REPLICATION OF BORJAS (2003): THE IMPACT OF IMMIGRATION ON NATIVE CITIZENS WAGES

ECON 6693: Masters Research Project

AUGUST 4, 2025

SUMMER 25

Anahi Frutiz & Brian McHatton

Introduction

In his working paper "The Labor Demand Curve is Downward Sloping," George Borjas investigates how immigration affects the wages of native U.S. workers. Borjas theorizes that immigration causes a downward pressure on labor demand. Through the use of Census and CPS data from the years 1960 to 2001, he supported his theory as he found that an increase in immigration was associated with lower native wages (-0.4).(Borjas 2003) "Immigration has been a major driving force behind U.S. labor force growth…according to data from the U.S. Census Bureau" (Gelatt 2024). Discussion surrounding the topic of immigration has remained relevant throughout the years, often a major policy dispute for political campaigns. More recently with a newly elected president there has been increased discussion surrounding policy regarding immigration. Immigration plays a key role in the U.S. labor market, and this study serves to provide insight as to whether that affect is positive or negative.

Other studies such as Ken-Hou Lin & Inbar Weiss, found "a large presence of immigrants concurs with lower native wages at the bottom of the labor market but higher native wages at the upper end." (Lin & Weiss, 2019) Therefor they maintained that immigration led to increased inequality amongst native wages.

Our replication investigates Borjas's main research question with more recent data which was collected from the Current Population Survey for the years 2002, 2012, and 2022. Through the application of his empirical framework with an updated dataset and methodology, we investigate whether Borjas' original findings hold in a more updated labor market. We seek to discover if immigration lowers the wages of native-born U.S. workers. Our hypothesis is that there is no significant change in log average weekly wages given a change in immigrant share population.

Data Summary

In our study we replicated Borjas (2003), and using IPUMS we aggregated three years of data: 2002, 2012, and 2022. The sample consisted of men aged 19-64 who are actively in the labor market in all 50 of the US states (excluding territories or districts like Washington D.C.). The sample was specifically selected to avoid individuals working in high school and those who may earn any form of retirement income (i.e. social security), which typically occurs at the age of 65.

The total number of sampled individuals was 1,682,141. Utilizing the St. Louis FRED *Consumer Price Index for All Urban Consumers: All items in the U.S. City Average (CPIAUCSL)*, we corrected the reported wage income (INCWAGE) to 2022 dollars, before ridding each year of INCWAGE outliers.

Once cleaned, we used INCWAGE and education (EDUC) variables pulled from IPUMS CPS database to generate two new variables: average weekly wage (AVGWAGEWK) and a corrected, more streamlined education variable (EDUC).

The variable AVGWAGEWK is an approximation for how much money an individual makes each week, calculated by:

$$AVGWAGEWK = \frac{INCWAGE}{52}$$

This notably does not account for hours worked in a year or weeks of unemployment. However, for the purpose of our study we will be using the log of AVGWAGEWK as the dependent variable.

EDUC was cleaned to estimate years of schooling, with less than an 8th grade education counting as zero years of education.

Dummy variables were created to describe educational attainment: those with 0-11 years of education were classified as DROPOUT, those with 12 years of education were classified as HSGRAD, those with greater than 12 but less than 16 years were classified as SOMECOL, and lastly those with 16 or more years of education were classified as COLGRAD.
Furthermore, we created an additional variable WORKEXP which approximated experience based on age and education:

$$WORKEXP = AGE - EDUC - 6$$
$$WORKEXP \geq 0$$

If a value was negative, it was instead changed to 0. This equation accounts for years of schooling and a constant of 6 years before primary schooling, meaning a high school graduate who is 18 years old and has completed 12 years of education would be defined as 18-years-old with 0 years of experience.

Using our newly created variable for work experience (WORKEXP) we sorted individuals into four experience categories and created dummies: WORKEXP0_5, WORKEXP6_10, WORKEXP11_15, and WORKEXP16_PLUS. While Borjas utilized eight work experience dummies we found that after 15 years of experience, the wage curves converged and were not meaningfully different, so we opted to reduce the number of dummies to maintain efficiency and degrees of freedom.

