Imputation Biases in the Estimation of the Wage Effects of Immigration Status:
Evidence from the National Agricultural Workers Survey


Tom Hertz
USDA, Economic Research Service thertz@ers.usda.gov

Imputation Biases in the Estimation of the Wage Effects of Immigration Status:
Evidence from the National Agricultural Workers Survey

Tom Hertz
USDA, Economic Research Service thertz@ers.usda.gov

Views expressed are those of the authors and should not be attributed to the Economic Research
Service or USDA.

INTRODUCTION & MOTIVATION

- The effect of immigration on the wages and employment prospects of US born workers is a topic of
  great economic and political relevance. Most economists find that net effects are positive, but that
  less well-educated US born workers may be negatively affected.

- Flows into the US of authorized versus unauthorized immigrants might be expected to have
  different effects on the labor market. Unauthorized workers earn somewhat lower hourly wages,
  holding observable productive characteristics constant, likely because they have less bargaining
  power, and more to lose from confrontations with employers or the law, than authorized workers.

- Most studies of the impact of immigration fail to control for legal status, and may thus be biased.
  Yet few datasets record legal status, making analysis difficult. One option is to impute that status
  from observable data. A recent example is Borjas (2017), who finds that the legal status wage
  penalty has declined in recent years:

**Trend in the wage penalty to unauthorized immigrant workers, 2001-2014**
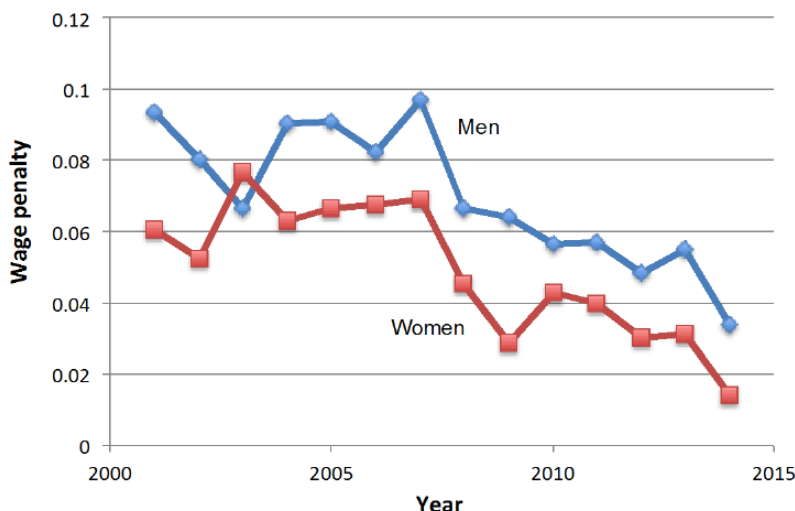


Source: Borjas (2017)

Notes [Borjas]: The wage penalty is the adjusted difference between the log hourly wage of
undocumented and legal immigrants calculated in the ACS. It is calculated using the Oaxaca
decomposition [see equations (2) and (3) in the text], adjusting for differences in age, educational
attainment, state of residence, years-since-migration, and country of birth. The Oaxaca decomposition is
conducted separately in each cross-section.

OBJECTIVES

- The first objective of this paper is to *validate the Borjas legal-status imputation algorithm*, by replicating it in a dataset which *also* collects actual (self-reported) legal immigration status. That dataset is the DOL's National Agricultural workers survey (NAWS), which covers crop farm workers only, from 1989 to the present. (Most recent public data are through 2014).

    - Under the maintained assumption that the self-reported data are correct, this allows us to calculate the error rate of the Borjas method: the share of the total population who are incorrectly classified as authorized, plus the share incorrectly classified as unauthorized.

- The second objective is to *assess the degree of attenuation bias in the estimated legal status wage penalty created by misclassification*.

    - When some (generally higher-earning) authorized workers are mistakenly classified as unauthorized while some (generally lower-earning) unauthorized workers are mistakenly classified as authorized, the estimated wage gap between them will be biased towards zero.

    - We ask: how large is this bias, and might a changing degree of bias over time explain Borjas' downward trend in the measured wage penalty?

