mu = 1)$.

At any point in time an agent can either be enrolled as a full-time student or working in low- and high-skilled sectors.[8] The high-skilled sector only hires high ability workers with college degrees.[9] Work is assumed to be an absorbing state with constant wage function $\tilde{w}(\mu, \tau)$ where $\tau \equiv T - t$ accounts for the amount of time left prior to graduation and T is

[8]Post-secondary education is a combination of both 2-year and 4-year colleges with dynamic patterns that involve dropouts and transfers across types of schools. Using data from the National Longitudinal Survey of 1972, Trachter (2010) shows that students enrolled at 4-year colleges either drop or remain at the current type of institution until graduation.

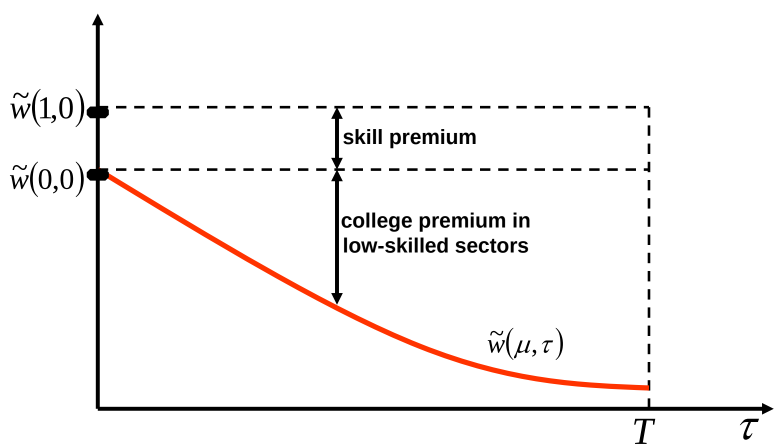
[9]The interest of this paper is to understand dropout behavior. Dropouts usually leave school and join low-skilled sectors.

the duration of college. We specify the function $\tilde{w}(\mu, \tau)$ as follows,

$$\tilde{w}(\mu, \tau) = \begin{cases} w(\tau) & \text{if } \tau > 0; \\ w(0) & \text{if } \tau = 0 \text{ and } \mu = 0; \\ w_1 & \text{if } \tau = 0 \text{ and } \mu = 1. \end{cases}$$

with $w(\tau_0) > w(\tau_1)$ if $\tau_0 < \tau_1$ and $w_1 > w(0)$. Therefore, the wage is increasing in the time spent in school and graduates of the high type enjoy higher wages. The function $\frac{w(\tau)}{w(T)}$ reflects the college premium in low-skilled sectors. A graphical representation of the wage function is depicted in Figure 1.[10]

Figure 1
Skill and College Premium



[10]The results of the paper still holds when we also allow the wage profile depend on workers' experience, given by the time spent on the job, and let the experience premium interact with the graduation status. See the end of the following subsection for details.

The evolution of the wealth level x is given by

$$\frac{dx}{dt} = \begin{cases} rx + \tilde{w}(\mu, \tau) - c & \text{if working;} \\ rx - a - c & \text{if enrolled in college.} \end{cases}$$

where a denotes per period cost of college education.

At every period of length dt in college, students are faced with an exam consisting of two questions. The first question is a multiple choice question and the second one is an essay question. High-ability students (i.e. $\mu = 1$) know the answer to the multiple choice question but need to study for the essay question. Let λ_1 denote the probability per unit of time that a student of type 1 answers the essay question correctly. Low-ability students (i.e. $\mu = 0$) think they know the answer to the multiple choice question but sometimes make a mistake when answering. Let λ_0 denote the probability per unit of time of low-ability individuals answering the multiple choice question incorrectly. Furthermore, low-ability students are not able to answer the essay question correctly.

When a student answers both questions wrong she receives a fail which reveals that she has low ability. When she gives correct answers to both questions she receives an excellent which reveals that she has high ability. When the students can answer only the multiple choice question correctly, she receives a pass and updates her belief according to Bayes' rule. Table 5 presents the probability of the outcome of a given exam conditional on the true type of the student.

We choose to model the learning process using discrete signals rather than continuous signals primarily because in our model the speed of learning depends on the type of the student. In a continuous Brownian Motion signal setting, this would be equivalent to having different volatilities for the signal process. However, Merton (1980) and Nelson and Foster (1994) point out that an observer of a continuous path generated by a diffusion process can estimate a constant or a smoothly time varying volatility term with arbitrary precision over an arbitrarily short period of calendar time provided she has access to arbitrarily high frequency data. Of course, introducing a discrete signal is not the only way to get around this problem but it is analytically more efficient than other possibilities such as introducing stochastic volatility or writing a discrete time model. These alternatives require belief updating about the mean and variance of the corresponding stochastic process which complicates the analysis without adding anything to the intuition.

