1   A non-intro Intro

Honestly, the first and foremost purpose of this paper is to explore the possibility of "big data" in applied econometrics.

The datafication has been a boom in the past decade. The volume of the data "public available" dwarfs that of the survey data the profession has built for a couple decades. It offers an exciting yet challenging frontier for emprical researcher to study subjects where data are either bad or outright none-existent.There are already good papers [1] where huge amount of structure data are collected and analyzed.

Although I expect it to be a fertile field for research, There are issues to be addressed. For one thing, these data are used in isolation, which limits its effectiveness. The other issue that labor economist may find the representativeness of the sample troubling. At this stage, the dataficiation is a highly selective process and it is hard to theorize what is the effective sampling scheme.

2   Motivation

Has your email box ever received spam email promising a quick degree? According to Allen Ezell [2], the degree mill industry has sold over a million fake diplomas and generated billions of dollars in sales. Yet very few empirical research has been done to study the impact of a fake degree.

Although holders of counter-feit diplomas are still elusive creatures to researcher, it is now possible to get some data on the degree holders from degree mills.

In this project, I will use scraped resume data from **Indeed.com** to do some descriptive analysis on the impact of obtaining a fake degree.

3   Data Collecting Procedure

3.1   List of College

GetEducated.com provides a database of acrreditation reports for 135 distance learning institutions.

Some background information on the accreditation in the US.

The highest accreditation authority is the regional accreditation agency, which is recognized by all legit accreditation agency. The next tier is national ac-

credidation agency, Department of Education and Distance Education & Training Council.

Other than these legitimate accreditation agency, there are numerous shady accreditation agency. Some of them have accreditation authority from foreign countries such as Cayman Island or Libya. Others are suspicious organization with shiny names such as National Board of Education.

However, I cannot decide which accreditation agency is legit and which is scam. Luckily, some state government have solved that problem for me. A few states issued consumer warning against some online college which bans the degree of these education institutions from its public service system.

Accordingly, I sort the list into four categories:

1. Degree Mill (53)

2. Dubious Acrreditation (28)

3. National Acrreditation (24)

4. Regional Acrreditation (29)

In the current version, I will use the regional accredication institution as one of the control group since it is a direct subsidy to the online degree mill. The other indirect substitute is community college. A list of all national community colleges is provided by **University of Texas at Austin**.

From this list, I extract all the community college that can be reliably tracked through a online search, which shrink the list to 1012 institutions. Possible as it is to do a sweep search, for the current version I sample 104 institutions randomly from the total list.

3.2   Resume Search

Indeed.com offers unlimited search in its alleged 100 million resume database. Because the resume stored on site is very structured, it can be colleged by scraping.

For now, I am scraping their name, the start time and end time of all self-reported employment spells, the time, name and institution of all self-reported degrees.

For the Degree mill group, it return over 10k resumes but only about 3400 are valid hit. For the online control group, it returned over 20k resumes while over 11k are valid hits. For the community college, it returned over 18k resumes

while about 10ks are valid hits.

It should be also noted that the requirement for validity also includes the resume reports the time of the degree obtained.

To which extent the self-report data can be trusted is left for the reader to judge, for now.

4 Preliminary Result

The data set I obtained has both advantages and disadvantages.

The advantage is rich work history, which is scarcely available for most public surveys on labor market. The disadvantage is the lack of detailed demographics variable. I do not have age, gender, or any other usual suspects.

That said, I still could answer a few interesting questions. For all the following figures, RED is for fake degrees, BLUE is for legit online degrees, GREEN is for community college.