METHODS

- We replicate the Borjas method in NAWS and compare results to actual (self-reported) legal status.

- Borjas' method looks at all foreign born, and classifies all of those with the characteristics listed in the table below as authorized; the remainder are assumed to be unauthorized.

- Receipt of public benefits for which the unauthorized are ineligible is a primary marker of legal status. But not all legal immigrants will apply for public benefits, leading to an undercount of the authorized population.

- Conversely, not all farm workers that meet Borjas' criteria are in fact authorized: for example, 11 percent of farm workers who immigrated prior to 1980 (which is enough to classify them as authorized under the Borjas criteria) report that they are unauthorized.

- Another caveat is that some of the NAWS characteristics are reported at the household level, not the individual level, which could result in misclassification for households that contain both authorized and unauthorized immigrants.

- Lastly, we run wage regressions that estimate the penalty to unauthorized legal status in NAWS, after adjusting for age, years of US farm experience, years of education, spoken English proficiency, gender, geographic region and year of immigration, using self-reported versus imputed legal status. This allows us to assess the degree of imputation bias.

**Borjas' indicators of legal immigration status and their analogue in NAWS**

| Borjas: ACS | This study: NAWS |
|---|---|
| Person is a citizen | Same |
| Person arrived before 1980 | Same |
| Person born in Cuba | Same |
| Person is currently covered by Medicaid, Medicare, or Military insurance | 1) Subject or spouse covered by "government-paid" health insurance [FY 2000 and later]; or<br>2) Medicare or Medicaid paid the majority of cost of person's last health care visit [FY1999 and later]; or<br>3) Household member received Medicaid services sometime in past two years, and household contains no children [FY1993 and later] |
| Person receives Social Security or SSI | Household member received Social Security or SSI |
| Person is a veteran, or is currently in the Armed Forces | Household member received Veteran's Pay |
| Person or spouse lives in public housing or receives rental subsidies* | Household member in low income housing |
| Person's spouse is a legal immigrant or citizen | Not yet implemented |
| Person works in the government sector; person's occupation requires some form of licensing (such as physicians, registered nurses, air traffic controllers, and lawyers) | Not applicable |
| Additional criteria, not referenced by Borjas but available in ACS** [referred to as "Expanded Criteria" in tables below] | 1) Household member received SNAP, and household contains no children<br>2) HH member received welfare payments<br>3) HH member received TANF [FY1999 and later]<br>4) HH member received Unemployment Insurance |

\* This question is not actually in the ACS, but is in the CPS which Borjas also uses.
\*\*Welfare/TANF payments are measured at the individual level in the ACS. Unemployment insurance income is likewise an individual variable in the ACS, but is measured together with "Other" income sources, including child support and alimony, not all of which necessarily require legal immigration status.
Source: Author's analysis.

FINDINGS

- Our replication of the Borjas imputation criteria correctly predicts the legal status of 74% of the NAWS sample. Using the expanded criteria, this rises to 80%. Accuracy is higher post 2001, but declining since about 2006.

- For 2001-14, the adjusted wage gap based on self-reported legal status averages 5.6 percent.

- Using imputed legal status this gap falls to 2.7 percent, implying a downward bias of 52 percent.

- This gap is larger when comparing the unauthorized to Legal Permanent Residents (80 percent) than to naturalized immigrants whose status is known by both methods (18 percent) [results not shown in figure].

**Imputed vs. self-reported immigration status: Immigrants only (cell percentages)**

### Borjas Criteria

All Fiscal Years

| | Self-reported | | |
|---|---|---|---|
| Imputed | Auth. | Unauth. | Total |
| Authorized | 24 | 6 | 31 |
| Unauthorized | 20 | 49 | 69 |
| Total | 44 | 56 | 100 |
| Percent correctly imputed: | | | 74 |

Fiscal Years 2001-2014

| | Self-reported | | |
|---|---|---|---|
| Imputed | Auth. | Unauth. | Total |
| Authorized | 20 | 7 | 28 |
| Unauthorized | 16 | 57 | 72 |
| Total | 36 | 64 | 100 |
| Percent correctly imputed: | | | 77 |