Table 5

Probabilities of different grades in exams

	Fail	Pass	Excellent
$\mu = 0$	$\lambda_0 dt$	$1 - \lambda_0 dt$	0
$\mu = 1$	0	$1 - \lambda_1 dt$	$\lambda_1 dt$

Each value in the table is the probability of a given grade in the exam per unit of time dt conditional of the student's true ability level.

A student initially enrolled in college chooses her consumption stream $\{c(t) : t \geq 0\}$ and whether to continue as a student or drop out and join the workforce, in order to maximize her time-separable expected discounted lifetime utility derived from consumption,

$$\mathbb{E}\left\{ \int_0^\infty e^{-\rho t} \frac{c(t)^{1-\gamma}}{1-\gamma} \middle| p(0), x(0) \right\}$$

where γ is the coefficient of Constant Relative Risk Aversion (CRRA).

We let $J(x, p, \tau)$ denote the value for a student with current wealth level x, prior p and $T - \tau$ time already spent in school. Also, $V(x, \mu, \tau)$ denotes the value for a worker with current wealth level x of true type μ that spent $T - \tau$ time in school.

3.1 The Problem of a Worker

An individual with $T - \tau$ years of schooling that joins the workforce maximizes her lifetime discounted utility $\int_0^\infty e^{-\rho t} \frac{c^{1-\gamma}}{1-\gamma} dt$ subject to the law of motion for wealth $\frac{dx}{dt} = rx + \tilde{w}(\mu, \tau) - c$. It will prove useful to characterize $W(\mu, \tau)$, the present discounted value of lifetime earnings. This object is simply

$$W(\mu, \tau) = \int_0^\infty e^{-rt} \tilde{w}(\mu, \tau) dt$$

The Hamilton-Jacobi-Bellman representation of the worker's problem is

$$\rho V(x, \mu, \tau) = \max_c \frac{c^{1-\gamma}}{1-\gamma} + V_x(x, \mu, \tau)(rx + \tilde{w}(\mu, \tau) - c)$$

which states that the flow value of being a worker has to be equal to the instant utility derived from consumption in addition to the change in value that happens through the change in wealth accrued from time passing by.

The first order condition of the worker's problem reads $c^{-\gamma} = V_x(x, \mu, \tau)$. Plugging it back and operating provides an equation that the maximized value function of a worker needs to satisfy,

$$\rho V_x(x,\mu,\tau) = \frac{\gamma}{1-\gamma}\left[V_x(x,\mu,\tau)\right]^{1-\frac{1}{\gamma}} + r(x+W(\mu,\tau))V_x(x,\mu,\tau) \tag{1}$$

with solution given by

$$V(x,\mu,\tau) = A\left(r\left[x+W(\mu,\tau)\right]\right)^{1-\gamma} \tag{2}$$

where

$$A \equiv \left[(1-\gamma)r\right]^{\gamma-1}\left[(\rho-(1-\gamma)r)\frac{1-\gamma}{\gamma}\right]^{-\gamma} \tag{3}$$

The term $x+W(\mu,\tau)$ in equation (2) is the wealth of the worker accounting for the discounted lifetime labor earnings.

Although we have assumed that the wages upon graduation are constant this solution also holds if the wages depend on-the-job experience once we redefine $\tilde{w}(\mu,\tau,t)$ as the worker's wage depending on how long he has been in the job market, t, and replace $W(\mu,\tau)$ with $W(\mu,\tau) = \int_0^\infty e^{-rt}\tilde{w}(\mu,\tau,t)dt$ in the solution. Therefore, our results are robust to different specifications of the lifetime wage profile for students with different duration of college education.

3.2 The problem of a student of type μ and the natural borrowing limit

The Hamilton-Jacobi-Bellman equation for the value function of a student of known type $\mu = 1$ is

$$\rho J(x, 1, \tau) = \max_c \frac{c^{1-\gamma}}{1-\gamma} + (rx - a - c) J_x(x, 1, \tau) + J_\tau(x, 1, \tau) \frac{d\tau}{dt} \tag{4}$$

subject to the terminal condition $J(x, 1, 0) = V(x, 1, 0)$ that states that the value upon graduation equates the value of being a worker in the high-skill sector. Also, we are using an implicit condition that guarantees that high-skilled students would never find it profitable to drop out. This condition will be derived later on this section.

This Hamilton-Jacobi-Bellman equation states that the desired return on being a student of high type ($\mu = 1$) with wealth level x and with distance τ to graduation equals the instant utility derived from consumption plus the change in value due to the change in wealth and distance to graduation, both due to the change accrued in time. Note that $\frac{d\tau}{dt} = -1$.