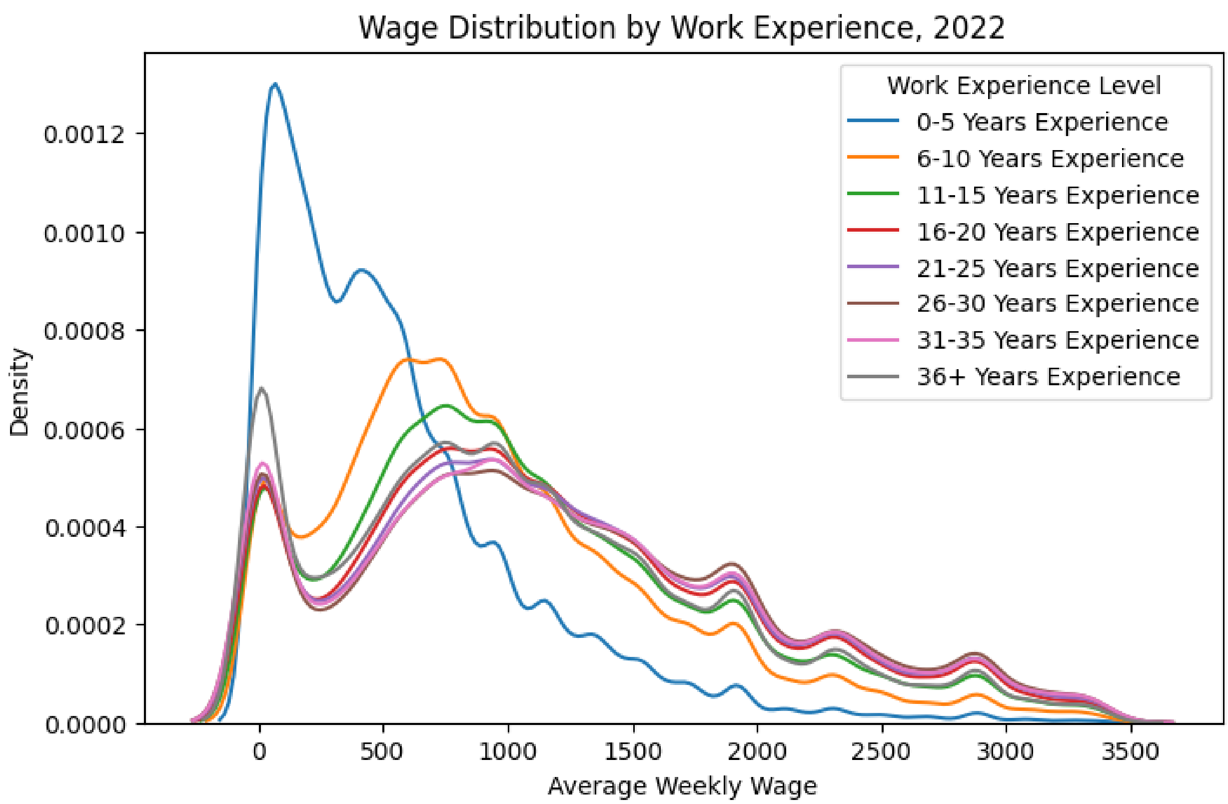

Figure 1: Wage Distribution by Work Experience, 2022

To control for interaction between education and experience, we made 16 interaction variables for each education and work experience level. In addition, we created state dummies based on the unique STATEFIP identifiers to yield a set of dummies categories as STATE_X.

We assessed the nativity of the sampled individuals (IMMIGRANT), classifying those as either native-born citizens (including citizens of the United States born abroad) or other, which includes naturalized citizens and non-citizens.

After identifying immigration status with the IMMIGRANT dummy, each U.S. citizen (IMMIGRANT = 0) was put into a group (i.e., "bucket"). Each bucket consists of native individuals with the same education, work experience and state values. These buckets each contain a minimum of at least 50 individuals. Buckets with less than 50 individuals were dropped to ensure viable point estimates and reduce within-group variation.

Following the creation of these buckets, we generated our dependent variable, IMMIGRANTPERCENT. This variable is the corresponding number of non-natives as a percentage of the population in their corresponding bucket (i.e., it is the weight of population for a given bucket (same education, work experience, and state) made up by immigrants, but immigrants were not included in the final data set, as our research is focused on native wages). IMMIGRANTPERCENT is a weighted variable that, over different years shows change in immigrant population for a given educational, work experience, and state category.

After aggregating the data, we had 1801 observation groups. We created STATEWEIGHT which is variable for the population weight of each state by number of observations in the data set for a given year.

