



Are Foreign IT Workers Paid Higher Than the Natives? A Replication Study Using Three US National Surveys

Gang Peng

Information Systems & Decision Sciences
College of Business and Economics
California State University Fullerton
gpeng@fullerton.edu

Rahul Bhaskar

Information Systems & Decision Sciences
College of Business and Economics
California State University Fullerton
rbhaskar@fullerton.edu

Abstract:

This study is a conceptual replication of Mithas and Lucas (2010) (hereafter ML). It investigates whether foreign information technology (IT) workers are paid higher than native IT workers. It replicates ML using three US national surveys: Current Population Survey, National Survey of College Graduates, and American Community Survey. While being able to obtain the same results as ML, this study shows that the estimated wage premium to foreign IT workers may differ across data sources used, predicting variables that are controlled, and the estimation methods applied. Further analysis using comparable subsamples reveals that sample size may also play a role in estimating the wage premium. This study enriches literature on wage differentials between foreign and native IT workers and deepens our understanding on the impact of foreign IT workers on the natives.

Keywords: Foreign IT workers, Native IT workers, Wage premium, Immigrants, Skills

The manuscript was received 3/21/2020 and was with the authors 6 months for 2 revisions.

1 Introduction

The impact of immigrants on the natives has been a significant and highly debated issue in many industrialized countries. In the US, a foreigner can immigrate to the US due to family ties or specific skills. In the 1950s, most of the immigrants to the US were from Europe, whereas during the 1990s and beyond the majority of immigrants came from Latin America and Asia (Gibson & Jung, 2006). Historically, immigrants to the US have been less educated, and many of the adult immigrants did not possess even a high school diploma (Friedberg, 2001). They primarily worked as cheap labor doing jobs natives would not like to do, such as agriculture and construction (Birgier, 2017; Orrenius & Zavodny, 2007). However, in recent decades, due to the acceleration of the knowledge economy and the fast pace of technological advancement, many skilled immigrants started to come into the US. Skilled immigrants or foreign workers are different from other types of immigrants—they carry significant human capital and play a critical role in today's fast-paced and knowledge-based economy (Kerr, 2013). Among all categories of skilled immigrants, probably the most prominent one is the IT immigrants or foreign IT workers. IT workers or IT professionals are non-executive personnel working in a firm's IT department (Wang & Kaarst-Brown, 2014). It is widely known that firms in the US have been hiring foreign IT workers through various programs to offset the shortage of IT workers in the country (Matloff, 2003).

With the inflow of foreign IT workers, both the general public and policy makers wonder how foreign IT workers have affected the native IT workers. Some researchers argue that the US is not short of skilled IT workers, and US employers prefer to hire foreign IT workers because they earn less than the natives and thus hiring foreign IT workers can help employers cut costs and increase profits (Hira, 2010; Matloff, 2004, 2013). Employers can potentially take advantage of foreign IT workers because it is difficult for them to change jobs, especially if they apply for permanent residency or the green card, with the help of their employers. Other researchers hold the opposite opinion and argue that foreign IT workers may earn a wage premium compared to the natives. In particular, Mithas & Lucas (2010) (hereafter ML) argued that such a wage premium can potentially arise from characteristics possessed by foreign IT workers and valued by their US employers, such as skills and expertise, travel flexibility and extended work hours, global perspectives and experience. ML used the salary survey data from *InformationWeek* for year 2000-2005 to explore the issue and found in one of their hypotheses, H1a, that foreign IT workers earn about 8.9% more than the native IT workers, and the sizable premium remains after a series of robustness tests are conducted.

Given that the finding from ML runs counter to the proposition that US firms hire foreign IT workers as cheap labor (Matloff, 2013), we intend to replicate the wage premium hypothesis (H1a) in ML to investigate if the wage premium for foreign IT workers is robust under different contexts. As commented by other researchers, replication studies can help generalize or extend the findings from original studies and advance knowledge accumulation in our field (Dennis, Brown, Wells, & Rai, 2020; Dennis & Valacich, 2014). In addition, such a replication can help deepen our understanding about wage compensation for IT workers and thus contribute to this important literature (Ang, Slaughter, & Ng, 2002; Matloff, 2003, 2013; Mithas & Lucas, 2010; Peng & Eunni, 2011; Whitaker, Mithas, & Liu, 2019). Therefore, in this study, we test the following hypothesis:

H1a: Everything else being equal, foreign IT workers are paid higher than native IT workers.

In H1a, we explicitly specify that, to estimate the wage premium to foreign IT workers, we need to isolate the wage differential due to workers' nationalities, and the impacts from other factors such as education, occupations, or work experience, should be partialled out as much as possible. We make use of three data sources to estimate the wage premium to foreign IT workers. We also conduct analysis by combining the three data sources together as well as using subsamples from the data sources. While being able to obtain the same results as ML, we find that estimation results may differ across data sources used, variables that are controlled, estimation methods applied, as well as the sample size of the analysis. We discuss possible causes of these issues and suggest future research directions for this important topic.

2 Data and Method

As defined by prior studies, conceptual replications “test exactly the same research questions or hypotheses, but use different measures, treatments, and/or analyses” (Dennis et al., 2020; Dennis & Valacich, 2014). ML examines whether foreign IT workers are paid higher than native IT workers. Specifically, they propose their H1a that “Compared to US citizens, non-US citizen IT professionals receive

a salary premium after controlling for their educational qualifications and work experience". In this study, we conduct a conceptual replication and test ML's H1a under three different contexts.

In literature, the phrase "foreign workers" is used in different ways. Some authors use it to strictly refer to non-US citizens who enter the US on temporary visas (Aobdia, Srivastava, & Wang, 2018; Mithas & Lucas, 2010), while others use it to mean workers not born in the US, so that even naturalized citizens are categorized as foreign workers (Hunt, 2011; Matloff, 2003). In this study, to be consistent with ML, we refer "foreign workers" as non-US citizens, which exclude foreign-born but naturalized US citizens. Most foreign workers enter the US on temporary visas. The primary method used is the H1B Visa, which allows employers within the US to temporarily hire foreign workers in specialty occupations that require theoretical and practical application of highly specialized knowledge in specific fields. The L1 visa is another non-immigrant visa available to employees of an international company with operations in the US. L1 visa allows intra-company transfers of foreign workers to a multinational corporation's US office if they have worked for the company for at least one year (Hira, 2010; Kerr & Lincoln, 2010). International students on F1 visa can also work in the US temporarily through Optional Practical Training (OPT). For more discussions on different types of visas and the immigration policies related to them, please refer to Hira (2010), Hunt (2011), and Mithas & Lucas (2010).

In this replication study, to test H1a, we make use of three US national surveys: Current Population Survey (CPS), National Survey of College Graduates (NSCG), and American Community Survey (ACS). The three surveys are conducted in different time spans. CPS is conducted monthly. ACS is released as a 1-year or 5-year survey and we make use of the 1-year or annual survey to be consistent with ML. The NSCG has been conducted in year 1993, 2003, 2010, 2013, 2015, and 2017. In their original study, ML used *InformationWeek* survey conducted from year 2000-2005. Thus, to be comparable to ML, we use the CPS and ACS surveys for 2000-2005 and the NSCG survey for 2003. We compare the time spans of the three surveys we use to ML in Table 1. Since wage premium to foreign IT workers is sensitive to visa policy and economic conditions over time, using data from the same time span as ML can help mitigate estimation discrepancies.

Study	Survey	Time Span
ML	InformationWeek	2000, 2001, 2002, 2003, 2004, 2005
This study	CPS	2000, 2001, 2002, 2003, 2004, 2005
	NSCG	2003
	ACS	2000, 2001, 2002, 2003, 2004, 2005

2.1 Current Population Survey (CPS)

CPS is a monthly survey of about 60,000 households conducted by the US Bureau of Labor Statistics. A household consists of all the people who occupy a housing unit, and it includes the related family members as well as unrelated people such as lodgers, foster children, wards, or those who share the housing unit.¹ The CPS has been conducted for more than 70 years, and it is the primary source of information on labor force characteristics of the US population.

The CPS surveys include basic monthly surveys as well as supplementary surveys on various topics. The basic monthly surveys provide a direct measurement of labor force, employment, wages, working hours, and demographic characteristics, etc. The supplementary surveys are used to gather in-depth information on specific aspects of the labor force, such as job tenure, occupation mobility, as well as computer and Internet use. CPS uses rotation groups: a given household is interviewed for four consecutive months, not interviewed for eight months, and then interviewed again for another four consecutive months, after which it leaves the sample permanently (Kostanich & Dipppo, 2002). CPS surveys have been used in many prior studies for IT workforce related issues (Burtch, Carnahan, & Greenwood, 2018; Levina & Xin, 2007; Peng & Eunni, 2011; Peng, Wang, & Han, 2018; Tambe & Hitt, 2012).

In this study, we make use of the CPS basic monthly surveys from 2000-2005.² We obtain the CPS data for each month and then aggregate them into annual data. We further extract immigration related data such as country of birth and citizenship for all correspondents, as well as data on employment status, work industry, occupation code, gender, age, education, work location, earning, and work hours, etc. Because rotation

¹ Please refer to <https://www.census.gov/programs-surveys/cps/technical-documentation/subject-definitions.html>.