The first order condition of this problem states that $c^{-\gamma} = J_x(x, 1, \tau)$. Solving for c and plugging back into equation (4) provides the maximized value function. We guess that the solution to this object is $J(x, 1, \tau) = A\left[rx + B(\tau)\right]^{1-\gamma}$, where A was is defined as in equation (3) and $B(\tau)$ needs to be solved for. Intuitively, $B(\tau)$ accounts for the decrease in value due to distance τ to graduation and terminal payoff. Plugging the guess into the maximized value function provides

$$rB(\tau) + B'(\tau) + ra = 0$$

with boundary condition $B(0) = rW(1, 0)$, which follows from the terminal condition of the problem presented in equation (4). This is an Ordinary Differential Equation in one variable with a terminal condition. The solution is $B(\tau) = (rW(1, 0) + a)\, e^{-r\tau} - a$ and therefore the value function of a student of type $\mu = 1$ is

$$J(x, 1, \tau) = A \left(rx - a + e^{-r\tau}\left(rW(1, 0) + a\right)\right)^{1-\gamma} \tag{5}$$

or,

$$J(x, 1, \tau) = A \left[r \left(x + e^{-r\tau} W(1, 0) - \frac{1 - e^{-r\tau}}{r} a\right)\right]^{1-\gamma}$$

where the term in paranthesis gives the net wealth of the worker accounting for the discounted value of future wages and the remaining tuition costs.

The next Lemma characterizes the condition such that students of the high type would not find it profitable to drop from school.

Lemma 1 *A student of type μ with current wealth level x and $T - \tau$ time spent in school will choose to remain as a student until $\tau = 0$ if $re^{-r\tau} W(1, 0) - a\left(1 - e^{-r\tau}\right) \geq rW(1, \tau)$.*

Lemma 1 follows from noting that for a student to remain in school it has to be the case that, for every value of τ, $J(x, 1, \tau) \geq V(x, 1, \tau)$. This condition simply tells that the graduation premium for a skilled student is high enough so that a student who knows she is skilled will remain in college until graduation. We assume throughout the paper that this condition holds.[11]

Another important assumption of the model is that students of type $\mu = 0$ and with $T - \tau$ time spent in school will always find it profitable to become dropouts. If this were not the case, for some values of τ there would be no dropouts by construction. The next Lemma guarantees that students with $\mu = 0$ will decide to drop and join the workforce, i.e. $J(x, 0, \tau) = V(x, 0, \tau)$.

Lemma 2 *A student that knows that she is of type $\mu = 0$ will drop out immediately if $a + rW(0, \tau) + W_\tau(0, \tau) > 0$.*

Proof. See Appendix A. ■

Intuitively, the marginal cost of schooling is adt and the marginal increase in the present value of earnings after an additional period of schooling is $e^{-rdt}W(0, \tau - dt) - W(0, \tau) = -[rW(0, \tau) + W_\tau(0, \tau)]\, dt + O(dt)^2$ where we used a Taylor series expansion. Subtracting increase in marginal earnings from marginal cost, dividing by dt and taking the limit as $dt \to 0$ gives the condition above.

Although our model does not entail any explicit borrowing constraint this last assumption leads to an implicit borrowing constraint for a student who does not know her type. Every student with $p < 1$ faces ex-ante a positive probability of receiving a shock that reveals that she has low ability which forces her to drop out and join the low-skilled workforce. Since the

[11] This condition at $\tau = T$ also guarantees that there are at least some students with optimistic prior beliefs willing to enroll in college.

marginal value of wealth for a dropout goes to infinity as x goes to $-W(0,\tau)$ the student will never borrow more than her discounted value of lifetime earnings $W(0,\tau)$. As long as the wage profile is common knowledge, as it is the case in our model, this "natural" borrowing constraint should also be the actual one because the student can always repay the borrowed money using his earnings when $x > -W(0,\tau)$.

3.3 The Problem of a Student of Unknown Type

The problem of a student of unknown type is more difficult because of three reasons. First, the wage upon graduation depends on the agent's true type. Second, arrival of information through exams generates beliefs' updating. Third, some students will become dropouts.

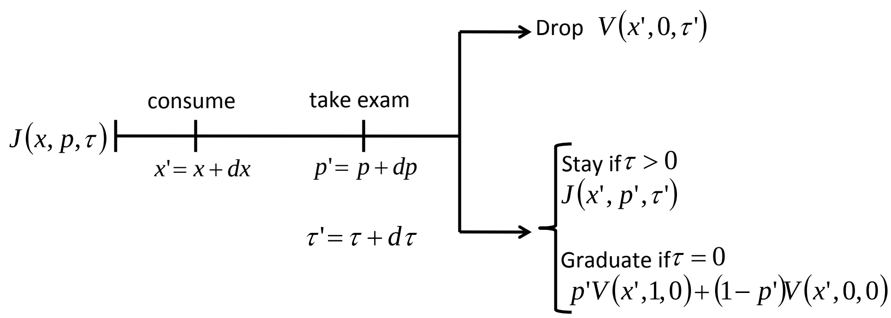
Before constructing the Hamilton-Jacobi-Bellman equation for this case first it is useful to consider how the information that arrives can be used to update beliefs. Consider a student with the belief $p(t)$ where $p(t)$ is the probability of being type 1 conditional on the information available at time t. Table 5 can be used to construct the posterior conditional on the grade during $(t, t+dt)$. When the grade is a fail it is clear that the student is of the low type and thus $p(t+dt) = 0$. When the grade is an excellent $p(t+dt) = 1$. Conditional on not receiving a fail or an excellent through period $(t, t+dt)$, a pass in the current exam implies that Bayes' rule can be used to update beliefs,