Empirical Strategy

Following the completion of organizing and cleaning our data set we ran a weighted least squares regression:

$$
\begin{aligned}
&Log(AVGWAGEWK_{ist}) \\
&\quad = \beta_0 + \beta_1 IMMIGRANTPERCENT_{ist} + \beta_2 DROPOUT_{ist} + \ \beta_3 SOMECOL_{ist} \\
&\quad + \beta_4 COLGRAD_{ist} + \beta_5 WORKEXP0\_5_{ist} + \beta_6 WORKEXP6\_10_{ist} \\
&\quad + \beta_7 WORKEXP11\_15_{ist} + \beta_8 WORKEXP16\_PLUS_{ist} \\
&\quad + \beta_9 (DROP\_WORKEXPX\_X)_{ist} + \beta_{10} (HSG\_WORKEXPX\_X)_{ist} \\
&\quad + \beta_{11} (SOME\_WORKEXPX\_X)_{ist} + \beta_{12} (COL\_WORKEXPX\_X)_{ist} + \sum_s \delta_s \\
&\quad \cdot STATE_S + \sum_t \gamma_t \cdot YEAR_t + \varepsilon_{ist}
\end{aligned}
$$

- $Log(AVGWAGEWK_{ist}) = Average\ weekly\ wage\ for\ group\ i\ in\ state\ s\ and\ year\ t$
- $IMMIGRANTPERCENT_{ist} = \ percentage\ of\ immigrants\ in\ the\ population$ $of\ a\ group\ \ with\ i\ set\ of\ education\ and\ experience, in\ states\ s, in\ year\ t$
- $Education\ dummies$: $DROPOUT, HSGRAD, SOMECOL, COLGRAD$
- $Work\ Experience\ dummies$: $WORKEXP0\_5, \ldots, WORKEXP16\_PLUS$
- $Interaction\ terms\ between\ education\ and\ experience$: $e.g., DROP\_WORKEXPX\_X$
- $STATE\ and\ YEAR\ are\ fixed\ effects$
- $\varepsilon_{ist} = error\ term$

In our regression, the log of average weekly wage was regressed on our variable IMMIGRANTPERCENT, controlled by education, work experience, state- and time-fixed effects, and weighted by STATEWEIGHT. Using 2022 as our base year, HSGRAD as base level of education and California as our base state.

Replication Approach

In our approach to replicate Borjas (2003), we maintained the core approach investigating if Immigration affects native wages. We utilized the same methodology implementing a weighted least squares regression using Current Population Survey (CPS) and Census data collected from IPUMS. In addition, we chose to only look at native male workers. As previously mentioned, we grouped individuals into skill-specific “buckets” based on education, work experience, and state. With our key independent variable being IMMIGRANTPERCENT we applied a weighted least squares regression like Borjas’s original model which includes controls for education, experience, state, and year fixed effects, with the inclusion of interaction terms. Our final dataset resulted in about 1,801 observations.

Key differences in our approach included an updated time period. In our replication we used more recent data looking specifically at 2002, 2012, and 2022 as opposed to the 1960-to-2001-timeframe utilized by Borjas, with our intention being to provide a more updated insight into Borjas research and see if his findings still stand.

In his approach Borjas used national-level data grouping workers into 40 skill groups based on 4 education categories and 8 experience categories. In our replication we sampled all 50 U.S. States using California as our base. With groups or "buckets" of at least 50 individuals sorted based on 4 education categories: drop out, high school graduate, some college, and college grad, and 4 experience categories: 0-5 years, 6-10 years, 11-15 years and 16 plus.
All together we maintained the core structure of Borjas's analysis. Adapting some aspects of methodology using updated data to reflect a more modern labor market.

Results
In our research we sought to discover the effect of immigration on the log of average weekly wages among native-born male workers aged 16-64. Our regression controls for education, work experience, state, and year fixed effects, and is weighted by each state's share of the sample population STATEWEIGHT.
Table 1 represents our main findings in comparison to those of Borjas (2003).

Study	Coefficient on Immigrant Share	Direction	Significance	R-Square
Borjas (2003)	-0.40	Negative	Strong	N/A
Frutiz, McHatton (2025)	+0.0042	Positive	Significant ($p<0.01$)	0.965

Table 1: Comparison of findings between Borjas (2003) & Frutiz, McHatton (2025)

With an R-squared of 0.965, our model performs well as approximately 96.5% of the variation in the log of average weekly wages is explained by the model. With IMMIGRANTPERCENT as our key independent variable which measures the share of immigrants within each education-experience-state group. We found that a 10-percentage point increase in immigrant share is associated with a 0.4% increase in log average weekly wages.