² Please refer to http://www.nber.org/data/cps_basic.html.

groups are used in CPS, there are duplicate observations across two contiguous months. However, since the earning data are only available in the 4th and the 8th month, which are separated by an eight-month break in between, observations with earning data do not have any duplicates in a single year. The IT workers in these surveys—the sample used for this study—are identified by the CPS occupation codes specified by the US Census Bureau, as shown in Table 2. From these surveys, we can identify the IT workers, native or foreign, as well as their wages. These data are ideal for testing H1a.

Year	CPS Occupation Code	Occupation Title
2000-2002	64	Computer systems analysts and scientists
	229	Computer programmers
	308	Computer operators
	309	(Computer) peripheral equipment operators
2003-2005	110	Computer and information systems manager
	1000	Computer scientists and systems analysts
	1010	Computer programmers
	1020	Computer software engineers
	1040	Computer support specialists
	1060	Database administrators
	1100	Network and computer systems administrators
	1110	Network systems and data communications analysts
1400	Computer hardware engineers	

In CPS, workers are paid on different schedules such as hourly, monthly, or annually, but all wage data are recorded as weekly pay. To be consistent with ML, we convert weekly wages to annual wages.³ We use the log annual wage as the dependent variable. The key independent variable is the binary variable *noncitizen*, indicating whether an IT worker is a US citizen or not. In other words, same as ML, we define foreign workers as noncitizens. Indeed, most of the media reports of foreign IT workers usually refer to noncitizen IT workers. Consistent with prior literature, we also include control variables such as education, work experience, worker location, gender, union member, marriage status, and race (Mithas & Lucas, 2010; Wang & Kaarst-Brown, 2014). CPS records workers' highest education in education grades or levels. Similar to ML, we control for the following education levels: PhD, professional degree (such as MD), master's degree, bachelor's degree, some college (including associate degree), and no college education. We use the group of no college education as the base education level for analysis. From education levels, we calculate the years of education (Jaeger, 1997). Consistent with prior literature, we use (age minus years of education minus 6) to measure work experience (Krueger, 1993; Mincer, 1974; Mithas & Lucas, 2010). Because wage levels vary across geographic regions, we also control for states of the workers (Hunt, 2011; Mithas & Lucas, 2010). CPS has indicators for full-time or part-time workers. To make the sample more homogenous, we restrict the sample to full-time workers only. Variable definitions are provided in Table 3.

Variable	Definition
Ln(wage)	Log annual wage for an IT worker
Noncitizen	Binary variable for a noncitizen IT worker. It equals 1 if the IT worker is not a US citizen and 0 otherwise.
PhD	Dummy variable representing that the highest education degree the IT worker received is a PhD
Professional	Dummy variable representing that the highest education degree the IT worker received is a professional degree
Master	Dummy variable representing that the highest education degree the IT worker received is a master's degree
Bachelor	Dummy variable representing that the highest education degree the IT worker received is a bachelor's degree
Some college	Dummy variable representing that the highest education degree the IT worker received is some college education
Experience	Work experience of the IT worker; calculated as (age–years of education–6)

³ There are seven pay schedules in CPS: hourly, weekly, bi-weekly, twice monthly, monthly, annually, and others. Among them, only weekly wages are recorded for employees. Hourly wages and usual work hours per week are also recorded for hourly paid workers. We calculate annual wages by multiplying the weekly wages by 52.

Experience ²	The square of work experience, divided by 100 to rescale
Male	Binary variable for the gender of the IT worker. It equals 1 for male and 0 otherwise.
Married	Binary variable for the marriage status of the IT worker. It equals 1 once married and 0 otherwise.
Union	Binary variable for the union status of the IT worker; It equals 1 for a union member and 0 otherwise.
White	Binary variable for the race white. It equals 1 for white and 0 otherwise.
Pay schedule dummies	Dummy variables for the pay schedules. There are 7 in total.
Year dummies	Dummy variables for the years in which the surveys are conducted. There are 6 in total.
Occupation dummies	Dummy variables for the specific IT occupation titles that the IT workers hold. There are 4 occupations for 2000-2003 and 8 for 2003-2005.
Industry dummies	Dummy variables for the categories of major industries in which the IT workers are employed. There are 22 industries for 2000-2003 and 13 for 2003-2005.
State dummies	Dummy variables for the US states (or D.C.) in which the IT workers live. There are 51 in total.

2.2 National Survey of College Graduates (NSCG)

The NSCG is a national survey conducted by the National Science Foundation (NSF), an independent agency of the US government.⁴ The US Census Bureau collects and processes the survey data for NSF. NSCG and its prototypes have been conducted since the 1970s and they provide data on the college graduates in the US, with a particular focus on those in the science and engineering workforce. NSCG samples random individuals who live in the US and have at least a bachelor's degree. These individuals are identified by ACS, which we discuss later. Individuals in the survey are asked about their education, occupations, work activities, salary, academic degrees and degree areas, age, citizenship, country of birth, and the year in which their principal jobs started, etc. NSCG has been conducted multiple times over the decades. To be comparable with ML, we make use of the survey conducted in year 2003. We identify IT workers using the NSCG occupation codes as listed in Table 4.

NSCG Occupation Code	Occupation Title
110510	Computer and information scientists, research
110530	Computer support specialists
110540	Computer system analysts
110550	Database administrators
110560	Network and computer systems administrators
110570	Network systems and data communications analysts
110580	Other computer information science occupations
110880	Computer engineers—software
540870	Computer engineers—hardware
621420	Computer and information systems managers
640520	Computer programmers

The same as in the CPS, NSCG asks individuals if they are US citizens, and hence we can derive variable *noncitizen* for foreign IT workers. NSCG asks for annual salary for workers, and we take its log as the dependent variable. Same as CPS, NSCG identifies full-time or part-time employees, and we restrict the sample to full-time employees only. We control for three broad employment industries: educational institutions, government, and business/industry. Other independent variables include age, marriage status, gender, race, firm size, occupation and industry dummies. NSCG asks for the year in which the correspondents started their principal jobs, and we use the time difference between year 2003 and this variable as the measure of work experience. Instead of tracking the state in which an individual lives in, NSCG tracks regions of the US, such as New England, Middle Atlantic and Pacific; therefore, we control for these region dummies. Regarding education, NSCG focus on college graduates, and the correspondents all have at least a bachelor's degree. Therefore, four education levels are identified: PhD, professional, master's degree, and bachelor's degree. Correspondingly, we measure education as four dummy variables and use bachelor's degree as the base level. The variable definitions are provided in Table 5.

⁴ The dataset is available at <https://www.nsf.gov/statistics/srvygrads/>.

Variable	Definition
Ln(wage)	Log annual wage for an IT worker
Noncitizen	Binary variable for a noncitizen IT worker. It equals 1 if the IT worker is not a US citizen and 0 otherwise.
PhD	Dummy variable representing that the highest education degree the IT worker received is a PhD
Professional	Dummy variable representing that the highest education degree the IT worker received is a professional degree
Master	Dummy variable representing that the highest education degree the IT worker received is a master's degree
Age	The age of the IT worker
Experience	Year 2003 minus the year starting the principal job
Experience ²	The square of work experience, divided by 100 to rescale
Male	Binary variable for the gender of the IT worker. It equals 1 for male and 0 otherwise.
Married	Binary variable for the marriage status of the IT worker. It equals 1 once married and 0 otherwise.
White	Binary variable for the race white. It equals 1 for white and 0 otherwise.
Firm size dummies	Dummy variable for the sizes of the firms in which the IT workers are employed. There are 8 in total.
Occupation dummies	Dummy variables for the specific IT occupations that the IT workers hold. There are 11 in total.
Industry dummies	Dummy variables for the specific industries in which the IT workers are employed. There are 3 in total.
Region dummies	Dummy variables for the regions in which the IT workers live. There are 9 in total.

2.3 American Community Survey (ACS)

The ACS is conducted by the US Census Bureau and it is currently the largest household survey in the US.⁵ ACS offers broad and comprehensive information on social, economic, and housing data, and is designed to provide this information at many levels of geography, particularly for local communities. Compared to CPS, ACS is conducted annually rather than monthly, and has a much larger sample size.⁶ Again, to be comparable to ML, we use six annual ACS surveys from year 2000-2005, and further restrict the sample to full-time employees only.⁷

Same as in CPS and NSCG, we identify IT workers from the occupation codes, which are listed in Table 6. Although occupation codes changed in 2003, matching them before and after 2003 is straightforward. For example, in Table 6, the occupation code for computer programmers is 101 during 2000-2002 and is changed to 1010 during year 2003-2005.

ACS Occupation Code	Occupation Title
11/110	Computer and information systems managers
100/1000	Computer scientists and systems analysts
101/1010	Computer programmers
102/1020	Computer software engineers
104/1040	Computer support specialists
106/1060	Database administrators
110/1100	Network and computer systems administrators
111/1110	Network systems and data communications analysts
140/1400	Computer hardware engineers
<i>Note:</i> the first part of the code is for 2000-2002, and the second part is the corresponding code for 2003-2005.	

⁵ The datasets are available at <https://www.census.gov/programs-surveys/acs/data/pums.html>.

⁶ The sample size of ACS is relatively small in early years, around 800,000 households, but has increased to over 3.5 million in recent years.