$$p(t+dt) = \frac{p(t)\left[1-\lambda_1 dt\right]}{p(t)\left[1-\lambda_1 dt\right] + \left[1-p(t)\right]\left[1-\lambda_0 dt\right]}$$

Substracting $p(t)$, dividing by dt, and taking the limit as $dt \to 0$ provides the Bayes' rule in its continuous time formulation,

$$\frac{dp}{dt} = -\left(\lambda_1 - \lambda_0\right) p(1-p) \tag{6}$$

Figure 2 Timeline



A student starts the period with wealth level x, prior p and remaining time in college τ. At the beginning of the period a student chooses her consumption level and thus provides the new value for wealth x'. Before the end of the period the student is faced with an exam used to produce the posterior p' and reduces the time left to graduation to τ'. At the beginning of next period the student chooses between dropping out or remaining as a student (or graduation is $\tau' = 0$).

Figure 2 describes the timeline of the problem of a student in a given period. The student enters the period with wealth level x, prior p and remaining time until graduation τ. Her value function is therefore $J(x, p, \tau)$. At the beginning of the period the student chooses her consumption level and thus produces the wealth level x' for next period. Before the end of the period she takes an exam and with the grade at hand updates her beliefs to p' using Bayesian updating. By the end of the period she accumulates time in school and therefore the distance to graduation is reduced to τ'. At the beginning of next period the student compares the value of remaining in school, that is either $J(x', p', \tau')$ if $\tau' > 0$ or

$p'V(x',1,0) + (1-p')V(x',0,0)$ if $\tau' = 0$ (i.e. graduation), with the value of joining the workforce $V(x',0,\tau)$ to decide between staying in college or becoming a dropout.

The Hamilton-Jacobi-Bellman equation for a student with current wealth level x, prior p and τ periods left in school is

$$\begin{aligned}
\rho J(x,p,\tau) &= \max_c \tfrac{c^{1-\gamma}}{1-\gamma} + (rx - a - c)J_x(x,p,\tau) - J_\tau(x,p,\tau) \\
&\quad -(\lambda_1 - \lambda_0)\, p(1-p)J_p(x,p,\tau) + \lambda_1 p\left[J(x,1,\tau) - J(x,p,\tau)\right] \\
&\quad +\lambda_0(1-p)\left[V(x,0,\tau) - J(x,p,\tau)\right]
\end{aligned} \tag{7}$$

This last equation states that the desired return on being a student with current wealth level x, prior p and accumulated time at school $T-\tau$ equals the instant utility derived from consumption plus (i) the change in value through the change in wealth, (ii) the change in value through the change in τ, and (iii) the change in value as different signals arrive. When a pass arrives the belief adjustment is continuous through the Bayesian updating of p; when an excellent arrives, expected with unconditional probability $\lambda_1 p$, the change in value is through a switch from having $p(t) = p$ to $p(t+dt) = 1$; and when a fail arrives, expected with unconditional probability $\lambda_0(1-p)$, the change in value is through a switch from having $p(t) = p$ to $p(t+dt) = 0$. Also note that the problem faced by a student of known type $\mu = 1$ presented in equation (4) is a particular case of the problem presented here, that follows by setting $p = 1$ in equation (7) and noting that $\frac{d\tau}{dt} = -1$.

The first boundary condition is degenerate because it implies J(x,pstar,0) to be different from V(x,0,0) which makes it inconsistent with the second boundary condition. This does not cause a problem for the solution.

A student faces the problem presented in equation (7) subject to a set of boundary conditions,

$$
\begin{aligned}
J(x,p,0) &= pV(x,1,0)+(1-p)V(x,0,0) \\
J(x,p^*(x,\tau),\tau) &= V(x,0,\tau) \\
J_p(x,p^*(x,\tau),\tau) &= 0 \\
J_x(x,p^*(x,\tau),\tau) &= V_x(x,0,\tau) \\
J_\tau(x,p^*(x,\tau),\tau) &= V_\tau(x,0,\tau)
\end{aligned} \tag{8}
$$

