Supplemental appendix to “The Effect of Immigration on Productivity: Evidence from U.S. States” by Giovanni Peri, *Review of Economics and Statistics,* Vol. 94, No. 1 (February 2012), pp. 348–358.

URL: http://www.mitpressjournals.org/doi/abs/10.1162/REST_a_00137

Online Appendix for the Paper:

The Effect of Immigration on Productivity:

Evidence from US States

Giovanni Peri (UC Davis and NBER)

August 13, 2010

1 Online Appendix: IPUMS Census Data

1) Eliminate people living in group quarters (military or convicts): those with the gq variable equal to 0, 3 or 4.

2) Eliminate people younger than 17. Since people of working age are defined by BLS as those 16 an older, and since the questions related to work variables pertain to the previous year, we consider 17 years of age as the cut-off.

3) Eliminate those who worked 0 weeks last year, which corresponds to wkswork2=0 in 1960 and 1970 and wkswork1=0 in 1980-1990-2000 and ACS.

4) Once we calculate experience as age-(time first worked), where (time first worked) is 16 for workers with no HS degree, 19 for HS graduates, 21 for workers with some college education and 23 for college graduates, we eliminate all those with experience <1 and >40.

5) Eliminate those workers who do not report valid salary income (999999) or report 0.

6) Eliminate the self-employed (keeping those for whom the variable CLASSWKD is between 20 and 28).

Construction of hours worked and employment by cell

To calculate the total amount of hours worked by natives and immigrants, male and female, in each education-experience cell, we add the hours worked by each person multiplied by her personal weight (PERWT) in the cell.

Construction of the average hourly wage by cell

In each cell we average the hourly wage of individuals, each weighted by the hours worked by the individual. Hence individuals with few hours worked (low job attachment) are correspondingly weighted little in the

calculation of the average wage of the group.

1.1 Individual Variables: Definition and Description

Education: Education groups in each year are defined using the variable EDUCREC which was built in order to consistently reflect the variables HIGRADE and EDUC99. In particular, we define as less educated (L) those with EDUCREC<=7, corresponding to a high school degree or less. Highly educated are those with EDUCREC>=8, corresponding to some college or more.

Experience: Defined as potential experience, assigns to each schooling group a certain age reflecting the beginning of their working life; in particular, the initial working ages are: 17 years for workers with no degree, 19 years for high school graduates, 21 years for those with some college education and 23 years for college graduates.

Immigration Status: In each year, only people who are not citizens or who were naturalized citizens are counted as immigrants. This is done using the variable CITIZEN and by attributing the status of foreign-born to people when the variable is equal to 2 or 3. In 1960, the variable is not available and the selection is done using the variable BPLD (birthplace, detailed) and by attributing the status of foreign-born to all of those for which BPLD>15000, except for the codes 90011 and 90021 which indicate U.S. citizens born abroad.

Weeks Worked in a Year: For the censuses 1960 and 1970 the variable used to define weeks worked in the last year is WKSWORK2, which defines weeks worked in intervals. We choose the median value for each interval so that we impute to individuals weeks worked in the previous year according to the following criteria: 6.5 weeks if wkswork2=1; 20 weeks if wkswork2=2; 33 weeks if wkswork2=3; 43.5 weeks if wkswork2=4; 48.5 weeks if wkswork2=5; 51 weeks if wkswork2==6. For the censuses 1980, 1990, 2000 and ACS we use the variable wkswork1 which records the exact number of weeks worked last year.

Hours Worked in a Week: For census years 1960 and 1970 the variable used is HRSWORK2 which measures the hours worked during the last week, using intervals. We attribute to each interval its median value and measure the number of hours per week worked by an individual according to the following criteria: 7.5 hours if hrswork2=1; 22 hours if hrswork2=2; 32 hours if hrswork2=3; 37 hours if hrswork2=4; 40 hours if hrswork2=5; 44.5 hours if hrswork2=6; 54 hours if hrswork2=7; 70 hours if hrswork2==8. For the censuses 1980, 1990, 2000 and ACS we use the variable UHRSWORK which records the exact number of hours worked in a usual week by a person.

Hours Worked in a Year: This is the measure of labor supply by an individual and it is obtained by multiplying Hours Worked in a Week by Weeks Worked in a Year, as defined above.

Yearly Wages: The yearly wage in constant 1999 US dollars is calculated as the variable INCWAGE multiplied by the price deflator suggested in the IPUMS, which is the one below. Recall that each census and

ACS is relative to the previous year so the deflators below are those to be applied to years 1960, 1970, 1980, and so on:

$Year$	1959	1969	1979	1989	1999	2005
$Deflator$	5.725	4.540	2.314	1.344	1.000	0.853

Topcodes for Yearly Wages: Following an established procedure we multiply the topcodes for yearly wages in 1960, 1970 and 1980 by 1.5.

