

Xu, Dafeng

Conference Paper

Acculturational Homophily in Friendships based on English-Name Usage: A Natural Experiment

55th Congress of the European Regional Science Association: "World Renaissance: Changing roles for people and places", 25-28 August 2015, Lisbon, Portugal

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Xu, Dafeng (2015) : Acculturational Homophily in Friendships based on English-Name Usage: A Natural Experiment, 55th Congress of the European Regional Science Association: "World Renaissance: Changing roles for people and places", 25-28 August 2015, Lisbon, Portugal, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/124717>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Acculturational Homophily in Friendships based on English-Name Usage: A Natural Experiment*

Dafeng Xu[†]

Abstract

This paper examines acculturational homophily in friendships of international students. Acculturation (also known as cultural assimilation) is measured by English-name usage. I use data from Renren.com, a Facebook-type social networking site based in China. The sample consists of students who have bachelor's degrees in China and have or will have graduate degrees in the U.S. On Renren, students have the option to add English names after their Chinese names, which quantifies acculturational characteristics. The difficulty of pronouncing the original Chinese name by native speakers of English creates a natural experiment on English-name usage. Using OLS and IV models, I find that a student who shows the English name online has more close friends who also show English names online. This provides empirical evidence that individual immigrants tend to have friends with similar acculturational characteristics.

Keywords: acculturation, homophily, name, friendship, immigration

*I have benefited from the comments of Francine D. Blau, Maria D. Fitzpatrick, Lawrence M. Kahn, Ravi Kanbur, Daniel T. Lichter, Eleonora Patacchini, Tony E. Smith, Nicolas R. Ziebarth, and seminar participants at Cornell University, Cornell Sociology Research Symposium, the North American Regional Science Congress (NARSC) and the European Regional Science Association (ERSA) annual meeting.

[†]Address: Sibley Hall 316, Cornell University, Ithaca, NY 14853. Email: dx42@cornell.edu.

1 Introduction

It has long been observed that immigrants receive supports from other immigrant friends (e.g., Winters et al., 2001; Munshi, 2003; Damm, 2009). A more fundamental question is: how do friendships form? Currarini et al. (2009) propose a theoretical model to show that *homophily*, the tendency of individuals in friendships to be similar, plays an important role in friendships. However, there is little prior research that empirically tests homophily in friendships due to the lack of data.

This paper attempts to study acculturational homophily in immigrants' friendships. As the first stage of assimilation (Gordon, 1964), acculturation (which is also commonly known as *cultural assimilation*) can be quantified based on the adoption and usage of the English name. Using a novel social networking data set retrieved from Renren.com¹, my sample consists of international graduate students in the U.S. who have bachelor's degrees in Chinese colleges and have (or will have) graduate degrees in U.S. schools. On Renren, a user has the option to add and show an English name after the Chinese name as the suffix, which allows us to study the acculturational characteristic: English-name usage. Based on the theory of homophily we may hypothesize that people tend to have friends with similar characteristics (e.g., Kandel, 1978; McPherson and Smith-Lovin, 1987; Schaefer et al., 2011). This leads to the empirical question in the paper: all else being equal, does an international student with online English-name usage have more close friends also with English names than a student without online English-name usage? Or in short, does acculturational homophily exist in friendships of international students?

It has been argued that naming convergence with natives is common among immigrants (e.g., Gerhards and Hans, 2009; Abramitzky et al., 2014, 2015). While not necessarily related to job skills, the adoption of Westernized (or Anglicized, in

¹The Renren.com Network is an online social networking site based in China, and is popular among university students. It is widely recognized as the Chinese version of Facebook.

the U.S.) names might help immigrants in the local labor market. Fryer and Levitt (2004) pioneer the study on effects of names, although with the main focus on minority groups: they find that the use of typical African American names is strongly correlated with socioeconomic status. Similarly, Bertrand and Mullainathan (2004) find that workers with “White names” receive more callbacks for job interviews. This conclusion also holds for immigration: renouncing foreign names helps immigrants in the labor market (Arai and Thoursie, 2009), and one possible explanation is that ethnic-sounding names cause discrimination in the labor market and the society (e.g., Shifman and Katz, 2005; Rubinstein and Brenner, 2014). As an effort of assimilation, using English names helps improve labor market experiences and overcome cultural barriers (Belot and Ederveen, 2012).

This paper connects two topics of demographic research: acculturation and homophily in friendships. So far, there are only a few prior empirical studies on homophily based on certain types of characteristics. The main reason for the scarcity of related research is that traditional social surveys do not pay much attention to friendships (Marmaros and Sacerdote, 2006). While some recent surveys ask questions about friendships (e.g., NLSY and Add Health), it is still difficult to acquire sufficient characteristics of friends to test homophily. Similar with this paper, most previous studies focus on university students using school administrative data (e.g., registrar data) or social networking data (e.g., Facebook) because these data provide some individual attributes of both the respondents and their friends. Marmaros and Sacerdote (2006) find that geographic proximity and race play crucial roles in friendship formation. Similarly, Mayer and Puller (2008) find that race is strongly related to social network formation on campus. More related to this paper, Girard et al. (forthcoming) employ a random experiment design and find that individual characteristics are robust predictors of friendships.

Another major challenge is that individual behaviors in friendships might not be

independent and isolated even conditional on a variety of observable characteristics. For example, unobservable attitudes and personalities are correlated with behaviors or characteristics that determine friendships, but might also affect friendship formation directly, which leads to the omitted variable problem (Burke and Sass, 2013). The reflection problem (Manski, 1993) also exists in the identification of homophily in friendships: in this context, a student’s English-name usage might be the cause or the outcome of his friends’ English-name usage — which cannot be determined by a simple OLS regression. One standard solution to this statistical challenge is the random assignment of peers. For example, some colleges have the policy that freshmen dormitories are randomly assigned, which creates experiments on peer effects (e.g., Marmaros and Sacerdote, 2001; Zimmerman, 2003; Foster, 2006; Stinebrickner and Stinebrickner, 2006). This empirical design, however, has two main drawbacks: first, such random assignments are conducted only in limited number of colleges, which threatens external validity; second, and more importantly, random assignments of peers do not actually create exogenous variation in individual characteristics, as well as behaviors in networks of friendships.

This paper overcomes the above challenges and sheds light on homophily in friendships using a unique social networking data set. The data set retrieved from Renren is similar with data from Facebook: researchers are able to obtain users’ age, gender, college and graduate school information, popularity on the social networking site, friend information, as well as English-name usage on Renren. The sample size is 7222. Observations of Chinese students come from 506 colleges in China and 69 graduate schools in the United States.

The data set used in this paper has two unique advantages. First, Renren users have the option to nominate up to ten *special friends*, which provides a natural definition of close friendships. The measure of close friends’ English-name usage is based on special friends, which makes more sense because non-close friends have dispropor-

tionately weak influences (Lin and Weinberg, 2014), which is an even more serious issue on online social networking sites where users accept friend requests from whom they do not know or are not familiar with. The second advantage is that different from other research, in this paper English-name usage can be predicted by a natural phonological experiment: all else being equal, a Chinese student is more likely to use an English name if his original Chinese name is difficult to be pronounced by native English speakers. This is because that the system of Romanized Chinese characters (i.e., the *pinyin* system) does not always well describe the pronunciation rule of Chinese (Bassetti, 2007); mispronunciations might cause embarrassment and discomfort for both Chinese and English speakers, and an effective solution is to use the English name. Based on this natural experiment design, I employ the instrumental variable (IV) strategy in which the indicator of the pronunciation difficulty serves as the IV for English-name usage.