The first equation gives the Terminal Condition (TC) and states that the value of being a student with no time left prior to graduation has to equal the expected value of being a worker. Note that with probability p a student expects to be of type $\mu = 1$ and therefore would earn lifetime discounted labor income $W(1,0)$ while with probability $1-p$ expects to be of type $\mu = 0$ and therefore earn lifetime discounted labor income $W(0,0)$. To understand the second to fifth equations $p^*(x,\tau)$ needs to be defined. Let $p^*(x,\tau)$ be the belief threshold such that students with $p \leq p^*(x,\tau)$ drop from school and join the workforce. The second equation states that a student with $p = p^*(x,\tau)$, wealth level x, and τ periods away from graduation, has to be indifferent between staying in school and dropping out and enjoying lifetime discounted labor income $W(0,\tau)$. This equation is also known as the Value Matching Condition (VMC). The third, fourth, and fifth equations are known as Smooth Pasting Conditions (SPC) required for optimality of $p^*(x,\tau)$.[12]

The first order condition of the problem presented in equation (7) is $c^{-\gamma} = J_x(x,p,\tau)$.

[12]For a treatment of Value Matching and Smooth Pasting Conditions see Dumas (1991) and Dixit (1992).

Plugging it back into equation (7) together with the terminal, value matching, and smooth pasting conditions provides the equation that the threshold $p^*(x,\tau)$ needs to satisfy,

$$\begin{aligned} \rho V(x,0,\tau) &= \tfrac{\gamma}{1-\gamma} V_x(x,0,\tau)^{1-\frac{1}{\gamma}} + (rx-a)V_x(x,0,\tau) \\ &\quad +\lambda_1 p^*(x,\tau)\left[J(x,1,\tau) - V(x,0,\tau)\right] - V_\tau(x,0,\tau) \end{aligned} \tag{9}$$

which we can rewrite as

$$\lambda_1 p^*\left(x,\tau\right)\left[J\left(x,1,\tau\right) - V\left(x,0,\tau\right)\right] = \left[a + rW\left(0,\tau\right) + W_\tau\left(0,\tau\right)\right] V_x\left(x,0,\tau\right).$$

This last equation provides provides intuition about the belief threshold $p^*\left(x,\tau\right)$. The left-side of this equation is the expected utility gain from delaying the dropout decision by dt, whereas the right side the marginal net utility loss due to delaying the dropout decision. The student chooses optimally when to drop out, by equalizing the marginal gain and loss from delaying the dropout decision.

Solving for $p^*(x,\tau)$ allows for a close-form representation of dropout threshold,

$$p^*(x,\tau) = \frac{a + rW(0,\tau) + W_\tau(0,\tau)}{\lambda_1} \frac{V_x(x,0,\tau)}{J(x,1,\tau) - V(x,0,\tau)} > 0 \tag{10}$$

provided that Lemma 2 holds. The threshold is decreasing in the wealth level. In other words,

$$\frac{\partial p^*(x,\tau)}{\partial x} < 0 \tag{11}$$

where the details of the calculations can be found in Appendix B.

Our main result is that, conditional on their beliefs, students from richer families drop out later and are less likely to drop out than poor students. The result that the threshold p^* is decreasing in the wealth level is not enough to argue this result because the consumption profiles during tenure in school can overcome the initial difference in wealth.[13] The next proposition and corollary deal with this.

Let $\widetilde{\tau}$ denote the time left to graduation at the moment the individual joins the workforce. For example, $\widetilde{\tau} = T$ if the individual joins the workforce directly after high-school graduation and $\widetilde{\tau} = 0$ if the individual joins the workforce with a college degree.

Proposition 1 *Let $x^i(0)$ and $x^j(0)$ denote the initial wealth levels at time 0 of students i and j. If $x^i(0) > x^j(0)$ then $E\{\widetilde{\tau}|\, x^i(0), p(0), \mu\} \leq E\{\widetilde{\tau}|\, x^j(0), p(0), \mu\}$. In other words, given a skill level μ and initial belief $p(0)$, richer students tend to drop out later and have longer expected tenures in college.*

Proof. See Appendix C. ■

Proposition 1 states that the expected length of college tenure $T - \tau$ is increasing in the initial wealth level of the student once conditioned for the initial prior $p(0)$ and the skill level of the student. This means that if we look at dropouts, students from richer families tend to drop later than those from more poorer families. Conditioning on the skill level of the student μ is important as the arrival of news is correlated with it.

[13] Miao and Wang (2007) entrepreneurial survival model is a special case of our model where they also show that the boundary p^* is decreasing in the wealth level. However, they immediately conclude that richer entrepreneurs survive longer without providing an explicit proof.

The next corollary extends the result to show that, conditioning in the initial prior $p(0)$ and skill level μ, students from richer families are less likely to drop out and, therefore, more likely to graduate.