Variable	**Coef**	**P> ltl**
Year_2002	0.07	0.000
Year_2012	-0.12	0.000

Table 2: Year fixed effects

At first glance there are notable shifts amongst our year dummy variables. *Table 2* shows that relative to 2022 wages were significantly higher in 2002 with a coefficient of **0.07**, while with a coefficient of **-0.12** wages dropped in 2012.

Additionally, we saw state-level fixed effects demonstrated a variation across regions. Using California as our base state, states such as Virginia (0.137), Washington (0.1400) and Wyoming (0.1526) showed significant and positive wage differentials compared to California. While other states like Mississippi (-0.0364) and Montana (-0.0574) demonstrated lower wages at the 10% confidence interval.

Unlike Borjas our results indicate a small but positive association between immigration and native wages in our modern labor market. This may suggest that the impact that immigration may have evolved over time or even that our slight differences in methods made a difference in our results.

Discussion and Reflection

In our own research we were able to successfully implement Borjas's regression framework. We constructed meaningful skill-experience groups and effectively adjusted for inflation as well as properly cleaned our dataset. Using Borjas (2003) as our reference guide we were able to replicate his study with minor adjustments to provide meaningful insight as to how immigration affects native-born citizens wages.

Through this eight week process we faced some challenges in our research. Having to adjust our data set over four times to ensure the proper variables were selected, ensuring variables were properly cleaned, and narrow down our data set to a reasonable number of variables. Being that Borjas's research is a working paper we were at times working with limited information regarding methodology. Through brainstorming and discussion our team made decisions that we felt were best suited for our replication.

Though our results deviated from Borjas (2003) as we found a positive relationship amongst immigration and native-born wages it is possible that results might vary simply due to time. Our results suggest that over time, immigration may have less adverse effects on wages. Another explanation might be a result of changes in immigration. Perhaps the slight adjustments in methodology and looking at immigration from a geographical standpoint as opposed to a national level might explain our differing results.

Future Research

There is room for future research to try and address endogeneity issues regarding regional selection bias. There is likely a bias that immigrants move to or are offered jobs in areas of the U.S. that are growing economically and have a demand for labor.

Additionally, we kept four education categories to better simulate Borjas's approach, however five education groupings may be appropriate to accommodate for tertiary education such as master's or professional degrees.

Lastly our data could be broadened or updated to include women and additionally focus on demographic differences such as race.

Appendix

Table A1. Variables & Codebook

The following codebook is a combination of the variables pulled from IPUMS including our constructed variables WORKEXP and YEAR_2022.

Variable Name	Description	Values/Codes
YEAR_2002	Dummy variable indicating year of observation	1 if the year is 2002, else 0
YEAR_2012	Dummy variable indicating year of observation	1 if the year is 2012, else 0
YEAR_2022	Dummy variable indicating year of observation	1 if the year is 2022, else 0
IMMIGRANTPERCENT	Percent of immigrants within each education x experience x state group	0 – 90.872%
DROPOUT	Dummy for individuals with less than high school diploma	1 if DROPOUT, 0 else
HSGRAD	Dummy for individuals with a high school diploma	1 if HSGRAD, 0 else
SOMECOL	Dummy for individuals with some college education	1 if SOMECOL, 0 else
COLGRAD	Dummy for individuals with a college degree	1 if COLGRAD, 0 else
WORKEXPX_X	Dummy variable for years of work experience, based on age and education	0-5, 6-10, 11-15, 16+
DROP_WORKEXPX_X	Interaction term: Dropout x Work Experience	0-5, 6-10, 11-15, 16+
SOME_WORKEXPX_X	Interaction term: Some College x Work Experience	0-5, 6-10, 11-15, 16+
HSG_WORKEXPX_X	Interaction term: High School grad x work experience	0-5, 6-10, 11-15, 16+
COL_WORKEXPX_X	Interaction term: College Graduate x Work Experience	0-5, 6-10, 11-15, 16+
STATE_X	State level dummy variables for geographic fixed effects	1 = resides in state X (e.g., Alabama); 0 = otherwise
AVGWAGEWK	Derived from INCWAGE variable by dividing by 52	Continuous: $216.90 - $1,840.71 (2022 dollars)

Figure A1. Regression Output (Full Model Results)

Below is a figure for the results our Weighted Least Squares (WLS) regression results estimating the effect of immigration percentage on log of average weekly wages. The following regression includes controls for education, experience, state fixed effects, year indicators, and interaction terms.