Our empirical findings can be summarized as follows. First, both the OLS and the IV regression indicate that all else being equal, a student who uses the English name online has more close friends who also use English names online. This implies that acculturational homophily based on English-name usage does exist in friendships of international students. Second, the IV model yields greater estimate, and the OLS estimate might be downward biased due to the potential endogeneity problem. Third, on average, a student with English-name usage has approximately one more close friend with English-name usage, conditional on that the student nominates at least one special friend on Renren.

This paper contributes to the existing literature along two dimensions. First, it uses a new data set to show evidence of homophily in friendships based on acculturational characteristics. This is an early study that focuses on cultural assimilation using online social networking data. Second, although the IV is simple (as I only create a dummy variable to describe the pronunciation difficulty), it effectively solves

the endogeneity problem and this paper proposes a novel research methodology: the natural experiment based on the phonological nature, which might be a powerful tool in future research on cultural identity or well-being of immigrants because linguistic properties of pre-immigration names generate exogenous variation in assimilation-related behaviors in the host country.

The remainder of the paper is organized as follows. Section 2 describes the background. Section 3 introduces the data set and empirical strategies. Section 4 reports empirical results. Section 5 concludes.

2 Background

This section discusses the institutional background of the paper. I first briefly discuss English education in China and Chinese students' preparation for graduate studies in the U.S. These are closely related to consequences of the adoption and usage of English names by Chinese students, which will be introduced in the second part of this section. I finally discuss the natural phonological experiment on English-name usage by Chinese students: the difficulty of pronouncing Chinese names. This is the basis of empirical strategies employed in this paper.

2.1 English Education in China

As English is one of the three mandatory subjects tested in the National Higher Education Entrance Examination, Chinese students start to learn English since at least middle school (i.e., seventh grade), and in most urban regions, since elementary school (before seventh grade). English education continues in the post-secondary stage as passing the College English Test (CET) is generally the requirement for graduation in most Chinese colleges. However, because the CET is substantially different from TOEFL and GRE tests (Gao, 2010), which are required by U.S. graduate schools,

Chinese college students planning to apply for graduate studies abroad usually need to make decisions in the early stages of college and start to receive extra English training no later than the beginning of the junior year².

College students receive English education mainly through three channels. First, Chinese universities provide formal English education, which includes oral English courses taught by native speakers of English who come from, e.g., the U.S.. As an important tool for teaching, students are usually asked to adopt English names in English courses (Gao et al., 2005; Edwards, 2006), although English-name usage is apparently not required outside of class. Second, students receive extra training of English, mainly focusing on TOEFL and GRE, in private educational institutions. Third, peers play important roles in English education. Students share information and learning experience in networks of friendships, which has long been observed in the education economic literature (e.g., Hoxby, 2000).

2.2 English-Name Usage

I now turn to the adoption and usage of English names among Chinese students. It has long been observed that the degree of assimilation, if measured by language proficiency, is positively correlated with labor market outcomes in the host country (e.g., Kossoudji, 1988; Chiswick, 1991; Berman et al., 2003; Chiswick and Miller, 2003). This is not surprising, as language proficiency is one component of job skills. But somewhat surprisingly, naming also matters in the labor market (Bertrand and Mullainathan, 2004; Fryer and Levitt, 2004; Arai and Thoursie, 2009), even if names are not necessarily related to job skills. This might be explained by discrimination against workers with foreign or exotic names in the labor market in the host country (e.g., Rubinstein and Brenner, 2014).

In this paper, I follow the way of Abramitzky et al. (2014, 2015) and quantify

²Early preparation for English tests training is especially important for college students who enter U.S. graduate schools before 2012, as the GRE test only held twice per year in China before that.

acculturation by English-name usage on the online social networking site. There are two points worth mentioning about English-name usage. First, English-name usage is not equivalent to English-name adoption. As discussed earlier, students usually adopt English names in college or even high school English courses in China, hence there is no much variation in the adoption of the English name. Second, showing the English name online is a stronger condition than using the English name outside of English class: a student who uses the English name in daily life does not necessarily put it on a networking site based in China; on the contrary, a student who puts the English name online should use the English name outside of class. Hence, students with English-name usage on the online social networking site are arguably frequent users of English names in daily life.

In the empirical analysis, I focus on students who have obtained their bachelor's degrees in China and have (or will have) graduate degrees in the U.S. My sample does not consist of students who 1) stay in China and do not continue graduate studies in the U.S., and 2) receive undergraduate education in the U.S. or other English-speaking countries (such as the U.K.). The first group of students work in China, and it makes little sense for them to use English names; the number of observations with English-name usage is small, and there is no much variation in online English-name usage. The second group of students come to English speaking countries before adulthood; migration occurs early and it is more like a common social rule to use English names in college life, which also results that there is no much variation in online English-name usage. The sample of international graduate students with bachelor's degrees in China provides adequate observations with English names as well as some variation in English-name usage on the online social networking site.

2.3 The Pronunciation Difficulty of Chinese Names

A simple comparison between students with or without online English-name usage might yield inconsistent results because using English names, as discussed in Section 1, is potentially endogenous. To establish the causal relationship between self usage of the English name and friends' usage of English names, an exogenous variable is needed to predict self usage. While ideally such an exogenous variable is constructed by the random experiment, this is not applicable in language teaching and education. In this paper I exploit a unique natural experiment based on phonological properties of Chinese characters.

The natural experiment on English-name usage is based on linguistic findings that some Chinese characters are difficult to be pronounced by native speakers of Western languages, which results that graduate students with difficult-to-pronounce Chinese names are more likely to use English names. The *pinyin* system in Chinese is designed to describe the pronunciation rule using the Roman alphabet (i.e., the Latin alphabet) and Chinese names are transcribed by Romanized characters in English. However, because of phonological differences between Chinese and English (as well as other Western languages), the *pinyin* system does not always precisely describe the pronunciation rule of Chinese³ (Bassetti, 2007).

Note that the term *pronunciation difficulty* that I use in this paper is with respect to native speakers of Western languages who have no systematic training in Chinese. If the pronunciation difficulty is with respect to Chinese speakers, then it is likely to observe systematic differences between students with and without difficult-to-pronounce names (in, e.g., family background), and such an experiment does not generate random assignments of difficult-to-pronounce names. However, it is unlikely for a Chinese family to consider the pronunciation difficulty of the name by English

³In Appendix A I present a list of Romanized Chinese characters that have different pronunciations in Chinese and in English (based on *pinyin*).

speakers at the time the child was born, which is more than 20 years ago in this context. Hence intuitively, the pronunciation difficulty is exogenous.

Consequences of name mispronunciations include embarrassment and discomfort for both Chinese and English speakers. One example involves the character *mèng*, which has the meaning of *dream*. This character is widely used in female given names, but is usually pronounced as *men* by English speakers⁴. A solution to this problem is to adopt and use an English name along with the original Chinese name. English-name usage in daily life is further reflected by English-name usage on the social network. This develops the research design in which the pronunciation difficulty serves as a natural experiment and predicts English-name usage in the Renren data set. In Section 3 and 4 I will introduce the instrumental variable strategy based on this natural experiment and analyze the validity of the IV.