Corollary 1 *Let $x^i(0)$ and $x^j(0)$ denote the wealth levels at time 0 of students i and j. If $x^i(0) > x^j(0)$ then $\Pr[\widetilde{\tau} = 0 | \, x^i(0), p(0), \mu] \geq \Pr[\widetilde{\tau} = 0 | \, x^j(0), p(0), \mu]$. That is, once conditioned on the initial prior $p(0)$ and skill level μ, richer students are more likely to graduate.*

Corollary 1 follows by noting that the proof of Proposition 1 also implies that, for any $\mathcal{T} \in [0, T]$, $\Pr[\widetilde{\tau} \leq \mathcal{T} | \, x^i(0), p(0), \mu] \geq \Pr[\widetilde{\tau} \leq \mathcal{T} | \, x^j(0), p(0), \mu]$ from where the result follows.

Both Proposition 1 and Corollary 1 are driven by the fact that ex-ante uncertainty regarding the outcome of college education cannot be diversified away which makes the college education more risky than financial investments in the model.[14] Given ability and beliefs, a student with higher wealth chooses to increase the number of dollars invested in the risky asset, i.e. stay longer in the college, if absolute risk aversion is decreasing as it is the case with the CRRA preferences.

[14] Of course, financial assets in real life are risky due to macroeconomic fluctuations not modeled in this paper. However, these macroeconomic fluctuations also affect the payoff of college education through its effect on wages, and college education is still riskier than financial assets due to undiversifiable idiosyncratic risk.

4 Conclusion

In this paper, we provide evidence regarding the skewed distribution of dropouts with respect to the student's family wealth. Poor students are more likely to drop and they tend to do so before rich students. We explore if a model, that incorporates education as a risky investment and Bayesian learning about own's unobserved ability, can explain the skewness in dropout behavior.

Our main results rely on the fact that the outcome of college education is subject to uncertainty against which the students cannot insure themselves. When we combine this fact with CRRA preferences so that the absolute risk aversion decreases with wealth we arrive to the conclusion that poor students are less willing to take the risk associated with college education. This mechanism generates skewness in dropout behavior.

We provide a closed-form characterization in close-form of the optimal policy of a student as a function of (i) the expected future income due to graduation (through the prior, p), and (ii) the direct and indirect costs of remaining in school (through the distance to graduation, τ). We exploit the simplicity of the model to show that the model is able to fit the data qualitatively: (i) poorer students are more likely to drop than rich students, and (ii) if the drop, they do it earlier.

To motivate the theory we run a series of reduced-form regressions, conditioning by measures of the unobserved ability and prior of the student. The regressions' results are in line with the model predictions. We estimate that poor students are at least 31% more likely to drop than rich students and if they do so, they tend to do it around an year earlier.

Our goal is not to say that borrowing constraints are not part of the story behind the high and skewed dropout rates. Instead, we provide a complementary story that is able to explain the skewness. Furthermore, our story is consistent with Stinebrickner and Stinebrickner (2008) where borrowing constraints are found not to be the main determinant of becoming a dropout, and Stinebrickner and Stinebrickner (2009) which shows that bad grades are a good predictor of dropout behavior in detriment of learning about scholastic taste.

Since explicit borrowing constraints are absent in our model, it also suggests that policies that are geared towards reducing borrowing constraints, such as student loan programs, are not likely to eliminate the differences in dropout rates of rich and poor students. Moreover, a direct subsidy to college education for poor students would not only increase their graduation rates but also their tenure in college, by both reducing cost of spending additional time in college and increasing the expected gain from delaying the dropout decision. The optimal subsidy would depend on the wealth level and the distribution of ability for a given wealth level, which is the topic of future research.

References

Chen, Stacey, "Is Investing in College Education Risky?," Discussion Papers 01-09, University at Albany, SUNY, Department of Economics 2001.

Dixit, Avinash, "The Art of Smooth Pasting (Now published in Fundamentals of Pure and Applied Economics 55 (Harwood Academic Publishers GbmH, Poststr. 22, 7000 Chur, Switzerland), 1993. ISBN: 3-7186-5384-2, ISSN: 0191," STICERD - Theoretical Economics Paper Series 252, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE 1992.

Dumas, Bernard, "Super contact and related optimality conditions," *Journal of Economic Dynamics and Control*, October 1991, *15* (4), 675–685.

Heckman, James J and Sergio Urzua, "The Option Value of Educational Choices and the Rate of Return to Educational Choices," working paper, Cowles Foundation Structural Conference, Yale University 2008.

Merton, Robert C., "On estimating the expected return on the market : An exploratory investigation," *Journal of Financial Economics*, December 1980, *8* (4), 323–361.

Miao, Jianjun and Neng Wang, "Experimentation under Uninsurable Idiosyncratic Risk: An Application to Entrepreneurial Survival," Working Paper 2007.

Mincer, Jacob A., *Schooling, Experience, and Earnings* number minc74-1. In 'NBER Books.', National Bureau of Economic Research, Inc, 1974.

Nelson, Daniel B and Dean P Foster, "Asymptotic Filtering Theory for Univariate ARCH Models," *Econometrica*, January 1994, *62* (1), 1–41.