### Expanded Criteria

All Fiscal Years

| | Self-reported | | |
|---|---|---|---|
| Imputed | Auth. | Unauth. | Total |
| Authorized | 32 | 8 | 41 |
| Unauthorized | 12 | 47 | 59 |
| Total | 44 | 56 | 100 |
| Percent correctly imputed: | | | 80 |

Fiscal Years 2001-2014

| | Self-reported | | |
|---|---|---|---|
| Imputed | Auth. | Unauth. | Total |
| Authorized | 27 | 9 | 36 |
| Unauthorized | 9 | 56 | 64 |
| Total | 36 | 64 | 100 |
| Percent correctly imputed: | | | 83 |

Source: Author's analysis.

**Percent correctly imputed by year: Immigrants only**
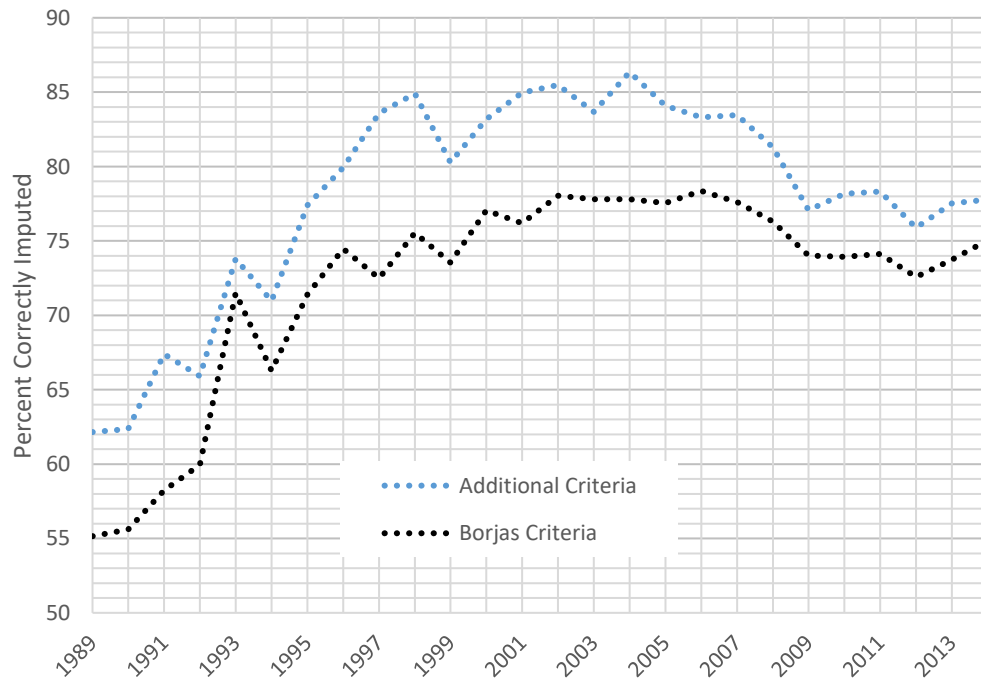


Source: Author's analysis.

**Adjusted log wage gap: unauthorized versus all authorized immigrants**



Note: dotted lines are quartic polynomial trend lines.

Source: Author's analysis.

CONCLUSIONS & DISCUSSION

- These results confirm that Borjas' estimates of the wage gap due to legal status are likely to be considerably attenuated.

- If so, the benefits of legalization to unauthorized workers may be larger than Borjas finds.

- Estimates of the effect of immigration on US born wages that cannot distinguish between authorized and unauthorized immigrants may be biased by this omission, if different subsets of the immigrant population (e.g. by geography or by age-education cells) that contribute to identification have different proportions of unauthorized workers.

- Our main finding is that the *level* of Borjas' estimates is likely biased downwards by imputation errors, perhaps by as much as 50 percent.

- We find some evidence of a downward trend in the penalty since 2007, whether using the imputed or the self-reported data, but it is much less clear than in the Borjas dataset.

REFERENCES

Borjas, George J. (2017) "The Earnings of Undocumented Immigrants." NBER Working Paper #23236.