3 Data and Empirical Strategies

3.1 Online Social Networking Data

I first introduce the online social networking data set retrieved from Renren.com. Founded in 2005, Renren is a provider of online social networking services based in China. Similar with Facebook, Renren requires users to provide personal biographical information. Researchers are also able to acquire the number of Renren friends and the number of time the personal page is visited by others Renren users. These two variables roughly describe the the popularity on the social networking site. Moreover, Renren has two unique features. First, a user has the option to add an English name after the original Chinese name as the suffix⁵, which quantifies acculturation status.

⁴Another related example for male names is *shí*, meaning *stone* or *time* in Chinese. This character is usually pronounced as *she* by English speakers.

⁵In fact, a Renren user can add any English word as the suffix. Ideally the English word after the Chinese name should be the user's English name, but the user might also put the name of the idol as the suffix (e.g., *Isaac Newton*), or other non-name English words (e.g., *Mathematics*). While

Second, a user has the option to nominate up to ten *special friends* shown on the personal Renren page, which defines close friendships. As for the definition of close friendships, there are two advantages of the Renren data set: 1) a Renren user’s close friendship is well defined because these special friends are nominated by the user himself; 2) a Renren user’s special friends usually only consist of his friends, significant other, or spouse, but not parents⁶ and siblings⁷ — who are naturally in and are not selected into the network of social relationships.

My sample consists of 7222 Chinese graduate students who have bachelor’s degrees in Chinese colleges and have or will have graduate degrees in the U.S. These Chinese students come from 506 colleges in China and 69 American graduate schools in Alabama (AL) through Missouri (MO)⁸. Table 1 and 2 report the basic summary statistics of observable characteristics and outcome variables.

In Table 1 I present the summary statistics of the full sample set, in which observations might list or do not list special friends on Renren. In the full sample set I find that 13.3% of all students (which is approximately 1000 students) show their English names online, and I define these users as frequent users of English names; other 86.7% students do not show English names on Renren. The average number of special friends nominated by users is 1.61. On average, Renren users in the sample receive slightly fewer than 10000 times of visits by other users and have about 600 online friends, although observations in the data set are fairly heterogeneous in these characteristics. For the numerical convenience, I will use the number of visits divided

it is easy to identify some non-English-name cases, it is unlikely to find out all of these cases, which leads to the measurement error problem: some non-name words are identified as English names and some English names are identified as non-name words. However, the instrumental variable model in this paper can help tackle this statistical issue.

⁶Unlike on Facebook, there are limited number of Chinese parents having accounts on Renren. This can be verified because we can observe special friends’ profile pictures and a friend with similar age with the user apparently cannot be the parent. In the data set I find that at least 99% of all observations do not list parents as special friends.

⁷Due to the “one child policy” in China, the case that a student has siblings is rare.

⁸Due to the data limitation, I am unable to obtain student account information from all graduate schools in the U.S. from Renren; the data set involves 69 graduate schools in states ranking before Missouri (MO) in alphabetical order.

by 10000 and the number of all friends divided by 100 as independent variables in regressions. In the full sample, about 49% of all students are male. On average, students have entered college in China 8 years ago.

Table 1: Summary Statistics: Full Sample

	Mean	Std. dev.	Min	Max
Independent Variables:				
English-name usage	0.133	(0.340)	0	1
# of special friends	1.610	(2.318)	0	10
(# of visits received)/10000	0.935	(1.288)	0.0003	25.268
(# of all friends)/100	5.859	(4.815)	0.02	185.540
Male dummy	0.489	(0.500)	0	1
Year since entering college	8.684	(1.826)	5	16
Category 1 Chinese college dummy	0.205	(0.403)	0	1
Category 2 Chinese college dummy	0.270	(0.444)	0	1
Category 1 US school dummy	0.140	(0.347)	0	1
Category 2 US school dummy	0.497	(0.500)	0	1
IV and Dependent Variables:				
Pronunciation difficulty indicator	0.422	0.494	0	1
# of special friends w/ Eng. names	0.106	0.371	0	4

Full sample includes all students with and without special friends on Renren.

Standard deviations are in parentheses. Number of observations: 7222.

This social networking data set allows me to control for school information in the empirical analysis. One simple approach is to introduce college and graduate school fixed effects. Alternatively, I partition Chinese colleges and U.S. graduate schools into three categories by school prestige in a coarse manner. For Chinese colleges, category 1 colleges include universities which are members of the C9 League⁹; category 2 colleges include universities which are sponsored by “Project 985”¹⁰ but are not members of the C9 League; category 3 colleges includes all other universities. For U.S. graduate schools, category 1 schools include universities that are in top 10

⁹Equivalent to the Ivy League in the U.S. and the Exzellenzinitiative in Germany, China’s C9 League comprises nine most renowned universities in Mainland China, including Peking University, Tsinghua University, Zhejiang University, University of Science and Technology of China, Fudan University, Nanjing University, Shanghai Jiao Tong University, Harbin Institute of Technology, and Xi’an Jiao Tong University.

¹⁰Project 985 is an official project initiated by national and local governments that allocates funding to 39 reputable research universities in Mainland China after careful evaluations on research and teaching quality.

of the US News Best Global University Rankings, plus all other Ivy League schools¹¹; category 2 schools include all other universities that are members of the Association of American Universities (AAU); category 3 schools include all other U.S. universities in the data set. I find about 20% of all students come from the C9 League, i.e., category 1 Chinese colleges, and 27% are from category 2 colleges. 14% of all students attend eight category 1 schools in the U.S. and around 50% of all students attend category 2 schools, i.e., all other AAU schools.

I finally turn to the instrumental variable and the dependent variable. In the full sample, 42.2% of all students have difficult-to-pronounce Chinese names — in other words, there are about 3000 students whose original Chinese names are hard to be pronounced by native speakers of Western languages. Among all students, the average number of special friends with English-name usage is 0.1. This small number, however, is somewhat underestimated because the number of special friends is unconditional on that the user nominates nonzero special friends.

In Table 2 I focus only on the sub-sample in which the observations nominate at least one special friend on Renren. 43.9% of all students (or 3171 students) have nonzero Renren special friends. Table 2 shows that students in this sub-sample set are systematically different from students without special friends on Renren. In the sub-sample, there are relatively more students with English-name usage on Renren. The average number of special friends, conditional on that students have at least one special friend, is close to 4. Compared with other students in the full sample, these students receive more visits from other Renren users and have more Renren friends. Also, there are relatively fewer male users in the sub-sample. The average year since

¹¹Category 1 U.S. graduate schools include following eight universities: Harvard University, Massachusetts Institute of Technology (MIT), University of California-Berkeley (UCB), Stanford University, California Institute of Technology (Caltech), University of California-Los Angeles (UCLA), University of Chicago, and Yale University. In theory, this category should also contain Columbia University, University of Pennsylvania, Cornell University, Brown University, and Dartmouth College. However, these universities are not included because I have no data on graduate students in New Hampshire, New York State, Pennsylvania, and Rhode Island.

entering college is 8.61. However, both data sets have similar composition of school categories, as shown by the summary statistics of school category fixed effects.