4.1 Does the Diploma Matter?

From figure 1, we know that over 80 % only obtained 1 degree from the category, although it is more likely to obtain a second degree for legit online program.
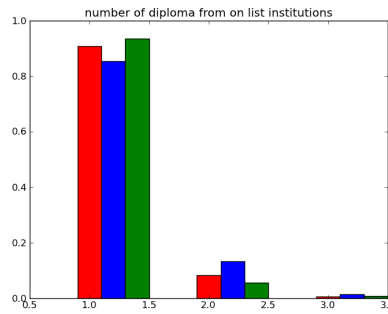

Figure 1: Number of Diploma obtained from category

From figure 2, we know that for 50 %, that is their only degree and for around 30%, the degree from the category is the 1 of the 2 degree listed on the resume.
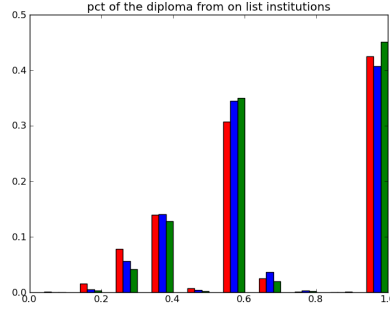
3

Figure 2: Percentage of Diploma obtained from category over all degrees

In figure 3, , rank 0 means the top degree on the education history list, rank 1 means the 2nd degree and so on. Assuming that people usually rank highest degree first, we can infer that for over 80% people, the degree from the category is their highest degree obtained.
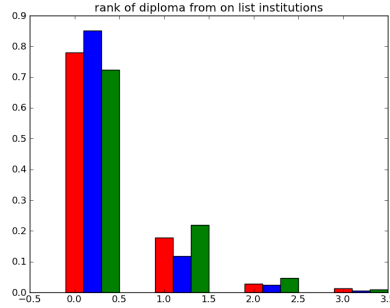

Figure 3: Rank of Diploma obtained from category over all degrees

Combining the analysis above, we can reasonably conclude that the diploma that will be studied is the backbone of the their education background, thus represents their education achievement in the labor market.

4.2   Before the degree

There are two interesting aspect of the profiles before they got the degree.

First of all, a majority of them got the degree while they are on a job. I define "on the job" means they report receiving the degree when they were on a job they reported on the resume.

4

Some notes on imputation. For those who only report the year of the degree, rather than the specific month, I impute month 12, For those who only report the year of the job, I impute the end month to be January. Such imputation gives the most conservative estimates for the on the job percentage. For those who do not report the date of either their degree or job, I code them as invalid(-1).

Although it looks like people who got legit online degree is more likely to be "on the job", I suspect part of the difference could be attribute to selective mis-reporting. Somehow folks from the other two group felt that they should not report the exact date of their degree. Although the incentive of doing so is quite imaginable for the holders of fake degree, it is not immediately clear why graduates from commnunity college also do that.
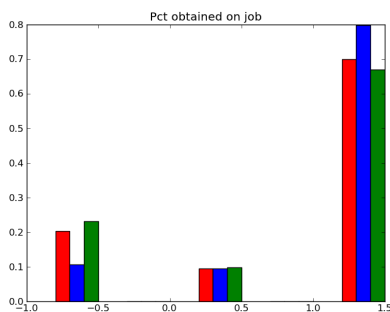

Figure 4: Whether or not they got the degree on the job. -1 = Invalid, 0 = No, 1 = Yes

The salient difference comes from the prior job experience. The graduates from a legit online education institution have more jobs than the graduates from community college or holders of fake diploma, as show in figure 5.

Here I need to clarify that for multiple degree holder, I define the "prior" to be prior to their first listed degree from the target institutions.
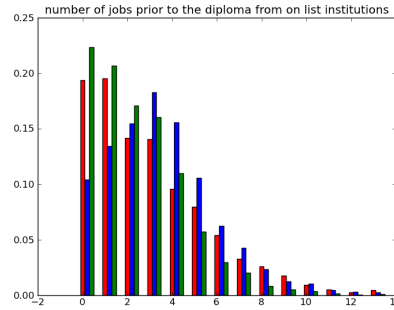
Figure 5: Number of Jobs prior to getting the diploma

However, more jobs does not mean longer working experience.It could be that each job has shorter history. However, if we actually code the prior job experience, we will arrive at the same conclusion. The graduates from legit online education institutions have lower concentration on the 0-3 years than the graduates from community college and fake degree holders.