⁷ ACS defines fulltime employee as those who are 16 years old or over and usually work 35 hours or more per week for 50 to 52 weeks per year: https://www.census.gov/topics/employment/labor-force/about/faq.html#par_textimage_735773790.

ACS asks for annual salaries, and to be consistent with ML, we use log annual wage as the dependent variable. ACS also asks if one is a US citizen or not, and thus allows us to generate variable *noncitizen*. Other independent variables such as education, age, work experience, gender, marriage status, and race are derived the same way as in CPS. We also control for occupation, industry, and state dummies. Variable definitions are provided in Table 7.

Variable	Definition
Ln(wage)	Log annual wage for an IT worker
Noncitizen	Binary variable for a noncitizen IT worker. It equals 1 if the IT worker is not a US citizen and 0 otherwise.
PhD	Dummy variable representing that the highest education degree the IT worker received is a PhD
Professional	Dummy variable representing that the highest education degree the IT worker received is a professional degree
Master	Dummy variable representing that the highest education degree the IT worker received is a master's degree
Bachelor	Dummy variable representing that the highest education degree the IT worker received is a bachelor's degree
Some college	Dummy variable representing that the highest education degree the IT worker received is some college education
Experience	Work experience of the IT worker; calculated as (age-education-6)
Experience ²	The square of work experience, divided by 100 to rescale
Male	Binary variable for the gender of the IT worker. It equals 1 for male and 0 otherwise.
Married	Binary variable for the marriage status of the IT worker. It equals 1 once married and 0 otherwise.
White	Binary variable for the race white. It equals 1 for white and 0 otherwise.
Year dummies	Dummy variables for the years in which the surveys are conducted. There are 6 in total.
Occupation dummies	Dummy variables for the specific IT occupation titles that the IT workers hold. There are 9 in total.
Industry dummies	Dummy variables for the specific industries in which the IT workers are employed. There are 17 in total.
State dummies	Dummy variables for the US states (or D.C.) in which the IT workers live. There are 51 in total.

2.4 Estimation Model

The same as in ML, we estimate the wage premium to foreign IT workers using the following ordinary least squares (OLS) model:

$$\ln(\text{wage}) = \alpha \cdot \text{noncitizen} + \mathbf{X}'\boldsymbol{\beta} + \varepsilon \quad (1)$$

where variable *wage* is annual wage, and *noncitizen* is a binary variable representing noncitizen IT workers. Vector \mathbf{X} represents other independent variables as defined for each data source. A positive and significant α would support H1a that foreign IT workers earn a wage premium over the natives.

3 Estimation Results

3.1 Results from CPS

As in ML, we first compare foreign (noncitizen) and native (citizen) IT workers regarding their key characteristics. To do this, following ML, we first use the employment cost index (ECI) to deflate wages across years to base year 1999.⁸ We then combine the 2000-2005 data together. Occupation and industry crosswalks are used across years. The characteristics of the data are shown in Table 8. All together there are 22,596 IT workers, among which 2,248 are noncitizens, about 10%. In comparison, ML has 51,363 IT workers, and 2,428 of them are noncitizens, about 4.73%.⁹ It shows that, without partialling out the impact

⁸ For each year, the December ECI is used: <https://www.bls.gov/bls/news-release/eci.htm>.

⁹ There are similarities and differences in the statistics between the three data sources and ML. For details, please refer to Table 23.

of other factors, the wage for foreign IT workers is higher than the natives, about 0.101 in log wage difference, or 10.6% higher in wage dollars ($e^{0.101} - 1 = 0.106$). This is slightly lower than in ML, which is 12.7%. Same as in ML, foreign workers tend to be younger, have more males, and have higher levels of education than the natives. Compared to ML, overall, employee in CPS have lower wages and less males, and tend to be less educated. Prior studies have shown that females and less educated workers tend to be paid less, and these can potentially explain why wage level is lower in CPS when compared to ML.

Attribute	All (N=22,596)	Noncitizen (N=2,248)	Citizen (N=20,348)
Log wage	10.724 (11.12)	10.815 (11.23)	10.714 (11.11)
PhD	0.012 (0.02)	0.036 (0.04)	0.010 (0.02)
Professional	0.006	0.017	0.005
Master	0.154 (0.15)	0.385 (0.30)	0.129 (0.14)
Bachelor	0.458 (0.45)	0.443 (0.43)	0.460 (0.45)
Some college	0.272 (0.15)	0.076 (0.06)	0.293 (0.15)
Age	38.96	33.80	39.53
Work experience	17.50 (19.29)	11.10 (13.43)	18.21 (19.58)
Male	0.709 (0.85)	0.772 (0.89)	0.702 (0.85)
Married	0.743	0.744	0.744
Union	0.044	0.022	0.047
White	0.802	0.339	0.853

Note: data in the parentheses are the corresponding statistics from ML.

The correlation matrix for the CPS data is presented in Table 9. It shows that log wage is positively correlated with noncitizen, higher levels of education, work experience, and male workers, but negatively with lower levels of education. These patterns are consistent with those in ML.

Variable	Mean (sd)	1	2	3	4	5	6	7	8	9	10	11	12
1. ln(wage)	10.724 (0.559)	—											
2. Noncitizen	0.100 (0.299)	0.054***	—										
3. PhD	0.012 (0.110)	0.062***	0.071***	—									
4. Professional	0.006 (0.078)	0.027***	0.048***	-0.009	—								
5. Master	0.154 (0.361)	0.174***	0.212***	-0.048***	-0.034***	—							
6. Bachelor	0.458 (0.498)	0.129***	-0.010	-0.102***	-0.072***	-0.393***	—						
7. Some college	0.272 (0.445)	-0.189***	-0.147***	-0.068***	-0.048***	-0.261***	-0.562***	—					
8. Experience	17.50 (10.34)	0.132***	-0.206***	0.006	-0.023***	-0.040***	-0.153***	0.107***	—				
9. Experience ²	4.133 (4.218)	0.079***	-0.185***	-0.006	-0.021***	-0.050***	-0.149***	0.105***	0.957***	—			
10. Male	0.709 (0.454)	0.171***	0.046***	0.042***	0.008	0.025***	0.048***	-0.031***	-0.073***	-0.063***	—		
11. Married	0.744 (0.437)	0.162***	0.001	0.027***	0.007	0.053***	-0.070***	0.017**	0.375***	0.297**	0.001	—	
12. Union	0.044 (0.206)	-0.047***	-0.037***	-0.012*	0.002	-0.020***	-0.035***	0.026***	0.093***	0.091***	-0.057***	0.011	—
13. White	0.802 (0.399)	-0.011	-0.386***	-0.041***	-0.028***	-0.152***	0.005	0.101***	0.141***	0.131***	0.046***	0.056***	-0.026***

Notes: N = 22,596; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The estimation results of Equation (1) using CPS 2000-2005 are shown in Table 10. Same as in ML, we report the heteroscedasticity-robust standard errors. Model 1 includes only the binary variable *noncitizen*, which represents a foreign IT worker. The coefficient is positive and significant ($b=0.101$, $p < 0.01$). This coefficient equals the wage premium when comparing the log wage levels between noncitizen and native IT workers without partialling out the impact of any other factors (see Table 8). However, this wage premium may arise from factors beyond worker nationality. So, in Model 2 to Model 4, we incrementally add more control variables.

The results from Model 2 and Model 3 are highly consistent with ML. For example, the coefficient on *noncitizen* is positive and highly significant ($p < 0.01$), and wage return to education is positive and significant, in the decreasing order from PhD, professional, master, bachelor, and some college degree. However, when the state dummies are entered in Model 4, the full model, the coefficient on *noncitizen* turns insignificant ($b = 0.19$, $p > 0.1$). This is one of the main differences between our findings and those from ML—the coefficient on *noncitizen* in ML remains positive and significant even after controlling for the state dummies in their robustness test of Table 8. To compare more closely with ML, the last model, Model 5, includes only the control variables used in Table 4 of ML.¹⁰ The coefficient on *noncitizen* is positive and highly significant ($b = 0.57$, $p < 0.01$). Therefore, Model 5 can successfully replicate ML's main result on H1a that foreign IT workers are paid higher than the natives.

Table 10. Estimation Result for CPS (2000-2005)

<i>Ln(wage)</i>	Model 1	Model 2	Model 3	Model 4	Model 5
Noncitizen	0.101*** (0.013)	0.064*** (0.013)	0.039*** (0.013)	0.019 (0.013)	0.057*** (0.013)
PhD		0.493*** (0.031)	0.433*** (0.031)	0.412*** (0.031)	0.580*** (0.032)
Professional		0.467*** (0.043)	0.415*** (0.042)	0.405*** (0.042)	0.527*** (0.044)
Master		0.457*** (0.015)	0.395*** (0.015)	0.381*** (0.015)	0.534*** (0.015)
Bachelor		0.347*** (0.013)	0.299*** (0.013)	0.293*** (0.013)	0.409*** (0.013)
Some college		0.110*** (0.013)	0.093*** (0.013)	0.092*** (0.013)	0.133*** (0.014)
Experience		0.032*** (0.001)	0.032*** (0.001)	0.030*** (0.001)	0.038*** (0.001)
Experience ²		-0.057*** (0.003)	-0.056*** (0.003)	-0.054*** (0.003)	-0.069*** (0.003)
Male		0.184*** (0.007)	0.166*** (0.007)	0.164*** (0.007)	0.198*** (0.007)
Married		0.068*** (0.008)	0.059*** (0.008)	0.070*** (0.008)	—
Union		-0.063*** (0.016)	-0.001 (0.017)	-0.027 (0.017)	—
White		0.038*** (0.009)	0.033*** (0.009)	0.056*** (0.009)	—
Control for year dummies?	No	Yes	Yes	Yes	Yes
Control for pay schedule dummies?	No	Yes	Yes	Yes	No
Control for occupation and industry dummies?	No	No	Yes	Yes	No
Control for state dummies?	No	No	No	Yes	No
R ²	0.003	0.225	0.255	0.274	0.177
<i>Notes:</i> N=22,596; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; heteroscedasticity-robust standard errors are reported in parentheses.					