Table 2: Summary Statistics: Sub-Sample, with ≥ 1 Special Friends

	Mean	Std. dev.	Min	Max
Independent Variables:				
English-name usage	0.154	(0.361)	0	1
# of special friends	3.668	(2.165)	1	10
(# of visits received)/10000	1.116	(1.418)	0.0007	25.102
(# of all friends)/100	6.322	(4.629)	0.04	142.180
Male dummy	0.420	(0.494)	0	1
Year since entering college	8.61	(1.69)	5	16
Category 1 Chinese college dummy	0.212	(0.409)	0	1
Category 2 Chinese college dummy	0.267	(0.443)	0	1
Category 1 US school dummy	0.135	(0.342)	0	1
Category 2 US school dummy	0.523	(0.500)	0	1
IV and Dependent Variables:				
Pronunciation difficulty indicator	0.425	0.494	0	1
# of special friends w/ Eng. names	0.241	0.531	0	4
% of special friends w/ Eng. names	0.062	0.159	0	1

Sub-sample includes students with at least one special friend on Renren.

Standard deviations are in parentheses. Number of observations: 3171.

In the second panel of Table 2 I find 42.5% of students in the sub-sample have difficult-to-pronounce Chinese names. Note that this ratio is very close to the ratio reported in Table 1, which is reasonable because the pronunciation difficulty of the Chinese name should be exogenous and unrelated to personal characteristics, including the number of special friends nominated. In this sub-sample, the average number of special friends with English names shown online is 0.241. Because the sub-sample only contains students who have at least one special friend, now I am able to define the proportion of special friends with English names. On average, about 6% of special friends show their English names on Renren.

In Table 3, I compare the instrument variable and outcome variables between students who show and do not show English names on Renren. There are approximately 1000 observations with English-name usage in the full sample, and 64.8% of them have difficult-to-pronounce Chinese names. In contrast, only 38.8% of students

not showing English names have difficult-to-pronounce Chinese names. The ratios of students with difficult-to-pronounce names are very similar with those shown in the panel for the sub-sample.

Table 3: Comparing Students with and without English Names

	Showing English names	Not showing English names
Full Sample:		
Pronunciation difficulty indicator	0.648 (0.478)	0.388 (0.487)
# of special friends w/ English names	0.401 (0.699)	0.060 (0.262)
Observations	960	6262
Sub-Sample:		
Pronunciation difficulty indicator	0.640 (0.480)	0.386 (0.487)
# of special friends w/ English names	0.787 (0.809)	0.141 (0.387)
% of special friends w/ English names	0.193 (0.233)	0.038 (0.128)
Observations	489	2682

Full sample includes all students with and without special friends on Renren.

Sub-sample includes students with at least one special friend on Renren.

Standard deviations are in parentheses.

In the full sample, the average number of special friends with online English-name usage is 0.4 among observations who show English names online. In contrast, students who do not show English names online only have 0.06 special friends with online English-name usage. The comparative qualitative pattern remains for the sub-sample in which students have at least one special friend, as shown in the second panel. Similarly, students with online English-name usage also have higher proportion of special friends with online English-name usage. The descriptive statistics in Table 3 roughly show acculturational homophily in friendships, although these results might be biased because English-name usage is not orthogonal.

3.2 Testing Homophily: OLS and IV Models

In this part, I present the regression models to test homophily in friendships of international students based on acculturational characteristics: online English-name usage. I first introduce the basic specification, and then discuss the potential statis-

tical challenge as well as the solution. The core empirical question is whether self usage of the English name online is causally related to the number of close friends with online English-name usage. To study this, I start with an OLS regression:

$$N_i^F = \alpha' \mathbf{X}_i + \beta E_i^S + \gamma T_i + \varepsilon_i \quad (1)$$

where i indexes individual. N_i^F is the number of close friends with online English-name usage, and close friends are defined by *special friends* nominated by the individual i . E_i^S is the main regressor, which is a dummy variable indicating whether i shows the English name on Renren. T_i is the total number of special friends that i nominates, ranging from 0 to 10.

In (1), \mathbf{X}_i is a set of control variables of personal characteristics and information, including gender, year since entering college, total number of Renren friends, and total number of visits received from other users. Besides, \mathbf{X}_i also includes variables of college and graduate school information. One way to control for school information is to introduce college and graduate school fixed effects. Alternatively, we can use school category fixed effects introduced in Section 3.1 as regressors. I find very similar empirical results using either ways to construct school variables in \mathbf{X}_i .

There are two major challenges with causal inference based on the specification in (1). The first challenge is the reflection problem (Manski, 1993). In this context, while E_i^S might have a positive effect on N_i^F , N_i^F might also reversely affect E_i^S through peer influences. The second challenge is the omitted variable bias problem, since there are unobservable personal attitudes and characteristics related to both E_i^S and N_i^F . These two challenges threaten the OLS estimate of β in (1), and in fact it is even difficult to determine the sign of bias.

To tackle the above statistical issues, I propose an IV strategy which relies on the idea of the natural experiment that the pronunciation difficulty of the Chinese name

exogenously affects English-name usage. This leads to the first-stage regression:

$$E_i^S = \gamma' \mathbf{X}_i + \tau D_i + \eta T_i + \epsilon_i \quad (2)$$

where D_i indicates the pronunciation difficulty: $D_i = 1$ if i 's Chinese name is considered to be difficult to be pronounced by native speakers of English, and $D_i = 0$ otherwise. In (2), D_i serves as the IV for the endogenous regressor E_i^S , and the fitted value is used in (1) to obtain the IV estimate.

There are two assumptions required to ensure the validity of the IV. First, D_i must be closely correlated with E_i^S so that the IV has the enough statistical power to predict the endogenous regressor. This can be verified by the first-stage regression, which will be presented in Section 4. Second, D_i must affect the outcome N_i^F only through E_i^S , i.e., the pronunciation difficulty should have the impact on the number of close friends with English-name usage only through self usage of the English name. This assumption of the exclusion restriction will be discussed in Section 3.3.

3.3 The Validity of the IV and the Exclusion Restriction

As introduced in previous sections, the empirical analysis relies on exogenous variation in English-name usage. The indicator of the pronunciation difficulty of the Chinese name is a natural candidate of the instrumental variable. In this part, I discuss the validity of this instrument.

The first condition of a valid IV is that it should well predict the endogenous variable. In Table 3 we have seen that English-name usage is positively correlated with the pronunciation difficulty indicator. This can also be found in Table 4, Row 1: in the full sample, a Chinese student with the difficult-to-pronounce Chinese name is significantly more likely to use a chosen English name on Renren. This conclusion remains after controlling for individual characteristics: in Section 4 I will report first-stage

regressions and show that the pronunciation difficulty of the Chinese name predicts English-name usage. In the second row I observe that a Chinese student with a difficult-to-pronounce Chinese name has significantly more special friends with online English-name usage. This result will be discussed again by reduced-form regressions reported in Section 4.

Table 4: Checking on Systematic Differences: The Full Sample

	w/o difficult-to-pronounce names	w/ difficult-to-pronounce names	<i>p</i> -value
English-name usage	0.081 (0.273)	0.204 (0.403)	***
# of special friends w/ Eng. names	0.081 (0.312)	0.139 (0.438)	***
# of special friends	1.581 (2.292)	1.650 (2.353)	n.s.
(# of visits)/10000	0.933 (1.315)	0.937 (1.250)	n.s.
(# of friends)/100	5.820 (4.544)	5.913 (5.163)	n.s.
Male	0.486 (0.500)	0.493 (0.500)	n.s.
Year since entering college	8.732 (1.822)	8.619 (1.828)	*
Category 1 Chinese college dummy	0.203 (0.402)	0.207 (0.405)	n.s.
Category 2 Chinese college dummy	0.271 (0.445)	0.269 (0.443)	n.s.
Category 1 US school dummy	0.138 (0.344)	0.144 (0.351)	n.s.
Category 2 US school dummy	0.498 (0.500)	0.495 (0.500)	n.s.
Observations	4173	3049	—

Full sample includes all students with and without special friends on Renren.