Stange, Kevin M, "An Empirical Examination of the Option Value of College Enrollment," working paper, University of California - Berkeley 2007.

Stinebrickner, Ralph and Todd Stinebrickner, "The Effect of Credit Constraints on the College Drop-Out Decision: A Direct Approach Using a New Panel Study," *American Economic Review*, December 2008, *98* (5), 2163–84.

Stinebrickner, Todd R. and Ralph Stinebrickner, "Learning about Academic Ability and the College Drop-out Decision," NBER Working Papers 14810, National Bureau of Economic Research, Inc March 2009.

Trachter, Nicholas, "Option Value and Transitions in a Model of Postsecondary Education," *working paper*, 2010.

Appendix

A Proof of Lemma 2

Let τ^* denote the threshold for τ so that students drop for $\tau < \tau^*$. A student of type $\mu = 0$ faces the problem given by

$$\rho J(x,0,\tau) = \max_c \frac{c^{1-\gamma}}{1-\gamma} + (rx - c - a)J_x(x,0,\tau) - J_\tau(x,0,\tau)$$

subject to the terminal condition $J(x,0,0) = V(x,0,0)$, boundary condition $J(x,0,\tau^*) = V(x,0,\tau^*)$, and smooth pasting conditions $J_x(x,0,\tau^*) = V_x(x,0,\tau^*)$ and $J_\tau(x,0,\tau^*) = V_\tau(x,0,\tau^*)$.

Plugging back the first order condition provides that

$$\rho V_x(x,0,\tau^*) = \frac{\gamma}{1-\gamma}\left[V_x(x,0,\tau^*)\right]^{1-\frac{1}{\gamma}} + (rx-a)V_x(x,0,\tau) - V_\tau(x,0,\tau^*)$$

Using equation (1) this equation can be reduced to $a + rW(0,\tau^*) + W_\tau(0,\tau^*) = 0$. Hence, if $a + rW(0,\tau^*) + W_\tau(0,\tau^*) > 0$ for all $\tau \le T$ the boundary condition for an interior dropout boundary is not satisfied. Moreover, the desired return from continuing education in terms of utility, the LHS of Bellman equation, is greater than the continuation value, the RHS of Bellman equation, at the default boundary and hence it is optimal to drop out immediately.

B Proof of $\frac{\partial p^*}{\partial x} < 0$

Differentiating the threshold p^* (see equation (10)) with respect to x provides

$$\begin{aligned}
\frac{\partial p^*}{\partial x} &= \frac{a+rW(0,\tau)+W_\tau(0,\tau)}{\lambda_1}\left[\frac{V_x(x,0,\tau)}{J(x,1,\tau)-V(x,0,\tau)}\frac{V_{xx}(x,0,\tau)}{V_x(x,0,\tau)} - \frac{V_x(x,0,\tau)}{J(x,1,\tau)-V(x,0,\tau)}\frac{J_x(x,1,\tau)-V_x(x,0,\tau)}{J(x,1,\tau)-V(x,0,\tau)}\right]\\
&= \frac{a+rW(0,\tau)+W_\tau(0,\tau)}{\lambda_1}\frac{V_x(x,0,\tau)}{J(x,1,\tau)-V(x,0,\tau)}\left[\frac{V_{xx}(x,0,\tau)}{V_x(x,0,\tau)} - \frac{J_x(x,1,\tau)-V_x(x,0,\tau)}{J(x,1,\tau)-V(x,0,\tau)}\right]\\
&= p^*(x,\tau)\left[\frac{V_{xx}(x,0,\tau)}{V_x(x,0,\tau)} - \frac{J_x(x,1,\tau)-V_x(x,0,\tau)}{J(x,1,\tau)-V(x,1,\tau)}\right]\\
&= -p^*(x,\tau)\left[\frac{\gamma r}{r[x+W(0,\tau)]} + (1-\gamma)r\frac{\left(r\left[x+e^{-r\tau}W(1,0)-\frac{a}{r}\left(1-e^{-r\tau}\right)\right]\right)^{-\gamma}-(r[x+W(0,\tau)])^{-\gamma}}{\left(r\left[x+e^{-r\tau}W(1,0)-\frac{a}{r}(1-e^{-r\tau})\right]\right)^{1-\gamma}-(r[x+W(0,\tau)])^{1-\gamma}}\right]\\
&= -p^*(x,\tau)\left[\frac{\gamma r}{r[x+W(0,\tau)]} + \frac{1}{x+W(0,\tau)}\frac{y^{-\gamma}-1}{\frac{y^{1-\gamma}-1}{1-\gamma}}\right]\\
&= -\frac{p^*(x,\tau)}{x+W(0,\tau)}\left[\gamma + \frac{y^{-\gamma}-1}{\frac{y^{1-\gamma}-1}{1-\gamma}}\right]
\end{aligned}$$

where $y \equiv \frac{x+e^{-r\tau}W(1,0)-\frac{a}{r}\left(1-e^{-r\tau}\right)}{x+W(0,\tau)}$, with $y \ge 1$ provided the condition on Lemma 1 holds. Next, it will be proved by contradiction that $\gamma + (1-\gamma)\frac{y^{-\gamma}-1}{y^{1-\gamma}-1} > 0$ when $y \ge 1$.