However, we decide to use results from Model 4 to test H1a. There are two main reasons: 1) We believe controlling for occupation and state dummies are necessary and important for estimating the wage premium to foreign IT workers.¹¹ Without controlling for occupation dummies, we may end up comparing the wage of a foreign programmer to that of a native web master. Similar, without controlling for state dummies, we may compare the wage of a foreign programmer living in Silicon Valley to that of a native programmer living in

¹⁰ ML present their main results in Table 4 to 7, and results of robustness tests in Table 8. Table 4 in ML has other independent variables such as *firm size* and *dot-com firm* which are not available from CPS, and we do not control for them in our study.

¹¹ ML do not control for occupation dummies in their main results because they are potential “bad controls”. However, we feel the context here is different from the example of Angrist & Pischke (2009), who estimate returns to education without controlling for occupation dummies—the connection between immigration status and occupation is not as strong, and thus the omitted variable bias is more of a concern here if occupation dummies are not controlled for.

the Midwest. 2) ML controlled for occupation and state dummies in their robustness tests, and the wage premium remains positive and highly significant. Thus, by controlling for occupation and state dummies, our study can more faithfully replicate ML. The result from Model 4 does not support H1a.

As in ML, we also estimate the wage premium to *noncitizen* using the propensity score matching (PSM) approach.¹² This approach compares wage level of noncitizens with that of the natives of comparable observed characteristics, i.e., all other independent variables excluding *noncitizen*. Conceptually, it tries to isolate the wage premium of foreign IT workers due to nationality by partialling out the impacts of all other factors. The estimated wage premium using variables in Model 4 is 0.002 ($p>0.1$), which does not support H1a either.

Next, we estimate Equation (1) for six years separately. The estimation results are shown in Table 11. For brevity, we only show the results from Model 4, the full model. It shows that during the six years from 2000-2005, the coefficient on *noncitizen* is only marginally significant in year 2001 ($b=0.055$, $p<0.1$).

Table 11. Estimation Result for Separate Years, CPS (2000-2005)						
<i>Ln(wage)</i>	2000	2001	2002	2003	2004	2005
Noncitizen	0.025 (0.028)	0.055* (0.030)	0.013 (0.028)	0.032 (0.046)	-0.024 (0.027)	0.024 (0.027)
PhD	0.514*** (0.075)	0.418*** (0.075)	0.460*** (0.080)	0.399*** (0.080)	0.356*** (0.073)	0.342*** (0.066)
Professional	0.426** (0.171)	0.384*** (0.113)	0.521*** (0.120)	0.347*** (0.078)	0.188*** (0.069)	0.541*** (0.100)
Master	0.425*** (0.036)	0.331*** (0.042)	0.400*** (0.034)	0.342*** (0.042)	0.320*** (0.032)	0.383*** (0.030)
Bachelor	0.321*** (0.031)	0.276*** (0.036)	0.295*** (0.031)	0.296*** (0.033)	0.212*** (0.029)	0.267*** (0.029)
Some college	0.113*** (0.033)	0.059 (0.040)	0.089*** (0.030)	0.085*** (0.031)	0.018 (0.029)	0.112*** (0.029)
Experience	0.032*** (0.004)	0.032*** (0.002)	0.029*** (0.003)	0.032*** (0.003)	0.030*** (0.003)	0.029*** (0.003)
Experience ²	-0.059*** (0.010)	-0.061*** (0.007)	-0.051*** (0.006)	-0.055*** (0.007)	-0.049*** (0.007)	-0.049*** (0.007)
Male	0.124*** (0.018)	0.159*** (0.018)	0.155*** (0.017)	0.181*** (0.021)	0.179*** (0.017)	0.140*** (0.016)
Married	0.049** (0.019)	0.057*** (0.021)	0.070*** (0.021)	0.066*** (0.022)	0.078*** (0.019)	0.086*** (0.018)
Union	-0.056 (0.053)	-0.014 (0.039)	0.044 (0.039)	-0.027 (0.046)	-0.015 (0.036)	0.003 (0.031)
White	0.035 (0.024)	0.032 (0.033)	0.033 (0.039)	0.098*** (0.029)	0.053*** (0.019)	0.067*** (0.022)
R ²	0.363	0.334	0.353	0.220	0.317	0.258
N	3,174	3,295	3,328	4,245	4,211	4,343

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$; heteroscedasticity-robust standard errors are reported in parentheses; other control variables include dummy variables for pay schedules, occupations, industries, and states.

3.2 Results from NSCG

Similar to CPS, we first compare foreign (noncitizen) and native (citizen) IT workers from NSCG 2003 in Table 12. There are 10,477 IT workers in total, and among them 1,503 are noncitizens, about 14.3%. Without partialling out the impact of other factors, foreign IT workers earn slightly more than the natives, about 0.036 in log wage difference, which is 3.7% higher in dollar amount ($e^{0.036} - 1 = 0.037$). Same as in CPS, compared to the natives, foreign IT workers in NSCG tend to be younger, have more males and less white, and have higher levels of education. Compared to ML, employees in NSCG have higher wage levels, higher education, and less males. Although workers in NSCG have a higher wage level for both foreign and native IT workers, the wage differential is actually smaller than in CPS.

¹² This is done using the *teffects psmatch* command in *Stata* 14 with *logit* model for ATE (the default options).

Attribute	All (10,477)	Noncitizen (1,503)	Citizen (8,974)
Log wage	11.182 (11.12)	11.212 (11.23)	11.176 (11.11)
PhD	0.041 (0.02)	0.089 (0.04)	0.033 (0.02)
Professional	0.003	0.003	0.003
Master	0.320 (0.15)	0.520 (0.30)	0.287 (0.14)
Bachelor	0.636 (0.45)	0.388 (0.43)	0.677 (0.45)
Age	41.56	36.19	42.46
Work experience	18.656 (19.29)	12.60 (13.43)	19.67 (19.58)
Male	0.733 (0.85)	0.787 (0.89)	0.724 (0.85)
Married	0.751	0.850	0.735
White	0.692	0.259	0.764

Note: data in the parentheses are corresponding statistics from ML.

The correlation matrix is presented in Table 13. It shows that log wage is positively correlated with noncitizen, higher levels of education (but not professional degree), work experience, and male workers. These patterns are consistent with those in CPS and/or ML.

Variable	Mean (sd)	1	2	3	4	5	6	7	8	9	10
1. ln(wage)	11.182 (0.558)	—									
2. Noncitizen	0.143 (0.351)	0.022**	—								
3. PhD	0.041 (0.198)	0.069***	0.099***	—							
4. Professional	0.003 (0.058)	-0.005	-0.005	-0.012	—						
5. Master	0.320 (0.467)	0.083***	0.176***	-0.142***	-0.040***	—					
6. Age	41.55 (8.960)	0.084***	-0.245***	0.079***	0.042***	0.062***	—				
7. Experience	6.258 (6.441)	0.077***	-0.159***	-0.019*	0.023**	-0.056***	0.378***	—			
8. Experience ²	0.806 (1.643)	0.076***	-0.138***	-0.008	0.022**	-0.044***	0.371***	0.931***	—		
9. Male	0.733 (0.442)	0.116***	0.051***	0.055***	0.013	-0.005	-0.004	-0.022**	-0.007	—	
10. Married	0.751 (0.432)	0.086***	0.094***	0.036***	-0.020***	0.067***	0.099***	0.017*	0.015	0.136***	—
11. White	0.691 (0.462)	-0.003	-0.383***	-0.066***	0.010	-0.171***	0.173***	0.130***	0.113***	0.039***	-0.044***

Notes: N = 10,477; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The estimation results for Equation (1) using NSCG 2003 data are shown in Table 14. The same models as in CPS are used there. Model 1 only controls for *noncitizen*, and the coefficient is marginally significant ($b=0.036$, $p < 0.1$). When more control variables are incrementally added in Model 2 through Model 4, the coefficient on *noncitizen* turns insignificant in all of them ($p > 0.1$). Finally, Model 5 controls for only the predictors from ML, and the coefficient on *noncitizen* remains insignificant ($b=0.27$, $p > 0.1$). Therefore, results from NSCG do not support H1a, not even from Model 5, which has less controls.

The same as for the CPS survey, we estimate the wage premium to *noncitizen* using PSM. The estimated wage premium using Model 4 is 0.078 ($p < 0.05$), which supports H1a. Thus, for the NSCG survey, the conclusions from OLS and PSM are not the same.