The first two rows show systematic differences in the name dummy and the outcome variable.

The following nine rows show systematic differences in control variables.

Standard deviations are in parentheses.

Unpaired *t* tests are employed. n.s.: $p \geq .1$; †: $p < .1$. *: $p < .05$; **: $p < .01$; ***: $p < .001$.

Another assumption for the IV is that the instrument should affect the outcome only through the endogenous regressor. Random assignments generate valid IVs in the sense that the treatment is randomly assigned. In the context of the natural experiment, it is ideal that the pronunciation difficulty is “randomized”: unrelated to characteristics. In the remaining rows in Table 4, I focus on systematic differences in other individual characteristics between students with and without difficult-to-pronounce names. I find no systematic difference between students with and without difficult-to-pronounce names for most characteristics; the only observable difference is that students with difficult-to-pronounce names entered college slightly earlier.

In Table 5 I turn to the sub-sample in which students have at least one special friend. The first three rows show differences in English-name usage and outcome variables. Similar with Table 4, this table shows the strong relationship between the pronunciation difficulty and English-name usage, as well as the number (and the proportion) of special friends with English-name usage.

Table 5: Checking on Systematic Differences: The Sub-Sample

	w/o difficult-to- pronounce names	w/ difficult-to- pronounce names	<i>p</i> -value
English-name usage	0.097 (0.295)	0.232 (0.422)	***
# of special friends w/ Eng. names	0.186 (0.451)	0.314 (0.616)	***
% of special friends w/ Eng. names	0.050 (0.145)	0.079 (0.174)	***
# of special friends	3.622 (2.154)	3.730 (2.180)	n.s.
(# of visits)/10000	1.095 (1.428)	1.144 (1.403)	n.s.
(# of friends)/100	6.323 (5.247)	6.321 (3.635)	n.s.
Male	0.423 (0.494)	0.417 (0.493)	n.s.
Year since entering college	8.640 (1.676)	8.581 (1.715)	n.s.
Category 1 Chinese college dummy	0.205 (0.404)	0.222 (0.416)	n.s.
Category 2 Chinese college dummy	0.268 (0.443)	0.265 (0.442)	n.s.
Category 1 US school dummy	0.131 (0.338)	0.140 (0.347)	n.s.
Category 2 US school dummy	0.518 (0.500)	0.528 (0.499)	n.s.
Observations	1822	1349	—

Sub-sample includes students with at least one special friend on Renren.

The first three rows show systematic differences in the name dummy and outcome variables.

The following nine rows show systematic differences in control variables.

Standard deviations are in parentheses.

Unpaired *t* tests are employed. n.s.: $p \geq .1$; †: $p < .1$. *: $p < .05$; **: $p < .01$; ***: $p < .001$.

I investigate other individual characteristics starting from Row 4. There is no significant difference in any personal attribute between students with and without difficult-to-pronounce original Chinese names. In particular, unlike in Table 4, here I find no statistically significant difference in the year since entering college between two groups of students. In both sample sets, our natural experiment creates an arguably valid instrument for usage of the English name, in the sense that the assignment of the pronunciation difficulty is nearly random.

Table 6: Other Possible Channels through which the IV Affects Outcomes

	Outcomes			Personal Characteristics		
	(1)	(2)	(3)	(4)	(5)	(6)
Pronunciation difficulty	0.003 (0.006)	0.009 (0.015)	-0.001 (0.005)	0.066 (0.053)	0.006 (0.009)	0.004 (0.008)
# of special friends	0.040*** (0.001)	0.039*** (0.004)	0.000 (0.001)	—	-0.004† (0.002)	-0.006** (0.002)
(# of visits)/10000	0.004 (0.003)	0.006 (0.004)	0.001 (0.002)	0.279*** (0.024)	0.015*** (0.004)	0.009* (0.003)
(# of all friends)/100	0.002* (0.001)	0.006** (0.002)	0.002* (0.001)	0.024*** (0.007)	0.013*** (0.001)	0.008*** (0.001)
Male	-0.001 (0.006)	-0.005 (0.015)	-0.004 (0.005)	-0.743*** (0.053)	0.074*** (0.009)	-0.007 (0.008)
Year after entry	-0.000 (0.002)	-0.001 (0.004)	0.000 (0.002)	-0.058*** (0.015)	0.023*** (0.003)	-0.006** (0.002)
Cat. 1 Chinese college	n.r.	n.r.	n.r.	n.r.	—	0.290*** (0.010)
Cat. 1 Chinese college	n.r.	n.r.	n.r.	n.r.	—	0.027** (0.009)
School category FE	Yes	Yes	Yes	Yes	No	—
R ²	0.126	0.065	0.006	0.072	0.045	0.141
Observations	6262	2682	2682	7222	7222	7222

Regression (1) - (3): all students who show English names online are **not** included.

Dependent variable in (1): the number of special friends showing English names, the full sample.

Dependent variable in (2): the number of special friends showing English names, the sub-sample, in which students have at least one special friend.

Dependent variable in (3): the proportion of special friends showing English names, the sub-sample.

Regression (4) - (6): the full sample (i.e., all students are included).

Dependent variable in (4): the total number of special friends.

Dependent variable in (5): the dummy indicating category 1 Chinese colleges.

Dependent variable in (6): the dummy indicating category 1 U.S. graduate schools.

Chinese school category dummies are added in (1) - (4), but coefficients are not reported (“n.r.”).

Standard errors are in parentheses. †: $p < .1$. *: $p < .05$; **: $p < .01$; ***: $p < .001$.

I finally do some additional checks on the validity of the IV and close this section. In Table 6, I first exclude all students with online English-name usage from the sample and run reduced-form regressions of outcomes on the instrument, i.e., the pronunciation difficulty indicator. The exclusion restriction is likely to be violated if the pronunciation difficulty still significantly affects the number of special friends with English-name usage: instead of self usage of the English name, there might be some other channels or mechanisms through which the IV affects outcomes because

students with English-name usage are now not in the sample. However, I find no evidence of this through Column 1 to 3 in Table 6: the pronunciation difficulty plays no role in friendship formation for students without online English-name usage.

In Column 4, 5, and 6, I examine whether the pronunciation difficulty is the predictor of other personal characteristics. If it can also predict these characteristics, we might worry about the validity of the IV because the IV is likely to be the predictor of related unobservable characteristics as well, which breaks the exclusion restriction. In Column 4 I regress the total number of special friends on the pronunciation difficulty, and find no evidence that students with difficult-to-pronounce Chinese names are more likely to nominate more special friends. In Column 5 I regress the category 1 Chinese college dummy on the pronunciation difficulty and other covariates. The IV strategy can be a problem if good universities are more likely to admit students with difficult-to-pronounce names. However, there is no evidence that the pronunciation difficulty has any major effect. This conclusion remains true when I test the admission of Chinese students by U.S. graduate schools in Column 6.