Hourly Wages: The hourly wage for an individual is constructed by dividing the yearly wage as defined above by the number of weeks worked in a year times the number of hours worked in a week.

2 Online Appendix: Construction of Physical Capital by State

In our construction of the state capital stocks we follow Garofalo and Yamarik (2002). This involved distributing the national capital stock by industry and year, obtained from the BEA (2008b), to each state and industry and year according to the percentage of value added for the state and industry and year in the national value added for that industry and year, obtained from the BEA (2008a). In other words, following the notation of the paper and denoting as j one industry, we constructed capital stocks for state s and industry j as :

$$K_{s,j}(t) = \left[\frac{Y_{s,j}(t)}{Y_j(t)}\right] K_j(t)$$

We then summed over all industries j, for each state s, in year t, to obtain a capital stock series by state and year. Finally, we used as a price deflator the implicit capital deflator, obtained from the aggregate BEA data to transform the capital stock series into real values. Furthermore, the value added data at the state level needed to be generated for all years using a concordance, as described below. That concordance left us with 19 industries that we use to attribute the capital stock. The industries are: Agriculture; Forestry; Fishing and Hunting; Mining; Utilities; Construction; Manufacturing; Wholesale Trade; Retail Trade; Transportation and Warehousing; Information, Finance and Insurance; Real Estate and Rental and Leasing; Professional, Scientific, and Technical Services; Management of Companies and Enterprises; Administrative and Waste Management Services; Educational Services; Health Care and Social Assistance; Arts, Entertainment, and Recreation; Accommodation and Food Services; Other Services, except Government.

Constructing the NAICS97 to SIC87 Concordance

The first step in generating the capital stock by state was to generate a crosswalk, or concordance, from NAICS97 to SIC87 using the Census Bureau's crosswalks at http://www.census.gov/epcd/ec97brdg/index.html. This step was necessary in order to extend the BEA's value added by state data to pre-1997 dates. The bridge from NAICS97 to SIC87 (NtoS) lists a NAICS code and then the corresponding SIC codes that go into it, and then the establishments, sales, payroll and employees per that combination. The file does not, however, list the

percentage of the SIC category which should be attributed to the NAICS code, and since there may be more than one NAICS code per SIC code, this information is needed. The HTML version on the website does list this percentage, but it is unfortunately not in the electronic file. This percentage can be calculated using the opposite bridge from SIC87 to NAICS97 (StoN). The StoN file contains the same variables as the NtoS file, but maps all the NAICS that go into a given SIC. Also available are the totals of the 4 categories (sales, etc.) for each SIC code, at different digit levels (2-digit, 3-digit, etc.).

We delete everything in the StoN file except the SIC totals (we delete the SIC to NAICS mappings). We then merge these to the NtoS file by SIC code, so that now the NtoS file has the mapping as before, but also includes the totals for each SIC value next to each NAICS-sic pair. Then the percentage can be calculated for each NAICS-sic combination by dividing the NAICS-SIC totals into the merged SIC totals. Since what we actually want is SIC2 to NAICS2, and the original mapping (NtoS) is actually SIC4 to NAICS6, before merging the SIC4 totals into the NtoS file we trimmed the NAICS codes down to 2 digits, and then summed up over the unique SIC4-NAICS2 combinations. We then trimmed the SIC4 values to SIC2, and summed over the unique SIC2-NAICS2 values. Finally, we merged in the SIC2 totals from the StoN file and calculated the percentage of each SIC2 that goes into each NAICS2.

Online Table Appendix

Table A1: Summary statistics: yearly growth in each decade, average across US states

Annual Growth Rate of:	***60's***	***70's***	***80's***	***90's***	***2000-2006***
Foreign Employment (as percentage of total)	0.1%	0.3%	0.25%	0.5%	0.5%
Total Employment ($\hat{N}$ *)*	2.5%	3.2%	2.1%	2.5%	1%
Gross Product per Worker ($\hat{y}$ *)*	+1.9%	+1.2%	+1.0%	+1.2%	+1.9%
Capital-Output ratio $(\hat{K} - \hat{Y})$	-0.5%	-0.3%	-0.7%	-0.7%	-0.1%
Total Factor Productivity $\hat{A}$	2.1%	0.1%	1.4%	1.5%	2.3%
Hours per Worker $\hat{x}$	-0.02%	1.2%	0.4%	0.3%	-0.04%
Share of Highly Educated $\hat{h}$	2.1%	3.0%	2.8%	0.9%	0.02%
Skill-bias of Technology $\hat{\beta}$	2.9%	2.6%	3.4%	0.8%	0.2%

Note: The variables are constructed as described in the text. We are reporting the yearly growth rates (averaging across states and years of the decade) by decade for each variable.