NSCG also identifies if a noncitizen holds a temporary work visa (for temporary worker) or a green card (for permanent resident). We create two dummy variables, *work visa* and *green card*, for noncitizens who hold either a work visa or a green card, and then estimate wage returns to them. The results are shown in Table 15. The coefficients on them are insignificant ($p > 0.1$) in all three models, and thus no evidence is found to

support the argument that foreign workers holding a temporary work visa or a green card earn a wage premium than the natives.¹³

<i>Ln(wage)</i>	Model 1	Model 2	Model 3	Model 4	Model 5
Noncitizen	0.036* (0.021)	0.022 (0.020)	-0.004 (0.019)	-0.008 (0.019)	0.027 (0.020)
PhD		0.192*** (0.022)	0.172*** (0.023)	0.161*** (0.022)	0.192*** (0.022)
Professional		-0.043 (0.062)	0.010 (0.061)	0.020 (0.061)	-0.056 (0.063)
Master		0.102*** (0.012)	0.074*** (0.012)	0.071*** (0.012)	0.103*** (0.012)
Age		0.003*** (0.0001)	0.004*** (0.0001)	0.004*** (0.001)	0.003*** (0.001)
Experience		0.003 (0.002)	0.004 (0.002)	0.004 (0.002)	0.003 (0.002)
Experience ²		0.006 (0.008)	0.004 (0.008)	0.004 (0.002)	0.001 (0.001)
Male		0.143*** (0.012)	0.117*** (0.011)	0.117*** (0.011)	0.153*** (0.012)
Married		0.070*** (0.012)	0.052*** (0.012)	0.058*** (0.012)	—
White		0.011 (0.011)	0.015 (0.011)	0.033** (0.011)	—
Control for firm size dummies?	No	Yes	Yes	Yes	Yes
Control for occupation and industry dummies?	No	No	Yes	Yes	No
Control for region dummies?	No	No	No	Yes	No
R ²	0.001	0.055	0.118	0.129	0.052

Notes: N=10,477; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; base education category is bachelor's degree.

<i>Ln(wage)</i>	Model 1	Model 2	Model 3
Work visa	-0.032 (0.035)	—	-0.030 (0.034)
Green card	—	0.014 (0.024)	0.010 (0.024)
PhD	0.161*** (0.022)	0.158*** (0.022)	0.159*** (0.022)
Professional	0.021 (0.061)	0.019 (0.061)	0.020 (0.061)
Master	0.071*** (0.012)	0.069*** (0.012)	0.070*** (0.012)
Age	0.003*** (0.0001)	0.004*** (0.0001)	0.004*** (0.001)
Experience	0.004* (0.002)	0.004* (0.002)	0.004* (0.002)
Experience ²	0.005 (0.008)	0.004 (0.008)	0.005 (0.008)
Male	0.118*** (0.011)	0.117*** (0.011)	0.118*** (0.011)
Married	0.058*** (0.012)	0.057*** (0.012)	0.057*** (0.012)

¹³ Although not the focus of this study, Table 15 corresponds to Table 5 in ML, which is used to test their H1b and H1c that foreign IT workers holding a work visa or a green card command a wage premium than the natives.

White	0.032*** (0.012)	0.036*** (0.011)	0.034*** (0.011)
R ²	0.129	0.129	0.129
Notes: N=10,477; ***p<0.01, **p<0.05, *p<0.1; base education category is bachelor's degree; other control variables include dummy variables for firm sizes, occupations, industries, and regions.			

3.3 Results from ACS

As in the CPS data, we first use the ECI to deflate wages across years to base year 1999 and then combine the 2000-2005 ACS data together. Similar to CPS, occupation and industry crosswalks are used.¹⁴ The comparison between foreign (noncitizen) and native (citizen) IT workers are shown in Table 16, and the correlation matrix in Table 17.

Attribute	All (N=76,500)	Noncitizen (N=6,762)	Native (69,738)
Log wage	10.862 (11.12)	10.955 (11.23)	10.852 (11.11)
PhD	0.016 (0.02)	0.049 (0.04)	0.013 (0.02)
Professional degree	0.006	0.016	0.005
Master's degree	0.170 (0.15)	0.408 (0.30)	0.147 (0.14)
Bachelor's degree	0.445 (0.45)	0.407 (0.43)	0.449 (0.45)
Some college	0.295 (0.15)	0.092 (0.06)	0.314 (0.15)
Age	40.17	34.78	40.69
Work experience	21.76 (19.29)	15.49 (13.43)	22.36 (19.58)
Male	0.725 (0.85)	0.799 (0.89)	0.718 (0.85)
Married	0.681	0.748	0.674
White	0.802	0.306	0.850
Note: the data in the parentheses are corresponding statistics from ML.			

Variables	Mean (sd)	1	2	3	4	5	6	7	8	9	10	11
1. ln(wage)	10.862 (0.511)	—										
2. Noncitizen	0.088 (0.284)	0.057***	—									
3. PhD	0.016 (0.127)	0.088***	0.080***	—								
4. Professional	0.006 (0.077)	0.022***	0.040***	-0.010***	—							
5. Master	0.170 (0.376)	0.198***	0.198***	-0.058***	-0.035***	—						
6. Bachelor	0.445 (0.497)	0.081***	-0.024***	-0.115***	-0.070***	-0.405***	—					
7. Some college	0.295 (0.456)	-0.202***	-0.138***	-0.083***	-0.050***	-0.293***	-0.579***	—				
8. Experience	21.76 (4.904)	0.175***	-0.192***	0.019***	-0.003	-0.010***	-0.165***	0.119***	—			
9. Experience ²	5.761 (4.904)	0.127***	-0.179***	0.014***	-0.002	-0.020***	-0.157***	0.115***	0.969***	—		
10. Male	0.725 (0.446)	0.149***	0.051***	0.034***	0.009***	0.016***	0.022***	-0.016***	-0.063***	-0.058***	—	
11. Married	0.681 (0.466)	0.176***	0.045***	0.036***	0.009***	0.069***	-0.029***	-0.019***	0.174***	0.130***	0.112***	—
12. White	0.802 (0.398)	-0.002	-0.388***	-0.048***	-0.019***	-0.143***	0.016***	0.097***	0.148***	0.140***	0.036***	0.014***
Notes: N = 76,500; ***p<0.01, **p<0.05, *p<0.1.												

There are 76,500 IT workers in total, and among them 6,762 are noncitizens, about 8.8%. Without partialling out the impact of other factors, foreign IT workers earn more than the natives, about 0.103 in log wage

¹⁴ The occupation and industry codes in ACS also changed starting January 2003, by adding a "0" to the previous codes. We convert the 2000-2002 codes to 2003-2005 codes before combining the data across years.

difference, or 11.2% more in dollar amount ($e^{0.103} - 1 = 0.108$). This is higher than in CPS (10.6%) and NSCG (3.7%), but lower than in ML (12.7%). Same as in ML, CPS, and NSCG, compared to the natives, foreign workers in ACS tend to be younger, have more males and less white, and have higher levels of education. Compared to ML, employees in ACS have less males and lower levels of education. These patterns are similar to CPS surveys. In addition, the correlation pattern is almost identical to CPS and/or NSCG.

Next, we present the estimation results in Table 18. The five models are constructed in the same way as in the CPS models. The results show that the coefficient of *noncitizen* is highly significant ($p < 0.01$) across all five models, supporting H1a.

<i>Ln(wage)</i>	Model 1	Model 2	Model 3	Model 4	Model 5
Noncitizen	0.102*** (0.006)	0.067*** (0.007)	0.030*** (0.007)	0.019*** (0.007)	0.067*** (0.006)
PhD		0.562*** (0.016)	0.508*** (0.016)	0.482*** (0.016)	0.572*** (0.017)
Professional		0.410*** (0.026)	0.381*** (0.025)	0.355*** (0.025)	0.418*** (0.026)
Master		0.497*** (0.008)	0.428*** (0.008)	0.410*** (0.008)	0.495*** (0.008)
Bachelor		0.349*** (0.007)	0.307*** (0.007)	0.300*** (0.007)	0.355*** (0.007)
Some college		0.105*** (0.007)	0.100*** (0.007)	0.099*** (0.007)	0.108*** (0.007)
Experience		0.041*** (0.001)	0.040*** (0.001)	0.039*** (0.001)	0.045*** (0.001)
Experience ²		-0.064*** (0.002)	-0.061*** (0.002)	-0.061*** (0.002)	-0.070*** (0.002)
Male		0.157*** (0.004)	0.139*** (0.004)	0.137*** (0.004)	0.169*** (0.004)
Married		0.086*** (0.004)	0.075*** (0.004)	0.084*** (0.004)	—
White		0.020*** (0.005)	0.014*** (0.004)	0.041*** (0.004)	—
Control for year dummies?	No	Yes	Yes	Yes	Yes
Control for occupation and industry dummies?	No	No	Yes	Yes	No
Control for state dummies?	No	No	No	Yes	No
R ²	0.003	0.195	0.257	0.281	0.190
Notes: N=76,500; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; heteroscedasticity-robust standard errors are reported in parentheses.					

We then estimate the wage premium to *noncitizen* using PSM. Using variables in Model 4, the estimated wage premium is 0.041 ($p < 0.05$), also supporting H1a.