4 Results

In this section, I report the empirical findings in this paper. I begin with first-stage and reduced-form regressions. I then move to OLS and IV regressions and present main results of this paper. I conclude the empirical part by some additional tests.

4.1 First-Stage and Reduced-Form Regressions

I begin this section with first-stage and reduced-form regressions. In Column 1, Table 7, I first regress the dummy of online English-name usage on the indicator of the pronunciation difficulty and other covariates. This regression shows the strong statistical connection between the pronunciation difficulty of the Chinese name and

English-name usage on Renren, and the F-statistic is far greater than 10. From the perspective of the first-stage relationship, the natural experiment on English-name usage generates a valid IV. In Column 2 I repeat the first-stage regression based on the sub-sample in which students nominate at least one special friend on Renren, and the qualitative pattern of the first-stage relationship remains.

Table 7: First-Stage Regressions and Reduced-Form Regressions

	First-Stage			Reduced-Form		
	(1)	(2)	(3)	(4)	(5)	(6)
Pronunciation difficulty	0.121*** (0.008)	0.132*** (0.013)	0.053*** (0.008)	0.120*** (0.018)	0.076*** (0.014)	0.029*** (0.006)
# of special friends	0.009*** (0.002)	0.012*** (0.003)	0.065*** (0.002)	0.062*** (0.004)	0.045*** (0.003)	0.000 (0.001)
(# of visits)/10000	0.025*** (0.004)	0.025*** (0.005)	0.014*** (0.003)	0.023** (0.007)	0.012* (0.006)	0.004† (0.002)
(# of all friends)/100	-0.000 (0.001)	0.000 (0.002)	0.002* (0.001)	0.006** (0.002)	0.007*** (0.002)	0.002* (0.001)
Male	-0.026** (0.008)	-0.026* (0.013)	-0.010 (0.008)	-0.025 (0.019)	-0.024† (0.014)	-0.010† (0.006)
Year since entry	-0.016*** (0.002)	-0.019*** (0.004)	-0.007** (0.002)	-0.021*** (0.005)	-0.014** (0.004)	-0.004* (0.001)
School category FE	Yes	Yes	Yes	Yes	Yes	Yes
R ²	—	—	0.194	0.119	0.109	0.022
First-stage F-statistic	48.97	24.48	—	—	—	—
Observations	7222	3171	7222	3171	3171	3171

Dependent variables in (1) and (2): whether the English name is shown on Renren.

Dependent variables in (3) and (4): the number of special friends showing English names online.

Dependent variable in (5): whether there is at least one special friend showing the English name.

Dependent variable in (6): the proportion of special friends showing English names online.

Regression (1) and (3): the full sample. Regression (2), (4), (5), and (6): the sub-sample.

Standard errors are in parentheses. †: $p < .1$. *: $p < .05$; **: $p < .01$; ***: $p < .001$.

From Column 3, I report results of reduced-form regressions. In column 3 I focus on the full sample and regress the number of special friends with English-name usage on the pronunciation difficulty indicator. Column 3 shows that a student with the difficult-to-pronounce Chinese name is associated with a greater number of special friends with online English names. In Column 4 I again focus only on the sub-sample and exclude students who do not nominate any close friend as special friends. The coefficient of the IV is nearly doubled. I repeat the exercise in Column 5 and 6

using different measures of the outcome: in Column 5 I find that a student with the difficult-to-pronounce Chinese name is more likely to have at least one special friend with online English-name usage, and in Column 6 I find that the pronunciation difficulty is positively associated with the proportion of special friends online English-name usage. In addition, in all regressions I find positive coefficients of the number of visits and the number of friends on Renren. This indicates that the popularity on the social network is positively correlated with close friends' English-name usage, but the coefficients are small. In addition, male students are less likely to have close friends with online English names, and the year since entering college is also negatively correlated with friends' usage of English names. Albeit statistically significant, the impacts of the above non-aculturational characteristics are moderate at best.

4.2 OLS and IV Regressions

I now turn to the main part of the empirical section. Using the number of special friends with online English-name usage as the outcome variable, in Table 8 I begin with an OLS regression of the outcome on the indicator of self usage of the English name on Renren. School category fixed effects and other covariates are included. The OLS regression in Column 1 shows the positive correlation between acculturational characteristics — English-name usage — of the individual and the close friends. This, however, does not indicate the causal relationship since English-name usage is potentially endogenous. Based on the natural experiment idea, in Column 2 I use the pronunciation difficulty of the original Chinese name to instrument for online English-name usage. The result shows that international students from China indeed tend to have friends with similar acculturational characteristics. Unconditional on that students show nonzero special friends, a student showing his English name online has nearly 0.5 more close friends also showing English names online on average. The OLS estimate in Column 1 is downward biased.

In Column 3 I run the similar OLS regression, but now using Chinese college and U.S. graduate school fixed effects. This is followed by the replication of the IV regression in Column 4 using these school fixed effects. The qualitative pattern in Column 1 and 2 is again observed in Column 3 and 4: acculturational homophily exists in friendships of graduate students from China, in the sense that self usage of the English name is positively correlated with close friends' usage of English names, and the OLS estimate is again downward biased.

Table 8: Testing Acculturational Homophily: OLS and IV Models

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
English name dummy	0.291*** (0.011)	0.434*** (0.064)	0.290*** (0.012)	0.520*** (0.096)	0.587*** (0.023)	0.907*** (0.128)
# of special friends	0.063*** (0.002)	0.061*** (0.002)	0.062*** (0.002)	-0.004 (0.003)	0.055*** (0.004)	0.051*** (0.004)
(# of visits)/10000	0.006† (0.004)	0.003 (0.004)	0.008* (0.004)	0.027*** (0.006)	0.008 (0.007)	-0.000 (0.008)
(# of all friends)/100	0.002* (0.001)	0.002* (0.001)	0.002* (0.001)	-0.000 (0.001)	0.006** (0.002)	0.006** (0.002)
Male	-0.002 (0.008)	0.001 (0.008)	0.001 (0.008)	0.039** (0.012)	-0.010 (0.017)	-0.002 (0.018)
Year since entry	-0.002 (0.008)	-0.000 (0.002)	-0.004 (0.002)	0.035** (0.004)	-0.010† (0.005)	-0.003 (0.006)
School category FE	Yes	Yes	No	No	Yes	Yes
School FE	No	No	Yes	Yes	No	No
R ²	0.258	—	0.324	—	0.260	—
Observations	7222	7222	7222	7222	3171	3171

IV: the pronunciation difficulty indicator is the IV for whether showing the English name online.

Dependent variables: the number of special friends showing English names online.

Regression (1) and (2): the full sample, with school category fixed effects.

Regression (3) and (4): the full sample, with school fixed effects.

Regression (5) and (6): the sub-sample, in which observations have at least one special friend on Renren, with school category fixed effects.

Standard errors are in parentheses. †: $p < .1$. *: $p < .05$; **: $p < .01$; ***: $p < .001$.

I conclude this table by focusing on the sub-sample in which students nominate at least one special friends on Renren. This will yield greater magnitudes of the OLS and IV estimate, as all observations who do not list close friends online are excluded. In Column 5 and 6, I report the OLS and IV regressions using school

category fixed effects. Acculturational homophily is again observed: students with English-name usage tend to have more close friends also with English-name usage. The OLS estimate is downward biased, and the IV regression in Column 6 indicates that on average, a student with online English-name usage has nearly one more close friend also uses the English name online. This magnitude is large relatively to the size of the close friendship — recall that the average number of special friends on Renren (conditional on that the student nominates at least one special friend) is around 3.5, as reported in Table 2. This shows that homophily based on acculturational characteristics plays an important role in friendships for international students.