WLS Regression Results

Dep. Variable:	AVGWAGEWK	R-squared:	0.965
Model:	WLS	Adj. R-squared:	0.964
Method:	Least Squares	F-statistic:	753.6
Date:	Fri, 01 Aug 2025	Prob (F-statistic):	0.00
Time:	22:02:48	Log-Likelihood:	1506.4
No. Observations:	1801	AIC:	-2883.
Df Residuals:	1736	BIC:	-2526.
Df Model:	64		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
IMMIGRANTPERCENT	0.0042	0.000	9.664	0.000	0.003	0.005
DROPOUT	3.1008	0.017	185.447	0.000	3.068	3.134
SOMECOL	3.3468	0.008	415.835	0.000	3.331	3.363
COLGRAD	3.7045	0.008	460.117	0.000	3.689	3.720
WORKEXP0_5	2.8058	0.007	426.983	0.000	2.793	2.819
WORKEXP6_10	3.2423	0.008	431.601	0.000	3.228	3.257
WORKEXP11_15	3.3709	0.008	427.990	0.000	3.355	3.386
DROP_WORKEXP0_5	-2.358e-15	2.65e-17	-88.870	0.000	-2.41e-15	-2.31e-15
DROP_WORKEXP6_10	1.946e-16	5.43e-17	3.582	0.000	8.8e-17	3.01e-16
DROP_WORKEXP11_15	-2.096e-15	2.26e-17	-92.906	0.000	-2.14e-15	-2.05e-15
DROP_WORKEXP16_PLUS	3.1008	0.017	185.447	0.000	3.068	3.134
SOME_WORKEXP0_5	-0.3080	0.005	-59.362	0.000	-0.318	-0.298
SOME_WORKEXP6_10	-0.0003	0.005	-0.060	0.952	-0.010	0.010
SOME_WORKEXP11_15	0.0646	0.005	12.403	0.000	0.054	0.075
SOME_WORKEXP16_PLUS	3.5905	0.009	383.705	0.000	3.572	3.609
HSG_WORKEXP0_5	3.0566	0.008	374.695	0.000	3.041	3.073
HSG_WORKEXP6_10	3.1630	0.009	368.421	0.000	3.146	3.180
HSG_WORKEXP11_15	3.2187	0.009	374.407	0.000	3.202	3.236
HSG_WORKEXP16_PLUS	6.7825	0.015	439.525	0.000	6.752	6.813
COL_WORKEXP0_5	0.0572	0.005	11.020	0.000	0.047	0.067
COL_WORKEXP6_10	0.0796	0.005	15.328	0.000	0.069	0.090
COL_WORKEXP11_15	0.0876	0.005	16.877	0.000	0.077	0.098
COL_WORKEXP16_PLUS	3.4800	0.009	386.099	0.000	3.462	3.498
STATE_Alabama	0.0294	0.021	1.407	0.160	-0.012	0.070
STATE_Alaska	0.1672	0.047	3.561	0.000	0.075	0.259
STATE_Arizona	0.0290	0.016	1.783	0.075	-0.003	0.061
STATE_Arkansas	-0.0199	0.025	-0.792	0.429	-0.069	0.029
STATE_Colorado	0.1182	0.018	6.715	0.000	0.084	0.153
STATE_Connecticut	0.1056	0.021	5.044	0.000	0.065	0.147
STATE_Delaware	0.0834	0.043	1.925	0.054	-0.002	0.168
STATE_Florida	-0.0479	0.011	-4.477	0.000	-0.069	-0.027
STATE_Georgia	0.0207	0.015	1.416	0.157	-0.008	0.049