Same as with CPS, we estimate Equation (1) for six years separately. The results are shown in Table 19. The coefficient on *noncitizen* is only marginally significant for year 2001 ($b = 0.027$, $p < 0.1$).

As an additional test, we combine the CPS 2000-2005, NSCG 2003, and ACS 2000-2005 data together and run an integrated test. The total sample size for this analysis is 109,573. Not all three of the surveys have the same set of control variables. Specifically, CPS and ACS do not have *firm size* and NSCG does not have groups of individuals receiving *some college education or below*. Therefore, we only control for the common variables. The results are shown in Table 20.

<i>Ln(wage)</i>	2000	2001	2002	2003	2004	2005
Noncitizen	0.008 (0.033)	0.027* (0.016)	0.022 (0.018)	0.023 (0.017)	-0.001 (0.018)	0.012 (0.011)
PhD	0.335*** (0.083)	0.471*** (0.042)	0.492*** (0.039)	0.460*** (0.039)	0.497*** (0.041)	0.456*** (0.027)
Professional	0.378*** (0.086)	0.289*** (0.063)	0.530*** (0.065)	0.279*** (0.057)	0.278*** (0.059)	0.345*** (0.045)
Master	0.413*** (0.032)	0.392*** (0.019)	0.429*** (0.021)	0.338*** (0.019)	0.412*** (0.021)	0.389*** (0.013)
Bachelor	0.304*** (0.029)	0.305*** (0.016)	0.353*** (0.018)	0.239*** (0.017)	0.282*** (0.018)	0.280*** (0.012)
Some College	0.079** (0.029)	0.094*** (0.016)	0.140*** (0.018)	0.079*** (0.016)	0.074*** (0.019)	0.099*** (0.012)
Experience	0.032*** (0.003)	0.036*** (0.002)	0.040*** (0.002)	0.038*** (0.002)	0.042*** (0.002)	0.040*** (0.001)
Experience ²	-0.045*** (0.006)	-0.056*** (0.004)	-0.065*** (0.004)	-0.059*** (0.004)	-0.068*** (0.004)	-0.062*** (0.003)
Male	0.128*** (0.016)	0.142*** (0.008)	0.143*** (0.009)	0.147*** (0.009)	0.144*** (0.010)	0.128*** (0.006)
Married	0.077*** (0.016)	0.086*** (0.009)	0.080*** (0.009)	0.078*** (0.009)	0.095*** (0.010)	0.090*** (0.006)
White	-0.026 (0.020)	0.048*** (0.011)	0.041*** (0.012)	0.027** (0.011)	0.056*** (0.012)	0.064*** (0.007)
R ²	0.318	0.311	0.314	0.300	0.304	0.293
N	3,770	12,145	10,598	11,540	11,460	26,987

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; other control variables include dummy variables for occupations, industries, and states.

<i>Ln(wage)</i>	Model 1	Model 2	Model 3	Model 4	Model 5
Noncitizen	0.108*** (0.006)	0.073*** (0.006)	0.030*** (0.006)	0.019** (0.006)	0.054*** (0.006)
PhD		0.310*** (0.012)	0.277*** (0.012)	0.254*** (0.011)	0.377*** (0.012)
Professional		0.178*** (0.021)	0.169*** (0.020)	0.151*** (0.020)	0.170*** (0.021)
Master		0.236*** (0.004)	0.198** (0.004)	0.184*** (0.004)	0.274*** (0.004)
Experience		0.031*** (0.001)	0.031*** (0.001)	0.030*** (0.001)	0.020*** (0.001)
Experience ²		-0.052*** (0.001)	-0.050*** (0.001)	-0.050*** (0.001)	-0.030*** (0.001)
Male		0.172*** (0.003)	0.152*** (0.003)	0.150*** (0.003)	0.179*** (0.003)
Married		0.096*** (0.003)	0.082*** (0.003)	0.093*** (0.003)	—
White		0.019*** (0.004)	0.016*** (0.004)	0.046*** (0.004)	—
Control for year dummies?	No	Yes	Yes	Yes	Yes
Control for data source dummies?	No	Yes	Yes	Yes	No
Control for occupation and industry dummies?	No	No	Yes	Yes	No
Control for state and region dummies?	No	No	No	Yes	No
R ²	0.004	0.159	0.223	0.248	0.101

Notes: N= 109,573 for all other models; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; base education category is bachelor's degree or below.

The coefficient on *noncitizen* is positive and highly significant ($p < 0.01$) from Model 1 to Model 3, but the significance level decreases slightly in Model 4 ($b = 0.19$, $p < 0.05$) when the state dummies are added. In Model 5, with less control variables, the result is highly significant ($b = 0.054$, $p < 0.01$). Overall, the result from the combined data support H1a.

Same as before, we estimate the wage premium to *noncitizen* using PSM. The estimated wage premium based on Model 4 is 0.031 ($p < 0.05$), also supporting H1a.

Using this combined dataset, we then estimate Equation (1) for the six years separately.¹⁵ The results from Table 21 show that the coefficient on *noncitizen* is significant in 2001, 2002, and 2005 only ($p < 0.05$).

<i>Ln(wage)</i>	2000	2001	2002	2003	2004	2005
Noncitizen	0.023 (0.023)	0.039*** (0.015)	0.032** (0.016)	0.004 (0.013)	-0.002 (0.015)	0.020** (0.010)
PhD	0.213*** (0.058)	0.279*** (0.035)	0.248*** (0.032)	0.228*** (0.019)	0.289*** (0.033)	0.242*** (0.022)
Professional	0.196** (0.084)	0.125** (0.055)	0.281*** (0.056)	0.100*** (0.038)	0.079* (0.045)	0.175*** (0.040)
Master	0.225*** (0.015)	0.194*** (0.011)	0.201*** (0.011)	0.131*** (0.008)	0.220*** (0.010)	0.184*** (0.007)
Experience	0.029*** (0.002)	0.033*** (0.002)	0.035*** (0.002)	0.021*** (0.001)	0.036*** (0.002)	0.035*** (0.001)
Experience ²	-0.050*** (0.006)	-0.056*** (0.003)	-0.061*** (0.004)	-0.031*** (0.002)	-0.060*** (0.003)	-0.057*** (0.002)
Male	0.147*** (0.012)	0.168*** (0.008)	0.167*** (0.009)	0.140*** (0.007)	0.161*** (0.009)	0.134*** (0.006)
Married	0.089*** (0.013)	0.095*** (0.008)	0.091*** (0.009)	0.085*** (0.007)	0.100*** (0.009)	0.095*** (0.006)
White	0.050*** (0.016)	0.052*** (0.010)	0.043*** (0.011)	0.035** (0.008)	0.055*** (0.010)	0.062*** (0.007)
R ²	0.241	0.240	0.239	0.250	0.268	0.256
N	6,944	15,440	13,926	26,262	15,671	31,330

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; other control variables include dummy variables for occupations, industries, states and regions.

In sum, we have replicated H1a using three different surveys, CPS (2000-2005), NSCG (2003), and ACS (2000-2005), as well as the combined dataset. We have reported the corresponding results in Table 10, 14, 18, and 20. We are able to successfully replicate the main findings from ML based on Model 5 from Table 10, 18, and 20—results from CPS, ACS, and the combined dataset support H1a that foreign IT workers earn a wage premium over the natives. However, results from Model 4, the full model, are quite different from ML. We summarize these findings in Table 22.

Data Source (sample size)	Estimation Approaches	Results of Testing H1a
CPS (22,596)	Six years combined	Not supported ($p > 0.1$)
	Separate years	Weakly supported ($p < 0.1$) only in 2001
	PSM	Not supported ($p > 0.1$)
NSCG (10,477)	One year	Not supported ($p > 0.1$)
	PSM	Supported ($p < 0.05$)
ACS (76,500)	Six years combined	Supported ($p < 0.01$)
	Separate years	Weakly supported ($p < 0.1$) only in 2001
	PSM	Supported ($p < 0.05$)
The combined dataset (109,573)	Six years combined	Supported ($p < 0.05$)
	Separate years	Supported ($p < 0.05$) in three of six years
	PSM	Supported ($p < 0.05$)

¹⁵ Year 2003 has data from all three surveys while other years have data from CPS and ACS surveys only.

First, the support of H1a from CPS is rather weak. All three tests in Table 22 using the three approaches are not quite supportive. Support from NSCG is split, with one result being supportive and the other one not. ACS results are very supportive, although H1a is weakly supported only in one of the six years and not supported in all other years. Support from the combined dataset improves further, with H1a is supported ($p < 0.05$) now in three of the six years, 2001, 2002, and 2005.

We would like to conclude this section with three observations: 1) There are discernable estimation differences across the three data sources. We discuss possible causes of these differences in the next section. 2) We also observe that estimation results differ depending on the variables that are controlled. It seems our results are very sensitive to whether or not the state dummies are controlled. This is different from ML where the wage premium is very robust to the state dummies. 3) OLS and PSM do not always give the same conclusion. For example, result from Model 4 in Table 14 does not support H1a ($p > 0.1$), but result using PSM from the same data does ($p < 0.05$).