4.3 Additional Tests

I conclude the empirical part by conducting several additional tests to check the robustness of our previous empirical findings.

In Table 9, I start with two regressions based on alternative measures of close friends' acculturation. Column 1 shows the IV regression of the proportion of special friends with English-name usage. The pronunciation difficulty serves as the IV for self usage of the English name. Focusing on the sub-sample in which observations have nonzero special friends, I find that using the English name on Renren increases the fraction of close friends showing English names online by 21.7%. In Column 2 I remeasure acculturational characteristics in a coarse manner by defining a dummy variable which equals 1 if there is at least one special friend showing the English name, and 0 otherwise. Again, I only focus on the sub-sample in which observations have nonzero special friends. The result in (2) shows that using the English name online increases the likelihood that there exists close friends also showing English names online by almost 60%.

In Column 3 to 6 I return to the outcome variable studied in Table 8 — the number of special friends with online English-name usage. In the following four IV

Table 9: Acculturational Homophily: Additional Tests

	Other Measures			Subpopulations		
	(1)	(2)	(3)	(4)	(5)	(6)
English name dummy	0.217*** (0.041)	0.578*** (0.096)	0.410*** (0.085)	0.435*** (0.073)	0.461*** (0.129)	0.499*** (0.078)
# of special friends	-0.002† (0.001)	0.038*** (0.003)	0.053*** (0.002)	0.062*** (0.002)	0.060*** (0.003)	0.066*** (0.002)
(# of visits)/10000	-0.001 (0.002)	-0.002 (0.006)	0.006 (0.005)	0.003 (0.004)	0.005 (0.005)	0.005 (0.005)
(# of all friends)/100	0.002** (0.001)	0.007*** (0.002)	0.003** (0.001)	0.002* (0.001)	0.002† (0.001)	0.001 (0.001)
Male	-0.004 (0.006)	-0.009 (0.013)	0.001 (0.011)	0.007 (0.010)	-0.015 (0.016)	-0.003 (0.011)
Year since entry	-0.000 (0.002)	-0.003 (0.004)	0.002 (0.003)	-0.003 (0.003)	0.000 (0.004)	-0.002 (0.005)
School category FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3171	3171	3429	4600	1934	4880

IV: the pronunciation difficulty indicator is the IV for whether showing the English name online.

Dependent variable in (1): the proportion of special friends showing English names online.

Dependent variable in (2): whether there is at least one special friend showing the English name.

Dependent variables in (3) to (6): the number of special friends showing English names online.

Regression (1) and (2): the sub-sample, in which observations have at least one special friend.

Regression (3): the sub-sample, in which students are from Chinese category 1 and 2 colleges.

Regression (4): the sub-sample, in which students are from U.S. category 1 and 2 schools.

Regression (5): the sub-sample, in which students are from both Chinese **and** U.S. cat. 1 schools.

Regression (6): the sub-sample, in which students entered the college in or after 2006.

Standard errors are in parentheses. †: $p < .1$. *: $p < .05$; **: $p < .01$; ***: $p < .001$.

regressions I investigate homophily in friendships in subpopulations¹². In Column 3 I study students who graduate from Chinese category 1 and 2 colleges. The IV estimate is slightly lower than that reported in Column 2, Table 8, although the causal relationship remains. In Column 4 I turn to students who enter U.S. category 1 and 2 graduate schools. Again, the magnitude of the coefficient is very close to the previous estimate in Table 8. In Column 5 I find slightly greater magnitude of acculturational homophily for students entering one of nine category 1 Chinese colleges *and* one of eight category 1 U.S. graduate schools. But overall, there is no much evidence that homophily in acculturational characteristics varies by school prestige — which

¹²While not reported here, I check all first-stage regressions and coefficients of the IV are statistically significant; all F-statistics are comfortably greater than 10.

is closely correlated with students' academic achievement and ability. In the last column, I focus on students entering Chinese colleges in or after 2006. These students are younger than the full population. The qualitative pattern remains, but the IV estimate becomes greater. This implies that the tendency to have friends with similar acculturational characteristics might be stronger among younger students.

5 Conclusion

This paper examines acculturational homophily in friendships of international graduate students based on English-name usage. I use the data from Renren.com, a Facebook-type online social networking site based in China. The sample consists of Chinese students attending college in China and graduate school in the U.S. Acculturation is quantified by English-name usage on this online networking site. The data allows us to observe this acculturational attribute because Renren users have the option to add an English name as the suffix following the original Chinese name. Another advantage of the data set is that close friendships are naturally defined as users have the option to nominate *special friends* on Renren by themselves. The specific empirical question in this paper is whether self usage the English name is positively correlated with close friends' English-name usage; in other words, whether international graduate students tend to have friends with similar characteristics of cultural assimilation.

It is difficult to test homophily in friendships because of the scarcity of the data as well as common statistical challenges in most network research. The reflection problem and the omitted variable problem might make the OLS estimate biased. To tackle these challenges, I exploit a natural experiment on English-name usage: all else being equal, a Chinese student with a difficult-to-pronounce Chinese name is more likely to use the English name, and further show the English name on Renren. In

Section 2 and 3 I find that the pronunciation difficulty is arguably randomized, and the exclusion restriction is intuitively convincing. With a strong statistical relationship between the pronunciation difficulty of the Chinese name and English-name usage, this natural experiment creates a valid instrumental variable for self usage of the English name on Renren.

The empirical findings in this paper can be summarized as follows. First, both the OLS and the IV regression show that a student with online English-name usage has more close friends also with online English-name usage. In other words, acculturational homophily does exist in friendships. The IV model yields greater estimate, and the OLS estimate might be downward biased. On average, a student showing the English name online has approximately one more close friend who also shows the English name online, conditional on that this student nominates at least one special friend. This magnitude of homophily is relatively large, considering that the average number of special friends in the data set is around 3.5.

This paper contributes to the existing literature by providing a new methodology for research on friendships. Based on linguistic properties of names, the natural experiment design solves the common endogeneity problem and verifies the existence of homophily based on acculturational characteristics in a causal manner. The instrumental variable is constructed in a simple way, yet the identification strategy provides an effective solution to major econometric challenges. That said, with a cross-sectional social networking data set, this paper has no statistical power to separately identify two major types of effects in causing acculturational homophily: peer influences and peer selection. New data sets with the panel structure might shed light on this identification question.

Appendix A: The Pronunciation Difficulty of Romanized Chinese Characters

The empirical analysis of this paper relies on the natural experiment on online English-name usage: all else being equal, a Chinese student whose original name is difficult to be pronounced by native speakers of English is more likely to use an English name. The pronunciation difficulty comes from the fact that the *pinyin* system, which romanizes Chinese characters, cannot well reflect pronunciation rules of Chinese due to phonological differences between Chinese and Western languages. The IV strategy based on the natural experiment design requires at least a dummy variable of the pronunciation difficulty. In this appendix, I introduce how this dummy variable is constructed.