Consider first the case where $\gamma < 1$. Suppose that $\gamma + (1-\gamma)\frac{y^{-\gamma}-1}{y^{1-\gamma}-1} < 0$. Because $\gamma \in (0,1)$ and hence $y^{1-\gamma} - 1 > 0$ we can multiply both sides of this inequality by $y^{1-\gamma} - 1$ to get $\gamma(y^{1-\gamma} - 1) + (1-\gamma)(y^{-\gamma} - 1) < 0$. The left-hand-side of this in equality is strictly increasing in y and therefore attains its minimum at $y = 1$ with value equal to 0. Therefore, $\gamma(y^{1-\gamma} - 1) + (1-\gamma)(y^{-\gamma} - 1) < 0$ and hence $\gamma + (1-\gamma)\frac{y^{-\gamma}-1}{y^{1-\gamma}-1} < 0$ is not possible.

Now consider the case where $\gamma > 1$. Suppose that $\gamma + (1-\gamma)\frac{y^{-\gamma}-1}{y^{1-\gamma}-1} < 0$. Because $\gamma > 1$ and hence $y^{1-\gamma} - 1 < 0$ we can multiply both sides of this inequality by $y^{1-\gamma} - 1$ to get $\gamma(y^{1-\gamma} - 1) + (1-\gamma)(y^{1-\gamma} - 1) > 0$. The left-hand-side of this equation is strictly decreasing in y and therefore attains its maximum at $y = 1$ with value equal to 0. Therefore, $\gamma(y^{1-\gamma} - 1) + (1-\gamma)(y^{1-\gamma} - 1) > 0$ and hence $\gamma + (1-\gamma)\frac{y^{-\gamma}-1}{y^{1-\gamma}-1} < 0$ is not possible.

As $\gamma + (1-\gamma)\frac{y^{-\gamma}-1}{y^{1-\gamma}-1} > 0$ for every $\gamma > 0$, $\frac{\partial p^*}{\partial x} < 0$.

C Proof of Proposition 1

Two students with the same skill level μ are equally likely to receive a fail or an excellent grade at any point in time. Therefore, although the grade will effect the behavior of an individual student it will not directly effect the average dropout behavior by the Law of Large Numbers. Therefore, we look at the dropout behavior of students that do not receive any signals that reveal their true types.

Suppose we have two students i and j with the same initial belief, i.e. $p^i(0) = p^j(0)$, and with initial wealth levels $x^i(0) > x^j(0)$ so that the first student is initially richer. There are two possible outcomes conditional on not receiving a signal. First, if $p(0)$ is high enough both students wait until they graduate which does not violate our proposition as $\widetilde{\tau}^i = \widetilde{\tau}^j = 0$. Second, if $p(0)$ is not high enough at least one of the students drops out. Let $t_0 < T$ be the first point in time when one of the students drop out. Then we have $p^i(t) = p^j(t)$ for all $t \leq t_0$ because $p^i(0) = p^j(0)$ and the belief evolution is the same for both students conditional on not receiving a signal. Moreover, the richer student does not drop out earlier than the poorer student, so that $\widetilde{\tau}^i \leq \widetilde{\tau}^j$, if $x^i(t) \geq x^j(t)$ for all $t \leq t_0$ as $p^*(x^i, \tau) \leq p^*(x^j, \tau)$ if $x^i \geq x^j$. Therefore, we can prove our proposition by showing that $x^i(t) \geq x^j(t)$ for all $t \leq t_0$. Suppose $x^i(t) < x^j(t)$ for some $t < t_0$. Then, since $x^i(t)$ and $x^j(t)$ have continuous paths there exists a $\bar{t} \leq t_0$ where $x^i(\bar{t}) = x^j(\bar{t})$ by the Intermediate Value Theorem. Moreover, since $p^i(t) = p^j(t)$ for all $t \leq t_0$ we have $p^i(\bar{t}) = p^j(\bar{t})$. As a result, both students' consumption decisions are synchronized from time $\bar{t}$ on because they are forward looking. Hence, $x^i(t) = x^j(t)$ for $\bar{t} \leq t \leq t_0$ which is a contradiction.

This analysis also implies that $\Pr[\widetilde{\tau} \leq \mathcal{T} | \, x^i(0), p(0), \mu] \geq \Pr[\widetilde{\tau} \leq \mathcal{T} | \, x^j(0), p(0), \mu]$ which we use for the proof of Corollary 1.