4 Discussion

There are several other studies that have used NSCG 2003 to investigate the wage differentials across various worker groups. However, they have not examined the wage premium to foreign IT workers exactly. For example, Matloff (2013) tried to estimate the wage differential between foreign-born and native IT workers and found that foreign-born IT workers are paid significantly lower than the natives. However, rather than using IT professionals as the sample, his study actually used workers holding an IT degree. The problem is that IT degree holders do not necessarily work as IT professionals. Furthermore, our analysis shows that, when not working as IT professionals, noncitizens typically earn less than the natives. Thus, using IT degree holders as the sample introduces downward bias in the estimation.

Drago (2015) studied the wage premium to H1B visa workers using the NSCG 2003 too. Different from our study, the sample he used includes workers of all occupations, not just IT workers. He found that while foreign workers in general earn less than the natives, workers in the technology industries, including both foreign and native workers, earn a premium than those in other industries. However, the study did not show if foreign workers in technology industries earn more or less than the natives. Indeed, the negative effect of being foreign workers and the positive effect of working in technology industries may well cancel out, and thus, the net impact of being foreign workers in the technology industries is inconclusive in his study.

Hunt (2011) also used NSCG 2003 to study the wage premium between foreign and native workers. She found that without any control variables, foreign workers earn a 2.9% wage premium than the natives. However, when more control variables are added, foreign workers earn as much as 8.1% less than the natives. She indeed found that immigrants from European countries earn higher wages than the natives, about 11.8%. Thus, it can be inferred that foreign workers from other countries, in aggregate, actually earn much less than the natives. The same as Drago (2015), Hunt (2011) did not examine the wage premium to foreign IT workers, but to all types of foreign workers.

In comparison, our study focuses on the wage premium earned by foreign IT workers. Results from NSCG 2003 data show that while the PSM result supports the wage premium hypothesis, the OLS result does not.

Three US national surveys are used in this replication study. One might wonder how well they corroborate with each other and with ML? To this purpose, we compare the characteristics of the three surveys with those of ML in Table 23.

Among the four data sources, NSCG stands out in two aspects: 1) The education levels are much higher than others. This is not surprising since correspondents in NSCG all have at least a bachelor's degree. 2) The raw wage premium to foreign IT workers from it (3.7%) is much lower than others. CPS and ACS are highly consistent on all characteristics. Compared with the others, ML has a much lower premium (4.73%), about half of others. ML used surveys conducted by *InformationWeek* magazine which had a distribution of around 400,000 back in early 2000s. There is reason to believe that many subscribers worked for large firms such as the *InformationWeek 500* firms. If this is the case, a sample of IT employees of large firms (such as *InformationWeek 500* firms) could certainly have different compensation packages compared with a sample of IT employees drawn from all firms.¹⁶

¹⁶ We thank one of the reviewers for suggesting this possible reason for sample differences between ML and others.

Characteristics	ML	CPS	NSCG	ACS
Sample size	51,363	22,596	10,477	76,500
Log wage for all IT workers	11.12	10.72	11.18	10.86
% of foreign IT workers	4.73	9.95	14.35	8.84
Wage premium to foreign IT workers (in %)	12.7%	10.6%	3.7%	10.8%
% of PhD degree	2	1.2	4.1	1.6
% of professional degree	—	0.6	0.3	0.6
% of master's degree	15	15.4	32	17.0
% of bachelor's degree	45	45.8	63.6	44.5
% of some college	15	27.2	—	29.5
Age	—	38.95	41.56	40.17
Work experience	19.29	17.50	18.656	21.76
% of males	85	70.7	73.3	72.5
% of married	—	74.3	75.1	68.1
% of white	—	80.2	69.2	80.2
<i>Notes:</i> NSCG is for 2003, and all other data sources are for 2000-2005; the wage premium is calculated in 1999 dollar and without any control variables.				

Besides the above differences in data characteristics, we also would like to add that there are some important differences in the design of CPS, NSCG, and ACS: 1) CPS is voluntary and has a smaller sample size while ACS is mandatory and has a much larger sample size.¹⁷ 2) NSCG is a random sampling for individuals who have at least a bachelor's degree, and these individuals are identified by ACS; however, responses to NSCG are voluntary.¹⁸ Thus, among the three surveys, ACS seems to be most representative of the US population because it has a much larger sample size and its responses are mandatory. In comparison, NSCG seems to have a better fit for the purpose of our study because IT professionals are more likely to hold at least a bachelor's degree.

One of the potential problems of the results in Table 20 is that the sample size, 109,573, is simply too large. It is well recognized that large sample studies (typically above 10,000) suffer from the “*p*-value problem”: as sample size increases, the *p*-value goes quickly to zero and the coefficient will always become statistically significant (Lin, Lucas, & Shmueli, 2013). In our analysis, this problem mainly arises from the large sample size of ACS surveys, with 76,500 workers, which is more than three times that of the CPS (see Table 22), making it hard to compare the results across different data sources.

To make the sample size of ACS more comparable to CPS and NSCG, we next restrict the sample size of ACS for further analysis. Year 2000 of ACS has relatively less employees and thus we use the original sample. For the rest of the years, we randomly select 30% of the original sample for 2001 to 2004, and 10% for 2005. We then use these subsamples for estimation for each year, for ACS 2000-2005 together, and for the combined dataset of CPS, NSCG, and ACS.

It is probably helpful to point out the advantages of using these restricted ACS subsamples. First, the size of the subsamples in each year is very close to that of CPS, making it more comparable to the results from CPS analysis. Second, the sample size of the dataset combining CPS, NSCG, and ACS, which is now 53,216, is about the same as that of ML, and this makes the findings from the combined dataset more comparable to ML.

The results for the combined ACS 2000-2005 are shown in Table 24, and those for separate years in Table 25. All controls are the same as in Table 18 and Table 19, corresponding to each model there. In Table 24, the coefficient on *noncitizen* from Model 4 is not significant ($b=0.016$, $p>0.1$), and therefore, H1a is not supported. The estimated wage premium to *noncitizen* using PSM is 0.054 ($p<0.1$), weakly supporting H1a. Moreover, in Table 25, the coefficient on *noncitizen* is not significant ($p>0.1$) across all six years.

¹⁷ Please refer to <https://www.census.gov/topics/education/school-enrollment/guidance/acs-vs-cps.html>.

¹⁸ Please refer to <https://www.census.gov/programs-surveys/nscg/respondent/faqs.html>.

<i>Ln(wage)</i>	Model 1	Model 2	Model 3	Model 4	Model 5
Noncitizen	0.106*** (0.013)	0.068*** (0.014)	0.027** (0.013)	0.016 (0.013)	0.072*** (0.013)
R ²	0.003	0.195	0.253	0.279	0.189

Notes: N=20,143; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

<i>Ln(wage)</i>	2000	2001	2002	2003	2004	2005
Noncitizen	0.008 (0.033)	0.025 (0.029)	0.014 (0.033)	0.038 (0.032)	-0.013 (0.035)	0.014 (0.033)
R ²	0.318	0.321	0.325	0.293	0.321	0.306
N	3,770	3,637	3,157	3,445	3,424	2,710

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Additionally, we re-estimate Equation (1) using the dataset combining CPS, NSCG, and ACS. The results are reported in Table 26 and 27. Again, all controls are the same as in Table 20 and 21, corresponding to each model there. In Model 4 of Table 26, the coefficient on *noncitizen* is not significant ($b=0.013$, $p>0.1$), not supporting H1a. In contrast, the estimated wage premium to *noncitizen* using PSM is 0.059 ($p<0.05$), supporting H1a. In Table 27, the coefficient on *noncitizen* is significant only in year 2001 ($b=0.047$, $p<0.05$).

<i>Ln(wage)</i>	Model 1	Model 2	Model 3	Model 4	Model 5
Noncitizen	0.115*** (0.009)	0.065*** (0.009)	0.025*** (0.009)	0.013 (0.009)	0.042*** (0.009)
R ²	0.004	0.178	0.232	0.254	0.107

Notes: N=53,216; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

<i>Ln(wage)</i>	2000	2001	2002	2003	2004	2005
Noncitizen	0.023 (0.023)	0.047** (0.022)	0.021 (0.024)	-0.002 (0.016)	-0.011 (0.022)	0.031 (0.022)
R ²	0.241	0.239	0.238	0.234	0.269	0.234
N	6,944	6,932	6,485	18,167	7,635	7,053

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Obviously, the results of hypothesis testing in Table 22 have changed with these ACS subsamples. We update the results and show them in Table 28. All changes are shown in bold with arrows representing directions of changes. Naturally, changes occur only when ACS subsamples are used.

Data Sources (sample size)	Estimation Approaches	Results of Testing H1a
CPS (22,596)	Six years combined	Not supported ($p>0.1$)
	Separate years	Weakly supported ($p<0.1$) only in 2001
	PSM	Not supported ($p>0.1$)
NSCG (10,477)	One year	Not supported ($p>0.1$)
	PSM	Supported ($p<0.05$)
ACS (20,143)	Six years combined	Supported ($p<0.01$) → not supported ($p>0.1$)
	Separate years	Weakly supported ($p<0.1$) only in 2001 → not supported ($p>0.1$) in all six years
	PSM	Supported ($p<0.05$) → weakly supported ($p<0.1$)
The combined dataset (53,216)	Six years combined	Supported ($p<0.01$) → not supported ($p>0.1$)
	Separate years	Supported ($p<0.05$) in three of six years → supported only in 2001 ($p<0.05$)
	PSM	Supported ($p<0.05$)

Several changes are observed: 1) The significance levels of all tests have come down, except the PSM result for the combined dataset. 2) Results from ACS are now very similar to those from CPS. This is not surprising, because they now have similar sample size, similar population characteristics, and similar survey questions. The similar results actually validate our findings from the two data sources. 3) The results from the combined dataset also have changed: the wage premium from Model 4 in Table 20 has turned from significant ($p < 0.05$) to insignificant ($p > 0.1$) in Table 26; the wage premium is significant in three years in Table 21, but now it is significant only in 2001 in Table 27.