To construct this variable, I introduce three types of Chinese characters that are hard to be pronounced. The first group of difficult-to-pronounce characters comes from the ambiguity between the alveolar nasal and the velar nasal. This involves all Chinese words with *pinyin* letters ending with *-ang* and *-eng*. For example, *ceng* is commonly pronounced as *cen* by native speakers of Western languages. However, I do not include any word with *pinyin* letters ending with *-ing* in this group since the pronunciation of *-ing* and *-in* are relatively similar.

The second group of difficult-to-pronounce characters comes from the uniqueness of Chinese characters. This involves all Chinese words with *pinyin* letters beginning with *x-*. These words have the unique pronunciation rule, which is not seen in English, although it is remotely similar with the pronunciation of *sch-* in German. But in general, *x-* words are hard to be pronounced by English speakers. For example, *xu* is commonly pronounced as *zu* by English speakers, which is substantially different from the standard way of pronunciation in Chinese.

The third group of difficult-to-pronounce characters comes from different rules of

pronouncing characters which are seen in both languages. This involves all Chinese words with *pinyin* letters ending with *-hi*, *-he*, and *-ue*. These linguistic elements are widely used in both English and Chinese, but pronunciation rules are different. For example, *shi* is commonly pronounced as *she* in English, which is distinct from the pronunciation in Chinese.

References

- [1] Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. “A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration.” *Journal of Political Economy*, 122(1), 467 - 506.
- [2] Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2015. “A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration.” manuscript.
- [3] Arai, Mahmood, and Peter Skogman Thoursie. 2009. “Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings.” *Journal of Labor Economics*, 27(1), 127 - 147.
- [4] Bassetti, Benedetta. 2007. “Effects of Hanyu Pinyin on Pronunciation in Learners of Chinese as a Foreign Language.” In *The Cognition, Learning and Teaching of Chinese Characters*, ed. Guder, A. and Jiang, X. and Wan, Y., 201 - 234. Beijing, China: Beijing Language and Culture University Press.
- [5] Belot, Michèle, and Sjef Ederveen. 2012. “Cultural Barriers in Migration between OECD Countries.” *Journal of Population Economics*, 25(3): 1077 - 1105.
- [6] Berman, Eli, Kevin Lang, and Erez Siniver. 2003. “Language-Skill Complementarity: Returns to Immigrant Language Acquisition.” *Labour Economics*, 10(3): 265 - 290.

- [7] Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review*, 94(4): 991 - 1013.
- [8] Burke, Mary A., and Tim R. Sass. 2013. "Classroom Peer Effects and Student Achievement." *Journal of Labor Economics*, 31(1), 51 - 82.
- [9] Chiswick, Barry R., and Paul W. Miller. 2003. "The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada." *Economics of Education Review*, 22(5): 469 - 480.
- [10] Chiswick, Barry R. 1991. "Speaking, Reading, and Earnings among Low-Skilled Immigrants." *Journal of Labor Economics*, 9(2): 149 - 170.
- [11] Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. 2009. "An Economic Model of Friendship: Homophily, Minorities, and Segregation." *Econometrica*, 77(4), 1003 - 1045.
- [12] Damm, Anna P. 2009. "Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence." *Journal of Labor Economics*, 27(2), 281 - 314.
- [13] Edwards, Rachel. 2006. "What's in a Name? Chinese Learners and the Practice of Adopting 'English' Names." *Language, Culture and Curriculum*, 19(1): 90 - 106.
- [14] Foster, Gigi. 2006. "It's Not Your Peers, and It's Not Your Friends: Some Progress toward Understanding the Educational Peer Effect Mechanism." *Journal of Public Economics*, 90, 8-9, 1455 - 1475.
- [15] Fryer, Roland G., and Stephen D. Levitt. 2004. "The Causes and Consequences of Distinctively Black Names." *Quarterly Journal of Economics*, 119(3): 767 - 805.

- [16] Gao, Fen. 2010. "What's Wrong with Current Chinese College English Assessment System? Reform or Not?" *International Education Studies*, 3(1): 34 - 37.
- [17] Gao Yihong, Cheng Ying, Zhao Yuan, and Zhou Yan. 2005. "Self-Identity Changes and English Learning among Chinese Undergraduates." *World Englishes*, 24(1): 39 - 51.
- [18] Gerhards, Jürgen, and Silke Hans. 2009. "From Hasan to Herbert: Name-Giving Patterns of Immigrant Parents between Acculturation and Ethnic Maintenance." *American Journal of Sociology*, 114(4), 1102 - 1128.
- [19] Girard, Yann, Florian Hett, and Daniel Schunk. *forthcoming*. "How Individual Characteristics Shape the Structure of Social Networks." *Journal of Economic Behavior & Organization*.
- [20] Gordon, Milton M. 1964. *Assimilation in American Life: The Role of Race, Religion, and National Origins*. New York: Oxford University Press.
- [21] Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper No. 7867.
- [22] Kandel, Denise B. 1978. "Homophily, Selection, and Socialization in Adolescent Friendships." *American Journal of Sociology*, 84(2), 427 - 436.
- [23] Kossoudji, Sherrie A. 1988. "English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men." *Journal of Labor Economics*, 6(2): 205 - 228.
- [24] Lin, Xu, and Bruce A. Weinberg. 2014. "Unrequited Friendship? How Reciprocity Mediates Adolescent Peer Effects." *Regional Science and Urban Economics*, 48: 144 - 153.

- [25] Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3), 531 - 542.
- [26] Marmaros, David, and Bruce Sacerdote. 2001. "Peer and Social Networks in Job Search." *European Economic Review*, 46(4-5), 870 - 879.
- [27] Marmaros, David, and Bruce Sacerdote. 2006. "How Do Friendships Form?" *Quarterly Journal of Economics*, 121(1), 79 - 119.
- [28] Mayer, Adalbert, and Steven L. Puller. 2008. "The Old Boy (and Girl) Network: Social Network Formation on University Campuses" *Journal of Public Economics*, 92(1-2), 329 - 347.
- [29] McPherson, J. Miller, and Lynn Smith-Lovin. 1987. "Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups." *American Sociological Review*, 52(3), 370 - 379.
- [30] Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics*, 118(2), 549 - 599.
- [31] Rubinstein, Yona, and Dror Brenner. 2014. "Pride and Prejudice: Using Ethnic-Sounding Names and Inter-Ethnic Marriages to Identify Labour Market Discrimination." *Review of Economic Studies*, 81(1): 389 - 425.
- [32] Schaefer, David R., Olga Kornienko, and Andrew M. Fox. 2011. "Misery Does Not Love Company: Network Selection Mechanisms and Depression Homophily." *American Sociological Review*, 76(5), 764 - 785.
- [33] Shifman, Limor, and Elihu Katz. 2005. "'Just Call Me Adonai': A Case Study of Ethnic Humor and Immigrant Assimilation." *American Sociological Review*, 70(5): 843 - 859.

- [34] Stinebrickner, Ralph, and Todd R. Stinebrickner. 2006. "What Can Be Learned About Peer Effects Using College Roommates? Evidence From New Survey Data and Students from Disadvantaged Backgrounds." *Journal of Public Economics*, 90, 8-9, 1435 - 1454.
- [35] Winters, Paul, Alain de Janvry, and Elisabeth Sadoulet. 2001. "Family and Community Networks in Mexico-U.S. Migration." *Journal of Human Resources*, 36(1), 159 - 184.
- [36] Zimmerman, David J. 2003. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment." *Review of Economics and Statistics*, 85(1), 9 - 23.