There is another very interesting observation here. The fake degree holders have fat right tail, meaning a large percentage of them getting their degree when they got more than 20 years work experience already. This supports the rumor that some of them got the degree for promotion reasons.
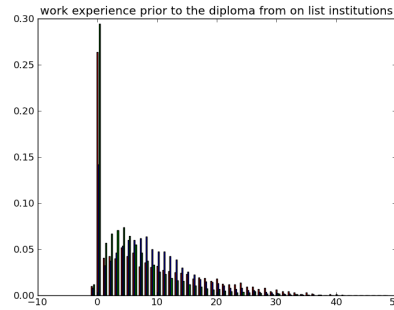

Figure 6: Years of work experience prior to getting the diploma

4.3   After the degree

So what happens after they got the degree?
First question I asked is that how many jobs they got after they obtained the degree. This is a tricky question since I have yet to take consideration of the time span since their graduation. From the naive analysis, we know that grau-

dates from legit online education institutions are more likely to hold on their jobs. However, a substantial proportion of three groups change their jobs at a higher than healthy rate.
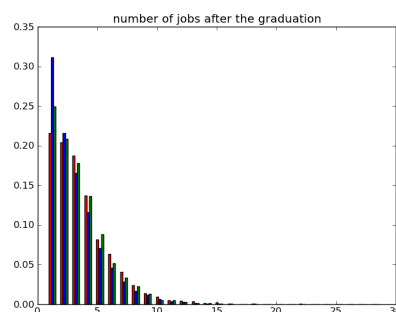

Figure 7: Number of Jobs post diploma

Another interesting question to ask since most of the people got their degrees on job, how soon do they leave that job title they start with? Notice here I only analyze on the job title, rather than the firm. They could get a promotion within the same firm. In principle, I will be able to answer that question. The other point I want to make is that I make the most conservative calculation and the cluster at month 12,24 and 36 are an artifact of combining year/month data with year data only.

The interesting finding here is that a significant portion of people change the title within a year from the time they got the diploma. Especially, there is a bump in the first three month. All three groups behave alike.
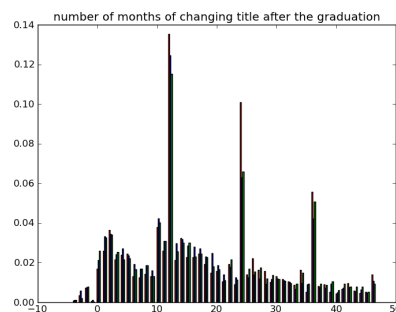

Figure 8: Month to Change Job Title post diploma

4.4 In 2008-2011 Financial Crisis

An interesting observation is to see what are their labor market performance between 2008 and 2011.

For the following analysis, I conditioned the sample on people who got their degree before 2008. Curretly, I define get laid off as having a job before 2008, which terminated after 2008 and before 2011. This will over-estimate their laid-off rate since I will count a job title changing promotion as a laid off, although there is good reason to believe that such error should be small during the crisis.

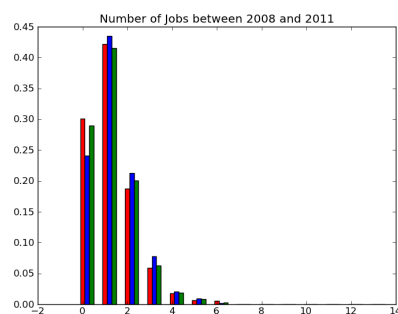|	Pct before 2008	Pct Laid off at least Once	Number of Jobs After laid off
Fake Degree	72.7%	63.8%	1.11
Legit Online College	34.6%	65.2%	1.25
Community College	47.8 %	58.5%	1.15

Figure 9: Number of Jobs in Crisis after laid off

References

[1] Alberto Cavallo, Brent Neiman, Roberto Rigobon, "Product Introductions, Currency Unions, and the Real Exchange Rate", *NBER Working Paper*, No. 18563, Nov. 2012

[2] Allen Ezell and John Bear, *Degree Mills: the Billion-dollar Industry That Has Sold over a Million Fake Diplomas*, Prometheus Books; Updated edition (January 24, 2012)