Figure A1. continued

STATE_Hawaii	0.0615	0.031	2.008	0.045	0.001	0.122
STATE_Idaho	-0.0057	0.029	-0.193	0.847	-0.063	0.052
STATE_Illinois	0.0825	0.013	6.208	0.000	0.056	0.109
STATE_Indiana	0.0846	0.018	4.673	0.000	0.049	0.120
STATE_Iowa	0.0412	0.023	1.765	0.078	-0.005	0.087
STATE_Kansas	0.0383	0.024	1.586	0.113	-0.009	0.086
STATE_Kentucky	0.0136	0.022	0.631	0.528	-0.029	0.056
STATE_Louisiana	0.0706	0.022	3.210	0.001	0.027	0.114
STATE_Maine	-0.0269	0.036	-0.749	0.454	-0.097	0.044
STATE_Maryland	0.1566	0.017	9.226	0.000	0.123	0.190
STATE_Massachusetts	0.1305	0.016	8.229	0.000	0.099	0.162
STATE_Michigan	0.0544	0.017	3.278	0.001	0.022	0.087
STATE_Minnesota	0.1242	0.019	6.612	0.000	0.087	0.161
STATE_Mississippi	-0.0364	0.026	-1.376	0.169	-0.088	0.016
STATE_Missouri	0.0552	0.019	2.879	0.004	0.018	0.093
STATE_Montana	-0.0574	0.039	-1.468	0.142	-0.134	0.019
STATE_Nebraska	0.0508	0.028	1.798	0.072	-0.005	0.106
STATE_Nevada	0.0519	0.022	2.347	0.019	0.009	0.095
STATE_New_Hampshire	0.1573	0.033	4.710	0.000	0.092	0.223
STATE_New_Jersey	0.0976	0.013	7.250	0.000	0.071	0.124
STATE_New_Mexico	-0.0487	0.032	-1.535	0.125	-0.111	0.014
STATE_New_York	0.0270	0.011	2.523	0.012	0.006	0.048
STATE_North_Carolina	-0.0004	0.015	-0.029	0.977	-0.030	0.029
STATE_North_Dakota	0.0851	0.043	1.994	0.046	0.001	0.169
STATE_Ohio	0.0668	0.017	4.026	0.000	0.034	0.099
STATE_Oklahoma	0.0178	0.022	0.812	0.417	-0.025	0.061
STATE_Oregon	0.0148	0.021	0.717	0.474	-0.026	0.055
STATE_Pennsylvania	0.0889	0.016	5.680	0.000	0.058	0.120
STATE_Rhode_Island	0.0802	0.038	2.093	0.037	0.005	0.155
STATE_South_Carolina	0.0066	0.020	0.332	0.740	-0.033	0.046
STATE_South_Dakota	-0.0034	0.041	-0.084	0.933	-0.084	0.078
STATE_Tennessee	-0.0148	0.018	-0.825	0.409	-0.050	0.020
STATE_Texas	0.0349	0.010	3.531	0.000	0.016	0.054
STATE_Utah	0.1333	0.022	6.040	0.000	0.090	0.177
STATE_Vermont	-0.0621	0.057	-1.090	0.276	-0.174	0.050
STATE_Virginia	0.1370	0.016	8.824	0.000	0.107	0.167
STATE_Washington	0.1400	0.016	8.954	0.000	0.109	0.171
STATE_West_Virginia	0.0180	0.032	0.567	0.571	-0.044	0.080
STATE_Wisconsin	0.0940	0.019	4.903	0.000	0.056	0.132
STATE_Wyoming	0.1526	0.058	2.625	0.009	0.039	0.267
YEAR_2002	0.0511	0.005	10.245	0.000	0.041	0.061
YEAR_2012	-0.1463	0.005	-29.906	0.000	-0.156	-0.137

Omnibus:	276.502	Durbin-Watson:	1.366
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1314.620
Skew:	-0.644	Prob(JB):	3.42e-286

Citations

Borjas, G. J. (2003). The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *The Quarterly Journal of Economics, 118*(4), 1335–1374. https://doi.org/10.1162/003355303322552810

Federal Reserve Bank of St. Louis. (n.d.). *Consumer price index for all urban consumers: All items in U.S. city average (CPIAUCSL) [Data set]*. FRED. https://fred.stlouisfed.org/series/CPIAUCSL

Gelatt, J. (2024, October). *Explainer: Immigrants and the U.S. economy*. Migration Policy Institute. https://www.migrationpolicy.org/content/explainer-immigrants-and-us-economy

IPUMS USA. (2025). *USA microdata extracts, 2022 and 2023*. https://usa.ipums.org/usa/

Lin, K.-H., & Weiss, I. (2019). Immigration and the wage distribution in the United States. *Demography, 56*(6), 2229–2252. https://doi.org/10.1007/s13524-019-00828-9