5 Conclusion

In this study, we replicate H1a in ML and test whether foreign IT workers earn a wage premium than the natives. While being able to obtain the same results as ML, this study shows that the estimated wage premium to foreign IT workers may differ across data sources used, the predicting variables controlled, the estimation methods applied, as well as sample sizes of the data sources. We recognize that exact replication is very hard to conduct for social science research because social events are not closed systems and identical social situations between the original and replication studies rarely exist. As commented by prior studies that replicability does not mean conclusive verification, and failure to replicate does not mean conclusive falsification (Dennis & Valacich, 2014; Tsang & Kwan, 1999). In this research, the three surveys we use draw samples from different populations, and there might be systematic differences in the samples the surveys target and the ways data are collected, and thus leading to the differences in estimation results. The various results from our replication study suggest that replication studies are not only necessary but also important since they can potentially help generalize or extend the findings of original studies. In this sense, we believe our study contributes to the literature on wage compensation to IT workers, enriches studies on wage differential between foreign and native IT workers, and deepens our understanding on the impact of foreign IT workers on the natives.

We would also like to add that an interesting dimension of compensation research is that worker compensation reflects the changing economic and political considerations, and thus findings from this line of research constantly evolve. To gain a better understanding of the patterns of changes and their underlying mechanisms, we suggest three possible venues for future research: 1) Since the economic and political environments have changed considerably over the years, scholars may use data from more recent years to examine the wage premium hypothesis to see whether and how the patterns found in ML have changed. 2) At the micro level, it is worthwhile to conduct longitudinal analysis from the career development perspective to investigate whether and how the potential wage gap between foreign and native IT workers may evolve over time (Joseph, Boh, Ang, & Slaughter, 2012). Such studies could shed more light on the wage gap observed at the aggregated level. 3) Scholars may also want to examine the different skill sets owned by foreign and native IT workers (Peng & Zhang, 2020), and investigate how wage premium may arise from their skill complementarity or substitution between the two groups of IT workers.

References

- Ang, S., Slaughter, S., & Ng, K. Y. (2002). Human capital and institutional determinants of information technology compensation: Modeling multilevel and cross-level interactions. *Management Science*, 48(11), 1427-1445.
- Angrist, J., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Aobdia, D., Srivastava, A., & Wang, E. Q. (2018). Are immigrants complements or substitutes? Evidence from the audit industry. *Management Science*, 64(5), 1997-2012.
- Birgier, D. P. (2017). Immigration, occupations, and native wages: Long time trends in the US. *Research in Social Stratification and Mobility*, 51, 41-55.
- Burtch, G., Carnahan, S., & Greenwood, B. N. (2018). Can you gig It? An empirical examination of the gig economy and entrepreneurial activity. *Management Science*, 64(12), 5497-5520.
- Dennis, A. R., Brown, S. A., Wells, T. M., & Rai, A. (2020). Replication crisis or replication reassurance: Results of the IS replication project. *MIS Quarterly*, 44(3), iii-vii.
- Dennis, A. R., & Valacich, J. S. (2014). A replication manifesto. *AIS Transactions on Replication Research*, 1(1), Article 1.
- Drago, D. M. (2015). Losing the best and the brightest: The disappearing wage premium for H-1B visa recipients. *Vanderbilt Journal of Entertainment & Technology Law*, 17(4), 1051-1084.
- Friedberg, R. M. (2001). The impact of mass migration on the Israeli labor market. *Quarterly Journal of Economics*, 116(4), 1373-1408.
- Gibson, C., & Jung, K. (2006). *Historical census statistics on the foreign-born population of the United States: 1850 to 2000*. <https://www.census.gov/content/dam/Census/library/working-papers/2006/demo/POP-twps0081.pdf>.
- Hira, R. (2010). *The H-1B and L-1 visa programs: Out of control*. Economic Policy Institute, Washington DC.
- Hunt, J. (2011). Which immigrants are most innovative and entrepreneurial? Distinctions by entry visa. *Journal of Labor Economics*, 29(3), 417-457.
- Jaeger, D. A. (1997). Reconciling the old and new Census Bureau education questions: Recommendations for researchers. *Journal of Business & Economic Statistics*, 15(3), 300-309.
- Joseph, D., Boh, W. F., Ang, S., & Slaughter, S. A. (2012). The career paths less (or more) traveled: A sequence analysis of it career histories, mobility patterns, and career success. *MIS Quarterly*, 36(2), 427-452.
- Kerr, W. R. (2013). *U.S. high-skilled immigration, innovation, and entrepreneurship: Empirical approaches and evidence*. Harvard Business School Working Paper, No. 14-017.
- Kerr, W. R., & Lincoln, W. F. (2010). The supply side of innovation: H-1B visa reforms and US ethnic invention. *Journal of Labor Economics*, 28(3), 473-508.
- Kostanich, D. L., & Dippo, C. S. (2002). *Current Population Survey: Design and Methodology*.
- Krueger, A. B. (1993). How computers have changed the wage structure: Evidence from microdata, 1984-1989. *Quarterly Journal of Economics*, 108(1), 33-60.
- Levina, N., & Xin, M. D. (2007). Comparing IT workers' compensation across country contexts: Demographic, human capital, and institutional factors. *Information Systems Research*, 18(2), 193-210.
- Lin, M. F., Lucas, H. C., & Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem. *Information Systems Research*, 24(4), 906-917.
- Matloff, N. (2003). On the need for reform of the H-1B non-immigrant work visa in computer-related occupations. *University of Michigan Journal of Law Reform*, 36, 815-915.
- Matloff, N. (2004). Globalization and the American IT worker. *Communications of the ACM*, 47(11), 27-29.

- Matloff, N. (2013). Immigration and the tech industry: As a labour shortage remedy, for innovation, or for cost saving? *Migration Letters*, 10(2), 211-228.
- Mincer, J. A. (1974). *Schooling, Experience, and Earnings* (1 ed.). New York, NY: Columbia University Press.
- Mithas, S., & Lucas, H. C. (2010). Are foreign IT workers cheaper? US visa policies and compensation of information technology professionals. *Management Science*, 56(5), 745-765.
- Orrenius, P. M., & Zavodny, M. (2007). Does immigration affect wages? A look at occupation-level evidence. *Labour Economics*, 14(5), 757-773.
- Peng, G., & Eunni, R. V. (2011). Computer skills, non-routine tasks, and wage premium: A longitudinal study. *Journal of Strategic Information Systems*, 20(4), 449-460.
- Peng, G., Wang, Y., & Han, G. H. (2018). Information technology and employment: The impact of job tasks and worker skills. *Journal of Industrial Relations*, 60(2), 201-223.
- Peng, G., & Zhang, D. D. (2020). *The complementarity between foreign and native IT workers: Evidence from an industry-level analysis*. Paper presented at the Americas Conference on Information Systems.
- Tambe, P., & Hitt, L. M. (2012). The productivity of information technology investments: New evidence from IT labor data. *Information Systems Research*, 23(3), 599-617.
- Tsang, E. W. K., & Kwan, K. M. (1999). Replication and theory development in organizational science: A critical realist perspective. *Academy of Management Review*, 24(4), 759-780.
- Wang, C. N., & Kaarst-Brown, M. L. (2014). The IT compensation challenge: Theorizing the balance among multi-level internal and external uncertainties. *Journal of the Association for Information Systems*, 15(3), 111-146.
- Whitaker, J., Mithas, S., & Liu, C. W. (2019). Beauty is in the eye of the beholder: Toward a contextual understanding of compensation of information technology professionals within and across geographies. *Information Systems Research*, 30(3), 892-911.

About the Authors

Gang Peng is an Associate Professor at College of Business and Economics, California State University, Fullerton. He earned his PhD from the University of Washington, Seattle, USA. His research interests primarily focus on the adoption, diffusion, usage, and impact of information technology. His research has appeared in *ACM Transactions on MIS*, *Decision Support Systems*, *Decision Sciences*, *Industrial Marketing Management*, *Information Systems Research*, *Journal of Management Information Systems*, *Journal of Industrial Relations*, and *Journal of Strategic Information Systems*, among others.

Rahul Bhaskar is a Professor at College of Business and Economics, California State University, Fullerton. He earned his PhD from the University of Wisconsin, Madison, USA. His research primarily focuses on Health Care Analytics and adoption of information technology. His research has appeared in *Communications of ACM*, *Interfaces*, *Information Systems Frontiers*, *Journal of Cases on Information Technology*, among others. He has been a recipient of grants from various organizations including Department of Justice to apply Artificial Intelligence and Machine Learning to Law Enforcement and Cybersecurity.

Copyright © 2021 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from ais@aisnet.org.